Twelve Publications

Jin-Yi Cai

1. Shuai Shao and Jin-Yi Cai: A Dichotomy for Real Boolean Holant Problems. In *Proc. of the 61st IEEE Annual Symposium on Foundations of Computer Science* (FOCS), 2020, pp. 1091-1102, doi: 10.1109/FOCS46700.2020.00105.
Full version at https://arxiv.org/pdf/2005.07906.pdf (89 pages).

2. Jin-Yi Cai and Zhiguo Fu: Holographic Algorithm with Matchgates Is Universal for Planar #CSP Over Boolean Domain. In *the 49th ACM Symposium on Theory of Computing (STOC)* 2017: 842-855. SIAM Journal on Computing https://doi.org/10.1137/17M1131672 (102 pages).

3. Jin-Yi Cai, Zhiguo Fu, Heng Guo and Tyson Williams. A Holant Dichotomy: Is the FKT Algorithm Universal? In *Proc. 56th IEEE Symposium on Foundations of Computer Science (FOCS)* 2015, pp. 1259–1276. To appear in Theory of Computing Systems (TOCS). Full version at http://arxiv.org/abs/1505.02993 (128 pages).

4. Jin-Yi Cai, Heng Guo and Tyson Williams. The Complexity of Counting Edge Colorings and a Dichotomy for Some Higher Domain Holant Problems. In *Proc. 55th IEEE Symposium on Foundations of Computer Science (FOCS)* 2014. pp. 601–610. *Research in the Mathematical Sciences*, (2016) 3:18 DOI 10.1186/s40687-016-0067-8 (77 pages).

5. Jin-Yi Cai and Xi Chen. Complexity of Counting CSP with Complex Weights.
J. ACM 64(3): 19:1-19:39 (2017).

6. Jin-Yi Cai, Xi Chen and Pinyan Lu. Graph Homomorphisms with Complex Values: A Dichotomy Theorem. *SIAM Journal on Computing*, (2013) 42(3), 924-1029 (106 pages).

7. Jin-Yi Cai, Xi Chen and Dong Li. Quadratic Lower Bound for Permanent vs. Determinant in any Characteristic. *Computational Complexity* 19(1): 37-56 (2010). A preliminary version appeared as: A quadratic lower bound for the permanent and determinant problem over any characteristic $\neq 2$. *The 40th Annual ACM Symposium on the Theory of Computing* (STOC) 2008. 491-498.

8. Jin-Yi Cai and Pinyan Lu. Holographic Algorithms: From Art to Science. *The Journal of Computer and System Sciences* 77(1): 41-61 (2011). A preliminary version appeared in *The 39th Annual ACM Symposium on the Theory of Computing* (STOC) 2007, 401–410.

9. Jin-Yi Cai. $S_2^p \subseteq \mathrm{ZPP}^{\mathrm{NP}}$. *The Journal of Computer and System Sciences* 73(1): 25-35 (2007). A preliminary version appeared in *The 42nd IEEE Symposium on Foundations of Computer Science* (FOCS), 2001, 620–628.

10. Jin-Yi Cai and D. Sivakumar. Sparse Hard sets for P: Resolution of a Conjecture of Hartmanis. *The Journal of Computer and System Sciences*, (special issue), Vol. **58**, 280–296 (1999). A preliminary version appeared in *The 36th Annual Symposium on Foundations of Computer Science* (FOCS), 1995, 362–371.

11. Jin-Yi Cai, Martin Fürer and Neil Immerman. An Optimal Lower Bound on the Number of Variables for Graph Identification. *Combinatorica* **12** (4) (1992), 389–410. Preliminary version in *The 30th Annual Symposium on Foundations of Computer Science*, (FOCS), 1989, 612–617.

12. Jin-Yi Cai. With Probability One, a Random Oracle Separates PSPACE from the Polynomial-time Hierarchy. *The Journal of Computer and System Sciences*, (special issue), Vol. **38**, No. 1, (1989) 68–85. A preliminary version appeared in *The 18th ACM Symposium on the Theory of Computing* (STOC), 1986, 21–29.

Paper 1

A Dichotomy for Real Boolean Holant Problems

Shuai Shao*
sh@cs.wisc.edu

Jin-Yi Cai*
jyc@cs.wisc.edu

Abstract

We prove a complexity dichotomy for Holant problems on the boolean domain with arbitrary sets of real-valued constraint functions. These constraint functions need not be symmetric nor do we assume any auxiliary functions as in previous results. It is proved that for every set $\mathcal{F}$ of real-valued constraint functions, Holant($\mathcal{F}$) is either P-time computable or #P-hard. The classification has an explicit criterion. This is the culmination of much research on this problem, and it uses previous results and techniques from many researchers. Some particularly intriguing concrete functions f_6, f_8 and their associated families with extraordinary closure properties related to Bell states in quantum information theory play an important role in this proof.

*Department of Computer Sciences, University of Wisconsin-Madison. Supported by NSF CCF-1714275.

Prologue

The young knight Siegfried and his tutor Jeyoda set out for their life journey together. Their aim is to pacify the real land of Holandica, to bring order and unity.

In the past decade, brave knights have battled innumerable demons and creatures, and have conquered the more familiar portion of Holandica, called Holandica Symmetrica. Along the way they have also been victorious by channeling the power of various deities known as Unary Oracles. In the past few years this brotherhood of the intrepid have also gained great power from the beneficent god Orieneuler and enhanced their skills in a more protected world ruled by Count Seaspie.

"But prepared we must be," Jeyoda reminds Siegfried, "arduous, our journey will become." As the real land of Holandica is teeming with unknowns, who knows what wild beasts and creatures they may encounter. Siegfried nods, but in his heart he is confident that his valor and power will be equal to the challenge.

They have recently discovered a treasure sword. This is their second gift from the Cathedral Orthogonia, more splendid and more powerful than the first. In their initial encounters with the minion creatures in their journey, the second sword from Cathedral Orthogonia proved to be invincible.

These initial victories laid the foundation for their journey, but also a cautious optimism sets in. Perhaps with their new powerful sword in hand, final victory will not be that far away.

Just as they savor these initial victories, things start to change. As they enter the Kingdom of Degree-Six everything becomes strange. Subliminally they feel the presence of a cunning enemy hiding in the darkness. Gradually they are convinced that this enemy possesses a special power that eludes the ongoing campaign, and in particular their magic sword. After a series of difficult and protracted battles with many twists and turns, their nemesis, the Lord of Intransigence slowly reveals his face. The Lord of Intransigence has a suit of magic armor, called the Bell Spell, that hides and protects him so well that the sword of Cathedral Orthogonia cannot touch him.

Siegfried and Jeyoda know that in order to conquer the Lord of Intransigence, they need all the skills and wisdom they have. Although the Lord of Intransigence has a strong armor, he has a weakness. The armor is maintained by four little elfs called the Bell Binaries. The next battle is tough and long. Siegfried and Jeyoda hit upon the idea of convincing the Bell Binaries to stage a mutiny. With his four little elfs turning against him, his armor loses its magic, and the Kingdom of Six-degree is conquered. In the aftermath of this victory, Siegfried and Jeyoda also collect some valuable treasures that will come in handy in their next campaign.

After defeating the Lord of Intransigence, Siegfried and Jeyoda enter the Land of Degree-Eight. Now they are very careful. After meticulous reconnaissance, they finally identify the most fearsome enemy, the Queen of the Night. Taking a page from their battle with the Lord of Intransigence they look for opportunities to gain help from within the enemy camp. However, the Queen of the Night has the strongest protective coat called the Strong Bell Spell. This time there is no way to summon help from within the Queen's own camp. In fact, her protective armor is so strong that any encounter with Siegfried and Jeyoda's sword makes her magically disappear in a puff of white smoke.

But, everyone has a weakness. For the Queen, her vanishing act also brings the downfall. After plotting the strategy for a long time, Siegfried and Jeyoda use a magical potion to create from nothing the helpers needed to defeat the Queen.

Buoyed by their victory, they summon their last strength to secure the Land of Degree-Eight and beyond. Finally they bring complete order to the entire real land of Holandica. At their celebratory banquet, they want to share the laurels with Knight Ming and Knight Fu who provided invaluable assistance in their journey; but the two brave and generous knights have retreated to their Philosopher's Temple and are nowhere to be found.

Contents

1 Introduction

Counting problems arise in many different fields, e.g., statistical physics, economics and machine learning. In order to study the complexity of counting problems, several natural frameworks have been proposed. Two well studied frameworks are counting constraint satisfaction problems (#CSP) [8, 22, 6, 12, 10] and counting graph homomorphisms (#GH) [21, 7, 24, 11] which is a special case of #CSP. These frameworks are expressive enough so that they can express many natural counting problems but also specific enough so that complete complexity classifications can be established.

Holant problems are a more expressive framework which generalizes #CSP and #GH. It is a broad class of sum-of-products computation. Unlike #CSP and #GH for which full complexity dichotomies have been established, the understanding of Holant problems, even restricted to the Boolean domain, is still limited. In this paper, we establish the first Holant dichotomy on the Boolean domain with arbitrary real-valued constraint functions. These constraint functions need not be symmetric nor do we assume any auxiliary functions (as in previous results).

A Holant problem on the Boolean domain is parameterized by a set of constraint functions, also called signatures; such a signature maps $\{0,1\}^n \to \mathbb{C}$ for some $n > 0$. Let $\mathcal{F}$ be any fixed set of signatures. A signature grid $\Omega = (G, \pi)$ over $\mathcal{F}$ is a tuple, where $G = (V, E)$ is a graph without isolated vertices, π labels each $v \in V$ with a signature $f_v \in \mathcal{F}$ of arity $\deg(v)$, and labels the incident edges $E(v)$ at v with input variables of f_v. We consider all 0-1 edge assignments σ, and each gives an evaluation $\prod_{v \in V} f_v(\sigma|_{E(v)})$, where $\sigma|_{E(v)}$ denotes the restriction of σ to $E(v)$.

Definition 1.1 (Holant problems)**.** *The input to the problem* $\text{Holant}(\mathcal{F})$ *is a signature grid* $\Omega = (G, \pi)$ *over* $\mathcal{F}$*. The output is the partition function*

$$\text{Holant}(\Omega) = \sum_{\sigma : E(G) \to \{0,1\}} \prod_{v \in V(G)} f_v(\sigma|_{E_{(v)}}).$$

Bipartite Holant problems $\text{Holant}(\mathcal{F} \mid \mathcal{G})$ *are Holant problems over bipartite graphs* $H = (U, V, E)$*, where each vertex in* U *or* V *is labeled by a signature in* $\mathcal{F}$ *or* $\mathcal{G}$ *respectively. When* $\{f\}$ *is a singleton set, we write* $\text{Holant}(\{f\})$ *as* $\text{Holant}(f)$ *and* $\text{Holant}(\{f\} \cup \mathcal{F})$ *as* $\text{Holant}(f, \mathcal{F})$*.*

Weighted #CSP is a special class of Holant problems. So are all weighted #GH. Other problems expressible as Holant problems include counting matchings and perfect matchings [28], counting weighted Eulerian orientations (#EO problems) [26, 14], computing the partition functions of six-vertex models [27, 16] and eight-vertex models [3, 13], and a host of other, if not almost all, vertex models from statistical physics [4]. It is proved that counting perfect matchings cannot be expressed by #GH [23, 17]. Thus, Holant problems are provably more expressive.

Progress has been made in the complexity classification of Holant problems. When all signatures are restricted to be *symmetric*, a full dichotomy is proved [18]. When asymmetric signatures are allowed, some dichotomies are proved for special families of Holant problems by assuming that certain auxiliary signatures are available, e.g., Holant^*, Holant^+ and Holant^c [19, 1, 20, 2]. Without assuming auxiliary signatures a Holant dichotomy is established for non-negative real-valued signatures [25], and for all real-valued signatures where a signature of odd arity is present [15]. In this paper, we prove a full complexity dichotomy for Holant problems with real values.

Theorem 1.2. *Let* $\mathcal{F}$ *be a set of real-valued signatures. If* $\mathcal{F}$ *satisfies the tractability condition* (T) *in Theorem 2.22, then* $\text{Holant}(\mathcal{F})$ *is polynomial-time computable; otherwise,* $\text{Holant}(\mathcal{F})$ *is #P-hard.*

This theorem is the culmination of a large part of previous research on dichotomy theorems on Holant problems, and it uses much of the previously established results and techniques. However, as it turned out, the journey to this theorem has been arduous. The overall plan of the proof is by induction on arities of signatures in $\mathcal{F}$. Since a dichotomy is proved when $\mathcal{F}$ contains a signature of odd arity, we only need to consider signatures of even arity. For signatures of small arity 2 or 4 (base cases) and large arity at least 10, we given an induction proof based on results of #CSP, #EO problems and eight-vertex models. However, two signatures f_6 and f_8 (and their associated families) of arity 6 and 8, are discovered which have extraordinary closure properties; we call them Bell properties [15]. These amazing signatures are wholly unexpected, and their existence presented a formidable obstacle to the induction proof.

All four binary Bell signatures (related to Bell states [5] in quantum information theory) are realizable from f_6 by gadget construction. We introduce Holant^b problems where the four binary Bell signatures are available. This is specifically to handle the signature f_6. We prove a #P-hardness result for Holant$^b(f_6, \mathcal{F})$. In this proof, we find other miraculous signatures with special structures such that all signatures realized from them by merging gadgets are affine signatures, while themselves are not affine signatures. In order to handle the signature f_8, we introduce Holant problems with limited appearance, where some signatures are only allowed to appear a limited number of times in all instances. We turn the obstacle of the closure property of f_8 in our favor to prove non-constructively a P-time reduction from Holant$^b(f_8, \mathcal{F})$ to Holant$(f_8, \mathcal{F})$. In fact, it is provable that except $=_2$, the other three binary Bell signatures are *not* realizable from f_8 by gadget construction. However, we show that we can realize, in the sense of a non-constructive complexity reduction, the desired binary Bell signatures which appear an unlimited number of times. This utilizes the framework where these signatures occur only a limited number of times. Then, we give a #P-hardness result for Holant$^b(f_8, \mathcal{F})$ similar to Holant$^b(f_6, \mathcal{F})$.

2 Preliminaries

2.1 Definitions and notations

Let f be a complex-valued signature. If $\overline{f(\alpha)} = f(\overline{\alpha})$ for all α where $\overline{f(\alpha)}$ denotes the complex conjugation of $f(\alpha)$ and $\overline{\alpha}$ denotes the bit-wise complement of α, we say f satisfies *Arrow Reversal Symmetry* (ARS). We may also use f^α to denote $f(\alpha)$. We use $\mathrm{wt}(\alpha)$ to denote the Hamming weight of α. The support $\mathscr{S}(f)$ of a signature f is $\{\alpha \in \mathbb{Z}_2^n \mid f(\alpha) \neq 0\}$. We say f has support of size k if $|\mathscr{S}(f)| = k$. If $\mathscr{S}(f) = \emptyset$, i.e., f is identically 0, we say f is a zero signature and denote it by $f \equiv 0$. Otherwise, f is a nonzero signature. Let $\mathscr{E}_n = \{\alpha \in \mathbb{Z}_2^n \mid \mathrm{wt}(\alpha) \text{ is even}\}$, and $\mathscr{O}_n = \{\alpha \in \mathbb{Z}_2^n \mid \mathrm{wt}(\alpha) \text{ is odd}\}$. A signature f of arity n has even or odd parity if $\mathscr{S}(f) \subseteq \mathscr{E}_n$ or $\mathscr{S}(f) \subseteq \mathscr{O}_n$ respectively. In both cases, we say that f has parity. Let $\mathscr{H}_{2n} = \{\alpha \in \mathbb{Z}_2^{2n} \mid \mathrm{wt}(\alpha) = n\}$. A signature f of arity $2n$ has half-weighted support if $\mathscr{S}(f) \subseteq \mathscr{H}_{2n}$. We call such a signature an *Eularian orientaion* (EO) signature. For $\alpha \in \mathbb{Z}_2^n$ and $1 \leqslant i \leqslant n$, we use α_i to denote the value of α on bit i.

Counting constraint satisfaction problems (#CSP) can be expressed as Holant problems. We use $=_n$ to denote the EQUALITY signature of arity n, which takes value 1 on the all-0 and all-1 inputs and 0 elsewhere. (We denote the n-bits all-0 and all-1 strings by $\vec{0}^n$ and $\vec{1}^n$ respectively. We may omit the superscript n when it is clear from the context.) Let $\mathcal{EQ} = \{=_1, =_2, \ldots, =_n, \ldots\}$ denote the set of all EQUALITY signatures.

Lemma 2.1 ([9]). $\#\mathrm{CSP}(\mathcal{F}) \equiv_T \mathrm{Holant}(\mathcal{EQ} \mid \mathcal{F})$.

We use $\neq_2$ to denote the binary DISEQUALITY signature with truth table $(0,1,1,0)$. We generalize this notion to signatures of higher arities. A signature f of arity $2n$ is called a DISEQUALITY signature of arity $2n$, denoted by $\neq_{2n}$, if $f = 1$ when $(x_1 \neq x_2) \wedge \ldots \wedge (x_{2n-1} \neq x_{2n})$, and 0 otherwise. By permuting its variables the DISEQUALITY signature of arity $2n$ also defines $(2n-1)(2n-3)\cdots 1$ functions which we also call DISEQUALITY signatures. These signatures are equivalent for the complexity of Holant problems; once we have one we have them all. Let $\mathcal{DEQ} = \{\neq_2, \neq_4, \ldots, \neq_{2n}, \ldots\}$ denote the set of all DISEQUALITY signatures.

We use $=_2^-$ to denote the binary signature $(1,0,0,-1)$ and $\neq_2^-$ to denote the binary signature $(0,1,-1,0)$. We may also write $=_2$ as $=_2^+$ and $\neq_2$ as $\neq_2^+$. Let $\mathcal{B} = \{=_2^+, =_2^-, \neq_2^+, \neq_2^-\}$. We call them Bell signatures which correspond to Bell states $|\Phi^+\rangle = |00\rangle + |11\rangle$, $|\Phi^-\rangle = |00\rangle - |11\rangle$, $|\Psi^+\rangle = |01\rangle + |10\rangle$ and $|\Psi^-\rangle = |01\rangle - |10\rangle$ in quantum information science [5].

A signature f of arity $n \geqslant 2$ can be expressed as a $2^k \times 2^{n-k}$ matrix $M_{S_k}(f)$ where S_k is a set of k many variables among all n variables of f. The matrix $M_{S_k}(f)$ lists all 2^n many entries of f with the assignments of variables in S_k[1] listed in lexicographic order (from $\vec{0}^k$ to $\vec{1}^k$) as row index and the assignments of the other $n-k$ many variables in lexicographic order as column index. In particular, f can be expressed as a $2 \times 2^{n-1}$ matrix $M_i(f)$ which lists the 2^n entries of f with the assignments of variable x_i as row index (from $x_i = 0$ to $x_i = 1$) and the assignments of the other $n-1$ variables in lexicographic order as column index. Then,

$$M_i(f) = \begin{bmatrix} f^{0,00\ldots0} & f^{0,00\ldots1} & \ldots & f^{0,11\ldots1} \\ f^{1,00\ldots0} & f^{1,00\ldots1} & \ldots & f^{1,11\ldots1} \end{bmatrix} = \begin{bmatrix} \mathbf{f}_i^0 \\ \mathbf{f}_i^1 \end{bmatrix},$$

where $\mathbf{f}_i^a$ denotes the row vector indexed by $x_i = a$ in $M_i(f)$. Similarly, f can also be expressed as a $4 \times 2^{n-2}$ matrix with the assignments of two variables x_i and x_j as row index. Then,

$$M_{ij}(f) = \begin{bmatrix} f^{00,00\ldots0} & f^{00,00\ldots1} & \ldots & f^{00,11\ldots1} \\ f^{01,00\ldots0} & f^{01,00\ldots1} & \ldots & f^{01,11\ldots1} \\ f^{10,00\ldots0} & f^{10,00\ldots1} & \ldots & f^{10,11\ldots1} \\ f^{11,00\ldots0} & f^{11,00\ldots1} & \ldots & f^{11,11\ldots1} \end{bmatrix} = \begin{bmatrix} \mathbf{f}_{ij}^{00} \\ \mathbf{f}_{ij}^{01} \\ \mathbf{f}_{ij}^{10} \\ \mathbf{f}_{ij}^{11} \end{bmatrix},$$

where $\mathbf{f}_{ij}^{ab}$ denotes the row vector indexed by $(x_i, x_j) = (a, b)$ in $M_{ij}(f)$. For $=_2$, it has the 2-by-2 signature matrix $M(=_2) = I_2 = \left[\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right]$. For $\neq_2$, $M(\neq_2) = N_2 = \left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right]$.

2.2 Holographic transformation

To introduce the idea of holographic transformation, it is convenient to consider bipartite graphs. For a general graph, we can always transform it into a bipartite graph while preserving the Holant value, as follows. For each edge in the graph, we replace it by a path of length two. (This operation is called the *2-stretch* of the graph and yields the edge-vertex incidence graph.) Each new vertex is assigned the binary EQUALITY signature $=_2$. Thus, we have Holant$(=_2 | \mathcal{F}) \equiv_T$ Holant$(\mathcal{F})$.

For an invertible 2-by-2 matrix $T \in \mathbf{GL}_2(\mathbb{C})$ and a signature f of arity n, written as a column vector (covariant tensor) $f \in \mathbb{C}^{2^n}$, we denote by $Tf = T^{\otimes n} f$ the transformed signature. For a signature set $\mathcal{F}$, define $T\mathcal{F} = \{Tf \mid f \in \mathcal{F}\}$ the set of transformed signatures. For signatures written as row vectors (contravariant tensors) we define fT^{-1} and $\mathcal{F}T^{-1}$ similarly. Whenever we

[1] Given a set of variables, without other specification, we always list them in the cardinal order i.e., from variables with the smallest index to the largest index.

write Tf or $T\mathcal{F}$, we view the signatures as column vectors; similarly for fT^{-1} or $\mathcal{F}T^{-1}$ as row vectors. We can also represent Tf as the matrix $M_{S_k}(Tf)$ with the assignments of variables in S_k as row index and the assignments of the other $n-k$ variables as column index. Then, we have $M_{S_k}(Tf) = T^{\otimes k} M_{S_k}(f)(T^{\mathtt{T}})^{\otimes n-k}$. Similarly, $M_{S_k}(fT^{-1}) = (T^{-1^{\mathtt{T}}})^{\otimes k} M_{S_k}(f)(T^{-1})^{\otimes n-k}$.

Let $T \in \mathbf{GL}_2(\mathbb{C})$. The holographic transformation defined by T is the following operation: given a signature grid $\Omega = (H, \pi)$ of Holant$(\mathcal{F} \mid \mathcal{G})$, for the same bipartite graph H, we get a new signature grid $\Omega' = (H, \pi')$ of Holant$(\mathcal{F}T^{-1} \mid T\mathcal{G})$ by replacing each signature in $\mathcal{F}$ or $\mathcal{G}$ with the corresponding signature in $\mathcal{F}T^{-1}$ or $T\mathcal{G}$.

Theorem 2.2 ([29]). *For every $T \in \mathbf{GL}_2(\mathbb{C})$,* Holant$(\mathcal{F} \mid \mathcal{G}) \equiv_T$ Holant$(\mathcal{F}T^{-1} \mid T\mathcal{G})$.

Therefore, a holographic transformation does not change the complexity of the Holant problem in the bipartite setting. Let $\mathbf{O}_2(\mathbb{R}) \subseteq \mathbb{R}^{2\times 2}$ be the set of all 2-by-2 real orthogonal matrices. We denote $\mathbf{O}_2(\mathbb{R})$ by $\mathbf{O}_2$. For all $Q \in \mathbf{O}_2$, since $(=_2)Q^{-1} = (=_2)$, Holant$(=_2 \mid \mathcal{F}) \equiv_T$ Holant$(=_2 \mid Q\mathcal{F})$.

A particular holographic transformation that will be commonly used in this paper is the transformation defined by $Z^{-1} = \frac{1}{\sqrt{2}} \left[\begin{smallmatrix} 1 & -\mathrm{i} \\ 1 & \mathrm{i} \end{smallmatrix}\right]$. Note that $(=_2)Z = (\neq_2)$. Thus, Holant$(=_2 \mid \mathcal{F}) \equiv_T$ Holant$(\neq_2 \mid Z^{-1}\mathcal{F})$. We denote $Z^{-1}\mathcal{F}$ by $\widehat{\mathcal{F}}$ and $Z^{-1}f$ by $\widehat{f}$. It is known that f and $\widehat{f}$ have the following relation.

Lemma 2.3 ([14]). *A (complex-valued) signature f is a real-valued signature iff $\widehat{f}$ satisfies* ARS.

We say a real-valued binary signature $f(x_1, x_2)$ is orthogonal if $M_1(f)M_1^{\mathtt{T}}(f) = \lambda I_2$ for some real $\lambda > 0$. Since $M_2(f) = M_1^{\mathtt{T}}(f)$, $M_1(f)M_1^{\mathtt{T}}(f) = \lambda I_2$ iff $M_2(f)M_2^{\mathtt{T}}(f) = \lambda I_2$. The following fact is easy to check.

Lemma 2.4. *A binary signature f is orthogonal or a zero signature iff $\widehat{f}$ has parity and* ARS.

Proof. Consider $M_1(f)$ and $M_1(\widehat{f}) = M_1(Z^{-1}f) = Z^{-1}M_1(f)(Z^{-1})^{\mathtt{T}}$. Then, $M_1(f) = \left[\begin{smallmatrix} a & b \\ -b & a \end{smallmatrix}\right]$ iff $M_1(\widehat{f}) = \left[\begin{smallmatrix} 0 & a+b\mathrm{i} \\ a-b\mathrm{i} & 0 \end{smallmatrix}\right]$, and $M_1(f) = \left[\begin{smallmatrix} a & b \\ b & -a \end{smallmatrix}\right]$ iff $M_1(\widehat{f}) = \left[\begin{smallmatrix} a-b\mathrm{i} & 0 \\ 0 & a+b\mathrm{i} \end{smallmatrix}\right]$. Also, $f \equiv 0$ iff $\widehat{f} \equiv 0$ which also has parity. □

Let $\mathcal{O}$ denote the set of all binary orthogonal signatures and the binary zero signature. Then, $\widehat{\mathcal{O}} = Z^{-1}\mathcal{O}$ is the set of all binary signatures with ARS and parity (including the binary zero signature). Note that $\mathcal{B} \subseteq \mathcal{O}$ and $\widehat{\mathcal{B}} \subseteq \widehat{\mathcal{O}}$. Here the transformed set

$$\widehat{\mathcal{B}} = \left\{\widehat{=_2^+}, \widehat{=_2^-}, \widehat{\neq_2^+}, \widehat{\neq_2^-}\right\} = \{\neq_2, =_2, (-\mathrm{i})\cdot =_2^-, \mathrm{i}\cdot \neq_2^-\}.$$

For every $Q \in \mathbf{O}_2$, let $\widehat{Q} = Z^{-1}QZ$. Then, $\widehat{Q}\widehat{\mathcal{F}} = (Z^{-1}QZ)(Z^{-1}\mathcal{F}) = Z^{-1}(Q\mathcal{F}) = \widehat{Q\mathcal{F}}$. Thus,

$$\text{Holant}(\neq_2 \mid \widehat{\mathcal{F}}) \equiv_T \text{Holant}(=_2 \mid \mathcal{F}) \equiv_T \text{Holant}(=_2 \mid Q\mathcal{F}) \equiv_T \text{Holant}(\neq_2 \mid \widehat{Q}\widehat{\mathcal{F}}).$$

Let $\widehat{\mathbf{O}_2} = \{\widehat{Q} = Z^{-1}QZ \mid Q \in \mathbf{O}_2\}$. Then, $\widehat{\mathbf{O}_2} = \{\left[\begin{smallmatrix} \alpha & 0 \\ 0 & \bar{\alpha} \end{smallmatrix}\right], \left[\begin{smallmatrix} 0 & \alpha \\ \bar{\alpha} & 0 \end{smallmatrix}\right] \mid \alpha \in \mathbb{C}, |\alpha| = 1\}$. Note that the notation $\widehat{\cdot}$ on a matrix $Q \in \mathbf{O}_2$ is *not* the same as the notation $\widehat{\cdot}$ on a signature $f \in \mathcal{O}$. Suppose that $M_1(f) = Q \in \mathbf{O}_2$. Since $Q = Z\widehat{Q}Z^{-1}$ and $Z^{-1}(Z^{-1})^{\mathtt{T}} = N_2$,

$$M_1(\widehat{f}) = Z^{-1}Q(Z^{-1})^{\mathtt{T}} = Z^{-1}(Z\widehat{Q}Z^{-1})(Z^{-1})^{\mathtt{T}} = \widehat{Q}N_2 \neq \widehat{Q}.$$

2.3 Signature factorization

Recall that by our definition, every (complex valued) signature has arity at least one. A nonzero signature g *divides* f denoted by $g \mid f$, if there is a signature h such that $f = g \otimes h$ (with possibly a permutation of variables) or there is a constant λ such that $f = \lambda \cdot g$. In the latter case, if $\lambda \neq 0$, then we also have $f \mid g$ since $g = \frac{1}{\lambda} \cdot f$. For nonzero signatures, if both $g \mid f$ and $f \mid g$, then they are nonzero constant multiples of each other, and we say g is an *associate* of f, denoted by $g \sim f$. In terms of this division relation, the notions of *irreducible* signatures and *prime* signatures have been defined. They are proved equivalent and thus, the *unique prime factorization* (UPF) of signatures is established [14].

A nonzero signature f is irreducible if there are no signatures g and h such that $f = g \otimes h$. A nonzero signature f is a prime signature if $f \mid g \otimes h$ implies that $f \mid g$ or $f \mid h$. These notions are equivalent. We say a signature f is reducible if $f = g \otimes h$, for some signatures g and h. All zero signatures of arity greater than 1 are reducible. A prime factorization of a signature f is $f = g_1 \otimes \ldots \otimes g_k$ up to a permutation of variables, where each g_i is irreducible.

Lemma 2.5 (Unique prime factorization [14]). *Every nonzero signature f has a prime factorization. If f has prime factorizations $f = g_1 \otimes \ldots \otimes g_k$ and $f = h_1 \otimes \ldots \otimes h_\ell$, both up to a permutation of variables, then $k = \ell$ and after reordering the factors we have $g_i \sim h_i$ for all i.*

Lemma 2.6 ([14]). *let f be a real-valued reducible signature, then there exists a factorization $f = g \otimes h$ such that g and h are both real-valued signatures.*

Equivalently, let $\widehat{f}$ be a reducible signature satisfying ARS*, then there exists a factorization $\widehat{f} = \widehat{g} \otimes \widehat{h}$ such that $\widehat{g}$ and $\widehat{h}$ both satisfy* ARS*.*

In the following, when we say that a real-valued reducible signature f has a factorization $g \otimes h$, we always assume that g and h are real-valued. Equivalently, when we say a signature $\widehat{f}$ satisfying ARS has a factorization $\widehat{g} \otimes \widehat{h}$, we always assume that $\widehat{g}$ and $\widehat{h}$ satisfy ARS.

For a signature set $\mathcal{F}$, we use $\mathcal{F}^{\otimes k}$ $(k \geqslant 1)$ to denote the set $\{\lambda \bigotimes_{i=1}^{k} f_i \mid \lambda \in \mathbb{R}\backslash\{0\}, f_i \in \mathcal{F}\}$. Here, λ denotes a normalization scalar. In this paper, we only consider the normalization by nonzero real constants. Note that $\mathcal{F}^{\otimes 1}$ contains all signatures obtained from $\mathcal{F}$ by normalization. We use $\mathcal{F}^{\otimes}$ to denote $\bigcup_{k=1}^{\infty} \mathcal{F}^{\otimes k}$.

If a vertex v in a signature grid is labeled by a reducible signature $f = g \otimes h$, we can replace the vertex v by two vertices v_1 and v_2 and label v_1 with g and v_2 with h, respectively. The incident edges of v become incident edges of v_1 and v_2 respectively according to the partition of variables of f in the tensor product of g and h. This does not change the Holant value. On the other hand, Lin and Wang proved that, from a real-valued reducible signature $f = g \otimes h \not\equiv 0$ we can freely replace f by g and h while preserving the complexity of a Holant problem.

Lemma 2.7 ([25]). *If a nonzero real-valued signature f has a real factorization $g \otimes h$, then*

$$\text{Holant}(g, h, \mathcal{F}) \equiv_T \text{Holant}(f, \mathcal{F}) \text{ and } \text{Holant}(\neq_2 | \widehat{g}, \widehat{h}, \widehat{F}) \equiv_T \text{Holant}(\neq_2 | \widehat{f}, \widehat{F})$$

for any signature set $\mathcal{F}$ $(\widehat{\mathcal{F}})$. We say g $(\widehat{g})$ and h $(\widehat{h})$ are realizable from f $(\widehat{f})$ by factorization.

2.4 Gadget construction

One basic tool used throughout the paper is gadget construction. An $\mathcal{F}$-gate is similar to a signature grid (G, π) for Holant$(\mathcal{F})$ except that $G = (V, E, D)$ is a graph with internal edges E and dangling

edges D. The dangling edges D define input variables for the $\mathcal{F}$-gate. We denote the regular edges in E by $1, 2, \ldots, m$ and the dangling edges in D by $m+1, \ldots, m+n$. Then the $\mathcal{F}$-gate defines a function f

$$f(y_1, \ldots, y_n) = \sum_{\sigma: E \to \{0,1\}} \prod_{v \in V} f_v(\hat{\sigma} \mid_{E(v)})$$

where $(y_1, \ldots, y_n) \in \{0,1\}^n$ is an assignment on the dangling edges, $\hat{\sigma}$ is the extension of σ on E by the assignment $(y_1, \ldots, y_m)$, and f_v is the signature assigned at each vertex $v \in V$. This function f is called the signature of the $\mathcal{F}$-gate. There may be no internal edges in an $\mathcal{F}$-gate at all. In this case, f is simply a tensor product of these signatures f_v, i.e., $f = \bigotimes_{v \in V} f_v$ (with possibly a permutation of variables). We say a signature f is *realizable* from a signature set $\mathcal{F}$ by gadget construction if f is the signature of an $\mathcal{F}$-gate. If f is realizable from a set $\mathcal{F}$, then we can freely add f into $\mathcal{F}$ while preserving the complexity (Lemma 1.3 in [9]).

Lemma 2.8 ([9]). *If f is realizable from a set $\mathcal{F}$, then* $\text{Holant}(f, \mathcal{F}) \equiv_T \text{Holant}(\mathcal{F})$.

Note that, if we view $\text{Holant}(=_2 \mid \mathcal{F})$ as the edge-vertex incidence graph form of $\text{Holant}(\mathcal{F})$, then it is equivalent to label every edge by $=_2$; similarly in the setting of $\text{Holant}(\neq_2 \mid \widehat{\mathcal{F}})$, every edge is labeled by $\neq_2$.

Lemma 2.9. *If f is realizable from a real-valued signature set $\mathcal{F}$ (in the setting of* $\text{Holant}(=_2 \mid \mathcal{F})$*), then f is also real-valued. Equivalently, if $\widehat{f}$ is realizable from a signature set $\widehat{\mathcal{F}}$ satisfying* ARS *(in the setting of* $\text{Holant}(\neq_2 \mid \widehat{\mathcal{F}})$*), then $\widehat{f}$ also satisfies* ARS.

A basic gadget construction is *merging*. In the setting of $\text{Holant}(=_2 \mid \mathcal{F})$, given a signature $f \in \mathcal{F}$ of arity n, we can connect two variables x_i and x_j of f using $=_2$, and this operation gives a signature of arity $n-2$. We use $\partial_{ij} f$ or $\partial^{+}_{ij} f$ to denote this signature and $\partial_{ij} f = f^{00}_{ij} + f^{11}_{ij}$, where f^{ab}_{ij}[1] denotes the signature obtained by setting $(x_i, x_j) = (a, b) \in \{0,1\}^2$. While in the setting of $\text{Holant}(\neq_2 \mid \widehat{\mathcal{F}})$, the above merging gadget is equivalent to connecting two variables x_i and x_j of $\widehat{f}$ using $\neq_2$. We denote the resulting signature by $\widehat{\partial}_{ij} \widehat{f}$ or $\widehat{\partial}^{+}_{ij} \widehat{f}$, and we have $\widehat{\partial_{ij} f} = \widehat{\partial}_{ij} \widehat{f} = \widehat{f}^{01}_{ij} + \widehat{f}^{10}_{ij}$. If $\neq_2$ is available (i.e., it either belongs to or can be realized from $\mathcal{F}$) in $\text{Holant}(=_2 \mid \mathcal{F})$, we can also connect two variables x_i and x_j of f using $\neq_2$. We denote the resulting signature by $\partial^{\widehat{+}}_{ij} f$. The merging gadget $\widehat{\partial}^{+}_{ij}$ is the same as $\partial^{\widehat{+}}_{ij}$, we use different notations to distinguish whether this gadget is used in the setting of $\text{Holant}(=_2 \mid \mathcal{F})$ or $\text{Holant}(\neq \mid \widehat{\mathcal{F}})$. Also, if $=^{-}_2$ and $\neq^{-}_2$ are available in $\text{Holant}(=_2 \mid \mathcal{F})$, then we can construct $\partial^{-}_{ij} f$ and $\partial^{\widehat{-}}_{ij} f$ by connecting x_i and x_j using $=^{-}_2$ and $\neq^{-}_2$ respectively. We also call ∂^{-}_{ij} and $\partial^{\widehat{-}}_{ij}$ merging gadgets. Without other specification, by default a merging gadget refers to ∂_{ij} in the setting of $\text{Holant}(=_2 \mid \mathcal{F})$. Similarly by default a merging gadget refers to $\widehat{\partial}_{ij}$ in the setting of $\text{Holant}(\neq_2 \mid \widehat{\mathcal{F}})$.

The following lemma gives a relation between a signature $\widehat{f}$ and signatures $\widehat{\partial}_{ij} \widehat{f}$.

Lemma 2.10 ([15]). *Let $\widehat{f}$ be a signature of arity $n \geqslant 3$. If $\widehat{f}(\alpha) \neq 0$ for some* $\text{wt}(\alpha) = k \neq 0$ *and $k \neq n$, then there is a pair of indices $\{i, j\}$ such that $\widehat{\partial}_{ij} \widehat{f}(\beta) \neq 0$ for some* $\text{wt}(\beta) = k - 1$. *In particular, if for all pairs of indices $\{i, j\}$, $\widehat{\partial}_{ij} \widehat{f} \equiv 0$, then $\widehat{f}(\alpha) = 0$ for all α with* $\text{wt}(\alpha) \neq 0$ *and n.*

[1] We use f^{ab}_{ij} to denote a function, and $\mathbf{f}^{ab}_{ij}$ to denote a vector that lists the truth table of f^{ab}_{ij} in a given order.

When $\widehat{f}$ is an EO signature satisfying ARS, the following relation between $\widehat{f}$ and $\widehat{\partial}_{ij}\widehat{f}$ can be easily obtained following the proofs of Lemmas 4.3 and 4.5 in [14]. Let $\mathcal{D} = \{\neq_2\}$. Then $\mathcal{D}^{\otimes} = \{\lambda \cdot (\neq_2)^{\otimes k} \mid \lambda \in \mathbb{R}\backslash\{0\}, k \geqslant 1\}$ is the set of tensor products of $\neq_2$ up to nonzero real scalars.

Lemma 2.11. *Let $\widehat{f}$ be a $2n$-ary* EO signature satisfying ARS.

- When $2n = 8$, if for all pairs of indices $\{i, j\}$, $\widehat{\partial}_{ij}\widehat{f} \in \mathcal{D}^{\otimes}$, and there exists some $\neq_2 (x_i, x_j)$ and two pairs of indices $\{u, v\}$ and $\{s, t\}$ where $\{u, v\} \cap \{s, t\} \neq \emptyset$ such that $\neq_2 (x_i, x_j) \mid \widehat{\partial}_{uv}\widehat{f}, \widehat{\partial}_{st}\widehat{f}$, then $\widehat{f} \in \mathcal{D}^{\otimes}$ and $\neq_2 (x_i, x_j) \mid \widehat{f}$.
- When $2n \geqslant 10$, if for all pairs of indices $\{i, j\}$, $\widehat{\partial}_{ij}\widehat{f} \in \mathcal{D}^{\otimes}$, then $\widehat{f} \in \mathcal{D}^{\otimes}$.

Another gadget construction that connects a nonzero binary signature b with a signature f is called *extending*. An extending gadget connects one variable of f with one variable of b using $=_2$ in the setting of Holant$(=_2 \mid \mathcal{F})$, and connects one variable of $\widehat{f}$ with one variable of $\widehat{b}$ using $\neq_2$ in the setting of Holant$(\neq_2 \mid \widehat{\mathcal{F}})$. By extending an irreducible signature using $=_2$ or $\neq_2$, we still get an irreducible signature. A particular extending gadget is to extend f with binary signatures in $\mathcal{B}^{\otimes 1}$ using $=_2$ in the setting of Holant$(\mathcal{F})$. We use $\{f\}^{\mathcal{B}}_{=_2}$ to denote the set of signatures realizable by extending some variables of f with binary signatures in $\mathcal{B}^{\otimes 1}$ using $=_2$ (recall that $\mathcal{B}^{\otimes 1}$ allows all nonzero real normalization scalars). Equivalently, this gadget is to extend $\widehat{f}$ with binary signatures in $\widehat{\mathcal{B}}$ using $\neq_2$ in the setting of Holant$(\neq_2 \mid \widehat{\mathcal{F}})$. We use $\{\widehat{f}\}^{\widehat{\mathcal{B}}}_{\neq_2}$ to denote the set of signatures realizable by extending some variables of $\widehat{f}$ with binary signatures in $\widehat{\mathcal{B}}^{\otimes 1}$ using $\neq_2$. If $\widehat{g} \in \{\widehat{f}\}^{\widehat{\mathcal{B}}}_{\neq_2}$, then we can say that the extending gadget by $\widehat{\mathcal{B}}$ defines a relation between $\widehat{g}$ and $\widehat{f}$. Clearly, by extending variables of $\widehat{f}$ with $\neq_2 \in \widehat{\mathcal{B}}$ (using $\neq_2$), we still get $\widehat{f}$. Thus, $\widehat{f} \in \{\widehat{f}\}^{\widehat{\mathcal{B}}}_{\neq_2}$. So this relation is reflexive. The following lemma shows that this relation is symmetric and transitive, thus it is an equivalence relation.

Lemma 2.12. *1. $\widehat{g} \in \{\widehat{f}\}^{\widehat{\mathcal{B}}}_{\neq_2}$ iff $\widehat{f} \in \{\widehat{g}\}^{\widehat{\mathcal{B}}}_{\neq_2}$. 2. If $\widehat{h} \in \{\widehat{g}\}^{\widehat{\mathcal{B}}}_{\neq_2}$ and $\widehat{g} \in \{\widehat{f}\}^{\widehat{\mathcal{B}}}_{\neq_2}$, then $\widehat{h} \in \{\widehat{f}\}^{\widehat{\mathcal{B}}}_{\neq_2}$.*

Proof. Note that for any $\widehat{b} \in \widehat{\mathcal{B}}^{\otimes 1}$, if we connect any variable of $\widehat{b}$ with another arbitrary variable of a copy of the same $\widehat{b}$ using $\neq_2$, then we get $\neq_2$ after normalization. Also, by extending a variable of $\widehat{f}$ with $\neq_2$ (using $\neq_2$), we still get $\widehat{f}$. Suppose that $\widehat{g} \in \{\widehat{f}\}^{\widehat{\mathcal{B}}}_{\neq_2}$, and it is realized by extending certain variables x_i of $\widehat{f}$ with certain $b_i \in \widehat{\mathcal{B}}$. Then, by extending each of these variables x_i of $\widehat{g}$ with exactly the same $b_i \in \widehat{\mathcal{B}}$, we will get $\widehat{f}$ after normalization. Thus, $\widehat{f} \in \{\widehat{g}\}^{\widehat{\mathcal{B}}}_{\neq_2}$. The other direction is proved by exchanging $\widehat{f}$ and $\widehat{g}$. Thus, $\widehat{g} \in \{\widehat{f}\}^{\widehat{\mathcal{B}}}_{\neq_2}$ iff $\widehat{f} \in \{\widehat{g}\}^{\widehat{\mathcal{B}}}_{\neq_2}$.

Also, note that for any $\widehat{b^1}, \widehat{b^2} \in \widehat{\mathcal{B}}^{\otimes 1}$, by connecting an arbitrary variable of $\widehat{b^1}$ with an arbitrary variable of $\widehat{b^2}$ using $\neq_2$, we still get a signature in $\widehat{\mathcal{B}}^{\otimes 1}$. Suppose that $\widehat{h}$ is realized by extending some variables x_i of $\widehat{g}$ with some $b_i^1 \in \widehat{\mathcal{B}}^{\otimes 1}$. We may assume every variable x_i of $\widehat{g}$ has been so connected as $\neq_2 \in \widehat{\mathcal{B}}^{\otimes 1}$. Similarly we can assume $\widehat{g}$ is realized by extending every variable x_i of $\widehat{f}$ with some $b_i^2 \in \widehat{\mathcal{B}}^{\otimes 1}$. Let b_i be the signature realized by connecting b_i^1 and b_i^2 (using $\neq_2$). Then, $\widehat{h}$ can be realized by extending each variable x_i of $\widehat{f}$ with $b_i \in \widehat{\mathcal{B}}^{\otimes 1}$. Thus, $\widehat{h} \in \{\widehat{f}\}^{\widehat{\mathcal{B}}}_{\neq_2}$. □

Remark: As a corollary, if $\widehat{g} \in \{\widehat{f}\}^{\widehat{\mathcal{B}}}_{\neq_2}$, then $\{\widehat{g}\}^{\widehat{\mathcal{B}}}_{\neq_2} = \{\widehat{f}\}^{\widehat{\mathcal{B}}}_{\neq_2}$.

Lemma 2.13. *Let $\widehat{b_1}(x_1, x_2), \widehat{b_2}(y_1, y_2) \in \widehat{\mathcal{O}}$. If by connecting the variable x_1 of $\widehat{b_1}$ and the variable y_1 of $\widehat{b_2}$ using $\neq_2$, we get $\lambda \cdot \neq_2 (x_2, y_2)$ for some $\lambda \in \mathbb{R}\backslash\{0\}$, then $\widehat{b_1} \sim \widehat{b_2}$. Moreover, by connecting the variable x_2 of $\widehat{b_1}$ and the variable y_2 of $\widehat{b_2}$, we will get $\lambda \cdot \neq_2 (x_1, y_1)$.*

Proof. We prove this lemma in the setting of Holant($\mathcal{F}$) after the transformation Z back. Now, $b_1 = Z\widehat{b_1} \in \mathcal{O}$ and $b_2 = Z\widehat{b_2} \in \mathcal{O}$.

Consider matrices $M_1(b_1) = M_2^{\mathtt{T}}(b_1)$ and $M_1(b_2) = M_2^{\mathtt{T}}(b_2)$. Since $b_1, b_2 \in \mathcal{O}$, both $M_1(b_1)$ and $M_1(b_2)$ are real multiples of real orthogonal matrices, of which there are two types, either rotations or reflections. For such matrices X, Y, to get $X^{\mathtt{T}}Y = \lambda I_2$ for some $\lambda \in \mathbb{R}\backslash\{0\}$, X and Y must be either both reflections, or both rotations of the same angle, up to nonzero real multiples. First suppose $M_1(b_1) = \left[\begin{smallmatrix} a & b \\ b & -a \end{smallmatrix}\right]$, reflection. Then by connecting x_1 of b_1 and y_1 of b_2 using $=_2$ we get $\lambda\cdot =_2 (x_2, y_2)$, i.e., $M_1^{\mathtt{T}}(b_1)M_1(b_2) = \lambda I_2$. This implies that b_2 is the same reflection up to a nonzero scalar, i.e., $b_2 \sim b_1$. Similarly, for a rotation $M_1(b_1) = \left[\begin{smallmatrix} a & b \\ -b & a \end{smallmatrix}\right]$, $M_1^{\mathtt{T}}(b_1)M_1(b_2) = \lambda I_2$ implies that b_2 is also a rotation of the same angle as b_1 up to a nonzero scalar, thus $b_2 \sim b_1$. In either case, by connecting the variable x_2 of b_1 and the variable y_2 of b_2, we will get

$$M_2^{\mathtt{T}}(b_1)M_2(b_2) = M_1(b_1)M_1^{\mathtt{T}}(b_2) = \lambda I_2.$$

This means that we get the signature $\lambda\cdot =_2 (x_1, y_1)$. The statement of the lemma follows from this after a Z^{-1} transformation. □

A gadget construction often used in this paper is *mating*. Given a real-valued signature f of arity $n \geqslant 2$, we connect two copies of f in the following manner: Fix a set S of $n - m$ variables among all n variables of f. For each $x_k \in S$, connect x_k of one copy of f with x_k of the other copy using $=_2$. The variables that are not in S are called dangling variables. In this paper, we only consider the case that $m = 1$ or 2. For $m = 1$, there is one dangling variable x_i. Then, the mating construction realizes a signature of arity 2, denoted by $\mathfrak{m}_i f$. It can be represented by matrix multiplication. We have

$$M(\mathfrak{m}_i f) = M_i(f)I_2^{\otimes(n-1)}M_i^{\mathtt{T}}(f) = \begin{bmatrix} \mathbf{f}_i^0 \\ \mathbf{f}_i^1 \end{bmatrix} \begin{bmatrix} \mathbf{f}_i^{0\mathtt{T}} & \mathbf{f}_i^{1\mathtt{T}} \end{bmatrix} = \begin{bmatrix} |\mathbf{f}_i^0|^2 & \langle \mathbf{f}_i^0, \mathbf{f}_i^1 \rangle \\ \langle \mathbf{f}_i^0, \mathbf{f}_i^1 \rangle & |\mathbf{f}_i^1|^2, \end{bmatrix} \tag{2.1}$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product and $|\cdot|$ denotes the norm defined by this inner product. (We will use the same notation $\langle \cdot, \cdot \rangle$ to denote the complex inner product (with conjugation) below. The notation is consistent.) Note that $|\langle \mathbf{f}_i^0, \mathbf{f}_i^1 \rangle|^2 \leqslant |\mathbf{f}_i^0|^2 |\mathbf{f}_i^1|^2$ by the Cauchy-Schwarz inequality. Similarly, in the setting of Holant($\neq_2 | \widehat{\mathcal{F}}$), the above mating operation is equivalent to connecting variables in S using $\neq_2$. We denote the resulting signature by $\widehat{\mathfrak{m}}_i \widehat{f}$, which is the same as $\widehat{\mathfrak{m}_i f}$, and we have

$$M(\widehat{\mathfrak{m}}_i \widehat{f}) = M_i(\widehat{f})N_2^{\otimes n-1}M_i^{\mathtt{T}}(\widehat{f}) = \begin{bmatrix} \widehat{\mathbf{f}}_i^0 \\ \widehat{\mathbf{f}}_i^1 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}^{\otimes(n-1)} \begin{bmatrix} \widehat{\mathbf{f}}_i^{0\mathtt{T}} & \widehat{\mathbf{f}}_i^{1\mathtt{T}} \end{bmatrix}.$$

Note that (in general complex-valued) $\widehat{f}$ satisfies the ARS since f is real, we have

$$N_2^{\otimes(n-1)}\widehat{\mathbf{f}}_i^{0\mathtt{T}} = (\widehat{f}^{0,11\ldots1}, \widehat{f}^{0,11\ldots0}, \ldots, \widehat{f}^{0,00\ldots0})^{\mathtt{T}} = (\overline{\widehat{f}^{1,00\ldots0}}, \overline{\widehat{f}^{1,00\ldots1}}, \ldots, \overline{\widehat{f}^{1,11\ldots1}}) = \overline{\widehat{\mathbf{f}}_i^1}^{\mathtt{T}}.$$

Thus, we have

$$M(\widehat{\mathfrak{m}}_i \widehat{f}) = \begin{bmatrix} \widehat{\mathbf{f}}_i^0 \\ \widehat{\mathbf{f}}_i^1 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}^{\otimes(n-1)} \begin{bmatrix} \widehat{\mathbf{f}}_i^{0\mathtt{T}} & \widehat{\mathbf{f}}_i^{1\mathtt{T}} \end{bmatrix} = \begin{bmatrix} \widehat{\mathbf{f}}_i^0 \\ \widehat{\mathbf{f}}_i^1 \end{bmatrix} \begin{bmatrix} \overline{\widehat{\mathbf{f}}_i^1}^{\mathtt{T}} & \overline{\widehat{\mathbf{f}}_i^0}^{\mathtt{T}} \end{bmatrix} = \begin{bmatrix} \langle \widehat{\mathbf{f}}_i^0, \widehat{\mathbf{f}}_i^1 \rangle & |\widehat{\mathbf{f}}_i^0|^2 \\ |\widehat{\mathbf{f}}_i^1|^2 & \langle \widehat{\mathbf{f}}_i^1, \widehat{\mathbf{f}}_i^0 \rangle \end{bmatrix}. \tag{2.2}$$

If there are two dangling variables x_i and x_j, we use $\mathfrak{m}_{ij} f$ and $\widehat{\mathfrak{m}}_{ij} \widehat{f}$ to denote the signatures realized by mating f using $=_2$ and mating $\widehat{f}$ using $\neq_2$ respectively.

With respect to mating gadgets, the following first order orthogonality was introduced.

Definition 2.14 (First order orthogonality [15]). *Let f be a complex-valued signature of arity $n \geqslant 2$. It satisfies the* first order orthogonality (1ST-ORTH) *if there exists some $\mu \neq 0$ such that for all indices $i \in [n]$, the entries of f satisfy the following equations*

$$|\mathbf{f}_i^0|^2 = |\mathbf{f}_i^1|^2 = \mu, \text{ and } \langle \mathbf{f}_i^0, \mathbf{f}_i^1 \rangle = 0.$$

Remark: When f is a real-valued signature, the inner product is just the ordinary dot product which can be represented by mating using $=_2$. Thus, f satisfies 1ST-ORTH iff there is some real $\mu \neq 0$ such that for all indices i, $M(\mathfrak{m}_i f) = \mu I_2$. On the other hand, when $\widehat{f}$ is a signature with ARS, by (2.2), the complex inner product can be represented by mating using $\neq_2$. Thus, $\widehat{f}$ satisfies 1ST-ORTH iff there is some real $\mu \neq 0$ such that for all i, $M(\widehat{\mathfrak{m}}_i \widehat{f}) = \mu N_2$. Moreover, f satisfies 1ST-ORTH iff $\widehat{f}$ satisfies it.

Lemma 2.15 ([15]). *Let f be a real-valued signature of arity n. If for all indices $i \in [n]$, $M(\mathfrak{m}_i f) = \mu_i I_2$ for some real $\mu_i \neq 0$, then f satisfies* 1ST-ORTH *(i.e., all μ_i have the same value).*

2.5 Tractable signatures

We give some known signature sets that define polynomial time computable (tractable) counting problems.

Definition 2.16. *Let $\mathscr{T}$ denote the set of tensor products of unary and binary signatures.*

Definition 2.17. *A signature on a set of variables X is of* product type *if it can be expressed as a product of unary functions, binary equality functions ($[1, 0, 1]$), and binary disequality functions ($[0, 1, 0]$), each on one or two variables of X. We use $\mathscr{P}$ to denote the set of product-type functions.*

Note that the product in Definition 2.17 are ordinary products of functions (not tensor products); in particular they may be applied on overlapping sets of variables.

Definition 2.18. *A signature $f(x_1, \ldots, x_n)$ of arity n is* affine *if it has the form*

$$\lambda \cdot \chi_{AX=0} \cdot \mathrm{i}^{Q(X)},$$

where $\lambda \in \mathbb{C}$, $X = (x_1, x_2, \ldots, x_n, 1)$, A is a matrix over $\mathbb{Z}_2$, $Q(x_1, x_2, \ldots, x_n) \in \mathbb{Z}_4[x_1, x_2, \ldots, x_n]$ is a multilinear polynomial with total degree $d(Q) \leqslant 2$ and the additional requirement that the coefficients of all cross terms are even, i.e., Q has the form

$$Q(x_1, x_2, \ldots, x_n) = a_0 + \sum_{k=1}^{n} a_k x_k + \sum_{1 \leq i < j \leq n} 2 b_{ij} x_i x_j,$$

and χ is a 0-1 indicator function such that $\chi_{AX=0}$ is 1 iff $AX = 0$. We use $\mathscr{A}$ to denote the set of all affine signatures.

If the support set $\mathscr{S}(f)$ is an affine linear subspace, then we say f has affine support. Clearly, any affine signature has affine support. Moreover, we have that any signature of product type has affine support [9]. When $\mathscr{S}(f)$ is affine, we can pick a set of free variables such that in $\mathscr{S}(f)$, every variable is an affine linear combination of free variables. Affine functions satisfy the following congruity or semi-congruity.

Lemma 2.19 ([9]). *Let $f(x_1,\dots,x_n)=(-1)^{Q(x_1,\dots,x_n)}\in\mathscr{A}$, and $y=x_n+L(x_1,\dots,L_{n-1})$ be a linear combination of variables $x_1,\dots,x_n$ that involves x_n. Define*

$$g(x_1,\dots,x_{n-1})=\frac{f_{y=0}(x_1,\dots,x_{n-1},y+L)}{f_{y=1}(x_1,\dots,x_{n-1},y+L)}=(-1)^{Q(x_1,\dots,x_{n-1},L)+Q(x_1,\dots,x_{n-1},L+1)}.$$

Then, g satisfies the following property.

- *(Congruity) $g\equiv 1$ or $g\equiv -1$, or*
- *(Semi-congruity) $g(x_1,\dots,x_{n-1})=(-1)^{L(x_1,\dots,x_{n-1})}$ where $L(x_1,\dots,x_{n-1})\in\mathbb{Z}_2[x_1,\dots,x_{n-1}]$ is an affine linear polynomial (degree $d(L)=1$).*

In particular, if $d(Q)=1$, then g has congruity.

Let $T_{\alpha^s}=\left[\begin{smallmatrix}1&0\\0&\alpha^s\end{smallmatrix}\right]$ where $\alpha=\frac{1+\mathrm{i}}{\sqrt{2}}$ and s is an integer.

Definition 2.20. *A signature f is local-affine if for each $\sigma=s_1s_2\dots s_n\in\{0,1\}^n$ in the support of f, $(T_{\alpha^{s_1}}\otimes T_{\alpha^{s_2}}\otimes\cdots\otimes T_{\alpha^{s_n}})f\in\mathscr{A}$. We use $\mathscr{L}$ to denote the set of local-affine signatures.*

Definition 2.21. *We say a signature set $\mathcal{F}$ is $\mathscr{C}$-transformable if there exists a $T\in\mathrm{GL}_2(\mathbb{C})$ such that $(=_2)(T^{-1})^{\otimes 2}\in\mathscr{C}$ and $T\mathcal{F}\subseteq\mathscr{C}$.*

This definition is important because if Holant($\mathscr{C}$) is tractable, then Holant($\mathcal{F}$) is tractable for any $\mathscr{C}$-transformable set $\mathcal{F}$. Then, the following tractable result is known [20, 2].

Theorem 2.22. *Let $\mathcal{F}$ be a set of complex valued signatures. Then* Holant($\mathcal{F}$) *is tractable if*

$$\mathcal{F}\subseteq\mathscr{T},\quad \mathcal{F}\text{ is }\mathscr{P}\text{-transformable},\quad \mathcal{F}\text{ is }\mathscr{A}\text{-transformable},\quad \text{or }\mathcal{F}\text{ is }\mathscr{L}\text{-transformable}.\qquad (\mathrm{T})$$

Lemma 2.23 ([15]). *Let $\mathcal{F}$ be a set of real-valued signatures. If $\mathcal{F}$ does not satisfy condition* (T), then for every $Q\in\mathbf{O}_2$, $Q\mathcal{F}$ also does not satisfy condition (T). Moreover, $\widehat{\mathcal{F}}\not\subseteq\mathscr{P}$ and $\widehat{\mathcal{F}}\not\subseteq\mathscr{A}$.

2.6 Hardness results and P-time reductions

We give some known hardness results. We state these results for our setting.

Theorem 2.24 ([20, 15]). *Let $\mathcal{F}$ be a set of real-valued signatures. If $\mathcal{F}$ does not satisfy condition* (T). Then for every $Q\in\mathbf{O}_2$ and every $k\geqslant 2$, $\#\mathrm{CSP}_2(Q\mathcal{F})$ and Holant($\neq_2|=_k,\widehat{Q}\widehat{\mathcal{F}}$) are #P-hard.

Theorem 2.25 ([15]). *Let $\mathcal{F}$ be a set of real-valued signatures containing a nonzero signature of odd arity. If $\mathcal{F}$ does not satisfy condition* (T), then Holant($\mathcal{F}$) is #P-hard.

The following reduction is obtained by *polynomial interpolation*.

Lemma 2.26 ([9]). *Let f and g be nonzero binary signatures with $M(f)=P^{-1}\left[\begin{smallmatrix}\lambda_1&0\\0&\lambda_2\end{smallmatrix}\right]P$ and $M(g)=P^{-1}\left[\begin{smallmatrix}1&0\\0&0\end{smallmatrix}\right]P$ for some invertible matrix P. If $\lambda_1\neq 0$ and $|\frac{\lambda_2}{\lambda_1}|\neq 1$, then*

$$\mathrm{Holant}(g,\mathcal{F})\leqslant_T\mathrm{Holant}(f,\mathcal{F})$$

for any signature set $\mathcal{F}$.

By Lemmas 2.7 and 2.26, we can always realize a nonzero unary signature from nonzero signatures not satisfying 1ST-ORTH. We have the following #P-hardness result.

Lemma 2.27. *Let $\mathcal{F}$ be a set of real-valued signatures containing a nonzero signature that does not satisfy* 1ST-ORTH. *If $\mathcal{F}$ does not satisfy condition* (T), then Holant($\mathcal{F}$) is #P-hard.

Proof. Consider $M_i(f)$ for all indices i. Clearly, $M(\mathfrak{m}_i f) = M_i(f)M_i^{\mathsf{T}}(f)$ is a real symmetric positive semi-definite matrix, which is diagonalizable with two non-negative real eigenvalues $\lambda_i \geqslant \mu_i \geqslant 0$. These two eigenvalues are not both zero since f is real valued and $f \not\equiv 0$, and so $M(\mathfrak{m}_i f) \neq 0$. Thus, $\lambda_i \neq 0$. Then, $|\frac{\mu_i}{\lambda_i}| = 1$ iff $\lambda_i = \mu_i$. In other words, $M(\mathfrak{m}_i f) = \mu_i I_2$ for some real $\mu_i \neq 0$.

Since f does not satisfy 1ST-ORTH, by Lemma 2.15, there is an index i such that $M(\mathfrak{m}_i f) \neq \mu_i I_2$ for any real $\mu_i \neq 0$. Thus, $M(\mathfrak{m}_i f)$ has two eigenvalues with different norms. By Lemma 2.26, we can realize a nonzero binary signature g such that $M(g)$ is degenerate. This implies that g can be factorized as a tensor product of two nonzero unary signatures. By Lemma 2.7, we can realize a nonzero unary signature and hence by Theorem 2.25, Holant($\mathcal{F}$) is #P-hard. □

We also need to use the results of *eight-vertex models* and *Eulerian Orientation* (EO) problems.

Theorem 2.28 ([13]). *Let $\widehat{f}$ be a signature with $M(\widehat{f}) = \begin{bmatrix} c & 0 & 0 & a \\ 0 & d & b & 0 \\ 0 & \overline{b} & \overline{d} & 0 \\ \overline{a} & 0 & 0 & \overline{c} \end{bmatrix}$. Then,* Holant($\neq_2 |\ \widehat{f}$) *is #P-hard in the following cases.*

- *$\widehat{f}$ has support 6,*
- *$\widehat{f}$ has support 4 and the nonzero entries of $M(\widehat{f})$ do not have the same norm, or*
- *$\widehat{f}$ has support 8, all nonzero entries of $M(\widehat{f})$ are positive real numbers and are not all equal.*

Theorem 2.29 ([14]). *Let $\widehat{\mathcal{F}}$ be a set of* EO signatures (i.e., with half-weighted support) satisfying ARS. Then Holant($\mathcal{DEQ} \mid \widehat{\mathcal{F}}$) is #P-hard unless $\widehat{\mathcal{F}} \subseteq \mathscr{P}$ or $\widehat{\mathcal{F}} \subseteq \mathscr{A}$.

The following reduction states that we can realize all $\mathcal{EQ}_2$ once we have $=_4$ in Holant($\mathcal{F}$).

Lemma 2.30 ([9]). $\#\mathrm{CSP}_2(\mathcal{F}) \leqslant_T \mathrm{Holant}(=_4, \mathcal{F})$.

The following reductions state that we can realize all $\mathcal{DEQ}$ once we have $\neq_4$ in Holant($\neq_2 |\ \widehat{\mathcal{F}}$).

Lemma 2.31 ([14]). $\mathrm{Holant}(\mathcal{DEQ} \mid \widehat{\mathcal{F}}) \leqslant_T \mathrm{Holant}(\neq_2 |\ \mathcal{DEQ}, \widehat{\mathcal{F}}) \leqslant_T \mathrm{Holant}(\neq_2 | \neq_4, \widehat{\mathcal{F}})$.

2.7 A summary of notations

We use the following Table 1 to summarize notations given in this section. In the left column, we list notations in Holant($=_2 |\ \mathcal{F}$) where $\mathcal{F}$ is a set of real-valued signatures, and in the right column, we list corresponding notations in Holant($\neq_2 |\ \widehat{\mathcal{F}}$) where $\widehat{\mathcal{F}} = Z^{-1}\mathcal{F}$ is the set of complex-valued signatures with ARS. Note that although $\mathcal{EO}$ also satisfies ARS, we will only use it in Holant($=_2 |\ \mathcal{F}$). Similarly, we will only use $\mathcal{DEQ}$ and $\mathcal{D}$ in Holant($\neq_2 |\ \widehat{\mathcal{F}}$) although it is real-valued.

Recall that $\mathcal{F}^{\otimes}$ denotes the set $\{\lambda \bigotimes_{i=1}^{k} f_i \mid \lambda \in \mathbb{R}\backslash\{0\}, k \geqslant 1, f_i \in \mathcal{F}\}$ for any signature set $\mathcal{F}$. We remark that both $\mathcal{O}^{\otimes}$ and $\widehat{\mathcal{O}}^{\otimes}$ contain all zero signatures of even arity since the binary zero signature is in $\mathcal{O}$ and $\widehat{\mathcal{O}}$. However, $\mathcal{B}^{\otimes}$, $\widehat{\mathcal{B}}^{\otimes}$, and $\mathcal{D}^{\otimes}$ do *not* contain any zero signatures.

In the following, without other specifications, we use f to denote a real-valued signature and $\mathcal{F}$ to denote a set of real-valued signatures. We use $\widehat{f}$ to denote a signature satisfying ARS and $\widehat{\mathcal{F}}$ to denote a set of such signatures. We use Q to denote a matrix in $\mathbf{O}_2$, and $\widehat{Q}$ to denote a matrix in $\widehat{\mathbf{O}_2}$. Clearly, if $\mathcal{F}$ is real-valued, then $Q\mathcal{F}$ is also real-valued. Equivalently, if $\widehat{\mathcal{F}}$ satisfies ARS, then $\widehat{Q}\widehat{\mathcal{F}} = \widehat{Q\mathcal{F}}$ also satisfies ARS.

Holant$(=_2 \mid \mathcal{F})$ where $\mathcal{F}$ is real-valued	Holant$(\neq_2 \mid \widehat{\mathcal{F}})$ where $\widehat{\mathcal{F}}$ satisfies ARS
$\mathcal{EQ} = \{=_1, =_2, \ldots, =_n, \ldots\}$	N/A
N/A	$\mathcal{DEQ} = \{\neq_2, \neq_4, \ldots, \neq_{2n}, \ldots\}$, $\mathcal{D} = \{\neq_2\}$
$\mathcal{O} = \{$binary orthogonal and zero signatures$\}$	$\widehat{\mathcal{O}} = \{$binary signatures with ARS and parity$\}$
$\mathcal{B} = \{=_2, =_2^-, \neq_2, \neq_2^-\}$	$\widehat{\mathcal{B}} = \{\neq_2, =_2, (-\mathrm{i})\cdot =_2^-, \mathrm{i}\cdot \neq_2^-\}$
a holographic transformation $Q\mathcal{F}$ by $Q \in \mathbf{O}_2$	a holographic transformation $\widehat{Q}\widehat{\mathcal{F}}$ by $\widehat{Q} \in \widehat{\mathbf{O}_2}$
a merging gadget $\partial_{ij} f = f_{ij}^{00} + f_{ij}^{11}$	a merging gadget $\widehat{\partial}_{ij} \widehat{f} = \widehat{f}_{ij}^{01} + \widehat{f}_{ij}^{10}$
extending gadgets $\{f\}_{=_2}^{\mathcal{B}}$ with $\mathcal{B}$	extending gadgets $\{\widehat{f}\}_{\neq_2}^{\widehat{\mathcal{B}}}$ with $\widehat{\mathcal{B}}$
a mating gadget $\mathfrak{m}_{ij} f = M_{ij}(f) I_2^{\otimes n-1} M_{ij}^{\mathrm{T}}(f)$	a mating gadget $\widehat{\mathfrak{m}}_{ij} \widehat{f} = M_{ij}(\widehat{f}) N_2^{\otimes n-1} M_{ij}^{\mathrm{T}}(\widehat{f})$

Table 1: Comparisons of notations in Holant$(=_2 \mid \mathcal{F})$ and Holant$(\neq_2 \mid \widehat{\mathcal{F}})$

3 Proof Organization

By Theorem 2.22, if $\mathcal{F}$ satisfies condition (T), then Holant$(\mathcal{F})$ is P-time computable. So, we only need to prove that Holant$(\mathcal{F})$ is #P-hard when $\mathcal{F}$ does not satisfy condition (T). If $\mathcal{F}$ contains a nonzero signature of odd arity, then by Theorem 2.25, we are done. In the following without other specifications, when refer to a real-valued signature set $\mathcal{F}$ or a corresponding signature set $\widehat{\mathcal{F}} = Z^{-1}\mathcal{F}$ satisfying ARS, we always assume that they consist of signatures of even arity, and $\mathcal{F}$ does not satisfy condition (T).

In Section 4, we generalize the notion of first order orthogonality (1ST-ORTH) to second order orthogonality (2ND-ORTH). This property plays a key role in our proof. We show that all irreducible signatures in $\mathcal{F}$ satisfy 2ND-ORTH, or else, we get #P-hardness based on results of #CSP problems, #EO problems and eight-vertex models (Lemma 4.4). We derive some consequences from the condition 2ND-ORTH for signatures with ARS. These will be used throughout in the proof.

In Section 5, we give the induction framework of the proof. Since $\mathcal{F}$ does not satisfy condition (T), $\mathcal{F} \not\subseteq \mathscr{T}$. Also since $\mathcal{O}^{\otimes} \subseteq \mathscr{T}$, and by Lemma 5.1, we may assume that $\mathcal{F}$ contains a signature f of arity $2n \geqslant 4$ where $f \notin \mathcal{O}^{\otimes}$. We want to achieve a proof of #P-hardness by induction on $2n$. When $2n = 2$, as a corollary of 1ST-ORTH, we show that Holant$(\mathcal{F})$ is #P-hard (Lemma 5.1). When $2n = 4$, by 2ND-ORTH, we show that Holant$(\mathcal{F})$ is #P-hard (Lemma 5.2).

In Sections 6 and 7, we handle the case of arity 6. Let $f \notin \mathcal{O}^{\otimes}$ be a 6-ary signature in $\mathcal{F}$. We show that Holant$(\mathcal{F})$ is #P-hard or the extraordinary signature which we named f_6 with the Bell property can be realized (Theorem 6.5). By gadget construction, all four Bell signatures $\mathcal{B}$ can be realized from f_6. Then we prove the #P-hardness of Holant$^b(f_6, \mathcal{F})$ = Holant$(\mathcal{B}, f_6, \mathcal{F})$ (Theorem 7.19 and Lemma 7.20). Combining these two results, we have Holant$(\mathcal{F})$ is #P-hard (Lemma 7.21).

In Section 8, we handle the case of arity 8. Let $f \notin \mathcal{O}^{\otimes}$ be an 8-ary signature in $\mathcal{F}$. We show that Holant$(\mathcal{F})$ is #P-hard or another extraordinary signature which we named f_8 with the strong Bell property can be realized (Theorem 8.5). One can prove that $\mathcal{B}$ cannot be realized from f_8 by gadget construction. However, by introducing Holant problems with limited appearance and using the strong Bell property of f_8, we show Holant$^b(f_8, \mathcal{F}) \leq_T$ Holant$(f_8, \mathcal{F})$ (Lemmas 8.10). Then, we

prove the #P-hardness of Holant$^b(f_8, \mathcal{F})$. Combining these results, we have Holant$(\mathcal{F})$ is #P-hard (Lemma 8.12).

In Section 9, we show that our induction proof works for signatures of arity $2n \geqslant 10$. Let $f \notin \mathcal{O}^{\otimes}$ be a $2n$-ary ($2n \geqslant 10$) signature in $\mathcal{F}$. Then, Holant$(\mathcal{F})$ is #P-hard or we can realize a signature of arity less than $2n$ that is not in $\mathcal{O}^{\otimes}$ (Lemma 9.1). Then, by a sequence of reductions of length independent of the problem instance size, we can eventually realize a signature of arity at most 8 that is not in $\mathcal{O}^{\otimes}$. Finally, combining Lemmas 5.1, 5.2, 7.21, 8.12 and 9.1, we finish the proof of Theorem 1.2. In the actual proof, for convenience, many results are proved in the setting of Holant$(\neq| \widehat{\mathcal{F}})$ which is equivalent to Holant$(\mathcal{F})$ under the Z^{-1} transformation.

4 Second Order Orthogonality

In this section, we generalize the notion of first order orthogonality (1ST-ORTH) to second order orthogonality (2ND-ORTH) (Definition 4.1). We show that for real-valued $\mathcal{F}$ that does not satisfy condition (T), every irreducible $f \in \mathcal{F}$ of arity at least 4 satisfies 2ND-ORTH, or otherwise Holant$(\mathcal{F})$ is #P-hard (Lemma 4.4). Then, we derive some consequences from the condition 2ND-ORTH for signatures with ARS. These will be used throughout in the following proof.

Definition 4.1 (Second order orthogonality)**.** *Let f be a complex-valued signature of arity $n \geqslant 4$. It satisfies the* second order orthogonality (2ND-ORTH) *if there exists some $\lambda \neq 0$ such that for all pairs of indices $\{i, j\} \subseteq [n]$, the entries of f satisfy*

$$|\mathbf{f}_{ij}^{00}|^2 = |\mathbf{f}_{ij}^{01}|^2 = |\mathbf{f}_{ij}^{10}|^2 = |\mathbf{f}_{ij}^{11}|^2 = \lambda, \qquad \text{and} \qquad \langle \mathbf{f}_{ij}^{ab}, \mathbf{f}_{ij}^{cd} \rangle = 0 \quad \text{for all } (a, b) \neq (c, d).$$

Remark: Similar to the remark of first order orthogonality (Definition 2.14), f satisfies 2ND-ORTH iff there is some $\lambda \neq 0$ such that for all (i, j), $M(\mathfrak{m}_{ij} f) = \lambda I_4 = \lambda I_2^{\otimes 2}$, and $\widehat{f}$ satisfies 2ND-ORTH iff there is some $\lambda \neq 0$ such that for all (i, j), $M(\widehat{\mathfrak{m}}_{ij} \widehat{f}) = \lambda N_4 = \lambda N_2^{\otimes 2}$. Moreover, f satisfies 2ND-ORTH iff $\widehat{f}$ satisfies it. Clearly, 2ND-ORTH implies 1ST-ORTH.

In the next, we will prove Lemma 4.4 based on dichotomies of #CSP problems, #EO problems and eight-vertex models. Since #EO problems and eight-vertex models are defined as special cases of the problem Holant$(\neq_2| \widehat{\mathcal{F}})$, for convenience, we will consider the problem Holant$(\neq_2| \widehat{\mathcal{F}})$ which is equivalent to Holant$(\mathcal{F})$. Recall that $\widehat{\mathcal{F}} = Z^{-1}\mathcal{F}$ satisfies ARS, and we assumed that $\mathcal{F}$ does not satisfy condition (T). We first give the following lemma.

Lemma 4.2. Holant$(\mathcal{DEQ} \mid \widehat{\mathcal{F}})$ *is #P-hard.*

Proof. Since $\mathcal{F}$ does not satisfy condition (T), by Lemma 2.23, $\widehat{\mathcal{F}} \not\subseteq \mathscr{P}$ and $\widehat{\mathcal{F}} \not\subseteq \mathscr{A}$. If $\widehat{\mathcal{F}}$ is a set of EO signatures, then by Theorems 2.29, Holant$(\mathcal{DEQ} \mid \widehat{\mathcal{F}})$ is #P-hard since $\widehat{\mathcal{F}} \not\subseteq \mathscr{P}$ and $\widehat{\mathcal{F}} \not\subseteq \mathscr{A}$. Thus, we may assume that there is a signature $\widehat{f} \in \widehat{\mathcal{F}}$ whose support is not half-weighted. Suppose that $\widehat{f}$ has arity $2n$. Since $\mathscr{S}(\widehat{f}) \not\subseteq \mathscr{H}_{2n}$, by ARS, there is an $\alpha \in \mathbb{Z}_2^{2n}$ with $\mathrm{wt}(\alpha) = k < n$ such that $\widehat{f}(\alpha) \neq 0$. We first show that we can realize a signature $\widehat{g}$ of arity $2n - 2k$ such that $\widehat{g}(\vec{0}) \neq 0$. If $\mathrm{wt}(\alpha) = k = 0$, then we are done. Otherwise, we have $n > k \geqslant 1$. Thus, $2n \geqslant 4$ and α has length at least 4. By Lemma 2.10, there is a pair of indices $\{i, j\}$ such that $\widehat{\partial}_{ij} \widehat{f}(\beta) \neq 0$ for some $\mathrm{wt}(\beta) = k - 1$. Clearly, $\widehat{\partial}_{ij} \widehat{f}$ has arity $2n - 2$. Since $0 \leqslant k - 1 < (2n - 2)/2$, $\widehat{\partial}_{ij} \widehat{f}$ is not an EO signature. Now we can continue this process, and by a chain of merging gadgets using $\neq_2$, we can realize a signature $\widehat{g}$ of arity $2m = 2n - 2k$ such that $\widehat{g}(\vec{0}) \neq 0$. Denote by $a = \widehat{g}(\vec{0})$.

Then, we connect all $2m$ variables of $\widehat{g}$ with $2m$ variables of $\neq_{4m}$ that always take the same value in $\mathscr{S}(\neq_{4m})$ using $\neq_2$. We get a signature $\widehat{h}$ of arity $2m$ where $\widehat{h}(\vec{0}) = a$, $\widehat{h}(\vec{1}) = \bar{a}$ by ARS, and $\widehat{h}(\gamma) = 0$ elsewhere. Then, consider the holographic transformation by $\widehat{Q} = \begin{bmatrix} \sqrt[2m]{\bar{a}} & 0 \\ 0 & \sqrt[2m]{a} \end{bmatrix} \in \widehat{\mathrm{O}}_2$. It transforms $\widehat{h}$ to $\neq_{2m}$, but does not change $\mathcal{DEQ}$. Thus,

$$\mathrm{Holant}(\mathcal{DEQ} \mid \widehat{h}, \widehat{\mathcal{F}}) \equiv_T \mathrm{Holant}(\mathcal{DEQ} \mid =_{2m}, \widehat{Q}\widehat{\mathcal{F}}).$$

Since $\widehat{\mathcal{F}}$ does not satisfy condition (T), by Theorem 2.24, $\mathrm{Holant}(\mathcal{DEQ} \mid =_{2m}, \widehat{Q}\widehat{\mathcal{F}})$ is #P-hard. Thus, $\mathrm{Holant}(\mathcal{DEQ} \mid \widehat{\mathcal{F}})$ is #P-hard. □

Then, we consider signatures $\widehat{\mathfrak{m}}_{ij}\widehat{f}$ realized by mating.

Lemma 4.3. *Let $\widehat{f} \in \widehat{\mathcal{F}}$ be a signature of arity $2n \geqslant 4$. Then,*

- $\mathrm{Holant}(\neq_2 \mid \widehat{\mathcal{F}})$ *is #P-hard, or*
- *for all pairs of indices $\{i, j\}$, there exists a nonzero binary signature $\widehat{b}_{ij} \in \widehat{\mathcal{O}}$ such that $\widehat{b}_{ij}(x_i, x_j) \mid \widehat{f}$ or $M(\widehat{\mathfrak{m}}_{ij}\widehat{f}) = \lambda_{ij} N_4$ for some real $\lambda_{ij} \neq 0$.*

Proof. If $\widehat{f} \equiv 0$, then the lemma holds trivially since for all $\{i, j\}$ and any $\widehat{b}_{ij} \neq 0$, $\widehat{b}_{ij}(x_i, x_j) \mid \widehat{f}$. Thus, we may assume that $f \not\equiv 0$.

If $\widehat{f}$ does not satisfy 1ST-ORTH, then f does not satisfy it. By Lemma 2.27, $\mathrm{Holant}(\neq_2 \mid \widehat{\mathcal{F}}) \equiv_T \mathrm{Holant}(=_2 \mid \mathcal{F})$ is #P-hard. Thus, we may assume that $\widehat{f}$ satisfies 1ST-ORTH. Then, for all indices i, we have

$$M(\widehat{\mathfrak{m}}_i\widehat{f}) = \begin{bmatrix} \langle \widehat{\mathbf{f}}_i^0, \widehat{\mathbf{f}}_i^1 \rangle & |\widehat{\mathbf{f}}_i^0|^2 \\ |\widehat{\mathbf{f}}_i^1|^2 & \langle \widehat{\mathbf{f}}_i^1, \widehat{\mathbf{f}}_i^0 \rangle \end{bmatrix} = \mu \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

For any variable x_i, we may take another variable x_j $(j \neq i)$ and partition the sum in the inner product $\langle \widehat{\mathbf{f}}_i^0, \widehat{\mathbf{f}}_i^1 \rangle = 0$ into two sums depending on whether $x_j = 0$ or 1. Also, by ARS we have

$$\langle \widehat{\mathbf{f}}_i^0, \widehat{\mathbf{f}}_i^1 \rangle = \langle \widehat{\mathbf{f}}_{ij}^{00}, \widehat{\mathbf{f}}_{ij}^{10} \rangle + \langle \widehat{\mathbf{f}}_{ij}^{01}, \widehat{\mathbf{f}}_{ij}^{11} \rangle = \langle \widehat{\mathbf{f}}_{ij}^{00}, \widehat{\mathbf{f}}_{ij}^{10} \rangle + \langle \overline{\widehat{\mathbf{f}}_{ij}^{10}}, \overline{\widehat{\mathbf{f}}_{ij}^{00}} \rangle = 2\langle \widehat{\mathbf{f}}_{ij}^{00}, \widehat{\mathbf{f}}_{ij}^{10} \rangle = 0.$$

Thus, for all pairs of indices $\{i, j\}$, $\langle \widehat{\mathbf{f}}_{ij}^{00}, \widehat{\mathbf{f}}_{ij}^{10} \rangle = 0$ and $\langle \widehat{\mathbf{f}}_{ij}^{01}, \widehat{\mathbf{f}}_{ij}^{11} \rangle = 0$. (Note that by exchanging i and j we also have $\langle \widehat{\mathbf{f}}_{ij}^{00}, \widehat{\mathbf{f}}_{ij}^{01} \rangle = 0$ and $\langle \widehat{\mathbf{f}}_{ij}^{10}, \widehat{\mathbf{f}}_{ij}^{11} \rangle = 0$.) Also by ARS, we have $|\widehat{\mathbf{f}}_{ij}^{00}|^2 = |\overline{\widehat{\mathbf{f}}_{ij}^{11}}|^2 = |\widehat{\mathbf{f}}_{ij}^{11}|^2$ and $|\widehat{\mathbf{f}}_{ij}^{01}|^2 = |\overline{\widehat{\mathbf{f}}_{ij}^{10}}|^2 = |\widehat{\mathbf{f}}_{ij}^{10}|^2$.

Now, consider $\widehat{\mathfrak{m}}_{ij}\widehat{f}$ for all pairs of indices $\{i, j\}$.

$$M(\widehat{\mathfrak{m}}_{ij}\widehat{f}) = \begin{bmatrix} \widehat{\mathbf{f}}_{ij}^{00} \\ \widehat{\mathbf{f}}_{ij}^{01} \\ \widehat{\mathbf{f}}_{ij}^{10} \\ \widehat{\mathbf{f}}_{ij}^{11} \end{bmatrix} \begin{bmatrix} \overline{\widehat{\mathbf{f}}_{ij}^{11}}^{\mathsf{T}} & \overline{\widehat{\mathbf{f}}_{ij}^{10}}^{\mathsf{T}} & \overline{\widehat{\mathbf{f}}_{ij}^{01}}^{\mathsf{T}} & \overline{\widehat{\mathbf{f}}_{ij}^{00}}^{\mathsf{T}} \end{bmatrix} = \begin{bmatrix} \langle \widehat{\mathbf{f}}_{ij}^{00}, \widehat{\mathbf{f}}_{ij}^{11} \rangle & 0 & 0 & |\widehat{\mathbf{f}}_{ij}^{00}|^2 \\ 0 & \langle \widehat{\mathbf{f}}_{ij}^{01}, \widehat{\mathbf{f}}_{ij}^{10} \rangle & |\widehat{\mathbf{f}}_{ij}^{01}|^2 & 0 \\ 0 & |\widehat{\mathbf{f}}_{ij}^{10}|^2 & \langle \widehat{\mathbf{f}}_{ij}^{10}, \widehat{\mathbf{f}}_{ij}^{01} \rangle & 0 \\ |\widehat{\mathbf{f}}_{ij}^{11}|^2 & 0 & 0 & \langle \widehat{\mathbf{f}}_{ij}^{11}, \widehat{\mathbf{f}}_{ij}^{00} \rangle \end{bmatrix}.$$

Note that $|\langle \widehat{\mathbf{f}}_{ij}^{00}, \widehat{\mathbf{f}}_{ij}^{11} \rangle| \leqslant |\widehat{\mathbf{f}}_{ij}^{00}| \cdot |\widehat{\mathbf{f}}_{ij}^{11}|$ by Cauchy-Schwarz inequality. Clearly, $\widehat{\mathfrak{m}}_{ij}\widehat{f}$ has even parity, and thus it represents a signature of the eight-vertex model. If there exists a pair of indices $\{i, j\}$ such that $\mathrm{Holant}(\neq_2 \mid \widehat{\mathfrak{m}}_{ij}\widehat{f})$ is #P-hard, then we are done since $\mathrm{Holant}(\neq_2 \mid \widehat{\mathfrak{m}}_{ij}\widehat{f}) \leqslant_T \mathrm{Holant}(\neq_2 \mid \widehat{\mathcal{F}})$. Thus, we may assume all $\widehat{\mathfrak{m}}_{ij}\widehat{f}$ belong to the tractable family for eight-vertex models. Clearly, by observing its antidiagonal entries of the matrix $M(\widehat{\mathfrak{m}}_{ij}\widehat{f})$, we have $\widehat{\mathfrak{m}}_{ij}\widehat{f} \not\equiv 0$ since $\widehat{f} \not\equiv 0$. By Theorem 2.28, there are three possible cases.

- There exists a pair $\{i,j\}$ such that $\widehat{\mathfrak{m}}_{ij}\widehat{f}$ has support of size 2. By Cauchy-Schwarz inequality, $M(\widehat{\mathfrak{m}}_{ij}\widehat{f})$ is either of the form $\lambda_{ij}\left[\begin{smallmatrix}0&0&0&1\\0&0&0&0\\0&0&0&0\\1&0&0&0\end{smallmatrix}\right]$ where $\lambda_{ij}=|\widehat{\mathbf{f}}_{ij}^{00}|^2=|\widehat{\mathbf{f}}_{ij}^{11}|^2\neq 0$ or $\lambda_{ij}\left[\begin{smallmatrix}0&0&0&0\\0&0&1&0\\0&1&0&0\\0&0&0&0\end{smallmatrix}\right]$ where $\lambda_{ij}=|\widehat{\mathbf{f}}_{ij}^{01}|=|\widehat{\mathbf{f}}_{ij}^{10}|\neq 0$. In both cases, $\neq_4$ is realizable since $\lambda_{ij}\neq 0$. The form that $\langle\widehat{\mathbf{f}}_{ij}^{01},\widehat{\mathbf{f}}_{ij}^{10}\rangle\neq 0$ while $|\widehat{\mathbf{f}}_{ij}^{01}|^2=|\widehat{\mathbf{f}}_{ij}^{10}|^2=0$ cannot occur since $|\langle\widehat{\mathbf{f}}_{ij}^{01},\widehat{\mathbf{f}}_{ij}^{10}\rangle|\leqslant|\widehat{\mathbf{f}}_{ij}^{01}||\widehat{\mathbf{f}}_{ij}^{10}|$. Also, the form that $\langle\widehat{\mathbf{f}}_{ij}^{00},\widehat{\mathbf{f}}_{ij}^{11}\rangle\neq 0$ while $|\widehat{\mathbf{f}}_{ij}^{00}|^2=|\widehat{\mathbf{f}}_{ij}^{11}|^2=0$ cannot occur. Since $\neq_4$ is available, by Lemma 2.31, Holant$(\mathcal{DEQ}\mid\widehat{\mathcal{F}})\leqslant_T$ Holant$(=_2|\widehat{\mathcal{F}})$. By Lemma 4.2, Holant$(=_2|\widehat{\mathcal{F}})$ is #P-hard.
- There exists a pair $\{i,j\}$ such that $\widehat{\mathfrak{m}}_{ij}\widehat{f}$ has support of size 8. We can rename the four variables of $\widehat{\mathfrak{m}}_{ij}\widehat{f}$ in a cyclic permutation. We use $\widehat{g}$ to denote this signature. Then $M(\widehat{g})=M_{12}(\widehat{g})=\left[\begin{smallmatrix}c&0&0&d\\0&a&b&0\\0&b&a&0\\\bar{d}&0&0&\bar{c}\end{smallmatrix}\right]$ where a and b are positive real numbers and c and d are nonzero complex numbers. Consider the signature $\widehat{\mathfrak{m}}_{12}\widehat{g}$ realized by mating $\widehat{g}$. We denote it by $\widehat{h}$. Then,
$$M(\widehat{h})=M(\widehat{g})N_4M^{\mathtt{T}}(\widehat{g})=\begin{bmatrix}2cd&0&0&|c|^2+|d|^2\\0&2ab&a^2+b^2&0\\0&a^2+b^2&2ab&0\\|c|^2+|d|^2&0&0&2\bar{c}\bar{d}\end{bmatrix}=\begin{bmatrix}c'&0&0&d'\\0&a'&b'&0\\0&b'&a'&0\\d'&0&0&\bar{c'}\end{bmatrix},$$
where a',b', and d' are positive real numbers and c' is a nonzero complex number. Suppose that the argument of c' is θ, i.e., $c'=|c'|e^{\mathrm{i}\theta}$.
Consider the holographic transformation by $\widehat{Q}=\left[\begin{smallmatrix}e^{-\mathrm{i}\theta/4}&0\\0&e^{\mathrm{i}\theta/4}\end{smallmatrix}\right]\in\widehat{\mathbf{O}_2}$. Then,
$$\text{Holant}(\neq_2|\,\widehat{h},\widehat{\mathcal{F}})\equiv_T\text{Holant}(\neq_2|\,\widehat{Q}\widehat{h},\widehat{Q}\widehat{\mathcal{F}}).$$
Note that $M(\widehat{Q}\widehat{h})=\left[\begin{smallmatrix}|c'|&0&0&d'\\0&a'&b'&0\\0&b'&a'&0\\d'&0&0&|c'|\end{smallmatrix}\right]$ where all entries are positive real numbers. Notice that all weight 2 entries of $\widehat{h}$ are unchanged in $\widehat{Q}\widehat{h}$. By Theorem 2.28, Holant$(\neq_2|\,\widehat{Q}\widehat{h})$ is #P-hard unless $a'=b'=|c'|=d'$. Thus, we may assume that $M(\widehat{Q}\widehat{h})=\left[\begin{smallmatrix}1&0&0&1\\0&1&1&0\\0&1&1&0\\1&0&0&1\end{smallmatrix}\right]$ up to normalization.
Notice that $M(Z(\widehat{Q}\widehat{h}))=Z^{\otimes 2}M(\widehat{Q}\widehat{h})(Z^{\mathtt{T}})^{\otimes 2}=\left[\begin{smallmatrix}1&0&0&0\\0&0&0&0\\0&0&0&0\\0&0&0&1\end{smallmatrix}\right]$, which is the arity-4 equality $(=_4)$.
Consider the holographic transformation by Z which transfers $\neq_2$ back to $=_2$. Remember that $\widehat{Q}=Z^{-1}QZ$. Then, $Z(\widehat{Q}\widehat{F})=Z(Z^{-1}QZ)(Z^{-1}\mathcal{F})=Q\mathcal{F}$. Since $\widehat{Q}\in\widehat{\mathbf{O}_2}$, we have $Q\in\mathbf{O}_2$. Thus,
$$\text{Holant}(\neq_2|\,\widehat{Q}\widehat{h},\widehat{Q}\widehat{\mathcal{F}})\equiv_T\text{Holant}(=_2|=_4,Q\mathcal{F}).$$
By Lemma 2.30, $\#\text{CSP}_2(Q\mathcal{F})\leqslant_T\text{Holant}(=_2|=_4,Q\mathcal{F})$. Since $\mathcal{F}$ does not satisfy condition (T) and $Q\in\mathbf{O}_2$, by Theorem 2.24, $\#\text{CSP}_2(Q\mathcal{F})$ is #P-hard. Thus, Holant$(\neq_2|\,\widehat{\mathcal{F}})$ is #P-hard.
- For all $\{i,j\}$, $\widehat{\mathfrak{m}}_{ij}\widehat{f}$ has support of size 4. By Cauchy-Schwarz inequality, $M(\widehat{\mathfrak{m}}_{ij}\widehat{f})$ is of the form $\left[\begin{smallmatrix}b&0&0&a\\0&0&0&0\\0&0&0&0\\a&0&0&\bar{b}\end{smallmatrix}\right]$ or $\left[\begin{smallmatrix}0&0&0&0\\0&b&a&0\\0&\bar{a}&\bar{b}&0\\0&0&0&0\end{smallmatrix}\right]$ where $a^2-|b|^2=0$, or the form $\lambda_{ij}\left[\begin{smallmatrix}0&0&0&1\\0&0&1&0\\0&1&0&0\\1&0&0&0\end{smallmatrix}\right]$ where $\lambda_{ij}=|\widehat{\mathbf{f}}_{ij}^{00}|^2=|\widehat{\mathbf{f}}_{ij}^{01}|\neq 0$. If $M(\mathfrak{m}_{ij}\widehat{f})=\lambda_{ij}N_4$, then we are done. Otherwise, $M(\widehat{\mathfrak{m}}_{ij}\widehat{f})$ has rank one. Hence, $M_{ij}(\widehat{f})$ also has rank one. Then, by observing the form of $M(\widehat{\mathfrak{m}}_{ij}\widehat{f})$ especially the all zero rows, $\widehat{f}$ can be factorized as $\widehat{b}_{ij}(x_i,x_j)\otimes\widehat{g}$ where $\widehat{b}_{ij}\in\widehat{\mathcal{O}}$ and $\widehat{g}$ is a signature on the other $n-2$ variables. Thus, we are done.

The lemma is proved. □

Remark: We give a restatement of Lemma 4.3 in the setting of Holant($\mathcal{F}$). Let $f \in \mathcal{F}$ be a signature of arity $2n \geqslant 4$. Then, Holant($\mathcal{F}$) is #P-hard, or for all pairs of indices $\{i, j\}$, there exists a nonzero binary signature $b_{ij} \in \mathcal{O}$ such that $b_{ij}(x_i, x_j) \mid f$ or $M(\mathfrak{m}_{ij} f) = \lambda_{ij} I_4$ for some real $\lambda_{ij} \neq 0$.

Now for an irreducible signature $\widehat{f}$ of arity $2n \geqslant 4$, we show that it satisfies 2ND-ORTH or we get #P-hardness.

Lemma 4.4. *Let $\widehat{f} \in \widehat{\mathcal{F}}$ be an irreducible signature of arity $2n \geqslant 4$. If $\widehat{f}$ does not satisfy* 2ND-ORTH, *then* Holant($\neq \mid \widehat{\mathcal{F}}$) *is #P-hard.*

Proof. Since $\widehat{f}$ is irreducible, by Lemma 4.3, $M(\widehat{\mathfrak{m}}_{ij} \widehat{f}) = \lambda_{ij} N_4$ for all $\{i, j\}$. Now, we show all λ_{ij} have the same value. If we connect further the two respective pairs of variables of $\mathfrak{m}_{ij} f$, which totally connects two copies of f, we get a value $4\lambda_{ij}$. This value clearly does not depend on the particular indices $\{i, j\}$. We denote the value λ_{ij} by λ. This value is nonzero because $\widehat{f}$ is irreducible. □

We derive some consequences from the condition 2ND-ORTH for signatures with ARS. Suppose that $\widehat{f}$ satisfies 2ND-ORTH. First, by definition we have $|\widehat{\mathbf{f}}_{ij}^{ab}|^2 = \lambda$ for any $(x_i, x_j) = (a, b) \in \{0, 1\}^2$. Given a vector $\widehat{\mathbf{f}}_{ij}^{ab}$, we can pick a third variable x_k and partition $\widehat{\mathbf{f}}_{ij}^{ab}$ into two vectors $\widehat{\mathbf{f}}_{ijk}^{ab0}$ and $\widehat{\mathbf{f}}_{ijk}^{ab1}$ according to $x_k = 0$ or 1. By setting $(a, b) = (0, 0)$, we have

$$|\widehat{\mathbf{f}}_{ij}^{00}|^2 = |\widehat{\mathbf{f}}_{ijk}^{000}|^2 + |\widehat{\mathbf{f}}_{ijk}^{001}|^2 = \lambda. \tag{4.1}$$

Similarly, we consider the vector $\widehat{\mathbf{f}}_{ik}^{00}$ and partition it according to $x_j = 0$ or 1. We have

$$|\widehat{\mathbf{f}}_{ik}^{00}|^2 = |\widehat{\mathbf{f}}_{ijk}^{000}|^2 + |\widehat{\mathbf{f}}_{ijk}^{010}|^2 = \lambda. \tag{4.2}$$

Comparing equations (4.1) and (4.2), we have $|\widehat{\mathbf{f}}_{ijk}^{001}|^2 = |\widehat{\mathbf{f}}_{ijk}^{010}|^2$. Moreover, by ARS, we have $|\widehat{\mathbf{f}}_{ijk}^{010}|^2 = |\widehat{\mathbf{f}}_{ijk}^{101}|^2$. Thus, we have $|\widehat{\mathbf{f}}_{ijk}^{001}|^2 = |\widehat{\mathbf{f}}_{ijk}^{101}|^2$. Note that the vector $\widehat{\mathbf{f}}_{jk}^{01}$ is partitioned into two vectors $\widehat{\mathbf{f}}_{ijk}^{001}$ and $\widehat{\mathbf{f}}_{ijk}^{101}$ according to $x_i = 0$ or 1. That is

$$|\widehat{\mathbf{f}}_{jk}^{01}|^2 = |\widehat{\mathbf{f}}_{ijk}^{001}|^2 + |\widehat{\mathbf{f}}_{ijk}^{101}|^2 = \lambda.$$

Thus, we have $|\widehat{\mathbf{f}}_{ijk}^{001}|^2 = |\widehat{\mathbf{f}}_{ijk}^{101}|^2 = \lambda/2$. Then, by equation (4.1), we have $|\widehat{\mathbf{f}}_{ijk}^{000}|^2 = \lambda/2$, and again by ARS, we also have $|\widehat{\mathbf{f}}_{ijk}^{111}|^2 = |\widehat{\mathbf{f}}_{ijk}^{000}|^2 = \lambda/2$. Note that indices i, j, k are picked arbitrarily, by symmetry, we have

$$|\widehat{\mathbf{f}}_{ijk}^{abc}|^2 = \lambda/2 \tag{4.3}$$

for all $(x_i, x_j, x_k) = (a, b, c) \in \{0, 1\}^3$.

Given a vector $\widehat{\mathbf{f}}_{ijk}^{abc}$, we can continue to pick a fourth variable x_ℓ and partition $\widehat{\mathbf{f}}_{ijk}^{abc}$ into two vectors $\widehat{\mathbf{f}}_{ijk\ell}^{abc0}$ and $\widehat{\mathbf{f}}_{ijk\ell}^{abc1}$ according to $x_\ell = 0$ or 1. By setting $(a, b, c) = (0, 0, 0)$, we have from (4.3)

$$|\widehat{\mathbf{f}}_{ijk}^{000}|^2 = |\widehat{\mathbf{f}}_{ijk\ell}^{0000}|^2 + |\widehat{\mathbf{f}}_{ijk\ell}^{0001}|^2 = \lambda/2. \tag{4.4}$$

Similarly, we consider the vector $\widehat{\mathbf{f}}^{001}_{ij\ell}$ and partition it according to $x_k = 0$ or 1. We have

$$|\widehat{\mathbf{f}}^{001}_{ij\ell}|^2 = |\widehat{\mathbf{f}}^{0001}_{ijk\ell}|^2 + |\widehat{\mathbf{f}}^{0011}_{ijk\ell}|^2 = \lambda/2. \tag{4.5}$$

Comparing equations (4.4) and (4.5), and also by ARS, we have

$$|\widehat{\mathbf{f}}^{0000}_{ijk\ell}|^2 = |\widehat{\mathbf{f}}^{0011}_{ijk\ell}|^2 = |\widehat{\mathbf{f}}^{1100}_{ijk\ell}|^2 = |\widehat{\mathbf{f}}^{1111}_{ijk\ell}|^2 \tag{4.6}$$

for all indices $\{i, j, k, \ell\}$. Similarly, we can get

$$|\widehat{\mathbf{f}}^{0001}_{ijk\ell}|^2 = |\widehat{\mathbf{f}}^{0010}_{ijk\ell}|^2 = |\widehat{\mathbf{f}}^{1101}_{ijk\ell}|^2 = |\widehat{\mathbf{f}}^{1110}_{ijk\ell}|^2. \tag{4.7}$$

By the definition of second order orthogonality, we also have

$$\langle \widehat{\mathbf{f}}^{ab}_{ij}, \widehat{\mathbf{f}}^{cd}_{ij} \rangle = 0 \tag{4.8}$$

for all variables x_i, x_j and $(a, b) \neq (c, d)$.

Equations (4.6), (4.7) and (4.8) will be used frequently in the analysis of signatures satisfying ARS and 2ND-ORTH. This is also a reason why we consider the problem in the setting under the Z^{-1} transformation, Holant($\neq_2 |\ \widehat{\mathcal{F}}$), where we can express these consequences of 2ND-ORTH elegantly, instead of Holant($\mathcal{F}$) which is logically equivalent. By combining 2ND-ORTH and ARS of the signature $\widehat{f}$, we get these simply expressed, thus easily applicable, conditions in terms of norms and inner products.

5 The Induction Proof: Base Cases $2n \leqslant 4$

In this section, we introduce the induction framework and handle the base cases (Lemmas 5.1 and 5.2). Recall that $\widehat{\mathcal{O}}$ denotes the set of binary signatures with ARS and parity (including the binary zero signature), and $\widehat{\mathcal{O}}^{\otimes}$ denotes the set of tensor products of signatures in $\widehat{\mathcal{O}}$. Since $\mathcal{F}$ does not satisfy condition (T), $\widehat{\mathcal{F}} \not\subseteq \mathscr{T}$. Also, since $\widehat{\mathcal{O}}^{\otimes} \subseteq \mathscr{T}$, $\widehat{\mathcal{F}} \not\subseteq \widehat{\mathcal{O}}^{\otimes}$. Thus, there is a nonzero signature $\widehat{f} \in \widehat{\mathcal{F}}$ of arity $2n$ such that $\widehat{f} \notin \widehat{\mathcal{O}}^{\otimes}$. We want to achieve a proof of #P-hardness by induction on $2n$. We first consider the base that $2n = 2$. Notice that a nonzero binary signature $\widehat{f}$ satisfies 1ST-ORTH iff its matrix form (as a 2-by-2 matrix) is orthogonal. Thus, $\widehat{f} \notin \widehat{\mathcal{O}}$ implies that it does not satisfy 1ST-ORTH. Then, we have the following result.

Lemma 5.1. *Let $\mathcal{F}$ contain a binary signature $f \notin \mathcal{O}^{\otimes}$. Then,* Holant($\mathcal{F}$) *is #P-hard.*
Equivalently, let $\widehat{\mathcal{F}}$ contain a binary signature $\widehat{f} \notin \widehat{\mathcal{O}}^{\otimes}$. Then, Holant($\neq_2 |\ \widehat{\mathcal{F}}$) *is #P-hard.*

Proof. We prove this lemma in the setting of Holant($\mathcal{F}$). Since $\mathcal{O}^{\otimes}$ contains the binary zero signature, $f \notin \mathcal{O}^{\otimes}$ implies that $f \not\equiv 0$. If f is reducible, then it is a tensor product of two nonzero unary signatures. By Lemma 2.7, we can realize a nonzero unary signature by factorization, and we are done by Theorem 2.25. Otherwise, f is irreducible. Since $f \notin \mathcal{O}^{\otimes}$, f does not satisfy 1ST-ORTH. By Lemma 2.27, Holant($\mathcal{F}$) is #P-hard. □

Then, the general induction framework is that we start with a signature $\widehat{f}$ of arity $2n \geqslant 4$ that is not in $\widehat{\mathcal{O}}^{\otimes}$, and realize a signature $\widehat{g}$ of arity $2k \leqslant 2n - 2$ that is also not in $\widehat{\mathcal{O}}^{\otimes}$, or otherwise we can directly show Holant($\neq_2 |\ \widehat{\mathcal{F}}$) is #P-hard. If we can reduce the arity down to 2 (by a sequence

of reductions of length independent of the problem instance size), then we have a binary signature $\widehat{b} \notin \widehat{\mathcal{O}}$. By Lemma 5.1, we are done.

For the inductive step, we first consider the case that $\widehat{f}$ is reducible. Suppose that $\widehat{f} = \widehat{f}_1 \otimes \widehat{f}_2$. If $\widehat{f}_1$ or $\widehat{f}_2$ have odd arity, then we can realize a signature of odd arity by factorization and we are done. Otherwise, $\widehat{f}_1$ and $\widehat{f}_2$ have even arity. Since $\widehat{f} \notin \widehat{\mathcal{O}}^{\otimes}$, we know $\widehat{f}_1$ and $\widehat{f}_2$ cannot both be in $\widehat{\mathcal{O}}^{\otimes}$. Then, we can realize a signature of lower arity that is not in $\widehat{\mathcal{O}}^{\otimes}$ by factorization. We are done. Thus, in the following we may assume that $\widehat{f}$ is irreducible. Then, we may further assume that $\widehat{f}$ satisfies 2ND-ORTH. Otherwise, we get #P-hardness by Lemma 4.4. We use merging with $\neq_2$ to realize signatures of arity $2n-2$ from $\widehat{f}$. Consider $\widehat{\partial}_{ij}\widehat{f}$ for all pairs of indices $\{i,j\}$. If there exists a pair $\{i,j\}$ such that $\widehat{\partial}_{ij}\widehat{f} \notin \widehat{\mathcal{O}}^{\otimes}$, then we can realize $\widehat{g} = \widehat{\partial}_{ij}\widehat{f}$ which has arity $2n-2$, and we are done. Thus, we may assume $\widehat{\partial}_{ij}\widehat{f} \in \widehat{\mathcal{O}}^{\otimes}$ for all $\{i,j\}$. We denote this property by $\widehat{f} \in \widehat{\int}\widehat{\mathcal{O}}^{\otimes}$. We want to achieve our induction proof based on these two properties: 2ND-ORTH and $\widehat{f} \in \widehat{\int}\widehat{\mathcal{O}}^{\otimes}$. We consider the case that $2n = 4$.

Lemma 5.2. *Let $\widehat{\mathcal{F}}$ contain a 4-ary signature $\widehat{f} \notin \widehat{\mathcal{O}}^{\otimes}$. Then,* Holant$(\neq_2 | \widehat{\mathcal{F}})$ *is #P-hard.*

Proof. Since $\widehat{f} \notin \widehat{\mathcal{O}}^{\otimes}$, $f \not\equiv 0$. First, we may assume that $\widehat{f}$ is irreducible. Otherwise, we can realize a nonzero unary signature or a binary signature that is not in $\widehat{\mathcal{O}}$. Then, by Theorem 2.25 and Lemma 5.1, we have #P-hardness. Since $\widehat{f}$ is irreducible, we may further assume that $\widehat{f}$ satisfies 2ND-ORTH. Otherwise, by Lemma 4.4, we get #P-hardness.

We consider binary signatures $\widehat{\partial}_{ij}\widehat{f}$ realized from $\widehat{f}$ by merging using $\neq_2$. Under the assumption that $\widehat{f}$ satisfies 2ND-ORTH, we will show that there exits a pair $\{i,j\}$ such that $\widehat{\partial}_{ij}\widehat{f} \notin \widehat{\mathcal{O}}$. Then by Lemma 5.1, we are done. For a contradiction, suppose that $\widehat{f} \in \widehat{\int}\widehat{\mathcal{O}}$ i.e., $\widehat{\partial}_{ij}\widehat{f} \in \widehat{\mathcal{O}}$ for all pairs $\{i,j\}$. Since $\widehat{f}$ satisfies 2ND-ORTH, by equations (4.6) and (4.7), we have $|\widehat{\mathbf{f}}^{0000}_{ijk\ell}| = |\widehat{\mathbf{f}}^{0011}_{ijk\ell}| = |\widehat{\mathbf{f}}^{1111}_{ijk\ell}|$ and $|\widehat{\mathbf{f}}^{0001}_{ijk\ell}| = |\widehat{\mathbf{f}}^{1110}_{ijk\ell}|$ respectively for any permutation (i,j,k,ℓ) of $(1,2,3,4)$. Thus all entries of $\widehat{f}$ on inputs of even weight $\{0,2,4\}$ have the same norm, and all entries of $\widehat{f}$ on inputs of odd weight $\{1,3\}$ have the same norm. We denote by ν_0 and ν_1 the norm squares of entries on inputs of even weight and odd weight, respectively.

Then, we consider the equation $\langle \widehat{\mathbf{f}}^{01}_{12}, \widehat{\mathbf{f}}^{10}_{12} \rangle = 0$ from (4.8) by taking $(i,j) = (1,2)$. We have

$$\langle \widehat{\mathbf{f}}^{01}_{12}, \widehat{\mathbf{f}}^{10}_{12} \rangle = \widehat{f}^{0100}\overline{\widehat{f}^{1000}} + f^{0101}\overline{\widehat{f}^{1001}} + \widehat{f}^{0110}\overline{\widehat{f}^{1010}} + \widehat{f}^{0111}\overline{\widehat{f}^{1011}} = 0.$$

(Here for clarity, we omitted the subscript 1234 of $\widehat{f}^{abcd}_{1234}$.) By ARS, we have $\widehat{f}^{0111}\overline{\widehat{f}^{1011}} = \overline{\widehat{f}^{1000}}\widehat{f}^{0100}$ and $\widehat{f}^{0110}\overline{\widehat{f}^{1010}} = \overline{\widehat{f}^{1001}}\widehat{f}^{0101}$. Thus, we have

$$\widehat{f}^{0100}\overline{\widehat{f}^{1000}} + \widehat{f}^{0101}\overline{\widehat{f}^{1001}} = 0. \tag{5.1}$$

Note that by taking norm, $|\widehat{f}^{0100}\overline{\widehat{f}^{1000}}| = \nu_1$ and $|\widehat{f}^{0101}\overline{\widehat{f}^{1001}}| = \nu_0$. Then, it follows that $\nu_0 = \nu_1$. Thus, all entries of $\widehat{f}$ have the same norm. We normalize the norm to be 1 since $\widehat{f} \not\equiv 0$.

Consider $\widehat{\partial}_{12}\widehat{f}$. We have

$$\widehat{\partial}_{12}\widehat{f} = (\widehat{f}^{0100} + \widehat{f}^{1000}, \quad \widehat{f}^{0101} + \widehat{f}^{1001}, \quad \widehat{f}^{0110} + \widehat{f}^{1010}, \quad \widehat{f}^{0111} + \widehat{f}^{1011}),$$

and by assumption $\widehat{\partial}_{12}\widehat{f} \in \widehat{\mathcal{O}}$. Thus, at least one of the two entries $\widehat{f}^{0100} + \widehat{f}^{1000}$ and $\widehat{f}^{0101} + \widehat{f}^{1001}$ is equal to zero. If $\widehat{f}^{0100} + \widehat{f}^{1000} = 0$, then we have

$$\widehat{f}^{0100}\overline{\widehat{f}^{1000}} = (-\widehat{f}^{1000})\overline{\widehat{f}^{1000}} = -|\widehat{f}^{1000}|^2 = -1.$$

Then, by equation (5.1), we have $\widehat{f}^{0101}\overline{\widehat{f}^{1001}} = 1$. Otherwise, $\widehat{f}^{0101} + \widehat{f}^{1001} = 0$. Then, we have $\widehat{f}^{0101}\overline{\widehat{f}^{1001}} = -1$ while $\widehat{f}^{0100}\overline{\widehat{f}^{1000}} = 1$. Thus, among these two products $\widehat{f}^{0100}\overline{\widehat{f}^{1000}}$ and $\widehat{f}^{0101}\overline{\widehat{f}^{1001}}$, exactly one is equal to 1, while the other is -1. Then, we have

$$\widehat{f}^{0100}\overline{\widehat{f}^{1000}}\widehat{f}^{0101}\overline{\widehat{f}^{1001}} = -1.$$

Similarly, by considering $\widehat{\partial}_{23}\widehat{f}$ and $\widehat{\partial}_{31}\widehat{f}$ respectively, we have

$$\widehat{f}^{0010}\overline{\widehat{f}^{0100}}\widehat{f}^{0011}\overline{\widehat{f}^{0101}} = -1, \qquad \text{and} \qquad \widehat{f}^{1000}\overline{\widehat{f}^{0010}}\widehat{f}^{1001}\overline{\widehat{f}^{0011}} = -1.$$

Multiply these three products, we have

$$|\widehat{f}^{0100}|^2|\widehat{f}^{0010}|^2|\widehat{f}^{1000}|^2|\widehat{f}^{0101}|^2|\widehat{f}^{0011}|^2|\widehat{f}^{1001}|^2 = (-1)^3 = -1.$$

A contradiction! □

Remark: In this proof, we showed that there is no irreducible 4-ary signature $\widehat{f}$ that satisfies both 2ND-ORTH and $\widehat{f} \in \widehat{\int}\widehat{\mathcal{O}}^{\otimes}$.

If Lemma 5.2 were to hold for signatures of arity $2n \geqslant 6$, i.e., there is no irreducible signature $\widehat{f}$ of $2n \geqslant 6$ such that $\widehat{f}$ satisfies both 2ND-ORTH and $\widehat{f} \in \widehat{\int}\widehat{\mathcal{O}}^{\otimes}$, then the induction proof holds and we are done. We show that this is true for signatures of arity $2n \geqslant 10$ in Section 9. However, there are extraordinary signatures of arity 6 and 8 with special closure properties (Bell properties) such that they satisfy both 2ND-ORTH and $\widehat{f} \in \widehat{\int}\widehat{\mathcal{O}}^{\otimes}$.

6 First Major Obstacle: 6-ary Signatures with Bell Property

We consider the following 6-ary signature $\widehat{f}_6$. We use χ_S to denote the indicator function on set S. Let

$$\widehat{f}_6 = \chi_S \cdot (-1)^{x_1x_2+x_2x_3+x_1x_3+x_1x_4+x_2x_5+x_3x_6}$$

where $S = \mathscr{S}(\widehat{f}_6) = \mathscr{E}_6 = \{\alpha \in \mathbb{Z}_2^6 \mid \mathrm{wt}(\alpha) \equiv 0 \bmod 2\}$. One can check that $\widehat{f}_6$ is irreducible, and $\widehat{f}_6$ satisfies both 2ND-ORTH and $\widehat{f} \in \widehat{\int}\widehat{\mathcal{O}}^{\otimes}$. $\widehat{f}_6$ has the following matrix form

$$M_{123,456}(\widehat{f}_6) = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & -1 & 1 & 0 & 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 & 1 & 0 & 0 & -1 \\ -1 & 0 & 0 & -1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & -1 & 0 & 0 & -1 \\ -1 & 0 & 0 & 1 & 0 & -1 & 1 & 0 \\ -1 & 0 & 0 & 1 & 0 & 1 & -1 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 \end{bmatrix}. \tag{6.1}$$

We use Figure 1 to visualize this matrix. A block with orange color denotes an entry $+1$ and a block with blue color denotes an entry -1. Other blank blocks are zeros.

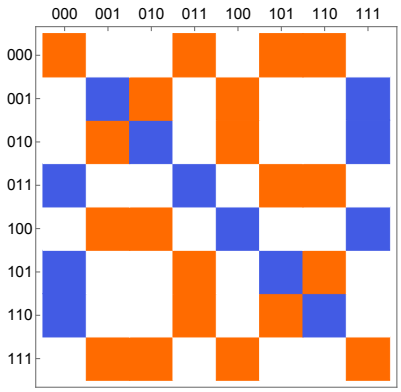


Figure 1: A visualization of $\widehat{f_6}$

6.1 The discovery of $\widehat{f_6}$

In this subsection, we show how this extraordinary signature $\widehat{f_6}$ was discovered. We prove that if $\widehat{\mathcal{F}}$ contains a 6-ary signature $\widehat{f}$ where $\widehat{f} \notin \widehat{\mathcal{O}}^{\otimes}$, then Holant$(\neq_2 | \widehat{\mathcal{F}})$ is #P-hard or $\widehat{f_6}$ is realizable from $\widehat{f}$ after a holographic transformation by some $\widehat{Q} \in \widehat{\mathbf{O}_2}$ (Theorem 6.5). The general strategy of this proof is to show that we can realize signatures with special properties from $\widehat{f}$ step by step (Lemmas 6.1, 6.2, 6.3 and 6.4), and finally we can realize $\widehat{f_6}$, or else we can realize signatures that lead to #P-hardness. So this $\widehat{f_6}$ emerges as essentially the unique (and true) obstacle to our proof of #P-hardness in this setting.

Lemma 6.1. *Suppose that $\widehat{\mathcal{F}}$ contains a 6-ary signature $\widehat{f} \notin \widehat{\mathcal{O}}^{\otimes}$. Then,* Holant$(\neq_2 | \widehat{\mathcal{F}})$ *is #P-hard, or an irreducible 6-ary signature $\widehat{f'}$ is realizable from $\widehat{f}$, where $\widehat{f'}(\alpha) = 0$ for all α with* $\mathrm{wt}(\alpha) = 2$ *or* 4*. Moreover, $\widehat{f'}$ is realizable by extending variables of $\widehat{f}$ with nonzero binary signatures in $\widehat{\mathcal{O}}$ that are realizable by factorization from $\widehat{\partial}_{12}\widehat{f}$.*

Proof. Since $\widehat{f} \notin \widehat{\mathcal{O}}^{\otimes}$, $\widehat{f} \not\equiv 0$. Again, we may assume that $\widehat{f}$ is irreducible. Otherwise, by factorization, we can realize a nonzero signature of odd arity, or a signature of arity 2 or 4 that is not in $\widehat{\mathcal{O}}^{\otimes}$. Then by Theorem 2.25, or Lemmas 5.1 or 5.2, we get #P-hardness. Under the assumption that $\widehat{f}$ is irreducible, we may further assume that $\widehat{f}$ satisfies 2ND-ORTH by Lemma 4.4. Also, we may assume that $\widehat{f} \in \int \widehat{\mathcal{O}}^{\otimes}$. Otherwise, there is a pair of indices $\{i, j\}$ such that the 4-ary signature $\widehat{\partial}_{ij}\widehat{f} \notin \widehat{\mathcal{O}}^{\otimes}$. Then by Lemma 5.2, Holant$(\neq_2 | \widehat{\mathcal{F}})$ is #P-hard.

If for all pairs of indices $\{i, j\}$, $\widehat{\partial}_{ij}\widehat{f} \equiv 0$, then by Lemma 2.10, we have $\widehat{f}(\alpha) = 0$ for all α with $\mathrm{wt}(\alpha) \neq 0$ and 6. Since $f \not\equiv 0$, clearly such a signature does not satisfy 2ND-ORTH. Contradiction. Otherwise, there is a pair of indices $\{i, j\}$ such that $\widehat{\partial}_{ij}\widehat{f} \not\equiv 0$. By renaming variables, without loss of generality, we assume that $\widehat{\partial}_{12}\widehat{f} \not\equiv 0$. Since $\widehat{\partial}_{12}\widehat{f} \in \widehat{\mathcal{O}}^{\otimes}$, in the UPF of $\widehat{\partial}_{12}\widehat{f}$, by renaming variables we assume that variables x_3 and x_4 appear in one nonzero binary signature $\widehat{b_1}(x_3, x_4) \in \widehat{\mathcal{O}}^{\otimes}$, and variables x_5 and x_6 appear in the other nonzero binary signature $\widehat{b_2}(x_5, x_6) \in \widehat{\mathcal{O}}^{\otimes}$. Thus, we have

$$\widehat{\partial}_{12}\widehat{f} = \widehat{b_1}(x_3, x_4) \otimes \widehat{b_2}(x_5, x_6) \not\equiv 0.$$

By Lemma 2.7, we know that these two binary signatures $\widehat{b_1}$ and $\widehat{b_2}$ are realizable by factorization. Note that for a nonzero binary signature $\widehat{b_i}(x_{2i+1}, x_{2i+2}) \in \widehat{\mathcal{O}}$ $(i \in \{1, 2\})$, if we connect the variable x_{2i+1} of two copies of $\widehat{b_i}(x_{2i+1}, x_{2i+2})$ using $\neq_2$ (mating two binary signatures), then we

get $\neq_2$ up to a nonzero scalar. We consider the following gadget construction G_1 on $\widehat{f}$. Recall that in the setting of Holant$(\neq|\ \widehat{\mathcal{F}})$, variables are connected using $\neq_2$. For $i \in \{1,2\}$, by a slight abuse of variable names, we connect the variable x_{2i+1} of $\widehat{f}$ with the variable x_{2i+1} of $\widehat{b_i}(x_{2i+1}, x_{2i+2})$. We get a signature $\widehat{f'}$ of arity 6. Such a gadget construction does not change the irreducibility of f. Thus, $\widehat{f'}$ is irreducible. Again, we may assume that $\widehat{f'} \in \widehat{\int}\widehat{\mathcal{O}}^{\otimes}$ and $\widehat{f'}$ satisfies 2ND-ORTH. Otherwise, we are done.

Consider $\widehat{\partial}_{12}\widehat{f'}$. Since the above gadget construction G_1 does not touch variables x_1 and x_2 of f, the operation of forming G_1 commutes with the merging operation $\widehat{\partial}_{12}$. Thus, $\widehat{\partial}_{12}\widehat{f'}$ can be realized by performing the gadget construction G_1 on $\widehat{\partial}_{12}\widehat{f}$, which connects each binary signature $\widehat{b_i}$ $(i \in \{1,2\})$ of $\widehat{\partial}_{12}\widehat{f}$ with another copy of itself using $\neq_2$ (in the mating fashion). Then, each $\widehat{b_i}$ in $\widehat{\partial}_{12}\widehat{f}$ is changed to $\neq_2$ up to a nonzero real scalar. After normalization and renaming variables, we have

$$\widehat{\partial}_{12}\widehat{f'} = (\neq_2)(x_3, x_4) \otimes (\neq_2)(x_5, x_6).$$

Since $\widehat{\partial}_{12}\widehat{f'} \in \mathcal{D}^{\otimes}$, for any $\{i,j\}$ disjoint with $\{1,2\}$ we have $\widehat{\partial}_{(ij)(12)}\widehat{f'} \in \mathcal{D}^{\otimes}$, and hence $\widehat{\partial}_{ij}\widehat{f'} \not\equiv 0$.

Now, we show that for all pairs of indices $\{i,j\}$, $\widehat{\partial}_{ij}\widehat{f'}$ has even parity. We first consider the case that $\{i,j\}$ is disjoint with $\{1,2\}$. Connect variables x_i and x_j of $\widehat{\partial}_{12}\widehat{f'}$ using $\neq_2$. Since $\widehat{\partial}_{12}\widehat{f'}$ has even parity, a merging gadget using $\neq_2$ will change the parity from even to odd. Thus, $\widehat{\partial}_{(ij)(12)}\widehat{f'}$ has odd parity. Consider $\widehat{\partial}_{ij}\widehat{f'}$. Remember that $\widehat{\partial}_{ij}\widehat{f'} \not\equiv 0$ since $\widehat{\partial}_{(ij)(12)}\widehat{f'} \not\equiv 0$. Since $\widehat{f'} \in \widehat{\int}\widehat{\mathcal{O}}^{\otimes}$, We have $\widehat{\partial}_{ij}\widehat{f'} \in \mathcal{O}^{\otimes}$. Thus, $\widehat{\partial}_{ij}\widehat{f'}$ has (either odd or even) parity. For a contradiction, suppose that it has odd parity. Then, $\widehat{\partial}_{(12)(ij)}\widehat{f'}$ has even parity since it is realized by merging using $\neq_2$. A signature that has both even parity and odd parity is identically zero. Thus $\widehat{\partial}_{(12)(ij)}\widehat{f'}$ is the zero signature. However, since $\widehat{\partial}_{(ij)(12)}\widehat{f'} \in \mathcal{D}^{\otimes}$, it is not the zero signature. Contradiction. Therefore, $\widehat{\partial}_{ij}\widehat{f'}$ has even parity for all $\{i,j\}$ disjoint with $\{1,2\}$.

Then, consider $\widehat{\partial}_{ij}\widehat{f'}$ for $\{i,j\} \cap \{1,2\} \neq \emptyset$. If $\{1,2\} = \{i,j\}$, then clearly, $\widehat{\partial}_{12}\widehat{f'}$ has even parity. Otherwise, without loss of generality, we may assume that $i = 1$ and $j \neq 2$. Consider $\widehat{\partial}_{1j}\widehat{f'}$ for $3 \leqslant j \leqslant 6$. If it is a zero signature, then it has even parity. Otherwise, $\widehat{\partial}_{1j}\widehat{f'} \not\equiv 0$. Since $\widehat{\partial}_{1j}\widehat{f'} \in \widehat{\mathcal{O}}^{\otimes}$, we assume that it has the following UPF

$$\widehat{\partial}_{1j}\widehat{f'} = \widehat{b'_1}(x_2, x_u) \otimes \widehat{b'_2}(x_v, x_w).$$

By connecting variables x_u and x_v of $\widehat{\partial}_{1j}\widehat{f'}$ using $\neq_2$, we get $\widehat{\partial}_{(uv)(1j)}\widehat{f'}$. Since the merging gadget connects two nonzero binary signatures in $\widehat{\mathcal{O}}$, the resulting signature is a nonzero binary signature. Thus, $\widehat{\partial}_{(uv)(1j)}\widehat{f'} \not\equiv 0$. Notice that $\{u,v\}$ is disjoint with $\{1,2\}$. As showed above, $\widehat{\partial}_{uv}\widehat{f'}$ has even parity. Then, $\widehat{\partial}_{(1j)(uv)}\widehat{f'}$ has odd parity. For a contradiction, suppose that $\widehat{\partial}_{1j}\widehat{f'}$ has odd parity. Then $\widehat{\partial}_{(uv)(1j)}\widehat{f'}$ has even parity. But a nonzero signature $\widehat{\partial}_{(uv)(1j)}\widehat{f'}$ cannot have both even parity and odd parity. Contradiction. Thus, $\widehat{\partial}_{1j}\widehat{f'}$ has even parity.

We have proved that $\widehat{\partial}_{ij}\widehat{f'}$ has even parity for all pairs of indices $\{i,j\}$. In other words, for all pairs of indices $\{i,j\}$ and all $\beta \in \mathbb{Z}_2^4$ with $\mathrm{wt}(\beta) = 1$ or 3, we have $(\widehat{\partial}_{ij}\widehat{f'})(\beta) = 0$. Then, by Lemma 2.10, $\widehat{f'}(\alpha) = 0$ for all α with $\mathrm{wt}(\alpha) = 2$ or 4. Clearly, $\widehat{f'}$ is realized by extending $\widehat{f}$ with nonzero binary signatures in $\widehat{\mathcal{O}}$ that are realized by factorization from $\widehat{\partial}_{12}\widehat{f}$. □

Lemma 6.2. *Suppose that $\widehat{\mathcal{F}}$ contains an irreducible 6-ary signature $\widehat{f'}$ where $\widehat{f'}(\alpha) = 0$ for all α with $\mathrm{wt}(\alpha) = 2$ or 4. Then,* Holant$(\neq_2|\ \widehat{\mathcal{F}})$ *is #P-hard, or $\mathscr{S}(\widehat{f'}) = \mathscr{O}_6 = \{\alpha \in \mathbb{Z}_2^6 \mid \mathrm{wt}(\alpha)$ is odd$\}$ and all nonzero entries of $\widehat{f'}$ have the same norm.*

Proof. Since $\widehat{f'}$ is irreducible, again we may assume that $\widehat{f'}$ satisfies 2ND-ORTH and $\widehat{f'} \in \widehat{\int}\widehat{\mathcal{O}}^{\otimes}$. Let $\{i,j,k,\ell\}$ be an arbitrarily chosen subset of indices from $\{1,\ldots,6\}$, and $\{m,n\}$ be the other two indices. Then by equation (4.7), and the condition that $\widehat{f'}$ vanishes at weight 2 and 4, we have

$$|\widehat{\mathbf{f}'}^{0001}_{ijk\ell}|^2 = |\widehat{f'}^{000100}_{ijk\ell mn}|^2 + |\widehat{f'}^{000111}_{ijk\ell mn}|^2 = |\widehat{f'}^{001000}_{ijk\ell mn}|^2 + |\widehat{f'}^{001011}_{ijk\ell mn}|^2 = |\widehat{\mathbf{f}'}^{0010}_{ijk\ell}|^2. \tag{6.2}$$

Also, by considering indices $\{k,\ell,m,n\}$, we have

$$|\widehat{\mathbf{f}'}^{0100}_{k\ell mn}|^2 = |\widehat{f'}^{000100}_{ijk\ell mn}|^2 + |\widehat{f'}^{110100}_{ijk\ell mn}|^2 = |\widehat{f'}^{001000}_{ijk\ell mn}|^2 + |\widehat{f'}^{111000}_{ijk\ell mn}|^2 = |\widehat{\mathbf{f}'}^{1000}_{k\ell mn}|^2. \tag{6.3}$$

By ARS, we have

$$|\widehat{f'}^{000111}_{ijk\ell mn}|^2 = |\widehat{f'}^{111000}_{ijk\ell mn}|^2, \tag{6.4}$$

and

$$|\widehat{f'}^{001011}_{ijk\ell mn}|^2 = |\widehat{f'}^{110100}_{ijk\ell mn}|^2. \tag{6.5}$$

By calculating (6.2) + (6.3) − (6.4) − (6.5), we have

$$|\widehat{f'}^{000100}_{ijk\ell mn}|^2 = |\widehat{f'}^{001000}_{ijk\ell mn}|^2. \tag{6.6}$$

By (6.2) − (6.6), we have

$$|\widehat{f'}^{000111}_{ijk\ell mn}|^2 = |\widehat{f'}^{001011}_{ijk\ell mn}|^2. \tag{6.7}$$

From (6.6), since the indices (i,j,k,ℓ,m,n) can be an arbitrary permutation of $(1,2,3,4,5,6)$, for all α, $\beta \in \mathbb{Z}_2^6$ with $\mathrm{wt}(\alpha) = \mathrm{wt}(\beta) = 1$, we have $|\widehat{f'}(\alpha)| = |\widehat{f'}(\beta)|$. The same statement holds for $\mathrm{wt}(\alpha) = \mathrm{wt}(\beta) = 3$, by (6.7).

Let $a = |\widehat{f'}(\vec{0}^6)|$; by ARS, $a = |\widehat{f'}(\vec{1}^6)|$ as well. It is the norm of entries of $\widehat{f'}$ on input of Hamming weight 0 and 6. We use b to denote the norm of entries of $\widehat{f'}$ on inputs of Hamming weight 1. By ARS, b is also the norm of entries of $\widehat{f'}$ on inputs of Hamming weight 5. We use c to denote the norm of entries of $\widehat{f'}$ on inputs of Hamming weight 3. Remember that by assumption, $|\widehat{f'}(\alpha)| = 0$ if $\mathrm{wt}(\alpha) = 2$ or 4.

By equation (4.6), we have

$$|\widehat{\mathbf{f}'}^{0000}_{1234}|^2 = a^2 + 2b^2 = |\widehat{\mathbf{f}'}^{0011}_{1234}|^2 = 2c^2.$$

Clearly, we have $0 \leqslant a, b \leqslant c$. If $c = 0$, then $a = b = 0$ which implies that $\widehat{f'}$ is a zero signature. This is a contradiction since $\widehat{f'}$ is irreducible. Therefore $c \neq 0$. We normalize c to 1. Then

$$a^2 + 2b^2 = 2.$$

We will show that $b = 1$ and $a = 0$. This will finish the proof of the lemma. For a contradiction, suppose that $b < 1$, then we also have $a > 0$.

Consider signatures $\widehat{f'}^{01}_{12}$, $\widehat{f'}^{10}_{12}$ and $\widehat{\partial}_{12}\widehat{f'} = \widehat{f'}^{01}_{12} + \widehat{f'}^{10}_{12}$. Since $\widehat{f'}(\alpha) = 0$ for all α with $\mathrm{wt}(\alpha) = 2$ or 4, $\widehat{f'}^{01}_{12}(\beta) = 0$ and $\widehat{f'}^{10}_{12}(\beta) = 0$ for all β with $\mathrm{wt}(\beta) = 1$ or 3. Thus, $\widehat{f'}^{01}_{12}$ and $\widehat{f'}^{10}_{12}$ have even parity. We also consider the complex inner product $\langle \widehat{\mathbf{f}'}^{01}_{12}, \widehat{\mathbf{f}'}^{10}_{12} \rangle$. First we build the following table.

In Table 2, we call these four rows by Row 1, 2, 3 and 4 respectively and these nine columns by Column 0, 1, ... and 8 respectively. We use $T_{i,j}$ to denote the cell in Row i and Column j. Table 2 is built as follows.

$\widehat{f'}_{12}^{01}$	$\widehat{f'}^{010000}$	$\widehat{f'}^{010011}$	$\widehat{f'}^{010101}$	$\widehat{f'}^{010110}$	$\widehat{f'}^{011001}$	$\widehat{f'}^{011010}$	$\widehat{f'}^{011100}$	$\widehat{f'}^{011111}$
$\widehat{f'}_{12}^{10}$	$\widehat{f'}^{100000}$	$\widehat{f'}^{100011}$	$\widehat{f'}^{100101}$	$\widehat{f'}^{100110}$	$\widehat{f'}^{101001}$	$\widehat{f'}^{101010}$	$\widehat{f'}^{101100}$	$\widehat{f'}^{101111}$
$\widehat{\partial}_{12}\widehat{f'}$	s_1	s_2	s_3	s_4	$\overline{s_4}$	$\overline{s_3}$	$\overline{s_2}$	$\overline{s_1}$
$\langle \widehat{\mathbf{f}}'^{01}_{12}, \widehat{\mathbf{f}}'^{10}_{12} \rangle$	p_1	p_2	p_3	p_4	p_4	p_3	p_2	p_1

Table 2: Entries of $\widehat{f'}_{12}^{01}$, $\widehat{f'}_{12}^{10}$, $\widehat{\partial}_{12}\widehat{f'}$ and pairwise product terms in $\langle \widehat{\mathbf{f}}'^{01}_{12}, \widehat{\mathbf{f}}'^{10}_{12} \rangle$ on even-weighed inputs

- In Row 1 and Row 2, we list the entries of signatures $\widehat{f'}_{12}^{01}$ and $\widehat{f'}_{12}^{10}$ that are on even-weighted inputs (excluding the first two bits that are pinned) respectively. Note that, those that did not appear are 0 entries on odd-weighted inputs (excluding the first two bits that are pinned) of the signatures $\widehat{f'}_{12}^{01}$ and $\widehat{f'}_{12}^{10}$, since $\widehat{f'}_{12}^{01}$ and $\widehat{f'}_{12}^{10}$ have even parity.
- In Row 3, we list the corresponding entries of the signature $\widehat{\partial}_{12}\widehat{f'} = \widehat{f'}_{12}^{01} + \widehat{f'}_{12}^{10}$, i.e., $T_{3,j} = T_{1,j} + T_{2,j}$ for $1 \leqslant j \leqslant 8$.
- In Row 4, we list the corresponding items in the complex inner product $\langle \widehat{\mathbf{f}}'^{01}_{12}, \widehat{\mathbf{f}}'^{10}_{12} \rangle$, i.e., $T_{4,j} = T_{1,j} \cdot \overline{T_{2,j}}$ for $1 \leqslant j \leqslant 8$.

For $1 \leqslant j \leqslant 8$, we consider the entry in $T_{1,j}$ and the entry in $T_{2,9-j}$. By ARS, we have $T_{1,j} = \overline{T_{2,9-j}}$ because their corresponding inputs are complement of each other. Thus,

$$T_{3,j} = T_{1,j} + T_{2,j} = \overline{T_{2,9-j}} + \overline{T_{1,9-j}} = \overline{T_{3,9-j}},$$

and

$$T_{4,j} = T_{1,j} \cdot \overline{T_{2,j}} = \overline{T_{2,9-j}} \cdot T_{2,9-j} = T_{4,9-j}.$$

We use $s_1, \ldots, s_4$ to denote the values in $T_{3,1}, \ldots, T_{3,4}$ and $p_1, \ldots, p_4$ to denote the values in $T_{4,1}, \ldots, T_{4,4}$. Correspondingly, the values in $T_{3,5}, \ldots, T_{3,8}$ are $\overline{s_4}, \ldots, \overline{s_1}$ and the values in $T_{4,5}, \ldots, T_{4,8}$ are $p_4, \ldots, p_1$. We also use x_j and y_j $(1 \leqslant j \leqslant 8)$ to denote the entries in $T_{1,j}$ and $T_{2,j}$ respectively.

By 2ND-ORTH, we have $\langle \widehat{\mathbf{f}}'^{01}_{12}, \widehat{\mathbf{f}}'^{10}_{12} \rangle = 2(p_1 + p_2 + p_3 + p_4) = 0$. Also we have $|p_1| = b^2$ and $|p_2| = |p_3| = |p_4| = 1$. Notice the fact that if $x_i + y_i = 0$, then $x_i \cdot \overline{y_i} = x_i \cdot \overline{-x_i} = -|x_i|^2 = -|x_i \cdot \overline{y_i}|$. Thus, if $s_1 = 0$ then $p_1 = -|p_1| = -b^2$ and for any $i = 2, 3, 4$, if $s_i = 0$ then $p_i = -1$. Note that $\widehat{\partial}_{12}\widehat{f'}(\beta) = \widehat{f'}_{12}^{01}(\beta) + \widehat{f'}_{12}^{10}(\beta) = 0$ for all β with $\mathrm{wt}(\beta) = 1$ or 3. Among all 16 entries of $\widehat{\partial}_{12}\widehat{f'}$, $s_1, \ldots, s_4, \overline{s_4}, \ldots, \overline{s_1}$ are those that are possibly nonzero. Since $\widehat{\partial}_{12}\widehat{f'} \in \widehat{\mathcal{O}}^{\otimes}$, it has support of size either 4 or 0. Thus, among s_1, s_2, s_3 and s_4, either exactly two of them are zero or they are all zero. There are three possible cases.

- $s_1 = s_2 = s_3 = s_4 = 0$. Then $p_1 + p_2 + p_3 + p_4 = -b^2 - 3 \leqslant -3 \neq 0$. Contradiction.
- $s_1 \neq 0$ and two of s_2, s_3 and s_4 are zero. Without loss of generality, we may assume that $s_2 = s_3 = 0$. Then $p_2 = p_3 = -1$. Since $p_1 + p_2 + p_3 + p_4 = 0$, we have $p_1 + p_4 = -p_2 - p_3 = 2$. Then, $2 = |p_1 + p_4| \leqslant |p_1| + |p_4| = b^2 + 1 < 2$. Contradiction.
- $s_1 = 0$ and one of s_2, s_3 and s_4 is zero. Without loss of generality, we may assume that $s_2 = 0$. Then $p_1 = -b^2$ and $p_2 = -1$. Thus, $p_3 + p_4 = -p_1 - p_2 = 1 + b^2 < 2$. Let $\theta = \arccos \frac{1+b^2}{2}$.

We know that $0 < \theta < \frac{\pi}{2}$. Recall that $|p_3| = |p_4| = 1$. Thus, $p_3 = e^{\pm \mathrm{i}\theta}$ and $p_4 = e^{\mp \mathrm{i}\theta}$ (and $p_3 = \overline{p_4}$).

Let $P = \{-1, e^{\mathrm{i}\theta}, e^{-\mathrm{i}\theta}\}$. Thus, $p_2, p_3, p_4 \in P$. Otherwise, we get a contradiction.

Now, we consider signatures $\widehat{\partial}_{ij}\widehat{f'}$ for all pairs of indices $\{i, j\}$. By symmetry, the same conclusion holds. In other words, let $\{i, j\}$ be an arbitrarily chosen pair of indices from $\{1, \ldots, 6\}$ and $\{k, \ell, m, n\}$ be the other four indices, and let $\beta \in \mathbb{Z}_2^4$ be an assignment on variables (x_k, x_ℓ, x_m, x_n) with $\mathrm{wt}(\beta) = 2$. Then, we have $\widehat{f'}^{01\beta}_{ijk\ell mn} \cdot \overline{\widehat{f'}^{10\beta}_{ijk\ell mn}} \in P$. Since the indices (i, j, k, ℓ, m, n) can be an arbitrary permutation of $(1, 2, 3, 4, 5, 6)$, we have $\widehat{f'}(\alpha) \cdot \overline{\widehat{f'}(\alpha')} \in P$ for any two assignments α and α' on the six variables where $\mathrm{wt}(\alpha) = \mathrm{wt}(\alpha') = 3$ and $\mathrm{wt}(\alpha \oplus \alpha') = 2$, because for any such two strings α and α', there exist two bit positions on which α and α' take values 01 and 10 respectively.

We consider the following three inputs $\alpha_1 = 100011$, $\alpha_2 = 010011$ and $\alpha_3 = 001011$ of $\widehat{f'}$. We have $\widehat{f'}(\alpha_1) \cdot \overline{\widehat{f'}(\alpha_2)} = q_{12} \in P$, $\widehat{f'}(\alpha_2) \cdot \overline{\widehat{f'}(\alpha_3)} = q_{23} \in P$ and $\widehat{f'}(\alpha_1) \cdot \overline{\widehat{f'}(\alpha_3)} = q_{13} \in P$. Recall that $|\widehat{f'}(\alpha_2)| = 1$ since $\mathrm{wt}(\alpha_2) = 3$. Then,

$$q_{12} \cdot q_{23} = \widehat{f'}(\alpha_1) \cdot \overline{\widehat{f'}(\alpha_2)} \cdot \widehat{f'}(\alpha_2) \cdot \overline{\widehat{f'}(\alpha_3)} = |\widehat{f'}(\alpha_2)|^2 \cdot \widehat{f'}(\alpha_1) \cdot \overline{\widehat{f'}(\alpha_3)} = q_{13} \in P.$$

However, since $0 < \theta < \frac{\pi}{2}$, it is easy to check that for any two (not necessarily distinct) elements in P, their product is not in P. Thus, we get a contradiction. This proves that $b = c = 1$ and $a = 0$.

Therefore we have proved that, $\mathscr{S}(\widehat{f'}) = \mathscr{O}_6$, and all its nonzero entries have the same norm that is normalized to 1. □

Lemma 6.3. *Suppose that $\widehat{\mathcal{F}}$ contains an irreducible 6-ary signature $\widehat{f'}$ where $\mathscr{S}(\widehat{f'}) = \mathscr{O}_6$ and $|\widehat{f'}(\alpha)| = 1$ for all $\alpha \in \mathscr{S}(\widehat{f'})$. Then,* $\mathrm{Holant}(\neq_2 | \widehat{\mathcal{F}})$ *is #P-hard, or after a holographic transformation by some $\widehat{Q} = \begin{bmatrix} \overline{\rho} & 0 \\ 0 & \rho \end{bmatrix} \in \widehat{\mathbf{O}_2}$ where $\rho = e^{\mathrm{i}\delta}$ and $0 \leqslant \delta < \pi/2$, an irreducible 6-ary signature $\widehat{f''}$ and $=_2$ are realizable from $\widehat{f'}$ where $\mathscr{S}(\widehat{f''}) = \mathscr{O}_6$ and there exists $\lambda = 1$ or i such that for all $\alpha \in \mathscr{S}(\widehat{f''})$, $\widehat{f''}(\alpha) = \pm\lambda$, i.e.,* $\mathrm{Holant}(\neq_2 | =_2, \widehat{f''}, \widehat{Q}\widehat{\mathcal{F}}) \leqslant_T \mathrm{Holant}(\neq_2 | \widehat{\mathcal{F}})$ *where $\widehat{f''} = \widehat{Q}\widehat{f'}$. Moreover, the nonzero binary signature $(\rho^2, 0, 0, \overline{\rho^2}) \in \widehat{\mathcal{O}}$ is realizable from $\widehat{\partial}_{ij}\widehat{f'}$ for some $\{i, j\}$.*

Proof. Again, we may assume that $\widehat{f'}$ satisfies 2ND-ORTH and $\widehat{f'} \in \widehat{\int}\widehat{\mathcal{O}}^{\otimes}$. We first show that there exists $\lambda = 1$ or i such that for all $\alpha \in \mathscr{S}(\widehat{f'})$ with $\mathrm{wt}(\alpha) = 3$, $\widehat{f''}(\alpha) = \pm\lambda$, or else we get #P-hardness.

Let's revisit Table 2. Now we have $|p_1| = |p_2| = |p_3| = |p_4| = 1$. Recall that for $1 \leqslant i \leqslant 4$, $s_i = 0$ implies that $p_i = -1$. Since $\widehat{\partial}_{12}\widehat{f'} \in \widehat{\mathcal{O}}^{\otimes 2}$, it has support of size 4 or 0. Thus, among s_1, s_2, s_3 and s_4, either exactly two of them are zero or they are all zero. If they are all zero, then we have $p_1 + p_2 + p_3 + p_4 = -4 \neq 0$. This is a contradiction to our assumption that $\widehat{f'}$ satisfies 2ND-ORTH. Thus, exactly two of s_1, s_2, s_3 and s_4 are zeros. Suppose that they are s_i and s_j. Recall that we use x_i and y_i $(1 \leqslant i \leqslant 8)$ to denote the entries in Row 1 and Row 2 of Table 2. Thus $|x_i| = |y_i| = 1$, for $1 \leqslant i \leqslant 8$. Since $s_i = x_i + y_i = 0$ and $s_j = x_j + y_j = 0$, we have $x_i = -y_i$, and $x_j = -y_j$. Also, since $s_i = s_j = 0$, we have $p_i = p_j = -1$. Let $\{\ell, k\} = \{1, 2, 3, 4\} \backslash \{i, j\}$. Then, by 2ND-ORTH, we have $p_\ell + p_k = -p_i - p_j = 2$. Since $|p_\ell| = |p_k| = 1$, we have $p_\ell = p_k = 1$. Note that $p_\ell = x_\ell \cdot \overline{y_\ell} = 1$ and also $1 = |y_\ell| = y_\ell \cdot \overline{y_\ell}$. Thus, we have $x_\ell = y_\ell$. Similarly, $x_k = y_k$. Thus, for all $1 \leqslant i \leqslant 8$, $x_i = \pm y_i$. Consider $\widehat{\partial}_{ij}\widehat{f'}$ for all pairs of indices $\{i, j\}$. By symmetry, the same conclusion holds. Thus, $\widehat{f}(\alpha) = \pm\widehat{f}(\alpha')$ for any two inputs α and α' on the six variables where $\mathrm{wt}(\alpha) = \mathrm{wt}(\alpha') = 3$ and $\mathrm{wt}(\alpha \oplus \alpha') = 2$. In particular, we have

$$\widehat{f'}^{000111} = \varepsilon_1 \widehat{f'}^{001011} = \varepsilon_2 \widehat{f'}^{011001} = \varepsilon_3 \widehat{f'}^{111000},$$

where $\varepsilon_1, \varepsilon_2, \varepsilon_3 = \pm 1$ independently. By ARS, we have $\widehat{f'}^{000111} = \overline{\widehat{f'}^{111000}}$.

- If $\widehat{f'}^{000111} = \widehat{f'}^{111000} = \overline{\widehat{f'}^{111000}}$, then $\widehat{f'}^{111000} = \pm 1$.
- If $\widehat{f'}^{000111} = -\widehat{f'}^{111000} = \overline{\widehat{f'}^{111000}}$, then $\widehat{f'}^{111000} = \pm \mathrm{i}$.

Thus, there exists $\lambda = 1$ or i such that $\widehat{f'}^{000111} = \pm\lambda$ and $\widehat{f'}^{111000} = \pm\lambda$. Consider any $\alpha \in \mathbb{Z}_2^6$ with $\mathrm{wt}(\alpha) = 3$. If $\alpha \in \{000111, 111000\}$, then clearly, $\widehat{f'}(\alpha) = \pm\lambda$. Otherwise, either $\mathrm{wt}(\alpha \oplus 000111) = 2$ or $\mathrm{wt}(\alpha \oplus 111000) = 2$. Then, $\widehat{f'}(\alpha) = \pm\lambda$. Thus, there exists $\lambda = 1$ or i such that for all $\alpha \in \mathbb{Z}_2^6$ with $\mathrm{wt}(\alpha) = 3$, $\widehat{f'}(\alpha) = \pm\lambda$.

Since $\widehat{f'}(\alpha) \neq 0$ for all α with $\mathrm{wt}(\alpha) = 1$, by Lemma 2.10, there exists a pair of indices $\{i, j\}$ such that $(\widehat{\partial}_{ij}\widehat{f'})^{0000} \neq 0$. Since $\widehat{\partial}_{ij}\widehat{f'} \in \widehat{\mathcal{O}}^{\otimes}$, it is of the form $(a, 0, 0, \bar{a}) \otimes (b, 0, 0, \bar{b})$, where $ab \neq 0$, since no other factorization form in $\widehat{\mathcal{O}}^{\otimes}$ has a nonzero value at 0000. By Lemma 2.7, we can realize the signature $\widehat{g} = (a, 0, 0, \bar{a})$. Here, we can normalize a to $e^{\mathrm{i}\theta}$ where $0 \leqslant \theta < \pi$. Then, let $\rho = e^{\mathrm{i}\theta/2}$. Clearly, $0 \leqslant \theta/2 < \pi/2$. Consider a holographic transformation by $\widehat{Q} = \left[\begin{smallmatrix} \overline{\rho} & 0 \\ 0 & \rho \end{smallmatrix}\right]$. Note that $(\neq_2)(\widehat{Q}^{-1})^{\otimes 2} = (\neq_2)$ and $\widehat{Q}^{\otimes 2}\widehat{g} = (1, 0, 0, 1)$. The holographic transformation by $\widehat{Q}$ does not change $\neq_2$, but transfers $\widehat{g} = (a, 0, 0, \bar{a})$ to $(=_2) = (1, 0, 0, 1)$. Thus, we have

$$\mathrm{Holant}(\neq_2 |\, \widehat{g}, \widehat{f'}, \widehat{\mathcal{F}}) \equiv_T \mathrm{Holant}(\neq_2 | =_2, \widehat{Q}\widehat{f'}, \widehat{Q}\widehat{\mathcal{F}}).$$

We denote $\widehat{Q}\widehat{f'}$ by $\widehat{f''}$. Note that $\widehat{Q}$ does not change those entries of $\widehat{f'}$ that are on half-weighted inputs. Thus, for all α with $\mathrm{wt}(\alpha) = 3$, we have $\widehat{f''}(\alpha) = \pm\lambda$ for some $\lambda = 1$ or i. Also, $\widehat{Q}$ does not change the parity and irreducibility of $\widehat{f'}$. Thus $\widehat{f''}$ has odd parity and $\widehat{f''}$ is irreducible. Again, we may assume that $\widehat{f''}$ satisfies 2ND-ORTH and $\widehat{f''} \in \widehat{\int}\widehat{\mathcal{O}}^{\otimes}$. Otherwise, we are done.

In the problem $\mathrm{Holant}(\neq_2 | =_2, \widehat{f''}, \widehat{Q}\widehat{\mathcal{F}})$, we can connect two $\neq_2$ on the LHS using $=_2$ on the RHS, and then we can realize $=_2$ on the LHS. Thus, we can use $=_2$ to merge variables of $\widehat{f''}$. Therefore, we may further assume $\widehat{f''} \in \widehat{\int}\widehat{\mathcal{O}}^{\otimes}$, i.e., $\partial_{ij}\widehat{f''} \in \widehat{\mathcal{O}}^{\otimes}$ for all pairs of indices $\{i, j\}$; otherwise, there exist two variables of $\widehat{f''}$ such that by merging these two variables using $=_2$, we can realize a 4-ary signature that is not in $\widehat{\mathcal{O}}^{\otimes}$, and then by Lemma 5.2 we are done.

Consider the signature $\partial_{12}\widehat{f''} = \widehat{f''}^{00}_{12} + \widehat{f''}^{11}_{12}$ and the inner product $\langle \widehat{\mathbf{f}''}^{00}_{12}, \widehat{\mathbf{f}''}^{11}_{12} \rangle$. Same as Table 2, we build the following Table 3.

$\widehat{f''}^{00}_{12}$	$\widehat{f''}^{000001}$	$\widehat{f''}^{000010}$	$\widehat{f''}^{000100}$	$\widehat{f''}^{000111}$	$\widehat{f''}^{001000}$	$\widehat{f''}^{001011}$	$\widehat{f''}^{001101}$	$\widehat{f''}^{001110}$
$\widehat{f''}^{11}_{12}$	$\widehat{f''}^{110001}$	$\widehat{f''}^{110010}$	$\widehat{f''}^{110100}$	$\widehat{f''}^{110111}$	$\widehat{f''}^{111000}$	$\widehat{f''}^{111011}$	$\widehat{f''}^{111101}$	$\widehat{f''}^{111110}$
$\partial_{12}\widehat{f''}$	t_1	t_2	t_3	t_4	$\overline{t_4}$	$\overline{t_3}$	$\overline{t_2}$	$\overline{t_1}$
$\langle \widehat{\mathbf{f}''}^{00}_{12}, \widehat{\mathbf{f}''}^{11}_{12} \rangle$	q_1	q_2	q_3	q_4	q_4	q_3	q_2	q_1

Table 3: Entries of $\widehat{f''}^{00}_{12}$, $\widehat{f''}^{11}_{12}$, $\partial_{12}\widehat{f''}$ and pair-wise product terms in $\langle \widehat{\mathbf{f}''}^{00}_{12}, \widehat{\mathbf{f}''}^{11}_{12} \rangle$ on odd-weighed inputs

Same as the proof of $x_i = \pm y_i$ for Table 2, we have $\widehat{f''}^{000001} = \pm\widehat{f''}^{110001}$. Since $\widehat{f''}^{110001} = \pm\lambda$, $\widehat{f''}^{000001} = \pm\lambda$, (here $\pm$ can be either $\pm$ or $\mp$). Consider $\partial_{ij}\widehat{f''}$ for all pairs of indices $\{i, j\}$. By

symmetry, the same conclusion holds. Thus, for every $\alpha \in \mathbb{Z}_2^6$ with $\mathrm{wt}(\alpha) = 1$, $\widehat{f''}(\alpha) = \pm\lambda$. Therefore, using ARS, there exists $\lambda = 1$ or $\mathbf{i}$ such that for all $\alpha \in \mathscr{S}(\widehat{f''})$, $\widehat{f''}(\alpha) = \pm\lambda$, and we have the reduction

$$\mathrm{Holant}(\neq_2|=_2, \widehat{f''}, \widehat{Q}\widehat{\mathcal{F}}) \leqslant_T \mathrm{Holant}(\neq_2| \widehat{\mathcal{F}})$$

for some $\widehat{Q} \in \widehat{\mathbf{O}_2}$. Clearly, $\widehat{f''} = \widehat{Q}\widehat{f'}$ where $\widehat{Q} = \left[\begin{smallmatrix} \bar{\rho} & 0 \\ 0 & \rho \end{smallmatrix}\right] \in \widehat{\mathbf{O}_2}$, and the nonzero binary signature $(\rho^2, 0, 0, \overline{\rho^2}) \in \widehat{\mathcal{O}}$ is realizable from $\widehat{\partial}_{ij}\widehat{f'}$ for some $\{i, j\}$. □

Finally, we go for the kill in the next lemma. Recall the signature $\widehat{f_6}$ defined in (6.1). This *Lord of Intransigence* at arity 6 makes its appearance in Lemma 6.4.

Lemma 6.4. *Suppose that $\widehat{\mathcal{F}}$ contains an irreducible 6-ary signature $\widehat{f''}$ where $\mathscr{S}(\widehat{f''}) = \mathscr{O}_6$, and there exists $\lambda = 1$ or $\mathbf{i}$ such that for all $\alpha \in \mathscr{S}(\widehat{f''})$, $\widehat{f''}(\alpha) = \pm\lambda$. Then,* $\mathrm{Holant}(\neq_2|=_2, \widehat{\mathcal{F}})$ *is #P-hard, or $\widehat{f_6}$ is realizable from $\widehat{f''}$ and $=_2$, i.e.,* $\mathrm{Holant}(\neq_2| \widehat{f_6}, \widehat{\mathcal{F}}) \leqslant_T \mathrm{Holant}(\neq_2|=_2, \widehat{\mathcal{F}})$. *Moreover, $\widehat{f_6}$ is realizable by extending variables of $\widehat{f''}$ with binary signatures in $\widehat{\mathcal{B}}$, i.e., $\widehat{f_6} \in \{\widehat{f''}\}^{\widehat{\mathcal{B}}}_{\neq_2}$.*

Proof. Again, we may assume that $\widehat{f''}$ satisfies 2ND-ORTH and $\widehat{f''} \in \widehat{\int}\widehat{\mathcal{O}}^{\otimes}$. Since $=_2$ is available on the RHS, given any signature $\widehat{f} \in \widehat{\mathcal{F}}$, we can extend any variable x_i of $\widehat{f}$ with $=_2 \in \widehat{\mathcal{B}}$ using $\neq_2$. This gives a signature $\widehat{g}$ where $\widehat{g}_i^0 = \widehat{f}_i^1$ and $\widehat{g}_i^1 = \widehat{f}_i^0$. We call this extending gadget construction the flipping operation on variable x_i. Clearly, it does not change the reducibility or irreducibility of $\widehat{f}$. But it changes the parity of $\widehat{f}$ if $\widehat{f}$ has parity. Once a signature $\widehat{f}$ is realizable, we can modify it by flipping some of its variables.

We first show that we can realize a signature $\widehat{f^*}$ from $\widehat{f''}$ having support $\mathscr{S}(\widehat{f^*}) = \mathscr{E}_6 = \{\alpha \in \mathbb{Z}_2^6 \mid \mathrm{wt}(\alpha) \equiv 0 \mod 2\}$, and $\widehat{f^*}(\alpha) = \pm 1$ for all $\alpha \in \mathscr{S}(\widehat{f^*})$. Remember that $=_2$ is available. If we connect $=_2$ with an arbitrary variable of $\widehat{f''}$ using $\neq_2$, we will change the parity of $\widehat{f''}$ from odd to even. If $\widehat{f''}(\alpha) = \pm 1$ for all $\alpha \in \mathscr{S}(\widehat{f''})$, then $\widehat{f^*}$ can be realized by flipping an arbitrary variable of $\widehat{f''}$. Otherwise, $\widehat{f''}(\alpha) = \pm\mathbf{i}$ for all $\alpha \in \mathscr{S}(\widehat{f''})$. Consider $\widehat{\partial}_{12}\widehat{f''}$. Look at Table 3. We use x_i and y_i $(1 \leqslant i \leqslant 8)$ to denote entries in Row 1 and 2. As we have showed, $x_i = \pm y_i$. Thus, $t_i = \pm 2\mathbf{i}$ or 0 for $1 \leqslant i \leqslant 4$. Remember that if $t_i = 0$ (i.e., $x_i = -y_i$), then $q_i = x_i \cdot \overline{y_i} = -x_i \cdot \overline{x_i} = -|x_i|^2 = -1$. If $t_i = 0$ for all $1 \leqslant i \leqslant 4$, then

$$\langle \widehat{\mathbf{f}''}_{12}^{00}, \widehat{\mathbf{f}''}_{12}^{11} \rangle = 2(q_1 + q_2 + q_3 + q_4) = -4 \neq 0.$$

This contradicts with our assumption that $\widehat{f''}$ satisfies 2ND-ORTH. Thus, t_i $(1 \leqslant i \leqslant 4)$ are not all zeros. Then $(\widehat{\partial}_{12}\widehat{f''}) \not\equiv 0$. Thus, $\mathscr{S}(\widehat{\partial}_{12}\widehat{f''}) \neq \emptyset$ and $(\widehat{\partial}_{12}\widehat{f''})(\alpha) = \pm 2i$ for all $\alpha \in \mathscr{S}(\widehat{\partial}_{12}\widehat{f''})$.

Since $\widehat{\partial}_{12}\widehat{f''} \in \widehat{\mathcal{O}}^{\otimes}$ and it has even parity, $\widehat{\partial}_{12}\widehat{f''}$ is of the form $2 \cdot (a, 0, 0, \bar{a}) \otimes (b, 0, 0, \bar{b})$ or $2 \cdot (0, a, \bar{a}, 0) \otimes (0, b, \bar{b}, 0)$, where the norms of a and b are normalized to 1. In both cases, we have $ab, \bar{a}b, a\bar{b}, \bar{a}\bar{b} \in \{\mathbf{i}, -\mathbf{i}\}$. Thus, $ab \cdot \bar{a}b = (a\bar{a})b^2 = b^2 = \pm 1$. Then, $b = \pm 1$ or $\pm\mathbf{i}$. If $b = \pm 1$, then $a = a\bar{b} \cdot b = \pm\mathbf{i}$. Similarly, if $b = \pm\mathbf{i}$, then $a = a\bar{b} \cdot b = \pm 1$. Thus, among a and b, exactly one is $\pm\mathbf{i}$. Thus, by factorization we can realize the binary signature $\widehat{g} = (\mathbf{i}, 0, 0, -\mathbf{i})$ or $(0, \mathbf{i}, -\mathbf{i}, 0)$ up to a scalar -1. Connecting an arbitrary variable of $\widehat{f}$ with a variable of $\widehat{g}$, we can get a signature which has parity and all its nonzero entries have value ± 1. If the resulting signature has even parity, then we get the desired $\widehat{f^*}$. If it has odd parity, then we can flip one of its variables to change the parity. Thus, we can realize a signature $\widehat{f^*}$ by extending variables of $\widehat{f''}$ with binary signatures in $\widehat{\mathcal{B}}^{\otimes}$ such that $\mathscr{S}(\widehat{f^*}) = \mathscr{E}_6$, and $\widehat{f^*}(\alpha) = \pm 1$ for all $\alpha \in \mathscr{S}(\widehat{f^*})$.

Consider the following 16 entries of $\widehat{f^*}$. In Table 4, we list 16 entries of $\widehat{f^*}$ with $x_1x_2x_3 = 000, 011, 101, 110$ as the row index and $x_4x_5x_6 = 000, 011, 101, 110$ as the column index. We also view these 16 entries in Table 4 as a 4-by-4 matrix denoted by $M_r(\widehat{f^*})$, and we call it the representative matrix of $\widehat{f^*}$. Note that for any $\alpha \in \mathscr{S}(\widehat{f^*})$ such that the entry $\widehat{f^*}(\alpha)$ does not appear in $M_r(\widehat{f^*})$, $\widehat{f^*}(\overline{\alpha})$ appears in $M_r(\widehat{f^*})$. Since $\widehat{f^*}(\alpha) = \pm 1 \in \mathbb{R}$, $\overline{\widehat{f^*}(\alpha)} = \widehat{f^*}(\alpha)$. By ARS, $\widehat{f^*}(\overline{\alpha}) = \overline{\widehat{f^*}(\alpha)} = \widehat{f^*}(\alpha)$. Thus, the 16 entries of the matrix $M_r(\widehat{f^*})$ listed in Table 4 gives a complete account for all the 32 nonzero entries of $\widehat{f^*}$.

$x_1x_2x_3$ \ $x_4x_5x_6$	000 (Col 1)	011 (Col 2)	101 (Col 3)	110 (Col 4)
000 (Row 1)	$\widehat{f^*}^{000000}$	$\widehat{f^*}^{000011}$	$\widehat{f^*}^{000101}$	$\widehat{f^*}^{000110}$
011 (Row 2)	$\widehat{f^*}^{011000}$	$\widehat{f^*}^{011011}$	$\widehat{f^*}^{011101}$	$\widehat{f^*}^{011110}$
101 (Row 3)	$\widehat{f^*}^{101000}$	$\widehat{f^*}^{101011}$	$\widehat{f^*}^{101101}$	$\widehat{f^*}^{101110}$
110 (Row 4)	$\widehat{f^*}^{110000}$	$\widehat{f^*}^{110011}$	$\widehat{f^*}^{110101}$	$\widehat{f^*}^{110110}$

Table 4: Representative entries of $\widehat{f^*}$

We use $(m_{ij})_{i,j=1}^4$ to denote the 16 entries of $M_r(\widehat{f^*})$. We claim that any two rows of $M_r(\widehat{f^*})$ are orthogonal; this follows from the fact that $\widehat{f^*}$ satisfies 2ND-ORTH and ARS. For example, consider the first two rows of $M_r(\widehat{f^*})$. By 2ND-ORTH, the inner product $\langle \widehat{\mathbf{f}^*}_{23}^{00}, \widehat{\mathbf{f}^*}_{23}^{11} \rangle$ for the real-valued $\widehat{f^*}$ is

$$\sum_{(x_1,x_4,x_5,x_6)\in\mathbb{Z}_2^4} \widehat{f^*}^{x_1 00 x_4x_5x_6} \widehat{f^*}^{x_1 11 x_4x_5x_6} = 0,$$

where the sum has 8 nonzero product terms. The first 4 terms given by $x_1 = 0$ are the pairwise products $m_{1j}m_{2j}$, for $1 \leqslant j \leqslant 4$. The second 4 terms are, by ARS, the pairwise products $m_{2j}m_{1j}$ in the reversal order of $1 \leqslant j \leqslant 4$, where we exchange row 1 with row 2 on the account of flipping the summation index x_1 from 0 to 1, and simultaneously flipping both x_2 and x_3. This shows that $\sum_{j=1}^4 m_{1j}m_{2j} = 0$. Similarly any two columns of $M_r(\widehat{f^*})$ are orthogonal.

Also, we consider the inner product $\langle \widehat{\mathbf{f}^*}_{14}^{00}, \widehat{\mathbf{f}^*}_{14}^{11} \rangle = 0$. It is computed using the following 16 entries in $M_r(\widehat{f^*})$, listed in Table 5.

$\widehat{f^*}^{000000} = m_{11}$	$\widehat{f^*}^{000011} = m_{12}$	$\widehat{f^*}^{010010} = m_{33}$	$\widehat{f^*}^{010001} = m_{34}$	$\widehat{f^*}^{001010} = m_{43}$	$\widehat{f^*}^{001001} = m_{44}$	$\widehat{f^*}^{011000} = m_{21}$	$\widehat{f^*}^{011011} = m_{22}$
$\widehat{f^*}^{100100} = m_{22}$	$\widehat{f^*}^{100111} = m_{21}$	$\widehat{f^*}^{110110} = m_{44}$	$\widehat{f^*}^{110101} = m_{43}$	$\widehat{f^*}^{101110} = m_{34}$	$\widehat{f^*}^{101101} = m_{33}$	$\widehat{f^*}^{111100} = m_{12}$	$\widehat{f^*}^{111111} = m_{11}$

Table 5: Pair-wise product terms in $\langle \widehat{\mathbf{f}^*}_{14}^{00}, \widehat{\mathbf{f}^*}_{14}^{11} \rangle$ on even-weighed inputs

Let $M_r(\widehat{f^*})_{[1,2]}$ be the 2-by-2 submatrix of $M_r(\widehat{f^*})$ by picking the first two rows and the first two columns, and $M_r(\widehat{f^*})_{[3,4]}$ be the 2-by-2 submatrix of $M_r(\widehat{f^*})$ by picking the last two rows and the last two columns. Indeed,

$$\begin{aligned}\langle \widehat{\mathbf{f}^*}_{14}^{00}, \widehat{\mathbf{f}^*}_{14}^{11} \rangle &= 2(\mathrm{perm}(M_r(\widehat{f^*})_{[1,2]}) + \mathrm{perm}(M_r(\widehat{f^*})_{[3,4]})) \\ &= 2(m_{11}m_{22} + m_{12}m_{21} + m_{33}m_{44} + m_{34}m_{43}) = 0.\end{aligned}$$

Then, we show that by renaming or flipping variables of $\widehat{f^*}$, we may modify $\widehat{f^*}$ to realize a signature whose representative matrix is obtained by performing row permutation, column permutation, or matrix transpose on $M_r(\widehat{f^*})$. First, if we exchange the names of variables (x_1, x_2, x_3) with variables (x_4, x_5, x_6), then the representative matrix $M_r(\widehat{f^*})$ will be transposed. Next, consider the group $\mathfrak{G}$ of permutations on the rows $\{1, 2, 3, 4\}$ effected by any sequence of operations of renaming and flipping variables in $\{x_1, x_2, x_3\}$. By renaming variables in $\{x_1, x_2, x_3\}$, we can switch any two rows among Row 2, 3 and 4. Thus S_3 on $\{2, 3, 4\}$ is contained in $\mathfrak{G}$. Also, if we flip both variables x_2 and x_3 of $\widehat{f^*}$, then for the realized signature, its representative matrix can be obtained by switching both the pair Row 1 and Row 2, and the pair Row 3 and Row 4 of $M_r(\widehat{f^*})$. Thus, the permutation $(12)(34) \in \mathfrak{G}$. It follows that $\mathfrak{G} = S_4$. Thus, by renaming or flipping variables of $\widehat{f^*}$, we can permute any two rows or any two columns of $M_r(\widehat{f^*})$, or transpose $M_r(\widehat{f^*})$. For the resulting signature, we may assume that its representative matrix A also satisfy $\mathrm{perm}(A_{[1,2]}) + \mathrm{perm}(A_{[3,4]}) = 0$, and any two rows of A are orthogonal and any two columns of A are orthogonal. Otherwise, we get #P-hardness. In the following, without loss of generality, we may modify $M_r(\widehat{f^*})$ by permuting any two rows or any two columns, or taking transpose. We show that it will give $M_r(\widehat{f_6})$, after a normalization by ± 1. In other words, $\widehat{f_6}$ is realizable from $\widehat{f^*}$ by renaming or flipping variables, up to a normalization by ± 1.

Consider any two rows, Row i and Row j, of $M_r(\widehat{f^*})$. Recall that every entry of $M_r(\widehat{f^*})$ is ± 1. We say that Row i and Row j differ in Column k if $m_{ik} \neq m_{jk}$, which implies that $m_{ik} = -m_{jk}$; otherwise, they are equal $m_{ik} = m_{jk}$. In the former case, $m_{ik} \cdot m_{jk} = -1$, and in the latter case $m_{ik} \cdot m_{jk} = 1$. Since Row i and Row j are orthogonal, they differ in exactly two columns and are equal in the other two columns. Similarly, for any two columns of $M_r(\widehat{f^*})$, they differ in exactly two rows and are equal in the other two rows. Depending on the number of -1 entries in each row and column of $M_r(\widehat{f^*})$, we consider the following two cases.

- Every row and column of $M_r(\widehat{f^*})$ has an odd number of -1 entries.
  Consider Row 1. It has either exactly three -1 entries or exactly one -1 entry. If it has three -1 entries, then we modify $M_r(\widehat{f^*})$ by multiplying the matrix with -1. This does not change the parity of the number of -1 entries in each row and each column. By such a modification, Row 1 has exactly one -1 entry. By permuting columns, we may assume that Row 1 is $(-1, 1, 1, 1)$. Consider the number of -1 entries in Rows 2, 3 and 4.
  - If they all have exactly one -1 entry, by orthogonality, the unique column locations of the -1 entry in each row must be pairwise distinct. Then, by possibly permuting rows 2, 3 and 4 we may assume that the matrix $M_r(\widehat{f^*})$ has the following form

$$M_r(\widehat{f^*}) = \begin{bmatrix} -1 & 1 & 1 & 1 \\ 1 & -1 & 1 & 1 \\ 1 & 1 & -1 & 1 \\ 1 & 1 & 1 & -1 \end{bmatrix}.$$

Then, $\operatorname{perm}(M_r(\widehat{f^*})_{[1,2]}) + \operatorname{perm}(M_r(\widehat{f^*})_{[3,4]}) = 2 + 2 = 4 \neq 0$. Contradiction.

- Otherwise, among Rows 2, 3 and 4, there is one that has three -1 entries. By permuting rows, we may assume that Row 2 has three -1 entries. Since Row 2 and Row 1 differ in two columns, the only $+1$ entry in Row 2 is not in Column 1. By possibly permuting Columns 2, 3 and 4, without loss of generality, we may assume that Row 2 is $(-1, 1, -1, -1)$. Then, we consider Column 3 and Column 4. Since every column has an odd number of -1 entries and $m_{13} = 1$ and $m_{23} = -1$, we have $m_{33} = m_{43}$, both $+1$ or -1. Similarly, $m_{34} = m_{44}$. Also, since Column 3 and Column 4 differ in exactly two rows, and $m_{13} = m_{14}$ and $m_{23} = m_{24}$, we have $m_{33} = -m_{34}$ and $m_{43} = -m_{44}$. Thus, $M_r(\widehat{f^*})_{[3,4]} = \pm \left[\begin{smallmatrix} 1 & -1 \\ 1 & -1 \end{smallmatrix}\right]$. In both cases, we have $\operatorname{perm}(M_r(\widehat{f^*})_{[1,2]}) = -2$. Notice that $M_r(\widehat{f^*})_{[1,2]} = \left[\begin{smallmatrix} -1 & 1 \\ -1 & 1 \end{smallmatrix}\right]$. Thus, $\operatorname{perm}(M_r(\widehat{f^*})_{[1,2]}) + \operatorname{perm}(M_r(\widehat{f^*})_{[3,4]}) = -4 \neq 0$. Contradiction.

- There is a row or a column of $M_r(\widehat{f^*})$ such that it has an even number of -1 entries. By transposing $M_r(\widehat{f^*})$, we may assume that it is a row, say Row i. For any other Row j, it differs with Row i in exactly two columns. Thus, Row j also has an even number of -1 entries. If all four rows of $M_r(\widehat{f^*})$ have exactly two -1 entries, then one can check that there are two rows such that one row is a scalar (± 1) multiple of the other, thus not orthogonal; this is a contradiction. Thus, there exists a row in which the number of -1 entries is 0 or 4. By permuting rows, we may assume that it is Row 1. Also, by possibly multiplying $M_r(\widehat{f^*})$ with -1, we may assume that all entries of Row 1 are $+1$. Thus, Row 1 is $(1, 1, 1, 1)$.
  By orthogonality, all other rows have exactly two -1 entries. By permuting columns (which does not change Row 1), we may assume that Row 2 is $(-1, -1, 1, 1)$. Then, consider Row 3. It also has exactly two -1 entries. Moreover, since Row 2 and Row 3 differ in 2 columns, among m_{31} and m_{32}, exactly one is -1. By permuting Column 1 and Column 2 (which does not change Row 1 and Row 2), we may assume that $m_{31} = -1$. Also, among m_{33} and m_{34}, exactly one is -1. By permuting Column 3 and Column 4 (still this will not change Row 1 and Row 2), we may assume that $m_{33} = -1$. Thus, Row 3 is $(-1, 1, -1, 1)$. Finally, consider Row 4. It also has two -1 entries. One can easily check that Row 4 has two possible forms, $(-1, 1, 1, -1)$ or $(1, -1, -1, 1)$. If Row 4 is $(1, -1, -1, 1)$, then,

$$M_r(\widehat{f^*}) = \begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & -1 & 1 & 1 \\ -1 & 1 & -1 & 1 \\ 1 & -1 & -1 & 1 \end{bmatrix}.$$

  Thus, $\operatorname{perm}(M_r(\widehat{f^*})_{[12]}) + \operatorname{perm}(M_r(\widehat{f^*})_{[34]}) = -4 \neq 0$. Contradiction.
  Thus, Row 4 is $(-1, 1, 1, -1)$. Then

$$M_r(\widehat{f^*}) = \begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & -1 & 1 & 1 \\ -1 & 1 & -1 & 1 \\ -1 & 1 & 1 & -1 \end{bmatrix}.$$

  This gives the desired $M_r(\widehat{f_6})$.

Therefore, $\widehat{f_6}$ is realizable from $\widehat{f^*}$.

Since $\widehat{f_6}$ is realized from $\widehat{f^*}$ by flipping (and permuting) variables, i.e., extending some variables of $\widehat{f^*}$ with $=_2$ (using $\neq_2$), we have $\widehat{f_6} \in \{\widehat{f^*}\}^{\widehat{\mathcal{B}}}_{\neq_2}$. Since $\widehat{f^*}$ is realized from $\widehat{f''}$ by extending some variables of $\widehat{f''}$ with signatures in $\widehat{\mathcal{B}}$, we have $\widehat{f^*} \in \{\widehat{f''}\}^{\widehat{\mathcal{B}}}_{\neq_2}$. By Lemma 2.12, we have $\widehat{f_6} \in \{\widehat{f''}\}^{\widehat{\mathcal{B}}}_{\neq_2}$. □

Theorem 6.5. *Suppose that $\widehat{\mathcal{F}}$ contains a 6-ary signature $\widehat{f} \notin \widehat{\mathcal{O}}^{\otimes}$. Then,*

- Holant$(\neq_2 | \widehat{\mathcal{F}})$ *is #P-hard, or*
- *there exists some $\widehat{Q} \in \widehat{\mathbf{O}_2}$ such that* Holant$(\neq_2 | \widehat{f_6}, \widehat{Q\mathcal{F}}) \leqslant_T$ Holant$(\neq_2 | \widehat{\mathcal{F}})$.

Proof. By Lemmas 6.1, 6.2 and 6.3, Holant$(\neq_2 | \widehat{\mathcal{F}})$ is #P-hard, or Holant$(\neq_2 | =_2, \widehat{f''}, \widehat{Q\mathcal{F}}) \leqslant_T$ Holant$(\neq_2 | \widehat{\mathcal{F}})$ for some $\widehat{Q}$ where $Q \in \mathbf{O}_2$, and some irreducible 6-ary signature $\widehat{f''}$ where $\mathscr{S}(\widehat{f''}) = \mathscr{E}_6$ and there exists $\lambda = 1$ or i such that for all $\alpha \in \mathscr{S}(\widehat{f''})$, $\widehat{f''}(\alpha) = \pm\lambda$. Remember that $\widehat{Q}\widehat{\mathcal{F}} = \widehat{Q\mathcal{F}}$ where $Q = Z\widehat{Q}Z^{-1} \in \mathbf{O}_2$. Clearly, $Q\mathcal{F}$ is a set of real-valued signatures of even arity. Since $\mathcal{F}$ does not satisfy condition (T), by Lemma 2.23, $Q\mathcal{F}$ also does not satisfy it. Then, by Lemma 6.4, Holant$(\neq_2 | =_2, \widehat{f''}, \widehat{Q\mathcal{F}})$ is #P-hard, or Holant$(\neq_2 | \widehat{f_6}, \widehat{Q\mathcal{F}}) \leqslant_T$ Holant$(\neq_2 | =_2, \widehat{f''}, \widehat{Q\mathcal{F}})$. Thus, Holant$(\neq_2 | \widehat{\mathcal{F}})$ is #P-hard, or Holant$(\neq_2 | \widehat{f_6}, \widehat{Q\mathcal{F}}) \leqslant_T$ Holant$(\neq_2 | \widehat{\mathcal{F}})$. □

Remark: Theorem 6.5 can be more succinctly stated as simply that a reduction

$$\text{Holant}(\neq_2 | \widehat{f_6}, \widehat{Q\mathcal{F}}) \leqslant_T \text{Holant}(\neq_2 | \widehat{\mathcal{F}})$$

exists, because when Holant$(\neq_2 | \widehat{\mathcal{F}})$ is #P-hard, the reduction exists trivially. However in keeping with the cadence of the other lemmas and theorems in this subsection, we list them as two cases.

Now, we want to show that Holant$(\neq_2 | \widehat{f_6}, \widehat{Q\mathcal{F}})$ is #P-hard for all $\widehat{Q} \in \widehat{\mathbf{O}_2}$ and all $\widehat{\mathcal{F}}$ where $\mathcal{F} = Z\widehat{\mathcal{F}}$ is a real-valued signature set that does not satisfy condition (T). If so, then we are done. Recall that for all $\widehat{Q} \in \widehat{\mathbf{O}_2}$, $\widehat{Q}\widehat{\mathcal{F}} = \widehat{Q\mathcal{F}}$ for some $Q \in \mathbf{O}_2$. Moreover, for all $Q \in \mathbf{O}_2$, and all real-valued $\mathcal{F}$ that does not satisfy condition (T), $Q\mathcal{F}$ is also a real-valued signature set that does not satisfy condition (T). Thus, it suffices for us to show that Holant$(\neq_2 | \widehat{f_6}, \widehat{\mathcal{F}})$ is #P-hard for all real-valued $\mathcal{F}$ that does not satisfy condition (T).

6.2 #P-hardness conditions and two properties of $\widehat{f_6}$

In this subsection, we give three conditions (Lemmas 6.6, 6.8 and 6.9) which can quite straightforwardly lead to the #P-hardness of Holant$(\neq_2 | \widehat{f_6}, \widehat{\mathcal{F}})$. We will extract two properties from $\widehat{f_6}$, the non-$\widehat{\mathcal{B}}$ hardness (Definition 6.7) and the realizability of $\widehat{\mathcal{B}}$ (Lemma 6.11). Later, we will prove the #P-hardness of Holant$(\neq_2 | \widehat{f_6}, \widehat{\mathcal{F}})$ based on these two properties.

Lemma 6.6. Holant$(\neq_2 | \widehat{f_6}, \widehat{\mathcal{F}})$ *is #P-hard if $\widehat{\mathcal{F}}$ contains a nonzero binary signature $\widehat{b} \notin \widehat{\mathcal{B}}^{\otimes}$.*

Proof. If $\widehat{b} \notin \widehat{\mathcal{O}}^{\otimes}$, then by Lemma 5.1, we are done. Otherwise, $\widehat{b} \in \widehat{\mathcal{O}}^{\otimes}$. Thus, $\widehat{b} = (a, 0, 0, \bar{a})$ or $\widehat{b} = (0, a, \bar{a}, 0)$. Since $\widehat{b} \not\equiv 0$, $a \neq 0$. We normalize the norm of a to 1. Since $\widehat{b} \notin \widehat{\mathcal{B}}^{\otimes}$, $a \neq \pm 1$ or $\pm\mathrm{i}$. We first consider the case that $\widehat{b}(y_1, y_2) = (0, a, \bar{a}, 0)$. Connecting variables x_1 and x_2 of $\widehat{f_6}$ with variables y_2 and y_1 of $\widehat{b}$ using $\neq_2$, we get a 4-ary signature $\widehat{g}$. We list the truth table of $\widehat{g}$ indexed by the assignments of variables (x_3, x_4, x_5, x_6) from 0000 to 1111.

$$\widehat{g} = (0, a + \bar{a}, -a + \bar{a}, 0, a - \bar{a}, 0, 0, -a - \bar{a}, -a - \bar{a}, 0, 0, -a + \bar{a}, 0, a - \bar{a}, a + \bar{a}, 0).$$

Since a has norm 1, and $a \neq \pm 1$ or $\pm \mathrm{i}$, $|a \pm \bar{a}| \neq 0$. Thus, $|\mathscr{S}(\widehat{g})| = 8$. Clearly, every 4-ary signature that is in $\widehat{\mathcal{O}}^{\otimes}$ has support of size 0 or 4. Thus, $\widehat{g} \notin \widehat{\mathcal{O}}^{\otimes}$. By Lemma 5.2, Holant$(\neq_2 | \widehat{f}_6, \widehat{\mathcal{F}})$ is #P-hard. We prove the case $\widehat{b}(y_1, y_2) = (a, 0, 0, \bar{a})$ similarly. By connecting variables x_1 and x_2 of $\widehat{f}_6$ with variables y_1 and y_2 of $\widehat{b}$ using $\neq_2$, we also get a 4-ary signature that is not in $\widehat{\mathcal{O}}^{\otimes}$. The lemma is proved. □

Definition 6.7. *We say a signature set $\widehat{\mathcal{F}}$ is non-$\widehat{\mathcal{B}}$ hard, if for any nonzero binary signature $\widehat{b} \notin \widehat{\mathcal{B}}^{\otimes}$, the problem* Holant$(\neq_2 | \widehat{b}, \widehat{\mathcal{F}})$ *is #P-hard. Correspondingly, we say that a signature set $\mathcal{F}$ is non-$\mathcal{B}$ hard, if for any nonzero binary signature $b \notin \mathcal{B}^{\otimes}$, the problem* Holant$(b, \mathcal{F})$ *is #P-hard.*

Clearly, Lemma 6.6 says that $\{\widehat{f}_6\} \cup \widehat{\mathcal{F}}$ is non-$\widehat{\mathcal{B}}$ hard for any $\widehat{\mathcal{F}}$ (where $\mathcal{F} = Z\widehat{\mathcal{F}}$ is a real-valued signature set that does not satisfy condition (T)). Before we give the other two #P-hardness conditions, we first explain why we introduce the notion of non-$\widehat{\mathcal{B}}$ hardness. We will extract two properties from $\widehat{f}_6$ to prove the #P-hardness of Holant$(\neq_2 | \widehat{f}_6, \widehat{\mathcal{F}})$. These are the non-$\widehat{\mathcal{B}}$ hardness and the realizability of $\widehat{\mathcal{B}}$. From Lemma 6.11[1] we get the redutcion Holant$(\neq_2 | \widehat{f}_6, \widehat{\mathcal{B}} \cup \widehat{\mathcal{F}}) \leqslant$ Holant$(\neq_2 | \widehat{f}_6, \widehat{\mathcal{F}})$. We will show that for any non-$\widehat{\mathcal{B}}$ hard set $\widehat{\mathcal{F}}$ where $\mathcal{F}$ does not satisfy condition (T), Holant$(\neq_2 | \widehat{\mathcal{B}} \cup \widehat{\mathcal{F}})$ is #P-hard (Theorem 7.19). This directly implies that Holant$(\neq_2 | \widehat{f}_6, \widehat{\mathcal{F}})$ is #P-hard when $\mathcal{F}$ does not satisfy condition (T). This slightly more general Theorem 7.19 will also be used when dealing with signatures of arity 8. Now, let us continue to give two more #P-hardness conditions without assuming the availability of $\mathcal{B}$ (Lemma 6.8 and 6.9).

Lemma 6.8. *Suppose that $\widehat{\mathcal{F}}$ is non-$\widehat{\mathcal{B}}$ hard. Then* Holant$(\neq_2 | \widehat{\mathcal{F}})$ *is #P-hard if $\widehat{\mathcal{F}}$ contains a nonzero 4-ary signature $\widehat{f} \notin \widehat{\mathcal{B}}^{\otimes}$.*

Proof. If $\widehat{f} \notin \widehat{\mathcal{O}}^{\otimes}$, then by Lemma 5.2, we are done. Otherwise, $\widehat{f} = \widehat{b}_1 \otimes \widehat{b}_2$, where the binary signatures $\widehat{b}_1, \widehat{b}_2 \in \widehat{\mathcal{O}}^{\otimes}$. Since $\widehat{f} \notin \widehat{\mathcal{B}}^{\otimes}$, $\widehat{b}_1$ and $\widehat{b}_2$ are not both in $\widehat{\mathcal{B}}^{\otimes}$. Then, we can realize a binary signature that is not in $\widehat{\mathcal{B}}^{\otimes}$ by factorization. Since $\widehat{\mathcal{F}}$ is non-$\widehat{\mathcal{B}}$ hard, we are done. □

Let $H = \frac{1}{\sqrt{2}} \left[\begin{smallmatrix} 1 & 1 \\ -1 & 1 \end{smallmatrix}\right]$. Then $\widehat{H} = Z^{-1}HZ = \left[\begin{smallmatrix} \frac{(1+\mathrm{i})}{\sqrt{2}} & 0 \\ 0 & \frac{(1-\mathrm{i})}{\sqrt{2}} \end{smallmatrix}\right] = \left[\begin{smallmatrix} e^{\mathrm{i}\pi/4} & 0 \\ 0 & e^{-\mathrm{i}\pi/4} \end{smallmatrix}\right]$. Let $\widehat{f_6^H} = \widehat{H}\widehat{f}_6$. Let $\widehat{\mathcal{F}_6} = \{\widehat{f}_6\}_{\neq_2}^{\widehat{B}}$ be the set of signature realizable by extending variables of $\widehat{f}_6$ with binary signatures in $\widehat{\mathcal{B}}$ using $\neq_2$, and $\widehat{\mathcal{F}_6^H} = \{\widehat{f_6^H}\}_{\neq_2}^{\widehat{B}}$ be the set of signature realizable by extending variables of $\widehat{f_6^H}$ with binary signatures in $\widehat{\mathcal{B}}$ using $\neq_2$. One can check that $\widehat{\mathcal{F}_6^H} = \widehat{H}\widehat{\mathcal{F}_6} \neq \widehat{\mathcal{F}_6}$.

Lemma 6.9. *Suppose that $\widehat{\mathcal{F}}$ is non-$\widehat{\mathcal{B}}$ hard. Then,* Holant$(\neq_2 | \widehat{\mathcal{F}})$ *is #P-hard if $\widehat{\mathcal{F}}$ contains a nonzero 6-ary signature $\widehat{f} \notin \widehat{\mathcal{B}}^{\otimes} \cup \widehat{\mathcal{F}_6} \cup \widehat{\mathcal{F}_6^H}$.*

Proof. If $\widehat{f}$ is reducible, since $\widehat{f} \notin \widehat{\mathcal{B}}^{\otimes}$, then by factorization, we can realize a nonzero signature of odd arity or a nonzero signature of arity 2 or 4 that is not in $\widehat{\mathcal{B}}^{\otimes}$. If we have a nonzero signature of odd arity, then we are done by Theorem 2.25. If we have a nonzero signature of 2, then we are done because $\widehat{\mathcal{F}}$ is non-$\widehat{\mathcal{B}}$ hard. If we have a nonzero signature of 4, then we are done by Lemma 6.8. Now we assume that $\widehat{f}$ is irreducible. In particular, being irreducible, $\widehat{f} \notin \widehat{\mathcal{O}}^{\otimes}$. For a contradiction, suppose that Holant$(\neq_2 | \widehat{\mathcal{F}})$ is not #P-hard. Then, by Theorem 6.5, $\widehat{f}_6$ is realizable from $\widehat{f}$. Remember that we realize $\widehat{f}_6$ from $\widehat{f}$ by realizing $\widehat{f}'$, $\widehat{f''}$ and $\widehat{f^*}$ (Lemmas 6.1, 6.3 and 6.4). We will trace back this process and show that they are all in $\widehat{\mathcal{F}_6} \cup \widehat{\mathcal{F}_6^H}$, which contradicts with the condition that $\widehat{f} \notin \widehat{\mathcal{F}_6} \cup \widehat{\mathcal{F}_6^H}$.

[1] This lemma and the following Theorem 7.19 are stated and proved in the setting of Holant$(\mathcal{F})$.

1. First, by Lemma 6.4, $\widehat{f_6} \in \{\widehat{f''}\}^{\widehat{\mathcal{B}}}_{\neq_2}$. Then, by Lemma 2.12, $\widehat{f''} \in \{\widehat{f_6}\}^{\widehat{\mathcal{B}}}_{\neq_2} = \widehat{\mathcal{F}_6}$.
2. Then, by Lemma 6.3, $\widehat{f''} = \widehat{Q}\widehat{f'}$ for some $\widehat{Q} = \begin{bmatrix} e^{-\mathrm{i}\delta} & 0 \\ 0 & e^{\mathrm{i}\delta} \end{bmatrix} \in \widehat{\mathbf{O}_2}$ where $0 \leqslant \delta < \pi/2$, and the binary signature $\widehat{b} = (e^{\mathrm{i}2\delta}, 0, 0, e^{-\mathrm{i}2\delta})$ is realizable from $\widehat{f'}$ where $\widehat{f'}$ is realizable from $\widehat{f}$. Thus, $\widehat{b}$ is realizable from $\widehat{\mathcal{F}}$. If $e^{\mathrm{i}2\delta} \neq \pm 1$ or $\pm\mathrm{i}$, then $\widehat{b} \notin \widehat{\mathcal{B}}^{\otimes}$. Since $\widehat{\mathcal{F}}$ is non-$\widehat{\mathcal{B}}$ hard, we get #P-harness. Contradiction. Otherwise, since $0 \leqslant \delta < \pi/2$, $e^{\mathrm{i}2\delta} = 1$ or i and then, $\delta = 0$ or $\pi/4$. If $\delta = 0$, then $e^{\mathrm{i}\delta} = e^{-\mathrm{i}\delta} = 1$ and $\widehat{f''} = \widehat{Q}\widehat{f'} = \widehat{f'}$. Thus, $\widehat{f'} \in \widehat{\mathcal{F}_6}$. If $\delta = \pi/4$, then $\widehat{f'} = \widehat{Q}^{-1}\widehat{f''}$ where $\widehat{Q}^{-1} = \begin{bmatrix} e^{\mathrm{i}\pi/4} & 0 \\ 0 & e^{-\mathrm{i}\pi/4} \end{bmatrix} = \widehat{H}$. Since $\widehat{f''} \in \widehat{\mathcal{F}_6}$, $\widehat{f'} = \widehat{H}\widehat{f''} \in \widehat{H}\widehat{\mathcal{F}_6} = \widehat{\mathcal{F}_6^H}$.
3. Finally, by Lemma 6.1, $\widehat{f'}$ is realized by extending variables of $\widehat{f}$ with nonzero binary signatures realized from $\widehat{\partial}_{12}\widehat{f}$. If we can realize a nonzero binary signature that is not in $\widehat{\mathcal{B}}^{\otimes 1}$ from $\widehat{\partial}_{12}\widehat{f}$ by factorization, then since $\widehat{\mathcal{F}}$ is non-$\widehat{\mathcal{B}}$ hard, we get #P-hardness. Contradiction. Thus, we may assume that all nonzero binary signatures realizable from $\widehat{\partial}_{12}\widehat{f}$ are in $\widehat{\mathcal{B}}^{\otimes 1}$. Then, $\widehat{f'}$ is realized by extending variables of $\widehat{f}$ with nonzero binary signatures in $\widehat{\mathcal{B}}^{\otimes 1}$. Thus, $\widehat{f'} \in \{\widehat{f}\}^{\widehat{\mathcal{B}}}_{\neq_2}$. By Lemma 2.12, $\widehat{f} \in \{\widehat{f'}\}^{\widehat{\mathcal{B}}}_{\neq_2}$. Since $\widehat{f'} \in \widehat{\mathcal{F}_6}$ or $\widehat{\mathcal{F}_6^H}$, $\widehat{f} \in \widehat{\mathcal{F}_6}$ or $\widehat{\mathcal{F}_6^H}$. Contradiction.

Thus, Holant$(\neq_2 | \widehat{\mathcal{F}})$ is #P-hard if $\widehat{\mathcal{F}}$ contains a nonzero 6-ary signature $\widehat{f} \notin \widehat{\mathcal{B}}^{\otimes} \cup \widehat{\mathcal{F}_6} \cup \widehat{\mathcal{F}_6^H}$. □

We go back to real-valued Holant problems under the Z-transformation. Consider the problem Holant$(f_6, \mathcal{F})$ where

$$f_6 = Z\widehat{f_6} = \chi_S \cdot (-1)^{x_1+x_2+x_3+x_1x_2+x_2x_3+x_1x_3+x_1x_4+x_2x_5+x_3x_6}$$

and $S = \mathscr{S}(f_6) = \mathscr{E}_6$. The signature f_6 has a quite similar matrix form to $\widehat{f_6}$.

$$M_{123,456}(f_6) = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & -1 & 0 & -1 & 0 & 0 & 1 \\ 0 & -1 & 1 & 0 & -1 & 0 & 0 & 1 \\ -1 & 0 & 0 & -1 & 0 & 1 & 1 & 0 \\ 0 & -1 & -1 & 0 & 1 & 0 & 0 & 1 \\ -1 & 0 & 0 & 1 & 0 & -1 & 1 & 0 \\ -1 & 0 & 0 & 1 & 0 & 1 & -1 & 0 \\ 0 & -1 & -1 & 0 & -1 & 0 & 0 & -1 \end{bmatrix}.$$

Since $\widehat{f_6^H} = \widehat{H}\widehat{f_6} = \widehat{Hf_6}$, $f_6^H = Z\widehat{f_6^H} = Hf_6$. Also, since $\widehat{\mathcal{F}_6} = \{\widehat{f_6}\}^{\widehat{\mathcal{B}}}_{\neq_2}$, $\mathcal{F}_6 = Z\widehat{\mathcal{F}_6} = \{f_6\}^{\mathcal{B}}_{=_2}$ is the set of signatures realizable by extending variables of f_6 with binary signatures in $\mathcal{B}$ using $=_2$. Similarly, since $\widehat{\mathcal{F}_6^H} = \{\widehat{f_6^H}\}^{\widehat{\mathcal{B}}}_{\neq_2}$, $\mathcal{F}_6^H = Z\widehat{\mathcal{F}_6^H} = \{f_6\}^{\mathcal{B}}_{=_2}$ is the set of signatures realizable by extending variables of f_6^H with binary signatures in $\mathcal{B}$ using $=_2$. Notice that $f_6 \in \mathscr{A}$ and $\mathcal{B} \subseteq \mathscr{A}$. Thus, $\mathcal{F}_6 \subseteq \mathscr{A}$. Also, the binary signature $(1, 1, -1, 1)$ with a signature matrix H is in $\mathscr{A}$. Thus, $f_6^H \in \mathscr{A}$ and then $\mathcal{F}_6^H \subseteq \mathscr{A}$. Also, $\mathscr{S}(f_6) = \mathscr{E}_6$ and one can check that $\mathscr{S}(f_6^H) = \mathscr{O}_6$. Thus, for every $f \in \mathcal{F}_6 \cup \mathcal{F}_6^H$, $\mathscr{S}(f) = \mathscr{E}_6$ or $\mathscr{O}_6$. Since f_6 and f_6^H satisfy 2ND-ORTH, one can easily check that every $f \in \mathcal{F}_6 \cup \mathcal{F}_6^H$ satisfies 2ND-ORTH.

We want to show that Holant$(f_6, \mathcal{F}) \equiv_T$ Holant$(\neq_2 | \widehat{f_6}, \widehat{\mathcal{F}})$ is #P-hard for all real-valued $\mathcal{F}$ that does not satisfy condition (T). By Lemma 6.6, $\{f_6\} \cup \mathcal{F}$ is non-$\mathcal{B}$ hard. We restate Lemmas 6.8 and 6.9 in the setting of Holant$(\mathcal{F})$ for non-$\mathcal{B}$ hard $\mathcal{F}$.

Corollary 6.10. *Suppose that $\mathcal{F}$ is non-$\mathcal{B}$ hard. Then,* Holant$(\mathcal{F})$ *is #P-hard if $\mathcal{F}$ contains a nonzero signature f of arity at most* 6 *where $f \notin \mathcal{B}^{\otimes} \cup \mathcal{F}_6 \cup \mathcal{F}_6^H$.*

Remark: Notice that $\mathcal{B}^{\otimes} \cup \mathcal{F}_6 \cup \mathcal{F}_6^H \subseteq \mathscr{A}$. Thus, for any non-$\mathcal{B}$ hard set $\mathcal{F}$, Holant($\mathcal{F}$) is #P-hard if $\mathcal{F}$ contains a nonzero signature f of arity at most 6 where $f \notin \mathscr{A}$.

Now, we show that all four binary signatures in $\mathcal{B}$ are realizable from f_6.

Lemma 6.11. Holant$(\mathcal{B}, f_6, \mathcal{F}) \leqslant$ Holant$(f_6, \mathcal{F})$.

Proof. Consider $\partial_{12} f_6$. Notice that

$$\begin{bmatrix} \mathbf{f_6}_{12}^{00} \\ \mathbf{f_6}_{12}^{11} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & -1 & 0 & -1 & 0 & 0 & 1 \\ -1 & 0 & 0 & 1 & 0 & 1 & -1 & 0 & 0 & -1 & -1 & 0 & -1 & 0 & 0 & -1 \end{bmatrix}.$$

Thus, $\partial_{12} f_6(x_3, x_4, x_5, x_6) = f_6{}_{12}^{00} + f_6{}_{12}^{11}$ has the truth table $(0,0,0,1,0,1,0,0,0,0,-1,0,-1,0,0,0)$. In other words, $\partial_{12} f_6(0011) = 1$, $\partial_{12} f_6(0101) = 1$, $\partial_{12} f_6(1010) = -1$, $\partial_{12} f_6(1100) = -1$, and $\partial_{12} f_6 = 0$ elsewhere. Then,

$$\mathscr{S}(\partial_{12} f_6) = \{(x_3, x_4, x_5, x_6) \in \mathbb{Z}_2^4 \mid x_3 \neq x_6 \wedge x_4 \neq x_5\},$$

and

$$\partial_{12} f_6(x_3, x_4, x_5, x_6) = (\neq_2^-)(x_3, x_6) \otimes (\neq_2)(x_4, x_5).$$

Thus, by factorization we can realize $\neq_2^-$ and $\neq_2$. Then connecting a variable of $\neq_2^-$ with a variable of $\neq_2$ (using $=_2$), we will get $=_2^-$. Thus, $\mathcal{B}$ is realizable from f_6. □

We define the problem Holant$^b(\mathcal{F})$ to be Holant$(\mathcal{B} \cup \mathcal{F})$. For all $\{i, j\}$ and every $b \in \mathcal{B}$, consider signatures $\partial_{ij}^b f_6$ (i.e., $\partial_{ij}^+ f_6$, $\partial_{ij}^{\widehat{+}} f_6$, $\partial_{ij}^- f_6$ and $\partial_{ij}^{\widehat{-}} f_6$) realized by merging variables x_i and x_j of f_6 using the binary signature b. If there were one that is not in $\mathcal{B}^{\otimes 2}$, then by Corollary 6.10, we would be done. However, it is observed in [15] that f_6 satisfies the following Bell property.

Definition 6.12 (Bell property)**.** *An irreducible signature f satisfies the Bell property if for all pairs of indices $\{i, j\}$ and every $b \in \mathcal{B}$, $\partial_{ij}^b f \in \mathcal{B}^{\otimes}$.*

It can be directly checked that

Lemma 6.13. *Every signature in $\mathcal{F}_6 \cup \mathcal{F}_6^H$ satisfies the Bell property.*

Now consider all possible gadget constructions. If we could realize a signature of arity at most 6 that is not in $\mathcal{B}^{\otimes} \cup \mathcal{F}_6 \cup \mathcal{F}_6^H$ from $\mathcal{B}$ and f_6 by any possible gadget, then by Corollary 6.10 there would be a somewhat more straightforward proof to our dichotomy theorem for the case of arity 6. However, after many failed attempts, we believe there is a more intrinsic reason why this approach cannot work. The following conjecture formulates this difficulty. This truly makes f_6 the *Lord of Intransigence* at arity 6.

Conjecture 6.14. *All nonzero signatures of arity at most 6 realizable from $\mathcal{B} \cup \{f_6\}$ are in $\mathcal{B}^{\otimes} \cup \mathcal{F}_6$. Also, all signatures of arity at most 6 realizable from $\mathcal{B} \cup \{f_6^H\}$ are in $\mathcal{B}^{\otimes} \cup \mathcal{F}_6^H$.*

So to prove the #P-hardness of Holant$^b(f_6, \mathcal{F})$, we have to make additional use of $\mathcal{F}$. In particular, we need to use a non-affine signature in $\mathcal{F}$.

7 The #P-hardness of $\text{Holant}^b(\mathcal{F})$

In this section, we prove that for all real-valued non-$\mathcal{B}$ hard set $\mathcal{F}$ that does not satisfy condition (T), $\text{Holant}^b(\mathcal{F})$ is #P-hard (Theorem 7.19). For any real-valued set $\mathcal{F}$ that does not satisfy condition (T), the set $\{f_6\} \cup \mathcal{F}$ is non-$\mathcal{B}$ hard, and since $\mathcal{B}$ is realizable from f_6, $\text{Holant}(f_6, \mathcal{F})$ is #P-hard by Theorem 7.19. Combining with Theorem 6.5, we show that $\text{Holant}(\neq_6 | \widehat{\mathcal{F}})$ is #P-hard if $\widehat{\mathcal{F}}$ contains a 6-ary signature that is not in $\widehat{\mathcal{O}}^{\otimes}$ (Lemma 7.21).

Since $\mathcal{F}$ does not satisfy condition (T), $\mathcal{F} \not\subseteq \mathscr{A}$. Thus, it contains a signature f of arity $2n$ that is not in $\mathscr{A}$. In the following, we will prove the #P-hardness of $\text{Holant}^b(\mathcal{F})$ where $\mathcal{F}$ is non-$\mathcal{B}$ hard by induction on $2n \geqslant 2$. For the base cases $2n \leqslant 6$, by Corollary 6.10 and the Remark after that, $\text{Holant}^b(\mathcal{F})$ is #P-hard. Then, starting with a signature of arity $2n \geqslant 8$ that is not in $\mathscr{A}$, we want to realize a signature of lower arity $2k \leqslant 2n-2$ that is also not in $\mathscr{A}$, or else we get #P-hardness directly. If we can reduce the arity down to at most 6, then we are done.

Let $f \notin \mathscr{A}$ be a nonzero signature of arity $2n \geqslant 8$. We first show that if f does not have parity, then we get #P-hardness (Lemma 7.1). Then, suppose that f has parity. If f is reducible, since f has even arity (as we assumed so starting from Section 3), it is a tensor product of two signatures of odd arity, or a tensor product of two signatures of even arity which are not both in $\mathscr{A}$ since $f \notin \mathscr{A}$. Thus, by factorization, we can realize a nonzero signature of odd arity and we get #P-hardness by Theorem 2.25, or we can realize a signature of lower even arity that is not in $\mathscr{A}$. Thus, we may assume that f is irreducible. Then by Lemma 4.4 and the Remark after Definition 4.1 we may assume f satisfies 2ND-ORTH.

Consider signatures $\partial^b_{ij} f$ (i.e., $\partial^+_{ij} f$, $\partial^-_{ij} f$, $\partial^{\widehat{+}}_{ij} f$ and $\partial^{\widehat{-}}_{ij} f$) realized by merging variables x_i and x_j of f using $b \in \mathcal{B}$ for all pairs of indices $\{i, j\}$ and every $b \in \mathcal{B}$. If there is one signature that is not in $\mathscr{A}$, then we have realized a signature of arity $2n-2$ that is not in $\mathscr{A}$. Otherwise, $\partial^b_{ij} f \in \mathscr{A}$ for all $\{i, j\}$ and every $b \in \mathcal{B}$. We denote this property by $f \in \int_{\mathcal{B}} \mathscr{A}$. Now, assuming that f has parity, f satisfies 2ND-ORTH and $f \in \int_{\mathcal{B}} \mathscr{A}$, we would like to reach a contradiction by showing that this would force f itself to belong to $\mathscr{A}$. However, quite amazingly, there do exist non-affine signatures that satisfy these stringent conditions. We will show how they are discovered and handled (Lemmas 7.9, 7.16 and 7.18).

In this section, all signatures are real-valued. When we say an entry of a signature has norm a, we mean it takes value $\pm a$. Since $\mathcal{B}$ is available in $\text{Holant}^b(\mathcal{F})$, if a signature f is realizable in $\text{Holant}^b(\mathcal{F})$, then we can realize all signatures in $\{f\}^{\mathcal{B}}_{=_2}$ that are realizable by extending f with $\mathcal{B}^{\otimes 1}$ (using $=_2$). If we extend the variable x_i of f with $\neq_2$, then we will get a signature g where $g^0_i = f^1_i$ and $g^1_i = f^0_i$. This is a flipping operation on the variable x_i. If we extend the variable x_i of f with $=^-_2$, then we will get a signature g where $g^0_i = f^0_i$ and $g^1_i = -f^1_i$. We call this a negating operation on the variable x_i. In the following, once f is realizable in $\text{Holant}^b(\mathcal{F})$, we may modify it by flipping or negating. This will not change the complexity of the problem.

7.1 Parity condition

We first show that if $\mathcal{F}$ contains a signature that does not have parity, then we can get #P-hardness.

Lemma 7.1. *Suppose that $\mathcal{F}$ is non-$\mathcal{B}$ hard and $\mathcal{F}$ contains a signature f of arity $2n$. If f does not have parity, then $\text{Holant}^b(\mathcal{F})$ is #P-hard.*

Proof. We prove this lemma by induction on $2n$. We first consider the base case that $2n = 2$. Since f has no parity, $f \notin \mathcal{B}$. Since $\mathcal{F}$ is non-$\mathcal{B}$ hard, $\text{Holant}^b(\mathcal{F})$ is #P-hard.

Now, suppose that $\text{Holant}^b(\mathcal{F})$ is #P-hard when $2n = 2k \geqslant 2$. Consider the case that $2n = 2k + 2 \geqslant 4$. We will show that we can realize a signature g of arity $2k$ with no parity from f, i.e., $\text{Holant}^b(g, \mathcal{F}) \leqslant_T \text{Holant}^b(\mathcal{F})$. Then by the induction hypothesis, we have $\text{Holant}^b(\mathcal{F})$ is #P-hard when $2n = 2k + 2$.

Since f has no parity, $f \not\equiv 0$. It has at least a nonzero entry. By flipping variables of f, we may assume that $f(\vec{0}^{2n}) = x \neq 0$. We denote $\vec{0}^{2n}$ by $\alpha = 000\delta$ where $\delta = \vec{0}^{2n-3}$. Since f has no parity and $f(\vec{0}^{2n}) \neq 0$, there exists an input α' with $\text{wt}(\alpha') \equiv 1 \pmod 2$ such that $f(\alpha') = x' \neq 0$. Since $2n \geqslant 4$, we can find three bits of α' such that on these three bits, the values of α' are the same. By renaming variables of f which gives a permutation of α', without loss of generality, we may assume that these are the first three bits, i.e, $\alpha'_1 = \alpha'_2 = \alpha'_3$.

We first consider the case that $\alpha'_1\alpha'_2\alpha'_3 = 000$. Then, $\alpha' = 000\delta'$ for some $\delta' \in \mathbb{Z}_2^{2n-3}$ where $\text{wt}(\delta') = \text{wt}(000\delta') = \text{wt}(\alpha') \equiv 1 \pmod 2$. We consider the following six entries of f.

$$x = f(000\delta), x' = f(000\delta'), y = f(011\delta), y' = f(011\delta'), z = f(101\delta), z' = f(101\delta').$$

Consider signatures $\partial^+_{23}f$ and $\partial^-_{23}f$ realized by connecting variables x_2 and x_3 of f using $=^+_2$ and $=^-_2$ respectively. Clearly, $\partial^+_{23}f$ and $\partial^-_{23}f$ have arity $2n-2$. If one of them has no parity, then we are done. Thus, we may assume that $\partial^+_{23}f$ and $\partial^-_{23}f$ both have parity. Note that $x + y$ and $x' + y'$ are entries of the signature $\partial^+_{23}f$ on inputs 0δ and $0\delta'$ respectively. Clearly, $\text{wt}(0\delta) = 0$ and $\text{wt}(0\delta') \equiv 1 \pmod 2$. Since $\partial^+_{23}f$ has parity, at least one of $x + y$ and $x' + y'$ is zero. Thus, we have $(x + y)(x' + y') = 0$. Also, note that $x - y$ and $x' - y'$ are entries of the signature $\partial^-_{23}f$ on inputs 0δ and $0\delta'$ respectively. Then, since $\partial^-_{23}f$ has parity, similarly we have $(x - y)(x' - y') = 0$. Thus,

$$(x + y)(x' + y') + (x - y)(x' - y') = 2(xx' + yy') = 0. \tag{7.1}$$

Consider signatures $\partial^+_{13}f$ and $\partial^-_{13}f$ realized by connecting variables x_1 and x_3 of f using $=_2$ and $=^-_2$ respectively. Again if one of them has no parity, then we are done. Suppose that $\partial^+_{13}f$ and $\partial^-_{13}f$ both have parity. Then, $(x + z)(x' + z') = 0$ since $x + z$ and $x' + z'$ are entries of $\partial^+_{13}f$ on inputs 0δ and $0\delta'$ respectively. Similarly, $(x - z)(x' - z') = 0$. Thus,

$$(x + z)(x' + z') + (x - z)(x' - z') = 2(xx' + zz') = 0. \tag{7.2}$$

Consider signatures $\partial^{\widehat{+}}_{12}f$ and $\partial^{\widehat{-}}_{12}f$ realized by connecting variables x_1 and x_2 of f using $\neq_2$ and $\neq^-_2$ respectively. Again if one of them has no parity, then we are done. Suppose that $\partial^{\widehat{+}}_{12}f$ and $\partial^{\widehat{-}}_{12}f$ both have parity. Then, $(y + z)(y' + z') = 0$ since $y + z$ and $y' + z'$ are entries of $\partial^{\widehat{+}}_{12}f$ on inputs 1δ and $1\delta'$ respectively, and $\text{wt}(1\delta) = 1$ and $\text{wt}(1\delta') \equiv 0 (\text{mod} 2)$. Similarly, $(y - z)(y' - z') = 0$. Thus,

$$(y + z)(y' + z') + (y - z)(y' - z') = 2(yy' + zz') = 0. \tag{7.3}$$

Then, consider (7.1) + (7.2) − (7.3). We have $xx' = 0$. However, since $x = f(\vec{0}^{2n}) \neq 0$ and $x' = f(\alpha') \neq 0$, $xx' \neq 0$. Contradiction.

For the case that $\alpha'_1\alpha'_2\alpha'_3 = 111$, we have $\alpha' = 111\delta'$ for some $\delta' \in \mathbb{Z}_2^{2n-3}$ where $\text{wt}(\delta') = \text{wt}(111\delta') - 3 = \text{wt}(\alpha') - 3 \equiv 0 \pmod 2$. We consider the following six entries of f.

$$x = f(000\delta), x' = f(111\delta'), y = f(011\delta), y' = f(100\delta'), z = f(101\delta), z' = f(010\delta').$$

We still consider signatures $\partial^+_{23}f$, $\partial^-_{23}f$, $\partial^+_{13}f$, $\partial^-_{13}f$, $\partial^{\widehat{+}}_{12}f$ and $\partial^{\widehat{-}}_{12}f$ and suppose that they all have parity. Then, similar to the above proof of the case $\alpha'_1\alpha'_2\alpha'_3 = 000$, we can show that $xx' = 0$. Contradiction.

Thus, among $\partial^+_{23}f$, $\partial^-_{23}f$, $\partial^+_{13}f$, $\partial^-_{13}f$, $\partial^{\widehat{+}}_{12}f$ and $\partial^{\widehat{-}}_{12}f$, at least one does not have parity. Thus, we realized a $2k$-ary signature with no parity. By our induction hypothesis, we are done. □

7.2 Norm condition

Under the assumptions that f has parity, f satisfies 2ND-ORTH and $f \in \int_{\mathcal{B}} \mathscr{A}$, we consider whether all nonzero entries of f have the same norm. In Lemma 7.9, we will show that the answer is yes, but only for signatures of arity $2n \geqslant 10$ (this lemma does not require $\mathcal{F}$ to be non-$\mathcal{B}$ hard). For a signature f of arity $2n = 8$, we show that either all nonzero entries of f have the same norm, or one of the following signatures g_8 or g'_8 is realizable. These two signatures are defined by $g_8 = \chi_S - 4 \cdot f_8$ and $g'_8 = q_8 - 4 \cdot f_8$, where

$$S = \mathscr{S}(q_8) = \mathscr{E}_8, \qquad q_8 = \chi_S (-1)^{\sum_{1 \leqslant i < j \leqslant 8} x_i x_j} \qquad \text{and}$$

$$f_8 = \chi_T \text{ with } T = \mathscr{S}(f_8) = \{(x_1, x_2, \ldots, x_8) \in \mathbb{Z}_2^8 \mid \; x_1 + x_2 + x_3 + x_4 = 0, \; x_5 + x_6 + x_7 + x_8 = 0, \; x_1 + x_2 + x_5 + x_6 = 0, \; x_1 + x_3 + x_5 + x_7 = 0\}. \tag{7.4}$$

It is here the function f_8 makes its first appearance, we dub it the *Queen of the Night*. Clearly, $g_8, g'_8 \notin \mathscr{A}$ since their nonzero entries have two different norms 1 and 3. One can check that g_8 and g'_8 have parity, g_8 and g'_8 satisfy 2ND-ORTH and $g_8, g'_8 \in \int_{\mathcal{B}} \mathscr{A}$. Thus, one cannot get a non-affine signature by connecting two variables of g_8 or g'_8 using signatures in $\mathcal{B}$. However, fortunately by merging two arbitrary variables of g_8 using $=_2$ and two arbitrary variables of g'_8 using $=_2^-$, we can get 6-ary irreducible signatures that do not satisfy 2ND-ORTH. Thus, we get #P-hardness.

The following Lemma 7.4 regarding the independence number of a family of special graphs is at the heart of the discovery of the signature g_8. It should be of independent interest.

Definition 7.2. *Define the graphs G_{2n} and H_{2n} as follows. The vertex set $V(G_{2n})$ is the set $\mathscr{E}_{2n}$ of all even weighted points in $\mathbb{Z}_2^{2n}$. The vertex set $V(H_{2n})$ is the set $\mathscr{O}_{2n}$ of all odd weighted points in $\mathbb{Z}_2^{2n}$. Two points $u, v \in \mathscr{E}_{2n}$ (or $\mathscr{O}_{2n}$) are connected by an edge iff* $\mathrm{wt}(u \oplus v) = 2$.

Let $\alpha(G_{2n})$ be the independence number of G_{2n} i.e, the size of a maximum independent set of G_{2n}, and $\alpha(H_{2n})$ be the independence number of H_{2n}. Let $S \subseteq [2n]$. We define φ_S be a mapping that flips the values on bits in S for all $u \in \mathscr{E}_{2n}$. In other words, suppose that $u' = \varphi_S(u)$. Then, $u'_i = \overline{u_i}$ if $i \in S$ and $u'_i = u_i$ if $i \notin S$ where u'_i and u_i are values of u and u on bit i respectively. For all S, clearly $\mathrm{wt}(u \oplus v) = 2$ iff $\mathrm{wt}(\varphi_S(u) \oplus \varphi_S(v)) = 2$. When $|S|$ is odd, $\varphi_S(\mathscr{E}_{2n}) = \mathscr{O}_{2n}$. One can easily check that φ_S gives an isomorphism between G_{2n} and H_{2n}. When $|S|$ is even, $\varphi_S(\mathscr{E}_{2n}) = \mathscr{E}_{2n}$. Then, φ_S gives an automorphism of G_{2n}. Also, by permuting these $2n$ bits, we can get an automorphism of G_{2n}. In fact, the automorphism group of G_{2n} is generated by these operations.

Lemma 7.3. *Let $2n \geqslant 6$. Every automorphism ψ of G_{2n} is a product $\varphi_S \circ \pi$ for some automorphism π induced by a permutation of $2n$ bits, and an automorphism φ_S given by flipping the values on some bits in a set S of even cardinality.*

Proof. Let ψ be an arbitrary automorphism of G_{2n}. Suppose $\psi(\vec{0}^{2n}) = u$. Let $S \subseteq [2n]$ be the index set where $u_i = 1$. Then $|S| = \mathrm{wt}(u)$ is even, and $\psi' = \varphi_S \circ \psi$ fixes $\vec{0}^{2n}$. Consider $\psi'(v)$ for all $v \in \mathscr{E}_{2n}$ of $\mathrm{wt}(v) = 2$. Since ψ' is an automorphism fixing $\vec{0}^{2n}$, $\psi'(v)$ has weight 2. We denote by e_{ij} the $2n$-bit string with $\mathrm{wt}(e_{ij}) = 2$ having 1's on bits i and j, for $1 \leqslant i < j \leqslant 2n$. Then, $e_{12} = 11\vec{0}^{2n-2}$. By a suitable permutation π of the variables, we have $\pi \circ \psi'(e_{12}) = e_{12}$, while still fixing $\vec{0}^{2n}$. We will show that $\pi \circ \psi' = \pi \circ \varphi_S \circ \psi$ is the identity mapping, i.e., $\pi \circ \varphi_S \circ \psi = 1_{G_{2n}}$. Then, $\psi = \varphi_S^{-1} \circ \pi^{-1}$. We are done.

For simplicity of notations, we reuse ψ to denote $\pi \circ \psi'$ in the following and we show that $\psi = 1_{G_{2n}}$. Consider e_{1i}, for $3 \leqslant i \leqslant 2n$. Note that $\psi(e_{1i})$ is some e_{st} and must have Hamming distance 2 to e_{12}. It is easy to see that the only possibilities are $s \in \{1,2\}$ and $t > 2$, i.e., from e_{12} we flip exactly one bit in $\{1,2\}$ and another bit in $\{3,\ldots,2n\}$. Suppose there are $i \neq i'$ $(i, i' \geqslant 3)$ such that $\psi(e_{1i}) = e_{1t}$ and $\psi(e_{1i'}) = e_{2t'}$. Since $\mathrm{wt}(e_{1i} \oplus e_{1i'}) = 2$, we must have $t = t'$. Since $2n \geqslant 6$, we can pick another $i'' \geqslant 3$ such that $i'' \neq i$ and i'. Then, this leads to a contradiction since $e_{1i''}$ must either be mapped to e_{1t} if $\psi(e_{1i''}) = e_{1t''}$, or be mapped to e_{2t} if $\psi(e_{1i''}) = e_{2t''}$; neither is possible. Thus either $\psi(e_{1i})$ is some e_{1t} for all $3 \leqslant i \geqslant 2n$, or is some e_{2t} for all $3 \leqslant i \geqslant 2n$. By a permutation of $\{1,2\}$ (which maintains the property that ψ fixes $\vec{0}^{2n}$ and e_{12}) we may assume it is the former. Then the mapping $i \mapsto t$ given by $\psi(e_{1i}) = e_{1t}$ for $3 \leqslant i \geqslant 2n$ defines a permutation of the variables for $3 \leqslant i \geqslant 2n$ (which again maintains the property that ψ fixes $\vec{0}^{2n}$ and e_{12}) and, after a permutation of the variables we may now assume that ψ fixes $\vec{0}^{2n}$ and all e_{1i}. For any $1 \leqslant i < j \leqslant 2n$, we have $\mathrm{wt}(\psi(e_{ij})) = 2$ and $\psi(e_{ij})$ has distance 2 from both $\psi(e_{1i}) = e_{1i}$ and $\psi(e_{1j}) = e_{1j}$. Then $\psi(e_{ij})$ must be obtained from e_{1i} by flipping exactly one bit in $\{1,i\}$ and another bit out of $\{1,i\}$. However, it cannot flip bit i which would result in some e_{1t} for some $t > 2$, because ψ already fixed e_{1t}. Thus, it flips bit 1 but not bit i. Similarly in view of e_{1j}, we must flip bit 1 but not bit j. Hence $\psi(e_{ij}) = e_{ij}$, and therefore ψ fixes all v with Hamming weight $\mathrm{wt}(v) \leqslant 2$.

Inductively assume ψ fixes all v of $\mathrm{wt}(v) \leqslant 2k$, for some $k \geqslant 1$. If $k < n$ we prove that ψ also fixes all v of $\mathrm{wt}(v) = 2k+2$. For notational simplicity we may assume $v = \vec{1}^{2k+2}\vec{0}^{2n-2k-2}$. As $2k+2 \geqslant 4$, we can choose $u = \vec{1}^{2k}00\vec{0}^{2n-2k-2}$ and $w = 00\vec{1}^{2k}\vec{0}^{2n-2k-2}$, and the two 00 in u and w among the first $2k+2$ bits are in disjoint bit positions. Clearly $\mathrm{wt}(\psi(v)) \geqslant 2k+2$ since all strings of wt $\leqslant 2k$ are fixed. Also since $\psi(v)$ has Hamming distance 2 from $\psi(u) = u$ and $\psi(w) = w$, it has weight exactly $2k+2$, and is obtained from u by flipping two bits from 00 to 11 in positions $> 2k$, as well as obtained from w by flipping two bits from 00 to 11 in positions in $\{1,2\} \cup \{t \mid t > 2k+2\}$. In particular, it is 1 in positions 1 to $2k$ (in view of u), and it is also 1 in positions 3 to $2k+2$. But together these positions cover all bits 1 to $2k+2$. Thus $\psi(v) = v$. This completes the induction, and proves the lemma for all $2n \geqslant 6$. □

Remark: The condition $2n \geqslant 6$ in Lemma 7.3 is necessary. Here is a counter example for $2n = 4$: ψ fixes 0000 and 1111, and it maps α to $\overline{\alpha}$ for all $\alpha \in \{0,1\}^4$ with $\mathrm{wt}(\alpha) = 2$. If ψ were to be expressible as $\varphi_S \circ \pi$, then since $\psi(0000) = 0000$, we have $S = \emptyset$. Then by $\psi(0011) = 1100$ and $\psi(0101) = 1010$, the permutation π must map variable x_1 to x_4. However this violates $\psi(1001) = 0110$.

Lemma 7.4. *Let $\{G_{2n}\}$ be the sequence of graphs defined above.*

- *If $2n = 8$, then $\alpha(G_8) = \frac{1}{8}|\mathscr{E}_8| = 2^4$, and the maximum independent set I_8 of G_8 is unique up to an automorphism, where*

$$\begin{aligned} I_8 = \{&00000000, 00001111, 00110011, 00111100, 01010101, 01011010, 01100110, 01101001,\\ &10010110, 10011001, 10100101, 10101010, 11000011, 11001100, 11110000, 11111111\}. \end{aligned}$$

- *If $2n \geqslant 10$, then $\alpha(G_{2n}) < \frac{1}{8}|\mathscr{E}_{2n}| = 2^{2n-4}$.*

Proof. We first consider the case $2n = 6$. One can check that the following set

$$I_6 = \{000000, 001111, 110011, 111100\}$$

is an independent set of G_6. Thus, $\alpha(G_6) \geqslant 4$. Next, we show that $\alpha(G_6) = 4$ and I_6 is the unique maximum independent set of $\alpha(G_6)$ up to an automorphism.

Let J_6 be an maximum independent set of G_6. Clearly, $|J_6| \geqslant 4$. After an automorphism of G_6 by flipping some bits, we may assume that $\vec{0}^6 \in J_6$. Then for any $u \in \mathscr{E}_6$ with $\mathrm{wt}(u) = 2$, $u \notin J_6$. If $\vec{1}^6 \in J_6$, then for any $u \in \mathscr{E}_6$ with $\mathrm{wt}(u) = 4$, $u \notin J_6$. Thus, J_6 is maximal with $|J_6| = 2 < 4$, a contradiction. Thus, we have $\vec{1}^6 \notin J_6$. Then all vertices in J_6, except $\vec{0}^6$ have hamming weight 4. After an automorphism by permuting bits (this will not change $\vec{0}^6$), we may assume that $u = 001111 \in J_6$. Consider some other $v \in J_6$ with $\mathrm{wt}(v) = 4$. If $v_1v_2 = 01$ or 10, then $\mathrm{wt}(v_3v_4v_5v_6) = 3$. Thus, $\mathrm{wt}(u \oplus v) = \mathrm{wt}(00 \oplus v_1v_2) + \mathrm{wt}(1111 \oplus v_3v_4v_5v_6) = 1 + 1 = 2$. Contradiction. The only $v \in J_6$ with $\mathrm{wt}(v) = 4$, and $v_1v_2 = 00$ is $v = 001111 = u$. Thus, $v_1v_2 = 11$, i.e., both bits of v are 1 where u is 00. After an automorphism by permuting bits in $\{3, 4, 5, 6\}$ (this will not change $\vec{0}^6$ and u), we may assume that $v = 110011 \in J_6$. For any other $w \in J_6$ with $\mathrm{wt}(w) = 4$, we must have $w_1w_2 = 11$ (by the same proof for the pair (u, v) applied to (u, w)), and also $w_3w_4 = 11$ (by the same proof for the pair (u, v) applied to (v, w)). Thus, $w = 111100$. Then, $J_6 = \{\vec{0}^6, u, v, w\} = I_6$ is maximal. Thus, $\alpha(G_6) = 4$ and I_6 is the unique maximum independent set of $\alpha(G_6)$ up to an automorphism.

Consider $2n = 8$. One can check that I_8 is an independent set of G_8. Thus, $\alpha(G_8) \geqslant 16$. We use G_8^{ab} to denote the subgraph of G_8 induced by vertices $\{u \in \mathscr{E}_8 \mid u_1u_2 = ab\}$ for $(a, b) \in \mathbb{Z}_2^2$. Clearly, G_8^{00} and G_8^{11} are isomorphic to G_6, and G_8^{01} and G_8^{10} are isomorphic to H_6. Since H_6 is isomorphic to G_6, G_8^{01} and G_8^{10} are also isomorphic to G_6. Let J_8 be a maximum independent set of G_8. Clearly, $|J_8| \geqslant |I_8| = 16$. Also, we use J_8^{ab} to denote the subset $\{u \in J_8 \mid u_1u_2 = ab\}$ for $(a, b) \in \mathbb{Z}_2^2$. Similarly, we can define I_8^{ab}. Since J_8 is an independent set of G_8, clearly, for every $(a, b) \in \mathbb{Z}_2^2$, J_8^{ab} is an independent set of G_8^{ab}. Since G_8^{ab} is isomorphic to G_6 and $\alpha(G_6) = 4$, thus $|J_8^{ab}| \leqslant 4$. Then $|J_8| \leqslant 16$. Thus, $|J_8| = 16$, and $|J_8^{ab}| = 4$ for every $(a, b) \in \mathbb{Z}_2^2$. Since the maximum independent set of G_6 is unique up to an automorphism of G_6, which can be extended to an automorphism of G_8 by fixing the first two bits, we may assume that

$$J_8^{00} = I_8^{00} = \{00000000, 00001111, 00110011, 00111100\}$$

after an automorphism of G_8.

Then, consider J_8^{01}. We show that for any $u \in J_8^{01}$, $u_3 \neq u_4$, $u_5 \neq u_6$ and $u_7 \neq u_8$. Otherwise, by switching the pair of bits $\{3, 4\}$ with $\{5, 6\}$ or $\{7, 8\}$ (this will not change J_8^{00}), we may assume that $u_3 = u_4$. Then $\mathrm{wt}(u_1u_2u_3u_4)$ is odd. Since $\mathrm{wt}(u)$ is even, $\mathrm{wt}(u_5u_6u_7u_8)$ is odd. Thus, either $u_5 = u_6$ or $u_7 = u_8$. Still by switching the pair $\{5, 6\}$ with $\{7, 8\}$ (again this will not change J_8^{00}), we may assume that $u_5 = u_6$. Then since $\mathrm{wt}(u_5u_6u_7u_8)$ is odd, we have $u_7 \neq u_8$. Then, one can check that there exists some $v \in J_8^{00}$ such that $v_3v_4v_5v_6 = u_3u_4u_5u_6$. Since $v_1 = v_2$ and $u_1 \neq u_2$, $\mathrm{wt}(u_1u_2 \oplus v_1v_2) = 1$. Also since $v_7 = v_8$ and $u_7 \neq u_8$, $\mathrm{wt}(u_7u_8 \oplus v_7v_8) = 1$. Thus, $\mathrm{wt}(u \oplus v) = \mathrm{wt}(u_1u_2 \oplus v_1v_2) + \mathrm{wt}(u_7u_8 \oplus v_7v_8) = 2$. This means that the vertices u and v are connected in the graph G_8, a contradiction. Thus, for any $u \in J_8^{01}$, $u_3 \neq u_4$, $u_5 \neq u_6$ and $u_7 \neq u_8$. By permuting bit 3 with bit 4, bit 5 with bit 6, and bit 7 with bit 8 (this will not change J_8^{00}), we may assume that $01010101 \in J_8^{01}$. Consider some other $w \in J_8^{01}$. Since $w_{2i+1} \neq w_{2i+2}$ for any $i = 1, 2$ or 3, the pair $w_{2i+1}w_{2i+2} = 01$ or 10. Among these three pairs, let k denote the number of pairs that are 01. If $k = 3$, then $w = 01010101$. Contraction. If $k = 2$, then $\mathrm{wt}(01010101 \oplus w) = 2$. Contradiction. If $k = 0$, then $w = 01101010$. One can check that $\{01010101, 01101010\}$ is already a maximal independent set of G_8^{01} and it has size $2 < 4$. Contradiction. Thus, $k = 1$. Then, w can

take $\binom{3}{1}$ possible values. Thus,

$$J_8^{01} \subseteq I_8^{01} = \{01010101, 01011010, 01100110, 01101001\}.$$

Since, $|J_8^{01}| = 4$, $J_8^{01} = I_8^{01}$.

Consider some $u \in J_8^{10}$. Similar to the proof of J_8^{01}, we can show that $u_3 \neq u_4$, $u_5 \neq u_6$ and $u_7 \neq u_8$. Thus, u can take 2^3 possible values. Moreover, for any $01u' \in J_8^{01}$, $10u' \notin J_8^{10}$. Thus, there are only four remaining values that u can take. Then,

$$J_8^{10} \subseteq I_8^{10} = \{10010110, 10011001, 10100101, 10101010\}.$$

Since $|J_8^{10}| = 4$, $J_8^{10} = I_8^{10}$.

Finally, consider J_8^{11}. We show that for any $u \in J_8^{11}$, $u_3 = u_4$, $u_5 = u_6$ and $u_7 = u_8$. Otherwise, by permuting the pair of bits $\{3, 4\}$ with $\{5, 6\}$ or $\{7, 8\}$ (one can check that this will not change J_8^{01} and J_8^{10}), we may assume that $u_3 \neq u_4$. Since $\mathrm{wt}(u)$ is even, between $\mathrm{wt}(u_5u_6)$ and $\mathrm{wt}(u_7u_8)$, exactly one is even and the other is odd. By permuting the pair of bits $\{5, 6\}$ with $\{7, 8\}$, we may further assume that $u_5 \neq u_6$ and $u_7 = u_8$. Then, one can check that there exists some $v \in J_8^{01}$ such that $u_3u_4u_5u_6 = v_3v_4v_5v_6$. Since $u_1 = u_2$ and $v_1 \neq v_2$, $\mathrm{wt}(u_1u_2 \oplus v_1v_2) = 1$. Also since $u_7 = u_8$ and $v_7 \neq v_8$, $\mathrm{wt}(u_7u_8 \oplus v_7v_8) = 1$. Thus, $\mathrm{wt}(u \oplus v) = \mathrm{wt}(u_1u_2 \oplus v_1v_2) + \mathrm{wt}(u_7u_8 \oplus v_7v_8) = 2$. Contradiction. Thus, for any $u \in J_8^{11}$, it can take 2^3 possible values. Moreover, for any $00u' \in J_8^{00}$, we have $11u' \notin J_8^{11}$. Thus, there are only four remaining values that u can take. Then,

$$J_8^{11} \subseteq I_8^{11} = \{11000011, 11001100, 11110000, 11111111\}.$$

Thus, after an automorphism, $J_8 = I_8$. In other words, I_8 is the unique maximum independent set of G_8 up to an automorphism.

Now, we consider the case $2n \geqslant 10$. For every $(a, b) \in \mathbb{Z}_2^2$, we define G_{2n}^{ab} to be the subgraph of G_{2n} induced by $\{u \in G_{2n} \mid u_1u_2 = ab\}$, and it is isomorphic to G_{2n-2}. Thus,

$$\alpha(G_{2n}) \leqslant \alpha(G_{2n}^{00}) + \alpha(G_{2n}^{01}) + \alpha(G_{2n}^{10}) + \alpha(G_{2n}^{11}) = 4\alpha(G_{2n-2}).$$

Then, $\alpha(G_{2n-2}) < 2^{(2n-2)-4}$ will imply that $\alpha(G_{2n}) < 2^{2n-4}$. Thus, in order to prove $\alpha(G_{2n}) < 2^{2n-4}$ for all $2n \geqslant 10$, it suffices to prove that $\alpha(G_{10}) < 2^{10-4}$. For a contradiction, suppose that $\alpha(G_{10}) \geqslant 2^{10-4}$. Let I be a maximum independent set of G_{10}. Then, $|I| \geqslant 2^6$. We define $I^{ab} = \{u \in I \mid u_1u_2 = ab\}$ for every $(a, b) \in \mathbb{Z}_2$. Since G_{10}^{ab} is isomorphic to G_8 and $\alpha(G_8) = 2^4$, $|I^{ab}| \leqslant 2^4$ for every $(a, b) \in \mathbb{Z}_2^2$. Then, $|I| \leqslant 4 \cdot 2^4$. Thus, $|I| = 2^6$ and $|I^{ab}| = 2^4$ for every $(a, b) \in \mathbb{Z}_2^2$. Since the maximum independent set of G_8 is unique up to an automorphism of G_8 which can be extended to an automorphism of G_{10} by fixing the first two bits, we may assume that $I^{00} = \{00u \mid u \in I_8\}$.

Consider I^{01}. Since $|I^{01}| \neq 0$, there exists some $01v \in I^{01}$. Since $\mathrm{wt}(v)$ is odd, among $\mathrm{wt}(v_3v_4)$, $\mathrm{wt}(v_5v_6)$, $\mathrm{wt}(v_7v_8)$ and $\mathrm{wt}(v_9v_{10})$, there is an odd number (either one or three) of pairs such that $\mathrm{wt}(v_{2i+1}v_{2i+2})$ $(1 \leqslant i \leqslant 4)$ is odd, i.e., $v_{2i+1} \neq v_{2i+2}$. In other words, there are exactly three pairs among v_3v_4, v_5v_6, v_7v_8 and v_9v_{10} such that the values inside each pair are all equal with each other or all distinct with each other. By permuting these pairs of bits $\{3, 4\}$, $\{5, 6\}$, $\{7, 8\}$ and $\{9, 10\}$ (this will not change I^{00}), we may assume that either $v_3 = v_4$, $v_5 = v_6$, $v_7 = v_8$ and $v_9 \neq v_{10}$, or $v_3 \neq v_4$, $v_5 \neq v_6$, $v_7 \neq v_8$ and $v_9 = v_{10}$. In both cases, one can check that there exists some $00u \in I^{00}$ such that $u_i = v_i$ for $i \in \{3, \ldots, 8\}$. Moreover, $u_9 = u_{10}$ if $v_9 \neq v_{10}$, and $u_9 \neq u_{10}$ if $v_9 = v_{10}$. Then, $\mathrm{wt}(00u \oplus 01v) = \mathrm{wt}(00 \oplus 01) + \mathrm{wt}(u_9u_{10} \oplus v_9v_{10}) = 2$. This contradiction proves that $\alpha(G_{10}) < 2^{10-4}$, and the lemma is proved. □

Remark: We remark that $I_8 = \mathscr{S}(f_8)$. Later, we will see that the signature f_8, this Queen of the Night, and its support $\mathscr{S}(f_8)$ have even more extraordinary properties.

We consider a particular gadget construction that will be used in our proof. Let h_4 be a 4-ary signature with signature matrix $M_{12,34}(h_4) = H_4 = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & -1 & 0 \\ 1 & 0 & 0 & -1 \end{bmatrix}$. Notice that $H_4 H_4^{\mathsf{T}} = H_4 H_4 = 2I_4$, and h_4 is an affine signature. The following is called an H_4 gadget construction on f, denoted by ${}^{H_4}_{ij}f$. This is the signature obtained by connecting variables x_3 and x_4 of h_4 with variables x_i and x_j of f using $=_2$, respectively. Note that ${}^{H_4}_{ij}f$ is not necessarily realizable from f since h_4 may not be available. However, we will analyze the structure of f by analyzing ${}^{H_4}_{ij}f$. For convenience, we consider $(i,j) = (1,2)$ and we use $\widetilde{f}$ to denote ${}^{H_4}_{12}f$. The following results (Lemmas 7.5 and 7.6) about $\widetilde{f} = {}^{H_4}_{12}f$ hold for all ${}^{H_4}_{ij}f$ by replacing $\{1,2\}$ with $\{i,j\}$. Note that $\widetilde{f}$ has the following signature matrix

$$M_{12}(\widetilde{f}) = \begin{bmatrix} \widetilde{\mathbf{f}}_{12}^{00} \\ \widetilde{\mathbf{f}}_{12}^{01} \\ \widetilde{\mathbf{f}}_{12}^{10} \\ \widetilde{\mathbf{f}}_{12}^{11} \end{bmatrix} = H_4 M_{12}(f) = \begin{bmatrix} \mathbf{f}_{12}^{00} + \mathbf{f}_{12}^{11} \\ \mathbf{f}_{12}^{01} + \mathbf{f}_{12}^{10} \\ \mathbf{f}_{12}^{01} - \mathbf{f}_{12}^{10} \\ \mathbf{f}_{12}^{00} - \mathbf{f}_{12}^{11} \end{bmatrix} = \begin{bmatrix} \partial_{12}^{+}\mathbf{f} \\ \partial_{12}^{+}\mathbf{f} \\ \partial_{12}^{-}\mathbf{f} \\ \partial_{12}^{-}\mathbf{f} \end{bmatrix}.$$

We give the following relations between f and $\widetilde{f}$.

Lemma 7.5. *1. If f has even parity then $\widetilde{f}$ also has even parity.*
2. If $f \in \mathscr{A}$, then $\widetilde{f} \in \mathscr{A}$.
3. If $M(\mathfrak{m}_{12} f) = \lambda I_4$ for some real $\lambda \neq 0$, then $M(\mathfrak{m}_{12}\widetilde{f}) = 2\lambda I_4$.
4. If $\partial_{12}^{+}f, \partial_{12}^{-}f, \partial_{12}^{\widehat{+}}f, \partial_{12}^{\widehat{-}}f \in \mathscr{A}$, then $\widetilde{f}_{12}^{00}$, $\widetilde{f}_{12}^{01}$, $\widetilde{f}_{12}^{10}$, $\widetilde{f}_{12}^{11} \in \mathscr{A}$.
5. For $\{u,v\}$ disjoint with $\{1,2\}$ and $b \in \mathcal{B}$, if $\partial_{uv}^{b} f \in \mathscr{A}$, then $\partial_{uv}^{b}\widetilde{f} \in \mathscr{A}$.

Proof. Since h_4 has even parity and $h_4 \in \mathscr{A}$, the first and second propositions hold.

If $M(\mathfrak{m}_{12} f) = \lambda I_4$, then $M(\mathfrak{m}_{12}\widetilde{f}) = M_{12}(\widetilde{f}) M_{12}^{\mathsf{T}}(\widetilde{f}) = H_4 M_{12}(f) M_{12}^{\mathsf{T}}(f) H_4^{\mathsf{T}} = \lambda H_4 I_4 H_4^{\mathsf{T}} = 2\lambda I_4$. The third proposition holds.

By the matrix form $M_{12}(\widetilde{f})$, the fourth proposition holds.

Since the H_4 gadget construction only touches variables x_1 and x_2 of f, it commutes with merging gadgets on variables other than x_1 and x_2. Thus $\partial_{ij}^{b}\widetilde{f} = \widetilde{\partial_{ij}^{b} f}$. For all $b \in \mathcal{B}$ and all $\{i,j\}$ disjoint with $\{1,2\}$, if $\partial_{ij}^{b} f \in \mathscr{A}$ where $\partial_{ij}^{b} f$ are signatures realized by connecting variables x_i and x_j of f using b, then $\partial_{ij}^{b}\widetilde{f} = \widetilde{\partial_{ij}^{b} f} \in \mathscr{A}$. The last proposition holds. □

Clearly, if $f \in \int_{\mathcal{B}} \mathscr{A}$, then $\widetilde{f}_{12}^{00}$, $\widetilde{f}_{12}^{01}$, $\widetilde{f}_{12}^{10}$, $\widetilde{f}_{12}^{11} \in \mathscr{A}$. Thus, for every $(a,b) \in \mathbb{Z}_2^2$, if $\widetilde{f}_{12}^{ab} \not\equiv 0$, then its nonzero entries have the same norm, denoted by n_{ab}. Let $n_{ab} = 0$ if $\widetilde{f}_{12}^{ab} \equiv 0$. We have the following results regarding these norms n_{ab}.

Lemma 7.6. *Let f be an irreducible signature of arity $2n \geqslant 6$. Suppose that f has even parity, f satisfies* 2ND-ORTH *and $f \in \int_{\mathcal{B}} \mathscr{A}$.*

1. For any $(a,b),(c,d) \in \mathbb{Z}_2^2$, there exists some $k \in \mathbb{Z}$ such that $n_{ab} = \sqrt{2}^{k} n_{cd} \neq 0$, and $n_{ab} = n_{cd}$ iff $|\mathscr{S}(\widetilde{f}_{12}^{ab})| = |\mathscr{S}(\widetilde{f}_{12}^{cd})|$.

2. *Furthermore, if* $\widetilde{f}_{12}^{00}(\vec{0}^{2n-2}) \neq 0$ *and* $n_{00} > n_{11}$, *then* $\mathscr{S}(\widetilde{f}_{12}^{11}) = \mathscr{E}_{2n-2}$[1] *and* $n_{ab} = n_{11}$ *or* $2n_{11}$ *for every* $(a,b) \in \mathbb{Z}_2^2$; *in particular,* $n_{00} = 2n_{11}$. *Symmetrically, if* $\widetilde{f}_{12}^{11}(\vec{0}^{2n-2}) \neq 0$ *and* $n_{00} < n_{11}$, *then* $\mathscr{S}(\widetilde{f}_{12}^{00}) = \mathscr{E}_{2n-2}$ *and* $n_{ab} = n_{00}$ *or* $2n_{00}$ *for every* $(a,b) \in \mathbb{Z}_2^2$, *and* $n_{11} = 2n_{00}$.

Proof. Since f satisfies 2ND-ORTH, $M(\mathfrak{m}_{12} f) = \lambda I_4$ for some real $\lambda \neq 0$. Then, by Lemma 7.5, $M(\mathfrak{m}_{12}\widetilde{f}) = 2\lambda I_4 \neq 0$. Thus, $|\widetilde{f}_{12}^{ab}|^2 = 2\lambda \neq 0$ for every $(a,b) \in \mathbb{Z}_2^2$. Also, since $f \in \int_{\mathcal{B}} \mathscr{A}$, by Lemma 7.5, for every $(a,b) \in \mathbb{Z}_2^2$, $\widetilde{f}_{12}^{ab} \in \mathscr{A}$. Thus, $\mathscr{S}(\widetilde{f}_{12}^{ab})$ is affine and $|\mathscr{S}(\widetilde{f}_{12}^{ab})| = 2^{k_{ab}}$ for some integer $k_{ab} \geqslant 0$. Note that

$$|\widetilde{f}_{12}^{ab}|^2 = n_{ab}^2 \cdot |\mathscr{S}(\widetilde{f}_{12}^{ab})| = n_{ab}^2 \cdot 2^{k_{ab}} \neq 0.$$

Thus, for any $(a,b),(c,d) \in \mathbb{Z}_2^2$, $n_{ab}^2 \cdot 2^{k_{ab}} = n_{cd}^2 \cdot 2^{k_{cd}} \neq 0$. Then, $n_{ab} = \sqrt{2}^k n_{cd} \neq 0$ where $k = k_{cd} - k_{ab} \in \mathbb{Z}$. Clearly, $k = 0$ iff $|\mathscr{S}(\widetilde{f}_{12}^{ab})| = 2^{k_{ab}} = 2^{k_{cd}} = |\mathscr{S}(\widetilde{f}_{12}^{cd})|$.

Now we prove the second part of this lemma. We give the proof for the case that $\widetilde{f}_{12}^{00}(\vec{0}^{2n-2}) \neq 0$ and $n_{00} > n_{11}$. The proof of the case that $\widetilde{f}_{12}^{11}(\vec{0}^{2n-2}) \neq 0$ and $n_{00} < n_{11}$ is symmetric. We first show that $\mathscr{S}(\widetilde{f}_{12}^{11}) = \mathscr{E}_{2n-2}$. For a contradiction, suppose that $\mathscr{S}(\widetilde{f}_{12}^{11}) \neq \mathscr{E}_{2n-2}$. Since f has even parity, by Lemma 7.5, $\widetilde{f}$ has even parity. Then, $\widetilde{f}_{12}^{11}$ also has even parity. Thus, $\mathscr{S}(\widetilde{f}_{12}^{11}) \subsetneq \mathscr{E}_{2n-2}$. There exists $\theta \in \mathscr{E}_{2n-2}$ such that $\theta \notin \mathscr{S}(\widetilde{f}_{12}^{11})$. Also, since $n_{11} \neq 0$, $\widetilde{f}_{12}^{11} \not\equiv 0$. Then, $\mathscr{S}(\widetilde{f}_{12}^{11}) \neq \emptyset$ and there exists $\delta \in \mathscr{E}_{2n-2}$ such that $\delta \in \mathscr{S}(\widetilde{f}_{12}^{11})$. Then, we can find a pair $\alpha, \beta \in \mathscr{E}_{2n-2}$ where $\mathrm{wt}(\alpha \oplus \beta) = 2$ such that one is in $\mathscr{S}(\widetilde{f}_{12}^{11})$ and the other one is not in $\mathscr{S}(\widetilde{f}_{12}^{11})$.

- If $\mathrm{wt}(\alpha) \neq \mathrm{wt}(\beta)$, then clearly the difference between their Hamming weights is 2 since $\mathrm{wt}(\alpha \oplus \beta) = 2$. Thus, α and β differ in two bits i, j on which one takes value 00 and the other takes value 11.
- If $\mathrm{wt}(\alpha) = \mathrm{wt}(\beta)$, then they differ in two bits i, j on which one takes value 01 and the other takes value 10. Without loss of generality, we assume that $\alpha_i\alpha_j = 01$ and $\beta_i\beta_j = 10$. They take the same value on other bits. Since $\alpha, \beta \in \mathscr{E}_{2n-2}$ and $2n \geqslant 6$, they have even Hamming weight and length at least 4. Thus, there is another bit k such that on this bit, $\alpha_k = \beta_k = 1$. Consider $\gamma \in \mathscr{E}_{2n-2}$ where $\gamma_i\gamma_j\gamma_k = 000$ and γ takes the same value as α and β on other bits. Clearly, $\mathrm{wt}(\gamma) + 2 = \mathrm{wt}(\alpha) = \mathrm{wt}(\beta)$. If $\gamma \in \mathscr{S}(\widetilde{f}_{12}^{11})$, then between α and β, we pick the one that is not in $\mathscr{S}(\widetilde{f}_{12}^{11})$. Otherwise, we pick the one that is in $\mathscr{S}(\widetilde{f}_{12}^{11})$. In both cases, we can get a pair of inputs in $\mathscr{E}_{2n-2}$ such that one is in $\mathscr{S}(\widetilde{f}_{12}^{11})$ and the other is not in $\mathscr{S}(\widetilde{f}_{12}^{11})$, and they have Hamming distance 2 as well as different Hamming weights.

Thus, we can always find a pair $\alpha, \beta \in \mathscr{E}_{2n-2}$ where $\mathrm{wt}(\alpha \oplus \beta) = 2$ and α, β differ in two bits i, j on which one takes value 00 and the other takes value 11, such that one is in $\mathscr{S}(\widetilde{f}_{12}^{11})$ and the other is not in $\mathscr{S}(\widetilde{f}_{12}^{11})$. Clearly, $\{i, j\}$ is disjoint with $\{1, 2\}$.

Consider signatures $\partial_{ij}^{+}\widetilde{f}$ and $\partial_{ij}^{-}\widetilde{f}$. Then, $\widetilde{f}(11\alpha) + \widetilde{f}(11\beta)$ is an entry of $\partial_{ij}^{+}\widetilde{f}$, and $\widetilde{f}(11\alpha) - \widetilde{f}(11\beta)$ is an entry of $\partial_{ij}^{-}\widetilde{f}$. Since between $\widetilde{f}(11\alpha)$ and $\widetilde{f}(11\beta)$, exactly one is nonzero and it has norm n_{11}, we have

$$|\widetilde{f}(11\alpha) + \widetilde{f}(11\beta)| = |\widetilde{f}(11\alpha) - \widetilde{f}(11\beta)| = n_{11}.$$

Thus, both $\partial_{ij}^{+}\widetilde{f}$ and $\partial_{ij}^{-}\widetilde{f}$ have an entry with norm n_{11}. Let $\delta \in \mathscr{E}_{2n}$ where $\delta_i\delta_j = 11$ and δ takes value 0 on other bits. Then, clearly, $\widetilde{f}(\vec{0}^{2n}) + \widetilde{f}(\delta)$ is an entry of $\partial_{ij}^{+}\widetilde{f}$, and $\widetilde{f}(\vec{0}^{2n}) - \widetilde{f}(\delta)$ is an entry of $\partial_{ij}^{-}\widetilde{f}$.

[1]Here, $\mathscr{E}_{2n-2} = \{(x_3, \ldots, x_8) \in \mathbb{Z}_2^6 \mid x_3 + \cdots + x_8 = 0\}$. When context is clear, we do not specify the variables of $\mathscr{E}_{2n-2}$ and also $\mathscr{O}_{2n-2}$.

- If $\widetilde{f}(\delta) \neq 0$, then $|\widetilde{f}(\delta)| = n_{00}$ since $\delta_1\delta_2 = 00$. Since $\widetilde{f}(\vec{0}^{2n}) \neq 0$, $|\widetilde{f}(\vec{0}^{2n})| = n_{00}$. Thus, between $\widetilde{f}(\vec{0}^{2n}) + \widetilde{f}(\delta)$ and $\widetilde{f}(\vec{0}^{2n}) - \widetilde{f}(\delta)$, one has norm $2n_{00}$ and the other is zero. Therefore, between $\partial^+_{ij}\widetilde{f}$ and $\partial^-_{ij}\widetilde{f}$, one signature has an entry with norm $2n_{00}$. Remember that both $\partial^+_{ij}\widetilde{f}$ and $\partial^-_{ij}\widetilde{f}$ have an entry with norm n_{11}. Clearly, $2n_{00} > n_{11}$. Thus, between $\partial^+_{ij}\widetilde{f}$ and $\partial^-_{ij}\widetilde{f}$, there is a signature that has two entries with different norms. Clearly, such a signature is not in $\mathscr{A}$. However, since $f \in \int_{\mathcal{B}} \mathscr{A}$, by Lemma 7.5, $\partial^+_{ij}\widetilde{f}, \partial^-_{ij}\widetilde{f} \in \mathscr{A}$. Contradiction.
- If $\widetilde{f}(\delta) = 0$, then $|\widetilde{f}(\vec{0}^{2n}) + \widetilde{f}(\delta)| = |\widetilde{f}(\vec{0}^{2n})| = n_{00} > n_{11}$. Thus, $\partial^+_{ij}\widetilde{f}$ has two nonzero entries with different norms n_{00} and n_{11}. Then, $\partial^+_{ij}\widetilde{f} \notin \mathscr{A}$. Contradiction.

Thus, $\mathscr{S}(\widetilde{f}^{11}_{12}) = \mathscr{E}_{2n-2}$.

Then, we show that $n_{ab} = n_{11}$ or $2n_{11}$ for any $(a,b) \in \mathbb{Z}_2^2$. Clearly, we may assume that $(a,b) \neq (1,1)$. For a contradiction, suppose that $n_{ab} \neq n_{11}$ and $2n_{11}$. First, we show that $|\mathscr{S}(\widetilde{f}^{ab}_{12})| < 2^{2n-3}$ and $n_{ab} > n_{11}$. Since f has parity, $\widetilde{f}^{ab}_{12}$ also has parity (either even or odd depending on whether $\mathrm{wt}(ab) = 0$ or 1). Thus $|\mathscr{S}(\widetilde{f}^{ab}_{12})| \leqslant |\mathscr{E}_{2n-2}| = |\mathscr{O}_{2n-2}| = 2^{2n-3}$. If the equality holds, then $n_{ab} = n_{11}$ since $n_{ab}^2|\mathscr{S}(\widetilde{f}^{ab}_{12})| = n_{11}^2|\mathscr{S}(\widetilde{f}^{11}_{12})|$ and $|\mathscr{S}(\widetilde{f}^{11}_{12})| = 2^{2n-3}$. Contradiction. Thus, $|\mathscr{S}(\widetilde{f}^{ab}_{12})| < 2^{2n-3}$ and also $n_{ab} > n_{11}$.

Depending on whether f^{ab}_{12} has even parity or odd parity, we can pick a pair of inputs α, β with $\mathrm{wt}(\alpha \oplus \beta) = 2$ from $\mathscr{E}_{2n-2}$ or $\mathscr{O}_{2n-2}$ such that exactly one is in $\mathscr{S}(f^{ab}_{12})$ and the other is not in $\mathscr{S}(f^{ab}_{12})$. Suppose that α and β differ in bits i, j. Depending on whether $\alpha_i = \alpha_j$ or $\alpha_i \neq \alpha_j$, we can connect variables x_i and x_j of $\widetilde{f}$ using $=^+_2$ and $=^-_2$, or $\neq^+_2$ and $\neq^-_2$. We will get two signatures $\partial^+_{ij}\widetilde{f}$ and $\partial^-_{ij}\widetilde{f}$, or $\partial^{\widehat{+}}_{ij}\widetilde{f}$ and $\partial^{\widehat{-}}_{ij}\widetilde{f}$. We consider the case that $\alpha_i = \alpha_j$. For the case that $\alpha_i \neq \alpha_j$, the analysis is the same by replacing $\partial^+_{ij}\widetilde{f}$ and $\partial^-_{ij}\widetilde{f}$ with $\partial^{\widehat{+}}_{ij}\widetilde{f}$ and $\partial^{\widehat{-}}_{ij}\widetilde{f}$ respectively.

Consider signatures $\partial^+_{ij}\widetilde{f}$ and $\partial^-_{ij}\widetilde{f}$. Then, $\widetilde{f}(ab\alpha) + \widetilde{f}(ab\beta)$ is an entry of $\partial^+_{ij}\widetilde{f}$, and $\widetilde{f}(ab\alpha) - \widetilde{f}(ab\beta)$ is an entry of $\partial^-_{ij}\widetilde{f}$. Since between α and β, exactly one is in $\mathscr{S}(f^{ab}_{12})$, between $\widetilde{f}(ab\alpha)$ and $\widetilde{f}(ab\beta)$, exactly one is nonzero and it has norm n_{ab}. Thus,

$$|\widetilde{f}(ab\alpha) + \widetilde{f}(ab\beta)| = |\widetilde{f}(ab\alpha) - \widetilde{f}(ab\beta)| = n_{ab}.$$

Both $\partial^+_{ij}\widetilde{f}$ and $\partial^-_{ij}\widetilde{f}$ have an entry with norm n_{ab}.

Let $\alpha', \beta' \in \mathscr{E}_{2n-2}$ where $\alpha'_i\alpha'_j = \alpha_i\alpha_j$, $\alpha'_k = \alpha'_i \oplus \alpha'_j$ for some $k \neq i, j$[1] and α' takes value 0 on other bits, and $\beta'_i\beta'_j = \beta_i\beta_j$, $\beta'_k = \beta'_i \oplus \beta'_j$ for the same $k \neq i, j$ and β' takes value 0 on other bits. Clearly, α' and β' differ in bits i and j and they differ in the same way as α and β. Then, $\widetilde{f}(11\alpha') + \widetilde{f}(11\beta')$ is an entry of $\partial^+_{ij}\widetilde{f}$, and $\widetilde{f}(11\alpha') - \widetilde{f}(11\beta')$ is an entry of $\partial^-_{ij}\widetilde{f}$. Since $\mathscr{S}(\widetilde{f}^{11}_{12}) = \mathscr{E}_{2n-2}$, both $\widetilde{f}(11\alpha')$ and $\widetilde{f}(11\beta')$ are nonzero and they have norm n_{11}. Thus, between $\widetilde{f}(11\alpha') + \widetilde{f}(11\beta')$ and $\widetilde{f}(11\alpha') - \widetilde{f}(11\beta')$, exactly one is zero and the other has norm $2n_{11}$. Thus, between signatures $\partial_{ij}\widetilde{f}$ and $\partial^-_{ij}\widetilde{f}$, there is a signature that has two entries with different norms $2n_{11}$ and n_{ab}. Such a signature is not in $\mathscr{A}$. However, since $f \in \int_{\mathcal{B}} \mathscr{A}$, by Lemma 7.5, $\partial_{ij}\widetilde{f}, \partial^-_{ij}\widetilde{f} \in \mathscr{A}$. Contradiction. Thus, $n_{ab} = n_{11}$ or $2n_{11}$ for any $(a,b) \in \mathbb{Z}_2^2$. □

We also give the following results about multilinear polynomials $F(x_1, \ldots, x_n) \in \mathbb{Z}_2[x_1, \ldots, x_n]$. We use $d(F)$ to denote the total degree of F. For $\{i,j\} \subseteq \{1, \ldots, n\} = [n]$, we use $F^{ab}_{ij} \in Z_2[\{x_1, \ldots, x_n\} \backslash \{x_i, x_j\}]$ to denote the polynomial obtained by setting $(x_i, x_j) = (a,b)$ in F.

[1] Since $2n - 2 \geqslant 4$, such a k exists. Here, $\alpha'_k = 0$ since $\alpha_i = \alpha_j$ in this case under discussion. For the case that $\alpha_i \neq \alpha_j$, we have $\alpha'_k = 1$.

Definition 7.7. *Let $F(x_1, \ldots, x_n) \in \mathbb{Z}_2[x_1, \ldots, x_n]$ be a multilinear polynomial. We say F is a complete quadratic polynomial if $d(F) = 2$ and for all $\{i, j\} \subseteq [n]$, the quadratic term $x_i x_j$ appears in F. We say F is a complete cubic polynomial if $d(F) = 3$ and for all $\{i, j, k\} \subseteq [n]$, the cubic term $x_i x_j x_k$ appears in F.*

Lemma 7.8. *Let $F(x_1, \ldots, x_n) \in \mathbb{Z}_2[x_1, \ldots, x_n]$ be a multilinear polynomial.*

1. *If for all $\{i, j\} \subseteq [n]$, $F^{00}_{ij} + F^{11}_{ij} \equiv 0$ or 1, and $F^{01}_{ij} + F^{10}_{ij} \equiv 0$ or 1, then $d(F) \leqslant 2$. Moreover, if $d(F) = 2$, then F is a complete quadratic polynomial.*
2. *If for all $\{i, j\} \subseteq [n]$, $d(F^{00}_{ij} + F^{11}_{ij}) \leqslant 1$, and $d(F^{01}_{ij} + F^{10}_{ij}) \leqslant 1$, then $d(F) \leqslant 3$. Moreover, if $d(F) = 3$, then F is a complete cubic polynomial.*

Proof. We prove the first part. The proof for the second part is similar which we omit here.

For all $\{i, j\} \subseteq [n]$, we write $F \in \mathbb{Z}_2[x_1, \ldots, x_n]$ as a polynomial of variables x_i and x_j.

$$F = X_{ij} x_i x_j + Y_{ij} x_i + Z_{ij} x_j + W_{ij}$$

where $X_{ij}, Y_{ij}, Z_{ij}, W_{ij} \in \mathbb{Z}_2[\{x_3, \ldots, x_n\} \backslash \{x_i, x_j\}]$. Then,

$$F^{00}_{ij} = W_{ij} \quad \text{and} \quad F^{11}_{ij} = X_{ij} + Y_{ij} + Z_{ij} + W_{ij}.$$

Thus, $X_{ij} + Y_{ij} + Z_{ij} = F^{00}_{ij} + F^{11}_{ij} \equiv 0$ or 1. Also,

$$F^{01}_{ij} = Z_{ij} + W_{ij} \quad \text{and} \quad F^{10}_{ij} = Y_{ij} + W_{ij}.$$

Thus, $Y_{ij} + Z_{ij} = F^{01}_{ij} + F^{10}_{ij} \equiv 0$ or 1. Then, $X_{ij} \equiv 0$ or 1 for all $\{i, j\}$. Thus, $d(F) \leqslant 2$.

Suppose that $d(F) = 2$. then F has at least a quadratic term $x_u x_v$ $(u \neq v)$. Without loss of generality, we assume that the term $x_1 x_2$ appears in F. We first show that for all $2 \leqslant j \leqslant n$, the quadratic term $x_1 x_j$ appears in F. Since $x_1 x_2$ is already in F, we may assume that $3 \leqslant j$. We write F as a polynomial of variables x_2 and x_j.

$$F = X_{2j} x_2 x_j + Y_{2j} x_2 + Z_{2j} x_j + W_{2j},$$

where $X_{2j}, Y_{2j}, Z_{2j}, W_{2j}$ do not involve x_2 and x_j. Since $x_1 x_2$ appears in F, x_1 appears in Y_{2j}. As we have proved above, $Y_{2j} + Z_{2j} \equiv 0$ or 1. Thus, x_1 also appears in Z_{2j}, which means that $x_1 x_j$ appears in F. Then, for all $2 \leqslant j \leqslant n$, $x_1 x_j$ appears in F.

Then, for all $2 \leqslant i < j \leqslant n$, we write F as a polynomial of variables x_1 and x_i.

$$F = X_{1i} x_1 x_i + Y_{1i} x_1 + Z_{1i} x_i + W_{1i},$$

where $X_{1i}, Y_{1i}, Z_{1i}, W_{1i}$ do not involve x_1 and x_i. Since $x_1 x_j$ appears in F, x_j appears in Y_{1i}. Since $Y_{1i} + Z_{1i} \equiv 0$ or 1, x_j also appears in Z_{1i}. Thus, $x_i x_j$ appears in F. Then, for all $2 \leqslant i < j \leqslant n$, the quadratic term $x_i x_j$ appears in F. Thus, for all $\{i, j\} \subseteq [n]$, $x_i x_j$ appears in F. □

Now, we are ready to take a major step towards Theorem 7.19.

Lemma 7.9. *Let $2n \geqslant 8$ and let $f \in \mathcal{F}$ be a $2n$-ary irreducible signature with parity. Then,*

- $\text{Holant}^b(\mathcal{F})$ *is #P-hard, or*
- *there is a signature $g \notin \mathscr{A}$ of arity $2k < 2n$ that is realizable from f and $\mathcal{B}$, or*
- *after normalization, $f(\alpha) = \pm 1$ for all $\alpha \in \mathscr{S}(f)$.*

Proof. Since f is irreducible, we may assume that it satisfies 2ND-ORTH. Otherwise, we get #P-hardness by Lemma 4.4. Also, we may assume that $f \in \int_{\mathcal{B}} \mathscr{A}$. Otherwise, we can realize a signature of arity $2n-2$ that is not in $\mathscr{A}$ by merging f using some $b \in \mathcal{B}$.

For any four entries x, y, z, w of f on inputs $\alpha, \beta, \gamma, \delta \in \mathbb{Z}_2^{2n}$ written in the form of a 2-by-2 matrix $\left[\begin{smallmatrix} x & y \\ z & w \end{smallmatrix}\right] = \begin{bmatrix} f(\alpha) & f(\beta) \\ f(\gamma) & f(\delta) \end{bmatrix}$, we say that such a matrix is a *distance-2 square* if there exist four bits i, j, k, ℓ such that $\alpha_i\alpha_j = \beta_i\beta_j = \overline{\gamma_i\gamma_j} = \overline{\delta_i\delta_j}$, $\alpha_k\alpha_\ell = \gamma_k\gamma_\ell = \overline{\beta_k\beta_\ell} = \overline{\delta_k\delta_\ell}$ and α, β, γ and δ take the same values on other bits. An equivalent description is that

$$\delta = \alpha \oplus \beta \oplus \gamma, \quad \mathrm{wt}(\alpha \oplus \beta) = 2, \quad \mathrm{wt}(\alpha \oplus \gamma) = 2 \quad \text{and} \quad \mathrm{wt}(\alpha \oplus \delta) = 4. \tag{7.5}$$

Indeed (7.5) is clearly satisfied by any distance-2 square. Conversely, suppose (7.5) holds. If we flip any bit i in all α, β, γ and δ, both (7.5) and the bitwise description are invariant, and thus we may assume $\alpha = \vec{0}^{2n}$. By $\mathrm{wt}(\alpha \oplus \gamma) = 2$, there exist two bits i, j such that $\gamma_i\gamma_j = 11$, and γ takes 0 on other bits. By $\mathrm{wt}(\alpha \oplus \beta) = 2$, there exits two bits k, ℓ such that $\beta_k\beta_\ell = 11$, and β takes 0 on other bits. Since $\delta = \alpha \oplus \beta \oplus \gamma$, $\mathrm{wt}(\beta \oplus \gamma) = \mathrm{wt}(\alpha \oplus \delta) = 4$. Thus, the bits i, j, k, ℓ are distinct four bits. Then, $\delta_i\delta_j\delta_k\delta_\ell = 1111$ and δ takes 0 on other bits. Thus, α, β, γ and δ satisfy the bitwise description of distance-2 squares.

We give an example of such a distance-2 square. Let

$$\begin{bmatrix} x & y \\ z & w \end{bmatrix} = \begin{bmatrix} f(\alpha) & f(\beta) \\ f(\gamma) & f(\delta) \end{bmatrix} = \begin{bmatrix} f(0001\theta) & f(0010\theta) \\ f(1101\theta) & f(1110\theta) \end{bmatrix}$$

where $\theta \in \mathbb{Z}_2^{2n-4}$ is an arbitrary binary string of length $2n-4$. In this example, $(i, j) = (1, 2)$ and $(k, \ell) = (3, 4)$. We show next that such a distance-2 square $\left[\begin{smallmatrix} x & y \\ z & w \end{smallmatrix}\right]$ has the property described in (7.6) $\sim$ (7.9).

By connecting variables x_1 and x_2 of f using $=_2^+$ and $=_2^-$ respectively, we get signatures $\partial_{12}^+ f$ and $\partial_{12}^- f$. By our assumption, $\partial_{12}^+ f$ and $\partial_{12}^- f$ are affine signatures. Note that, $x+z$ and $y+w$ are entries of $\partial_{12}^+ f$ on inputs 01θ and $10\theta \in \mathbb{Z}_2^{2n-2}$. Since $\partial_{12}^+ f \in \mathscr{A}$, if $x+z$ and $y+w$ are both nonzero, then they have the same norm. Thus, we have $(x+z)(y+w) = 0$ or $(x+z)^2 = (y+w)^2$. Similarly, $x-z$ and $y-w$ are entries of $\partial_{12}^- f \in \mathscr{A}$. Thus, we have $(x-z)(y-w) = 0$ or $(x-z)^2 = (y-w)^2$.

Also, by connecting variables x_3 and x_4 of f using $\neq_2$ and $\neq_2^-$ respectively, we get signatures $\widehat{\partial}_{34}^+ f$ and $\widehat{\partial}_{34}^- f$ that are affine signatures. Note that, $x+y$ and $z+w$ are entries of $\widehat{\partial}_{34}^+ f$ on inputs 00θ and 11θ. Since $\widehat{\partial}_{34}^+ f \in \mathscr{A}$, we have $(x+y)(z+w) = 0$ or $(x+y)^2 = (z+w)^2$. Similarly, $x-y$ and $z-w$ are entries of $\widehat{\partial}_{34}^- f$. Then, we have $(x-y)(z-w) = 0$ or $(x-y)^2 = (z-w)^2$.

Now, consider an arbitrary distance-2 square $\left[\begin{smallmatrix} x & y \\ z & w \end{smallmatrix}\right] = \begin{bmatrix} f(\alpha) & f(\beta) \\ f(\gamma) & f(\delta) \end{bmatrix}$. Depending on whether $\alpha_i = \alpha_j$ or $\alpha_i \neq \alpha_j$, we can use $=_2^+$ and $=_2^-$, or $\neq_2^+$ and $\neq_2^-$ respectively, to connect variables x_i and x_j of f to produce two signatures $\partial_{ij}^+ f$ and $\partial_{ij}^- f$, or $\widehat{\partial}_{ij}^+ f$ and $\widehat{\partial}_{ij}^- f$ in either case, such that $x \pm z$ and $y \pm w$ are both entries of the resulting two signatures. Since the two resulting signatures are in affine, we have

$$(x+z)(y+w) = 0 \quad \text{or} \quad (x+z)^2 = (y+w)^2, \tag{7.6}$$

and

$$(x-z)(y-w) = 0 \quad \text{or} \quad (x-z)^2 = (y-w)^2. \tag{7.7}$$

Similarly, by connecting variables x_k and x_ℓ of f using either $=_2^\pm$ or $\neq_2^\pm$, we have

$$(x+y)(z+w) = 0 \quad \text{or} \quad (x+y)^2 = (z+w)^2 \tag{7.8}$$

and

$$(x-y)(z-w)=0 \quad \text{or} \quad (x-y)^2=(z-w)^2. \tag{7.9}$$

Now, we show that by solving equations (7.6) $\sim$ (7.9), every distance-2 square has one of the following forms (after normalization, row or column permutation, multiplying a -1 scalar of one row or one column, and taking transpose)

$$\underbrace{\begin{bmatrix}0&0\\0&0\end{bmatrix},\begin{bmatrix}1&0\\0&0\end{bmatrix},\begin{bmatrix}1&1\\0&0\end{bmatrix},\begin{bmatrix}1&0\\0&1\end{bmatrix},\begin{bmatrix}1&1\\1&1\end{bmatrix},\begin{bmatrix}1&1\\1&-1\end{bmatrix}}_{\text{type I}},\quad \underbrace{\begin{bmatrix}1&a\\a&1\end{bmatrix}(a>1)}_{\text{type II}},\quad \text{or}\quad \underbrace{\begin{bmatrix}1&1\\3&-1\end{bmatrix}}_{\text{type III}}.$$

We say that the first six forms are type I, and the other two are type II and type III respectively. These forms listed above are canonical forms of each type.

Let $\left[\begin{smallmatrix}x&y\\z&w\end{smallmatrix}\right]$ be a distance-2 square. Consider

$$p=(x+y)(z+w)(x+z)(y+w)(x-y)(z-w)(x-z)(y-w).$$

- If $p=0$, then among its eight factors (four sums and four differences), at least one factor is zero. By taking transpose and row permutation, we may assume that $x+y=0$ or $x-y=0$. If $x+y=0$, then by multiplying the column $\left[\begin{smallmatrix}y\\w\end{smallmatrix}\right]$ with -1, we can modify this distance-2 square to get $x-y=0$. Thus, we may assume that $x-y=0$. If $x=y=0$, then by (7.6), we have $z=0$ or $w=0$, or $z=\pm w$. Thus, after normalizing operations of row and column permutation and multiplication by -1, we reach the following canonical forms $\left[\begin{smallmatrix}0&0\\0&0\end{smallmatrix}\right]$, $\left[\begin{smallmatrix}1&0\\0&0\end{smallmatrix}\right]$ or $\left[\begin{smallmatrix}1&1\\0&0\end{smallmatrix}\right]$. Otherwise, $x=y\neq 0$. Consider $q=(x+z)(y+w)(x-z)(y-w)$.
  - If $q=0$, then among its four factors (two sums and two differences), at least one is zero. By column permutation on the matrix $\left[\begin{smallmatrix}x&y\\z&w\end{smallmatrix}\right]$ and multiplying the row (z,w) with -1 (which does not change the values of x and y), we may assume that $x-z=0$. Thus, $x=y=z\neq 0$. We normalize their values to 1. Then by (7.6), $1+w=0$ or $1+w=\pm 2$. Thus, $w=-1,1$ or -3. If $w=\pm 1$, then $\left[\begin{smallmatrix}x&y\\z&w\end{smallmatrix}\right]$ has the canonical form $\left[\begin{smallmatrix}1&1\\1&1\end{smallmatrix}\right]$ or $\left[\begin{smallmatrix}1&1\\1&-1\end{smallmatrix}\right]$. If $w=-3$, then $\left[\begin{smallmatrix}x&y\\z&w\end{smallmatrix}\right]=\left[\begin{smallmatrix}1&1\\1&-3\end{smallmatrix}\right]$ which has the canonical form $\left[\begin{smallmatrix}1&1\\3&-1\end{smallmatrix}\right]$ (Type III).
  - If $q\neq 0$, then $(x+z)(y+w)\neq 0$ and $(x-z)(y-w)\neq 0$. By equations (7.6) and (7.7), $(x+z)^2=(y+w)^2$ and $(x-z)^2=(y-w)^2$. Thus, $xz=yw$. Since $x=y\neq 0$, $z=w$. If $z=w=0$, then this gives the canonical form $\left[\begin{smallmatrix}1&1\\0&0\end{smallmatrix}\right]$. Otherwise, $z=w\neq 0$. Then $z+w\neq 0$ and hence by (7.8), $z+w=\pm(x+y)$. Since $z=w$ and $x=y$, we get $z=\pm x$. Thus, $x+z=0$ or $x-z=0$. Contradiction.
- If $p\neq 0$, then all its eight factors are nonzero. Thus by (7.6) $\sim$ (7.9), $(x+z)^2=(y+w)^2$, $(x-z)^2=(y-w)^2$, $(x+y)^2=(z+w)^2$ and $(x-y)^2=(z-w)^2$. By solving these equations, we have $x^2=w^2$, $y^2=z^2$, and $xy=zw$. If $x=y=z=w=0$, then it gives the canonical form $\left[\begin{smallmatrix}0&0\\0&0\end{smallmatrix}\right]$. Otherwise, by permuting rows and columns, we may assume that $x\neq 0$ and $|x|$ is the smallest among the norms of nonzero entries in $\left[\begin{smallmatrix}x&y\\z&w\end{smallmatrix}\right]$. We normalize x to 1. Since $x^2=w^2$, we get $w=\pm 1$. By multiplying the row (z,w) with -1 (which does not change $xy=zw$), we may assume that $w=1$. Then, $xy=zw$ implies that $y=z$. If $y=z=0$, then $\left[\begin{smallmatrix}x&y\\z&w\end{smallmatrix}\right]$ has the canonical form $\left[\begin{smallmatrix}1&0\\0&1\end{smallmatrix}\right]$. Otherwise, since $|x|=1$ is the smallest norm among nonzero entries, $y=z=\pm a$ where $a\geqslant 1$. If $a=1$ (i.e., $y=z=\pm 1$), then $\left[\begin{smallmatrix}x&y\\z&w\end{smallmatrix}\right]$ has the canonical form $\left[\begin{smallmatrix}1&1\\1&1\end{smallmatrix}\right]$. If $a>1$, then $\left[\begin{smallmatrix}x&y\\z&w\end{smallmatrix}\right]$ has the canonical form of Type II.

Thus, every distance-2 square has a canonical form of Type I, II or III.

Note that given a particular distance-2 square of f, by normalization, and renaming or flipping or negating variables of f, we can always modify this distance-2 square to get its canonical form.

Clearly, for signatures of arity at least 4, distance-2 squares exist. We consider the following two cases according to which types of distance-2 squares appear in f.

Case 1. All distance-2 squares in f are of type I.

We show that (after normalization) $f(\alpha) = \pm 1$ for all $\alpha \in \mathscr{S}(f)$. Since $f \not\equiv 0$, it has at least one nonzero entry. By normalization, we may assume that 1 is the smallest norm of all nonzero entries of f. Then by flipping variables of f, we may assume that $f(\vec{0}^{2n}) = 1$. For a contradiction, suppose that there is some $\beta \in \mathscr{S}(f)$ such that $f(\beta) \neq \pm 1$. Then by our assumption that 1 is the smallest norm and $|f(\beta)| \neq 1$, we have $|f(\beta)| > 1$. Also, since f has parity and $\vec{0}^{2n} \in \mathscr{S}(f)$, f has even parity. Thus, $\mathrm{wt}(\beta) \equiv 0 \pmod 2$. By renaming variables of f, we may assume that $\beta = \vec{1}^{2m}\vec{0}^{2n-2m}$, for some $m \geqslant 1$. (This does not affect the normalization $f(\vec{0}^{2n}) = 1$). Then, we show that for all $\alpha = \delta\vec{0}^{2n-2m}$ where $\delta \in \mathbb{Z}_2^{2m}$, $f(\alpha) = \pm 1$. We prove this by induction on $\mathrm{wt}(\delta)$. This will lead to a contradiction when $\mathrm{wt}(\delta) = 2m$, since $|f(\beta)| = |f(\vec{1}^{2m}\vec{0}^{2n-2m})| \neq 1$.

Since $f(\vec{0}^{2n}) = 1$, we may assume $\mathrm{wt}(\delta) \geqslant 2$. We first consider the base case that $\mathrm{wt}(\delta) = 2$. By renaming the first $2m$ variables, without loss of generality, we may assume that $\delta = 11\vec{0}^{2m-2}$ and then $\alpha = 11\vec{0}^{2n-2} = 11000\vec{0}^{2n-4}$. This renaming will not change β. Consider the following distance-2 square

$$\begin{bmatrix} x & y \\ z & w \end{bmatrix} = \begin{bmatrix} f(00000\vec{0}^{2n-4}) & f(11000\vec{0}^{2n-4}) \\ f(00110\vec{0}^{2n-4}) & f(11110\vec{0}^{2n-4}) \end{bmatrix}.$$

Recall our assumption that every distance-2 square is of type I. Here $x = f(\vec{0}^{2n})$, and $y = f(\alpha)$. Since $x = 1$, $\left[\begin{smallmatrix} x & y \\ z & w \end{smallmatrix}\right]$ being of type I implies that $y = 0$ or ± 1 (the normalization steps include possibly multiplying a row or a column by -1). We want to show that $|y| = 1$; for a contradiction, suppose that $y = 0$. We consider the following two extra entries of f, where $\bar{\delta} = 00\vec{1}^{2m-2}$.

$$x' = f(\bar{\delta}\vec{0}^{2n-2m}) = f(00\vec{1}^{2m-2}\vec{0}^{2n-2m}) \quad \text{and} \quad y' = f(\beta) = f(11\vec{1}^{2m-2}\vec{0}^{2n-2m}).$$

By connecting variables x_1 and x_2 of f using $=_2$ and $=_2^-$, we get signatures $\partial_{12}f$ and ∂_{12}^-f respectively. Note that both $x + y$ and $x' + y'$ are entries of $\partial_{12}f$. Since $\partial_{12}f \in \mathscr{A}$, we have $(x+y)(x'+y') = 0$ or $(x+y)^2 = (x'+y')^2$. We can also consider ∂_{12}^-f and get $(x-y)(x'-y') = 0$ or $(x-y)^2 = (x'-y')^2$. Since $x = 1$ and $y = 0$, we have

$$\Big[x' + y' = 0 \ \text{ or } \ (x' + y')^2 = 1\Big] \quad \text{and} \quad \Big[x' - y' = 0 \ \text{ or } \ (x' - y')^2 = 1\Big].$$

Recall that $|y'| = |f(\beta)| > 1$. Clearly $x' + y' = 0$ and $x' - y' = 0$ cannot be both true, otherwise $y' = 0$. Suppose one of them is true, then $x' = \pm y'$. And at least one of $(x'+y')^2 = 1$ or $(x'-y')^2 = 1$ holds. So either $|x' + y'| = 1$ or $|x' - y'| = 1$. Substituting $x' = \pm y'$ we reach a contradiction to $|y'| > 1$. So neither $x' + y' = 0$ nor $x' - y' = 0$ holds. Then $(x' + y')^2 = 1$ and $(x' - y')^2 = 1$. Subtracting them, we get $x'y' = 0$, and since $y' \neq 0$, we get $x' = 0$. But then this contradicts $|y'| > 1$ and $(x' + y')^2 = 1$. Therefore, $y \neq 0$. Then, $y = \pm 1$. Thus, $y = f(\delta\vec{0}^{2n-2m}) = \pm 1$ for all δ with $\mathrm{wt}(\delta) = 2$.

If $2m = 2$, then the induction is finished. Otherwise, $2m > 2$. Inductively for some $2k \geqslant 2$, we assume that $f(\theta\vec{0}^{2n-2m}) = \pm 1$ for all $\theta \in \mathbb{Z}_2^{2m}$ with $\mathrm{wt}(\theta) \leqslant 2k < 2m$. Let δ be such that $\mathrm{wt}(\delta) = 2k + 2 \leqslant 2m$ and we show that $f(\delta\vec{0}^{2n-2m}) = \pm 1$. Since $\mathrm{wt}(\delta) = 2k + 2 \geqslant 4$, we can find four bits of δ such that the values of δ are 1 on these four bits. Without loss of generality, we assume that they are the first four bits, i.e. $\delta = 1111\delta'$ where $\delta' \in \mathbb{Z}_2^{2m-4}$. Consider the following

distance-2 square

$$\begin{bmatrix} x & y \\ z & w \end{bmatrix} = \begin{bmatrix} f(0000\delta'\vec{0}^{2n-2m}) & f(0011\delta'\vec{0}^{2n-2m}) \\ f(1100\delta'\vec{0}^{2n-2m}) & f(1111\delta'\vec{0}^{2n-2m}) \end{bmatrix}.$$

Clearly, three entries in this distance-2 square have input strings of weight at most $2k$, namely $\mathrm{wt}(0000\delta'\vec{0}^{2n-2m}) = 2k-2$, and $\mathrm{wt}(0011\delta'\vec{0}^{2n-2m}) = \mathrm{wt}(1100\delta'\vec{0}^{2n-2m}) = 2k$. By our induction hypothesis, $x, y, z \in \{1, -1\}$. Then, since the distance-2 square $\left[\begin{smallmatrix} x & y \\ z & w \end{smallmatrix}\right]$ is of type I, we have $w = f(\delta\vec{0}^{2n-2m}) = \pm 1$. The induction is complete. This finishes the proof of Case 1.

Case 2. There is a type II or type III distance-2 square in f.

This is the case where signatures g_8 and g_8' appear. We handle this case in two steps.

Step 1. We show that after flipping variables of f, $\mathscr{S}(f) = \mathscr{E}_{2n}$, and after normalization $f(\alpha) = \pm 1$ or ± 3 for all $\alpha \in \mathscr{S}(f)$. Let $\mathscr{S}_3(f) = \{\alpha \in \mathscr{S}(f) \mid f(\alpha) = \pm 3\}$. We also show that $|\mathscr{S}_3(f)| = 2^{2n-4} = \frac{1}{8}|\mathscr{S}(f)|$, and for any distinct $\alpha, \beta \in \mathscr{S}_3(f)$, $\mathrm{wt}(\alpha \oplus \beta) \geqslant 4$.

We first consider the case that there is a Type II distance-2 square in f. We show that the only possible Type II distance-2 square in f has the canonical form $\left[\begin{smallmatrix} 1 & 3 \\ 3 & 1 \end{smallmatrix}\right]$. Suppose that a distance-2 square of Type II appears in f. By flipping and negating variables, we modify f such that this distance-2 square is in its canonical form $\left[\begin{smallmatrix} 1 & a \\ a & 1 \end{smallmatrix}\right]$ $(a > 1)$. Also, by flipping variables and renaming variables, we may assume that this distance-2 square appears on inputs α, β, γ and δ where

$$\begin{bmatrix} f(\alpha) & f(\beta) \\ f(\gamma) & f(\delta) \end{bmatrix} = \begin{bmatrix} f(0000\vec{0}^{2n-4}) & f(0011\vec{0}^{2n-4}) \\ f(1100\vec{0}^{2n-4}) & f(1111\vec{0}^{2n-4}) \end{bmatrix} = \begin{bmatrix} 1 & a \\ a & 1 \end{bmatrix}.$$

Then, we consider the entries of $\widetilde{f}$ on inputs α, β, γ and δ. We have

$$\begin{bmatrix} \widetilde{f}(\alpha) & \widetilde{f}(\beta) \\ \widetilde{f}(\gamma) & \widetilde{f}(\delta) \end{bmatrix} = \begin{bmatrix} f(\alpha)+f(\gamma) & f(\beta)+f(\delta) \\ f(\alpha)-f(\gamma) & f(\beta)-f(\delta) \end{bmatrix} = \begin{bmatrix} 1+a & 1+a \\ 1-a & a-1 \end{bmatrix}.$$

Since $a > 1$, clearly $1 + a \neq 0$, $1 - a \neq 0$ and $|1 + a| > |1 - a|$. Since f has parity and $f(\vec{0}^{2n}) = 1$, f has even parity. By Lemma 7.6(2), $\mathscr{S}(\widetilde{f}_{12}^{11}) = \mathscr{E}_{2n-2}$ and $|1 + a| = 2|1 - a|$. Since $a > 1$, we have $1 + a = 2(a - 1)$. Then, $a = 3$. Thus, the only possible Type II distance-2 square in f has the canonical form $\left[\begin{smallmatrix} 1 & 3 \\ 3 & 1 \end{smallmatrix}\right]$.

Under the assumption that a Type II distance-2 square appears in f and $\left[\begin{smallmatrix} f(\alpha) & f(\beta) \\ f(\gamma) & f(\delta) \end{smallmatrix}\right] = \left[\begin{smallmatrix} 1 & 3 \\ 3 & 1 \end{smallmatrix}\right]$, we have $\left[\begin{smallmatrix} \widetilde{f}(\alpha) & \widetilde{f}(\beta) \\ \widetilde{f}(\gamma) & \widetilde{f}(\delta) \end{smallmatrix}\right] = \left[\begin{smallmatrix} 4 & 4 \\ -2 & 2 \end{smallmatrix}\right]$. As showed above, by Lemma 7.6(2), $\mathscr{S}(\widetilde{f}_{12}^{11}) = \mathscr{E}_{2n-2}$ and $n_{01}, n_{10} = 2$ or 4. We first prove

> **Claim 1.** *$\mathscr{S}(f_{12}^{00}) = \mathscr{S}(f_{12}^{11}) = \mathscr{E}_{2n-2}$, $f_{12}^{00}(\theta), f_{12}^{11}(\theta) = \pm 3$ or ± 1 for all $\theta \in \mathscr{E}_{2n-2}$, and $|\mathscr{S}_3(f_{12}^{00})| + |\mathscr{S}_3(f_{12}^{11})| = 2^{2n-5}$.*

Remember that $\widetilde{f}_{12}^{00}, \widetilde{f}_{12}^{11} \in \mathscr{A}$. For any of them, its nonzero entries have the same norm. Since $\widetilde{f}(\alpha) = \widetilde{f}(000\vec{0}^{2n-2}) = 1 + 3 = 4$ and $\mathscr{S}(\widetilde{f}_{12}^{00}) \subseteq \mathscr{E}_{2n-2}$, for every $\theta \in \mathscr{E}_{2n-2}$, $\widetilde{f}(00\theta) = \pm 4$ or 0. Also, since $\widetilde{f}(\gamma) = \widetilde{f}(110\vec{0}^{2n-2}) = 1 - 3 = -2$, and $\mathscr{S}(\widetilde{f}_{12}^{11}) = \mathscr{E}_{2n-2}$, for every $\theta \in \mathscr{E}_{2n-2}$, $\widetilde{f}(11\theta) = \pm 2$. Then,

$$f(00\theta) = \frac{\widetilde{f}(00\theta) + \widetilde{f}(11\theta)}{2} = \frac{(\pm 4) + (\pm 2)}{2} \quad \text{or} \quad \frac{0 + (\pm 2)}{2}.$$

Thus, $f(00\theta) = \pm 3$ or ± 1 for every $\theta \in \mathscr{E}_{2n-2}$. Also,

$$f(11\theta) = \frac{\widetilde{f}(00\theta) - \widetilde{f}(11\theta)}{2} = \frac{(\pm 4) - (\pm 2)}{2} \quad \text{or} \quad \frac{0 - (\pm 2)}{2}.$$

Thus, $f(11\theta) = \pm 3$ or ± 1 for every $\theta \in \mathscr{E}_{2n-2}$. Additionally note that, for any $\theta \in \mathscr{E}_{2n-2}$ if $\widetilde{f}(00\theta) = \pm 4$, then of the two values $f(00\theta)$ and $f(11\theta)$, exactly one is ± 3 and the other one is ± 1; if $\widetilde{f}(00\theta) = 0$, then $f(00\theta) = \pm 1$ and $f(11\theta) = \pm 1$. Since

$$|\widetilde{f}_{12}^{00}|^2 = 4^2 \cdot |\mathscr{S}(\widetilde{f}_{12}^{00})| = |\widetilde{f}_{12}^{11}|^2 = 2^2 \cdot |\mathscr{S}(\widetilde{f}_{12}^{11})| = 2^2 \cdot |\mathscr{E}_{2n-2}|,$$

we have $|\mathscr{S}(\widetilde{f}_{12}^{00})| = |\mathscr{E}_{2n-2}|/4 = 2^{2n-5}$. Thus, there are exactly 2^{2n-5} entries of $\widetilde{f}_{12}^{00}$ having value ± 4, which give arise to exactly 2^{2n-5} many entries of value ± 3 among all entries of f_{12}^{00} and f_{12}^{11}. Claim 1 has been proved.

Next, we prove

Claim 2. *$\mathscr{S}(f_{12}^{01}) = \mathscr{S}(f_{12}^{10}) = \mathscr{O}_{2n-2}$, $f_{12}^{01}(\theta), f_{12}^{10}(\theta) = \pm 3$ or ± 1 for all $\theta \in \mathscr{O}_{2n-2}$, and $|\mathscr{S}_3(f_{12}^{01})| + |\mathscr{S}_3(f_{12}^{10})| = 2^{2n-5}$.*

We have $\widetilde{f}(\vec{0}^{2n}) = 4$. We have $n_{00} = 4$ and $n_{11} = 2$. Also recall that we have showed that $n_{01}, n_{10} = 2$ or 4, by Lemma 7.6(2). There are three cases.

- $n_{01} = n_{10} = 2$. Since $n_{11} = n_{01} = 2$ and

  $$|\widetilde{f}_{12}^{11}|^2 = n_{11}^2 \cdot |\mathscr{S}(\widetilde{f}_{12}^{11})| = n_{01}^2 \cdot |\mathscr{S}(\widetilde{f}_{12}^{01})| = |\widetilde{f}_{12}^{01}|^2,$$

  we have

  $$|\mathscr{S}(\widetilde{f}_{12}^{01})| = |\mathscr{S}(\widetilde{f}_{12}^{11})| = |\mathscr{E}_{2n-2}| = 2^{2n-3}.$$

  Since $\widetilde{f}$ has even parity, $\mathscr{S}(\widetilde{f}_{12}^{01}) \subseteq \mathscr{O}_{2n-2}$. As $|\mathscr{O}_{2n-2}| = 2^{2n-3}$, we get $\mathscr{S}(\widetilde{f}_{12}^{01}) = \mathscr{O}_{2n-2}$. Similarly, we can show that $\mathscr{S}(\widetilde{f}_{12}^{10}) = \mathscr{O}_{2n-2}$. Let $\zeta = 01100\vec{0}^{2n-4}$ and $\eta = 10100\vec{0}^{2n-4}$. Then, $\widetilde{f}(\zeta) = \pm 2$ and $\widetilde{f}(\eta) = \pm 2$. Note that

  $$f(\zeta) = \frac{\widetilde{f}(\zeta) + \widetilde{f}(\eta)}{2} \quad \text{and} \quad f(\eta) = \frac{\widetilde{f}(\zeta) - \widetilde{f}(\eta)}{2}.$$

  If $\widetilde{f}(\zeta) = \widetilde{f}(\eta)$, then $f(\zeta) = \pm 2$ and $f(\eta) = 0$. If $\widetilde{f}(\zeta) = -\widetilde{f}(\eta)$, then $f(\zeta) = 0$ and $f(\eta) = \pm 2$. We first consider the case that $f(\zeta) = \pm 2$. Let $\xi = 10010\vec{0}^{2n-4}$. Consider the following distance-2 square.

  $$\begin{bmatrix} f(\alpha) & f(\zeta) \\ f(\xi) & f(\delta) \end{bmatrix} = \begin{bmatrix} f(00000\vec{0}^{2n-4}) & f(01100\vec{0}^{2n-4}) \\ f(10010\vec{0}^{2n-4}) & f(11110\vec{0}^{2n-4}) \end{bmatrix} = \begin{bmatrix} 1 & \pm 2 \\ * & 1 \end{bmatrix}.$$

  Clearly, it is not of type I nor type III. Also, it is not of type II with the canonical form $\left[\begin{smallmatrix} 1 & 3 \\ 3 & 1 \end{smallmatrix}\right]$. Contradiction. If $f(\eta) = \pm 2$, then similarly by considering the distance-2 square $\left[\begin{smallmatrix} f(\alpha) & f(\eta) \\ f(\tau) & f(\delta) \end{smallmatrix}\right]$ where $\tau = 01010\vec{0}^{2n-4}$, we get a contradiction.

- $n_{01} = n_{10} = 4$. We still consider

$$f(\zeta) = \frac{\widetilde{f}(\zeta) + \widetilde{f}(\eta)}{2} \quad \text{and} \quad f(\eta) = \frac{\widetilde{f}(\zeta) - \widetilde{f}(\eta)}{2}, \quad \text{where } \zeta = 01100\vec{0}^{2n-4} \text{ and } \eta = 10100\vec{0}^{2n-4}.$$

Then, as ζ has leading bits 01 and η has leading bits 10,

$$f(\zeta) = \frac{(\pm 4) + (\pm 4)}{2}, \frac{(\pm 4) + 0}{2} \text{ or } \frac{0+0}{2} \quad \text{and} \quad f(\eta) = \frac{(\pm 4) - (\pm 4)}{2}, \pm\frac{(\pm 4) - 0}{2} \text{ or } \frac{0-0}{2}.$$

Thus, $f(\zeta), f(\eta) = \pm 4, \pm 2$ or 0. If $f(\zeta)$ or $f(\eta) = \pm 4, \pm 2$, then by considering the distance-2 square $\begin{bmatrix} f(\alpha) & f(\zeta) \\ f(\xi) & f(\delta) \end{bmatrix}$ or $\begin{bmatrix} f(\alpha) & f(\eta) \\ f(\tau) & f(\delta) \end{bmatrix}$, we still get a contradiction. Thus we have $f(\zeta) = f(\eta) = 0$. Then, consider the signature ${}^{H_4}_{23}f$, denoted by $\widetilde{f}'$. Since f has even parity, f satisfies 2ND-ORTH and $f \in \int_{\mathcal{B}} \mathscr{A}$, $\widetilde{f}'$ has even parity, $\widetilde{f}'^{00}_{23}, \widetilde{f}'^{01}_{23}, \widetilde{f}'^{10}_{23}, \widetilde{f}'^{11}_{23} \in \mathscr{A}$. Let $n'_{00}, n'_{01}, n'_{10}$ and n'_{11} denote the norms of nonzero entries in $\widetilde{f}'^{00}_{23}, \widetilde{f}'^{01}_{23}, \widetilde{f}'^{10}_{23}$, and $\widetilde{f}'^{11}_{23}$ respectively. Notice that

$$\widetilde{f}'(\alpha) = \widetilde{f}'(\vec{0}^{2n}) = f(00000\vec{0}^{2n-4}) + f(01100\vec{0}^{2n-4}) = f(\alpha) + f(\zeta) = 1 + 0 = 1.$$

Thus, $n'_{00} = 1$. Also, notice that

$$\widetilde{f}'(\gamma) = \widetilde{f}'(11000\vec{0}^{2n-4}) = f(10100\vec{0}^{2n-4}) - f(11000\vec{0}^{2n-4}) = f(\eta) - f(\gamma) = 0 - 3 = -3.$$

Thus, $n'_{10} = 3$. But by Lemma 7.6(1), $n'_{00} = \sqrt{2}^k n'_{10}$ for some $k \in \mathbb{Z}$. However, clearly, $3 \neq \sqrt{2}^k$ for any $k \in \mathbb{Z}$. Contradiction.
- Therefore exactly one of n_{01} and n_{10} is 2 and the other is 4. Let $(a, b) = (0, 1)$ or $(1, 0)$ be such that $n_{ab} = 2$. Since $n_{11} = 2$ and $|\mathscr{S}(\widetilde{f}^{11}_{12})| = |\mathscr{E}_{2n-2}| = 2^{2n-3}$, we have $|\mathscr{S}(\widetilde{f}^{ab}_{12})| = 2^{2n-3}$. Since $\widetilde{f}$ has even parity, $\widetilde{f}^{ab}_{12}$ has odd parity, thus $\mathscr{S}(\widetilde{f}^{ab}_{12}) = \mathscr{O}_{2n-2}$. Then, similar to the proof of f^{00}_{12} and f^{11}_{12}, we can show that for every $\theta \in \mathscr{O}_{2n-2}$, $f^{01}_{12}(\theta), f^{10}_{12}(\theta) = \pm 3$ or ± 1. Also, among f^{01}_{12} and f^{10}_{12}, exactly 2^{2n-5} many entries are ± 3.

This completes the proof of Claim 2.

Thus, combining Claim 1 and Claim 2, $\mathscr{S}(f) = \mathscr{E}_{2n}$, $f(\alpha) = \pm 1$ or ± 3 for all $\alpha \in \mathscr{S}(f)$, and $|\mathscr{S}_3(f)| = 2^{2n-4} = \frac{1}{8}|\mathscr{S}(f)|$. Also remember that by our assumption, $f(\vec{0}^{2n}) = 1$.

Now, we show that for any distinct $\alpha, \beta \in \mathscr{S}_3(f)$, $\mathrm{wt}(\alpha \oplus \beta) \geqslant 4$. For a contradiction, suppose that $\alpha, \beta \in \mathscr{S}_3(f)$ and $\mathrm{wt}(\alpha \oplus \beta) = 2$, and they differ at bits i and j. By renaming variables, without loss of generality, we may assume that $\{i, j\} = \{1, 2\}$. This renaming does not change the value of $f(\vec{0}^{2n}) = 1$. Since $f(110\vec{0}^{2n-2}) = \pm 1$ or ± 3, of the values $f(000\vec{0}^{2n-2}) + f(110\vec{0}^{2n-2})$ and $f(000\vec{0}^{2n-2}) - f(110\vec{0}^{2n-2})$, which are respectively an entry of $\widetilde{f}^{00}_{12}$ and an entry of $\widetilde{f}^{11}_{12}$, at least one has norm 2. Thus, among n_{00} and n_{11}, at least one is 2. Since $f(\alpha) = \pm 3$ and $f(\beta) = \pm 3$, among $f(\alpha) + f(\beta)$ and $f(\alpha) - f(\beta)$, exactly one has norm 6 and the other has norm 0. Clearly, $f(\alpha) + f(\beta)$ and $f(\alpha) - f(\beta)$ are entries of $\widetilde{f}$ since α and β differ at bits 1 and 2. Thus, among n_{00}, n_{01}, n_{10} and n_{11}, one has norm 6. By Lemma 7.6(1), $2 = \sqrt{2}^k \cdot 6$ for some $k \in \mathbb{N}$. Contradiction. This proves that for any distinct $\alpha, \beta \in \mathscr{S}_3(f)$, $\mathrm{wt}(\alpha \oplus \beta) \geqslant 4$.

We have established the goal laid out in Step 1 of Case 2 under the assumption that there is a Type II distance-2 square in f.

Finally, within Step 1 of Case 2, we consider the case that a type III distance-2 square appears in f. By flipping and negating variables, we modify f such that this distance-2 square is in its

canonical form $\left[\begin{smallmatrix}1 & 3\\ 1 & -1\end{smallmatrix}\right]$. Also, by flipping variables and renaming variables, still we may assume that this distance-2 square appears on inputs α, β, γ and δ where

$$\begin{bmatrix} f(\alpha) & f(\beta) \\ f(\gamma) & f(\delta) \end{bmatrix} = \begin{bmatrix} f(00000\vec{0}^{2n-4}) & f(00110\vec{0}^{2n-4}) \\ f(11000\vec{0}^{2n-4}) & f(11110\vec{0}^{2n-4}) \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 3 & -1 \end{bmatrix}.$$

Then, we consider the entries of $\widetilde{f}$ on inputs α, β, γ and δ. We have

$$\begin{bmatrix} \widetilde{f}(\alpha) & \widetilde{f}(\beta) \\ \widetilde{f}(\gamma) & \widetilde{f}(\delta) \end{bmatrix} = \begin{bmatrix} f(\alpha)+f(\gamma) & f(\beta)+f(\delta) \\ f(\alpha)-f(\gamma) & f(\beta)-f(\delta) \end{bmatrix} = \begin{bmatrix} 4 & 0 \\ -2 & 2 \end{bmatrix}.$$

Then exactly in the same way as the above proof when $\left[\begin{smallmatrix}\widetilde{f}(\alpha) & \widetilde{f}(\beta)\\ \widetilde{f}(\gamma) & \widetilde{f}(\delta)\end{smallmatrix}\right] = \left[\begin{smallmatrix}4 & 4\\ -2 & 2\end{smallmatrix}\right]$, we can show that the same result holds. Thus, $\mathscr{S}(f) = \mathscr{E}_{2n}$, $f(\alpha) = \pm 1$ or ± 3 for all $\alpha \in \mathscr{S}(f)$, $|\mathscr{S}_3(f)| = 2^{2n-4} = \frac{1}{8}|\mathscr{S}(f)|$, and for any distinct $\alpha, \beta \in \mathscr{S}(f)$ with $\mathrm{wt}(\alpha \oplus \beta) = 2$, α and β cannot be both in $\mathscr{S}_3(f)$.

This finishes the proof of Step 1 of Case 2.

Step 2. Now we show that either g_8 or g_8' is realizable from f. We will show that they are both irreducible and do not satisfy 2ND-ORTH, which gives #P-hardness.

We define a graph G_{2n} with vertex set $\mathscr{E}_{2n}$, and there is an edge between α and β if $\mathrm{wt}(\alpha \oplus \beta) = 2$. I.e., we view every $\alpha \in \mathscr{E}_{2n}$ as a vertex, and the edges are distance 2 neighbors in Hamming distance. Then, $\mathscr{S}_3(f)$ is an independent set of G_{2n}. Remember that $2n \geqslant 8$ by the hypothesis of the lemma. If $2n \geqslant 10$, then by Lemma 7.4, $|\mathscr{S}_3(f)| < \frac{1}{8}|\mathscr{S}(f)|$. Contradiction. Thus, $2n = 8$. After renaming and flipping variables, we may assume that $\mathscr{S}_3(f) = I_8 = \mathscr{S}(f_8)$. For brevity of notation, let $S = \mathscr{E}_8$ and $T = \mathscr{S}(f_8)$. We can pick $(x_1, \ldots, x_7)$ as a set of free variables of $S = \mathscr{E}_8$. Then, there exists a multilinear polynomial $F(x_1, \ldots, x_7) \in \mathbb{Z}_2[x_1, \ldots, x_7]$, and a multilinear polynomial $G(x_1, \ldots, x_8) \in \mathbb{Z}_2[x_1, \ldots, x_8]$ that is viewed as a representative for its image in the quotient algebra $\mathbb{Z}_2[x_1, \ldots, x_8]/(P_1, P_2, P_3, P_4)$ where P_1, P_2, P_3, P_4 are the four linear polynomials in (7.4) such that T is decided by $P_1 = P_2 = P_3 = P_4 = 0$, such that

$$f = \chi_S(-1)^{F(x_1,\ldots,x_7)} + 4\chi_T(-1)^{G(x_1,\ldots,x_8)}.$$

We note that such multilinear polynomials $F(x_1, \ldots, x_7)$ and $G(x_1, \ldots, x_8)$ exist: For any point in $S \setminus T$ we can choose a unique value $s \in \mathbb{Z}_2$ which represents the ± 1 value of f as $(-1)^s$, and for any point in $T \subseteq S$ we can choose unique values $t \in \mathbb{Z}_2$ and $s' \in \mathbb{Z}_2$ such that $(-1)^{s'} + 4(-1)^t$ represents the ± 3 value of f.

For $\{i, j\} \subseteq [7] = \{1, \ldots, 7\}$, remember that $F_{ij}^{ab} \in Z_2[\{x_1, \ldots, x_7\} \backslash \{x_i, x_j\}]$ is the function obtained by setting $(x_i, x_j) = (a, b)$ in F. Similarly, we can define G_{ij}^{ab} with respect to $P_1 = P_2 = P_3 = P_4 = 0$ (any assignment of $(x_i, x_j) = (a, b)$ is consistent with $P_1 = P_2 = P_3 = P_4 = 0$ which defines T). We make the following claim about F_{ij}^{ab}.

Claim 3. *For all $\{i, j\} \subseteq [7]$, $F_{ij}^{00} + F_{ij}^{11} \equiv 0$ or 1, and also $F_{ij}^{01} + F_{ij}^{10} \equiv 0$ or 1.*

We first show how this claim will let us realize g_8 or g_8', and lead to #P-hardness. Then, we give a proof of Claim 3. By Claim 3 and Lemma 7.8, the degree $d(F) \leqslant 2$.

- If $d(F) \leqslant 1$, then F is an affine linear combination of variables $x_1, \ldots, x_7$, i.e., $F = \lambda_0 + \sum_{i=1}^{7} \lambda_i x_i$ where $\lambda_i \in \mathbb{Z}_2$ for $0 \leqslant i \leqslant 7$. Notice that if we negate the variable x_i of f, we will get a signature $f'(x_1, \ldots, x_8) = (-1)^{x_i} f(x_1, \ldots, x_8)$. For every x_i appearing in F (i.e., $\lambda_i = 1$), we negate the variable x_i of f. Also, if $\lambda_0 = 1$, then we normalize f by a scalar -1. Then, we get a signature
$$f' = \chi_S \cdot 1 + 4\chi_T (-1)^{G'(x_1,\ldots,x_8)}.$$
This will not change the support of f and also norms of entries of f. Thus, $f'(\alpha) = \pm 3$ or ± 1 for all $\alpha \in \mathscr{S}(f') = \mathscr{E}_8$. Then, for every $\alpha \in T$, $f'(\alpha) = 1 + 4(-1)^{G'(\alpha)} = \pm 3$, which implies that $(-1)^{G'(\alpha)} = -1$ and $f'(\alpha) = -3$, because $1 + 4 = 5$ cannot be an entry of f'. Therefore, $f' = \chi_S - 4\chi_T = g_8$. Thus, g_8 is realizable from f.
By merging variables x_1 and x_5 of g_8 using $=_2$, we can get a 6-ary signature h. We rename variables x_2, x_3, x_4 to x_1, x_2, x_3 and variables x_6, x_7, x_8 to x_4, x_5, x_6 (The choice of merging x_1 and x_5 is just for a simple renaming of variables). Then after normalization by a scalar $1/2$, h has the following signature matrix
$$M_{123,456}(h) = A = \begin{bmatrix} -1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & -1 & 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & -1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & -1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & -1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & -1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & -1 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 & -1 \end{bmatrix}.$$
Consider the inner product $\langle \mathbf{h}_{14}^{00}, \mathbf{h}_{14}^{11} \rangle$. One can check that
$$\langle \mathbf{h}_{14}^{00}, \mathbf{h}_{14}^{11} \rangle = \sum_{1 \leqslant i,j \leqslant 4} A_{i,j} \cdot A_{i+4,j+4} = 8 \neq 0.$$
(This is the sum of pairwise products of every entry in the upper left 4×4 submatrix of A with the corresponding entry of the lower right 4×4 submatrix of A.) In fact, notice that $h(\overline{\alpha}) = \overline{h(\alpha)} = h(\alpha)$. By considering the representative matrix $M_r(h)$ of h (see Table 4), we have
$$M_r(h) = \begin{bmatrix} -1 & 1 & 1 & 1 \\ 1 & -1 & 1 & 1 \\ 1 & 1 & -1 & 1 \\ 1 & 1 & 1 & -1 \end{bmatrix}.$$
Then,
$$\langle \mathbf{h}_{14}^{00}, \mathbf{h}_{14}^{11} \rangle = 2(\mathrm{perm}(M_r(h)_{[1,2]}) + \mathrm{perm}(M_r(h)_{[3,4]})) = 2(2+2) = 8 \neq 0.$$
Also, since $\mathscr{S}(h) = \mathscr{E}_6$, it is easy to see that h is irreducible. Since h does not satisfy 2ND-ORTH, we get #P-hardness.
- If $d(F) = 2$, then by Lemma 7.8, for all $\{i, j\} \subseteq [7]$, $x_i x_j$ appears in F. Then, $F = \sum_{1 \leqslant i < j \leqslant 7} x_i x_j + L$ where L is an affine linear combination of variables $x_1, \ldots, x_7$. Since on the support $\mathscr{S}(f) = \mathscr{E}_8$, $x_1 + \cdots + x_8 = 0$, and on Boolean inputs $x_8^2 = x_8$, we can substitute F by $F' = F + x_8(x_1 + \cdots + x_8) - (x_8^2 - x_8) = \sum_{1 \leqslant i < j \leqslant 8} x_i x_j + L + x_8$ (all arithmetic

mod 2). This will not change the signature f. Then, by negating variables of f that appear as linear terms in F' and normalization with a scalar ± 1, we get a signature

$$f' = \chi_S(-1)^{\sum_{1\leqslant i<j\leqslant 8} x_i x_j} + 4\chi_T(-1)^{G'(x_1,\ldots,x_8)} = q_8 + 4\chi_T(-1)^{G'(x_1,\ldots,x_8)}.$$

where $q_8 = \chi_S(-1)^{\sum_{1\leqslant i<j\leqslant 8} x_i x_j}$ (see form (7.4)). For every $\alpha \in T$, since $\mathrm{wt}(\alpha) = 0, 4$ or 8, it is easy to see that $q_8(\alpha) = (-1)^{\binom{\mathrm{wt}(\alpha)}{2}} = 1$. Thus, $(-1)^{G'(\alpha)}$ must be -1 in order to get $1 - 4 = -3$, of norm 3 for f'. The other choice would give $1 + 4 = 5$ to be an entry of f', a contradiction. Therefore, $f'(\alpha) = q_8 - 4\chi_T = g'_8$. Thus, g'_8 is realizable from f.
By merging variables x_1 and x_5 of g'_8 using $=_2^-$, we can get a 6-ary signature h'. After renaming variables (same as we did for h) and normalization by a scalar $-1/2$, we have

$$M_{123,456}(h') = B = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & -1 & 1 & 0 & 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 & 1 & 0 & 0 & -1 \\ 1 & 0 & 0 & 1 & 0 & -1 & -1 & 0 \\ 0 & 1 & 1 & 0 & -1 & 0 & 0 & -1 \\ 1 & 0 & 0 & -1 & 0 & 1 & -1 & 0 \\ 1 & 0 & 0 & -1 & 0 & -1 & 1 & 0 \\ 0 & -1 & -1 & 0 & -1 & 0 & 0 & -1 \end{bmatrix}.$$

Consider the inner product $\langle \mathbf{h}'^{00}_{14}, \mathbf{h}'^{11}_{14} \rangle$. One can check that

$$\langle \mathbf{h}'^{00}_{14}, \mathbf{h}'^{11}_{14} \rangle = \sum_{1\leqslant i,j\leqslant 4} B_{i,j} \cdot B_{i+4,j+4} = -8 \neq 0.$$

Also, since $\mathscr{S}(h') = \mathscr{E}_6$, it is easy to see that h' is irreducible. Since h' does not satisfy 2ND-ORTH, we get #P-hardness.

This completes the proof of Step 2, and the proof of the lemma, modulo Claim 3.

Now, we prove Claim 3 that for all $\{i, j\} \subseteq [7]$, $F^{00}_{ij} + F^{11}_{ij} \equiv 0$ or 1 and $F^{01}_{ij} + F^{10}_{ij} \equiv 0$ or 1. For simplicity of notation, we prove this for $\{i, j\} = \{1, 2\}$. The proof for arbitrary $\{i, j\}$ is the same by replacing $\{1, 2\}$ by $\{i, j\}$. Since $f \in \int_{\mathcal{B}} \mathscr{A}$, $\widetilde{f}^{00}_{12}, \widetilde{f}^{01}_{12}, \widetilde{f}^{10}_{12}, \widetilde{f}^{11}_{12} \in \mathscr{A}$. Remember all nonzero entries in $\widetilde{f}^{ab}_{12}$ have the same norm, denoted by n_{ab}. We first show that between $\widetilde{f}^{00}_{12}$ and $\widetilde{f}^{11}_{12}$, exactly one has support $\mathscr{E}_{2n-2}$ and its nonzero entries have norm 2 and the other has nonzero entries of norm 4, and between $\widetilde{f}^{01}_{12}$ and $\widetilde{f}^{10}_{12}$, exactly one has support $\mathscr{O}_{2n-2}$ and its nonzero entries have norm 2 and the other has nonzero entries of norm 4. (This is not what we have proved in Step 1 where $\{1, 2\}$ is a pair of particularly chosen indices. Here $\{1, 2\}$ means an arbitrary pair $\{i, j\}$.)

Consider $f^{00}_{12}(\vec{0}^6)$ and $f^{11}_{12}(\vec{0}^6)$. By Step 1 of Case 2 and Lemma 7.4, we may assume that $\mathscr{S}_3(f) = \mathscr{S}(f_8)$ (after flipping and renaming variables). We have $00\vec{0}^6 \in \mathscr{S}_3(f)$ and $11\vec{0}^6 \notin \mathscr{S}_3(f)$. Thus, $f^{00}_{12}(\vec{0}^6) = \pm 3$ and $f^{11}_{12}(\vec{0}^6) = \pm 1$. (This is true when replacing $\{1, 2\}$ by an arbitrary pair of indices $\{i, j\}$.) Thus, between

$$\widetilde{f}^{00}_{12}(\vec{0}^6) = f^{00}_{12}(\vec{0}^6) + f^{11}_{12}(\vec{0}^6) \quad \text{and} \quad \widetilde{f}^{11}_{12}(\vec{0}^6) = f^{00}_{12}(\vec{0}^6) - f^{11}_{12}(\vec{0}^6),$$

one has norm 2 and the other has norm 4. They are both nonzero. Then, between n_{00} and n_{11}, one is 2 and the other is 4. By Lemma 7.6(2), between $\widetilde{f}^{00}_{12}$ and $\widetilde{f}^{11}_{12}$, the one whose nonzero entries

have norm 2 has support $\mathscr{E}_6$, and moreover n_{01} and $n_{10} = 2$ or 4. Since there exists $(a,b) = (0,0)$ or $(1,1)$ such that

$$|\widetilde{f}_{12}^{ab}|^2 = n_{ab}^2 \cdot |\mathscr{S}(\widetilde{f}_{12}^{ab})| = 2^2 \cdot |\mathscr{E}_6|,$$

for $\widetilde{f}_{12}^{cd}$ where $(c,d) = (0,1)$ or $(1,0)$, if $n_{cd} = 2$, then $|\mathscr{S}(\widetilde{f}_{12}^{cd})| = |\mathscr{E}_6| = |\mathscr{O}_6|$. Since $\widetilde{f}_{12}^{cd}$ has odd parity, $\mathscr{S}(\widetilde{f}_{12}^{cd}) \subseteq \mathscr{O}_6$. Thus, $|\mathscr{S}(\widetilde{f}_{12}^{cd})| = 2^{2n-3}$ implies that $\mathscr{S}(\widetilde{f}_{12}^{cd}) = \mathscr{O}_6$.

- If $n_{01} = n_{10} = 2$, then $\mathscr{S}(\widetilde{f}_{12}^{01}) = \mathscr{S}(\widetilde{f}_{12}^{10}) = \mathscr{O}_6$. For an arbitrary $\theta \in \mathscr{O}_6$,

$$f(01\theta) = \frac{\widetilde{f}(01\theta) + \widetilde{f}(10\theta)}{2} = \frac{(\pm 2) + (\pm 2)}{2} \quad \text{and} \quad f(10\theta) = \frac{\widetilde{f}(01\theta) - \widetilde{f}(10\theta)}{2} = \frac{(\pm 2) - (\pm 2)}{2}.$$

  Thus, between $f(01\theta)$ and $f(10\theta)$, exactly one has norm 2 and the other has norm 0. This gives a contradiction since every nonzero entry of f has norm 1 or 3.
- If $n_{01} = n_{10} = 4$, then still consider $f(01\theta)$ and $f(10\theta)$ for an arbitrary $\theta \in \mathscr{O}_6$. We know that $f(01\theta), f(10\theta) = \pm 4, \pm 2$ or 0. The case that $f(01\theta) = 0$ or $f(10\theta) = 0$ cannot occur since $\mathscr{S}(f) = \mathscr{E}_{2n}$ and clearly, $01\theta, 10\theta \in \mathscr{E}_{2n}$. Thus, $f(01\theta), f(10\theta) = \pm 4, \pm 2$. Still, we get a contradiction since every nonzero entry of f has norm 1 or 3.
- Thus, between n_{01} and n_{10}, one is 2 and the other is 4.

Then, between $\widetilde{f}_{12}^{01}$ and $\widetilde{f}_{12}^{10}$, exactly one has support $\mathscr{O}_6$ and its nonzero entries have norm 2, and the other has nonzero entries of norm 4.

Now, we show that $F_{12}^{00} + F_{12}^{11} \equiv 0$ or 1. We first consider the case that between $\widetilde{f}_{12}^{00}$ and $\widetilde{f}_{12}^{11}$, $\widetilde{f}_{12}^{11} = f_{12}^{00} - f_{12}^{11}$ is the signature whose support is $\mathscr{E}_6$ and nonzero entries have norm 2; the case where it is $\widetilde{f}_{12}^{00}$ will be addressed shortly. Let S_0 be the subspace in $\mathbb{Z}_2^6$ obtained by setting $x_1 = x_2 = 0$ in $S = \mathscr{S}(f) = \mathscr{E}_8$, and S_1 be the subspace in $\mathbb{Z}_2^6$ obtained by setting $x_1 = x_2 = 1$. Similarly, we can define T_0 and T_1, replacing S in the definition by $T = \mathscr{S}_3(f) = I_8$. Clearly, $S_0 = S_1 = \{(x_3, \ldots, x_8) \in \mathbb{Z}_2^6 \mid x_3 + \cdots x_8 = 0\} = \mathscr{E}_6$. Also, one can check that T_0 is disjoint with T_1. Then

$$f_{12}^{00} = \chi_{S_0}(-1)^{F_{12}^{00}(x_3,\ldots,x_7)} + 4\chi_{T_0}(-1)^{G_{12}^{00}(x_3,\ldots,x_8)},$$

and

$$f_{12}^{11} = \chi_{S_1}(-1)^{F_{12}^{11}(x_3,\ldots,x_7)} + 4\chi_{T_1}(-1)^{G_{12}^{11}(x_3,\ldots,x_8)}.$$

Thus,

$$\widetilde{f}_{12}^{11} = \chi_{\mathscr{E}_6}((-1)^{F_{12}^{00}(x_3,\ldots,x_7)} - (-1)^{F_{12}^{11}(x_3,\ldots,x_7)}) + 4\chi_{T_0}(-1)^{G_{12}^{00}(x_3,\ldots,x_8)} - 4\chi_{T_1}(-1)^{G_{12}^{11}(x_3,\ldots,x_8)}.$$

Since $\mathscr{S}(\widetilde{f}_{12}^{11}) = \mathscr{E}_6$ and $n_{11} = 2$, $\widetilde{f}_{12}^{11}(\theta) = \pm 2$ for every $\theta \in \mathscr{E}_6$. If $\theta \notin T_0 \cup T_1$, then

$$\widetilde{f}_{12}^{11}(\theta) = (-1)^{F_{12}^{00}(\theta)} - (-1)^{F_{12}^{11}(\theta)} = \pm 2.$$

If $\theta \in T_0 \cup T_1$, then it belongs to exactly one of T_0 or T_1,

$$\widetilde{f}_{12}^{11}(\theta) = (-1)^{F_{12}^{00}(\theta)} - (-1)^{F_{12}^{11}(\theta)} + 4a = \pm 2,$$

where $a = \pm 1$. In this case, the sum of the first two terms is still $(-1)^{F_{12}^{00}(\theta)} - (-1)^{F_{12}^{11}(\theta)} = \pm 2$, because the only other possible value for $(\pm 1) - (\pm 1)$ is 0 and then we would have $4a = \pm 2$, a contradiction. Thus, for every $(x_3, \ldots, x_7) \in \mathbb{Z}_2^5$ which decides every $(x_3, \ldots, x_8) \in \mathscr{E}_6$ by $x_8 = x_3 + \cdots + x_7$,

$$(-1)^{F_{12}^{00}(x_3,\ldots,x_7)} - (-1)^{F_{12}^{11}(x_3,\ldots,x_7)} = \pm 2.$$

This implies that

$$(-1)^{F_{12}^{00}(x_3,\ldots,x_7)} = -(-1)^{F_{12}^{11}(x_3,\ldots,x_7)}.$$

Thus,

$$(-1)^{F_{12}^{00}(x_3,\ldots,x_7)+F_{12}^{11}(x_3,\ldots,x_7)} = -1.$$

Then, $F_{12}^{00} + F_{12}^{11} \equiv 1$.

Now we address the case that (between $\widetilde{f}_{12}^{00}$ and $\widetilde{f}_{12}^{11}$) it is $\widetilde{f}_{12}^{00} = f_{12}^{00} + f_{12}^{11}$ the signature whose support is $\mathscr{E}_6$ and nonzero entries have norm 2. Then similarly for every $(x_3, \ldots, x_7) \in \mathbb{Z}_2^5$, which determines every $(x_3, \ldots, x_8) \in \mathscr{E}_6$,

$$(-1)^{F_{12}^{00}(x_3,\ldots,x_7)} + (-1)^{F_{12}^{11}(x_3,\ldots,x_7)} = \pm 2.$$

This implies that

$$(-1)^{F_{12}^{00}(x_3,\ldots,x_7)} = (-1)^{F_{12}^{11}(x_3,\ldots,x_7)}.$$

Thus,

$$(-1)^{F_{12}^{00}(x_3,\ldots,x_7)+F_{12}^{11}(x_3,\ldots,x_7)} = 1$$

Then, $F_{12}^{00} + F_{12}^{11} \equiv 0$.

We have proved that, $F_{12}^{00} + F_{12}^{11} \equiv 0$ or 1.

Also, consider $\widetilde{f}_{12}^{01}$ and $\widetilde{f}_{12}^{10}$. One of them is a signature whose support is $\mathscr{O}_{2n-2}$ and nonzero entries have norm 2. Then similarly, for every $(x_3, \ldots, x_7) \in \mathbb{Z}^5$ which decides every $(x_3, \ldots, x_8) \in \mathscr{O}_6$ by $x_8 = 1 + x_3 + \cdots + x_7$,

$$(-1)^{F_{12}^{01}(x_3,\ldots,x_7)} + (-1)^{F_{12}^{10}(x_3,\ldots,x_7)} = \pm 2,$$

or

$$(-1)^{F_{12}^{01}(x_3,\ldots,x_7)} - (-1)^{F_{12}^{10}(x_3,\ldots,x_7)} = \pm 2.$$

Then, $F_{12}^{01} + F_{12}^{10} \equiv 0$ or $F_{12}^{01} + F_{12}^{10} \equiv 1$. The above proof holds for all $\{i, j\} \subseteq [7]$. Thus, for all $\{i, j\} \subseteq [7]$, $F_{ij}^{00} + F_{ij}^{11} \equiv 0$ or 1, and $F_{ij}^{01} + F_{ij}^{10} \equiv 0$ or 1. □

Remark: The above proof does not require $\mathcal{F}$ to be non-$\mathcal{B}$ hard.

7.3 Support condition

Then, by further assuming that nonzero entries of f have the same norm, we show that f has affine support or we can get the #P-hardness for non-$\mathcal{B}$ hard set $\mathcal{F}$ (Lemma 7.16). Here, we do require $\mathcal{F}$ to be non-$\mathcal{B}$ hard.

We first give one more result about $\widetilde{f}$. Remember that if $f \in \int_{\mathcal{B}} \mathscr{A}$, then $\widetilde{f}_{12}^{00}$, $\widetilde{f}_{12}^{01}$, $\widetilde{f}_{12}^{10}$, $\widetilde{f}_{12}^{11} \in \mathscr{A}$, and n_{ab} denotes the norm of nonzero entries of $\widetilde{f}_{12}^{ab}$. Let $\widetilde{\mathcal{B}} = \left\{\widetilde{=_2^+}, \widetilde{=_2^-}, \widetilde{\neq_2^+}, \widetilde{\neq_2^-}\right\}$ where $\widetilde{=_2^+} = (2,0,0,0)$, $\widetilde{=_2^-} = (0,0,0,2)$, $\widetilde{\neq_2^+} = (0,2,0,0)$ and $\widetilde{\neq_2^-} = (0,0,2,0)$. Signatures in $\widetilde{\mathcal{B}}$ are obtained by performing the H_4 gadget construction on binary signatures in $\mathcal{B}$.

Lemma 7.10. *Let f be an irreducible signature of arity $2n \geqslant 6$ with the following properties.*

1. *f has even parity, f satisfies* 2ND-ORTH, *and $f \in \int_{\mathcal{B}} \mathscr{A}$;*

2. *for all* $\{i,j\}$ *disjoint with* $\{1,2\}$ *and every* $b \in \mathcal{B}$, *either* $M(\mathfrak{m}_{12}(\partial^b_{ij} f)) = \lambda^b_{ij} I_4$ *for some real* $\lambda^b_{ij} \neq 0$, *or there exists a nonzero binary signature* $g^b_{ij} \in \mathcal{B}$ *such that* $g^b_{ij}(x_1, x_2) \mid \partial^b_{ij} f$.

If $\mathscr{S}(\widetilde{f}^{01}_{12}) = \mathscr{S}(\widetilde{f}^{10}_{12})$, $n_{00} > n_{01} > 0$, *then* $\mathscr{S}(\widetilde{f}^{01}_{12}) = \mathscr{O}_{2n-2}$.

Proof. We first analyze the second property of f, i.e., the property about $\partial^b_{ij} f$.

- If $M(\mathfrak{m}_{12}(\partial^b_{ij} f)) = \lambda^b_{ij} I_4$, by Lemma 7.5, then $M(\mathfrak{m}_{12}(\widetilde{\partial^b_{ij} f})) = 2\lambda^b_{ij} I_4$. Since $\{i,j\}$ is disjoint with $\{1,2\}$, the H_4 gadget on variables x_1 and x_2 commutes with the merging gadget on variables x_i and x_j. Thus, $\widetilde{\partial^b_{ij} f} = \partial^b_{ij} \widetilde{f}$. Let $(\partial^b_{ij} \widetilde{f})^{ab}_{12}$ be the signature obtained by setting variables x_1 and x_2 of $\partial^b_{ij} \widetilde{f}$ to a and b, and $\partial^b_{ij}(\widetilde{f}^{ab}_{12})$ be the signature obtained by merging variables x_i and x_j of $\widetilde{f}^{ab}_{12}$. Again, since $\{1,2\}$ and $\{i,j\}$ are disjoint, $(\partial^b_{ij} \widetilde{f})^{ab}_{12} = \partial^b_{ij}(\widetilde{f}^{ab}_{12})$. We denote them by $\partial^b_{ij} \widetilde{f}^{ab}_{12}$. Then, since $M(\mathfrak{m}_{12}(\widetilde{\partial^b_{ij} f})) = M(\mathfrak{m}_{12}(\partial^b_{ij} \widetilde{f})) = 2\lambda^b_{ij} I_4$,

$$|\partial^b_{ij} \widetilde{\mathbf{f}}^{00}_{12}|^2 = |\partial^b_{ij} \widetilde{\mathbf{f}}^{01}_{12}|^2 = |\partial^b_{ij} \widetilde{\mathbf{f}}^{10}_{12}|^2 = |\partial^b_{ij} \widetilde{\mathbf{f}}^{11}_{12}|^2 = 2\lambda^b_{ij} \neq 0.$$

- If $g^b_{ij}(x_1, x_2) \mid \partial^b_{ij} f$, i.e, $\partial^b_{ij} f = g^b_{ij}(x_1, x_2) \otimes h$, then $\widetilde{\partial^b_{ij} f} = \partial^b_{ij} \widetilde{f} = \widetilde{g^b_{ij}}(x_1, x_2) \otimes h$. Since $g^b_{ij} \in \mathcal{B}$, $\widetilde{g^b_{ij}} \in \widetilde{\mathcal{B}}$. By the form of signatures in $\widetilde{\mathcal{B}}$, among $\partial^b_{ij} \widetilde{f}^{00}_{12}$, $\partial^b_{ij} \widetilde{f}^{01}_{12}$, $\partial^b_{ij} \widetilde{f}^{10}_{12}$ and $\partial^b_{ij} \widetilde{f}^{11}_{12}$, at most one is a nonzero signature.

Combining the above two cases we have that, among $\partial^b_{ij} \widetilde{f}^{00}_{12}$, $\partial^b_{ij} \widetilde{f}^{01}_{12}$, $\partial^b_{ij} \widetilde{f}^{10}_{12}$ and $\partial^b_{ij} \widetilde{f}^{11}_{12}$, if at least two of them are nonzero signatures then they are all nonzero signatures.

Now, we show that $\mathscr{S}(\widetilde{f}^{01}_{12}) = \mathscr{O}_{2n-2}$. Since f has even parity, $\widetilde{f}$ also has even parity. Then, $\widetilde{f}^{01}_{12}$ has odd parity, i.e., $\mathscr{S}(\widetilde{f}^{01}_{12}) \subseteq \mathscr{O}_{2n-2}$. For a contradiction, suppose that $\mathscr{S}(\widetilde{f}^{01}_{12}) \subsetneq \mathscr{O}_{2n-2}$. Since $n_{01} > 0$, $\mathscr{S}(\widetilde{f}^{01}_{12}) \neq \emptyset$. Then, we can pick a pair of inputs $\alpha, \beta \in \mathscr{O}_{2n-2}$ with $\mathrm{wt}(\alpha \oplus \beta) = 2$ such that $\alpha \in \mathscr{S}(\widetilde{f}^{01}_{12})$ and $\beta \notin \mathscr{S}(\widetilde{f}^{01}_{12})$. Also, since $\mathscr{S}(\widetilde{f}^{01}_{12}) = \mathscr{S}(\widetilde{f}^{10}_{12})$, $\alpha \in \mathscr{S}(\widetilde{f}^{10}_{12})$ and $\beta \notin \mathscr{S}(\widetilde{f}^{10}_{12})$. Thus, $|\widetilde{f}^{01}_{12}(\alpha)| = n_{01}$ and $|\widetilde{f}^{01}_{12}(\beta)| = 0$, and $|\widetilde{f}^{10}_{12}(\alpha)| = n_{10}$ and $|\widetilde{f}^{10}_{12}(\beta)| = 0$. Suppose that α and β differ in bits i and j. Clearly, $\{i,j\}$ is disjoint with $\{1,2\}$. Depending whether $\alpha_i = \alpha_j$ or $\alpha_i \neq \alpha_j$, we connect variables x_i and x_j of $\widetilde{f}$ using $=^+_2$ or $\neq^+_2$. We get signatures $\partial^+_{ij} \widetilde{f}$ or $\partial^{\widehat{+}}_{ij} \widetilde{f}$ respectively. We consider the case that $\alpha_i = \alpha_j$; in this case $\{\alpha_i\alpha_j, \beta_i\beta_j\} = \{00, 11\}$. For the case that $\alpha_i \neq \alpha_j$, the analysis is the same by replacing $\partial^+_{ij} \widetilde{f}$ with $\partial^{\widehat{+}}_{ij} \widetilde{f}$.

Consider $\partial^+_{ij} \widetilde{f}$. Then, because $\{\alpha_i\alpha_j, \beta_i\beta_j\} = \{00, 11\}$, $\widetilde{f}^{01}_{12}(\alpha) + \widetilde{f}^{01}_{12}(\beta)$ and $\widetilde{f}^{10}_{12}(\alpha) + \widetilde{f}^{10}_{12}(\beta)$ are entries of $\partial^+_{ij} \widetilde{f}$; more precisely, they are entries of $\partial^+_{ij} \widetilde{f}^{01}_{12}$ and $\partial^+_{ij} \widetilde{f}^{10}_{12}$ respectively. Since $\widetilde{f}^{01}_{12}(\beta) = \widetilde{f}^{10}_{12}(\beta) = 0$, we have

$$|\widetilde{f}^{01}_{12}(\alpha) + \widetilde{f}^{01}_{12}(\beta)| = |\widetilde{f}^{01}_{12}(\alpha)| = n_{01} \neq 0, \quad \text{and} \quad |\widetilde{f}^{10}_{12}(\alpha) + \widetilde{f}^{10}_{12}(\beta)| = |\widetilde{f}^{10}_{12}(\alpha)| = n_{10} \neq 0.$$

Thus, $\partial^+_{ij} \widetilde{f}^{01}_{12}$ has a nonzero entry with norm n_{01}, and then $\partial^+_{ij} \widetilde{f}^{01}_{12} \not\equiv 0$. Also, we have $\partial^+_{ij} \widetilde{f}^{10}_{12} \not\equiv 0$. Thus at least two among $\partial^+_{ij} \widetilde{f}^{00}_{12}$, $\partial^+_{ij} \widetilde{f}^{01}_{12}$, $\partial^+_{ij} \widetilde{f}^{10}_{12}$ and $\partial^+_{ij} \widetilde{f}^{10}_{12}$ are nonzero, it follows that all of them are nonzero signatures.

Then $\partial^+_{ij} \widetilde{f}^{00}_{12} \not\equiv 0$. Let $\partial^+_{ij} \widetilde{f}^{00}_{12}(\gamma)$ be a nonzero entry of $\partial^+_{ij} \widetilde{f}^{00}_{12}$. Then, $\partial^+_{ij} \widetilde{f}^{00}_{12}(\gamma) = \widetilde{f}^{0000}_{12ij}(\gamma) + \widetilde{f}^{0011}_{12ij}(\gamma) \neq 0$.[1] Clearly, $\widetilde{f}^{0000}_{12ij}(\gamma)$ and $\widetilde{f}^{0011}_{12ij}(\gamma)$ are entries of $\widetilde{f}^{00}_{12}$, and they have norm n_{00} or 0. Thus, $\partial^+_{ij} \widetilde{f}^{00}_{12}(\gamma)$ has norm $2n_{00}$ or n_{00}. Also, $\partial^+_{ij} \widetilde{f}^{00}_{12}(\gamma)$ is an entry of $\partial^+_{ij} \widetilde{f}$ on the input 00γ. Thus,

[1] For the case that $\alpha_i \neq \alpha_j$, $\partial^+_{ij} \widetilde{f}^{00}_{12}(\gamma) = \widetilde{f}^{0000}_{12ij}(\gamma) + \widetilde{f}^{0011}_{12ij}(\gamma)$ will be replced by $\partial^{\widehat{+}}_{ij} \widetilde{f}^{00}_{12}(\gamma) = \widetilde{f}^{0001}_{12ij}(\gamma) + \widetilde{f}^{0010}_{12ij}(\gamma)$.

$\partial_{ij}^{+}\widetilde{f}$ has a nonzero entry with norm $2n_{00}$ or n_{00}. Since $n_{00} > n_{01}$, both $2n_{00}$ and n_{00} are not equal to n_{01}. Thus, $\partial_{ij}^{+}\widetilde{f}$ has two nonzero entries with different norms. Such a signature is not in $\mathscr{A}$. However, since $f \in \int_{\mathcal{B}} \mathscr{A}$, by Lemma 7.5, $\partial_{ij}^{+}\widetilde{f} \in \mathscr{A}$. Contradiction. Thus, $\mathscr{S}(\widetilde{f}_{12}^{01}) = \mathscr{O}_{2n-2}$. □

We also give a result about the edge partition of complete graphs into two complete tripartite subgraphs. This result should also be of independent interest. We say a graph $G = (V, E)$ is tripartite if $V = V_1 \sqcup V_2 \sqcup V_3$ and every $e \in E$ is between distinct V_i and V_j. Here $\sqcup$ denotes disjoint union. The parts V_i are allowed to be empty. It is a complete tripartite graph if every pair between distinct V_i and V_j is an edge.

Definition 7.11. *Let K_n be the complete graph on n vertices. We say K_n has a* tripartite 2-partition *if there exist complete tripartite subgraphs T_1 and T_2 such that $\{E(T_1), E(T_2)\}$ is a partition of $E(K_n)$, i.e., $E(K_n) = E(T_1) \sqcup E(T_2)$. We say T_1 and T_2 are witnesses of a tripartite 2-partition of K_n.*

Lemma 7.12. *K_n has a tripartite 2-partition iff $n \leqslant 5$. For $n = 5$, up to an automorphism of K_5, there is a unique tripartite 2-partition where T_1 is a triangle on $\{v_1, v_2, v_3\}$ and T_2 is the complete tripartite graph with parts $\{v_1, v_2, v_3\}$, $\{v_4\}$ and $\{v_5\}$.*

Proof. Let T be a complete tripartite graph. Let $G_{2,1}$ be the union of K_2 and an isolated vertex. We first prove the following two claims.

Claim 1. *$G_{2,1}$ is not an induced subgraph of T.*

For a contradiction, suppose $G_{2,1} = (V, E)$ is an induced subgraph of T, where $V = \{v_1, v_2, v_3\}$, and $E = \{(v_1, v_2)\}$. Then, v_1 and v_2 belong to distinct parts of T. Since $(v_1, v_3), (v_2, v_3) \notin E(T)$, v_1 and v_3 belong to the same part of T, and so are v_2 and v_3. Thus, v_1 and v_2 belong to the same part of T. This contradiction proves Claim 1.

Claim 2. *K_4 is not an induced subgraph of T.*

For a contradiction, suppose K_4 on $V = \{v_1, v_2, v_3, v_4\}$ is an induced subgraph of T. Then, for any two distinct vertices $v_i, v_j \in V$, the edge $(v_i, v_j) \in K_4$ shows that v_i and v_j belong to distinct parts in T. But T has at most three distinct nonempty parts. This contradiction proves Claim 2.

Now, we prove this lemma. The cases $n = 1, 2, 3$ are trivial. When $n = 4$, we have the following two tripartite 2-partitions of K_4, with $V(T_1) = \{v_1\} \sqcup \{v_2\} \sqcup \{v_3\}$ and $V(T_2) = \{v_1, v_2, v_3\} \sqcup \{v_4\} \sqcup \emptyset$, or alternatively with $V(T_1') = \{v_1\} \sqcup \{v_2\} \sqcup \emptyset$ and $V(T_2') = \{v_1, v_2\} \sqcup \{v_3\} \sqcup \{x_4\}$.

We consider $n \geqslant 5$. Suppose K_n has a tripartite 2-partition with complete tripartite subgraphs $T_1 = (V_1, E_1)$ and $T_2 = (V_2, E_2)$. We write (A_i, B_i, C_i) for the three parts of T_i, $i = 1, 2$.

Clearly $V = V_1 \cup V_2$, as all vertices of V must appear in either T_1 or T_2, for otherwise any edge incident to $v \in V \setminus (V_1 \cup V_2)$ is not in $E_1 \cup E_2$. If all parts of both T_1 and T_2 have size at most 1, then $|E_1 \sqcup E_2| \leqslant 6 < |K_5| \leqslant |K_n|$, a contradiction. So at least one part, say A_1, has size $|A_1| \geqslant 2$, and we let $a, a' \in A_1$. Then, $(a, a') \notin E_1$. Thus, $(a, a') \in E_2$ and $a, a' \in V_2$.

We show that $(V_1 \setminus A_1) \cap (V_2 \setminus A_1) = \emptyset$. Otherwise, there exists $v \in (V_1 \setminus A_1) \cap (V_2 \setminus A_1)$. Then, edges $(v, a), (v, a') \in E_1$. Thus, among edges $(v, a), (v, a')$ and (a, a') of K_n, (a, a') is the only one in T_2. Since $v, a, a' \in V_2$, $G_{2,1}$ is an induced subgraph of T_2. A violation of Claim 1.

If both $V_1 \setminus A_1$ and $V_2 \setminus A_1$ are nonempty, then an edge in K_n between $u \in V_1 \setminus A_1$ and $v \in V_2 \setminus A_1$ is in neither E_1 nor E_2, since $u \notin V_2$ and $v \notin V_1$. This is a contradiction. If $V_1 \setminus A_1 = \emptyset$, then

$E_1 = \emptyset$, and then all edges of K_n belong to T_2, which violates Claim 2. So $V_2 \setminus A_1 = \emptyset$. Since $V = V_1 \cup V_2$, $V_2 \setminus A_1 = \emptyset$ implies that $V = V_1$.

Clearly $V_1 \setminus A_1 = B_1 \sqcup C_1$. If $|B_1| \geqslant 2$, then there exists some $\{u, v\} \subseteq B_1 \subseteq V_1 \setminus A_1$, which is disjoint from V_2. Thus $\{u, v\} \notin E_1 \sqcup E_2$, a contradiction. Hence $|B_1| \leqslant 1$. Similarly $|C_1| \leqslant 1$. Finally, if $|A_1| \geqslant 4$, then there is a K_4 inside A_1 which must be an induced subgraph of T_2, a violation of Claim 2. Thus $|A_1| \leqslant 3$. It follows that $n \leqslant 5$ since $V = V_1 = A_1 \sqcup B_1 \sqcup C_1$. If $n = 5$, then $|A_1| = 3$ and $|B_1| = |C_1| = 1$. After relabeling vertices, we may assume that $A_1 = \{v_1, v_2, v_3\}$, $B_1 = \{v_4\}$ and $C_1 = \{v_5\}$. Then, we have $A_2 = \{v_1\}$, $B_2 = \{v_2\}$ and $C_2 = \{v_3\}$. This gives the unique tripartite 2-partition of K_5. □

We will apply Lemma 7.12 to multilinear $\mathbb{Z}_2$-polynomials. Remember that we take the reduction of polynomials in $\mathbb{Z}_2[x_1, \ldots, x_n]$ modulo the ideal generated by $\{x_i^2 - x_i \mid i \in [n]\}$ replacing any F by its unique multilinear representative.

Definition 7.13. *Let $F(x_1, \ldots, x_n) \in \mathbb{Z}_2[x_1, \ldots, x_n]$ be a complete quadratic polynomial. We say F has a twice-linear 2-partition if there exist $L_1, L_2, L_3, L_4 \in \mathbb{Z}_2[x_1, \ldots, x_n]$ where $d(L_1) = d(L_2) = d(L_3) = d(L_4) \leqslant 1$ such that $F = L_1 \cdot L_2 + L_3 \cdot L_4$.*

Lemma 7.12 gives the following result about multilinear $\mathbb{Z}_2$-polynomials.

Lemma 7.14. *Let $F(x_1, \ldots, x_n) \in \mathbb{Z}_2[x_1, \ldots, x_n]$ be a complete quadratic polynomial. For $n \geqslant 6$, F does not have a twice-linear 2-partition. For $n = 5$, F has a twice-linear 2-partition $F = L_1 \cdot L_2 + L_3 \cdot L_4$ iff (after renaming variables) the cross terms of $L_1 \cdot L_2$ and $L_3 \cdot L_4$ correspond to the unique tripartite 2-partition of K_5, and we have $L_1 \cdot L_2 = (x_1 + x_2 + a)(x_2 + x_3 + b)$ and $L_3 \cdot L_4 = (x_1 + x_2 + x_3 + x_4 + c)(x_1 + x_2 + x_3 + x_5 + d)$ for some $a, b, c, d \in \mathbb{Z}_2$.*

Proof. We first analyze the quadratic terms that appear in a product of two linear polynomials. We use $x_i \in L$ to denote that a linear term x_i appears in a linear polynomial L. Let L_1 and L_2 be two linear polynomials.

Let $U_1 = \{x_i \mid x_i \in L_1, x_i \notin L_2\}$, $U_2 = \{x_i \mid x_i \in L_1, x_i \in L_2\}$, and $U_3 = \{x_i \mid x_i \notin L_1, x_i \in L_2\}$. Then,

$$L_1 = \sum_{x_i \in U_1} x_i + \sum_{x_j \in U_2} x_j + a, \quad \text{and} \quad L_2 = \sum_{x_j \in U_2} x_j + \sum_{x_k \in U_3} x_k + b$$

for some $a, b \in \mathbb{Z}_2^2$. The quadratic terms in $L_1 \cdot L_2$ are from

$$\Big(\sum_{x_i \in U_1} x_i + \sum_{x_j \in U_2} x_j\Big) \cdot \Big(\sum_{x_j \in U_2} x_j + \sum_{x_k \in U_3} x_k\Big)$$

which are enumerated in

$$\sum_{x_i \in U_1, x_j \in U_2} x_i x_j + \sum_{x_i \in U_1, x_k \in U_3} x_i x_k + \sum_{x_j \in U_2, x_k \in U_3} x_j x_k.$$

Note that each term x_i^2 for $i \in U_2$ is replaced by x_i (thus no longer counted as a quadratic term) as we calculate modulo the ideal generated by $\{x_i^2 - x_i \mid i \in [n]\}$, and every pairwise cross product term $x_i x_j$ for $i, j \in U_2$ and $i \neq j$ disappears since it appears exactly twice.

If we view variables $x_1, \ldots, x_n$ as n vertices and each quadratic term $x_i x_j$ as an edge between vertices x_i and x_j, then the quadratic terms in $L_1 \cdot L_2$ are the edges of a complete tripartite subgraph

T of K_n (the parts of a tripartite graph could be empty) and $V(T) = U_1 \sqcup U_2 \sqcup U_3$. Therefore, $L_1 \cdot L_2$ is one of the two terms of a twice-linear 2-partition of a complete quadratic polynomial over n variables iff T is one tripartite complete graph in a tripartite 2-partition of the complete graph K_n. By Lemma 7.12, a tripartite 2-partition does not exist for K_n when $n \geqslant 6$. Thus, F does not have twice-linear partition when $n \geqslant 6$. When $n = 5$, the tripartite 2-partition of K_5 is unique up to relabeling. One tripartite complete graph of this tripartite 2-partition is a triangle, and we may assume it is on $\{x_1, x_2, x_3\}$. Then, we take $L_1 \cdot L_2 = (x_1 + x_2 + a)(x_2 + x_3 + b)$ for some $a, b \in \mathbb{Z}_2^2$, and $L_3 \cdot L_4 = (x_1 + x_2 + x_3 + x_4 + c)(x_1 + x_2 + x_3 + x_5 + d)$ for some $c, d \in \mathbb{Z}_2^2$. Thus, a complete quadratic polynomial $F(x_1, \ldots, x_5)$ over 5 variables has a twice-linear 2-partition iff (after renaming variables) $F = L_1 \cdot L_2 + L_3 \cdot L_4$. □

Now, we are ready to make a further major step towards Theorem 7.19. We first give a preliminary result.

Lemma 7.15. *Let f be a $2n$-ary signature, where $2n \geqslant 4$. If $f \in \int_{\mathcal{B}} \mathscr{A}$ and $|f(\alpha)| = 1$ for all $\alpha \in \mathscr{S}(f)$, then for all $\{i, j\} \subseteq [2n]$, $\mathscr{S}(f_{ij}^{00}) = \mathscr{S}(f_{ij}^{11})$ or $\mathscr{S}(f_{ij}^{00}) \cap \mathscr{S}(f_{ij}^{11}) = \emptyset$, and $\mathscr{S}(f_{ij}^{01}) = \mathscr{S}(f_{ij}^{10})$ or $\mathscr{S}(f_{ij}^{01}) \cap \mathscr{S}(f_{ij}^{10}) = \emptyset$.*

Proof. We first prove that for all $\{i, j\} \subseteq [2n]$, $\mathscr{S}(f_{ij}^{00}) = \mathscr{S}(f_{ij}^{11})$ or $\mathscr{S}(f_{ij}^{00}) \cap \mathscr{S}(f_{ij}^{11}) = \emptyset$. For a contradiction, suppose that there exist $\alpha, \beta \in \mathbb{Z}_2^{2n-2}$ such that $\alpha \in \mathscr{S}(f_{ij}^{00}) \cap \mathscr{S}(f_{ij}^{11})$ and $\beta \in \mathscr{S}(f_{ij}^{00}) \Delta \mathscr{S}(f_{ij}^{11})$, where Δ denotes the symmetric difference between two sets. Consider signatures $\partial_{ij}^{+} f$ and $\partial_{ij}^{-} f$. Then, $f_{ij}^{00}(\alpha) + f_{ij}^{11}(\alpha)$ and $f_{ij}^{00}(\beta) + f_{ij}^{11}(\beta)$ are entries of $\partial_{ij}^{+} f$, and $f_{ij}^{00}(\alpha) - f_{ij}^{11}(\alpha)$ and $f_{ij}^{00}(\beta) - f_{ij}^{11}(\beta)$ are entries of $\partial_{ij}^{-} f$. Since $\alpha \in \mathscr{S}(f_{ij}^{00}) \cap \mathscr{S}(f_{ij}^{11})$, $f_{ij}^{00}(\alpha) = \pm 1$ and $f_{ij}^{11}(\alpha) = \pm 1$. Then between $f_{ij}^{00}(\alpha) + f_{ij}^{11}(\alpha)$ and $f_{ij}^{00}(\alpha) - f_{ij}^{11}(\alpha)$, exactly one has norm 2 and the other is 0. However, since $\beta \in \mathscr{S}(f_{ij}^{00}) \Delta \mathscr{S}(f_{ij}^{11})$, between $f_{ij}^{00}(\beta)$ and $f_{ij}^{11}(\beta)$, exactly one is 0 and the other has norm 1. Thus, $|f_{ij}^{00}(\beta) + f_{ij}^{11}(\beta)| = |f_{ij}^{00}(\beta) - f_{ij}^{11}(\beta)| = 1$. Then, between $\partial_{ij}^{+} f$ and $\partial_{ij}^{-} f$, there is a signature that has an entry of norm 1 and an entry of norm 2. Clearly, such a signature is not in $\mathscr{A}$. However, since $f \in \int_{\mathcal{B}} \mathscr{A}$, we have $\partial_{ij}^{+} f, \partial_{ij}^{-} f \in \mathscr{A}$. Contradiction.

By considering signatures $\partial_{ij}^{\widehat{+}} f$ and $\partial_{ij}^{\widehat{-}} f$, similarly we can show that $\mathscr{S}(f_{ij}^{01}) = \mathscr{S}(f_{ij}^{10})$ or $\mathscr{S}(f_{ij}^{01}) \cap \mathscr{S}(f_{ij}^{10}) = \emptyset$. □

The next lemma is a major step.

Lemma 7.16. *Suppose that $\mathcal{F}$ is non-$\mathcal{B}$ hard. Let $f \in \mathcal{F}$ be an irreducible $2n$-ary ($2n \geqslant 8$) signature with parity. Then,*

- Holant$^b(\mathcal{F})$ *is #P-hard, or*
- *there is a signature $g \notin \mathscr{A}$ of arity $2k < 2n$ that is realizable from f and $\mathcal{B}$, or*
- *f has affine support.*

Proof. Again, we may assume that f satisfies 2ND-ORTH and $f \in \int_{\mathcal{B}} \mathscr{A}$. Also, by Lemma 7.9, we may assume that $f(\alpha) = \pm 1$ for all $\alpha \in \mathscr{S}(f)$ after normalization.

For any four distinct binary strings $\alpha, \beta, \gamma, \delta \in \mathbb{Z}_2^{2n}$ with $\alpha \oplus \beta \oplus \gamma = \delta$, we define a *score* $T(\alpha, \beta, \gamma, \delta) = (t_1, t_2, t_3)$ which are the values of $\mathrm{wt}(\alpha \oplus \beta) = \mathrm{wt}(\gamma \oplus \delta), \mathrm{wt}(\alpha \oplus \gamma) = \mathrm{wt}(\beta \oplus \delta)$ and $\mathrm{wt}(\alpha \oplus \delta) = \mathrm{wt}(\beta \oplus \gamma)$ ordered from the smallest to the largest. We order the scores lexicographically, i.e., we say $T = (t_1, t_2, t_3) < T' = (t_1', t_2', t_3')$ if $t_1 < t_1'$, or $t_2 < t_2'$ when $t_1 = t_1'$, or $t_3 < t_3'$ when $t_1 = t_1'$ and $t_2 = t_2'$. Note that since $\alpha, \beta, \gamma, \delta$ are distinct, the smallest value of the score T is

$(2,2,2)$. We say that $(\alpha,\beta,\gamma,\delta)$ where $\alpha\oplus\beta\oplus\gamma=\delta$ forms a *non-affine quadrilateral* of f if exactly three of them are in $\mathscr{S}(f)$ and the fourth is not.

For a contradiction, suppose that $\mathscr{S}(f)$ is not affine. Then, f has at least a non-affine quadrilateral. Among all non-affine quadrilaterals of f, we pick the one $(\alpha,\beta,\gamma,\delta)$ with the minimum score $T_{\min}=T(\alpha,\beta,\gamma,\delta)=(t_1,t_2,t_3)$. Without loss of generality, we may assume that among α,β,γ and δ, δ is the one that is not in $\mathscr{S}(f)$.

We first consider the case that $(2,2,2)<T_{\min}$. We prove that we can realize a non-affine signature from f by merging. Depending on the values of $T_{\min}$, there are three cases.

- $t_1\geqslant 4$. Without loss of generality, we may assume that $t_1=\mathrm{wt}(\alpha\oplus\beta)$. (Note that even though we have named δ as the one not belonging to $\mathscr{S}(f)$, since $\alpha\oplus\beta\oplus\gamma\oplus\delta=0$, we can name them so that $t_1=\mathrm{wt}(\alpha\oplus\beta)$.) Then, there are at least four bits on which α and β differ. Among these four bits, there are at least two bits on which γ is identical to α or β. Without loss of generality, we assume that these are the first two bits and $\gamma_1\gamma_2=\alpha_1\alpha_2$. We have $\beta_1\beta_2=\overline{\alpha_1\alpha_2}$, and as $\delta=\alpha\oplus\beta\oplus\gamma$, we have $\delta_1\delta_2=\overline{\alpha_1\alpha_2}$. Also by flipping variables, we may assume that $\alpha=\vec{0}^{2n}=00\vec{0}^{2n-2}$. Then, $\beta=11\beta^*$, $\gamma=00\gamma^*$ and $\delta=11\delta^*$ where $\beta^*,\gamma^*,\delta^*\in\mathbb{Z}_2^{2n-2}$ and $\delta^*=\beta^*\oplus\gamma^*$. We consider the following eight inputs of f.

$$\begin{array}{cccc}\alpha=00\alpha^* & \alpha'=11\alpha^* & \beta'=00\beta^* & \beta=11\beta^*\\ \gamma=00\gamma^* & \gamma'=11\gamma^* & \delta'=00\delta^* & \delta=11\delta^*\end{array}$$

  Note that $\gamma'=\alpha\oplus\alpha'\oplus\gamma$, and $\mathrm{wt}(\alpha\oplus\alpha')=2<t_1$. Then,

$$T(\alpha,\alpha',\gamma,\gamma')<T(\alpha,\beta,\gamma,\delta).$$

  By our assumption that $T(\alpha,\beta,\gamma,\delta)$ is the minimum score among all non-affine quadrilaterals of f, $(\alpha,\alpha',\gamma,\gamma')$ is not a non-affine quadrilateral of f. Since $\alpha,\gamma\in\mathscr{S}(f)$, α' and γ' are either both in $\mathscr{S}(f)$ or both not in $\mathscr{S}(f)$. Also, note that $\gamma'=\alpha'\oplus\beta\oplus\delta$, and $\mathrm{wt}(\alpha'\oplus\beta)=\mathrm{wt}(\alpha\oplus\beta)-2=t_1-2<t_1$. Then,

$$T(\alpha',\beta,\gamma',\delta)<T(\alpha,\beta,\gamma,\delta).$$

  Again since $T(\alpha,\beta,\gamma,\delta)$ is the minimum score among all non-affine quadrilaterals of f, $(\alpha',\beta,\gamma',\delta)$ is not a non-affine quadrilateral. Since $\beta\in\mathscr{S}(f)$ and $\delta\notin\mathscr{S}(f)$, α' and γ' are not both in $\mathscr{S}(f)$. Thus, $\alpha',\gamma'\notin\mathscr{S}(f)$. Similarly, $(\beta',\beta,\delta',\delta)$ and $(\alpha,\beta',\gamma,\delta')$ are not non-affine quadrilaterals of f, since their scores are less than $T(\alpha,\beta,\gamma,\delta)$. Since $\beta\in\mathscr{S}(f)$ and $\delta\notin\mathscr{S}(f)$, we cannot have both $\beta',\delta'\in\mathscr{S}(f)$ from considering $(\beta',\beta,\delta',\delta)$, and then from $(\alpha,\beta',\gamma,\delta')$, we cannot have exactly one of β',δ' is in $\mathscr{S}(f)$. Thus, both $\beta',\delta'\notin\mathscr{S}(f)$. In other words, we have $f(\alpha')=f(\beta')=f(\gamma')=f(\delta')=0$.
  Consider the signature $\partial_{12}f$. Then, $f(\alpha)+f(\alpha')$, $f(\beta)+f(\beta')$, $f(\gamma)+f(\gamma')$ and $f(\delta)+f(\delta')$ are entries of $\partial_{12}f$ on inputs α^*, β^*, γ^* and δ^* respectively. Since $f(\alpha)+f(\alpha')=f(\alpha)\neq 0$, $f(\beta)+f(\beta')=f(\beta)\neq 0$ and $f(\gamma)+f(\gamma')=f(\gamma)\neq 0$, $\alpha^*,\beta^*,\gamma^*\in\mathscr{S}(\partial_{12}f)$. Meanwhile we have $f(\delta)+f(\delta')=0+0=0$, thus $\delta^*\notin\mathscr{S}(\partial_{12}f)$. Thus, $(\alpha^*,\beta^*,\gamma^*,\delta^*)$ is a non-affine quadrilateral of $\partial_{12}f$. Then, $\partial_{12}f$ is a non-affine signature of arity $2n-2$. Contradiction.
- $t_1=2$ and $t_2\geqslant 4$. Without loss of generality, we assume that $\mathrm{wt}(\alpha\oplus\gamma)=2$ and $\mathrm{wt}(\alpha\oplus\beta)=t_2\geqslant 4$. (Again, using $\alpha\oplus\beta\oplus\gamma\oplus\delta=0$, a moment reflection shows that this is indeed without loss of generality, even though we have named $\delta\notin\mathscr{S}(f)$.) Again by flipping variables, we may assume that $\alpha=\vec{0}^{2n}$. Then, $\mathrm{wt}(\gamma)=2$ and $\mathrm{wt}(\beta)\geqslant 4$. Take four bits where $\beta_i=1$, for

at most two of these we can have $\gamma_i = 1$, thus there exist two other bits of these four bits (we may assume that they are the first two bits) such that $\gamma_1\gamma_2 = 00$ and $\beta_1\beta_2 = 11$. Then, $\alpha = 00\alpha^*$, $\beta = 11\beta^*$, $\gamma = 00\gamma^*$, and $\delta = 11\delta^*$ by $\delta = \alpha \oplus \beta \oplus \gamma$, where $\beta^*, \gamma^*, \delta^* \in \mathbb{Z}_2^{2n-2}$, $\mathrm{wt}(\beta^*) \geqslant 2$, $\mathrm{wt}(\gamma^*) = 2$ and $\delta^* = \beta^* \oplus \gamma^*$. Still, we consider the following eight inputs of f.

$$\begin{array}{cccc} \alpha = 00\alpha^* & \alpha' = 11\alpha^* & \beta' = 00\beta^* & \beta = 11\beta^* \\ \gamma = 00\gamma^* & \gamma' = 11\gamma^* & \delta' = 00\delta^* & \delta = 11\delta^* \end{array}$$

Note that $\mathrm{wt}(\alpha \oplus \gamma) = 2$ and $\mathrm{wt}(\alpha \oplus \alpha') = 2 < t_2$. Then,

$$T(\alpha, \alpha', \gamma, \gamma') < T(\alpha, \beta, \gamma, \delta).$$

Then similarly since $T(\alpha, \beta, \gamma, \delta)$ is the minimum, $(\alpha, \alpha', \gamma, \gamma')$ is not a non-affine quadrilateral. Since $\alpha, \gamma \in \mathscr{S}(f)$, α' and γ' are either both in $\mathscr{S}(f)$ or both not in it. Also, note that $\mathrm{wt}(\alpha' \oplus \gamma') = 2$ and $\mathrm{wt}(\alpha' \oplus \beta) = \mathrm{wt}(\alpha \oplus \beta) - 2 = t_2 - 2 < t_2$. Then,

$$T(\alpha', \beta, \gamma', \delta) < T(\alpha, \beta, \gamma, \delta).$$

Thus, $(\alpha', \beta, \gamma', \delta)$ is not a non-affine quadrilateral. Since $\beta \in \mathscr{S}(f)$ and $\delta \notin \mathscr{S}(f)$, α' and γ' are not both in $\mathscr{S}(f)$. Thus, $\alpha', \gamma' \notin \mathscr{S}(f)$. Similarly, by considering $(\beta', \beta, \delta', \delta)$ and $(\alpha, \beta', \gamma, \delta')$, we know that they are not non-affine quadrilaterals. Thus, $\beta', \delta' \notin \mathscr{S}(f)$. In other words, we have $f(\alpha') = f(\beta') = f(\gamma') = f(\delta') = 0$. Still consider the signature $\partial_{12} f$. We have $\partial_{12} f \notin \mathscr{A}$. Contradiction.

- $t_1 = 2$, $t_2 = 2$ and $t_3 = 4$. In this case, by the definition of distance-2 squares (equation (7.5)), $\begin{bmatrix} f(\alpha) & f(\beta) \\ f(\gamma) & f(\delta) \end{bmatrix}$ forms a distance-2 square. Clearly, it is not of type I, II or III since exactly one entry of this square is zero. As proved in Lemma 7.9, since f has a distance-2 square that is not type I, II or III, then we can realize a non-affine signature by merging. Contradiction.

Now, we consider the case that $T_{\min} = (2, 2, 2)$.

Then, we show that $|\mathscr{S}(f)| = 2^{2n-2}$. We consider the non-affine quadrilateral $(\alpha, \beta, \gamma, \delta)$ with score $T = (2, 2, 2)$. By renaming and flipping variables, without loss of generality, we may assume that

$$\begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix} = \begin{bmatrix} 0000\vec{0}^{2n-3} & 0110\vec{0}^{2n-3} \\ 1100\vec{0}^{2n-3} & 1010\vec{0}^{2n-3} \end{bmatrix},$$

and δ is the only one among four that is not in $\mathscr{S}(f)$. By normalization, we may assume that $f(\alpha) = 1$. If $f(\gamma) = -1$, then we negate the variable x_1 of f. This keeps f_1^0 unchanged but changes f_1^1 to $-f_1^1$, so this does not change the value of $f(\alpha)$, but changes the value of $f(\gamma)$ to 1. Thus, without loss of generality, we may assume that $f(\alpha) = f(\gamma) = 1$. Clearly, f has even parity. Consider the signature $\widetilde{f}$ by the H_4 gadget applied on variables x_1 and x_2 of f. We have $\widetilde{f}_{12}^{00}(\vec{0}^{2n-2}) = f(\alpha) + f(\gamma) = 2$ and $\widetilde{f}_{12}^{01}(1\vec{0}^{2n-3}) = f(\beta) + f(\delta) = f(\beta)$ since $f(\delta) = 0$. Remember that since $f \in \int_{\mathcal{B}} \mathscr{A}$, by Lemma 7.5, for all $(a, b) \in \mathbb{Z}_2^2$, $\widetilde{f}_{12}^{ab} \in \mathscr{A}$ and we use n_{ab} to denote the norm of its nonzero entries. Thus, $n_{00} = 2$ and $n_{01} = 1$. Also, we have $f(\beta) \neq 0$ which is the same as $1\vec{0}^{2n-3} \in \mathscr{S}(f_{12}^{01})$, and $f(\delta) = 0$ which is the same as $1\vec{0}^{2n-3} \notin \mathscr{S}(f_{12}^{10})$. By Lemma 7.15, $\mathscr{S}(f_{12}^{01}) \cap \mathscr{S}(f_{12}^{10}) = \emptyset$. Remember that $\widetilde{f}_{12}^{01} = f_{12}^{01} + f_{12}^{10}$ and $\widetilde{f}_{12}^{10} = f_{12}^{01} - f_{12}^{10}$. Then,

$$\mathscr{S}(\widetilde{f}_{12}^{01}) = \mathscr{S}(f_{12}^{01}) \cup \mathscr{S}(f_{12}^{10}) = \mathscr{S}(\widetilde{f}_{12}^{10}).$$

Consider signatures $\partial_{ij}^b f$ for all $\{i,j\}$ disjoint with $\{1,2\}$ and every $b \in \mathcal{B}$. By Lemma 4.3 and its remark, we may assume that either $M(\mathfrak{m}_{12}(\partial_{ij}^b f)) = \lambda_{ij}^b I_4$ for some real $\lambda_{ij}^b \neq 0$, or there exists a nonzero binary signature $g_{ij}^b \in \mathcal{O}$ such that $g_{ij}^b(x_1, x_2) \mid \partial_{ij}^b f$. Otherwise, we get #P-hardness.

Consider the case that $g_{ij}^b(x_1, x_2) \mid \partial_{ij}^b f$. If $\partial_{ij}^b f \equiv 0$, then we can let $g_{ij}^b \in \mathcal{B}$ since a zero signature can be divided by any nonzero binary signature. If $\partial_{ij}^b f \not\equiv 0$, we can realize g_{ij}^b by factorization. If $g_{ij}^b \notin \mathcal{B}^{\otimes 1}$, then we get #P-hardness since $\mathcal{F}$ is non-$\mathcal{B}$ hard. Thus, we may assume that $g_{ij}^b \in \mathcal{B}$ after normalization. Therefore, for all $\{i,j\}$ disjoint with $\{1,2\}$ and every $b \in \mathcal{B}$, we may assume that either $M(\mathfrak{m}_{12}(\partial_{ij}^b f)) = \lambda_{ij}^b I_4$ for some real $\lambda_{ij}^b \neq 0$, or there exists a nonzero binary signature $g_{ij}^b \in \mathcal{B}$ such that $g_{ij}^b(x_1, x_2) \mid \partial_{ij}^b f$. Then, by Lemma 7.10, $\mathscr{S}(\widetilde{f}_{12}^{01}) = \mathscr{O}_{2n-2}$. Thus, $|\mathscr{S}(\widetilde{f}_{12}^{01})| = 2^{2n-3}$.

Now consider again the signature f. Since f satisfies 2ND-ORTH, and all its nonzero entries have norm 1, for any $(a,b) \in \mathbb{Z}_2^2$, $|\mathbf{f}_{12}^{ab}|^2 = |\mathscr{S}(f_{12}^{ab})|$. Then,

$$|\mathscr{S}(f_{12}^{00})| = |\mathscr{S}(f_{12}^{01})| = |\mathscr{S}(f_{12}^{10})| = |\mathscr{S}(f_{12}^{11})|.$$

Remember that $\mathscr{S}(f_{12}^{01}) \cap \mathscr{S}(f_{12}^{10}) = \emptyset$, and $\mathscr{S}(\widetilde{f}_{12}^{01}) = \mathscr{S}(f_{12}^{01}) \cup \mathscr{S}(f_{12}^{10})$. Then, $\mathscr{S}(f_{12}^{00})$ and $\mathscr{S}(f_{12}^{01})$ form an equal size partition of $\mathscr{S}(\widetilde{f}_{12}^{01})$. Thus, $|\mathscr{S}(f_{12}^{01})| = |\mathscr{S}(f_{12}^{10})| = \frac{1}{2}|\mathscr{S}(\widetilde{f}_{12}^{01})| = 2^{2n-4}$. Also, $|\mathscr{S}(f_{12}^{00})| = |\mathscr{S}(f_{12}^{11})| = 2^{2n-4}$. Therefore,

$$|\mathscr{S}(f)| = |\mathscr{S}(f_{12}^{00})| + |\mathscr{S}(f_{12}^{01})| + |\mathscr{S}(f_{12}^{10})| + |\mathscr{S}(f_{12}^{11})| = 4 \cdot 2^{2n-4} = 2^{2n-2}.$$

Since all nonzero entries of f have norm 1, $|\mathbf{f}|^2 = |\mathscr{S}(f)| = 2^{2n-2}$. Also, since f satisfies 2ND-ORTH, for all $\{i,j\} \in [2n]$ and all $(a,b) \in \mathbb{Z}_2^2$, $|\mathbf{f}_{ij}^{ab}| = \frac{1}{4}|\mathbf{f}|^2 = 2^{2n-4}$.

We denote $\mathscr{S}(f)$ by S. Since f has even parity, for every $(x_1, \ldots, x_{2n}) \in S$, $x_1 + \cdots + x_{2n} = 0$, i.e., $S \subseteq \mathscr{E}_{2n}$. Let $F(x_1, \ldots, x_{2n-1}) \in \mathbb{Z}_2[x_1, \ldots, x_{2n-1}]$ be the multilinear polynomial such that

$$F(x_1, \ldots, x_{2n-1}) = \begin{cases} 1, & (x_1, \ldots, x_{2n-1}, x_{2n}) \in S \\ 0, & (x_1, \ldots, x_{2n-1}, x_{2n}) \notin S \end{cases} \quad \text{where} \quad x_{2n} = \sum_{i=1}^{2n-1} x_i.$$

Then, $S = \{(x_1, \ldots, x_{2n}) \in \mathscr{E}_{2n} \mid F(x_1, \ldots, x_{2n-1}) = 1\}$.

Now, we show that for all $\{i,j\} \subseteq [2n-1]$, $F_{ij}^{00} + F_{ij}^{11} \equiv 0$ or 1, and also $F_{ij}^{01} + F_{ij}^{10} \equiv 0$ or 1. For simplicity of notations, we prove this for $\{i,j\} = \{1,2\}$. The proof for arbitrary $\{i,j\}$ is the same by replacing $\{1,2\}$ by $\{i,j\}$. Consider

$$S_0 = \mathscr{S}(f_{12}^{00}) = \{(x_3, \ldots, x_{2n}) \in \mathscr{E}_{2n-2} \mid F_{12}^{00}(x_3, \ldots, x_{2n-1}) = 1\},$$

and

$$S_1 = \mathscr{S}(f_{12}^{11}) = \{(x_3, \ldots, x_{2n}) \in \mathscr{E}_{2n-2} \mid F_{12}^{11}(x_3, \ldots, x_{2n-1}) = 1\}.$$

Then,

$$S_0 \cap S_1 = \{(x_3, \ldots, x_{2n}) \in \mathscr{E}_{2n-2} \mid F_{12}^{00} \cdot F_{12}^{11} = 1\},$$

and

$$S_0 \cup S_1 = \{(x_3, \ldots, x_{2n}) \in \mathscr{E}_{2n-2} \mid F_{12}^{00} + F_{12}^{11} + F_{12}^{00} \cdot F_{12}^{11} = 1\}.$$

By Lemma 7.15, $S_0 = S_1$ or $S_0 \cap S_1 = 0$.

- If $S_0 = S_1$, then for every $(x_3, \dots, x_{2n-1}) \in \mathbb{Z}_2^{2n-3}$ which decides every $(x_3, \dots, x_{2n}) \in \mathscr{E}_{2n-2}$ by $x_{2n} = x_3 + \cdots + x_{2n-1}$,

$$F_{12}^{00}(x_3, \dots, x_{2n-1}) = F_{12}^{11}(x_3, \dots, x_{2n-1}).$$

  Thus, $F_{12}^{00} + F_{12}^{11} \equiv 0$.
- If $S_0 \cap S_1 = \emptyset$, then since $|S_0| = |S_1| = 2^{2n-4}$ (which is still true when replacing $\{1, 2\}$ by an arbitrary $\{i, j\}$), $|S_0 \cup S_1| = |S_0| + |S_1| = 2^{2n-3}$. Since $S_0 \cup S_1 \subseteq \mathscr{E}_{2n-2}$ and $|\mathscr{E}_{2n-2}| = 2^{2n-3}$, $S_0 \cup S_1 = \mathscr{E}_{2n-2}$. Thus, for every $(x_3, \dots, x_{2n-1}) \in \mathbb{Z}_2^{2n-3}$ which decides every $(x_3, \dots, x_{2n}) \in \mathscr{E}_{2n-2}$ by $x_{2n} = x_3 + \cdots + x_{2n-1}$,

$$F_{12}^{00}(x_3, \dots, x_{2n-1}) \cdot F_{12}^{11}(x_3, \dots, x_{2n-1}) = 0,$$

  and

$$F_{12}^{00}(x_3, \dots, x_{2n-1}) + F_{12}^{11}(x_3, \dots, x_{2n-1}) + F_{12}^{00} \cdot F_{12}^{11}(x_3, \dots, x_{2n-1}) = 1.$$

  Thus, $F_{12}^{00} + F_{12}^{11} \equiv 1$.

Similarly, we can show that $F_{12}^{01} + F_{12}^{10} \equiv 0$ or 1. Therefore, for all $\{i, j\} \subseteq [2n-1]$, $F_{ij}^{00} + F_{ij}^{11} \equiv 0$ or 1 and $F_{ij}^{01} + F_{ij}^{10} \equiv 0$ or 1. By Lemma 7.8, $d(F) \leqslant 2$.

If $d(F) \leqslant 1$, then clearly, $S = \{(x_1, \dots, x_{2n}) \in \mathscr{E}_{2n} \mid F(x_1, \dots, x_{2n-1}) = 1\}$ is an affine linear space. Thus, f has affine support. Otherwise, $d(F) = 2$. By Lemma 7.8, F is a complete quadratic polynomial. Consider signatures f_{12}^{00} and f_{12}^{11}. Remember that $f(0000\vec{0}^{2n-3}) = f(1100\vec{0}^{2n-3}) = 1$. Thus, $\vec{0}^{2n-2} \in S_0 \cap S_1 \neq \emptyset$. Then, $S_0 = S_1$. Let

$$S_+ = \{\alpha \in S_0 \mid f_{12}^{00}(\alpha) = f_{12}^{11}(\alpha)\} \quad \text{and} \quad S_- = \{\alpha \in S_0 \mid f_{12}^{00}(\alpha) = -f_{12}^{11}(\alpha)\}.$$

Then, as f takes ± 1 values on its support, $S_+ = \mathscr{S}(\partial_{12}^+ f)$ and $S_- = \mathscr{S}(\partial_{12}^- f)$. Since $\partial_{12}^+ f, \partial_{12}^- f \in \mathscr{A}$, S_+ and S_- are affine linear subspaces of $\mathscr{E}_{2n-2}$. Also, by 2ND-ORTH, $\langle \mathbf{f}_{12}^{00}, \mathbf{f}_{12}^{11} \rangle = |S_+| - |S_-| = 0$. Thus, $|S_+| = |S_-| = \frac{1}{2}|S_0| = 2^{2n-5}$. Since $|\mathscr{E}_{2n-2}| = 2^{2n-3}$, S_+ is a an affine linear subspaces of $\mathscr{E}_{2n-2}$ decided by two affine linear constraints $L_1^+ = 1$ and $L_2^+ = 1$. (Here both L_1^+ and L_2^+ are *affine* linear forms which may have nonzero constant terms, but we write the constraints as $L_1^+ = 1$ and $L_2^+ = 1$.) In other words,

$$S_+ = \{(x_3, \dots, x_{2n}) \in \mathscr{E}_{2n-2} \mid L_1^+ = 1 \text{ and } L_2^+ = 1\} = \{(x_3, \dots, x_{2n}) \in \mathscr{E}_{2n-2} \mid L_1^+ \cdot L_2^+ = 1\}.$$

Since for every $(x_3, \dots, x_{2n}) \in \mathscr{E}_{2n-2}$, $x_3 + \cdots + x_{2n} = 0$, we may substitute the appearance of x_{2n} in L_1^+ and L_2^+ by $x_3 + \cdot + x_{2n-1}$. Thus, we may assume that $L_1^+, L_2^+ \in \mathbb{Z}_2[x_3, \dots, x_{2n-1}]$, and $d(L_1^+) = d(L_2^+) = 1$. Similarly, there exist $L_1^-, L_2^- \in \mathbb{Z}_2[x_3, \dots, x_{2n-1}]$ with $d(L_1^-) = d(L_2^-) = 1$ such that

$$S_- = \{(x_3, \dots, x_{2n}) \in \mathscr{E}_{2n-2} \mid L_1^- = 1 \text{ and } L_2^- = 1\} = \{(x_3, \dots, x_{2n}) \in \mathscr{E}_{2n-2} \mid L_1^- \cdot L_2^- = 1\}.$$

Clearly, $S_+ \cap S_- = \emptyset$. Then

$$S_+ \cup S_- = \{(x_3, \dots, x_{2n}) \in \mathscr{E}_{2n-2} \mid L_1^+ \cdot L_2^+ + L_1^- \cdot L_2^- = 1\}.$$

Remember that

$$S_0 = S_+ \cup S_- = \{(x_3, \dots, x_{2n}) \in \mathscr{E}_{2n-2} \mid F_{12}^{00} = 1\}.$$

Thus, $L_1^+ \cdot L_2^+ + L_1^- \cdot L_2^- = F_{12}^{00}$. Since for all $1 \leqslant i < j \leqslant 2n-1$, the quadratic term $x_i x_j$ appears in F, for all $3 \leqslant i < j \leqslant 2n-1$, the quadratic term $x_i x_j$ appears in F_{12}^{00}. Thus, $F_{12}^{00} \in \mathbb{Z}_2[x_3, \ldots, x_{2n-1}]$ is a complete quadratic polynomial over $2n-3$ variables and it has a twice-linear 2-partition. Since $2n \geqslant 8$, $2n-3 \geqslant 5$. By Lemma 7.14, we have $2n-3 = 5$, and after renaming variables,

$$F = (x_3 + x_4 + a)(x_4 + x_5 + b) + (x_3 + x_4 + x_5 + x_6 + c)(x_3 + x_4 + x_5 + x_7 + d)$$

where $a, b, c, d \in \mathbb{Z}_2$. Without loss of generality, we may assume that $L_1^+ \cdot L_2^+ = (x_3 + x_4 + a)(x_4 + x_5 + b)$. Then,

$$S_+ = \mathscr{S}(\partial_{12}^+ f) = \{(x_3, \ldots, x_8) \in \mathscr{E}_{2n-2} \mid x_3 = x_4 + a \text{ and } x_4 = x_5 + b\},$$

for some $a, b \in \mathbb{Z}_2$.

Clearly $\partial_{12}^+ f$ is a 6-ary signature and $|\mathscr{S}(\partial_{12}^+ f)| = 2^{5-2} = 2^3$. We show that $\partial_{12}^+ f \notin \mathcal{B}^{\otimes 3} \cup \mathcal{F}_6 \cup \mathcal{F}_6^H$. Then, by Corollary 6.10, we get #P-hardness. Since the support of a signature in $\mathcal{F}_6 \cup \mathcal{F}_6^H$ is either $\mathscr{E}_6$ or $\mathscr{O}_6$ whose sizes are both 2^5. Thus, $\partial_{12}^+ f \notin \mathcal{F}_6 \cup \mathcal{F}_6^H$. For any 6-ary signature g in $\mathcal{B}^{\otimes 3}$, its 6 variables can be divided into three independent pairs such that on the support $\mathscr{S}(g)$, the values of variables inside each pair do not rely on the values of variables of other pairs. Thus, if we pick any three variables in $\mathscr{S}(g)$, the degree of freedom of them is at least 2; more precisely, there are at least 4 assignments on these three variables which can be extended to an input in $\mathscr{S}(g)$. However, in $\mathscr{S}(\partial_{12}^+ f)$, the degree of freedom of variables x_3, x_4, x_5 is only 1, namely there are only two assignments on x_3, x_4, x_5 that can be extended to an input in $\mathscr{S}(\partial_{12}^+ f)$. Thus, $\partial_{12}^+ f \notin \mathcal{B}^{\otimes 3}$. This completes the proof of Lemma 7.16. □

7.4 Affine signature condition

Finally, by further assuming that f has affine support, we consider whether f itself is an affine signature. We prove that this is true only for signature of arity $2n \geqslant 10$. For signature f of arity $2n = 8$, we show that either $f \in \mathscr{A}$ or the following signature is realizable.

$$h_8 = \chi_T \cdot (-1)^{x_1x_2x_3 + x_1x_2x_5 + x_1x_3x_5 + x_2x_3x_5}, \text{ where } T = \mathscr{S}(h_8) = \mathscr{S}(f_8).$$

Note that in the support $\mathscr{S}(f_8)$ (see its definition (7.4) for this *Queen of the Night* f_8), by taking x_1, x_2, x_3, x_5 as free variables, the remaining 4 variables are mod 2 sums of $\binom{4}{3}$ subsets of $\{x_1, x_2, x_3, x_5\}$. Clearly, h_8 is not affine, but it has affine support and all its nonzero entries have the same norm. One can check that h_8 satisfies 2ND-ORTH and $h_8 \in \int_{\mathcal{B}} \mathscr{A}$. But fortunately, we show that by merging h_8, we can realize a 6-ary signature that is not in $\mathcal{B}^{\otimes} \cup \mathcal{F}_6 \cup \mathcal{F}_6^H$. By Corollary 6.10, we are done.

After we give one more result about multilinear boolean polynomials, we make our final step towards Theorem 7.19.

Lemma 7.17. *Let $F(x_1, \ldots, x_n) \in \mathbb{Z}_2[x_1, \ldots, x_n]$ be a complete cubic polynomial, $L(x_2, \ldots, x_n) \in \mathbb{Z}_2[x_2, \ldots, x_n]$ and $d(L) \leqslant 1$. If we substitute x_1 by $x_{n+1} + L(x_2, \ldots, x_n)$ in F to get F', and suppose $F'(x_2, \ldots, x_{n+1}) = F(x_{n+1} + L, x_2, \ldots, x_n) \in \mathbb{Z}_2[x_2, \ldots, x_{n+1}]$ is also a complete cubic polynomial, then*

- *If $n \geqslant 5$, then L must be a constant $\epsilon = 0$ or 1.*
- *If $n = 4$, then L must be either ϵ, or of the form $x_i + x_j + \epsilon$, for some $\epsilon = 0$ or 1, for some $\{i, j\} \in \{2, 3, 4\}$.*

Proof. Since $F(x_1, \ldots, x_n)$ is a complete cubic polynomial, we can write it as

$$F(x_1, \ldots, x_n) = x_1 \cdot \sum_{2 \leqslant i < j \leqslant n} x_i x_j + \sum_{2 \leqslant i < j < k \leqslant n} x_i x_j x_k + G(x_1, \ldots, x_n)$$

where $d(G) \leqslant 2$. Then,

$$F'(x_2, \ldots, x_n, x_{n+1}) = (x_{n+1} + L) \cdot \sum_{2 \leqslant i < j \leqslant n} x_i x_j + \sum_{2 \leqslant i < j < k \leqslant n} x_i x_j x_k + G(x_{n+1} + L, \ldots, x_n).$$

Let $G'(x_2, \ldots, x_n, x_{n+1}) = G(x_{n+1} + L, \ldots, x_n)$. Since $d(L) \leqslant 1$ and $d(G) \leqslant 2$, $d(G') \leqslant 2$. Then, there is no cubic term in $G'(x_2, \ldots, x_n, x_{n+1})$. Since $F'(x_2, \ldots, x_n, x_{n+1})$ is a complete cubic polynomial over variables $(x_2, \ldots, x_n, x_{n+1})$ and $x_{n+1} \cdot \sum_{2 \leqslant i < j \leqslant n} x_i x_j + \sum_{2 \leqslant i < j < k \leqslant n} x_i x_j x_k$ already gives every cubic term over $(x_2, \ldots, x_n, x_{n+1})$ exactly once, there is no cubic term in $L \cdot \sum_{2 \leqslant i < j \leqslant n} x_i x_j$ (after taking module 2). If $L \equiv 0$ or 1, then we are done. Otherwise, there is a variable that appears in L. Without loss of generality, we may assume that $x_2 \in L$ (i.e., x_2 appears in L).

Let $Q(x_3, \ldots, x_n) = \sum_{3 \leqslant i < j \leqslant n} x_i x_j \in \mathbb{Z}_2[x_3, \ldots, x_n]$. Since $n \geqslant 4$, we have $Q \not\equiv 0$. For every $x_i x_j \in Q$, since $x_2 \in L$, the cubic term $x_2 x_i x_j$ will appear in $L \cdot \sum_{2 \leqslant i < j \leqslant n} x_i x_j$. To cancel it, exactly one between $x_i \cdot x_2 x_j$ and $x_j \cdot x_2 x_i$ must also appear in $L \cdot \sum_{2 \leqslant i < j \leqslant n} x_i x_j$. Thus, exactly one between x_i and x_j appears in L.

If $n \geqslant 5$, then $x_3 x_4, x_4 x_5, x_3 x_5 \in Q$. Thus, exactly one between x_3 and x_4 is in L, exactly one between x_4 and x_5 is in L, and exactly one between x_3 and x_5 is in L. Clearly, this is a contradiction.

If $n = 4$, then $Q = x_3 x_4$. Either x_3 or x_4 appears in L. Thus, L is a sum of two variables among $\{x_2, x_3, x_4\}$ plus a constant 0 or 1. □

Lemma 7.18. *Let $\mathcal{F}$ be non-$\mathcal{B}$ hard. Let $f \in \mathcal{F}$ be an irreducible $2n$-ary ($2n \geqslant 8$) signature with parity. Then,*

- $\text{Holant}^b(\mathcal{F})$ *is #P-hard, or*
- *there is a signature $g \notin \mathscr{A}$ of arity $2k < 2n$ that is realizable from f and $\mathcal{B}$, or*
- $f \in \mathscr{A}$.

Proof. Again, we may assume that f satisfies 2ND-ORTH and $f \in \int_{\mathcal{B}} \mathscr{A}$. Also by Lemmas 7.9 and 7.16, we may assume that $f(\alpha) = \pm 1$ for all $\alpha \in \mathscr{S}(f)$ and $\mathscr{S}(f)$ is an affine linear space. Let $\{x_1, \ldots, x_m\}$ be a set of free variables of $\mathscr{S}(f)$. Then, on the support $\mathscr{S}(f)$, every variable x_i ($1 \leqslant i \leqslant 2n$) is expressible as a unique affine linear combination over $\mathbb{Z}_2$ of these free variables, i.e., $x_i = L_i(x_1, \ldots, x_m) = \lambda_i^0 + \lambda_i^1 x_1 + \ldots + \lambda_i^m x_m$, where $\lambda_i^0, \ldots, \lambda_i^m \in \mathbb{Z}_2$. Clearly, for $1 \leqslant i \leqslant m$, $L(x_i) = x_i$. Then,

$$\begin{aligned} \mathscr{S}(f) &= \{(x_1, \ldots, x_{2n}) \in \mathbb{Z}_2^{2n} \mid x_1 = L_1, \ldots, x_{2n} = L_{2n}\} \\ &= \{(x_1, \ldots, x_{2n}) \in \mathbb{Z}_2^{2n} \mid x_{m+1} = L_{m+1}, \ldots, x_{2n} = L_{2n}\}. \end{aligned}$$

Also, let $I(x_i) = \{1 \leqslant k \leqslant m \mid \lambda_i^k = 1\}$. Clearly, for $1 \leqslant i \leqslant m$, $I(x_i) = \{i\}$. For $m+1 \leqslant i \leqslant 2n$, we show that $|I_{x_i}| \geqslant 2$. For a contradiction, suppose that there exists $m + 1 \leqslant i \leqslant 2n$ such that $|I_{x_i}| = 0$ or 1. If $|I_{x_i}| = 0$, then x_i takes a constant value in $\mathscr{S}$. Then, among f_i^0 and f_i^1, one is a zero signature. Thus, f is reducible. Contradiction. If $|I_{x_i}| = 1$, then $x_i = x_k$ or $x_k + 1$ for some free variable x_k. Then, among f_{ik}^{00}, f_{ik}^{01}, f_{ik}^{10} and f_{ik}^{11}, two are zero signatures. Thus, f does not satisfy 2ND-ORTH. Contradiction.

Since $f(\alpha) = \pm 1$ for all $\alpha \in \mathscr{S}(f)$ and each $\alpha \in \mathscr{S}(f)$ can be uniquely decided by its value on the first m free variables, there exists a unique multilinear boolean polynomial $F(x_1, \ldots, x_m) \in \mathbb{Z}_2[x_1, \ldots, x_m]$ such that

$$f(x_1, \ldots, x_m, \ldots, x_{2n}) = \chi_S(-1)^{F(x_1,\ldots,x_m)}$$

where $S = \mathscr{S}(f)$. If $d(F) \leqslant 2$, then clearly $f \in \mathscr{A}$. We are done. Thus, we may assume that $d(F) > 2$ and hence $m > 2$. Remember that F_{ij}^{ab} denotes the polynomial obtained by setting variables (x_i, x_j) of F to $(a, b) \in \mathbb{Z}_2^2$. Then, $f_{ij}^{ab} = (-1)^{F_{ij}^{ab}}$ on $\mathscr{S}(f)$. We will show that for all $i, j \in [m]$, $d(F_{ij}^{00} + F_{ij}^{11}) \leqslant 1$ and $d(F_{ij}^{01} + F_{ij}^{10}) \leqslant 1$. For brevity of notation, we prove this for $\{i, j\} = \{1, 2\}$. The proof for arbitrary $\{i, j\}$ is the same by replacing $\{1, 2\}$ with $\{i, j\}$. We first show that $d(F_{ij}^{00} + F_{ij}^{11}) \leqslant 1$. We use S_0 to denote $\mathscr{S}(f_{ij}^{00})$ and S_1 to denote $\mathscr{S}(f_{ij}^{11})$. By Lemma 7.15, there are two cases, $S_0 = S_1$ or $S_0 \cap S_1 = \emptyset$.

- Suppose that $S_0 = S_1$. For convenience, we use L_i^0 to denote $(L_i)_{12}^{00}$ and L_i^1 to denote $(L_i)_{12}^{11}$. Then,

  $$S_0 = \{(x_3, \ldots, x_{2n}) \in \mathbb{Z}_2^{2n-2} \mid x_{m+1} = L_{m+1}^0, \ldots, x_{2n} = L_{2n}^0\}$$
  $$S_1 = \{(x_3, \ldots, x_{2n}) \in \mathbb{Z}_2^{2n-2} \mid x_{m+1} = L_{m+1}^1, \ldots, x_{2n} = L_{2n}^1\}.$$

  So $L_i^0 \equiv L_i^1$ for all $i \geqslant m+1$. Thus, either $\{1, 2\} \subseteq I(x_i)$ or $\{1, 2\} \cap I(x_i) = \emptyset$ for $i \geqslant m+1$. Let $S_+ = \{\alpha \in S_0 \mid f_{ij}^{00}(\alpha) = f_{ij}^{11}(\alpha)\}$ and $S_- = \{\alpha \in S_0 \mid f_{ij}^{00}(\alpha) = -f_{ij}^{11}(\alpha)\}$. Then, $\langle \mathbf{f}_{ij}^{00}, \mathbf{f}_{ij}^{11} \rangle = 1 \cdot |S_+| - 1 \cdot |S_-| = 0$. Since $S_0 = S_+ \cup S_-$, $|S_+| = |S_-| = \frac{1}{2}|S_0|$. Note that $\mathscr{S}(\partial_{12} f) = S_+$ and $\mathscr{S}(\partial_{12}^- f) = S_-$. By our assumption that $f \in \int_{\mathcal{B}} \mathscr{A}$, $\partial_{12} f, \partial_{12}^- f \in \mathscr{A}$. Thus, both S_+ and S_- are affine linear subspaces of $S_0 = S_1$. Since $|S_+| = |S_-| = |S_0|/2$, there exists an (affine) linear polynomial $L(x_3, \ldots, x_{2n})$ such that

  $$S_+ = \{(x_3, \ldots, x_{2n}) \in S_0 \mid L(x_3, \ldots, x_{2n}) = 0\},$$

  and

  $$S_- = \{(x_3, \ldots, x_{2n}) \in S_0 \mid L(x_3, \ldots, x_{2n}) = 1\}.$$

  For $(x_3, \ldots, x_{2n}) \in S_0$, and $i \geqslant m+1$, we can substitute the variable x_i that appears in $L(x_3, \ldots, x_{2n})$ with $L_i^0 \equiv L_i^1$. Then, we get an (affine) linear polynomial $L'(x_3, \ldots, x_m) \in \mathbb{Z}_2[x_1, \ldots, x_m]$ such that $L'(x_3, \ldots, x_m) = L(x_3, \ldots, x_m, x_{m+1}, \ldots, x_{2n})$ for $(x_3, \ldots, x_{2n}) \in S_0$. Thus,

  $$S_+ = \{(x_3, \ldots, x_{2n}) \in S_0 \mid L'(x_3, \ldots, x_m) = 0\},$$

  and

  $$S_- = \{(x_3, \ldots, x_{2n}) \in S_0 \mid L'(x_3, \ldots, x_m) = 1\}.$$

  Note that as $|S_+| = |S_-| > 0$, the affine linear polynomial L' is non-constant, i.e., $d(L') = 1$. Then, for every $(x_3, \ldots, x_m) \in \mathbb{Z}_2^{m-2}$,

  $$(-1)^{F_{12}^{00}(x_3,\ldots,x_m)} = (-1)^{F_{12}^{11}(x_3,\ldots,x_m)} \text{ if } L'(x_3, \ldots, x_m) = 0$$

  and

  $$(-1)^{F_{12}^{00}(x_3,\ldots,x_m)} = -(-1)^{F_{12}^{11}(x_3,\ldots,x_m)} \text{ if } L'(x_3, \ldots, x_m) = 1.$$

  Thus,

  $$(-1)^{F_{12}^{00}(x_3,\ldots,x_m) + F_{12}^{11}(x_3,\ldots,x_m)} = (-1)^{L'(x_3,\ldots,x_m)}.$$

  Therefore, $F_{12}^{00}(x_3, \ldots, x_m) + F_{12}^{11}(x_3, \ldots, x_m) \equiv L'(x_3, \ldots, x_m)$. Then, $d(F_{12}^{00} + F_{12}^{11}) = 1$.

- Suppose that $S_0 \cap S_1 = \emptyset$. Then, there exists a variable x_i where $i \geqslant m+1$ such that between $\{1,2\}$, exactly one index is in $I(x_i)$. Without loss of generality, we may assume that $i = m+1$, $1 \in I(x_{m+1})$ and $2 \notin I(x_{m+1})$. Then, $x_{m+1} = x_1 + K(x_3, \ldots, x_m)$ where $K \in \mathbb{Z}_2[x_3, \ldots, x_m]$ is an (affine) linear polynomial. Consider S_0.

$$S_0 = \{(x_3, \ldots, x_{2n}) \in \mathbb{Z}_2^{2n-2} \mid x_1 = x_2 = 0, x_{m+1} = x_1 + K, x_{m+2} = L_{m+2} \ldots, x_{2n} = L_{2n}\}.$$

Since $x_1 = x_2$ on S_0, for every $i \geqslant m+2$, if x_1 or x_2 appear in L_i, we substitute each one of them with $x_{m+1} + K$. We get a linear polynomial $K_i \in \mathbb{Z}_2[x_3, \ldots, x_m, x_{m+1}]$. Then, for every $(x_3, \ldots, x_{2n}) \in S_0$, $L_i = K_i$. Thus,

$$S_0 = \{(x_3, \ldots, x_{2n}) \in \mathbb{Z}_2^{2n-2} \mid x_{m+1} + K = 0, x_{m+2} = K_{m+2} \ldots, x_{2n} = K_{2n}\}.$$

Similarly, we have

$$S_1 = \{(x_3, \ldots, x_{2n}) \in \mathbb{Z}_2^{2n-2} \mid x_{m+1} + K = 1, x_{m+2} = K_{m+2} \ldots, x_{2n} = K_{2n}\}.$$

Let $S_\cup = S_0 \cup S_1$. Then,

$$S_\cup = \{(x_3, \ldots, x_{2n}) \in \mathbb{Z}_2^{2n-2} \mid x_{m+2} = K_{m+2} \ldots, x_{2n} = K_{2n}\}.$$

Thus, we can pick $x_3, \ldots, x_m, x_{m+1}$ as a set of free variables of $S_\cup$.
Consider $g = \partial_{12} f$. Clearly, $\mathscr{S}(g) = S_\cup$ since $S_0 \cap S_1 = \emptyset$. Then, there exists a unique multilinear boolean polynomial $G(x_3, \ldots, x_{m+1}) \in \mathbb{Z}_2[x_3, \ldots, x_{m+1}]$ such that

$$g(x_3, \ldots, x_{2n}) = \chi_{S_\cup} \cdot (-1)^{G(x_3, \ldots, x_{m+1})}.$$

For every $(x_3, \ldots, x_{2n}) \in S_0$ that is uniquely decided by $(0, 0, x_3, \ldots, x_m) \in \{(0,0)\} \times \mathbb{Z}_2^{m-2}$, $x_{m+1} = K(x_3, \ldots, x_m)$ and $f_{12}^{00}(x_3, \ldots, x_{2n}) = g(x_3, \ldots, x_{2n})$. Thus, for every $(x_3, \ldots, x_m) \in \mathbb{Z}_2^{m-2}$,

$$(-1)^{F_{12}^{00}(x_3, \ldots, x_m)} = (-1)^{G(x_3, \ldots, x_m, K)}.$$

Also, for every $(x_3, \ldots, x_{2n}) \in S_1$ that is uniquely decided by $(1, 1, x_3, \ldots, x_m) \in \{(1,1)\} \times \mathbb{Z}_2^{m-2}$, $x_{m+1} = K(x_3, \ldots, x_m) + 1$, and $f_{12}^{11}(x_3, \ldots, x_{2n}) = g(x_3, \ldots, x_{2n})$. Thus, for every $(x_3, \ldots, x_m) \in \mathbb{Z}_2^{m-2}$,

$$(-1)^{F_{12}^{11}(x_3, \ldots, x_m)} = (-1)^{G(x_3, \ldots, x_m, K+1)}.$$

Thus, $F_{12}^{00}(x_3, \ldots, x_m) \equiv G(x_3, \ldots, x_m, K)$ and $F_{12}^{11}(x_3, \ldots, x_m) \equiv G(x_3, \ldots, x_m, K+1)$.
Since $f \in \int_{\mathcal{B}} \mathscr{A}$, $g = \partial_{12} f \in \mathscr{A}$. Thus,

$$g'(x_3, \ldots, x_m, x_{m+1}) = (-1)^{G(x_3, \ldots, x_m, x_{m+1})}$$

is also in $\in \mathscr{A}$. Let $y = x_{m+1} + K(x_3, \ldots, x_m) \in \mathbb{Z}[x_3, \ldots, x_{m+1}]$ be an affine linear combination of variables $x_3, \ldots, x_{m+1}$. Since $g \in \mathscr{A}$, by Lemma 2.19,

$$d[G(x_3, \ldots, x_m, K) + G(x_3, \ldots, x_m, K+1)] \leqslant 1.$$

Thus, $d(F_{12}^{00} + F_{12}^{11}) \leqslant 1$. Also if $d(G) = 1$, then by Lemma 2.19

$$d(F_{12}^{00} + F_{12}^{11}) = 0, \text{ i.e., } F_{12}^{00} + F_{12}^{11} \equiv 0 \text{ or } 1. \tag{7.10}$$

Similarly, we can show that $d(F_{12}^{01} + F_{12}^{10}) \leqslant 1$. Thus, for all $i, j \in [m]$, $d(F_{ij}^{00} + F_{ij}^{11}) \leqslant 1$ and $d(F_{ij}^{01} + F_{ij}^{10}) \leqslant 1$. By Lemma 7.8, $d(F) \leqslant 3$.

If $d(F) \leqslant 2$, then clearly $f \in \mathscr{A}$. We are done. Otherwise, $d(F) = 3$ and by Lemma 7.8, F is a complete cubic multilinear polynomial over m variables. If we pick another set X of m free variables and substitute variables of F by variables in X, then we will get a cubic multilinear polynomial F' over variables in X. Same as the analysis of F, F' is also a complete cubic polynomial. In particular, consider the variable x_{m+1}. Recall that $|I(x_{m+1})| \geqslant 2$. Without loss of generality, we assume that $1 \in I(x_{m+1})$. Then, $x_{m+1} = x_1 + L(x_2, \ldots, x_m)$ where $L(x_2, \ldots, x_m)$ is an affine linear combination of variables $x_2, \ldots, x_m$. We substitute x_1 in F by $x_{m+1} + L$, and we get a complete cubic multilinear polynomial $F'(x_2, \ldots, x_{m+1}) \in \mathbb{Z}_2[x_2, \ldots, x_{m+1}]$. By Lemma 7.17, if $m \geqslant 5$, then $x_{m+1} = x_1$ or $x_{m+1} = \overline{x_1}$. Thus, $I(x_{m+1}) = \{1\}$. This contradicts with $|I(x_{m=1})| \geqslant 2$. Thus, $m \leqslant 4$.

If $m = 4$, then by Lemma 7.17, $x_5 = x_1 + \epsilon$, or $x_5 = x_1 + x_i + x_j + \epsilon$, where $\epsilon = 0$ or 1, for some $2 \leqslant i < j \leqslant 4$. Since $|I(x_5)| \geqslant 2$, the case that $x_5 = x_1 + \epsilon$ is impossible. Similarly, for $i \geqslant m + 2$, the variable x_i is a sum of three variables in $\{x_1, x_2, x_3, x_4\}$ plus a constant 0 or 1. If there exist x_i and x_j for $5 \leqslant i < j \leqslant 2n$ such that $I(x_i) = I(x_j)$. Then, $x_i = x_j$ or $\overline{x_j}$. Thus, among f_{ij}^{00}, f_{ij}^{01}, f_{ij}^{10} and f_{ij}^{11}, two are zero signatures. Thus, f does not satisfy 2ND-ORTH. Contradiction. Thus, $I(x_i) \neq I(x_j)$ for any $5 \leqslant i < j \leqslant 2n$. There are only $\binom{4}{3} = 4$ ways to pick three variables from $\{x_1, x_2, x_3, x_4\}$. Thus, $2n \leqslant 4 + 4 = 8$. By the hypothesis $2n \geqslant 8$ of the lemma, we have $2n = 8$. Clearly, $|\mathscr{S}(f)| = 2^4 = 16$. Due to 2ND-ORTH, for all $\{i, j\} \in [8]$, $|\mathscr{S}(f_{ij}^{00})| = |\mathscr{S}(f_{ij}^{01})| = |\mathscr{S}(f_{ij}^{10})| = |\mathscr{S}(f_{ij}^{11})| = 4$.

- If there exists $\{i, j\}$ such that $\mathscr{S}(f_{ij}^{00}) = \mathscr{S}(f_{ij}^{11})$, then for any point α in $\mathscr{S}(f_{ij}^{00}) = \mathscr{S}(f_{ij}^{11})$, regardless whether $f_{ij}^{00}(\alpha) = f_{ij}^{11}(\alpha)$ or $f_{ij}^{00}(\alpha) = -f_{ij}^{11}(\alpha)$, either $\alpha \in \mathscr{S}(\partial_{ij}^{+} f)$ or $\alpha \in \mathscr{S}(\partial_{ij}^{-} f)$. Thus,
$$\mathscr{S}(\partial_{ij}^{+} f) \cup \mathscr{S}(\partial_{ij}^{-} f) = \mathscr{S}(f_{ij}^{00}) = \mathscr{S}(f_{ij}^{11}).$$
Also, by 2ND-ORTH,
$$\langle \mathbf{f}_{ij}^{00}, \mathbf{f}_{ij}^{11} \rangle = |\mathscr{S}(\partial_{ij}^{-} f)| - |\mathscr{S}(\partial_{ij}^{+} f)| = 0.$$
Thus, $|\mathscr{S}(\partial_{ij}^{+} f)| = |\mathscr{S}(\partial_{ij}^{-} f)| = 2$. Note that every 6-ary signature in $\mathcal{B}^{\otimes}$ has support of size 8, and every signature in $\mathcal{F}_6$ and $\mathcal{F}_6^H$ has support of size 32. Thus, $\partial_{ij}^{+} f \notin \mathcal{B} \cup \mathcal{F}_6 \cup \mathcal{F}_6^H$. Then, by Corollary 6.10, we get #P-hardness. Similarly, if there exists $\{i, j\}$ such that $\mathscr{S}(f_{ij}^{01}) = \mathscr{S}(f_{ij}^{10})$, then we have $|\mathscr{S}(\widehat{\partial_{ij}^{+}} f)| = |\mathscr{S}(\widehat{\partial_{ij}^{-}} f)| = 2$. Thus, $\widehat{\partial_{ij}^{+}} f \notin \mathcal{B}^{\otimes} \cup \mathcal{F}_6 \cup \mathcal{F}_6^H$. Again, we get #P-hardness.
- Otherwise, for all $\{i, j\} \in [8]$, $\mathscr{S}(f_{ij}^{00}) \cap \mathscr{S}(f_{ij}^{11}) = \emptyset$ and $\mathscr{S}(f_{ij}^{01}) \cap \mathscr{S}(f_{ij}^{10}) = \emptyset$. Then, $\mathscr{S}(\partial_{ij}^{+} f) = \mathscr{S}(f_{ij}^{00}) \cup \mathscr{S}(f_{ij}^{11})$. Thus, $|\mathscr{S}(\partial_{ij}^{+} f)| = 8$. Clearly, $\partial_{ij}^{+} f \notin \mathcal{F}_6 \cup \mathcal{F}_6^H$. If $\partial_{ij}^{+} f \notin \mathcal{B}^{\otimes 3}$, then we get #P-hardness. For a contradiction, suppose that $\partial_{ij}^{+} f \in \mathcal{B}^{\otimes 3}$. Then,
$$\partial_{ij}^{+} f = \chi_{\mathscr{S}(\partial_{ij}^{+} f)}(-1)^{G_{ij}^{+}} \quad \text{where} \quad d(G_{ij}^{+}) = 1.$$
As we proved above in equation (7.10), $F_{ij}^{00} + F_{ij}^{11} \equiv 0$ or 1. Similarly, suppose $\widehat{\partial_{ij}^{+}} f \in \mathcal{B}^{\otimes 3}$, and we can show that $F_{ij}^{01} + F_{ij}^{10} \equiv 0$ or 1. Thus, for all $\{i, j\} \subseteq [8]$, $F_{ij}^{00} + F_{ij}^{11} \equiv 0$ or 1 and $F_{ij}^{01} + F_{ij}^{10} \equiv 0$ or 1. Then, by Lemma 7.8, $d(F) \leqslant 2$. Contradiction.

Suppose that $m = 3$. Remember that for $4 \leqslant i \leqslant 2n$, $|I(x_i)| \geqslant 2$. Thus, x_i is a sum of at least two variables in $\{x_1, x_2, x_3\}$ plus a constant 0 or 1. Again, if there exist x_i and x_j for $4 \leqslant i < j \leqslant 2n$

such that $I(x_i) = I(x_j)$, then among f_{ij}^{00}, f_{ij}^{01}, f_{ij}^{10} and f_{ij}^{11}, two are zero signatures. Contradiction. Thus, $I(x_i) \neq I(x_j)$ for any $4 \leqslant i < j \leqslant 2n$. There are $\binom{3}{2} + \binom{3}{3} = 4$ different ways to pick at least two variables from $\{x_1, x_2, x_3\}$. Thus, $2n \leqslant 3 + 4 = 7$. Contradiction. □

Theorem 7.19. *Suppose that $\mathcal{F}$ is non-$\mathcal{B}$ hard. Then,* $\text{Holant}^b(\mathcal{F})$ *is #P-hard.*

Proof. Since $\mathcal{F}$ does not satisfy condition (T), $\mathcal{F}$ contains a signature $f \notin \mathscr{A}$. Suppose that f has arity $2n$. We prove this theorem by induction on $2n$.

If $2n = 2, 4$ or 6, then by Corollary 6.10 and its remark, $\text{Holant}^b(\mathcal{F})$ is #P-hard.

Inductively assume for some $2k \geqslant 6$, $\text{Holant}^b(\mathcal{F})$ is #P-hard when $2n \leqslant 2k$. We consider the case that $2n = 2k + 2 \geqslant 8$. First, suppose that f is reducible. If it is a tensor product of two signatures of odd arity, then we can realize a signature of odd arity by factorization. We get #P-hardness by Theorem 2.25. Otherwise, it is a tensor product of two signatures of even arity that are not both in $\mathscr{A}$ since $f \notin \mathscr{A}$. Then, we can realize a non-affine signature of arity $2m \leqslant 2k$ by factorization. By our induction hypothesis, we get #P-hardness. Thus, we may assume that f is irreducible. If f has no parity, then we get #P-hardness by Lemma 7.1. Thus, we may also assume that f has parity. Then by Lemma 7.18, $\text{Holant}^b(\mathcal{F})$ is #P-hard, or we can realize a non-affine signature of arity $2m \leqslant 2k$. By our induction hypothesis, we get #P-hardness. □

Since $\mathcal{B}$ is realizable from f_6 and $\{f_6\} \cup \mathcal{F}$ is non-$\mathcal{B}$ hard for any real-valued $\mathcal{F}$ that does not satisfy condition (T), we have the following result.

Lemma 7.20. $\text{Holant}^b(f_6, \mathcal{F})$ *is #P-hard.*

Combining Theorem 6.5 and Lemma 7.20, we have the following result. This concludes Sections 6 and 7, and we are done with the arity 6 case.

Lemma 7.21. *If $\widehat{\mathcal{F}}$ contains a signature $\widehat{f}$ of arity 6 and $\widehat{f} \notin \widehat{\mathcal{O}}^{\otimes}$, then* $\text{Holant}(\neq_2 | \widehat{\mathcal{F}})$ *is #P-hard.*

8 Final Obstacle: an 8-ary Signature with Strong Bell Property

We have seen some extraordinary properties of the signature f_8. Now, we formally analyze it. Remember that $f_8 = \chi_T$ where

$$\begin{aligned} T = &\mathscr{S}(f_8) = \{(x_1, x_2, \ldots, x_8) \in \mathbb{Z}_2^8 \mid \ x_1 + x_2 + x_3 + x_4 = 0,\ x_5 + x_6 + x_7 + x_8 = 0, \\ &\qquad\qquad x_1 + x_2 + x_5 + x_6 = 0,\ x_1 + x_3 + x_5 + x_7 = 0\}. \\ =&\{00000000, 00001111, 00110011, 00111100, 01010101, 01011010, 01100110, 01101001, \\ &\ 10010110, 10011001, 10100101, 10101010, 11000011, 11001100, 11110000, 11111111\}. \end{aligned} \tag{8.1}$$

One can see that $\mathscr{S}(f_8)$ has the following structure: the sums of the first four variables, and the last four variables are both even; the assignment of the first four variables are either identical to, or complement of the assignment of the last four variables. Another interesting description of $\mathscr{S}(f_8)$ is as follows: One can take 4 variables, called them y_1, y_2, y_3, y_4. Then on the support the remaining 4 variables are mod 2 sums of $\binom{4}{3}$ subsets of $\{y_1, y_2, y_3, y_4\}$, and y_1, y_2, y_3, y_4 are free variables. (However, the 4 variables (y_1, y_2, y_3, y_4) cannot be taken as (x_1, x_2, x_3, x_4) in the above description (8.1). But one *can* take $(y_1, y_2, y_3, y_4) = (x_1, x_2, x_3, x_5)$. More specifically, one can take any 3 variables x_i, x_j, x_k from $\{x_1, \ldots, x_8\}$ first as free variables, which excludes one unique other

x_ℓ from the remainder set $X' = \{x_1, \ldots, x_8\} \setminus \{x_i, x_j, x_k\}$, and *then* one can take any one variable $x_r \in X'$ as the 4th free variable. *Then* the remaining 4 variables are the mod 2 sums of $\binom{4}{3}$ subsets of the 4 free variables $\{x_i, x_j, x_k, x_r\}$, and in particular $x_\ell = x_i + x_j + x_k$, on $\mathscr{S}(f_8)$.) We give the following Figure 2 to visualize the signature matrix $M_{1234}(f_8)$. A block with orange color denotes an entry $+1$. Other blank blocks are zeros.

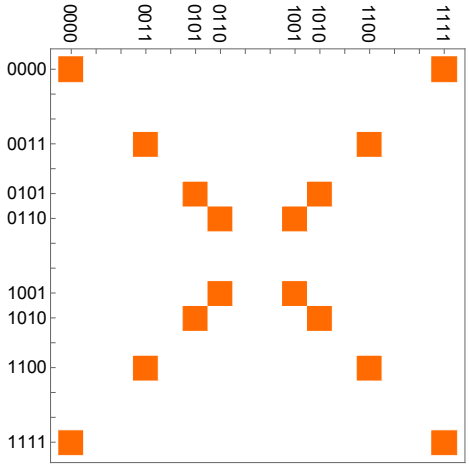


Figure 2: A visualization of f_8, which happens to be the same as $\widehat{f_8} = Z^{-1} f_8$

One can check that f_8 satisfies both 2ND-ORTH and $f_8 \in \int \mathcal{O}^{\otimes}$. Also, f_8 is unchanged under the holographic transformation by Z^{-1}, i.e., $\widehat{f_8} = Z^{-1} f_8 = f_8$.

8.1 The discovery of $\widehat{f_8}$

In this subsection, we show how this extraordinary signature $\widehat{f_8}$ was discovered. We use the notation $\widehat{f_8}$ since we consider the problem Holant($\neq_2$| $\widehat{\mathcal{F}}$) for complex-valued $\widehat{\mathcal{F}}$ satisfying ARS. We prove that if $\widehat{\mathcal{F}}$ contains an 8-ary signature $\widehat{f}$ where $\widehat{f} \notin \widehat{\mathcal{O}}^{\otimes}$, then Holant($\neq_2$| $\widehat{\mathcal{F}}$) is #P-hard or $\widehat{f_8}$ is realizable from $\widehat{f}$ (Theorem 8.5).

Remember that $\mathcal{D} = \{\neq_2\}$. Then $\mathcal{D}^{\otimes} = \{\lambda \cdot (\neq_2)^{\otimes n} \mid \lambda \in \mathbb{R}\setminus\{0\}, n \geqslant 1\}$ is the set of tensor products of binary disequalities $\neq_2$ up to a nonzero real scalar. If for all pairs of indices $\{i, j\}$, $\widehat{\partial}_{ij} \widehat{f} \in \mathcal{D}^{\otimes}$, then we say $\widehat{f} \in \widehat{\int} \mathcal{D}^{\otimes}$. Clearly, if $\widehat{f} \in \mathcal{D}^{\otimes}$ and $\widehat{f}$ has arity greater than 2, then $\widehat{f} \in \widehat{\int} \mathcal{D}^{\otimes}$. We first show the following result for signatures of arity at least 8.

Lemma 8.1. *Let $\widehat{f} \notin \widehat{\mathcal{O}}^{\otimes}$ be a signature of arity $2n \geqslant 8$ in $\widehat{\mathcal{F}}$. Then,*

- Holant($\neq_2$| $\widehat{\mathcal{F}}$) *is #P-hard, or*
- *there is a signature $\widehat{g} \notin \widehat{\mathcal{O}}^{\otimes}$ of arity $2k \leqslant 2n-2$ that is realizable from $\widehat{f}$, or*
- *there is an irreducible signature $\widehat{f^*} \in \widehat{\int} \mathcal{D}^{\otimes}$ of arity $2n$ that is realizable from $\widehat{f}$.*

Proof. Since $\widehat{f} \notin \widehat{\mathcal{O}}^{\otimes}$, $\widehat{f} \not\equiv 0$. Again, we may assume that $\widehat{f}$ is irreducible. Otherwise, by factorization, we can realize a nonzero signature of odd arity and we get #P-hardness by Theorem 2.25, or we can realize a signature of lower even arity that is not in $\widehat{\mathcal{O}}^{\otimes}$ and we are done. Under the assumption that $\widehat{f}$ is irreducible, we may further assume that $\widehat{f}$ satisfies 2ND-ORTH by Lemma 4.4. Consider signatures $\widehat{\partial}_{ij} \widehat{f}$ for all pairs of indices $\{i, j\}$. If there exists a pair $\{i, j\}$ such that $\widehat{\partial}_{ij} \widehat{f} \notin \widehat{\mathcal{O}}^{\otimes}$, then let $\widehat{g} = \widehat{\partial}_{ij} \widehat{f}$, and we are done. Thus, we may also assume that $\widehat{f} \in \widehat{\int} \widehat{\mathcal{O}}^{\otimes}$.

If for all pairs of indices $\{i, j\}$, we have $\widehat{\partial}_{ij}\widehat{f} \equiv 0$. Then, by Lemma 2.10, $\widehat{f}(\alpha) = 0$ for all α with $\mathrm{wt}(\alpha) \neq 0$ or $2n$. Since $f \not\equiv 0$ and by ARS, $|\widehat{f}(\vec{0}^{2n})| = |\widehat{f}(\vec{1}^{2n})| \neq 0$. Clearly, such a signature does not satisfy 2ND-ORTH. Contradiction. Thus, without loss of generality, we assume that $\widehat{\partial}_{12}\widehat{f} \not\equiv 0$. Since $\widehat{\partial}_{12}\widehat{f} \in \widehat{\mathcal{O}}^{\otimes}$, without loss of generality, we may assume that in the UPF of $\widehat{\partial}_{12}\widehat{f}$, variables x_3 and x_4 appear in one binary signature $b_1(x_3, x_4)$, x_5 and x_6 appear in one binary signature $b_2(x_5, x_6)$ and so on. Thus, we have

$$\widehat{\partial}_{12}\widehat{f} = \widehat{b_1}(x_3, x_4) \otimes \widehat{b_2}(x_5, x_6) \otimes \widehat{b_3}(x_7, x_8) \otimes \ldots \otimes \widehat{b_{n-1}}(x_{2n-1}, x_{2n}).$$

By Lemma 2.7, all these binary signatures $\widehat{b_1}$, $\widehat{b_2}$, $\ldots$, $\widehat{b_{n-1}}$ are realizable from f by factorization. Note that for nonzero binary signatures $\widehat{b_i}(x_{2i+1}, x_{2i+2})$ $(1 \leqslant i \leqslant n-1)$, if we connect the variable x_{2i+1} of two copies of $\widehat{b_i}(x_{2i+1}, x_{2i+2})$ using $\neq_2$ (mating two binary signatures), then we get $\neq_2$ up to a scalar. We consider the following gadget construction on $\widehat{f}$. Recall that in the setting of Holant$(\neq| \widehat{\mathcal{F}})$, variables are connected using $\neq_2$. For $1 \leqslant i \leqslant n-1$, by a slight abuse of names of variables, we connect the variable x_{2i+1} of $\widehat{f}$ with the variable x_{2i+1} of $\widehat{b_i}(x_{2i+1}, x_{2i+2})$ using $\neq_2$. We get a signature $\widehat{f'}$ of arity $2n$. (Note that, as a complexity reduction using factorization (Lemma 2.7), we can only apply it a constant number of times. However, the arity $2n$ of $\widehat{f}$ is considered a constant, as $\widehat{f} \in \widehat{\mathcal{F}}$, which is independent of the input size of a signature grid to the problem Holant$(\neq_2| \widehat{\mathcal{F}})$.) We denote this gadget construction by G_1 and we write $\widehat{f'}$ as $G_1 \circ \widehat{f}$. G_1 is constructed by extending variables of $\widehat{f}$ using binary signatures realized from $\widehat{\partial}_{12}\widehat{f}$. It does not change the irreducibility of $\widehat{f}$. Thus, $\widehat{f'}$ is irreducible since $\widehat{f}$ is irreducible. Similarly, we may assume that $\widehat{f'} \in \widehat{\int}\widehat{\mathcal{O}}^{\otimes}$. Otherwise, we are done.

Consider the signature $\widehat{\partial}_{12}\widehat{f'}$. Since the above gadget construction G_1 does not touch variables x_1 and x_2 of f, G_1 commutes with the merging gadget $\widehat{\partial}_{12}$. (Succinctly, the commutativity can be expressed as $\widehat{\partial}_{12}\widehat{f'} = \widehat{\partial}_{12}(G_1 \circ \widehat{f}) = G_1 \circ \widehat{\partial}_{12}\widehat{f}$.) Thus, $\widehat{\partial}_{12}\widehat{f'}$ can be realized by performing the gadget construction G_1 on $\widehat{\partial}_{12}\widehat{f}$, which connects each binary signature $\widehat{b_i}(x_{2i+1}, x_{2i+2})$ in the UPF of $\widehat{\partial}_{12}\widehat{f}$ with another copy of $\widehat{b_i}(x_{2i+1}, x_{2i+2})$ (in the mating fashion). Thus, each binary signature $\widehat{b_i}$ in $\widehat{\partial}_{12}\widehat{f}$ is changed to $\neq_2$ up to a nonzero scalar after this gadget construction G_1. After normalization and renaming variables, we have

$$\widehat{\partial}_{12}\widehat{f'} = (\neq_2)(x_3, x_4) \otimes (\neq_2)(x_5, x_6) \otimes (\neq_2)(x_7, x_8) \otimes \ldots \otimes (\neq_2)(x_{2n-1}, x_{2n}). \tag{8.2}$$

Thus, $\widehat{\partial}_{12}\widehat{f'} \in \mathcal{D}^{\otimes}$. Moreover, for all pairs of indices $\{i, j\}$ disjoint with $\{1, 2\}$, we have

$$\widehat{\partial}_{(ij)(12)}\widehat{f'} \in \mathcal{D}^{\otimes}, \quad \text{and hence} \quad \widehat{\partial}_{(ij)(12)}\widehat{f'} \not\equiv 0. \tag{8.3}$$

A fortiori, for all pairs of indices $\{i, j\}$ disjoint with $\{1, 2\}$, $\widehat{\partial}_{ij}\widehat{f'} \not\equiv 0$.

Now, we show that we can realize an irreducible signature $\widehat{f^*}$ of arity $2n$ from $\widehat{f'}$ such that $\widehat{f^*} \in \widehat{\int}\mathcal{D}^{\otimes}$. We first prove the following claim.

> **Claim.** *Let $\widehat{h} \in \widehat{\int}\widehat{\mathcal{O}}^{\otimes}$ be a signature of arity $2n \geqslant 8$. If $\widehat{\partial}_{ij}\widehat{h} \in \mathcal{D}^{\otimes}$ for all $\{i, j\}$ disjoint with $\{1, 2\}$, then $\widehat{h} \in \widehat{\int}\mathcal{D}^{\otimes}$.*

Clearly, we only need to show that $\widehat{\partial}_{1k}\widehat{h} \in \mathcal{D}^{\otimes}$ for all $2 \leqslant k \leqslant 2n$. Then, by symmetry we also have $\widehat{\partial}_{2k}\widehat{h} \in \mathcal{D}^{\otimes}$ for $k = 1$ and all $3 \leqslant k \leqslant 2n$. This will prove $\widehat{h} \in \widehat{\int}\mathcal{D}^{\otimes}$. Consider $\widehat{\partial}_{1k}\widehat{h}$ for an

arbitrary $2 \leqslant k \leqslant 2n$. Since for all $\{i,j\}$ disjoint with $\{1,2\}$, we have $\widehat{\partial}_{ij}\widehat{h} \in \mathcal{D}^{\otimes}$, a fortiori for all $\{i,j\}$ disjoint with $\{1,2\} \cup \{k\}$,

$$\widehat{\partial}_{(1k)(ij)}\widehat{h} \in \mathcal{D}^{\otimes}. \tag{8.4}$$

Since $\widehat{h}$ has arity $2n \geqslant 8$, we can pick a pair of indices $\{i,j\}$ disjoint with $\{1,2\} \cup \{k\}$. Since $\widehat{\partial}_{(1k)(ij)}\widehat{h} \in \mathcal{D}^{\otimes}$, which is nonzero, a fortiori we have $\widehat{\partial}_{1k}\widehat{h} \not\equiv 0$. So we may consider the UPF of $\widehat{\partial}_{1k}\widehat{h}$, which is known to be in $\widehat{\mathcal{O}}^{\otimes}$. For a contradiction, suppose that there is a binary signature $\widehat{b_1}$ (as a factor of $\widehat{\partial}_{1k}\widehat{h}$) such that $\widehat{b_1}$ is not an associate of $\neq_2$. Among the two variables in the scope of $\widehat{b_1}$, at least one is not x_2. We pick such a variable x_s where $x_s \neq x_2$. Then, we consider another binary signature $\widehat{b_2}$ in the UPF of $\widehat{\partial}_{1k}\widehat{h}$.

- If $\widehat{b_2} = \lambda \cdot \neq_2$, for some nonzero scalar λ, then we pick a variable x_t in the scope of $\widehat{b_2}$ that is not x_2. Consider $\widehat{\partial}_{(st)(1k)}\widehat{h}$. When merging variables x_s and x_t of $\widehat{\partial}_{1k}\widehat{h}$, we connect the variable x_s of $\widehat{b_1}$ with the variable x_t of $\lambda \cdot \neq_2$, and the resulting binary signature is just $\lambda \cdot \widehat{b_1}$, which is not an associate of $\neq_2$. Thus, we have $\widehat{\partial}_{(st)(1k)}\widehat{h} \notin \mathcal{D}^{\otimes}$.
- Otherwise, $\widehat{b_2}$ is not an associate of $\neq_2$. Since $\widehat{\partial}_{1k}\widehat{h}$ has arity $2n-2 \geqslant 6$, we can find another binary signature $\widehat{b_3}$ in the UPF of $\widehat{\partial}_{1k}\widehat{h}$. We pick a variable x_t in the scope of $\widehat{b_3}$ that is not x_2. Consider $\widehat{\partial}_{(st)(1k)}\widehat{h}$. Now, when merging variables x_s and x_t of $\widehat{\partial}_{1k}\widehat{h}$, the binary signature $\widehat{b_2}$ is untouched. Thus, we have $\widehat{b_2} \mid \widehat{\partial}_{(st)(1k)}\widehat{h}$, which implies that $\widehat{\partial}_{(st)(1k)}\widehat{h} \notin \mathcal{D}^{\otimes}$.

Note that in both cases, $\{s,t\} \cap (\{1,2\} \cup \{k\}) = \emptyset$. Therefore the two cases above both contradict (8.4) by picking $\{i,j\} = \{s,t\}$. Thus, $\widehat{\partial}_{1k}\widehat{h} \in \mathcal{D}^{\otimes}$ for all $2 \leqslant k \leqslant 2n$. Then similarly, we can show that $\widehat{\partial}_{2k}\widehat{h} \in \mathcal{D}^{\otimes}$ for all $3 \leqslant k \leqslant 2n$. This finishes the proof of our Claim.

Remember that $\widehat{\partial}_{ij}\widehat{f'} \not\equiv 0$ for all $\{i,j\}$ disjoint with $\{1,2\}$. We consider the UPF of $\widehat{\partial}_{ij}\widehat{f'}$. Since $\widehat{f'} \in \widehat{\int}\widehat{\mathcal{O}}^{\otimes}$, there are two cases depending on whether variables x_1 and x_2 appear in one binary signature or two distinct binary signatures.

Case 1. For every $\{i,j\}$ disjoint with $\{1,2\}$, in the UPF of $\widehat{\partial}_{ij}\widehat{f'}$, x_1 and x_2 appear in one nonzero binary signature $\widehat{b_{ij}}(x_1,x_2) \in \widehat{\mathcal{O}}$. In other words, for every $\{i,j\}$ disjoint with $\{1,2\}$,

$$\widehat{\partial}_{ij}\widehat{f'} = \widehat{b_{ij}}(x_1,x_2) \otimes \widehat{g_{ij}}, \quad \text{for some } \widehat{g_{ij}} \not\equiv 0.$$

(These factors $\widehat{b_{ij}}$ and $\widehat{g_{ij}}$ are nonzero since $\widehat{\partial}_{ij}\widehat{f'} \not\equiv 0$.) Then, $\widehat{g_{ij}} \sim \widehat{\partial}_{(12)(ij)}\widehat{f'}$, and by (8.3), we have $\widehat{g_{ij}} \in \mathcal{D}^{\otimes}$. Also for $\{k,\ell\}$ disjoint with both $\{i,j\}$ and $\{1,2\}$, $\widehat{\partial}_{(k\ell)(ij)}\widehat{f'} \not\equiv 0$ since $\widehat{\partial}_{(12)(k\ell)(ij)}\widehat{f'} = \widehat{\partial}_{(ij)(k\ell)(12)}\widehat{f'} \not\equiv 0$.

We first show that for any two pairs $\{i,j\} \neq \{k,\ell\}$ that are both disjoint with $\{1,2\}$, $\widehat{b_{ij}}(x_1,x_2) \sim \widehat{b_{k\ell}}(x_1,x_2)$. If $\{i,j\}$ is disjoint with $\{k,\ell\}$, then $\widehat{b_{ij}}(x_1,x_2) \mid \widehat{\partial}_{(k\ell)(ij)}\widehat{f'}$ and $\widehat{b_{k\ell}}(x_1,x_2) \mid \widehat{\partial}_{(ij)(k\ell)}\widehat{f'}$. Since $\widehat{\partial}_{(k\ell)(ij)}\widehat{f'} = \widehat{\partial}_{(ij)(k\ell)}\widehat{f'} \not\equiv 0$, by Lemma 2.5, we have $\widehat{b_{ij}}(x_1,x_2) \sim \widehat{b_{k\ell}}(x_1,x_2)$. Otherwise, $\{i,j\}$ and $\{k,\ell\}$ are not disjoint. Since $\widehat{f'}$ has arity $\geqslant 8$, we can find another pair of indices $\{s,t\}$ such that it is disjoint with $\{1,2\} \cup \{i,j\} \cup \{k,\ell\}$. Then, by the above argument, we have $\widehat{b_{ij}}(x_1,x_2) \sim \widehat{b_{st}}(x_1,x_2)$, and $\widehat{b_{st}}(x_1,x_2) \sim \widehat{b_{k\ell}}(x_1,x_2)$. Thus, $\widehat{b_{ij}}(x_1,x_2) \sim \widehat{b_{k\ell}}(x_1,x_2)$. We can use a binary signature $\widehat{b}(x_1,x_2)$ to denote these binary signature $\widehat{b_{ij}}(x_1,x_2)$ for all $\{i,j\}$ disjoint with $\{1,2\}$. Then, $\widehat{b}(x_1,x_2) \mid \widehat{\partial}_{ij}\widehat{f'}$ for all $\{i,j\}$ disjoint with $\{1,2\}$. Also, $\widehat{b}(x_1,x_2)$ is realizable from $\widehat{f'}$ by merging and factorization.

Then, we consider the following gadget construction G_2 on $\widehat{f'}$. By a slight abuse of variable names, we connect the variable x_1 of $\widehat{f'}$ with the variable x_1 of $\widehat{b}(x_1,x_2)$ and we get a signature

$\widehat{f^*}$. Clearly, G_2 is constructed by extending variables of $\widehat{f'}$. It does not change the irreducibility of $\widehat{f'}$. Thus, $\widehat{f^*}$ is irreducible. Again, we may assume that $\widehat{f^*} \in \widehat{\int}\widehat{\mathcal{O}^{\otimes}}$. Consider $\widehat{\partial}_{ij}\widehat{f^*}$ for all $\{i,j\}$ disjoint with $\{1,2\}$. Since the above gadget construction G_2 only touches the variable x_1 of f', it commutes with the merging operation $\widehat{\partial}_{ij}$. Thus, $\widehat{\partial}_{ij}\widehat{f^*}$ can be realized by performing the gadget construction G_2 on $\widehat{\partial}_{ij}\widehat{f'}$, i.e., connecting the binary signature $\widehat{b}(x_1,x_2)$ in the UPF of $\widehat{\partial}_{ij}\widehat{f'}$ with itself (in the mating fashion), which changes $\widehat{b}(x_1,x_2)$ to $\neq_2$ up to some nonzero scalar λ_{ij}. Thus, for all $\{i,j\}$ disjoint with $\{1,2\}$, after renaming variables, we have

$$\widehat{\partial}_{ij}\widehat{f^*} = \lambda_{ij} \cdot (\neq_2)(x_1,x_2) \otimes \widehat{g_{ij}} \in \mathcal{D}^{\otimes}.$$

Thus, $\widehat{\partial}_{ij}\widehat{f^*} \in \mathcal{D}^{\otimes}$ for all $\{i,j\}$ disjoint with $\{1,2\}$. By our Claim, $\widehat{f^*} \in \widehat{\int}\mathcal{D}^{\otimes}$. We are done with Case 1.

Case 2. There is a pair of indices $\{i,j\}$ disjoint with $\{1,2\}$ such that x_1 and x_2 appear in two distinct nonzero binary signatures $\widehat{b'_1}(x_1,x_u)$ and $\widehat{b'_2}(x_2,x_v)$ in the UPF of $\widehat{\partial}_{ij}\widehat{f'}$. In other words, there exits $\{i,j\}$ such that

$$\widehat{\partial}_{ij}\widehat{f'} = \widehat{b'_1}(x_1,x_u) \otimes \widehat{b'_2}(x_2,x_v) \otimes \widehat{h_{ij}}, \text{ for some } \widehat{h_{ij}} \not\equiv 0. \tag{8.5}$$

Since $\widehat{h_{ij}} \mid \widehat{\partial}_{(12)(ij)}\widehat{f'}$ and $\widehat{\partial}_{(12)(ij)}\widehat{f'} \in \mathcal{D}^{\otimes}$, we have $\widehat{h_{ij}} \in \mathcal{D}^{\otimes}$. Also, after merging variables x_1 and x_2 (using $\neq_2$) in $\widehat{\partial}_{ij}\widehat{f'}$, variables x_u and x_v form a binary disequality up to a nonzero scalar (this binary signature on x_u and x_v must be a binary disequality because we already know $\widehat{\partial}_{(12)(ij)}\widehat{f'} \in \mathcal{D}^{\otimes}$). In other words, by connecting the variable x_1 of $\widehat{b'_1}(x_1,x_u)$ and the variable x_2 of $\widehat{b'_2}(x_2,x_v)$ (using $\neq_2$), we get $\lambda \cdot \neq_2 (x_u,x_v)$ for some $\lambda \neq 0$. By Lemma 2.13, we have $\widehat{b'_1} \sim \widehat{b'_2}$. Also, connecting the variable x_u of $\widehat{b'_1}$ and the variable x_v of $\widehat{b'_2}$ (using $\neq_2$) will give the binary signature $\lambda \cdot \neq_2 (x_1,x_2)$ as well.

We consider the following gadget construction G_3 on $\widehat{f'}$. By a slight abuse of variable names, we connect variables x_1 and x_2 of $\widehat{f'}$ with the variable x_1 of $\widehat{b'_1}$ and x_2 of $\widehat{b'_2}$ using $\neq_2$ respectively. We get a signature $\widehat{f^*}$. Again, $\widehat{f^*}$ is irreducible since the gadget construction G_3 does not change the irreducibility of $\widehat{f'}$. Also, we may assume that $\widehat{f^*} \in \widehat{\int}\widehat{\mathcal{O}^{\otimes}}$. Otherwise, we are done. Consider $\widehat{\partial}_{ij}\widehat{f^*}$. Similarly, by the commutitivity of the gadget construction G_3 and the merging gadget $\widehat{\partial}_{ij}$, $\widehat{\partial}_{ij}\widehat{f^*}$ can be realized by connecting variables x_1 and x_2 of $\widehat{\partial}_{ij}\widehat{f'}$ with the variable x_1 of $\widehat{b'_1}$ and the variable x_2 of $\widehat{b'_2}$ respectively. After renaming variables, we have

$$\widehat{\partial}_{ij}\widehat{f^*} = \lambda_{ij} \cdot (\neq_2)(x_1,x_u) \otimes (\neq_2)(x_2,x_v) \otimes \widehat{h}_{ij} \in \mathcal{D}^{\otimes}. \tag{8.6}$$

We now show that $\widehat{\partial}_{12}\widehat{f^*} \in \mathcal{D}^{\otimes}$. Note that it is realized in the following way; we first connect variables x_1 and x_2 of $\widehat{f'}$ with the variable x_1 of $\widehat{b'_1}(x_1,x_u)$ and the variable x_2 of $\widehat{b'_2}(x_2,x_v)$ respectively (using $\neq_2$) to get $\widehat{f^*}$, and then after renaming variables x_u and x_v to x_1 and x_2 respectively, we merge them using $\neq_2$ (see Figure 3(a)). By associativity of gadget constructions, we can change the order; we first connect the variable x_u of $\widehat{b'_1}(x_1,x_u)$ with the variable x_v of $\widehat{b'_2}(x_2,x_v)$ (using $\neq_2$), and then we use the resulting binary signature to connect variables x_1 and x_2 of $\widehat{f'}$ (edges are connected using $\neq_2$). Note that connecting x_u of $\widehat{b'_1}(x_1,x_u)$ with x_v of $\widehat{b'_2}(x_2,x_v)$ gives $\lambda \cdot \neq_2$ up to a nonzero scalar λ, and $\lambda \cdot \neq_2$ is unchanged by extending both of its two variables with $\neq_2$

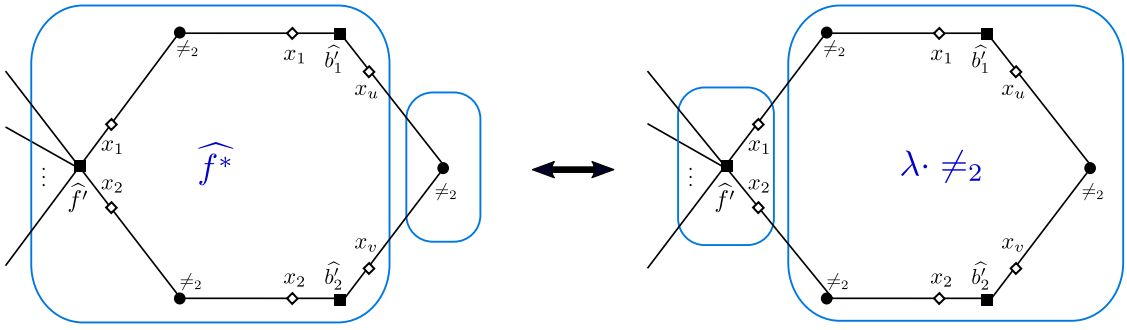


Figure 3: Gadget constructions of $\widehat{\partial}_{12}\widehat{f^*}$ and $\widehat{\partial}_{12}\widehat{f'}$

(see Figure 3(b)). Thus, $\widehat{\partial}_{12}\widehat{f^*}$ is actually realized by merging x_1 and x_2 of $\widehat{f'}$ (using $\neq_2$) up to a nonzero scalar. Thus, we have $\widehat{\partial}_{12}\widehat{f^*} \sim \widehat{\partial}_{12}\widehat{f'}$, and hence $\widehat{\partial}_{12}\widehat{f^*} \in \mathcal{D}^{\otimes}$, by the form (8.2) of $\widehat{\partial}_{12}\widehat{f'}$.

Then, we show that $\widehat{\partial}_{st}\widehat{f^*} \in \mathcal{D}^{\otimes}$ for all pairs of indices $\{s,t\}$ disjoint with $\{1,2,i,j\}$ and $\{s,t\} \neq \{u,v\}$ where u and v are named in (8.6). Clearly, $\widehat{\partial}_{st}\widehat{f^*} \not\equiv 0$ since $\widehat{\partial}_{(st)(12)}\widehat{f^*} \in \mathcal{D}^{\otimes}$. We first show that in the UPF of $\widehat{\partial}_{st}\widehat{f^*}$, x_1 and x_2 appear in two distinct nonzero binary signatures. Otherwise, for a contradiction, suppose that there is a nonzero binary signature $\widehat{b^*}(x_1,x_2)$ such that $\widehat{b^*}(x_1,x_2) \mid \widehat{\partial}_{st}\widehat{f^*}$. Then, $\widehat{b^*}(x_1,x_2) \mid \widehat{\partial}_{(ij)(st)}\widehat{f^*} = \widehat{\partial}_{(st)(ij)}\widehat{f^*} \not\equiv 0$. By the form (8.6) of $\widehat{\partial}_{ij}\widehat{f^*}$, the only way that x_1 and x_2 can form a nonzero binary signature in $\widehat{\partial}_{(st)(ij)}\widehat{f^*}$ is that the merging gadget is actually merging x_u and x_v. Thus, $\{s,t\} = \{u,v\}$. Contradiction. Therefore, for some i' and j', we have

$$\widehat{\partial}_{st}\widehat{f^*} = \widehat{b^*_{st1}}(x_1,x_{i'}) \otimes \widehat{b^*_{st2}}(x_2,x_{j'}) \otimes \widehat{h_{st}}, \tag{8.7}$$

for some $\widehat{b^*_{st1}}(x_1,x_{i'}), \widehat{b^*_{st2}}(x_2,x_{j'}), \widehat{h_{st}} \not\equiv 0$ since $\widehat{\partial}_{st}\widehat{f^*} \not\equiv 0$. Since $\widehat{h_{st}} \mid \widehat{\partial}_{(12)(st)}\widehat{f^*}$ and $\widehat{\partial}_{(12)(st)}\widehat{f^*} \in \mathcal{D}^{\otimes}$, we have $\widehat{h_{st}} \in \mathcal{D}^{\otimes}$. Also, by Lemma 2.13, $\widehat{b^*_{st1}} \sim \widehat{b^*_{st2}}$. For a contradiction, suppose that $\widehat{\partial}_{st}\widehat{f^*} \notin \mathcal{D}^{\otimes}$, then $\widehat{b^*_{st1}}(x_1,x_{i'}) \not\sim (\neq_2)$, and $\widehat{b^*_{st2}}(x_2,x_{j'}) \not\sim (\neq_2)$. Consider the signature $\widehat{\partial}_{(st)(ij)}\widehat{f^*}$. Since $\{s,t\} \neq \{u,v\}$, by the form (8.6) of $\widehat{\partial}_{ij}\widehat{f^*}$, x_1 and x_2 appear in two binary signatures in the UPF of $\widehat{\partial}_{(st)(ij)}\widehat{f^*}$. Remember that $\widehat{\partial}_{(st)(ij)}\widehat{f^*} = \widehat{\partial}_{(ij)(st)}\widehat{f^*}$. By the form (8.7) of $\widehat{\partial}_{st}\widehat{f^*}$, if $\{i',j'\} = \{i,j\}$, then, after merging x_i and x_j of $\widehat{\partial}_{st}\widehat{f^*}$, x_1 and x_2 will form a new binary signature in $\widehat{\partial}_{(ij)(st)}\widehat{f^*}$. Contradiction. Thus, $\{i',j'\} \neq \{i,j\}$. Then, when merging x_i and x_j of $\widehat{\partial}_{st}\widehat{f^*}$, among $\widehat{b^*_{st1}}(x_1,x_{i'})$ and $\widehat{b^*_{st2}}(x_2,x_{j'})$, at least one binary signature is untouched. Thus, $\widehat{\partial}_{(ij)(st)}\widehat{f^*}$ has a factor that is not an associate of $\neq_2$. A contradiction with $\widehat{\partial}_{(ij)(st)}\widehat{f^*} \in \mathcal{D}^{\otimes}$, which is a consequence of (8.6). Thus, $\widehat{\partial}_{st}\widehat{f^*} \in \mathcal{D}^{\otimes}$.

Then, we show that $\widehat{\partial}_{uv}\widehat{f^*} \in \mathcal{D}^{\otimes}$. Recall the form (8.6) of $\widehat{\partial}_{ij}\widehat{f^*}$. Clearly, $\{u,v\}$ is disjoint with $\{1,2,i,j\}$. Also, $\widehat{\partial}_{uv}\widehat{f^*} \not\equiv 0$ since $\widehat{\partial}_{(ij)(uv)}\widehat{f^*} \in \mathcal{D}^{\otimes}$. Consider the UPF of $\widehat{\partial}_{uv}\widehat{f^*}$.

- If x_1 and x_2 appear in one nonzero binary signature $\widehat{b^*_{uv}}(x_1,x_2)$, then

  $$\widehat{\partial}_{uv}\widehat{f^*} = \widehat{b^*_{uv}}(x_1,x_2) \otimes \widehat{g_{uv}} \qquad \text{for some } \widehat{g_{uv}} \not\equiv 0.$$

  Then, we have $\widehat{g_{uv}} \sim \widehat{\partial}_{(12)(uv)}\widehat{f^*} \in \mathcal{D}^{\otimes}$ since $\widehat{\partial}_{12}\widehat{f^*} \in \mathcal{D}^{\otimes}$. Also, since $\widehat{b^*_{uv}}(x_1,x_2) \mid \widehat{\partial}_{(ij)(uv)}\widehat{f^*} \in \mathcal{D}^{\otimes}$, we have $\widehat{b^*_{uv}}(x_1,x_2) \in \mathcal{D}^{\otimes}$. Hence, $\widehat{\partial}_{uv}\widehat{f^*} \in \mathcal{D}^{\otimes}$.

- If x_1 and x_2 appear in two distinct nonzero binary signatures $\widehat{b^*_{uv1}}(x_1, x_{i'})$ and $\widehat{b^*_{uv2}}(x_2, x_{j'})$, then
$$\widehat{\partial}_{uv}\widehat{f^*} = \widehat{b^*_{uv1}}(x_1, x_{i'}) \otimes \widehat{b^*_{uv2}}(x_2, x_{j'}) \otimes \widehat{h_{uv}} \qquad \text{for some } \widehat{h_{uv}} \not\equiv 0.$$
Then, we have $\widehat{h_{uv}} \in \mathcal{D}^{\otimes}$ since $\widehat{\partial}_{(12)(uv)}\widehat{f^*} \in \mathcal{D}^{\otimes}$. By the form (8.6) of $\widehat{\partial}_{ij}\widehat{f^*}$, after merging variables x_u and x_v of $\widehat{\partial}_{ij}\widehat{f^*}$, variables x_1 and x_2 form a binary $\neq_2$ in $\widehat{\partial}_{(uv)(ij)}\widehat{f^*} = \widehat{\partial}_{(ij)(uv)}\widehat{f^*}$. On the other hand, by the form of $\widehat{\partial}_{uv}\widehat{f^*}$, the only way that x_1 and x_2 form a binary after merging two variables in $\widehat{\partial}_{uv}\widehat{f^*}$ is to merge $x_{i'}$ and $x_{j'}$. Thus, we have $\{i', j'\} = \{i, j\}$. Since $\widehat{f^*}$ has arity $2n \geqslant 8$, we can find another pair of indices $\{s, t\}$ disjoint with $\{1, 2, i, j, u, v\}$. When merging variables x_s and x_t in $\widehat{\partial}_{uv}\widehat{f^*}$, binary signatures $\widehat{b^*_{uv1}}(x_1, x_{i'})$ and $\widehat{b^*_{uv2}}(x_2, x_{j'})$ are untouched. Thus, we have $\widehat{b^*_{uv1}}(x_1, x_{i'}) \otimes \widehat{b^*_{uv2}}(x_2, x_{j'}) \mid \widehat{\partial}_{(st)(uv)}\widehat{f^*}$. As showed above, we have $\widehat{\partial}_{st}\widehat{f^*} \in \mathcal{D}^{\otimes}$ and then $\widehat{\partial}_{(st)(uv)}\widehat{f^*} \in \mathcal{D}^{\otimes}$. Thus, $\widehat{b^*_{uv1}}(x_1, x_{i'}) \otimes \widehat{b^*_{uv2}}(x_2, x_{j'}) \in \mathcal{D}^{\otimes}$ and then $\widehat{\partial}_{uv}\widehat{f^*} \in \mathcal{D}^{\otimes}$.

So far, we have shown that $\widehat{\partial}_{12}\widehat{f^*} \in \mathcal{D}^{\otimes}$, $\widehat{\partial}_{ij}\widehat{f^*} \in \mathcal{D}^{\otimes}$ and $\widehat{\partial}_{st}\widehat{f^*} \in \mathcal{D}^{\otimes}$ for all $\{s, t\}$ disjoint with $\{1, 2, i, j\}$. If we can further show that $\widehat{\partial}_{ik}\widehat{f^*} \in \mathcal{D}^{\otimes}$ for all $k \neq 1, 2, i, j$, and then symmetrically $\widehat{\partial}_{jk}\widehat{f^*} \in \mathcal{D}^{\otimes}$ for all $k \neq 1, 2, i, j$, then $\widehat{\partial}_{st}\widehat{f^*} \in \mathcal{D}^{\otimes}$ for all $\{s, t\}$ disjoint with $\{1, 2\}$. Thus, by our Claim, $\widehat{f^*} \in \widehat{\int}\mathcal{D}^{\otimes}$. This will finish the proof of Case 2.

Now we prove $\widehat{\partial}_{ik}\widehat{f^*} \in \mathcal{D}^{\otimes}$ for all $k \neq 1, 2, i, j$. Since $\widehat{\partial}_{(ik)(12)}\widehat{f^*} \in \mathcal{D}^{\otimes}$, we have $\widehat{\partial}_{ik}\widehat{f^*} \not\equiv 0$. So we can consider the UPF of $\widehat{\partial}_{ik}\widehat{f^*}$.

- If x_1 and x_2 appear in one nonzero binary signature, then
$$\widehat{\partial}_{ik}\widehat{f^*} = \widehat{b^*_{ik}}(x_1, x_2) \otimes \widehat{g_{ik}} \qquad \text{for some } \widehat{g_{ik}} \in \mathcal{D}^{\otimes}.$$
Here, $\widehat{g_{ik}} \in \mathcal{D}^{\otimes}$ since $\widehat{\partial}_{(ik)(12)}\widehat{f^*} \in \mathcal{D}^{\otimes}$. Since $\widehat{f^*}$ has arity $2n \geqslant 8$, we can pick a pair of indices $\{s, t\}$ disjoint with $\{1, 2, i, j, k\}$, and merge variables x_s and x_t of $\widehat{\partial}_{ik}\widehat{f^*}$. Then, $\widehat{b^*_{ik}}(x_1, x_2) \mid \widehat{\partial}_{(st)(ik)}\widehat{f^*}$. Since $\widehat{\partial}_{st}\widehat{f^*} \in \mathcal{D}^{\otimes}$, $\widehat{\partial}_{(st)(ik)}\widehat{f^*} = \widehat{\partial}_{(ik)(st)}\widehat{f^*} \in \mathcal{D}^{\otimes}$. Thus, $\widehat{b^*_{ik}}(x_1, x_2) \in \mathcal{D}^{\otimes}$ and then $\widehat{\partial}_{ik}\widehat{f^*} \in \mathcal{D}^{\otimes}$.
- If x_1 and x_2 appear in two nonzero distinct binary signatures, then
$$\widehat{\partial}_{ik}\widehat{f^*} = \widehat{b^*_{ik1}}(x_1, x_p) \otimes \widehat{b^*_{ik2}}(x_2, x_q) \otimes \widehat{h_{ik}} \qquad \text{for some } \widehat{h_{ik}} \in \mathcal{D}^{\otimes}.$$
Again, here $\widehat{h_{ik}} \in \mathcal{D}^{\otimes}$ since $\widehat{\partial}_{(ik)(12)}\widehat{f^*} \in \mathcal{D}^{\otimes}$. By connecting variables x_1 and x_2 of $\widehat{\partial}_{ik}\widehat{f^*}$, x_p and x_q will form a binary disequality up to a nonzero scalar (this binary signature is disequality because we know that $\widehat{\partial}_{(ik)(12)}\widehat{f^*} \in \mathcal{D}^{\otimes}$). By Lemma 2.13, as the type of binary signatures, $\widehat{b^*_{ik1}} \sim \widehat{b^*_{ik2}}$. Between x_p and x_q, at least one of them is not x_j; suppose that it is x_p. We pick a variable x_r in the scope of $\widehat{h_{ik}}$ that is also not x_j (such a variable x_r exists as $2n \geqslant 8$). Then, by merging x_p and x_r of $\widehat{\partial}_{ik}\widehat{f^*}$, the binary signature $\widehat{b^*_{ik2}}(x_2, x_q)$ is untouched. Since $\{p, r\}$ is disjoint with $\{1, 2, i, j\}$, we have $\widehat{b^*_{ik2}}(x_2, x_q) \mid \widehat{\partial}_{(ik)(pr)}\widehat{f^*} \in \mathcal{D}^{\otimes}$. Thus, we have $\widehat{b^*_{ik2}}(x_2, x_q) \in \mathcal{D}^{\otimes}$ and so does $\widehat{b^*_{ik1}}(x_1, x_p)$, since we have shown that they are associates as the type of binary signatures. Thus, $\widehat{\partial}_{ik}\widehat{f^*} \in \mathcal{D}^{\otimes}$.

As remarked earlier, by symmetry, we also have $\widehat{\partial}_{jk}\widehat{f^*} \in \mathcal{D}^{\otimes}$ for all $k \neq 1, 2, i, j$. Thus, we are done with Case 2.

Thus, an irreducible signature $\widehat{f^*} \in \widehat{\int}\mathcal{D}^{\otimes}$ of arity $2n$ is realized from $\widehat{f}$. □

Remark: Since $\widehat{f^*}$ is realized from $\widehat{f}$ by gadget construction, $\widehat{f^*}$ satisfies ARS as $\widehat{f}$ does.

We first give a condition (Lemma 8.3) in which we can quite straightforwardly get the #P-hardness of Holant$(\neq_2 \mid \widehat{f}, \widehat{\mathcal{F}})$ by 2ND-ORTH given $\widehat{f} \in \widehat{\int}\mathcal{D}^{\otimes}$ is an irreducible 8-ary signature.

Lemma 8.2. *Let $\widehat{f} = a(1,0)^{\otimes 2n} + \bar{a}(0,1)^{\otimes 2n} + (\neq_2)(x_i, x_j) \otimes \widehat{g_{\mathrm{h}}}$ be an irreducible $2n$-ary signature, where $2n \geqslant 4$ and $\widehat{g_{\mathrm{h}}}$ is a nonzero* EO *signature (i.e., with half-weighted support) of arity $2n-2$. Then, $\widehat{f}$ does not satisfy* 2ND-ORTH.

Proof. By renaming variables, without loss of generality, we may assume that $\{i, j\} = \{1, 2\}$.

For any input $00\beta \neq \vec{0}^{2n}$ of $\widehat{f}$, we have $\widehat{f}(00\beta) = (\neq_2)(0,0) \cdot \widehat{g_{\mathrm{h}}}(\beta) = 0$. Thus,

$$|\widehat{f_{12}^{00}}|^2 = \sum_{\beta \in \mathbb{Z}_2^{2n-2}} |\widehat{f}(00\beta)|^2 = |\widehat{f}(\vec{0}^{2n})|^2.$$

On the other hand, since both $(\neq_2)(x_1, x_2)$ and $\widehat{g_{\mathrm{h}}}$ are nonzero EO signatures, $(\neq_2)(x_1, x_2) \otimes \widehat{g_{\mathrm{h}}}$ is a nonzero EO signature. Then, we can pick an input $01\gamma \in \mathbb{Z}_2^{2n}$ with $\mathrm{wt}(01\gamma) = n$ such that $\widehat{f}(01\gamma) = (\neq_2)(0,1) \cdot \widehat{g_{\mathrm{h}}}(\gamma) \neq 0$. Since $\gamma \in \mathbb{Z}_2^{2n-2}$, and $\mathrm{wt}(\gamma) = n - 1 \geqslant 1$, there exists a bit γ_i in γ such that $\gamma_i = 0$. Without loss of generality, we may assume that $01\gamma = 010\gamma'$. Then,

$$|\widehat{f_{13}^{00}}|^2 \geqslant |\widehat{f}(\vec{0}^{2n})|^2 + |\widehat{f}(010\gamma')|^2 > |\widehat{f}(\vec{0}^{2n})|^2 = |\widehat{f_{12}^{00}}|^2.$$

Note that the constant λ for the norm squares must be the same for all index pairs $\{i, j\} \subseteq [2n]$ in order to satisfy 2ND-ORTH in Definition 4.1. Thus, $\widehat{f}$ does not satisfy 2ND-ORTH. □

Lemma 8.3. *Let $\widehat{f} \in \widehat{\int}\mathcal{D}^{\otimes}$ be an irreducible 8-ary signature in $\widehat{\mathcal{F}}$. If there exists a binary disequality $(\neq_2)(x_i, x_j)$ and two pairs of indices $\{u, v\}$ and $\{s, t\}$ where $\{u, v\} \cap \{s, t\} \neq \emptyset$ such that $(\neq_2)(x_i, x_j) \mid \widehat{\partial}_{uv}\widehat{f}$ and $(\neq_2)(x_i, x_j) \mid \widehat{\partial}_{st}\widehat{f}$, then* Holant$(\neq_2 \mid \widehat{\mathcal{F}})$ *is #P-hard.*

Proof. For all pairs of indices $\{i, j\}$, since $\widehat{\partial}_{ij}\widehat{f} \in \mathcal{D}^{\otimes}$, $\mathscr{S}(\widehat{\partial}_{ij}\widehat{f})$ is on half-weight. By Lemma 2.10, we have $\widehat{f}(\alpha) = 0$ for all $\mathrm{wt}(\alpha) \neq 0, 4, 8$. Suppose that $\widehat{f}(\vec{0}^8) = a$ and by ARS $\widehat{f}(\vec{1}^8) = \bar{a}$. We can write $\widehat{f}$ in the following form

$$\widehat{f} = a(1,0)^{\otimes 8} + \bar{a}(0,1)^{\otimes 8} + \widehat{f_{\mathrm{h}}},$$

where $\widehat{f_{\mathrm{h}}}$ is an EO signature of arity 8.

Clearly, $\widehat{\partial}_{ij}\widehat{f} = \widehat{\partial}_{ij}\widehat{f_{\mathrm{h}}}$ for all $\{i, j\}$. Then, $\widehat{f_{\mathrm{h}}} \in \widehat{\int}\mathcal{D}^{\otimes}$ since $\widehat{f} \in \widehat{\int}\mathcal{D}^{\otimes}$. In addition, since there exists a binary disequality $(\neq_2)(x_i, x_j)$ and two pairs of indices $\{u, v\}$ and $\{s, t\}$ where $\{u, v\} \cap \{s, t\} \neq \emptyset$ such that $(\neq_2)(x_i, x_j) \mid \widehat{\partial}_{uv}\widehat{f_{\mathrm{h}}}, \widehat{\partial}_{st}\widehat{f_{\mathrm{h}}}$, by Lemma 2.11, $\widehat{f_{\mathrm{h}}} \in \mathcal{D}^{\otimes}$ and $(\neq_2)(x_i, x_j) \mid \widehat{f_{\mathrm{h}}}$. Thus,

$$\widehat{f} = a(1,0)^{\otimes 8} + \bar{a}(0,1)^{\otimes 8} + (\neq_2)(x_i, x_j) \otimes \widehat{g_{\mathrm{h}}},$$

where $\widehat{g_{\mathrm{h}}} \in \mathcal{D}^{\otimes}$ is a nonzero EO signature or arity 6 since $\widehat{f_{\mathrm{h}}} \in \mathcal{D}^{\otimes}$. By Lemma 8.2, $\widehat{f}$ does not satisfy 2ND-ORTH. Thus, Holant$(\neq_2 \mid \widehat{\mathcal{F}})$ is #P-hard by Lemma 4.4. □

For signatures in $\mathcal{D}^{\otimes}$, we give the following property. Now we adopt the following notation for brevity. We use (i, j) to denote the binary disequality $(\neq_2)(x_i, x_j)$ on variables x_i and x_j.

Lemma 8.4. *Let $\widehat{f} \in \mathcal{D}^{\otimes}$ be a signature of arity at least 6. If there exist $\{u, v\} \neq \{s, t\}$ such that $(i, j) \mid \widehat{\partial}_{uv}\widehat{f}$ and $(i, j) \mid \widehat{\partial}_{st}\widehat{f}$, then $(i, j) \mid \widehat{f}$.*

Proof. For a contradiction, suppose that $(i,j) \nmid \widehat{f}$. Thus x_i and x_j appear in two separate disequalities in the UPF of $\widehat{f}$. Since $\widehat{f} \in \mathcal{D}^{\otimes}$, there exists $\{\ell,k\}$ such that $(i,\ell) \otimes (j,k) \mid \widehat{f}$. By merging two variables of $\widehat{f}$, the only way to make x_i and x_j to form a binary disequality is by merging x_ℓ and x_k. By the hypothesis of the lemma, $\{\ell,k\} = \{u,v\} = \{s,t\}$. Contradiction. □

Theorem 8.5. *Let $\widehat{f} \notin \widehat{\mathcal{O}^{\otimes}}$ be a signature of arity 8 in $\widehat{\mathcal{F}}$. Then*

- $\mathrm{Holant}(\neq_2 \mid \widehat{\mathcal{F}})$ *is #P-hard, or*
- *there exists some $\widehat{Q} \in \widehat{\mathbf{O}_2}$ such that* $\mathrm{Holant}(\neq_2 \mid \widehat{f}_8, \widehat{Q}\widehat{\mathcal{F}}) \leqslant_T \mathrm{Holant}(\neq_2 \mid \widehat{\mathcal{F}})$.

Proof. By Lemma 8.1, we may assume that an irreducible signature $\widehat{f^*}$ of arity 8 where $\widehat{f^*} \in \widehat{\int}\mathcal{D}^{\otimes}$ is realizable from $\widehat{f}$, and $\widehat{f^*}$ also satisfies ARS. Otherwise, $\mathrm{Holant}(\neq_2 \mid \widehat{\mathcal{F}})$ is #P-hard or we can realize a signature $\widehat{g} \notin \widehat{\mathcal{O}^{\otimes}}$ of arity 2, 4 or 6. Then, by Lemmas 5.1, 5.2 and 7.21, we get #P-hardness. We will show that $\widehat{f}_8$ is realizable from $\widehat{f^*}$, or otherwise we get #P-hardness. For brevity of notation, we rename $\widehat{f^*}$ by $\widehat{f}$. We first show that after renaming variables by applying a suitable permutation to $\{1,2,\ldots,8\}$, for all $\{i,j\} \subseteq \{1,2,3,4\}$, $(\ell,k) \mid \widehat{\partial}_{ij}\widehat{f}$ where $\{\ell,k\} = \{1,2,3,4\}\backslash\{i,j\}$. Furthermore, we show that either $\mathrm{Holant}(\neq_2 \mid \widehat{\mathcal{F}})$ is #P-hard, or

$$(5,6) \mid \widehat{\partial}_{12}\widehat{f}, \quad (5,7) \mid \widehat{\partial}_{13}\widehat{f}, \quad (6,7) \mid \widehat{\partial}_{23}\widehat{f}, \text{ and } \quad (1,2) \mid \widehat{\partial}_{56}\widehat{f} \text{ or } (1,3) \mid \widehat{\partial}_{56}\widehat{f}. \tag{8.8}$$

Consider $\widehat{\partial}_{12}\widehat{f}$. Since $\widehat{f} \in \widehat{\int}\mathcal{D}^{\otimes}$, $\widehat{\partial}_{12}\widehat{f} \in \mathcal{D}^{\otimes}$. By renaming variables, without loss of generality, we may assume that

$$\widehat{\partial}_{12}\widehat{f} = \lambda_{12} \cdot (3,4) \otimes (5,6) \otimes (7,8), \tag{8.9}$$

for some $\lambda_{12} \in \mathbb{R} \setminus \{0\}$. Then, consider $\widehat{\partial}_{34}\widehat{f}$. $\widehat{\partial}_{56}\widehat{f}$, and $\widehat{\partial}_{78}\widehat{f}$. There are two cases.

- Case 1. $(1,2) \mid \widehat{\partial}_{34}\widehat{f}, \widehat{\partial}_{56}\widehat{f}$ and $\widehat{\partial}_{78}\widehat{f}$. Then we can write $\widehat{\partial}_{56}\widehat{f} = (1,2) \otimes \widehat{h}$ for some $\widehat{h} \in \mathcal{D}^{\otimes}$. Clearly, $\widehat{h} \sim \widehat{\partial}_{(12)(56)}\widehat{f}$. By the form (8.9) and commutativity, $\widehat{\partial}_{(12)(56)}\widehat{f} \sim (3,4) \otimes (7,8)$. Thus, $\widehat{h} \sim (3,4) \otimes (7,8)$. Then, for some $\lambda_{56} \in \mathbb{R} \setminus \{0\}$,

  $$\widehat{\partial}_{56}\widehat{f} = \lambda_{56} \cdot (1,2) \otimes (3,4) \otimes (7,8). \tag{8.10}$$

  Similarly, we have

  $$\widehat{\partial}_{78}\widehat{f} = \lambda_{78} \cdot (1,2) \otimes (3,4) \otimes (5,6),$$

  and

  $$\widehat{\partial}_{34}\widehat{f} = \lambda_{34} \cdot (1,2) \otimes (5,6) \otimes (7,8),$$

  for some $\lambda_{78}, \lambda_{34} \in \mathbb{R} \setminus \{0\}$.
  Let $\widehat{g} = (1,2) \otimes (3,4)$. Let $\{i,j\} \subseteq \{1,2,3,4\}$ and $\{\ell,k\} = \{1,2,3,4\}\backslash\{i,j\}$. If we merge variables x_i and x_j of $\widehat{g}$, i.e., if we form $\widehat{\partial}_{ij}\widehat{g}$, then clearly variables x_ℓ and x_k will form a disequality. Thus, for all $\{i,j\} \subseteq \{1,2,3,4\}$, $(\ell,k) \mid \widehat{\partial}_{ij}\widehat{g}$. Then, $(\ell,k) \mid \widehat{\partial}_{ij}\widehat{g} \otimes (7,8) \sim \widehat{\partial}_{(ij)(56)}\widehat{f}$ by (8.10), and similarly $(\ell,k) \mid \widehat{\partial}_{ij}\widehat{g} \otimes (5,6) \sim \widehat{\partial}_{(ij)(78)}\widehat{f}$. By Lemma 8.4, $(\ell,k) \mid \widehat{\partial}_{ij}\widehat{f}$.
- Case 2. Among $\widehat{\partial}_{34}\widehat{f}$, $\widehat{\partial}_{56}\widehat{f}$, and $\widehat{\partial}_{78}\widehat{f}$, there is at least one signature that is not divisible by $(1,2)$. Without loss of generality, suppose that $(1,2) \nmid \widehat{\partial}_{56}\widehat{f}$. Since $\widehat{\partial}_{56}\widehat{f} \in \mathcal{D}^{\otimes}$, there exists $\{u,v\}$ disjoint from $\{1,2,5,6\}$ such that $(1,u) \otimes (2,v) \mid \widehat{\partial}_{56}\widehat{f}$. Then, by merging variables x_1 and x_2 of $\widehat{\partial}_{56}\widehat{f}$, we have $(u,v) \mid \widehat{\partial}_{(12)(56)}\widehat{f}$; comparing it to $\widehat{\partial}_{(56)(12)}\widehat{f}$ using the form of (8.9) and by unique factorization we get $\{u,v\} = \{3,4\}$ or $\{7,8\}$. Without loss of generality (i.e., this is still within the freedom of our naming variables subject to the choices made so far),

we may assume that $\{u,v\} = \{3,4\}$ and furthermore, $u = 3$ and $v = 4$. Then, for some $\lambda'_{56} \in \mathbb{R} \setminus \{0\}$,

$$\widehat{\partial}_{56}\widehat{f} = \lambda'_{56} \cdot (1,3) \otimes (2,4) \otimes (7,8). \tag{8.11}$$

Then, consider $\widehat{\partial}_{78}\widehat{f}$. We show that $(5,6) \mid \widehat{\partial}_{78}\widehat{f}$. Otherwise, there exists $\{s,t\}$ disjoint from $\{5,6,7,8\}$ such that $(5,s) \otimes (6,t) \mid \widehat{\partial}_{78}\widehat{f}$. By merging two variables of $\widehat{\partial}_{78}\widehat{f}$, the only way to make x_5 and x_6 form a binary disequality is to merge x_s and x_t. By the form (8.9), $(5,6) \mid \widehat{\partial}_{(12)(78)}\widehat{f}$. Thus, $\{s,t\} = \{1,2\}$. From $(5,s) \otimes (6,t) \mid \widehat{\partial}_{78}\widehat{f}$, and $\{s,t\} = \{1,2\}$ we know that x_1 and x_2 will form a binary disequality in $\widehat{\partial}_{(56)(78)}\widehat{f}$. Thus, $(1,2) \mid \widehat{\partial}_{(56)(78)}\widehat{f}$. However, by (8.11) $\widehat{\partial}_{(56)(78)}\widehat{f} \sim (1,3) \otimes (2,4)$. This is a contradiction to UPF. Thus, $\widehat{\partial}_{78}\widehat{f} = (5,6) \otimes \widehat{g'}$ and $\widehat{g'} \sim \widehat{\partial}_{(56)(78)}\widehat{f} \sim (1,3) \otimes (2,4)$. Then, for some $\lambda'_{78} \in \mathbb{R} \setminus \{0\}$,

$$\widehat{\partial}_{78}\widehat{f} = \lambda'_{78} \cdot (1,3) \otimes (2,4) \otimes (5,6). \tag{8.12}$$

Let $\{i,j\} \subseteq \{1,2,3,4\}$ and $\{\ell,k\} = \{1,2,3,4\} \setminus \{i,j\}$. If we merge variables x_i and x_j of $\widehat{g'}$, which is an associate of $(1,3) \otimes (2,4)$, then clearly variables x_ℓ and x_k will form a disequality. Thus, for all $\{i,j\} \subseteq \{1,2,3,4\}$, $(\ell,k) \sim \widehat{\partial}_{ij}\widehat{g'}$. Then, $(\ell,k) \mid \widehat{\partial}_{ij}\widehat{g'} \otimes (7,8) \sim \widehat{\partial}_{(ij)(56)}\widehat{f}$ (by (8.11)) and $(\ell,k) \mid \widehat{\partial}_{ij}\widehat{g'} \otimes (5,6) \sim \widehat{\partial}_{(ij)(78)}\widehat{f}$ (by (8.12)). By Lemma 8.4, $(\ell,k) \mid \widehat{\partial}_{ij}\widehat{f}$.

Thus, in both cases, we have $(\ell,k) \mid \widehat{\partial}_{ij}\widehat{f}$ where $\{i,j\} \sqcup \{\ell,k\} = \{1,2,3,4\}$ is an arbitrary disjoint union of two pairs. Now, we show that in both cases, (with possibly switching the names x_7 and x_8, which we are still free to do), we can have

$$(5,6) \mid \widehat{\partial}_{12}\widehat{f}, \quad (5,7) \mid \partial_{13}\widehat{f}, \quad (6,7) \mid \widehat{\partial}_{23}\widehat{f}. \tag{8.13}$$

Clearly, by the form (8.9), we have $(5,6) \mid \widehat{\partial}_{12}\widehat{f}$. Consider $\partial_{13}\widehat{f}$. We already know that $(2,4) \mid \partial_{13}\widehat{f}$ (in both cases). If $(5,6) \mid \widehat{\partial}_{13}\widehat{f}$, then since $(5,6) \mid \widehat{\partial}_{12}\widehat{f}$ and $\{1,2\} \cap \{1,3\} \neq \emptyset$, by Lemma 8.3, Holant$(\neq_2 \mid \widehat{\mathcal{F}})$ is #P-hard. Thus, $(5,7) \mid \widehat{\partial}_{13}\widehat{f}$ or $(5,8) \mid \widehat{\partial}_{13}\widehat{f}$. By renaming variables x_7 and x_8, we may assume that in both cases

$$\widehat{\partial}_{13}\widehat{f} = (2,4) \otimes (5,7) \otimes (6,8). \tag{8.14}$$

This renaming will not change any of the above forms of $\widehat{\partial_{ij}}\widehat{f}$. Consider $\widehat{\partial}_{23}\widehat{f}$. We already have $(1,4) \mid \widehat{\partial}_{23}\widehat{f}$. We know $\widehat{\partial}_{23}\widehat{f} \in \mathcal{D}^{\otimes}$, and so in its UPF, $(6,r) \mid \widehat{\partial}_{23}\widehat{f}$, for some $r \in [8] \setminus \{1,2,3,4,6\}$. If $(5,6) \mid \widehat{\partial}_{23}\widehat{f}$, then since $(5,6) \mid \widehat{\partial}_{12}\widehat{f}$ and $\{1,2\} \cap \{2,3\} \neq \emptyset$, by Lemma 8.3, we get #P-hardness. If $(6,8) \mid \widehat{\partial}_{23}\widehat{f}$, then since $(6,8) \mid \widehat{\partial}_{13}\widehat{f}$ by (8.14) and $\{1,3\} \cap \{2,3\} \neq \emptyset$, again by Lemma 8.3, we get #P-hardness. Thus, we may assume that $r = 7$ and $(6,7) \mid \widehat{\partial}_{23}\widehat{f}$. Therefore, we have established (8.13) in both cases. Furthermore, in Case 1, we have $(1,2) \mid \widehat{\partial}_{56}\widehat{f}$ by form (8.10), and in Case 2, we have $(1,3) \mid \widehat{\partial}_{56}\widehat{f}$ by form (8.11).

Now, we show that for any $\alpha \in \mathbb{Z}_2^4$ with $\mathrm{wt}(\alpha) = 1$, $\widehat{f}^{\alpha}_{1234} \equiv 0$. Since $(3,4) \mid \widehat{\partial}_{12}\widehat{f}$, $(\widehat{\partial}_{12}\widehat{f})^{00}_{34} \equiv 0$. Since $\{1,2\}$ is disjoint with $\{3,4\}$,

$$(\widehat{\partial}_{12}\widehat{f})^{00}_{34} = \widehat{\partial}_{12}(\widehat{f}^{00}_{34}) = \widehat{f}^{0100}_{1234} + \widehat{f}^{1000}_{1234} \equiv 0. \tag{8.15}$$

Since $(1,4) \mid \widehat{\partial}_{23}\widehat{f}$,

$$(\widehat{\partial}_{23}\widehat{f})^{00}_{14} = \widehat{\partial}_{23}(\widehat{f}^{00}_{14}) = \widehat{f}^{0010}_{1234} + \widehat{f}^{0100}_{1234} \equiv 0. \tag{8.16}$$

Since $(1,3) \mid \widehat{\partial}_{24}\widehat{f}$,

$$(\widehat{\partial}_{13}\widehat{f})^{00}_{24} = \widehat{\partial}_{13}(\widehat{f}^{00}_{24}) = \widehat{f}^{0010}_{1234} + \widehat{f}^{1000}_{1234} \equiv 0. \tag{8.17}$$

Comparing (8.15), (8.16) and (8.17), we have

$$\widehat{f}^{1000}_{1234} = \widehat{f}^{0100}_{1234} = \widehat{f}^{0010}_{1234} \equiv 0.$$

Since $(2,3) \mid \widehat{\partial}_{14}\widehat{f}$,

$$(\widehat{\partial}_{14}\widehat{f})^{00}_{23} = \widehat{\partial}_{14}(\widehat{f}^{00}_{23}) = \widehat{f}^{0001}_{1234} + \widehat{f}^{1000}_{1234} \equiv 0.$$

Plug in $\widehat{f}^{1000}_{1234} \equiv 0$, we have $\widehat{f}^{0001}_{1234} \equiv 0$. Thus for any $\alpha \in \mathbb{Z}_2^4$ with $\mathrm{wt}(\alpha) = 1$, we have $\widehat{f}^{\alpha}_{1234} \equiv 0$.

Also, for $\alpha \in \mathbb{Z}_2^4$ with $\mathrm{wt}(\alpha) = 3$ and any $\beta \in \mathbb{Z}_2^4$, by ARS we have,

$$\widehat{f}^{\alpha}_{1234}(\beta) = \overline{\widehat{f}^{\overline{\alpha}}_{1234}(\overline{\beta})} = 0$$

since $\mathrm{wt}(\overline{\alpha}) = 1$. Thus, for any $\alpha \in \mathbb{Z}_2^4$ with $\mathrm{wt}(\alpha) = 3$, we also have $\widehat{f}^{\alpha}_{1234} \equiv 0$.

Let $\alpha \in \mathbb{Z}_2^4$ be an assignment of the first four variables of f, and $\beta \in \mathbb{Z}_2^4$ be an assignment of the last four variables of f. Thus, for any $\alpha, \beta \in \mathbb{Z}_2^4$, $\widehat{f}(\alpha\beta) = 0$ if $\mathrm{wt}(\alpha) = 1$ or 3. Also, since $\widehat{f} \in \widehat{\int}\mathcal{D}^{\otimes}$, by Lemma 2.10, $\widehat{f}(\alpha\beta) = 0$ if $\mathrm{wt}(\alpha) + \mathrm{wt}(\beta) \neq 0, 4$ and 8. Then, we show that for any $\alpha\beta \in \mathscr{S}(\widehat{f})$,

$$|\widehat{f}(\alpha\beta)| = |\widehat{f}(\overline{\alpha}\beta)| = |\widehat{f}(\alpha\overline{\beta})| = |\widehat{f}(\overline{\alpha}\overline{\beta})|.$$

By ARS, $|\widehat{f}(\alpha\beta)| = |\widehat{f}(\overline{\alpha}\overline{\beta})|$ and $|\widehat{f}(\overline{\alpha}\beta)| = |\widehat{f}(\alpha\overline{\beta})|$. So, we only need to show that

$$|\widehat{f}(\alpha\beta)| = |\widehat{f}(\alpha\overline{\beta})|. \tag{8.18}$$

Pick an arbitrary $\{i,j\} \subseteq \{1,2,3,4\}$ and an arbitrary $\{u,v\} \subseteq \{5,6,7,8\}$. Let $\{\ell,k\} = \{1,2,3,4\}\backslash\{i,j\}$ and $\{s,t\} = \{5,6,7,8\}\backslash\{u,v\}$. Since $\widehat{f}$ satisfies 2ND-ORTH, by equation (4.6), we have $|\widehat{\mathbf{f}}^{0000}_{ijuv}|^2 = |\widehat{\mathbf{f}}^{0011}_{ijuv}|^2$. Since $\widehat{f}(\alpha\beta) = 0$ if $\mathrm{wt}(\alpha) = 1$ or 3, or $\mathrm{wt}(\alpha) + \mathrm{wt}(\beta) \neq 0, 4$ and 8, we get the equation,

$$|\widehat{f}^{00000000}_{ij\ell kuvst}|^2 + |\widehat{f}^{00110011}_{ij\ell kuvst}|^2 = |\widehat{f}^{00111100}_{ij\ell kuvst}|^2 + |\widehat{f}^{00001111}_{ij\ell kuvst}|^2. \tag{8.19}$$

Note that for $|\widehat{\mathbf{f}}^{0000}_{ijuv}|^2$, since we set $x_ix_j = 00$, the only possible nonzero terms are for $x_\ell x_k = 00$ or 11; furthermore, as we also set $x_ux_v = 00$, then $x_sx_t = 00$ if $x_\ell x_k = 00$, and $x_sx_t = 11$ if $x_\ell x_k = 11$. The situation is similar for $|\widehat{\mathbf{f}}^{0011}_{ijuv}|^2$.

Also, by considering $|\widehat{\mathbf{f}}^{0000}_{ijst}|^2 = |\widehat{\mathbf{f}}^{0011}_{ijst}|^2$, we have

$$|\widehat{f}^{00000000}_{ij\ell kuvst}|^2 + |\widehat{f}^{00111100}_{ij\ell kuvst}|^2 = |\widehat{f}^{00110011}_{ij\ell kuvst}|^2 + |\widehat{f}^{00001111}_{ij\ell kuvst}|^2. \tag{8.20}$$

Comparing equations (8.19) and (8.20), we have

$$|\widehat{f}^{00000000}_{ij\ell kuvst}|^2 = |\widehat{f}^{00001111}_{ij\ell kuvst}|^2, \quad \text{and} \quad |\widehat{f}^{00110011}_{ij\ell kuvst}|^2 = |\widehat{f}^{00111100}_{ij\ell kuvst}|^2.$$

Also, by ARS,

$$|\widehat{f}^{11111111}_{ij\ell kuvst}|^2 = |\widehat{f}^{11110000}_{ij\ell kuvst}|^2.$$

As (i,j,k,ℓ) is an arbitrary permutation of $(1,2,3,4)$ and (u,v,s,t) is an arbitrary permutation of $(5,6,7,8)$, and $\widehat{f}(\alpha\beta)$ vanishes if $\mathrm{wt}(\alpha) + \mathrm{wt}(\beta) \neq 0, 4$ and 8, the above have established (8.18) for any $\alpha, \beta \in \mathbb{Z}_2^4$. Hence, for all $\alpha, \beta \in \mathbb{Z}_2^4$,

$$|\widehat{f}(\alpha\beta)| = |\widehat{f}(\overline{\alpha}\beta)| = |\widehat{f}(\alpha\overline{\beta})| = |\widehat{f}(\overline{\alpha}\overline{\beta})|.$$

$x_1x_2x_3x_4$ \ $x_5x_6x_7x_8$	$\alpha_1 = 0110$ (Col 1)	$\alpha_2 = 1010$ (Col 2)	$\alpha_3 = 1100$ (Col 3)
$\alpha_1 = 0110$ (Row 1)	$m_{11} = \widehat{f}^{01100110}$	$m_{12} = \widehat{f}^{01101010}$	$m_{13} = \widehat{f}^{01101100}$
$\alpha_2 = 1010$ (Row 2)	$m_{21} = \widehat{f}^{10100110}$	$m_{22} = \widehat{f}^{10101010}$	$m_{23} = \widehat{f}^{10101100}$
$\alpha_3 = 1100$ (Row 3)	$m_{31} = \widehat{f}^{11000110}$	$m_{32} = \widehat{f}^{11001010}$	$m_{33} = \widehat{f}^{11001100}$

Table 6: Representative entries of $\widehat{f}$ in terms of norms

Note that $\widehat{f}$ has at most $4 + \binom{4}{2} \cdot \binom{4}{2} = 40$ many possibly non-zero entries. In terms of norms, these 40 entries can be represented by $\widehat{f}^{\vec{0}^8}$ and the following 9 entries in Table 6. In other words, for every $\alpha\beta \in \mathbb{Z}_2^8$ where $\mathrm{wt}(\alpha) \equiv \mathrm{wt}(\beta) \equiv 0 \pmod 2$ and $\mathrm{wt}(\alpha) + \mathrm{wt}(\beta) \equiv 0 \pmod 4$, exactly one entry among $\widehat{f}(\alpha\beta)$, $\widehat{f}(\overline{\alpha}\beta)$, $\widehat{f}(\alpha\overline{\beta})$ and $\widehat{f}(\overline{\alpha}\overline{\beta})$ appears in Table 6. We also view these 9 entries in Table 6 as a 3-by-3 matrix denoted by $M = (m_{ij})_{i,j=1}^3$.

Let $\widehat{f}^{\vec{0}^8} = a$. First we show that

$$|m_{i,1}|^2 + |m_{i,2}|^2 + |m_{i,3}|^2 = |a|^2, \quad \text{for } i = 1, 2, 3. \tag{8.21}$$

and

$$|m_{1,j}|^2 + |m_{2,j}|^2 + |m_{3,j}|^2 = |a|^2, \quad \text{for } j = 1, 2, 3. \tag{8.22}$$

Let (i, j, k) be an arbitrary permutation of $(1, 2, 3)$. Again, by equation (4.6), $|\widehat{\mathbf{f}}^{0110}_{ijk8}|^2 = |\widehat{\mathbf{f}}^{0000}_{ijk8}|^2$. Then, we have

$$|\widehat{f}^{01100110}_{ijk45678}|^2 + |\widehat{f}^{01101010}_{ijk45678}|^2 + |\widehat{f}^{01101100}_{ijk45678}|^2 = |\widehat{f}^{00000000}_{ijk45678}|^2 = |a|^2.$$

By taking $(i, j, k) = (1, 2, 3), (2, 1, 3)$ and $(3, 1, 2)$, we get equations (8.21) for $i = 1, 2, 3$ respectively. Similarly, by considering $|\widehat{\mathbf{f}}^{0110}_{4ijk}|^2 = |\widehat{\mathbf{f}}^{0000}_{4ijk}|^2$ where (i, j, k) is an arbitrary permutation of $(5, 6, 7)$, we get equations (8.22).

Also, since $(5, 6) \mid \widehat{\partial}_{12}\widehat{f}$, we have $\widehat{\partial}_{12}\widehat{f}(x_3, \ldots, x_8) = 0$ if $x_5 = x_6$. Notice that

$$m_{13} + m_{23} = \widehat{f}^{01101100} + \widehat{f}^{10101100}$$

is an entry of $\widehat{\partial}_{12}\widehat{f}$ on the input 101100. Thus, $m_{13} + m_{23} = 0$. Also, since $(5, 7) \mid \widehat{\partial}_{13}\widehat{f}$, we have

$$m_{12} + m_{32} = 0.$$

Since $(6, 7) \mid \widehat{\partial}_{23}\widehat{f}$, we have

$$m_{21} + m_{31} = 0.$$

Let $x = |m_{13}| = |m_{23}|$, $y = |m_{12}| = |m_{32}|$, and $z = |m_{21}| = |m_{31}|$. Plug x, y, z into equations (8.21) and (8.22). We have

$$\begin{aligned} |m_{11}|^2 + y^2 + x^2 = &\quad |m_{11}|^2 + z^2 + z^2 \\ = z^2 + |m_{22}|^2 + x^2 = &\quad y^2 + |m_{22}|^2 + y^2 \\ = z^2 + y^2 + |m_{33}|^2 = &\quad x^2 + x^2 + |m_{33}|^2. \end{aligned}$$

Thus, $x = y = z$ and $|m_{11}| = |m_{22}| = |m_{33}|$. Consider

$$m_{11} + m_{21} = \widehat{f}^{01100110} + \widehat{f}^{10100110} \quad \text{and} \quad m_{12} + m_{22} = \widehat{f}^{01101010} + \widehat{f}^{10101010}.$$

They are entries of $\widehat{\partial}_{12}\widehat{f}$ on inputs 100110 and 101010. By form (8.9) of $\widehat{\partial}_{12}\widehat{f}$, we have

$$m_{11} + m_{21} = m_{12} + m_{22} \in \mathbb{R}\backslash\{0\}.$$

Remember that we also have $(1,2) \mid \widehat{\partial}_{56}\widehat{f}$ or $(1,3) \mid \widehat{\partial}_{56}\widehat{f}$.

We first consider the case that $(1,3) \mid \widehat{\partial}_{56}\widehat{f}$. Then

$$m_{21} + m_{22} = \widehat{f}^{10100110} + \widehat{f}^{10101010} = 0.$$

Thus,

$$m_{11} + m_{21} = m_{12} - m_{21} \in \mathbb{R}\backslash\{0\}.$$

Since $|m_{12}| = |m_{21}|$, $|m_{22}| = |m_{11}|$ and $m_{21} + m_{22} = 0$,

$$|m_{12}| = |m_{21}| = |m_{22}| = |m_{11}|.$$

Thus, $m_{11} = \overline{m_{21}}$ and $m_{12} = -\overline{m_{21}}$. Let $\mathfrak{Re}(x)$ the real part of a number x. Then,

$$\mathfrak{Re}(m_{11}) + \mathfrak{Re}(m_{21}) = 2\mathfrak{Re}(m_{21}) = \mathfrak{Re}(m_{12}) - \mathfrak{Re}(m_{21}) = -2\mathfrak{Re}(m_{21}).$$

Thus, $\mathfrak{Re}(m_{21}) = 0$. Then, $\mathfrak{Re}(m_{11}) = \mathfrak{Re}(m_{21}) = 0$. Thus, $m_{11} + m_{21} \notin \mathbb{R}\backslash\{0\}$ since $\mathfrak{Re}(m_{11} + m_{21}) = 0$. Contradiction.

Now, we consider the case that $(1,2) \mid \widehat{\partial}_{56}\widehat{f}$. Then

$$m_{31} + m_{32} = \widehat{f}^{11000110} + \widehat{f}^{11001010} = 0.$$

Since $m_{12} + m_{32} = 0$ and $m_{21} + m_{31} = 0$, we have $m_{12} = -m_{21}$. Thus, we have

$$m_{11} + m_{21} = m_{12} + m_{22} = m_{22} - m_{21} \in \mathbb{R}\backslash\{0\}.$$

Taking the imaginary part, $\mathfrak{Im}(m_{11}) + \mathfrak{Im}(m_{21}) = \mathfrak{Im}(m_{22}) - \mathfrak{Im}(m_{21}) = 0$. Adding the two, we get $\mathfrak{Im}(m_{11}) + \mathfrak{Im}(m_{22}) = 0$, and thus, $m_{11} + m_{22} \in \mathbb{R}$. Since $|m_{11}| = |m_{22}|$, $m_{11} = \overline{m_{22}}$. Then, $\mathfrak{Re}(m_{11}) = \mathfrak{Re}(m_{22})$. Also, since $m_{11} + m_{21} = m_{22} - m_{21} \in \mathbb{R}\backslash\{0\}$,

$$\mathfrak{Re}(m_{11}) + \mathfrak{Re}(m_{21}) = \mathfrak{Re}(m_{22}) - \mathfrak{Re}(m_{21}) = \mathfrak{Re}(m_{11}) - \mathfrak{Re}(m_{21}) \neq 0.$$

Thus, $\mathfrak{Re}(m_{21}) = 0$, and $\mathfrak{Re}(m_{11}) \neq 0$. Suppose that $m_{21} = d\mathfrak{i}$ for some $d \in \mathbb{R}$. Then there exists $c \in \mathbb{R}\backslash\{0\}$ such that $m_{11} = c - d\mathfrak{i}$ and then $m_{22} = c + d\mathfrak{i}$. Remember that $m_{21} + m_{31} = 0$. Thus, $m_{31} = -d\mathfrak{i}$. Consider

$$m_{11} + m_{31} = \widehat{f}^{01100110} + \widehat{f}^{11000110} = c - 2d\mathfrak{i}.$$

It is an entry of the signature $\widehat{\partial}_{13}\widehat{f}$. Since $\widehat{\partial}_{13}\widehat{f} \in \mathcal{D}^{\otimes}$, $c - 2d\mathfrak{i} \in \mathbb{R}$. Thus, $d = 0$. Then, $m_{21} = 0$ and $m_{11} \in \mathbb{R}$. Thus,

$$x = |m_{13}| = |m_{23}| = y = |m_{12}| = |m_{32}| = z = |m_{21}| = |m_{31}| = 0,$$

and

$$|m_{11}| = |m_{22}| = |m_{33}| = |a| = |\widehat{f}(\vec{0})|.$$

Since $\widehat{f} \not\equiv 0$, $a \neq 0$. Thus,

$$\mathscr{S}(\widehat{f}) = \{\delta\delta, \delta\overline{\delta}, \overline{\delta}\delta, \overline{\delta\delta} \in \mathbb{Z}_2^8 \mid \delta = 0000, \alpha_1, \alpha_2, \alpha_3\},$$

where $\alpha_1, \alpha_2, \alpha_3$ are named in Table 6. It is easy to see that $\mathscr{S}(\widehat{f}) = \mathscr{S}(\widehat{f_8})$. Since $m_{11} \in \mathbb{R}$, and $|m_{11}| = |a| \neq 0$, we can normalize it to 1. Since, $\widehat{\partial}_{12}\widehat{f} \in \mathcal{D}^{\otimes}$, we have

$$1 = \widehat{f}(\alpha_1\alpha_1) + \widehat{f}(\alpha_2\alpha_1) = \widehat{f}(\alpha_1\alpha_2) + \widehat{f}(\alpha_2\alpha_2) = \widehat{f}(\alpha_1\overline{\alpha_1}) + \widehat{f}(\alpha_2\overline{\alpha}_1) = \widehat{f}(\alpha_1\overline{\alpha_2}) + \widehat{f}(\alpha_2\overline{\alpha}_2).$$

Since, $\widehat{f}(\alpha_2\alpha_1) = \widehat{f}(\alpha_1\alpha_2) = \widehat{f}(\alpha_2\overline{\alpha}_1) = \widehat{f}(\alpha_1\overline{\alpha_2}) = 0$,

$$\widehat{f}(\alpha_1\alpha_1) = \widehat{f}(\alpha_2\alpha_2) = \widehat{f}(\alpha_1\overline{\alpha}_1) = \widehat{f}(\alpha_2\overline{\alpha_2}) = 1.$$

Similarly, since $\widehat{\partial}_{13}\widehat{f} \in \mathcal{D}^{\otimes}$,

$$\widehat{f}(\alpha_1\alpha_1) = \widehat{f}(\alpha_3\alpha_3) = \widehat{f}(\alpha_1\overline{\alpha}_1) = \widehat{f}(\alpha_3\overline{\alpha_3}) = 1.$$

By ARS, we have

$$1 = \overline{\widehat{f}(\alpha_1\alpha_1)} = \widehat{f}(\overline{\alpha_1\alpha_1}) = \widehat{f}(\overline{\alpha_1}\alpha_1) = \widehat{f}(\overline{\alpha_2\alpha_2}) = \widehat{f}(\overline{\alpha_2}\alpha_2) = \widehat{f}(\overline{\alpha_3\alpha_3}) = \widehat{f}(\overline{\alpha_3}\alpha_3).$$

Also, since $\widehat{\partial}_{15}\widehat{f} \in \mathcal{D}^{\otimes}$,

$$1 = \widehat{f}(\alpha_1\overline{\alpha_1}) = \widehat{f}^{01101001} + \widehat{f}^{11100001} = \widehat{f}^{00001111} + \widehat{f}^{10000111} = \widehat{f}^{00001111}.$$

Then, by ARS, $\widehat{f}^{11110000} = \overline{\widehat{f}^{00001111}} = 1$. Thus, $\widehat{f}(\gamma) = 1$ for any $\gamma \in \mathscr{S}(\widehat{f})$ with $\mathrm{wt}(\gamma) = 4$. Remember that $\widehat{f}(\vec{0}^8) = a$ where $|a| = 1$. Then, $\widehat{f}(\vec{1}^8) = \overline{a}$ by ARS. Suppose that $a = e^{\mathrm{i}\theta}$. Let $\widehat{Q} = \left[\begin{smallmatrix} \rho & 0 \\ 0 & \overline{\rho} \end{smallmatrix}\right] \in \widehat{\mathbf{O}_2}$ where $\rho = e^{-\mathrm{i}\theta/8}$. Consider the holographic transformation by $\widehat{Q}$. $\widehat{Q}$ does not change the entries of $\widehat{f}$ on half-weighed inputs, but change the values of $\widehat{f}(\vec{0}^8)$ and $\widehat{f}(\vec{1}^8)$ to 1. Thus, $\widehat{Q}\widehat{f} = \widehat{f_8}$. Then, Holant$(\neq_2 |\ \widehat{f_8}, \widehat{Q}\widehat{\mathcal{F}}) \leqslant_T$ Holant$(\neq_2 |\ \widehat{\mathcal{F}})$. □

Now, we want to show that Holant$(\neq_2 |\ \widehat{f_8}, \widehat{Q}\widehat{\mathcal{F}})$ is #P-hard for all $\widehat{Q} \in \widehat{\mathbf{O}_2}$ and all $\widehat{\mathcal{F}}$ where $\mathcal{F} = Z\widehat{\mathcal{F}}$ is a real-valued signature set that does not satisfy condition (T). If so, then we are done. Recall that for all $\widehat{Q} \in \widehat{\mathbf{O}_2}$, $\widehat{Q}\widehat{\mathcal{F}} = \widehat{Q\mathcal{F}}$ for some $Q \in \mathbf{O}_2$. Moreover, for all $Q \in \mathbf{O}_2$, and all real-valued $\mathcal{F}$ that does not satisfy condition (T), $Q\mathcal{F}$ is also a real-valued signature set that does not satisfy condition (T). Thus, it suffices for us to show that Holant$(\neq_2 |\ \widehat{f_8}, \widehat{\mathcal{F}})$ is #P-hard for all real-valued $\mathcal{F}$ that does not satisfy condition (T).

The following Lemma shows that $\widehat{f_8}$ gives non-$\widehat{\mathcal{B}}$ hardness (Definition 6.7).

Lemma 8.6. Holant$(\neq_2 |\ \widehat{f_8}, \widehat{\mathcal{F}})$ *is #P-hard if* $\widehat{\mathcal{F}}$ *contains a nonzero binary signature* $\widehat{b} \notin \widehat{\mathcal{B}}^{\otimes}$. *Equivalently,* Holant$(f_8, \mathcal{F})$ *is #P-hard if* $\mathcal{F}$ *contains a nonzero binary signature* $b \notin \mathcal{B}^{\otimes}$.

Proof. We prove this lemma in the setting of Holant$(\neq_2 |\ \widehat{f_8}, \widehat{\mathcal{F}})$. If $\widehat{b} \notin \widehat{\mathcal{O}}^{\otimes}$, then by Lemma 5.1, we get #P-hardness. Thus, we may assume that $\widehat{b} \in \widehat{\mathcal{O}}^{\otimes}$. Then, $\widehat{b}$ has parity. We first consider the case that $\widehat{b}$ has even parity, i.e., $\widehat{b} = (a, 0, 0, \bar{a})$. Since $\widehat{b} \not\equiv 0$, $a \neq 0$. We can normalize a to $e^{\mathrm{i}\theta}$ where $0 \leqslant \theta < \pi$. Then $\bar{a} = e^{-\mathrm{i}\theta}$. Since $\widehat{b} \notin \widehat{\mathcal{B}}$, $a \neq \pm 1$ and $a \neq \pm\mathrm{i}$. Thus, $\theta \neq 0$ and $\theta \neq \frac{\pi}{2}$.

We connect variables x_1 and x_5 of $\widehat{f_8}$ with the two variables of $\widehat{b}$ (using $\neq_2$), and we get a 6-ary signature denoted by $\widehat{g}$. We rename variables x_2, x_3, x_4 of $\widehat{g}$ to x_1, x_2, x_3 and variables x_6, x_7, x_8 to x_4, x_5, x_6. Then, $\widehat{g}$ has the following signature matrix

$$M_{123,456}(\widehat{g}) = \begin{bmatrix} e^{-\mathrm{i}\theta} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & e^{\mathrm{i}\theta} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & e^{\mathrm{i}\theta} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & e^{-\mathrm{i}\theta} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & e^{\mathrm{i}\theta} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & e^{-\mathrm{i}\theta} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & e^{-\mathrm{i}\theta} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & e^{\mathrm{i}\theta} \end{bmatrix}.$$

Now, we show that $\widehat{g} \notin \widehat{\mathcal{O}}^{\otimes}$. For a contradiction, suppose that $\widehat{g} \in \widehat{\mathcal{O}}^{\otimes}$. Notice that $\mathscr{S}(\widehat{g}) = \{(x_1, \ldots, x_6) \in \mathbb{Z}_2^6 \mid x_1 = x_4,\ x_2 = x_5 \text{ and } x_3 = x_6\}$. Then, we can write $\widehat{g}$ as

$$\widehat{g} = \widehat{b_1}(x_1, x_4) \otimes \widehat{b_2}(x_2, x_5) \otimes \widehat{b_3}(x_3, x_6),$$

where $\widehat{b_1} = (e^{\mathrm{i}\theta_1}, 0, 0, e^{-\mathrm{i}\theta_1})$, $\widehat{b_2} = (e^{\mathrm{i}\theta_2}, 0, 0, e^{-\mathrm{i}\theta_2})$ and $\widehat{b_3} = (e^{\mathrm{i}\theta_3}, 0, 0, e^{-\mathrm{i}\theta_3})$. Then notice that

$$\widehat{g}^{000000} = e^{-\mathrm{i}\theta} = \widehat{b_1}(0,0) \cdot \widehat{b_2}(0,0) \cdot \widehat{b_3}(0,0) = e^{\mathrm{i}(\theta_1+\theta_2+\theta_3)},$$

and

$$\widehat{g}^{011011} = e^{-\mathrm{i}\theta} = \widehat{b_1}(0,0) \cdot \widehat{b_2}(1,1) \cdot \widehat{b_3}(1,1) = e^{\mathrm{i}(\theta_1-\theta_2-\theta_3)}.$$

By multiplying the above two equations, we have

$$e^{-\mathrm{i}2\theta} = e^{\mathrm{i}(\theta_1+\theta_2+\theta_3)} \cdot e^{\mathrm{i}(\theta_1-\theta_2-\theta_3)} = e^{\mathrm{i}2\theta_1}.$$

Also, notice that

$$\widehat{g}^{001001} = e^{\mathrm{i}\theta} = \widehat{b_1}(0,0) \cdot \widehat{b_2}(0,0) \cdot \widehat{b_3}(1,1) = e^{\mathrm{i}(\theta_1+\theta_2-\theta_3)},$$

and

$$\widehat{g}^{010010} = e^{\mathrm{i}\theta} = \widehat{b_1}(0,0) \cdot \widehat{b_2}(1,1) \cdot \widehat{b_3}(0,0) = e^{\mathrm{i}(\theta_1-\theta_2+\theta_3)}.$$

By multiplying them, we have

$$e^{\mathrm{i}2\theta} = e^{\mathrm{i}(\theta_1+\theta_2-\theta_3)} \cdot e^{\mathrm{i}(\theta_1-\theta_2+\theta_3)} = e^{\mathrm{i}2\theta_1}.$$

Thus, $e^{\mathrm{i}2\theta} = e^{-\mathrm{i}2\theta}$. Then, $e^{\mathrm{i}4\theta} = 1$. Since, $\theta \in [0, \pi)$, $\theta = 0$ or $\frac{\pi}{2}$. Contradiction. Thus, $\widehat{g} \notin \widehat{\mathcal{O}}^{\otimes}$. By Lemma 7.21, we get #P-hardness.

Now, suppose that $\widehat{b}$ has odd parity, i.e., $\widehat{b}(y_1, y_2) = (0, e^{\mathrm{i}\theta}, e^{-\mathrm{i}\theta}, 0)$ where $\theta \in [0, \pi)$ after normalization. We still consider the 6-ary signature $\widehat{g'}$ that is realized by connecting variables x_1 and x_5 of $\widehat{f_8}$ with the two variables y_1 and y_2 of $\widehat{b}$ (using $\neq_2$). Then, after renaming variables, $\widehat{g'}$ has the following signature matrix

$$M_{123,456}(\widehat{g'}) = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & e^{-\mathrm{i}\theta} \\ 0 & 0 & 0 & 0 & 0 & 0 & e^{\mathrm{i}\theta} & 0 \\ 0 & 0 & 0 & 0 & 0 & e^{\mathrm{i}\theta} & 0 & 0 \\ 0 & 0 & 0 & 0 & e^{-\mathrm{i}\theta} & 0 & 0 & 0 \\ 0 & 0 & 0 & e^{\mathrm{i}\theta} & 0 & 0 & 0 & 0 \\ 0 & 0 & e^{-\mathrm{i}\theta} & 0 & 0 & 0 & 0 & 0 \\ 0 & e^{-\mathrm{i}\theta} & 0 & 0 & 0 & 0 & 0 & 0 \\ e^{\mathrm{i}\theta} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Similarly, we can show that $\widehat{g'} \notin \widehat{\mathcal{O}}^{\otimes}$. Thus, by Lemma 7.21, we get #P-hardness. □

We go back to real-valued Holant problems under the Z-transformation. Consider the problem $\text{Holant}(f_8, \mathcal{F})$. Remember that $f_8 = \widehat{f_8}$. We observe that, by Lemma 8.6 the set $\{f_8\} \cup \mathcal{F}$ is non-$\mathcal{B}$ hard, according to Definition 6.7. Then if we apply Theorem 7.19 to the set $\{f_8\} \cup \mathcal{F}$ we see that $\text{Holant}^b(f_8, \mathcal{F})$ is #P-hard. Now *if we were able to show* that $\mathcal{B}$ is realizable from f_8 then we would be done, since by Theorem 8.5, we either already have the #P-hardness for $\text{Holant}(\mathcal{F})$, or we can realize f_8 from $\mathcal{F}$, and thus the following reduction chain holds

$$\text{Holant}^b(f_8, \mathcal{F}) \leqslant_T \text{Holant}(f_8, \mathcal{F}) \leqslant_T \text{Holant}(\mathcal{F}).$$

Thus we get the #P-hardness of $\text{Holant}(\mathcal{F})$ in either way.

However, since f_8 has even parity and all its entries are non-negative, all gadgets realizable from f_8 have even parity and have non-negative entries. Thus, $=_2^-$, $\neq_2$ and $\neq_2^-$ *cannot* be realized from f_8 by gadget construction. In fact, it is observed in [15] that f_8 satisfies the following strong Bell property.

Definition 8.7. *A signature f satisfies the strong Bell property if for all pairs of indices $\{i, j\}$, and every $b \in \mathcal{B}$, the signature $\partial_{ij}^b f$ realized by merging x_i and x_j of f using b is in $\{b\}^{\otimes}$.*

8.2 Holant problems with limited appearance and a novel reduction

In this subsection, *not using gadget construction* but critically based on the strong Bell property of f_8, we prove that $\text{Holant}^b(f_8, \mathcal{F}) \leqslant_T \text{Holant}(f_8, \mathcal{F})$ in a novel way. We define the following Holant problems with limited appearance.

Definition 8.8. *Let $\mathcal{F}$ be a signature set containing a signature f. The problem $\text{Holant}(f^{\leqslant k}, \mathcal{F})$ contains all instances of $\text{Holant}(\mathcal{F})$ where the signature f appears at most k times.*

Lemma 8.9. *For any $b \in \mathcal{B}$, $\text{Holant}(b, f_8, \mathcal{F}) \leqslant_T \text{Holant}(b^{\leqslant 2}, f_8, \mathcal{F})$.*

Proof. Consider an instance Ω of $\text{Holant}(b, f_8, \mathcal{F})$. Suppose that b appears n times in Ω. If $n \leqslant 2$, then Ω is already an instance of $\text{Holant}(b^{\leqslant 2}, f_8, \mathcal{F})$. Otherwise, $n \geqslant 3$. Consider the gadget $\partial_{ij}^b f_8$ realized by connecting two variables x_i and x_j of f_8 using b. (This gadget uses b only once.) Since f_8 satisfies the strong Bell property, $\partial_{ij}^b f_8 = b^{\otimes 3}$. Thus, by replacing three occurrences of b in Ω by $\partial_{ij}^b f_8$, we can reduce the number of occurrences of b by 2. We carry out this replacement a linear number of times to obtain an equivalent instance of $\text{Holant}(b^{\leqslant 2}, f_8, \mathcal{F})$, of size linear in Ω. □

Now, we are ready to prove the reduction $\text{Holant}^b(f_8, \mathcal{F}) \leqslant_T \text{Holant}(f_8, \mathcal{F})$. Note that if $\text{Holant}(f_8, \mathcal{F})$ is #P-hard, then the reduction holds trivially. For any $b \in \mathcal{B}$, if we connect a variable of b with a variable of another copy of b using $=_2$, we get $\pm(=_2)$. Also, for any $b_1, b_2 \in \mathcal{B}$ where $b_1 \neq b_2$ if we connect the two variables of b_1 with the two variables of b_2, we get a value 0.

Lemma 8.10. $\text{Holant}^b(f_8, \mathcal{F}) \leqslant_T \text{Holant}(f_8, \mathcal{F})$.

Proof. We prove this reduction in two steps.

Step 1. There exists a signature $b_1 \in \mathcal{B} \backslash \{=_2\}$ such that $\text{Holant}(b_1, f_8, \mathcal{F}) \leqslant_T \text{Holant}(f_8, \mathcal{F})$.

We consider *all* binary and 4-ary signatures realizable by gadget constructions from $\{f_8\} \cup \mathcal{F}$. If a binary signature $g \notin \mathcal{B}$ is realizable from $\{f_8\} \cup \mathcal{F}$, then by Lemma 8.6, $\text{Holant}(f_8, \mathcal{F})$ is #P-hard,

and we are done. If a binary signature $g \in \mathcal{B}\backslash\{=_2\}$ is realizable from $\{f_8\} \cup \mathcal{F}$, then we are done by choosing $b_1 = g$. So we may assume that all binary signatures g realizable from $\{f_8\} \cup \mathcal{F}$ are $=_2$ (up to a scalar) or the zero binary signature, i.e., $g = \mu \cdot (=_2)$ for some $\mu \in \mathbb{R}$. Similarly, if a nonzero 4-ary signature $h \notin \mathcal{B}^{\otimes 2}$ is realizable, then we have Holant$(f_8, \mathcal{F})$ is #P-hard, by Lemma 6.8, as Lemma 8.6 says the set $\{f_8\} \cup \mathcal{F}$ is non-$\mathcal{B}$ hard. If a nonzero 4-ary signature $h \in \mathcal{B}^{\otimes 2}\backslash\{=_2\}^{\otimes 2}$ is realizable, then we can realize a binary signature $b_1 \in \mathcal{B}\backslash\{=_2\}$ by factorization, and we are done. Thus, we may assume that all 4-ary signatures h realizable from $\{f_8\} \cup \mathcal{F}$ are $(=_2)^{\otimes 2}$ or the 4-ary zero signature, i.e., $h = \lambda \cdot (=_2)^{\otimes 2}$ for some $\lambda \in \mathbb{R}$.

Now, let b_1 be a signature in $\mathcal{B}\backslash\{=_2\}$. We show that Holant$(b_1^{\leqslant 2}, f_8, \mathcal{F}) \leqslant_T$ Holant$(f_8, \mathcal{F})$. Consider an instance Ω of Holant$(b_1^{\leqslant 2}, f_8, \mathcal{F})$.

- If b_1 does not appear in Ω, then Ω is already an instance of Holant$(f_8, \mathcal{F})$.
- If b_1 appears exactly once in Ω (we may assume it does connect to itself), then we may consider *the rest of* Ω *that connects to* b_1 as a gadget realized from $\{f_8\} \cup \mathcal{F}$, which must have signature $\lambda \cdot (=_2)$, for some $\lambda \in \mathbb{R}$. Connecting the two variables of b_1 by $(=_2)$ for every $b_1 \in \mathcal{B}\backslash\{=_2\}$ will always gives 0. Thus, Holant$(\Omega) = 0$.
- Suppose b_1 appears exactly twice in Ω. It is easy to handle when the two copies of b_1 form a gadget of arity 0 or 2 to the rest of Ω. We may assume they are connected to the rest of Ω in such a way that the rest of Ω forms a 4-ary gadget h realized from $\{f_8\} \cup \mathcal{F}$. We can name the four dangling edges of h in any specific ordering as (x_1, x_2, x_3, x_4). Then

$$h(x_1, x_2, x_3, x_4) = \lambda \cdot (=_2)(x_1, x_j) \otimes (=_2)(x_k, x_\ell)$$

  for some partition $\{1, 2, 3, 4\} = \{1, j\} \sqcup \{k, \ell\}$, and some $\lambda \in \mathbb{R}$. (Note that while we have named four specific dangling edges as (x_1, x_2, x_3, x_4), the specific partition $\{1, 2, 3, 4\} = \{1, j\} \sqcup \{k, \ell\}$ and the value λ are unknown at this point.) We consider the following three instances Ω_{12}, Ω_{13}, and Ω_{14}, where Ω_{1s} ($s \in \{2, 3, 4\}$) is the instance formed by merging variables x_1 and x_s of h using $=_2$, and merging the other two variables of h using $=_2$ (see Figure 4 where $h_1 = h_2 = (=_2)$ and $h = \lambda \cdot h_1 \otimes h_2$). Since h is a gadget realized from $\{f_8\} \cup \mathcal{F}$, Ω_{12}, Ω_{13}, and Ω_{14} are instances of Holant$(f_8, \mathcal{F})$. Note that Holant$(\Omega_{1s}) = 4\lambda$ when $s = j$ and Holant$(\Omega_{1s}) = 2\lambda$ otherwise. Thus, by computing Holant(Ω_{1s}) for $s \in \{2, 3, 4\}$, we can get λ, and if $\lambda \neq 0$ the partition $\{1, j\} \sqcup \{k, \ell\}$ of the four variables. Thus we can get the exact structure of the 4-ary gadget h. In either case (whether $\lambda = 0$ or not), we can compute the value of Holant(Ω).

Thus, Holant$(b_1^{\leqslant 2}, f_8, \mathcal{F}) \leqslant_T$ Holant$(f_8, \mathcal{F})$. By Lemma 8.9, Holant$(b_1, f_8, \mathcal{F}) \leqslant_T$ Holant$(f_8, \mathcal{F})$.

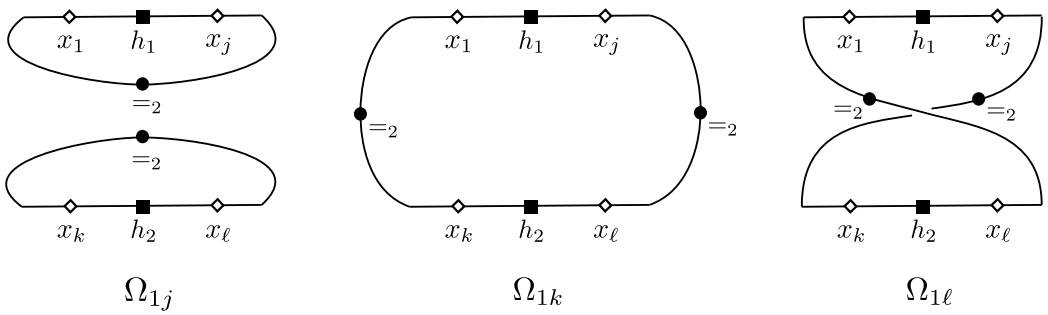


Figure 4: Instances Ω_{1j}, Ω_{1k} and $\Omega_{1\ell}$

Step 2. For any $b_1 \in \mathcal{B}\backslash\{=_2\}$, we have $\text{Holant}^b(f_8, \mathcal{F}) \leqslant_T \text{Holant}(b_1, f_8, \mathcal{F})$.

We show that we can get another $b_2 \in \mathcal{B}\backslash\{=_2, b_1\}$, i.e., for some binary signature $b_2 \in \mathcal{B}\backslash\{=_2, b_1\}$ we have the reduction $\text{Holant}(b_2, b_1, f_8, \mathcal{F}) \leqslant_T \text{Holant}(b_1, f_8, \mathcal{F})$. Then, by connecting one variable of b_1 and one variable of b_2 using $=_2$, we get the third signature in $\mathcal{B}\backslash\{b_1, b_2\}$. Then, the lemma is proved. The proof is similar to the proof in Step 1. We consider *all* binary and 4-ary gadgets realizable from $\{b_1, f_8\} \cup \mathcal{F}$. Still, we may assume that all realizable binary signatures are of the form $\mu \cdot (=_2)$ or $\mu \cdot b_1$ for some $\mu \in \mathbb{R}$, and all realizable 4-ary signatures are of form $\lambda \cdot (=_2)^{\otimes 2}$, $\lambda \cdot b_1^{\otimes 2}$ or $\lambda \cdot (=_2) \otimes b_1$ for some $\lambda \in \mathbb{R}$. Otherwise, we can show that $\text{Holant}(b_1, f_8, \mathcal{F})$ is #P-hard or we realize a signature $b_2 \in \mathcal{B}\backslash\{=_2, b_1\}$ directly by gadget construction.

Then, let b_2 be an arbitrary signature in $\mathcal{B}\backslash\{=_2, b_1\}$. We show that

$$\text{Holant}(b_2^{\leqslant 2}, b_1, f_8, \mathcal{F}) \leqslant_T \text{Holant}(b_1, f_8, \mathcal{F}).$$

Consider an instance Ω of $\text{Holant}(b_2^{\leqslant 2}, b_1, f_8, \mathcal{F})$. If b_2 does not appear in Ω, then Ω is already an instance of $\text{Holant}(b_1, f_8, \mathcal{F})$. If b_2 appears exactly once in Ω, then it is connected with a binary gadget g where $g = \mu \cdot (=_2)$ or $g = \mu \cdot b_1$. In both cases, the evaluation is 0. Thus, $\text{Holant}(\Omega) = 0$. Suppose b_2 appears exactly twice in Ω. Again it is easy to handle the case if the rest of Ω forms a gadget of arity 0 or 2 to the two occurrences of b_2. So we may assume the two occurrences of b_2 are connected to a 4-ary gadget $h = \lambda \cdot (=_2)^{\otimes 2}$, $\lambda \cdot b_1^{\otimes 2}$ or $\lambda \cdot (=_2) \otimes b_1$. We denote the four variables of h by (x_1, x_2, x_3, x_4), by an arbitrary ordering of the four dangling edges. Then $h(x_1, x_2, x_3, x_4) = \lambda \cdot h_1(x_1, x_j) \otimes h_2(x_k, x_\ell)$ where $h_1, h_2 \in \{=_2, b_1\}$, for some λ and $\{j, k, \ell\} = \{2, 3, 4\}$. (Note that at the moment the values λ and j, k, ℓ are unknown.) We consider the following three instances Ω_{12}, Ω_{13} and Ω_{14}, where Ω_{1s} ($s \in \{2, 3, 4\}$) is the instance formed by connecting variables x_1 and x_s of h using $=_2$, and connecting the other two variables of h using $=_2$ (again see Figure 4). Clearly, Ω_{12}, Ω_{13} and Ω_{14} are instances of $\text{Holant}(b_1, f_8, \mathcal{F})$. Consider the evaluations of these instances. We have three cases.

- If $h_1 = h_2 = (=_2)$, then $\text{Holant}(\Omega_{1s}) = 4\lambda$ when $s = j$ and $\text{Holant}(\Omega_{1s}) = 2\lambda$ when $s \neq j$.
- If $h_1 = h_2 = b_1$, then $\text{Holant}(\Omega_{1s}) = 0$ when $s = j$. If $M(b_1)$ is the 2 by 2 matrix form for the binary signature b_1 where we list its first variable as row index and second variable as column index, then we have $\text{Holant}(\Omega_{1k}) = \lambda \cdot \text{tr}(M(b_1)M(b_1)^{\mathtt{T}})$, and $\text{Holant}(\Omega_{1\ell}) = \lambda \cdot \text{tr}(M(b_1)^2)$, where tr denotes trace. For $b_1 = (=_2^-)$ or $(\neq_2^+)$, the matrix $M(b_1)$ is symmetric, and the value $\text{Holant}(\Omega_{1s}) = 2\lambda$ in both cases $s = k$ or $s = \ell$. For $b_1 = (\neq_2^-)$, $M(b_1)^{\mathtt{T}} = -M(b_1)$, and we have $\text{Holant}(\Omega_{1k}) = 2\lambda$, and $\text{Holant}(\Omega_{1\ell}) = -2\lambda$.
- If one of h_1 and h_2 is $=_2$ and the other is b_1, then $\text{Holant}(\Omega_{1s}) = 0$ for all $s \in \{j, k, \ell\}$.

Thus, if the values of $\text{Holant}(\Omega_{1s})$ for $s \in \{2, 3, 4\}$ are not all zero, then $\lambda \neq 0$ and the third case is impossible, and we can tell whether h is in the form $\lambda \cdot (=_2)^{\otimes 2}$ or $\lambda \cdot (b_1)^{\otimes 2}$. Moreover we can get the exact structure of h, i.e., the value λ and the decomposition form of h_1 and h_2. Otherwise, the values of $\text{Holant}(\Omega_{1s})$ for $s \in \{2, 3, 4\}$ are all zero. Then we can write $h = \lambda \cdot (=_2)(x_1, x_j) \otimes b_1(x_k, x_\ell)$ or $h = \lambda \cdot b_1(x_1, x_j) \otimes (=_2)(x_k, x_\ell)$, including possibly $\lambda = 0$, which means $h \equiv 0$. (Note that if $\lambda \neq 0$, this uniquely identifies that we are in the third case; if $\lambda = 0$ then this form is still formally valid, even though we cannot say this uniquely identifies the third case. But when $\lambda = 0$ all three cases are the same, i.e., $h \equiv 0$.) At this point we still do not know the exact value of λ and the decomposition form of h.

We further consider the following three instances Ω'_{12}, Ω'_{13} and Ω'_{14}, where Ω'_{1s} ($s \in \{2, 3, 4\}$) is the instance formed by connecting variables x_1 and x_s of h using b_1, and connecting the other two variables of h using $=_2$. (In other words, we replace the labeling $=_2$ of the edge that is connected

to the variable x_1 in each instance illustrated in Figure 4 by b_1.) It is easy to see that Ω'_{12}, Ω'_{13} and Ω'_{14} are instances of $\text{Holant}(b_1, f_8, \mathcal{F})$. Consider the evaluations of these instances.

- If $h_1 = (=_2)(x_1, x_j)$, then $\text{Holant}(\Omega'_{1s}) = 0$ when $s = j$. Also we have $\text{Holant}(\Omega_{1k}) = \lambda \cdot \text{tr}(M(b_1)^2)$, and $\text{Holant}(\Omega_{1\ell}) = \lambda \cdot \text{tr}(M(b_1)M(b_1)^{\mathrm{T}})$. For $b_1 = (=_2^-)$ or $\neq_2^+$, the matrix $M(b_1)$ is symmetric, and the value $\text{Holant}(\Omega_{1s}) = 2\lambda$ in both cases $s = k$ or $s = \ell$. For $b_1 = (\neq_2^-)$, $M(b_1)^{\mathrm{T}} = -M(b_1)$, and we have $\text{Holant}(\Omega_{1k}) = -2\lambda$, and $\text{Holant}(\Omega_{1\ell}) = 2\lambda$.
- If $h_1 = b_1(x_1, x_j)$, then $\text{Holant}(\Omega'_{1s}) = 4\lambda$ when $s = j$ and $\text{Holant}(\Omega'_{1s}) = 2\lambda$ when $s \neq j$.

Thus, by further computing $\text{Holant}(\Omega'_{1s})$ for $s \in \{2, 3, 4\}$, we can get the exact structure of h.

Therefore, by querying $\text{Holant}(b_1, f_8, \mathcal{F})$ at most 6 times, we can compute h exactly. Then, we can compute $\text{Holant}(\Omega)$ easily. Thus, $\text{Holant}(b_2^{\leqslant 2}, b_1, f_8, \mathcal{F}) \leqslant_T \text{Holant}(b_1, f_8, \mathcal{F})$. By Lemma 8.9, $\text{Holant}(b_2, b_1, f_8, \mathcal{F}) \leqslant_T \text{Holant}(b_1, f_8, \mathcal{F})$. The other signature in $\mathcal{B}\backslash\{=_2, b_1, b_2\}$ can be realized by connecting b_1 and b_2. Thus, $\text{Holant}^b(f_8, \mathcal{F}) \leqslant_T \text{Holant}(b_1, f_8, \mathcal{F})$.

Therefore, $\text{Holant}^b(f_8, \mathcal{F}) \leqslant_T \text{Holant}(f_8, \mathcal{F})$. □

Since $\text{Holant}^b(f_8, \mathcal{F}) \leqslant_T \text{Holant}(f_8, \mathcal{F})$ and $\{f_8\} \cup \mathcal{F}$ is non-$\mathcal{B}$ hard for any real-valued $\mathcal{F}$ that does not satisfy condition (T), by Theorem 7.19, we have the following result.

Lemma 8.11. $\text{Holant}(f_8, \mathcal{F})$ *is #P-hard.*

Combining Theorem 8.5 and Lemma 8.11, we have the following result.

Lemma 8.12. *If* $\widehat{\mathcal{F}}$ *contains a signature* $\widehat{f}$ *of arity* 8 *and* $\widehat{f} \notin \widehat{\mathcal{O}}^{\otimes}$*, then* $\text{Holant}(\neq_2 | \widehat{\mathcal{F}})$ *is #P-hard.*

9 The Induction Proof: $2n \geqslant 10$

Now, we show that our induction framework works for signatures of arity $2n \geqslant 10$.

Lemma 9.1. *If* $\widehat{\mathcal{F}}$ *contains a signature* $\widehat{f}$ *of arity* $2n \geqslant 10$ *and* $\widehat{f} \notin \widehat{\mathcal{O}}^{\otimes}$*, then,*

- $\text{Holant}(\neq_2 | \widehat{\mathcal{F}})$ *is #P-hard, or*
- *a signature* $\widehat{g} \notin \widehat{\mathcal{O}}^{\otimes}$ *of arity* $2k \leqslant 2n - 2$ *is realizable from* $\widehat{f}$*.*

Proof. By Lemma 8.1, we may assume that an irreducible signature $\widehat{f^*}$ of arity $2n \geqslant 10$ where $\widehat{f^*} \in \widehat{\int}\mathcal{D}^{\otimes}$ is realizable, and $\widehat{f^*}$ satisfies ARS. We show that $\widehat{f^*}$ does not satisfy 2ND-ORTH, and hence we get #P-hardness.

For all pairs of indices $\{i, j\}$, since $\widehat{\partial}_{ij}\widehat{f^*} \in \mathcal{D}^{\otimes}$, $\mathscr{S}(\widehat{\partial}_{ij}\widehat{f^*})$ is on half-weight. By Lemma 2.10, we have $\widehat{f^*}(\alpha) = 0$ for all $\text{wt}(\alpha) \neq 0, n, 2n$. Suppose that $\widehat{f^*}(\vec{0}^{2n}) = a$ and $\widehat{f^*}(\vec{1}^{2n}) = \bar{a}$ by ARS. We can write $\widehat{f^*}$ in the following form

$$\widehat{f^*} = a(1,0)^{\otimes 2n} + \bar{a}(0,1)^{\otimes 2n} + \widehat{f^*_{\text{h}}}.$$

where $\widehat{f^*_{\text{h}}}$ is an EO signature of arity $2n \geqslant 10$.

Clearly, $\partial_{ij}\widehat{f^*} = \partial_{ij}\widehat{f^*_{\text{h}}}$ for all $\{i, j\}$. Then, $\widehat{f^*_{\text{h}}} \in \widehat{\int}\mathcal{D}^{\otimes}$ since $\widehat{f^*} \in \widehat{\int}\mathcal{D}^{\otimes}$. Since $\widehat{f^*_{\text{h}}}$ is an EO signature of arity at least 10 and $\widehat{f^*_{\text{h}}} \in \widehat{\int}\mathcal{D}^{\otimes}$, by Lemma 2.11, we have $\widehat{f^*_{\text{h}}} \in \mathcal{D}^{\otimes}$. Recall that all signatures in $\mathcal{D}^{\otimes}$ are nonzero by definition. Pick some $\{i, j\}$ such that $(\neq_2)(x_i, x_j) \mid \widehat{f^*_{\text{h}}}$. Then,

$$\widehat{f^*} = a(1,0)^{\otimes 2n} + \bar{a}(0,1)^{\otimes 2n} + \widehat{b^*}(x_i, x_j) \otimes \widehat{g^*_{\text{h}}},$$

where $\widehat{g^*_{\text{h}}} \in \mathcal{D}^{\otimes}$ is a nonzero EO signature since $\widehat{f^*_{\text{h}}} \in \mathcal{D}^{\otimes}$. By Lemma 8.2, $\widehat{f^*}$ does not satisfy 2ND-ORTH. Thus, $\text{Holant}(\neq_2 | \widehat{\mathcal{F}})$ is #P-hard by Lemma 4.4. □

Remark: Indeed, following from our proof, we can also show that there is no irreducible signature $\widehat{f}$ of arity $2n \geqslant 10$ that satisfies both 2ND-ORTH and $\widehat{f} \in \widehat{\int}\widehat{\mathcal{O}}^{\otimes}$.

Finally, we give the proof of Theorem 1.2. We restate it here.

Theorem 9.2. *Let $\mathcal{F}$ be a set of real-valued signatures. If $\mathcal{F}$ satisfies the tractability condition (T) in Theorem 2.22, then* Holant$(\mathcal{F})$ *is polynomial-time computable; otherwise,* Holant$(\mathcal{F})$ *is #P-hard.*

Proof. By Theorem 2.22, if $\mathcal{F}$ satisfies condition (T), then Holant$(\mathcal{F})$ is P-time computable. Suppose that $\mathcal{F}$ does not satisfy condition (T). If $\mathcal{F}$ contains a nonzero signature of odd arity, then by Theorem 2.25, Holant$(\mathcal{F})$ is #P-hard. We show Holant$(\neq_2|\ \widehat{\mathcal{F}}) \equiv_T$ Holant$(\mathcal{F})$ is #P-hard when $\mathcal{F}$ is a set of signatures of even arity. Since $\mathcal{F}$ does not satisfy condition (T), $\widehat{\mathcal{F}} \not\subseteq \mathscr{T}$. Since $\widehat{\mathcal{O}}^{\otimes} \subseteq \mathscr{T}$, there is a signature $\widehat{f} \in \widehat{\mathcal{F}}$ of arity $2n$ such that $\widehat{f} \notin \widehat{\mathcal{O}}^{\otimes}$. We prove this theorem by induction on $2n$.

When $2n \leqslant 8$, by Lemmas 5.1, 5.2, 7.21, 8.12, Holant$(\neq_2|\ \widehat{\mathcal{F}})$ is #P-hard.

Inductively, suppose for some $2k \geqslant 8$, if $2n \leqslant 2k$, then Holant$(\neq_2|\ \widehat{\mathcal{F}})$ is #P-hard. We consider $2n = 2k+2 \geqslant 10$. By Lemma 9.1, Holant$(\neq_2|\ \widehat{\mathcal{F}})$ is #P-hard, or Holant$(\neq|\ \widehat{g}, \widehat{\mathcal{F}}) \leqslant_T$ Holant$(\neq|\ \widehat{\mathcal{F}})$ for some $\widehat{g} \notin \widehat{\mathcal{O}}^{\otimes}$ of arity $\leqslant 2k$. By the induction hypothesis, Holant$(\neq|\ \widehat{g}, \widehat{\mathcal{F}})$ is #P-hard. Thus, Holant$(\neq_2|\ \widehat{\mathcal{F}})$ is #P-hard. □

Acknowledgement

We would like to thank Professor Mingji Xia for pointing out that the strong Bell property can be used to prove Lemma 8.9, which led to the inspiration to prove Lemma 8.10. Without this inspiration this paper may languish for much more time. We would also like to thank Professor Zhiguo Fu for many valuable discussions and verification of parts of the proof. The proof of the present paper is built on his substantial previous work. Despite their invaluable contributions, they generously declined our invitation for co-authorship.

References

[1] Miriam Backens. A new holant dichotomy inspired by quantum computation. In *Proceedings of the 44th International Colloquium on Automata, Languages, and Programming*, pages 16:1–16:14, 2017.

[2] Miriam Backens. A complete dichotomy for complex-valued holant^c. In *Proceedings of the 45th International Colloquium on Automata, Languages, and Programming*, pages 12:1–12:14, 2018.

[3] Rodney J Baxter. Eight-vertex model in lattice statistics. *Physical Review Letters*, 26(14):832, 1971.

[4] Rodney J Baxter. The six and eight-vertex models revisited. *Journal of statistical physics*, 116(1-4):43–66, 2004.

[5] John S Bell. On the einstein podolsky rosen paradox. *Physics Physique Fizika*, 1(3):195, 1964.

[6] Andrei Bulatov, Martin Dyer, Leslie Ann Goldberg, Markus Jalsenius, Mark Jerrum, and David Richerby. The complexity of weighted and unweighted #csp. *Journal of Computer and System Sciences*, 78(2):681–688, 2012.

[7] Andrei Bulatov and Martin Grohe. The complexity of partition functions. *Theoretical Computer Science*, 348(2-3):148–186, 2005.

[8] Andrei A Bulatov. The complexity of the counting constraint satisfaction problem. *Journal of the ACM (JACM)*, 60(5):1–41, 2013.

[9] Jin-Yi Cai and Xi Chen. *Complexity Dichotomies for Counting Problems: Volume 1, Boolean Domain*. Cambridge University Press, 2017.

[10] Jin-Yi Cai and Xi Chen. Complexity of counting csp with complex weights. *Journal of the ACM (JACM)*, 64(3):1–39, 2017.

[11] Jin-Yi Cai, Xi Chen, and Pinyan Lu. Graph homomorphisms with complex values: A dichotomy theorem. *SIAM Journal on Computing*, 42(3):924–1029, 2013.

[12] Jin-Yi Cai, Xi Chen, and Pinyan Lu. Nonnegative weighted# csp: An effective complexity dichotomy. *SIAM Journal on Computing*, 45(6):2177–2198, 2016.

[13] Jin-Yi Cai and Zhiguo Fu. Complexity classification of the eight-vertex model. *arXiv preprint arXiv:1702.07938*, 2017.

[14] Jin-Yi Cai, Zhiguo Fu, and Shuai Shao. Complexity of counting weighted eulerian orientations with ars. *arXiv preprint arXiv:1904.02362*, 2019.

[15] Jin-Yi Cai, Zhiguo Fu, and Shuai Shao. From holant to quantum entanglement and back. *arXiv preprint arXiv:2004.05706 (to appear in ICALP 2020)*, 2020.

[16] Jin-Yi Cai, Zhiguo Fu, and Mingji Xia. Complexity classification of the six-vertex model. *Information and Computation*, 259:130–141, 2018.

[17] Jin-Yi Cai and Artem Govorov. Perfect matchings, rank of connection tensors and graph homomorphisms. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 476–495. SIAM, 2019.

[18] Jin-Yi Cai, Heng Guo, and Tyson Williams. A complete dichotomy rises from the capture of vanishing signatures. *SIAM Journal on Computing*, 45(5):1671–1728, 2016.

[19] Jin-Yi Cai, Pinyan Lu, and Mingji Xia. Dichotomy for holant* problems of boolean domain. In *Proceedings of the twenty-second annual ACM-SIAM symposium on Discrete Algorithms*, pages 1714–1728. SIAM, 2011.

[20] Jin-Yi Cai, Pinyan Lu, and Mingji Xia. Dichotomy for real holant^c problems. In *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1802–1821. SIAM, 2018.

[21] Martin Dyer and Catherine Greenhill. The complexity of counting graph homomorphisms. *Random Structures & Algorithms*, 17(3-4):260–289, 2000.

[22] Martin Dyer and David Richerby. An effective dichotomy for the counting constraint satisfaction problem. *SIAM Journal on Computing*, 42(3):1245–1274, 2013.

[23] Michael Freedman, László Lovász, and Alexander Schrijver. Reflection positivity, rank connectivity, and homomorphism of graphs. *Journal of the American Mathematical Society*, 20(1):37–51, 2007.

[24] Leslie Ann Goldberg, Martin Grohe, Mark Jerrum, and Marc Thurley. A complexity dichotomy for partition functions with mixed signs. *SIAM Journal on Computing*, 39(7):3336–3402, 2010.

[25] Jiabao Lin and Hanpin Wang. The complexity of boolean holant problems with nonnegative weights. *SIAM Journal on Computing*, 47(3):798–828, 2018.

[26] Milena Mihail and Peter Winkler. On the number of eulerian orientations of a graph. *Algorithmica*, 16(4-5):402–414, 1996.

[27] Linus Pauling. The structure and entropy of ice and of other crystals with some randomness of atomic arrangement. *Journal of the American Chemical Society*, 57(12):2680–2684, 1935.

[28] Leslie G Valiant. The complexity of enumeration and reliability problems. *SIAM Journal on Computing*, 8(3):410–421, 1979.

[29] Leslie G Valiant. Holographic algorithms. *SIAM Journal on Computing*, 37(5):1565–1594, 2008.

Paper 2

SIAM J. COMPUT.
Vol. 0, No. 0, pp. STOC17-50–STOC17-151

HOLOGRAPHIC ALGORITHM WITH MATCHGATES IS UNIVERSAL FOR PLANAR #CSP OVER BOOLEAN DOMAIN*

JIN-YI CAI† AND ZHIGUO FU‡

Abstract. We prove a complexity classification theorem that classifies all counting constraint satisfaction problems (#CSP) over Boolean variables into exactly three classes: (1) polynomial-time solvable; (2) #P-hard for general instances but solvable in polynomial time over planar structures; and (3) #P-hard over planar structures. The classification applies to all finite sets of local, *not necessarily symmetric*, constraint functions on Boolean variables that take algebraic complex values. It is shown that Valiant's holographic algorithm with matchgates is a *universal* strategy for all problems in class (2).

Key words. #CSP, holographic algorithms, matchgates

AMS subject classifications. 68Q25, 68Q17

DOI. 10.1137/17M1131672

1. Introduction.

Half a century ago, the Fisher–Kasteleyn–Temperley (FKT) algorithm was discovered [41, 33, 35]. The FKT algorithm can count the number of perfect matchings (dimers) over planar graphs in polynomial time. This was a milestone in a long and beautiful sequence of work in statistical physics involving Lenz, Ising, Onsager, Yang, Lee, Fisher, Temperley, Kasteleyn, Baxter, Lieb, and Wilson [31, 39, 48, 49, 37, 41, 33, 35, 1, 38], with contributions from many others. The central question is what constitutes an "exactly solved model." The basic conclusion from physicists is that for some "systems" their partition functions are exactly solvable for planar structures but appear intractable for higher dimensions. However, exactly what does it mean to be intractable? Physicists did not have a formal notion of intractability.

This notion is supplied by complexity theory. Following the P versus NP theory, in 1979 Leslie Valiant [42] defined the class #P for counting problems. Most interesting combinatorial counting problems are included in this broad class. Sum-of-product computations, such as partition functions studied in physics, and counting constraint satisfaction problems are included in #P (or by a polynomial-time reduction when the output is not an integer), and #P-hardness is at least as hard as NP-hardness. In particular, counting perfect matchings in general graphs is #P-complete [43].

But are there other surprises like the FKT algorithm? If so, can they solve any #P-hard problems? In two seminal papers [44, 46], Valiant introduced *matchgates* and *holographic algorithms*. These holographic algorithms use a quantum-like superposition to achieve fantastic cancellations, which produce polynomial-time algorithms to solve a number of concrete problems that would seem to require exponential time to compute. The first ingredient of holographic algorithms is the FKT algorithm. The second ingredient is a tensor-theoretic transformation that establishes a quantitative

*Received by the editors May 30, 2017; accepted for publication (in revised form) June 12, 2019; published electronically November 5, 2019.
https://doi.org/10.1137/17M1131672
Funding: The second author was supported by NSFC-61872076 and "Fundamental Research Funds for Central Universities."
†Computer Sciences, University of Wisconsin-Madison, Madison, WI 53706 (jyc@cs.wisc.edu).
‡College of Computer Science and Information Technology, Northeast Normal University, Changchun, Jilin, 130117, China (fuzg432@nenu.edu.cn).

equivalence of two seemingly different counting problems. This holographic reduction in general does not preserve solutions between the two problems in a 1-1 fashion. These transformations establish a duality similar in spirit to the Fourier transform and its inverse.

As these novel algorithms solve problems that appear so close to being #P-hard, they naturally raise the question of whether they can solve #P-hard problems in polynomial time. In the past 10 to 15 years significant progress has been made in the understanding of these remarkable algorithms [6, 8, 10, 11, 14, 15, 18, 27, 36, 45, 46, 47]. In an interesting twist, it turns out that the idea of a holographic reduction is not only a powerful technique to design new and unexpected algorithms but also an indispensable tool to classify the inherent complexity of counting problems, in particular to understand the limit and scope of holographic algorithms based on matchgates [12, 13, 26, 30, 28, 19, 8, 27, 9, 6]. This study has produced a number of complexity dichotomy theorems. These classify *every* problem expressible in a framework as either solvable in P or #P-hard, with nothing in between.

One such framework is called #CSP problems. A #CSP problem on Boolean variables is specified by a set of local constraint functions $\mathcal{F}$. Each function $f \in \mathcal{F}$ has an arity k and maps $\{0,1\}^k \to \mathbb{C}$. (For consideration of models of computation, we restrict function values to be algebraic numbers. Unweighted #CSP problems are defined by 0-1 valued constraint functions.) An instance of #CSP($\mathcal{F}$) on Boolean variables is specified by a finite set of variables $X = \{x_1, x_2, \ldots, x_n\}$, each taking value 0 or 1, and a finite sequence of constraints $\mathcal{S}$ from $\mathcal{F}$, each applied to an ordered sequence of variables from X. Every instance can be described by a bipartite graph where the left-hand side (LHS) nodes are variables X, the right-hand side (RHS) nodes are constraints $\mathcal{S}$, and the connections between them specify occurrences of variables in constraints in the input instance. The output of this instance is $\sum_\sigma \prod_{f \in \mathcal{S}} f|_\sigma$, a sum over all $\sigma : X \to \{0,1\}$ of products of all constraints in $\mathcal{S}$ evaluated according to σ. In the unweighted 0-1 case, each such product contributes a 1 if σ satisfies all constraints in $\mathcal{S}$, and 0 otherwise. In the general case, the output is a weighted sum of 2^n terms. #CSP is a very expressive framework for locally specified counting problems. A spin system is a special case where there is one single binary constraint in $\mathcal{F}$, and possibly one or more unary constraints when there are "external fields."

We prove in this paper that holographic algorithms with matchgates form a *universal* strategy for problems expressible in this framework that are #P-hard in general but solvable in polynomial time on planar graphs. More specifically, we prove the following classification theorem.

THEOREM 1.1. *For any set of constraint functions $\mathcal{F}$ over Boolean variables, each taking complex values and not necessarily symmetric, #CSP($\mathcal{F}$) belongs to exactly one of three categories according to $\mathcal{F}$:* (1) *It is polynomial-time solvable.* (2) *It is polynomial-time solvable over planar graphs but #P-hard over general graphs.* (3) *It is #P-hard over planar graphs. Moreover, category* (2) *consists precisely of those problems that are holographically reducible to the FKT algorithm.*

This theorem finally settles the full reach of the power of Valiant's holographic algorithms in the #CSP framework over Boolean variables. Several results preceded this. The three most direct predecessors are as follows:

(I) In [15] it is shown that Theorem 1.1 holds if every function in $\mathcal{F}$ is real-valued and *symmetric*.

The value of a symmetric function is invariant when the input values are permuted. For practical problems, symmetric constraint functions cover a lot of

ground. However from the viewpoint of the classification program for counting complexity, to be symmetric is quite a stringent restriction. A constraint function on n Boolean variables requires 2^n output values to specify, while a symmetric one needs only $n+1$ values.

(II) Guo and Williams [27] generalize [15] to the case where functions in $\mathcal{F}$ are complex-valued, but they must still be *symmetric*.
Complex numbers form the natural setting to discuss the power of these problems. Many problems, even though they are real-valued, are shown to be equivalent under a holographic reduction which goes through $\mathbb{C}$, and their inherent complexity is only understood by an analysis in $\mathbb{C}$ on quantities such as eigenvalues.

(III) If one ignores planarity, [20] proves a complexity dichotomy.
This result itself generalizes previous results by Creignou and Hermann [21] for the case when all constraint functions are 0-1 valued, by Dyer, Goldberg, and Jerrum [22] for nonnegative valued constraint functions, and by Bulatov et al. [3] for real-valued constraint functions of mixed signs.

The classification in Theorem 1.1, especially the claim that holographic reductions followed by the FKT are universal for category (2), is by no means self-evident. In fact, such a sweeping claim should invite skepticism. Nowhere in complexity theory of decision problems at the P versus NP level are we aware of such a provable universal algorithmic strategy for a broad class of problems, in the sense that for every problem in this class it is solvable in polynomial time iff it is solvable by this particular strategy. Moreover, in the study of holographic algorithms, an even broader class than #CSP of locally specified sum-of-product computations has been introduced [13], called Holant problems. It turns out that counting perfect matchings is naturally expressible as a Holant problem but not as a #CSP problem. Very recently we discovered that for planar Holant problems a corresponding universality statement as in Theorem 1.1 is *false* [6]. For planar Holant problems, a holographic reduction to the FKT is *not* universal; there are other #P-hard problems that become tractable, i.e., polynomial time computable, on planar structures, and they are not holographically reducible to the FKT.

The class of Holant problems turns out to be more than just a separate framework providing a cautionary reference to Theorem 1.1. In fact they form the main arena in which we carry out the proof of Theorem 1.1. A basic idea in this proof is a holographic transformation between the #CSP setting and the Holant setting via the Hadamard transformation $H_2 = \frac{1}{\sqrt{2}}\left[\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}\right]$. This transformation is similar to the Fourier transform. Certain properties are easier to handle in one setting, while others are easier after a transform. We will go back and forth.

In subsection 2.10 we give a more detailed account of the strategies used and a proof outline. Among the techniques used are a derivative operator ∂, a Tableau Calculus, and arity reductions. An overall philosophy is that various tractable constraint functions of different families cannot mix. Then the truth of Theorem 1.1 itself, precisely because it is such a complete statement without any exceptions, guides the choices made in various constructions. As a proof strategy, this is pretty dicey or at least self-serving. Essentially we want the validity of the very statement we want to prove to provide its own guarantee of success in every step in its proof. If there were other tractable problems, e.g., as in the case of planar Holant [6], where different classes of tractable constraints can indeed mix, then we would be stuck. Luckily, the vision is correct for planar #CSP. And therefore, the self-serving plan becomes a

reliable guide to the proof—a bit self-fulfilling.

2. Preliminaries.

2.1. Problems and definitions. In this paper, $\mathfrak{i} = e^{\mathfrak{i}\frac{\pi}{2}}$ denotes a square root of -1; thus $\mathfrak{i}^2 = -1$. The symbol e_i denotes the string where the ith bit is 1, and all other bits are 0. In a string $a_1 \cdots \hat{a}_i \cdots a_n \in \{0,1\}^{n-1}$, $\hat{a}_i$ means that the ith bit a_i is deleted. For $\alpha, \beta \in \{0,1\}^n$, $\alpha \oplus \beta$ denotes the bitwise XOR of α and β.

Even though our focus in this paper is on planar counting #CSP problems, most of the proof needs to be carried out in the framework of Holant problems [13, 17]. A Holant problem is specified by a set of local constraint functions, also called *signatures*. In this paper, we investigate complex-valued planar #CSP problems over Boolean variables, and thus all signatures in the corresponding Holant problems are of the form $\{0,1\}^n \to \mathbb{C}$. For consideration of models of computation, functions output complex algebraic values. If a function takes n input variables, it is said to be of arity n; we also use $f_{x_1 x_2 \cdots x_n}$ to denote $f(x_1, x_2, \ldots, x_n)$.

Graphs may have self-loops and parallel edges. A graph without self-loops or parallel edges is a *simple* graph. Fix a set of local constraint functions $\mathcal{F}$. A *signature grid* $\Omega = (G, \pi)$ consists of a graph $G = (V, E)$, and a mapping π which maps each vertex $v \in V$ to some $f_v \in \mathcal{F}$ of arity $\deg(v)$, and maps its incident edges $E(v)$ bijectively to the input variables of f_v. We say that Ω is a *planar signature grid* if G is a plane graph, and in addition to the vertex map, π specifies one edge of $E(v)$ to be the first input variable to f_v. All other edges of $E(v)$ are ordered counterclockwise from the planar embedding starting from that edge and map to the ordered list of input variables of f_v. The Holant problem on instance Ω is to evaluate

$$\text{Holant}(\Omega; \mathcal{F}) = \sum_{\sigma: E \to \{0,1\}} \prod_{v \in V} f_v(\sigma \mid_{E(v)}),$$

where $\sigma \mid_{E(v)}$ denotes the restriction of σ to $E(v)$. We write G in place of Ω when π is clear from the context.

A signature f of arity n can be specified by listing its values in lexicographical order as in a truth table, which is a vector in $\mathbb{C}^{2^n}$, or as a tensor in $(\mathbb{C}^2)^{\otimes n}$. A symmetric signature f of arity n takes values depending only on the Hamming weight of the input and can be expressed as $[f_0, f_1, \ldots, f_n]$, where f_w is the value of f on inputs of Hamming weight w. An example is the EQUALITY signature $(=_n) = [1, 0, \ldots, 0, 1]$ of arity n. Another example is the EXACT-ONE signature $[0, 1, \ldots, 0, 0]$ corresponding to the PERFECT MATCHING constraint. Freedman, Lovász, and Schrijver [25] proved that counting the number of perfect matchings cannot be expressed as a partition function of graph homomorphism with positive real vertex weights and real edge weights. This is a special case of #CSP (the definitions are in subsection 2.3). However, in this paper, we will solve problems in #CSP on planar instances by holographic reductions to the problem of counting perfect matchings.

A Holant problem is parametrized by a set of signatures.

DEFINITION 2.1. *Given a set of signatures $\mathcal{F}$, we define the counting problem* Holant($\mathcal{F}$) *as follows:*

Input: A signature grid $\Omega = (G, \pi)$*;*

Output: Holant($\Omega; \mathcal{F}$)*.*

The problem Pl-Holant($\mathcal{F}$) *is defined similarly using a planar signature grid.*

A signature f of arity n is *degenerate* if there exist unary signatures $u_j \in \mathbb{C}^2$ $(1 \le j \le n)$ such that $f = u_1 \otimes \cdots \otimes u_n$. Using a degenerate signature is equivalent to

replacing it by its n unary signatures, each on its corresponding edge. A symmetric degenerate signature has the form $u^{\otimes n}$. Replacing a signature $f \in \mathcal{F}$ by a constant multiple cf, where $c \neq 0$, does not change the complexity of Holant($\mathcal{F}$). It introduces a factor c^m to Holant($\Omega; \mathcal{F}$), where f occurs m times in Ω.

We allow $\mathcal{F}$ to be an infinite set. For Holant($\mathcal{F}$) or Pl-Holant($\mathcal{F}$) to be tractable, the problem must be computable in polynomial time even when the description of the signatures in the input Ω is included in the input size. On the other hand, we say Holant($\mathcal{F}$) or Pl-Holant($\mathcal{F}$) is #P-hard if there exists a finite subset of $\mathcal{F}$ for which the problem is #P-hard. In this paper we focus on planar problems, and so we say a signature set $\mathcal{F}$ is tractable (resp., #P-hard) if the corresponding counting problem Pl-Holant($\mathcal{F}$) is tractable (resp., #P-hard). For a signature f, we say f is tractable (resp., #P-hard) if $\{f\}$ is. We follow the usual conventions about polynomial-time Turing reduction $\leq_T$ and polynomial-time Turing equivalence $\equiv_T$.

2.2. Holographic reduction. To introduce the idea of holographic reductions, it is convenient to consider bipartite graphs. For a general graph, we can always transform it into a bipartite graph while preserving the Holant value, as follows. For each edge in the graph, we replace it by a path of length two. (This operation is called the 2-*stretch* of the graph and yields the edge-vertex incidence graph.) Each new vertex is assigned the binary EQUALITY signature $(=_2) = [1, 0, 1]$.

We use Holant $(\mathcal{F} \mid \mathcal{G})$ to denote the Holant problem over signature grids with a bipartite graph $H = (U, V, E)$, where each vertex in U or V is assigned a signature in $\mathcal{F}$ or $\mathcal{G}$, respectively. Signatures in $\mathcal{F}$ are considered as row vectors (or covariant tensors); signatures in $\mathcal{G}$ are considered as column vectors (or contravariant tensors). We denote by Holant$(\Omega; \mathcal{F} \mid \mathcal{G})$ the value of the Holant problem Holant $(\mathcal{F} \mid \mathcal{G})$ on the signature grid Ω. Similarly, Pl-Holant $(\mathcal{F} \mid \mathcal{G})$ denotes the Holant problem over signature grids with a planar bipartite graph, and Pl-Holant$(\Omega; \mathcal{F} \mid \mathcal{G})$ its value on the signature grid Ω.

For an invertible 2×2 matrix $T \in \mathrm{GL}_2(\mathbb{C})$ and a signature f of arity n, written as a column vector (contravariant tensor) $f \in \mathbb{C}^{2^n}$, we denote by $T^{-1}f = (T^{-1})^{\otimes n} f$ the transformed signature. For a signature set $\mathcal{F}$, define $T^{-1}\mathcal{F} = \{T^{-1}f \mid f \in \mathcal{F}\}$ as the set of transformed signatures. For signatures written as row vectors (covariant tensors) we define $\mathcal{F}T$ similarly. Whenever we write $T^{-1}f$ or $T^{-1}\mathcal{F}$, we view the signatures as column vectors; similarly, whenever we write fT or $\mathcal{F}T$, we view them as row vectors. In the special case of the Hadamard matrix $H_2 = \frac{1}{\sqrt{2}} \left[\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}\right]$, we also define $\widehat{\mathcal{F}} = H_2\mathcal{F}$. Note that $H_2 = H_2^{-1} = H_2^{\mathsf{T}}$ is orthogonal. Since constant factors are immaterial, for convenience we sometimes drop the factor $\frac{1}{\sqrt{2}}$ when using H_2.

Let $T \in \mathrm{GL}_2(\mathbb{C})$. The holographic transformation defined by T is the following operation: given a signature grid $\Omega = (H, \pi)$ of Holant $(\mathcal{F} \mid \mathcal{G})$, for the same bipartite graph H, we get a new grid $\Omega' = (H, \pi')$ of Holant $\left(\mathcal{F}T \mid T^{-1}\mathcal{G}\right)$ by replacing each signature in $\mathcal{F}$ or $\mathcal{G}$ with the corresponding signature in $\mathcal{F}T$ or $T^{-1}\mathcal{G}$.

THEOREM 2.2 (Valiant's Holant theorem [46]). *For any $T \in \mathrm{GL}_2(\mathbb{C})$,*

$$\mathrm{Holant}(\Omega; \mathcal{F} \mid \mathcal{G}) = \mathrm{Holant}(\Omega'; \mathcal{F}T \mid T^{-1}\mathcal{G}).$$

Therefore, a holographic transformation does not change the complexity of the Holant problem in the bipartite setting.

2.3. Counting constraint satisfaction problems and Holant$(\widehat{\mathcal{EQ}}, \widehat{\mathcal{F}})$. Counting constraint satisfaction problems (#CSP) can be defined as a special case of Holant problems. An instance of #CSP($\mathcal{F}$) is presented as a bipartite graph. There is

one node for each variable and for each occurrence of constraint functions, respectively. Connect a constraint node to a variable node if the variable appears in that occurrence of constraint, with a labeling on the edges for the order of these variables. This bipartite graph is also known as the *constraint graph*. If we label each variable node with an EQUALITY function and consider every edge as a variable, then the #CSP problem is just the Holant problem on this bipartite graph. Thus $\#\mathrm{CSP}(\mathcal{F}) \equiv_T \text{Holant}\,(\mathcal{EQ} \mid \mathcal{F})$, where $\mathcal{EQ} = \{=_1, =_2, =_3, \dots\}$ is the set of EQUALITY signatures of all arities. By restricting to planar constraint graphs, we have the planar #CSP framework, which we denote by Pl-#CSP. The construction above also shows that $\text{Pl-}\#\mathrm{CSP}(\mathcal{F}) \equiv_T \text{Pl-Holant}\,(\mathcal{EQ} \mid \mathcal{F})$.

For any positive integer d, the problem $\#\mathrm{CSP}^d(\mathcal{F})$ is the same as $\#\mathrm{CSP}(\mathcal{F})$ except that every variable appears a multiple of d times. Thus,

$$\text{Pl-}\#\mathrm{CSP}^d(\mathcal{F}) \equiv_T \text{Pl-Holant}\,(\mathcal{EQ}_d \mid \mathcal{F}), \tag{2.1}$$

where $\mathcal{EQ}_d = \{=_d, =_{2d}, =_{3d}, \dots\}$ is the set of EQUALITY signatures of arities that are multiples of d. For $d = 1$, we just have #CSP problems. For $d = 2$, these are #CSP problems where every variable appears an even number of times. If $d \in \{1, 2\}$, then we further have

$$\text{Pl-}\#\mathrm{CSP}^d(\mathcal{F}) \equiv_T \text{Pl-Holant}\,(\mathcal{EQ}_d \mid \mathcal{F}) \equiv_T \text{Pl-Holant}(\mathcal{EQ}_d, \mathcal{F}). \tag{2.2}$$

The first equivalence is by (2.1). The reduction from left to right in the second equivalence is trivial. For the other direction of the second equivalence, we take a signature grid for $\text{Pl-Holant}(\mathcal{EQ}_d, \mathcal{F})$ and create a bipartite signature grid for $\text{Pl-Holant}\,(\mathcal{EQ}_d \mid \mathcal{F})$ such that both signature grids have the same Holant value up to an easily computable factor. If two signatures in $\mathcal{F}$ are assigned to adjacent vertices, then we subdivide all edges between them and assign the binary EQUALITY signature $(=_2) \in \mathcal{EQ}_d$ to all new vertices. Suppose EQUALITY signatures $(=_n), (=_m) \in \mathcal{EQ}_d$ are assigned to adjacent vertices connected by k edges. If $n = m = k$, then we simply remove these two vertices. The Holant of the resulting signature grid differs from the original by a factor of 2. Otherwise, we contract all k edges, merge the two vertices, and assign $(=_{n+m-2k}) \in \mathcal{EQ}_d$ to the new vertex.

By the holographic transformation defined by the matrix $H_2 = \frac{1}{\sqrt{2}} \left[\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}\right]$ (or equivalently without the nonzero factor $\frac{1}{\sqrt{2}}$ since this does not affect the complexity), we have

$$\text{Pl-}\#\mathrm{CSP}(\mathcal{F}) \equiv_T \text{Pl-Holant}(\widehat{\mathcal{EQ}}, \widehat{\mathcal{F}}), \tag{2.3}$$

where $\widehat{\mathcal{EQ}} = \{[1, 0], [1, 0, 1], [1, 0, 1, 0], \dots\}$ (where we ignore nonzero factors) and $\widehat{\mathcal{F}} = H_2\mathcal{F}$. Note that (2.3) holds for both the bipartite expression $\text{Pl-Holant}(\widehat{\mathcal{EQ}} \mid \widehat{\mathcal{F}})$ as well as the nonbipartite expression $\text{Pl-Holant}(\widehat{\mathcal{EQ}}, \widehat{\mathcal{F}})$ on the RHS. To see that, the bipartite expression follows from Theorem 2.2 and the fact that $H_2^{-1} = H_2$. For the nonbipartite expression, by (2.2) we have

$$\text{Pl-}\#\mathrm{CSP}(\mathcal{F}) \equiv_T \text{Pl-Holant}(\mathcal{EQ}, \mathcal{F}) \equiv_T \text{Pl-Holant}\,(=_2 \mid \mathcal{EQ}, \mathcal{F}) \equiv_T \text{Pl-Holant}\left(=_2 \mid \widehat{\mathcal{EQ}}, \widehat{\mathcal{F}}\right),$$

where we used the fact that H_2 is orthogonal, and hence $(=_2)H_2^{\otimes 2} = (=_2)$. This equivalence (2.3) plays a central role in our proof.

The next lemma shows that if we have $(=_4)$ in $\text{Pl-Holant}(\widehat{\mathcal{EQ}}, \widehat{\mathcal{F}})$, then we can construct $(=_{2k})$ for any $k \in \mathbb{Z}^+$.

LEMMA 2.3.

$$\text{Pl-Holant}(\widehat{\mathcal{EQ}}, \mathcal{EQ}_2, \widehat{\mathcal{F}}) \equiv_{\text{T}} \text{Pl-Holant}(\widehat{\mathcal{EQ}}, (=_4), \widehat{\mathcal{F}}).$$

Proof. One direction is trivial since $(=_4) \in \mathcal{EQ}_2$. For the other direction we use induction. For $k = 1$, we have $(=_2) \in \widehat{\mathcal{EQ}}$. For $k = 2$, we have $(=_4)$ given. Assume that we have $(=_{2(k-1)})$. Then, connecting $(=_{2(k-1)})$ and $(=_4)$ by one edge, we get $(=_{2k})$. ∎

By (2.2), we have

$$\text{Pl-\#CSP}^2(\widehat{\mathcal{EQ}}, \widehat{\mathcal{F}}) \equiv_T \text{Pl-Holant}(\widehat{\mathcal{EQ}}, \mathcal{EQ}_2, \widehat{\mathcal{F}}).$$

Thus Lemma 2.3 implies that

$$\text{Pl-\#CSP}^2(\widehat{\mathcal{EQ}}, \widehat{\mathcal{F}}) \leq_T \text{Pl-Holant}(\widehat{\mathcal{EQ}}, (=_4), \widehat{\mathcal{F}}).$$

DEFINITION 2.4. *The crossover function $\mathfrak{X}$ is a signature of arity 4 which satisfies $f_{0000} = f_{1111} = f_{0101} = f_{1010} = 1$ and $f_\alpha = 0$ for all other $\alpha \in \{0,1\}^4$.*

The crossover function $\mathfrak{X}$ on (x_1, x_2, x_3, x_4) is the tensor product of two binary EQUALITY functions $(=_2)$ on (x_1, x_3) and on (x_2, x_4). If we can obtain $\mathfrak{X}$ (by some construction or reduction) in Pl-\#CSP($\mathcal{F}$), then we can reduce \#CSP($\mathcal{F}$) to Pl-\#CSP($\mathcal{F}$). The same is true for Holant($\widehat{\mathcal{EQ}}, \widehat{\mathcal{F}}$) and Pl-Holant($\widehat{\mathcal{EQ}}, \widehat{\mathcal{F}}$). Moreover, note that $H_2^{\otimes 4}(\mathfrak{X}) = \mathfrak{X}$, because an orthogonal transformation does not change a binary EQUALITY function $(=_2)$. So we can obtain $\mathfrak{X}$ in Pl-\#CSP($\mathcal{F}$) iff we can obtain $\mathfrak{X}$ in Pl-Holant($\widehat{\mathcal{EQ}}, \widehat{\mathcal{F}}$).

2.4. Realization. One basic notion used throughout the paper is realization. We say a signature f is *realizable* or *constructible* from a signature set $\mathcal{F}$ if (informally speaking) there is a gadget with some dangling edges such that each vertex is assigned a signature from $\mathcal{F}$, and the resulting graph, when viewed as a black-box signature with inputs on the dangling edges, is exactly f. If f is realizable from a set $\mathcal{F}$, then we can freely add f into $\mathcal{F}$ while preserving the complexity.

Formally, this notion is defined by an $\mathcal{F}$-gate. An $\mathcal{F}$-gate is similar to a signature grid (G, π) for Holant($\mathcal{F}$) except that $G = (V, E, D)$ is a graph with some dangling edges D. The dangling edges define external variables for the $\mathcal{F}$-gate. (See Figure 1 for an example.) We denote the ordinary edges in E by $1, 2, \ldots, m$ and the dangling edges in D by $m+1, \ldots, m+n$. Then we can define a function f for this $\mathcal{F}$-gate as

$$f(y_1, \ldots, y_n) = \sum_{x_1, \ldots, x_m \in \{0,1\}} H(x_1, \ldots, x_m, y_1, \ldots, y_n),$$

where $(y_1, \ldots, y_n) \in \{0,1\}^n$ is an assignment on the dangling edges and $H(x_1, \ldots, x_m, y_1, \ldots, y_n)$ is the value of the signature grid on an assignment of all edges in G, which is the product of evaluations at all vertices in V. We also call this function f the signature of the $\mathcal{F}$-gate.

An $\mathcal{F}$-gate is called planar if the underlying graph G is a planar graph, and the dangling edges, ordered counterclockwise corresponding to the order of the input variables, are in the outer face in a planar embedding. A planar $\mathcal{F}$-gate can be used in a planar signature grid as if it is just a single vertex with the particular signature.

PROPOSITION 2.5. *If $\mathcal{F}$ is a signature set and g is the signature of some planar $\mathcal{F}$-gate, then*

$$\text{Pl-Holant}(\mathcal{F}, g) \leq_T \text{Pl-Holant}(\mathcal{F}).$$

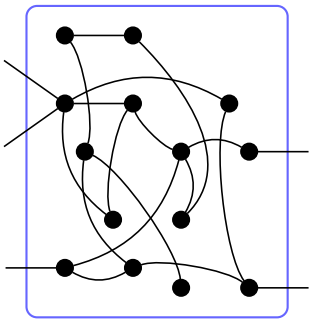


FIG. 1. *An $\mathcal{F}$-gate with five dangling edges.*

Proof. The reduction is simple. Given any signature grid Ω as an instance of Pl-Holant$(\mathcal{F}, g)$, by replacing every occurrence of g by the $\mathcal{F}$-gate, we get an instance Ω' of Pl-Holant$(\mathcal{F})$. Since the $\mathcal{F}$-gate for g has constant size, the size of Ω' is at most linear in the size of Ω. Since the signature of the $\mathcal{F}$-gate is g, the Holant values for these two signature grids are the same. ∎

When a gadget has an asymmetric signature, we place a diamond on the edge corresponding to the first input. The remaining inputs are ordered counterclockwise around the vertex.

DEFINITION 2.6 (derivative). *Let f and g be two signatures of arities n and m, respectively, and $n > m$. We connect all m input edges $(1 \le j \le m)$ of g to m consecutive edges of f in a clockwise order, indexed $i-1, \dots, i-m \pmod n$. The* derivative signature $\partial_g^{\{i-1,\dots,i-m\}}(f)$ *is the signature of this planar $\{f, g\}$-gate of arity $n-m$, whose variables are the unmatched variables of f in the original counterclockwise order starting with the first unmatched variable. (The clockwise order of edges of f to be matched with the counterclockwise order of edges of g ensures planarity.)*

If f is symmetric, we will simply write $\partial_g(f)$ since the derivative signature is independent of the choice of i in this case. Moreover, if $kn < m$ and we connect k copies of g to f, which is the same as forming $\partial_g(f)$ sequentially k times, the resulting repeated derivative signature *is denoted by $\partial_g^k(f)$.*

For a unary signature u, we can connect a copy of u to each edge of f indexed by a subset $S \subset [n]$, and we also denote the resulting signature by $\partial_u^S(f)$.

For convenience, we use $f^{x_i=0}$ to denote $\partial_{[1,0]}^{\{i\}} f$ and $f^{x_i=1}$ to denote $\partial_{[0,1]}^{\{i\}} f$.

For a signature f of arity n, we can partition its variables $\{x_1, x_2, \dots, x_n\}$ into $\{x_{i_1}, x_{i_2}, \dots, x_{i_s}\}$ and $\{x_{i_{s+1}}, x_{i_{s+2}}, \dots, x_{i_n}\}$ and then list the values of f as a $2^s \times 2^{n-s}$ matrix $M(f)$ with the entry $f(x_1, x_2, \dots, x_n)$ indexed by row index $x_{i_1} x_{i_2} \cdots x_{i_s}$ and column index $x_{i_{s+1}} x_{i_{s+2}} \cdots x_{i_n}$. The rows and columns are ordered lexicographically. We also denote the matrix by $M_{x_{i_1} \cdots x_{i_s}, x_{i_{s+1}} \cdots x_{i_n}}(f)$ when we need to specify the names of the variables. We call this matrix a signature matrix of f. For example, we use the signature matrix

$$M_{x_1,x_2x_3}(f) = \begin{bmatrix} f_{000} & f_{001} & f_{010} & f_{011} \\ f_{100} & f_{101} & f_{110} & f_{111} \end{bmatrix}$$

to denote a ternary signature, and we use the signature matrix

$$M_{x_1x_2,x_4x_3}(f) = \begin{bmatrix} f_{0000} & f_{0010} & f_{0001} & f_{0011} \\ f_{0100} & f_{0110} & f_{0101} & f_{0111} \\ f_{1000} & f_{1010} & f_{1001} & f_{1011} \\ f_{1100} & f_{1110} & f_{1101} & f_{1111} \end{bmatrix}$$

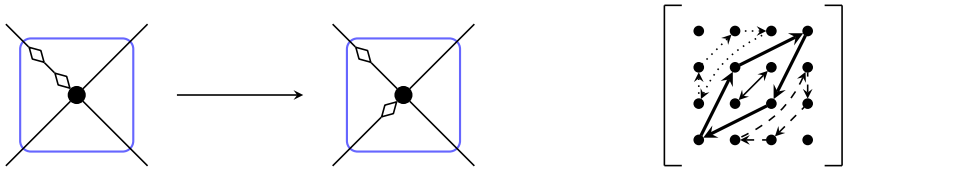
to denote a signature of arity 4. Note that in $M_{x_1x_2,x_4x_3}(f)$, the rows are indexed by x_1x_2, and the columns are indexed by x_4x_3 (not x_3x_4), both in lexicographic order. This reversal of column index ensures that the signature matrix of linking two arity 4 signatures in a planar setting is simply the matrix product of the two signature matrices.

For example, the signature matrix of the crossover function $\mathfrak{X}$ is

$$M_{x_1x_2,x_4x_3}(\mathfrak{X}) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

where the nonzero values are $f_{0000} = f_{0101} = f_{1010} = f_{1111} = 1$, with the four variables ordered in counterclockwise cyclic order.

When we rotate a signature, the transformation of its signature matrix is depicted in Figure 2.



(a) A counterclockwise rotation (b) Movement of signature matrix entries

FIG. 2. *The movement of the entries in the signature matrix of a signature of arity* 4 *under a counterclockwise rotation of the input edges. Entries of Hamming weight* 1 *are in the dotted cycle, entries of Hamming weight* 2 *are in the two solid cycles (one has length* 4 *and the other one is a swap), and entries of Hamming weight* 3 *are in the dashed cycle.*

2.5. Tractable signature sets. We define some signatures that are known to be tractable [8, 27]. These form three families: affine signatures, product-type signatures, and matchgate signatures.

Affine signatures.

DEFINITION 2.7. *For a signature f of arity n, the support of f is*

$$\mathrm{supp}(f) = \{(x_1, x_2, \ldots, x_n) \in \mathbb{Z}_2^n \mid f(x_1, x_2, \ldots, x_n) \neq 0\}.$$

DEFINITION 2.8. *Let f be a signature of arity n. We say f has affine support of dimension k if* $\mathrm{supp}(f)$ *is an affine subspace of dimension k over $\mathbb{Z}_2$, i.e., there is a matrix A over $\mathbb{Z}_2$ such that $f(x_1, x_2, \ldots, x_n) \neq 0$ iff $AX = 0$, where $X = (x_1, x_2, \ldots, x_n, 1)$ and the affine space $\{(x_1, x_2, \ldots, x_n) \in \mathbb{Z}_2^n \mid AX = 0\}$ has dimension k.*

For a signature of arity n with affine support of dimension k, let $X = \{x_{i_1}, x_{i_2}, \ldots, x_{i_k}\}$ be a set of free variables where $i_1 < i_2 < \cdots < i_k$. Then on $\mathrm{supp}(f)$, $f(x_1, x_2, \ldots, x_n)$ is uniquely determined by the input on X.

DEFINITION 2.9. *If f has affine support of dimension k, and $X = \{x_{i_1}, x_{i_2}, \ldots, x_{i_k}\}$ is a set of k free variables, then we define the* compressed signature $\underline{f_X}$ *of f for X to be a signature of arity k, such that $\underline{f_X}(x_{i_1}, \ldots, x_{i_k}) = f(x_1, x_2, \ldots, x_n)$, where $(x_1, x_2, \ldots, x_n) \in \mathrm{supp}(f)$. When it is clear from the context, we omit X and use $\underline{f}$ to denote $\underline{f_X}$.*

DEFINITION 2.10. *A signature $f(x_1, \ldots, x_n)$ of arity n is* affine *if the following hold:*

- *it has affine support;*
- *if $(x_1, x_2, \ldots, x_n) \in \mathrm{supp}(f)$, then $f(x_1, x_2, \ldots, x_n) = \lambda \mathrm{i}^{Q(X)}$, where $X = (x_1, x_2, \ldots, x_n, 1)$, $\lambda \in \mathbb{C}$, and $Q(x_1, x_2, \ldots, x_n) \in \mathbb{Z}_4[x_1, x_2, \ldots, x_n]$ is a quadratic (total degree at most 2) multilinear polynomial with the additional requirement that the coefficients of all cross terms are even, i.e., Q has the form*

$$Q(x_1, x_2, \ldots, x_n) = a_0 + \sum_{k=1}^{n} a_k x_k + \sum_{1 \le i < j \le n} 2b_{ij} x_i x_j.$$

We use $\mathscr{A}$ to denote the set of all affine signatures.

Note that to say a signature is affine requires more than it having an affine support. Historically, Dyer, Goldberg, and Jerrum [22] called a signature *pure affine* when it is a constant multiple of the characteristic function of an affine linear subspace. When the functions take nonnegative values, this notion coincides with the definition of affine signatures given here. In the general setting, the requirement that there be a quadratic polynomial with even coefficients for all cross terms is more subtle; but it is necessary to make Theorems 2.33 and 2.34 work.

In [20], there is an alternative definition for affine signatures.

DEFINITION 2.11. *A signature $f(x_1, x_2, \ldots, x_n)$ of arity n is* affine *if it has the form*

$$\lambda \cdot \chi_{AX=0} \cdot \mathrm{i}^{\sum_{j=1}^{k} \langle \mathbf{v}_j, X \rangle},$$

where $\lambda \in \mathbb{C}$, $X = (x_1, x_2, \ldots, x_n, 1)$, A is a matrix over $\mathbb{Z}_2$, $\mathbf{v}_j$ is a vector over $\mathbb{Z}_2$, and χ is a 0-1 *indicator signature such that $\chi_{AX=0}$ is 1 iff $AX = 0$. Note that the dot product $\langle \mathbf{v}_j, X \rangle$ is calculated over $\mathbb{Z}_2$ with a* 0-1 *output in $\mathbb{Z}$, while the summation $\sum_{j=1}^{k}$ on the exponent of $\mathrm{i} = \sqrt{-1}$ is evaluated as a sum mod 4 of* 0-1 *terms.*

Definitions 2.10 and 2.11 are equivalent. To see this we observe that each $\langle \mathbf{v}_j, X \rangle$ as an integer sum (mod 2) can be replaced by $(\langle \mathbf{v}_j, X \rangle)^2$ as an integer sum (mod 4) since $N \equiv 0, 1 \pmod 2$ iff $N^2 \equiv 0, 1 \pmod 4$, respectively, for any integer N. After this, all cross terms have even coefficients and all square terms x_s^2 can be replaced by x_s since $x_s = 0, 1$. Conversely, we can express $Q \pmod 4$ as a sum of squares of affine forms of X, using the condition that all cross terms have even coefficients.

The following lemma shows that for a $\{\pm 1, \pm \mathrm{i}\}$-valued signature of arity k, there exists a unique multilinear polynomial $P(x_1, \ldots, x_k) \mod 4$ such that $f(x_1, \ldots x_k) = \mathrm{i}^{P(x_1, \ldots, x_k)}$. Thus if there exists a multilinear polynomial $P(x_1, \ldots, x_k)$ such that $f(x_1, \ldots x_k) = \mathrm{i}^{P(x_1, \ldots, x_k)}$, and P has total degree greater than 2 or has a cross term with an odd coefficient, then $f \notin \mathscr{A}$.

LEMMA 2.12. *Let f be a signature of arity k taking values in $\{\pm 1, \pm \mathrm{i}\}$. Then there exists a unique multilinear polynomial $P(x_1, \ldots, x_k) \in \mathbb{Z}_4[x_1, \ldots, x_k]$ such that $f(x_1, \ldots x_k) = \mathrm{i}^{P(x_1,\ldots,x_k)}$.*

Similarly, if f is a signature of arity k taking values in $\{\pm 1\}$, then there exists a unique multilinear polynomial $P(x_1, \ldots, x_k) \in \mathbb{Z}_2[x_1, \ldots, x_k]$ such that $f(x_1, \ldots x_k) = (-1)^{P(x_1,\ldots,x_k)}$.

Proof. We prove the first statement. The proof for the second is similar, and so we omit it here.

For any input $\alpha = a_1 \ldots a_k \in \{0,1\}^k$, there exists $r_\alpha \in \{0,1,2,3\}$ such that $f_\alpha = \mathrm{i}^{r_\alpha}$ since f takes values in $\{\pm 1, \pm \mathrm{i}\}$. Let $P(x_1, \ldots, x_k) = \sum_{\alpha \in \{0,1\}^k} r_\alpha \prod_{i=1}^k \widetilde{x_i} \in \mathbb{Z}_4[x_1, \ldots, x_k]$, where $\alpha = a_1 \ldots a_k$ and $\widetilde{x_i} = x_i$ if $a_i = 1$ and $1 - x_i$ if $a_i = 0$. Then $f(x_1, \ldots x_k) = \mathrm{i}^{P(x_1,\ldots,x_k)}$.

Now we prove that $P(x_1, \ldots, x_k)$ is unique. It is equivalent to prove that if f is the constant 1 function, then $P(x_1, \ldots, x_k) = 0$ in $\mathbb{Z}_4[x_1, \ldots, x_k]$. For a contradiction, suppose $r \prod_{i \in S} x_i$ is a nonzero term in $P(x_1, \ldots, x_k)$ with minimum $|S|$. Set $x_i = 1$ for all $i \in S$ and all other $x_i = 0$. Then P evaluates to $r \neq 0$ in $\mathbb{Z}_4$, and f evaluates to $\mathrm{i}^r \neq 1$. This is a contradiction. ∎

By Lemma 2.12, we directly have the following corollary.

COROLLARY 2.13. *Let $f \in \mathscr{A}$ be a signature of arity n with affine support of dimension k. Let $X = \{x_{i_1}, \ldots, x_{i_k}\}$ be a set of free variables. Then there exists a unique $Q(X) \in \mathbb{Z}_4[X]$ such that*

$$f(x_1, x_2, \ldots, x_n) = \mathrm{i}^{Q(X)}$$

for $(x_1, x_2, \ldots, x_n) \in \mathrm{supp}(f)$, where $Q(X)$ is a quadratic multilinear polynomial and the coefficients of cross terms are even.

COROLLARY 2.14. *Let f be a signature of arity n having affine support of dimension k. Suppose f takes values in $\{0, \pm 1, \pm \mathrm{i}\}$, and $X = \{x_{i_1}, \ldots, x_{i_k}\}$ is a set of free variables. Then $f \in \mathscr{A}$ iff $\underline{f} \in \mathscr{A}$, where $\underline{f}$ is the compressed signature of f for X.*

Proof. Note that if $\underline{f}(x_{i_1}, \ldots, x_{i_k}) = \mathrm{i}^{Q(x_{i_1},\ldots,x_{i_k})}$, then $f(x_1, \ldots, x_n) = \mathrm{i}^{Q(x_{i_1},\ldots,x_{i_k})}$ for $(x_1, \ldots, x_n) \in \mathrm{supp}(f)$. So $f \in \mathscr{A}$ iff $\underline{f} \in \mathscr{A}$. ∎

The next two lemmas allow us to easily determine if a binary or ternary signature is affine.

LEMMA 2.15. *Let f be a binary signature and $M_{x_1,x_2}(f) = \begin{bmatrix} f_{00} & f_{01} \\ f_{10} & f_{11} \end{bmatrix} = \begin{bmatrix} 1 & b \\ c & d \end{bmatrix} = \begin{bmatrix} 1 & \mathrm{i}^r \\ \mathrm{i}^s & d \end{bmatrix}$, where $r, s \in \{0,1,2,3\}$. Then $f \in \mathscr{A}$ iff $d = \pm \mathrm{i}^{r+s}$. In particular, if $b, c, d \in \{1, -1\}$, then $f \in \mathscr{A}$.*

Proof. Let $Q(x_1, x_2) = sx_1 + rx_2$ if $d = \mathrm{i}^{r+s}$, and $sx_1 + rx_2 + 2x_1x_2$ if $d = -\mathrm{i}^{r+s}$. Then $f(x_1, x_2) = \mathrm{i}^{Q(x_1,x_2)}$. Thus $f \in \mathscr{A}$ by Definition 2.10. Conversely, if $f(x_1, x_2) = \mathrm{i}^{Q(x_1,x_2)}$, for some $Q(x_1, x_2) = a_0 + a_1x_1 + a_2x_2 + 2b_{12}x_1x_2 \in \mathbb{Z}_4[x_1, x_2]$, then we have $a_0 = 0$ by $f_{00} = 1$, $a_1 = s$ by $f_{10} = \mathrm{i}^s$, and $a_2 = r$ by $f_{01} = \mathrm{i}^r$. Thus $f_{11} = \mathrm{i}^{r+s+2b_{12}} = \pm \mathrm{i}^{r+s}$. ∎

LEMMA 2.16. *Let f be a ternary signature and*

$$M_{x_1,x_2x_3}(f) = \begin{bmatrix} f_{000} & f_{001} & f_{010} & f_{011} \\ f_{100} & f_{101} & f_{110} & f_{111} \end{bmatrix} = \begin{bmatrix} 1 & \mathrm{i}^r & \mathrm{i}^s & \epsilon_1 \mathrm{i}^{r+s} \\ \mathrm{i}^t & \epsilon_2 \mathrm{i}^{r+t} & \epsilon_3 \mathrm{i}^{s+t} & \epsilon_4 \mathrm{i}^{r+s+t} \end{bmatrix},$$

where $r, s, t \in \{0,1,2,3\}$ and $\epsilon_i \in \{1, -1\}$ for $1 \leq i \leq 4$. Then $f \in \mathscr{A}$ iff $\epsilon_1\epsilon_2\epsilon_3\epsilon_4 = 1$.

Proof. We can write $\epsilon_i = (-1)^{a_i}$ for $a_i \in \{0,1\}$ and let

$$Q = tx_1+sx_2+rx_3+2a_1x_2x_3+2a_2x_3x_1+2a_3x_1x_2+2(a_1+a_2+a_3+a_4)x_1x_2x_3 \in \mathbb{Z}_4[x_1,x_2,x_3].$$

Then $f(x_1,x_2,x_3) = \mathfrak{i}^{Q(x_1,x_2,x_3)}$.

By Lemma 2.12, $f \in \mathscr{A}$ iff Q is a multilinear quadratic polynomial and the coefficients of the cross terms are even. Thus $f \in \mathscr{A}$ iff $2(a_1+a_2+a_3+a_4) \equiv 0 \pmod 4$. This is equivalent to $a_1+a_2+a_3+a_4 \equiv 0 \pmod 2$, i.e., $\epsilon_1\epsilon_2\epsilon_3\epsilon_4 = 1$. ∎

In addition, we often use the following facts, which can be derived from Definition 2.10 directly.

PROPOSITION 2.17. *The following hold:*

- *For any signature $f \in \mathscr{A}$, up to a nonzero factor, all nonzero entries are powers of $\mathfrak{i}$. In particular, they have the same norm.*
- *If a signature has only one nonzero entry, then it is affine. In particular, $[1,0],[0,1] \in \mathscr{A}$.*
- *If a signature f has only two nonzero entries f_α and f_β, then the support of f is affine. Moreover, in this case $f \in \mathscr{A}$ iff $f_\alpha^4 = f_\beta^4$.*
- *$[1,a],[1,0,a]$ are affine iff $a^4 = 0$ or 1.*

The following lemma is useful in proving arity reductions for nonaffine signatures.

LEMMA 2.18. *Let f be a signature of arity n with affine support of dimension $k \geq 4$. If $f^{x_i=0} \in \mathscr{A}$ and $f^{x_i=1} \in \mathscr{A}$ for all $1 \leq i \leq n$, then $f \in \mathscr{A}$.*

Proof. Let $X = \{x_{i_1}, x_{i_2}, \ldots, x_{i_k}\}$ be a set of free variables of f, and let $\underline{f}$ be the compressed signature of f for X. Since both $f^{x_{i_1}=0}$ and $f^{x_{i_1}=1}$ are affine, $\underline{f}^{x_{i_1}=0}$ and $\underline{f}^{x_{i_1}=1}$ are affine by Corollary 2.14. By Corollary 2.13 there exist $Q_0(x_{i_2},\ldots,x_{i_k})$ and $Q_1(x_{i_2},\ldots,x_{i_k})$ such that $\underline{f}(0,x_{i_2},\ldots,x_{i_k}) = \mathfrak{i}^{Q_0(x_{i_2},\ldots,x_{i_k})}$ and $\underline{f}(1,x_{i_2},\ldots,x_{i_k}) = \mathfrak{i}^{Q_1(x_{i_2},\ldots,x_{i_k})}$, where Q_0 and Q_1 are quadratic multilinear polynomials in $\mathbb{Z}_4[x_{i_2},\ldots,x_{i_k}]$, and the coefficients of all cross terms are even.

Let $Q(x_{i_1},x_{i_2},\ldots,x_{i_k}) = (1-x_1)Q_0 + x_1Q_1$; then $\underline{f}(x_{i_1},x_{i_2},\ldots,x_{i_k}) = \mathfrak{i}^{Q(x_{i_1},x_{i_2},\ldots,x_{i_k})}$. Note that the total degree of Q is at most 3. If Q has total degree at most 2 and the coefficients of all cross terms are even, then $\underline{f}$ is affine. Thus f is affine by Corollary 2.14, and we are done.

Otherwise, either there is a cross term $x_{i_s}x_{i_t}$ $(1 \leq s < t \leq k)$ with odd coefficient a_{st} or there is a term $x_1x_{i_s}x_{i_t}$ $(2 \leq s < t \leq k)$ with coefficient $a_{1st} \neq 0$ in Q. Since $k \geq 4$, there exists some $r \in [k]\backslash\{1,s,t\}$. Then $Q^{x_{i_r}=0} = Q(x_{i_1},\ldots,x_{i_{r-1}},0,x_{i_{r+1}},\ldots,x_{i_k})$ has a cross term $x_{i_s}x_{i_t}$ with odd coefficient a_{st} or a term $x_1x_{i_s}x_{i_t}$ with coefficient $a_{1st} \neq 0$. Note that $\underline{f}^{x_{i_r}=0} = \mathfrak{i}^{Q(x_{i_1},\ldots,x_{i_{r-1}},0,x_{i_{r+1}},\ldots,x_{i_k})}$. Thus $\underline{f}^{x_{i_r}=0}$ is not affine by Lemma 2.12. So $f^{x_{i_r}=0}$ is not affine. This is a contradiction. ∎

Let

$$\mathscr{F}_1 = \left\{\lambda\left([1,0]^{\otimes k} + \mathfrak{i}^r[0,1]^{\otimes k}\right) | \lambda \in \mathbb{C}, k = 1,2,\ldots, r = 0,1,2,3\right\},$$
$$\mathscr{F}_2 = \left\{\lambda\left([1,1]^{\otimes k} + \mathfrak{i}^r[1,-1]^{\otimes k}\right) | \lambda \in \mathbb{C}, k = 1,2,\ldots, r = 0,1,2,3\right\},$$
$$\mathscr{F}_3 = \left\{\lambda\left([1,\mathfrak{i}]^{\otimes k} + \mathfrak{i}^r[1,-\mathfrak{i}]^{\otimes k}\right) | \lambda \in \mathbb{C}, k = 1,2,\ldots, r = 0,1,2,3\right\}.$$

It is known that the set of nondegenerate symmetric signatures in $\mathscr{A}$ consists of precisely the nonzero signatures in $\mathscr{F}_1 \bigcup \mathscr{F}_2 \bigcup \mathscr{F}_3$ with arity at least 2 $(\lambda \neq 0)$ [20, 4].

Product-type signatures.

DEFINITION 2.19. *A signature on a set of variables X is of* product type *if it can be expressed as a product of unary functions, binary equality functions* ($[1,0,1]$)*, and binary disequality functions* ($[0,1,0]$)*, each on one or two variables of X. We use $\mathscr{P}$ to denote the set of product-type functions.*

A symmetric signature of the form $[a,0,\ldots,0,b]$ is called a GENERALIZED EQUALITY.

PROPOSITION 2.20 (cf. Lemma A.1 in the full version of [30]). *If f is a symmetric signature in $\mathscr{P}$, then f is either degenerate, binary* DISEQUALITY $(\neq_2) = [0,1,0]$ *up to a constant scalar, or $[a,0,\ldots,0,b]$ for some $a,b \in \mathbb{C}$.*

COROLLARY 2.21. $[1,0,1,0] \notin \mathscr{P}$.

We will use Corollary 2.21 in the proof of Theorem 4.9.

Definition 2.19 is succinct and is from [20]. But to deal with asymmetric signatures, an alternative definition of $\mathscr{P}$ given in [16] is useful. This is given below in Definition 2.22. To state it we need some notation.

Suppose f is a signature of arity n and $\mathcal{I} = \{I_1, I_2, \ldots, I_k\}$ is a partition of $[n]$. If $f(X) = \prod_{j=1}^{k} f_j(X|_{I_j})$ for some signatures $f_1, f_2, \ldots, f_k$, where $X = \{x_1, x_2, \ldots, x_n\}$ and $X|_{I_j} = \{x_s | s \in I_j\}$ (we also denote it by X_j), then we say f can be decomposed as a tensor product of $f_1, f_2, \ldots, f_k$. We denote such a function by $f = \bigotimes_{\mathcal{I}}(f_1, f_2, \ldots, f_k)$. If each f_j is the signature of some $\mathcal{F}$-gate, then $\bigotimes_{\mathcal{I}}(f_1, f_2, \ldots, f_k)$ is the signature of the $\mathcal{F}$-gate which is the disjoint union of the $\mathcal{F}$-gates for f_j, with variables renamed and ordered according to $\mathcal{I}$. (This is not necessarily a planar $\mathcal{F}$-gate even when the $\mathcal{F}$-gates for all f_j are planar, unless the sets $I_1, I_2, \ldots, I_k$ partition $[n]$ in order.) When the indexing is clear, we also use the notation $f_1 \otimes f_2 \otimes \cdots \otimes f_k$. Note that this tensor product notation $\otimes$ is consistent with notation for tensor product of matrices. We say a signature set $\mathcal{F}$ is closed under tensor product if for any partition $\mathcal{I} = \{I_1, I_2\}$ and any $f, g \in \mathcal{F}$ on X_1 and X_2, respectively, we have $\bigotimes_{\mathcal{I}}(f, g) \in \mathcal{F}$. The tensor closure $\langle \mathcal{F} \rangle$ of $\mathcal{F}$ is the minimum set containing $\mathcal{F}$, closed under tensor product.

DEFINITION 2.22. *Let $\mathcal{E}$ be the set of all signatures f such that* $\mathrm{supp}(f)$ *is contained in two antipodal points; i.e., if f has arity n, then f is zero except on (possibly) two complementary inputs $\alpha = (a_1, a_2, \ldots, a_n)$ and $\overline{\alpha} = (\overline{a}_1, \overline{a}_2, \ldots, \overline{a}_n) = (1 - a_1, 1 - a_2, \ldots, 1 - a_n)$. Then $\mathscr{P} = \langle \mathcal{E} \rangle$.*

We claim that Definition 2.19 is equivalent to Definition 2.22. If $f \in \mathcal{E}$, then its support $\mathrm{supp}(f) \subseteq \{\alpha, \bar{\alpha}\}$ for some $\alpha = a_1 a_2 \cdots a_n \in \{0,1\}^n$. We may assume that $a_1 = 0$. Then

$$f = [f_\alpha, f_{\bar{\alpha}}](x_1) \prod_{i=1}^{n-1} f_i(x_i, x_{i+1}),$$

where $[f_\alpha, f_{\bar{\alpha}}](x_1)$ is a unary function on x_1, and for all $1 \le i \le n-1$, $f_i = [1,0,1]$ if $a_i = a_{i+1}$ and $f_i = [0,1,0]$ if $a_i \neq a_{i+1}$. This implies that f is a product of the unary function $[f_\alpha, f_{\bar{\alpha}}]$, and binary functions $[1,0,1]$ and $[0,1,0]$. Thus all functions in $\langle \mathcal{E} \rangle$ are products of unary functions, and binary functions $[1,0,1]$ and $[0,1,0]$. Conversely, if f has arity n and is a product of unary functions, and binary functions $[1,0,1]$ and $[0,1,0]$, then there exist $S \subseteq \{(i,j) \mid i,j \in [n], \text{ and } i < j\}$ and $S' \subseteq [n]$ such that $f = \prod_{(i,j)\in S} h_{ij}(x_i, x_j) \prod_{\ell \in S'} u_\ell(x_\ell)$, where $h_{ij} = [1,0,1]$ or $[0,1,0]$, and u_ℓ are unary functions. Let $\mathcal{I} = \{I_1, I_2, \ldots, I_k\}$ be the partition of $[n]$ such that for any $i, j \in [n]$,

i, j are in the same I_c iff i and j belong to the same connected component of the graph $([n], S)$. Let

$$H_c = \prod_{i,j \in I_c, (i,j) \in S} h_{ij}(x_i, x_j) \prod_{\ell \in I_c \cap S'} u_\ell(x_\ell)$$

for $1 \le c \le k$. Then $H_c \in \mathcal{E}$ and

$$f = \bigotimes_{\mathcal{I}} (H_1, H_2, \ldots, H_k).$$

This implies that $f \in \langle \mathcal{E} \rangle$. This concludes the justification that Definition 2.19 is equivalent to Definition 2.22.

Given a function $f(X)$, if it is the product of two functions g and h on disjoint proper subsets of variables of X, then $f = g \otimes h$. Clearly every function $f(X)$ has a decomposition as a tensor product $f_1 \otimes f_2 \otimes \cdots \otimes f_k$ where each f_i is not further expressible as a tensor product of functions on disjoint proper subsets. If f is not identically 0, then such a *primitive decomposition* is unique up to a nonzero constant factor. To see this, suppose $f = f_1 \otimes f_2 \otimes \cdots \otimes f_k = g_1 \otimes g_2 \otimes \cdots \otimes g_\ell$ are two such decompositions. Since f is not identically 0, all f_i are not identically 0. For any i, there is a partial assignment for f to all variables in X except those in f_i, such that the resulting function is a nonzero constant multiple of f_i. This gives an expression $c_i f_i(X_i) = g'_1 \otimes g'_2 \otimes \cdots \otimes g'_\ell$ where $c_i \neq 0$ and each g'_j is on a disjoint subset X_{ij} of X_i. By the assumption that f_i is not further expressible as a tensor product of functions on disjoint proper subsets, the only possibility is that all but one $X_{ij} = \emptyset$. It follows that there is one (unique) j such that X_i is a subset of the variables of g_j. By symmetry, for every j, the set of variable of g_j is a subset of the variables of some $f_{i'}$. As the X_i are disjoint, $i' = i$. Hence there is a 1-1 correspondence of these subsets, and so $k = \ell$, and the corresponding subsets are equal. After renaming these functions and subsets, there are nonzero constants c'_i such that $f_i = c'_i g_i$ $(1 \le i \le k)$.

We will consider the primitive decomposition of signatures in $\mathscr{P}$. We claim that for any function $f \in \mathcal{E}$ of arity at least 2, f is nondegenerate iff $|\operatorname{supp}(f)| = 2$.

To justify this claim, one direction is trivial: if $|\operatorname{supp}(f)| = 0$ or 1, then f is identically 0 or is a product of unary functions and thus degenerate. Conversely, suppose f is degenerate, and $f = u_1(x_1) \otimes \cdots \otimes u_n(x_n)$. If any u_i is identically 0, then $|\operatorname{supp}(f)| = 0$. Otherwise, if every u_i is a nonzero multiple of $[1, 0]$ or $[0, 1]$, then $|\operatorname{supp}(f)| = 1$. Otherwise, some u_i has the form $[a, b]$ with $ab \neq 0$. As $n \ge 2$ there is another $u_j = [c, d]$, where c or $d \neq 0$. Without loss of generality, $c \neq 0$. Then there are two points $a_1 a_2 \ldots a_n$ and $a'_1 a'_2 \ldots a'_n \in \operatorname{supp}(f)$, where $a_i = 0, a_j = 0$ and $a'_i = 1, a'_j = 0$. This contradicts $f \in \mathcal{E}$. This concludes the justification.

By Definitions 2.19 and 2.22, for any $f \in \mathscr{P}$ not identically 0, its primitive decomposition exists and is unique up to a constant factor and is a product of unary functions and nondegenerate functions in $\mathcal{E}$ with $|\operatorname{supp}(f)| = 2$.

DEFINITION 2.23. *Let $f \in \mathscr{P}$, where f is not identically zero. There exist a partition $\mathcal{I} = \{I_1, I_2, \ldots, I_k\}$ of $[n]$ and signatures $f_1, f_2, \ldots, f_k \in \mathcal{E}$, where each f_i is a unary signature or f_i is nondegenerate, such that*

$$f(X) = \prod_{j=1}^{k} f_j(X|_{I_j}). \tag{2.4}$$

We call (2.4) *a primitive decomposition of f.*

To define a compatibility relation on functions in $\mathscr{P}$, we need to first define a notion of compatible partitions of $[n]$.

DEFINITION 2.24. *Two partitions $\mathcal{I} = \{I_1, I_2, \ldots, I_k\}$ and $\mathcal{J} = \{J_1, J_2, \ldots, J_\ell\}$ of $[n]$ are* compatible *if for any $i \in [k]$, $j \in [\ell]$, I_i and J_j satisfy one of the following conditions:*

- $I_i = J_j$;
- $I_i \bigcap J_j = \emptyset$;
- $|I_i| = 1$ *and* $I_i \subseteq J_j$;
- $|J_j| = 1$ *and* $J_j \subseteq I_i$.

An equivalent condition is that if $I_i \bigcap J_j \neq \emptyset$, and $|I_i| \geq 2$ and $|J_j| \geq 2$, then $I_i = J_j$. Yet another equivalent condition is that for any $i \in [k]$, $j \in [\ell]$,

$$I_i \bigcap J_j \neq \emptyset \implies I_i = J_j \text{ or } \Big[|I_i| = 1 \text{ (and thus } I_i \subseteq J_j)\Big] \text{ or } \Big[|J_j| = 1 \text{ (and thus } J_j \subseteq I_i)\Big].$$

Thus, two partitions $\mathcal{I}$ and $\mathcal{J}$ are not compatible iff there exist I_i and J_j such that

$$I_i \cap J_j \neq \emptyset \quad \text{and} \quad I_i \neq J_j \quad \text{and} \quad |I_i| \geq 2 \quad \text{and} \quad |J_j| \geq 2. \tag{2.5}$$

DEFINITION 2.25. *For $f, g \in \mathscr{P}$ not identically zero, we say f, g have* compatible type *if in the primitive decompositions of f and g,*

$$f(X) = \prod_{i=1}^{k} f_i(X|_{I_i}), \qquad g(X) = \prod_{j=1}^{\ell} g_j(X|_{J_j}),$$

- *the partitions $\mathcal{I} = \{I_1, I_2, \ldots, I_k\}$ and $\mathcal{J} = \{J_1, J_2, \ldots, J_\ell\}$ are compatible;*
- *for I_i with $|I_i| \geq 2$ and so supp$(f_i) = \{\alpha, \bar{\alpha}\}$, either there exists J_j such that $I_i = J_j$ and f_i, g_j have the same support, or there exist $\{J_{j_1}, J_{j_2}, \ldots, J_{j_{|I_i|}}\}$ such that $I_i = \bigcup_{t=1}^{|I_i|} J_{j_t}$, $|J_{j_t}| = 1$ for $1 \leq t \leq |I_i|$, and the support of $\prod_{t=1}^{|I_i|} g_{j_t}$ is the singleton set $\{\alpha\}$ or $\{\bar{\alpha}\}$;*
- *for J_j with $|J_j| \geq 2$ and so supp$(g_j) = \{\beta, \bar{\beta}\}$, either there exists I_i such that $J_j = I_i$ and f_i, g_j have the same support, or there exist $\{I_{i_1}, I_{i_2}, \ldots, I_{i_{|J_j|}}\}$ such that $J_j = \bigcup_{s=1}^{|J_j|} I_{i_s}$, $|I_{i_s}| = 1$ for $1 \leq s \leq |J_j|$, and the support of $\prod_{s=1}^{|J_j|} f_{i_s}$ is the singleton set $\{\beta\}$ or $\{\bar{\beta}\}$.*

As primitive decompositions are unique up to constant factors, Definition 2.25 is well defined; it does not depend on these constant factors.

LEMMA 2.26. *Suppose $f, g, h \in \mathscr{P}$ have arity n and any two of them have compatible type. Then there exist a partition $\mathcal{L} = \{L_1, L_2, \ldots, L_\ell\}$ of $[n]$ and signatures $f_1, f_2, \ldots, f_\ell$, $g_1, g_2, \ldots, g_\ell$, and $h_1, h_2, \ldots, h_\ell$, where $f_i, g_i, h_i \in \mathcal{E}$ for $1 \leq i \leq \ell$ and* $\mathrm{supp}(f_i), \mathrm{supp}(g_i), \mathrm{supp}(h_i) \subseteq \{\alpha_i, \bar{\alpha_i}\}$ *for some $\alpha_i \in \{0,1\}^{|L_i|}$, such that $f(X) = \prod_{i=1}^{\ell} f_i(X|_{L_i})$, $g(X) = \prod_{i=1}^{\ell} g_i(X|_{L_i})$, and $h(X) = \prod_{i=1}^{\ell} h_i(X|_{L_i})$.*

Remark. The binary relation that two signatures of arity n in $\mathscr{P}$ have compatible type is not transitive. For example, if $f = f_1(x_1)f_2(x_2, x_3), g = g_1(x_1)g_2(x_2)g_3(x_3)$, $h = h_1(x_1, x_2)h_2(x_3)$, where $f_1 = [1, 0], f_2 = [1, 0, 1], g_1 = g_2 = g_3 = [1, 0]$, $h_1 = [1, 0, 1], h_2 = [1, 0]$, then the two pairs (f, g) and (g, h) are both compatible. But the pair (f, h) is not compatible.

Proof. We prove by induction on n. For $n = 1$, f, g, h are all unary signatures. The lemma is true trivially. Inductively assume the lemma is true for $n' < n$, and we prove the lemma for $n \geq 2$.

Let $f(X) = \prod_{j=1}^{p} F_j(X|_{I_j})$, $g(X) = \prod_{j=1}^{q} G_j(X|_{J_j})$, and $h(X) = \prod_{j=1}^{r} H_j(X|_{K_j})$ be the primitive decompositions of f, g, h, respectively, where $\mathcal{I} = \{I_1, I_2, \ldots, I_p\}$, $\mathcal{J} = \{J_1, J_2, \ldots, J_q\}$, and $\mathcal{K} = \{K_1, K_2, \ldots, K_r\}$ are three partitions of $[n]$. If all $|I_i| = |J_j| = |K_k| = 1$ $(1 \leq i \leq p, 1 \leq j \leq q, 1 \leq k \leq r)$, then $p = q = r = n$. We can rename the sets so that $I_i = J_i = K_i = \{i\}$ and let $\mathcal{L} = \mathcal{I}$ and $f_i = F_i, g_i = G_i, h_i = H_i$ for $1 \leq i \leq n$.

Otherwise, without loss of generality, we assume that $|I_1| \geq 2$. With respect to I_1, we will take out some suitable factors of g to form a function G_1'; similarly, we will take out some suitable factors of h to form a function H_1'. More precisely, since f and g have compatible type, by the definition of primitive decomposition, either there exists $j \in [q]$ such that $J_j = I_1$ and the support of G_j is the same as F_1, or there exist J_{j_s} such that $|J_{j_s}| = 1$ $(1 \leq s \leq |I_1|)$ and $I_1 = \bigcup_{s=1}^{|I_1|} J_{j_s}$, and the support of $\prod_{s=1}^{|I_1|} G_{j_s}$ is a singleton subset of the support of F_1. Then we let $G_1' = G_j$ or $G_1' = \prod_{s=1}^{|I_1|} G_{j_s}$ according to the two cases. For h, we set H_1' similarly.

Let f', g', and h' be defined by the product of those factors other than those of F_1, G_1', and H_1' in the respective primitive decompositions of f, g, and h. Then each pair of f', g', h' have compatible type, and all have arity $n - |I_1|$. By induction, there exist a partition $\mathcal{L} = \{L_2, \ldots, L_\ell\}$ of $[n] \setminus I_1$ and signatures $f_2, \ldots, f_\ell$, $g_2, \ldots, g_\ell$, and $h_2, \ldots, h_\ell$, where $f_i, g_i, h_i \in \mathcal{E}$ for $2 \leq i \leq \ell$ and $\mathrm{supp}(f_i), \mathrm{supp}(g_i), \mathrm{supp}(h_i) \subseteq \{\alpha_i, \bar{\alpha_i}\}$ for some $\alpha_i \in \{0,1\}^{|L_i|}$, such that $f'(X) = \prod_{i=2}^{\ell} f_i(X|_{L_i})$, $g'(X) = \prod_{i=2}^{\ell} g_i(X|_{L_i})$, and $h'(X) = \prod_{i=2}^{\ell} h_i(X|_{L_i})$. Then we finish the proof by letting $f_1 = F_1$, $g_1 = G_1'$, $h_1 = H_1'$, and $L_1 = I_1$. ∎

Matchgate signatures. Matchgates were introduced by Valiant [44, 45] to give polynomial-time algorithms for a collection of counting problems over planar graphs. As the name suggests, problems expressible by matchgates can be reduced to computing a weighted sum of perfect matchings. The latter problem is tractable over planar graphs by Kasteleyn's algorithm [35], a.k.a. the FKT algorithm [41, 33]. These counting problems are naturally expressed in the Holant framework using *matchgate signatures*. We give the definition of matchgate signatures and describe some properties that will be needed in this paper. More details can be found in [7, 4].

Let $G = (V, E, W)$ be a weighted undirected plane graph. A *matchgate* Γ is a tuple (G, X) where $X \subseteq V$ is a set of external nodes, ordered counterclockwise on the external face. Γ is called an odd (resp., even) matchgate if it has an odd (resp., even) number of nodes.

Each matchgate Γ with n external nodes is assigned a *matchgate signature* $(\Gamma^\alpha)_{\alpha \in \{0,1\}^n}$ with 2^n entries,

$$\Gamma^{i_1 i_2 \cdots i_n} = \mathrm{PerfMatch}(G - Z) = \sum_{M} \prod_{(i,j) \in M} w_{ij},$$

where $Z \subseteq X$ is a subset of external nodes having the characteristic sequence $\chi_Z = i_1 i_2 \cdots i_n$, the graph $G - Z$ is obtained from G by removing Z and its incident edges, and the sum is over all perfect matchings M of $G - Z$.

An entry Γ^α is called an even (resp., odd) entry if the Hamming weight $\mathrm{wt}(\alpha)$ is even (resp., odd). It is known (see [7] and [4]) that matchgate signatures are characterized by the following two sets of conditions. (1) The parity requirements: Either

all even entries are 0 or all odd entries are 0. This is due to the fact that every perfect matching takes an even number of nodes. (2) A set of matchgate identities (MGIs) defined as follows: For any $\alpha \in \{0,1\}^n$ and any position vector $P = \{p_1, p_2, \ldots, p_\ell\}$, where $p_1 < p_2 < \cdots < p_\ell$ (we also use P to denote the bit string with 1 in the p_ith bits for $1 \le i \le \ell$ and 0 elsewhere),

$$\sum_{i=1}^{\ell} (-1)^i \Gamma^{\alpha + e_{p_i}} \Gamma^{\alpha + P + e_{p_i}} = 0 \tag{2.6}$$

(alternating sum by flipping in sequence the bits p_i and the bits in $P \setminus \{p_i\}$), where $\alpha + \beta$ denotes the XOR of α and β.

Actually in [7] it is shown that MGIs imply the Parity Condition. But in practice, it is easier to apply the Parity Condition first.

We use $\mathscr{M}$ to denote the set of all matchgate signatures; thus Pl-Holant($\mathscr{M}$) is tractable.

DEFINITION 2.27. *A signature satisfies the even (resp., odd) Parity Condition if all nonzero entries have even (resp., odd) Hamming weight.*

PROPOSITION 2.28. *A unary signature $[a, b] \in \mathscr{M}$ iff it is $[1, 0]$ or $[0, 1]$ up to a scalar.*

We will use Proposition 2.28 in the proof of Theorem 3.12.

LEMMA 2.29 (cf. Lemmas 2.3 and 2.4 in [5]). *If f has arity ≤ 3, then $f \in \mathscr{M}$ iff f satisfies the Parity Condition.*

If f has arity 4 and f satisfies the even Parity Condition, i.e.,

$$M_{x_1x_2,x_4x_3}(f) = \begin{bmatrix} f_{0000} & 0 & 0 & f_{0011} \\ 0 & f_{0110} & f_{0101} & 0 \\ 0 & f_{1010} & f_{1001} & 0 \\ f_{1100} & 0 & 0 & f_{1111} \end{bmatrix},$$

then $f \in \mathscr{M}$ iff

$$f_{0000}f_{1111} - f_{1100}f_{0011} + f_{1010}f_{0101} - f_{1001}f_{0110} = 0.$$

(This is the MGI with $\alpha = 1000$ and $P = \{1, 2, 3, 4\}$.) Equivalently, $f \in \mathscr{M}$ iff

$$\det \begin{bmatrix} f_{0000} & f_{0011} \\ f_{1100} & f_{1111} \end{bmatrix} = \det \begin{bmatrix} f_{0110} & f_{0101} \\ f_{1010} & f_{1001} \end{bmatrix}.$$

2.6. Transformable signature sets. An important definition involving a holographic transformation is the notion of a signature set being transformable.

DEFINITION 2.30. *We say a pair of signature sets $(\mathcal{G}, \mathcal{F})$ is $\mathscr{C}$-transformable for* Holant $(\mathcal{G} \mid \mathcal{F})$ *if there exists $T \in \mathbf{GL}_2(\mathbb{C})$ such that $\mathcal{G}T \subseteq \mathscr{C}$ and $T^{-1}\mathcal{F} \subseteq \mathscr{C}$.*

For $\mathcal{G} = \{(=_2)\}$, Holant $((=_2) \mid \mathcal{F}) \equiv_T$ Holant$(\mathcal{F})$*, we say simply that $\mathcal{F}$ is $\mathscr{C}$-transformable. For $\mathcal{G} = \mathcal{EQ}$,* Holant $(\mathcal{EQ} \mid \mathcal{F}) \equiv_T \#\mathrm{CSP}(\mathcal{F})$*, we say that $\mathcal{F}$ is $\mathscr{C}$-transformable for #CSP. We define similarly for #CSP^d when $\mathcal{G} = \mathcal{EQ}_d$. The definitions also work in the planar case.*

Notice that if Pl-Holant($\mathscr{C}$) is tractable, and $(\mathcal{G}, \mathcal{F})$ is $\mathscr{C}$-transformable, then Pl-Holant $(\mathcal{G} \mid \mathcal{F})$ is tractable by a holographic transformation. For example, consider

$H_2 = \frac{1}{\sqrt{2}}\left[\begin{smallmatrix}1 & 1\\ 1 & -1\end{smallmatrix}\right]$, with $H_2^{-1} = H_2$. Recall the notation $\widehat{\mathcal{F}} = H_2\mathcal{F}$. It is an important fact that $\widehat{\mathcal{EQ}} \subset \mathcal{M}$ (see [7, 4]); this can be verified both by MGIs as well as by direct constructions. Also $H_2\widehat{\mathcal{M}} = \mathcal{M}$. Thus Pl-#CSP($\widehat{\mathcal{M}}$) is tractable, since Pl-Holant($\mathcal{M}$) is tractable. We list some important families of signatures specific to the Pl-#CSP and Pl-#CSP^2 frameworks. First we have

$$\widehat{\mathcal{P}} = H_2\mathcal{P} \quad \text{and} \quad \widehat{\mathcal{M}} = H_2\mathcal{M}.$$

Note that $\mathcal{A}$ is unchanged under the transformation by H_2; thus there is no need to define $\widehat{\mathcal{A}}$. We have $\widehat{\mathcal{EQ}} \subset \mathcal{A} \cap \mathcal{M}$. Thus $\mathcal{A}$ is $\mathcal{A}$-transformable and $\widehat{\mathcal{M}}$ is $\mathcal{M}$-transformable, respectively, for Pl-#CSP.

DEFINITION 2.31. *Let* $\mathcal{R}_k = \left\{\left[\begin{smallmatrix}1 & 0\\ 0 & \omega\end{smallmatrix}\right] \mid \omega^k = 1\right\}$ *be a set of diagonal matrices of order dividing* k *and* $\mathcal{T}_k = \mathcal{R}_{2k} \setminus \mathcal{R}_k = \left\{\left[\begin{smallmatrix}1 & 0\\ 0 & \omega\end{smallmatrix}\right] \mid \omega^k = -1\right\}$. *Let* $\mathcal{A}^\dagger = \mathcal{T}_4\mathcal{A}$ *and* $\widehat{\mathcal{M}}^\dagger = \mathcal{T}_2\widehat{\mathcal{M}}$ *be the sets of signatures transformed by* $\mathcal{T}_4$ *from the affine family* $\mathcal{A}$ *and transformed by* $\mathcal{T}_2$ *from* $\widehat{\mathcal{M}}$*, respectively.*

Note that $\mathcal{P}$ is unchanged under any diagonal matrix. Thus there is no need to define $\mathcal{P}^\dagger$.

We claim that $\mathcal{A}^\dagger$, as well as $\widehat{\mathcal{M}}$ and $\widehat{\mathcal{M}}^\dagger$, are, respectively, both $\mathcal{A}$-transformable and $\mathcal{M}$-transformable for Pl-#CSP^2. To see this, let $T = \left[\begin{smallmatrix}1 & 0\\ 0 & \omega\end{smallmatrix}\right] \in \mathcal{T}_4$ with $\omega^4 = -1$; then $T^{-1}\mathcal{A}^\dagger = \mathcal{A}$ and $(=_{2n})T^{\otimes 2n} \in \mathcal{A}$. Hence $\mathcal{A}^\dagger$ is $\mathcal{A}$-transformable for Pl-#CSP^2. Similarly, for $T = \left[\begin{smallmatrix}1 & 0\\ 0 & \pm 1\end{smallmatrix}\right] = T^{-1}$, $TH_2 = \frac{1}{\sqrt{2}}\left[\begin{smallmatrix}1 & 1\\ \pm 1 & \mp 1\end{smallmatrix}\right]$ is either H_2 or $H_2\left[\begin{smallmatrix}0 & 1\\ 1 & 0\end{smallmatrix}\right]$, and $\left[\begin{smallmatrix}0 & 1\\ 1 & 0\end{smallmatrix}\right]\mathcal{M} = \mathcal{M}$. Thus $T^{-1}\widehat{\mathcal{M}} = TH_2\mathcal{M} = H_2\mathcal{M} = \widehat{\mathcal{M}}$, and $(TH_2)^{-1}\widehat{\mathcal{M}} = H_2^{-1}T^{-1}\widehat{\mathcal{M}} = H_2^{-1}\widehat{\mathcal{M}} = \mathcal{M}$, so $\widehat{\mathcal{M}}$ is $\mathcal{M}$ transformed under TH_2. Also note that for all such T, we have $(=_{2n})(TH_2)^{\otimes 2n} \in \mathcal{M}$. Hence $\widehat{\mathcal{M}}$ is $\mathcal{M}$-transformable for Pl-#CSP^2. Finally, let $T' = \left[\begin{smallmatrix}1 & 0\\ 0 & \omega\end{smallmatrix}\right] \in \mathcal{T}_2$ with $\omega^2 = -1$. We have $T'^{-1}\widehat{\mathcal{M}}^\dagger = \left[\begin{smallmatrix}1 & 0\\ 0 & \pm 1\end{smallmatrix}\right]\widehat{\mathcal{M}} = \widehat{\mathcal{M}}$. Thus $(T'H_2)^{-1}\widehat{\mathcal{M}}^\dagger = H_2T'^{-1}\widehat{\mathcal{M}}^\dagger = H_2\widehat{\mathcal{M}} = \mathcal{M}$. Also $(=_{2n})(T'H_2)^{\otimes 2n} \in \mathcal{M}$. Hence $\widehat{\mathcal{M}}^\dagger$ is $\mathcal{M}$-transformable for Pl-#CSP^2.

Note that the set of nondegenerate symmetric signatures in $\mathcal{A}^\dagger$ is precisely the nonzero signatures ($\lambda \neq 0$) in $\mathcal{F}_1^\dagger \bigcup \mathcal{F}_2^\dagger$ with arity at least 2, where $\mathcal{F}_1^\dagger$ and $\mathcal{F}_2^\dagger$ are two families of signatures defined as

$$\mathcal{F}_1^\dagger = \left\{\lambda\left([1,0]^{\otimes k} + \mathrm{i}^r[0,1]^{\otimes k}\right) \mid \lambda \in \mathbb{C}, k = 1,2,\ldots, r = 0,1,2,3\right\}, \text{ and}$$
$$\mathcal{F}_2^\dagger = \left\{\lambda\left([1,\xi]^{\otimes k} + \mathrm{i}^r[1,-\xi]^{\otimes k}\right) \mid \lambda \in \mathbb{C}, \xi^4 = -1, k = 1,2,\ldots, r = 0,1,2,3\right\}.$$

Proposition 2.32 can be directly checked.

PROPOSITION 2.32. *The following hold:*

- *A unary signature is in* $\widehat{\mathcal{M}}$ *iff it is* $\lambda[1, \pm 1], \lambda \in \mathbb{C}$.
- *A unary signature is in* $\widehat{\mathcal{M}}^\dagger$ *iff it is* $\lambda[1, \pm \mathrm{i}], \lambda \in \mathbb{C}$.
- $[1,0,1,0] \notin \mathcal{A}^\dagger$.

2.7. Some known dichotomies. Here we list several known dichotomies. The first is for #CSP without planarity. The other two are about planar #CSP (and #CSP^2) but restricted to *symmetric* signatures.

THEOREM 2.33 (Theorem 3.1 in [20]). *Let* $\mathcal{F}$ *be any set of complex-valued signatures in Boolean variables. Then* $\#\,\mathrm{CSP}(\mathcal{F})$ *is* #P*-hard unless* $\mathcal{F} \subseteq \mathcal{A}$ *or* $\mathcal{F} \subseteq \mathcal{P}$*, in which case the problem is computable in polynomial time.*

By (2.3), this result can be rephrased for Holant problems.

THEOREM 2.33′. *Let $\widehat{\mathcal{F}}$ be any set of complex-valued signatures in Boolean variables. Then* Holant$(\widehat{\mathcal{EQ}}, \widehat{\mathcal{F}})$ *is* #P*-hard unless $\widehat{\mathcal{F}} \subseteq \widehat{\mathscr{A}}$ or $\widehat{\mathcal{F}} \subseteq \widehat{\mathscr{P}}$, in which case the problem is computable in polynomial time.*

The next theorem is a dichotomy for Pl-#CSP problems over symmetric signatures.

THEOREM 2.34 (Theorem 19 in [27]). *Let $\mathcal{F}$ be any set of* symmetric*, complex-valued signatures in Boolean variables. Then* Pl-#CSP$(\mathcal{F})$ *is* #P*-hard unless $\mathcal{F} \subseteq \mathscr{A}$, $\mathcal{F} \subseteq \mathscr{P}$, or $\mathcal{F} \subseteq \widehat{\mathscr{M}}$, in which case the problem is computable in polynomial time.*

By (2.3), it can be restated for Pl-Holant problems.

THEOREM 2.34′. *Let $\widehat{\mathcal{F}}$ be any set of* symmetric*, complex-valued signatures in Boolean variables. Then* Pl-Holant$(\widehat{\mathcal{EQ}}, \widehat{\mathcal{F}})$ *is* #P*-hard unless $\widehat{\mathcal{F}} \subseteq \widehat{\mathscr{A}}$, $\widehat{\mathcal{F}} \subseteq \widehat{\mathscr{P}}$, or $\widehat{\mathcal{F}} \subseteq \widehat{\mathscr{M}}$, in which case the problem is computable in polynomial time.*

The following theorem is a dichotomy theorem for Pl-#CSP^2 problems over *symmetric* signatures. By (2.2) for $d = 2$, we have Pl-#CSP$^2(\mathcal{F}) \equiv_{\mathrm{T}}$ Pl-Holant$(\mathcal{EQ}_2, \mathcal{F})$. Thus the theorem can be equivalently stated for Pl-Holant$(\mathcal{EQ}_2, \mathcal{F})$. Note that this equivalence is not by a holographic transformation. However, when we apply it later in this paper, we actually use it on the RHS of the equivalence by a holographic transformation Pl-#CSP$(\mathcal{F}) \equiv_{\mathrm{T}}$ Pl-Holant$(\widehat{\mathcal{EQ}}, \widehat{\mathcal{F}})$, when we can construct $\mathcal{EQ}_2$ in the RHS.

THEOREM 2.35 (Theorem A.2 in [6]). *Let $\mathcal{F}$ be any set of* symmetric*, complex-valued signatures in Boolean variables. Then* Pl-#CSP$^2(\mathcal{F})$*, equivalently* Pl-Holant$(\mathcal{EQ}_2, \mathcal{F})$*, is #P-hard unless $\mathcal{F} \subseteq \mathscr{P}$, $\mathcal{F} \subseteq \mathscr{A}$, $\mathcal{F} \subseteq \mathscr{A}^{\dagger}$, $\mathcal{F} \subseteq \widehat{\mathscr{M}}$, or $\mathcal{F} \subseteq \widehat{\mathscr{M}}^{\dagger}$, in which case the problem is computable in polynomial time.*

Note that Theorem 2.34 (and Theorem 2.35) are applicable only for symmetric signatures. The main theorem of the present paper is to generalize Theorem 2.34 to be valid for all, not necessarily symmetric, signatures over Boolean variables.

2.8. Some lemmas. In this subsection, we prove some simple lemmas. The next lemma shows that flipping any input variable of a signature f does not change its membership in $\mathscr{P}$, or $\mathscr{A}$, or $\mathscr{M}$.

LEMMA 2.36. *Let $g(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_n) = f(x_1, \ldots, x_{i-1}, \overline{x_i}, x_{i+1}, \ldots, x_n)$; then for $\mathscr{C} \in \{\mathscr{P}, \mathscr{A}, \mathscr{M}\}$, $f \in \mathscr{C}$ iff $g \in \mathscr{C}$.*

Proof. Note that $[0, 1, 0] \in \mathscr{P} \cap \mathscr{A} \cap \mathscr{M}$, and f is obtained from g by flipping x_i. It follows easily by definition of $\mathscr{C} \in \{\mathscr{P}, \mathscr{A}, \mathscr{M}\}$ that $f \in \mathscr{C}$ iff $g \in \mathscr{C}$. ∎

The following lemma shows how to use $[1, 0, 1, 0] \in \widehat{\mathcal{EQ}}$ and $[0, 1]^{\otimes 2}$ to get $[0, 1, 0]^{\otimes 2}$, then how to flip any two variables that are not necessarily adjacent, while preserving planarity.

LEMMA 2.37. *In* Pl-Holant$(\widehat{\mathcal{EQ}}, [0, 1]^{\otimes 2}, f)$*, if f has arity n, then for any $s \neq t \in [n]$, we can construct g such that $g(x'_1, x'_2, \ldots, x'_n) = f(x_1, x_2, \ldots, x_n)$, where $x'_k = \overline{x_k}$ for $k \in \{s, t\}$ and $x'_k = x_k$ otherwise. Moreover, for $\mathscr{C} \in \{\mathscr{P}, \mathscr{A}, \mathscr{M}\}$, $f \in \mathscr{C}$ iff $g \in \mathscr{C}$.*

Proof. That $f \in \mathscr{C}$ iff $g \in \mathscr{C}$ follows from Lemma 2.36.

Note that we have $[1, 0, 1, 0] \in \widehat{\mathcal{EQ}}$, and $[0, 1]^{\otimes 2}$. Since $\partial_{[0,1]}([1, 0, 1, 0]) = [0, 1, 0]$, by connecting $[0, 1]^{\otimes 2}$ to two disjoint copies of $[1, 0, 1, 0]$, we get $[0, 1, 0]^{\otimes 2}$. Note that this is a planar gadget where two adjacent pairs of variables are flipped. This function

is $(x_1 \neq x_2) \wedge (x_3 \neq x_4)$. After a rotation of 90° we also get $(x_4 \neq x_1) \wedge (x_2 \neq x_3)$, which we will denote as D^2.

Without loss of generality, we assume that $t > s$. If $t - s = 1$, then x_s and x_{s+1} are adjacent variables and we can directly apply D^2 to flip both x_s and x_{s+1}. In general (see Figure 3 for an illustration), we let $h^{(0)} = f$ and, for $1 \leq j \leq t-s$, define $h^{(j)}(x_1^{(j)}, x_2^{(j)}, \ldots, x_n^{(j)}) = h^{(j-1)}(x_1^{(j-1)}, x_2^{(j-1)}, \ldots, x_n^{(j-1)})$, where $x_i^{(j)} = \overline{x_i^{(j-1)}}$ for $i \in \{s+j-1, s+j\}$ and $x_i^{(j)} = x_i^{(j-1)}$ for all others. Then we are done by letting $g = h^{(t-s)}$. In effect, all variables x_i with $s < i < t$ are flipped twice. ∎

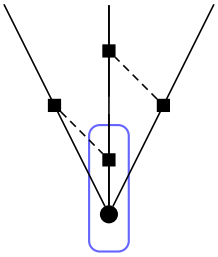


FIG. 3. *Flipping two variables of f that are not adjacent by $[0,1,0]^{\otimes 2}$ while preserving planarity. The circle vertex is labeled f and squares are $[0,1,0]$. A pair of squares connected by a dashed line forms $[0,1,0]^{\otimes 2}$.*

The following lemma implies that in Pl-Holant$(\widehat{\mathcal{EQ}}, f)$, where $f \notin \mathscr{A}$ or $f \notin \mathscr{M}$, we can assume that $f_{00\cdots 0} = 1$. Moreover, if f satisfies the Parity Condition, we can assume it satisfies the *even* Parity Condition.

LEMMA 2.38. *For $\mathscr{C} = \mathscr{A}$ or $\mathscr{M}$, if $\widehat{\mathcal{F}}$ contains a signature $f \notin \mathscr{C}$ of arity n, then we can construct a function $f' \notin \mathscr{C}$ of arity n with $f'_{00\cdots 0} = 1$ such that*

$$\text{Pl-Holant}(\widehat{\mathcal{EQ}}, f', \widehat{\mathcal{F}}) \leq_{\text{T}} \text{Pl-Holant}(\widehat{\mathcal{EQ}}, \widehat{\mathcal{F}}).$$

Moreover, if f satisfies the Parity Condition, then f' satisfies the even Parity Condition, and if f takes values in $\{0,1\}$ ($\{0,1,-1\}$), then f' also takes values in $\{0,1\}$ ($\{0,1,-1\}$).

Proof. If $f_{00\cdots 0} \neq 0$, then we simply normalize f by setting $f' = f/f_{00\cdots 0}$. So we suppose $f_{00\cdots 0} = 0$. By $f \notin \mathscr{C}$, clearly f is not identically 0. Let α be an element of $\text{supp}(f)$ of minimum weight and $S = \{i \mid 1 \leq i \leq n,$ the ith bit of α is $0\}$. Since we have $[1,0] \in \widehat{\mathcal{EQ}}$, we can get $\partial^S_{[1,0]}(f) = [0,1]^{\otimes \text{wt}(\alpha)}$. Depending on whether $\text{wt}(\alpha)$ is odd or even, we can take $\partial_{=_2}$ on $[0,1]^{\otimes \text{wt}(\alpha)}$ repeatedly and obtain either $[0,1]$ or $[0,1]^{\otimes 2}$, respectively. Since we have $[1,0,1,0] \in \widehat{\mathcal{EQ}}$, we can get either $\partial_{[0,1]}([1,0,1,0]) = [0,1,0]$ or $[0,1,0]^{\otimes 2}$.

If $\text{wt}(\alpha)$ is odd, and this includes the case when f satisfies the odd Parity Condition, we have $[0,1,0]$ and can flip any variable of f individually. By flipping all variables in $[n] \setminus S$, and normalizing, we obtain f' with the required property. In particular, if f satisfies the odd Parity Condition, then f' satisfies the even Parity Condition.

If $\text{wt}(\alpha)$ is even, and this includes the case when f satisfies the even Parity Condition, we have $[0,1]^{\otimes 2}$ and $[0,1,0]^{\otimes 2}$. By Lemma 2.37 we can flip any two variables of f. By applying the construction in Lemma 2.37 simultaneously on $\text{wt}(\alpha)/2$ pairs of variables of f, we can transform f to f' by a planar construction so that

$f'_{00\cdots0} = f_\alpha \neq 0$. By normalizing, we obtain the required f' with $f'_{00\cdots0} = 1$. In particular, if f satisfies the even Parity Condition, then f' also satisfies the even Parity Condition. We get f' from f by flipping some variables in all cases. Thus if f takes values in $\{0,1\}$ ($\{0,1,-1\}$), then f' also takes values in $\{0,1\}$ ($\{0,1,-1\}$). ∎

2.9. Interpolation. Polynomial interpolation is a powerful tool in the study of counting problems. In this subsection, we give the following two lemmas by polynomial interpolation.

LEMMA 2.39. *If $x \in \mathbb{C}$ with norm $|x| \neq 0, 1$, then for any $a, b \in \mathbb{C}$, we have*

$$\text{Pl-Holant}(\mathcal{EQ}, [a,b], \mathcal{F}) \leq_{\text{T}} \text{Pl-Holant}(\mathcal{EQ}, [1,x], \mathcal{F}).$$

Proof. Note that for any $k \in \mathbb{Z}^+$, we have $\partial^k_{[1,x]}(=_{k+1}) = [1, x^k]$. Consider an instance Ω of Pl-Holant$(\mathcal{EQ}, [a,b], \mathcal{F})$. Let S be the subset of vertices assigned $[a,b]$, and suppose that $|S| = n$. By replacing each occurrence of $[a,b]$ with $[1,x^k]$, we construct a sequence of instances Ω_k of Pl-Holant$(\mathcal{EQ}, [1,x], \mathcal{F})$.

We stratify the assignments in Ω based on the assignment to $[a,b]$. Let c_ℓ be the sum over all assignments of products of evaluations at all vertices other than those from S such that exactly ℓ occurrences of $[a,b]$ have their respective incident edges assigned 1 (and the other $n-\ell$ are assigned 0). Then

$$\text{Pl-Holant}(\Omega) = \sum_{0 \leq \ell \leq n} a^{n-\ell} b^\ell c_\ell,$$

and the value of the planar Holant on Ω_k, for $1 \leq k \leq n+1$, is

$$\text{Pl-Holant}(\Omega_k) = \sum_{0 \leq \ell \leq n} x^{k\ell} c_\ell.$$

This is a linear system with unknowns c_ℓ and a Vandermonde coefficient matrix. Since $|x| \notin \{0,1\}$, x^k are all distinct ($1 \leq k \leq n+1$), which implies that the Vandermonde matrix has full rank. Therefore, we can solve the linear system in polynomial time for the unknown c_ℓ's and obtain the value of Pl-Holant(Ω). ∎

LEMMA 2.40. *Suppose $\mathcal{F}$ contains a signature f of arity 4 with*

$$M_{x_1x_2,x_4x_3}(f) = \begin{bmatrix} a & 0 & 0 & b \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ c & 0 & 0 & d \end{bmatrix} \quad \text{or} \quad M_{x_1x_2,x_4x_3}(f) = \begin{bmatrix} a & 0 & 0 & 0 \\ 0 & b & 0 & 0 \\ 0 & 0 & c & 0 \\ 0 & 0 & 0 & d \end{bmatrix},$$

where $\left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right]$ has full rank. Then

$$\text{Pl-Holant}(=_4, \mathcal{F}) \leq_{\text{T}} \text{Pl-Holant}(\mathcal{F}).$$

Proof. If

$$M_{x_1x_2,x_4x_3}(f) = \begin{bmatrix} a & 0 & 0 & 0 \\ 0 & b & 0 & 0 \\ 0 & 0 & c & 0 \\ 0 & 0 & 0 & d \end{bmatrix},$$

then after a rotation we have the signature

$$M_{x_4x_1,x_3x_2}(f) = \begin{bmatrix} a & 0 & 0 & b \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ c & 0 & 0 & d \end{bmatrix}.$$

From one signature we can obtain the other signature by a rotation (see Figure 2). So it suffices to prove the lemma for the first form of f.

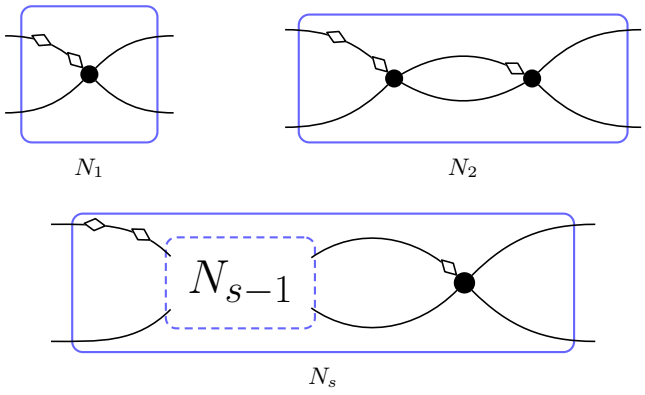


FIG. 4. *Linear recursive construction used for interpolation.*

Consider an instance Ω of Pl-Holant$(=_4, \mathcal{F})$. Suppose $=_4$ appears n times in Ω. We construct from Ω a sequence of instances Ω_s of Pl-Holant$(\mathcal{F})$ indexed by $s \geq 1$. We obtain Ω_s from Ω by replacing each occurrence of $=_4$ with the gadget N_s in Figure 4 with f assigned to all vertices. In Ω_s, the edge corresponding to the ith variable of N_s connects to the same location as the edge corresponding to the ith variable of $=_4$ in Ω. In Figure 4, we place a diamond on the edge corresponding to the first variable. The remaining variables are ordered counterclockwise around the vertex.

By the Jordan normal form of $\left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right]$, there exists $P = \left[\begin{smallmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{smallmatrix}\right]$ such that

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} = P \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} P^{-1},$$

or, when there is a double root, we can normalize it to 1, and then

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} = P \begin{bmatrix} 1 & \lambda \\ 0 & 1 \end{bmatrix} P^{-1},$$

where all of $\lambda_1, \lambda_2, \lambda$ are nonzero. This implies that

$$M_{x_1x_2,x_4x_3}(f) = \begin{bmatrix} p_{00} & 0 & 0 & p_{01} \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ p_{10} & 0 & 0 & p_{11} \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda_2 \end{bmatrix} \begin{bmatrix} p_{00} & 0 & 0 & p_{01} \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ p_{10} & 0 & 0 & p_{11} \end{bmatrix}^{-1} \tag{2.7}$$

or

$$M_{x_1x_2,x_4x_3}(f) = \begin{bmatrix} p_{00} & 0 & 0 & p_{01} \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ p_{10} & 0 & 0 & p_{11} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & \lambda \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} p_{00} & 0 & 0 & p_{01} \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ p_{10} & 0 & 0 & p_{11} \end{bmatrix}^{-1}. \tag{2.8}$$

Let

$$T = \begin{bmatrix} p_{00} & 0 & 0 & p_{01} \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ p_{10} & 0 & 0 & p_{11} \end{bmatrix}$$

and let f_s be the signature of the gadget N_s.

For (2.7), $M_{x_1x_2,x_4x_3}(f_s) = T\Lambda^s T^{-1}$, where

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda_2 \end{bmatrix} \quad \text{and} \quad \Lambda^s = \begin{bmatrix} \lambda_1^s & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda_2^s \end{bmatrix}.$$

If there exists $d \in \mathbb{Z}^+$ such that $(\frac{\lambda_2}{\lambda_1})^d = 1$, then f_d is $=_4$ up to a nonzero scalar, and we are done. Otherwise, for any $i, j \in \mathbb{Z}$, $(\frac{\lambda_2}{\lambda_1})^i \neq (\frac{\lambda_2}{\lambda_1})^j$ if $i \neq j$. We can view our construction of Ω_s as replacing $=_4$ by 3 signatures, with matrices T, Λ^s, T^{-1}, respectively. This does not change the Holant value. The Holant value on Ω is also unchanged by replacing $=_4$ with $T, (=_4), T^{-1}$ in sequence. We stratify assignments in Ω based on assignment values to the n occurrences of the new $(=_4)$, each sandwiched between T and T^{-1}. Note that we only need to consider the assignments to $(=_4)$ that assign

- (0, 0, 0, 0) i many times,
- (1, 1, 1, 1) j many times

such that $i + j = n$, since any other assignment contributes 0 to the Holant sum. Let c_{ij} be the sum over all such assignments of the products of evaluations (including the contributions from T, T^{-1}) in Ω. Then we have

$$\text{Pl-Holant}(\Omega) = \sum_{i+j=n} c_{ij},$$

and

$$\text{Pl-Holant}(\Omega_s) = \sum_{i+j=n} c_{ij}\lambda_1^{is}\lambda_2^{js} = \lambda_1^{ns} \sum_{i+j=n} c_{ij} \left(\frac{\lambda_2}{\lambda_1}\right)^{js}.$$

Note that the same set of values c_{ij} occurs in Pl-Holant(Ω_s) independent of s. Then we get a Vandermonde system with unknowns $c_{n-j,j}$. Since $(\frac{\lambda_2}{\lambda_1})^j \neq (\frac{\lambda_2}{\lambda_1})^{j'}$ if $j \neq j'$, this coefficient matrix has full rank. Therefore, we can solve the linear system in polynomial time and obtain the value of Holant(Ω). This implies that

$$\text{Pl-Holant}(=_4, \mathcal{F}) \leq_{\text{T}} \text{Pl-Holant}(\mathcal{F}).$$

For (2.8),

$$M_{x_1x_2,x_4x_3}(f_s) = T\Lambda^s T^{-1}, \tag{2.9}$$

where

$$\Lambda = \begin{bmatrix} 1 & 0 & 0 & \lambda \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

and

$$\Lambda^s = \begin{bmatrix} 1 & 0 & 0 & s\lambda \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Similarly, we can view our construction of Ω_s as replacing $=_4$ by 3 signatures, with matrices T, Λ^s, T^{-1}, respectively. This does not change the Holant value. We also consider replacing each $=_4$ by $T(=_4)T^{-1}$ in Ω. We group together all assignments in the Ω according to the way the new $(=_4)$'s are assigned.

- (0, 0, 0, 0) or (1, 1, 1, 1) i many times,
- (0, 0, 1, 1) j many times

such that $i + j = n$. Let c'_{ij} be the sum over all such assignments of the products of evaluations (including the contributions from T, T^{-1}) in Ω. Then we have

$$\text{Pl-Holant}(\Omega) = c'_{n0},$$

and

$$\text{Pl-Holant}(\Omega_s) = \sum_{i+j=n} c'_{ij}(s\lambda)^j.$$

Again note that the same set of values c'_{ij} occurs in Pl-Holant(Ω_s), independent of s. Then we get a Vandermonde system with unknown c''_j, where $c''_j = c'_{n-j,j}\lambda^j$ (for $0 \le j \le n$). The coefficient matrix (s^j) has full rank. Therefore, we can solve the linear system in polynomial time and obtain the value of Holant(Ω). This shows that

$$\text{Pl-Holant}(=_4, \mathcal{F}) \le_{\text{T}} \text{Pl-Holant}(\mathcal{F}),$$

and we finish the proof. ∎

2.10. Outline of the proof. We now give an outline of the proof of the main dichotomy, Theorem 6.1, and also explain some overall vision that guided our proof. An important technique is to view our counting problems in the dual perspectives of planar #CSP *and* planar Holant problems; i.e., we make essential use of the equivalence Pl-#CSP$(\mathcal{F}) \equiv_T$ Pl-Holant$(\widehat{\mathcal{EQ}}, \widehat{\mathcal{F}})$. Some questions are easier to handle in one framework, while others are easier in the other.

We aim to prove Theorem 6.1. Our overall vision is that the classification in Theorem 2.34 should be valid for general, not necessarily symmetric, signatures. Thus we want to show that either $\mathcal{F} \subseteq \mathscr{A}$, or $\mathcal{F} \subseteq \mathscr{P}$, or $\mathcal{F} \subseteq \widehat{\mathscr{M}}$, or else Pl-#CSP$(\mathcal{F})$ is #P-hard. In the Pl-Holant$(\widehat{\mathcal{EQ}}, \widehat{\mathcal{F}})$ setting, the tractability condition is expressed as $\widehat{\mathcal{F}} \subseteq \mathscr{A}$, or $\widehat{\mathcal{F}} \subseteq \widehat{\mathscr{P}}$, or $\widehat{\mathcal{F}} \subseteq \mathscr{M}$.

Note that $\mathscr{A}$ is invariant under the transformation; i.e., $\widehat{\mathscr{A}} = \mathscr{A}$. However, $\widehat{\mathscr{P}}$ is more difficult to reason about than $\mathscr{P}$, while $\mathscr{M}$ is easier than $\widehat{\mathscr{M}}$ to handle. The former suggests that we carry out our proof in the Pl-#CSP framework, while the latter suggests the opposite—that we do so in the Pl-Holant framework instead.

One necessary condition for $\mathscr{M}$ is the Parity Condition. If some signature in $\widehat{\mathcal{F}}$ violates the Parity Condition, then we have established that $\widehat{\mathcal{F}} \not\subseteq \mathscr{M}$, eliminating one of the three possible tractable cases. If we prove the theorem in the Pl-#CSP framework, we can avoid discussing the more difficult class $\widehat{\mathscr{M}}$, leaving only two tractable cases $\mathcal{F} \subseteq \mathscr{A}$ and $\mathcal{F} \subseteq \mathscr{P}$. On the other hand, if every signature in $\widehat{\mathcal{F}}$ satisfies the Parity Condition, then we have the lucky situation (Proposition 7.12) that all signatures in $\mathcal{F} \cap \mathscr{P}$ are already in $\mathscr{A}$. This is equivalent to $\widehat{\mathcal{F}} \cap \widehat{\mathscr{P}} \subseteq \mathscr{A}$, and therefore $\widehat{\mathcal{F}} \subseteq \widehat{\mathscr{P}}$ already implies $\widehat{\mathcal{F}} \subseteq \mathscr{A}$, with the consequence that we do not need to specifically discuss the tractability condition $\widehat{\mathcal{F}} \subseteq \widehat{\mathscr{P}}$. Thus in this case we can avoid discussing the irksome class $\widehat{\mathscr{P}}$, leaving only the other two tractable cases $\widehat{\mathcal{F}} \subseteq \mathscr{A}$ and $\widehat{\mathcal{F}} \subseteq \mathscr{M}$.

Therefore, we break the proof into two main cases according to whether $\widehat{\mathcal{F}}$ satisfies the Parity Condition or not. If not, we want to show that Pl-#CSP($\mathcal{F}$) is #P-hard unless $\mathcal{F} \subseteq \mathscr{A}$ or $\mathcal{F} \subseteq \mathscr{P}$ (Theorem 3.12). If yes, we want to prove, in the Pl-Holant setting for $\mathscr{A}$ and $\mathscr{M}$, namely Pl-Holant($\widehat{\mathcal{EQ}}, \widehat{\mathcal{F}}$) is #P-hard unless $\widehat{\mathcal{F}} \subseteq \mathscr{A}$ or $\widehat{\mathcal{F}} \subseteq \mathscr{M}$ (Theorem 5.7).

In the first main case where $\widehat{\mathcal{F}}$ fails the Parity Condition, from any signature in $\widehat{\mathcal{F}}$ violating the Parity Condition, we can construct a gadget with a signature of the simplest form that violates the Parity Condition, namely a unary signature $[1, w]$ with $w \neq 0$, in the Pl-Holant setting Pl-Holant($\widehat{\mathcal{EQ}}, \widehat{\mathcal{F}}$). Any signature that violates the Parity Condition is a witness that $\widehat{\mathcal{F}} \not\subseteq \mathscr{M}$, or equivalently $\mathcal{F} \not\subseteq \widehat{\mathscr{M}}$. If $\mathcal{F} \subseteq \mathscr{A}$ or $\mathcal{F} \subseteq \mathscr{P}$, then the problem Pl-#CSP($\mathcal{F}$) is tractable. Otherwise, there exist some signatures $f, g \in \mathcal{F}$ such that $f \notin \mathscr{A}$ and $g \notin \mathscr{P}$. We would like to construct some *symmetric* signatures from these that are also nonaffine and nonproduct type, respectively, and then apply Theorem 2.34. For the nonproduct type we will do so in the Pl-#CSP($\mathcal{F}$) setting to avoid having to deal with $\widehat{\mathscr{P}}$. For the nonaffine signatures, we can do so in either the Pl-#CSP framework or the Pl-Holant framework as $\mathscr{A}$ is invariant, $\widehat{\mathscr{A}} = \mathscr{A}$.

However, the main difficulty in this plan is that it is generally difficult to construct *symmetric* signatures from *asymmetric* signatures in a *planar* fashion, especially for arity greater than 3. Therefore, a main engine of the proof is *arity reduction*. Starting from a nonproduct type signature of arity $n > 3$, we construct in the Pl-#CSP setting a nonproduct type signature of arity $n - 1$. Then, in an arduous proof (given in section 7), we show how to construct, from any nonproduct type signature of arity 3, *either* a binary nonproduct type signature *or* a *symmetric* and nonproduct type signature of arity 3. Lemma 3.3 turns a binary nonproduct type signature into a *symmetric and nonproduct type* signature. These constructions will need suitable unary signatures which will be constructed starting with that $[1, w]$ constructed in Pl-Holant($\widehat{\mathcal{EQ}}, \widehat{\mathcal{F}}$). The derivative operator (Definition 2.6) will be used throughout.

For the construction of nonaffine signatures, we will employ a *Tableau Calculus*. This is adapted from Dyer and Richerby [23], to whom it should be credited. Again we will carry out an arity reduction proof, this time all the way down to arity 1. We prove that with the help of unary signatures $[1, 0], [0, 1], [1, x]$ with any complex number $x \neq 0$, we can get a unary nonaffine signature from any nonaffine signature of higher arity in the Pl-Holant setting (Lemma 3.9). This proof heavily depends on the Tableau Calculus. Then we construct $[1, 0], [0, 1], [1, x]$ by shuttling between Pl-#CSP($\mathcal{F}$) and Pl-Holant($\widehat{\mathcal{EQ}}, \widehat{\mathcal{F}}$). There is an exceptional case where all signatures in $\mathcal{F}$ are $\{0, 1\}$-valued in Pl-#CSP($\mathcal{F}$). In this case, we cannot construct $[1, 0], [0, 1], [1, x]$ simultaneously. We resolve this case separately. For $\{0, 1\}$-valued $\mathcal{F}$, we actually also cannot construct all the unary signatures in the arity reduction proof for nonproduct type if we *only* assume the existence of some $g \in \mathcal{F} \setminus \mathscr{P}$. However, if we have both $g \in \mathcal{F} \setminus \mathscr{P}$ and some $f \in \mathcal{F} \setminus \mathscr{A}$, we can use f to produce the needed unary signatures to help the arity reduction on g. All of these use the Tableau Calculus.

The second main case is when all signatures in $\widehat{\mathcal{F}}$ satisfy the Parity Condition. In this case, if $\widehat{\mathcal{F}} \subseteq \mathscr{A}$, or $\widehat{\mathcal{F}} \subseteq \widehat{\mathscr{P}}$, or $\widehat{\mathcal{F}} \subseteq \mathscr{M}$, then the problem is tractable. These are the exact tractability criteria according to the dichotomy theorem to be proved, Theorem 6.1. However, due to the Parity Condition, there are really only two consequential conditions here, $\widehat{\mathcal{F}} \subseteq \mathscr{A}$ and $\widehat{\mathcal{F}} \subseteq \mathscr{M}$; the containment $\widehat{\mathcal{F}} \subseteq \widehat{\mathscr{P}}$ is subsumed by $\widehat{\mathcal{F}} \subseteq \mathscr{A}$. Therefore, we want to prove that if $\widehat{\mathcal{F}} \not\subseteq \mathscr{A}$ and $\widehat{\mathcal{F}} \not\subseteq \mathscr{M}$, then

Pl-Holant$(\widehat{\mathcal{EQ}}, \widehat{\mathcal{F}})$ is #P-hard.

Again a natural idea is to construct nonaffine and nonmatchgate *symmetric* signatures from any such asymmetric signatures, and then we can apply the known dichotomy Theorem 2.34. The main difficulty of this approach lies in dealing with nonmatchgate signatures. Note that both $\widehat{\mathcal{F}}$ and $\widehat{\mathcal{EQ}}$ (being a subset of $\mathscr{M}$) satisfy the Parity Condition, and therefore the signature of any construction from an $(\widehat{\mathcal{EQ}} \cup \widehat{\mathcal{F}})$-gate must also satisfy the Parity Condition. By Lemma 2.29, any signature of arity at most 3 is a matchgate signature iff it satisfies the Parity Condition. Hence all constructible nonmatchgate signatures have arity ≥ 4. But it is difficult to construct a symmetric signature from any asymmetric signature of arity ≥ 4 while preserving planarity.

So we take an alternative approach. For a given nonmatchgate signature, we first prove that we can get a nonmatchgate signature f of arity 4. Then we can construct either the crossover function $\mathfrak{X}$ or $(=_4)$ from f. If we have the crossover function $\mathfrak{X}$, we can finish the proof by the nonplanar #CSP dichotomy Theorem 2.33. If we have $(=_4)$, then we can construct $(=_{2k})$ for any $k \in \mathbb{Z}^+$ in Pl-Holant$(\widehat{\mathcal{EQ}}, \widehat{\mathcal{F}})$. Thus we get all $\mathcal{EQ}_2$. This implies that (by (2.2))

$$\text{Pl-Holant}(\widehat{\mathcal{EQ}}, \widehat{\mathcal{F}}) \equiv_{\mathrm{T}} \text{Pl-Holant}(\mathcal{EQ}_2, \widehat{\mathcal{EQ}}, \widehat{\mathcal{F}}) \equiv_{\mathrm{T}} \text{Pl-\#CSP}^2(\widehat{\mathcal{EQ}}, \widehat{\mathcal{F}}). \tag{2.10}$$

Now comes a "cognitive dissonance." Everything in (2.10) is usually considered to be on the RHS in the equivalence

$$\text{Pl-\#CSP}(\mathcal{F}) \equiv_{\mathrm{T}} \text{Pl-Holant}(\widehat{\mathcal{EQ}}, \widehat{\mathcal{F}}).$$

But now by the last form in (2.10) it will be treated as a Pl-#CSP^2 problem with the function set $\widehat{\mathcal{EQ}} \cup \widehat{\mathcal{F}}$:

$$\begin{array}{c} \text{Pl-\#CSP}(\mathcal{F}) \equiv_{\mathrm{T}} \text{Pl-Holant}(\widehat{\mathcal{EQ}}, \widehat{\mathcal{F}}) \\ \| |_{\mathrm{T}} \\ \text{Pl-\#}\,\text{CSP}^2(\widehat{\mathcal{EQ}}, \widehat{\mathcal{F}}). \end{array}$$

A Pl-#CSP^2 problem is more in line with a Pl-#CSP problem. For Pl-#CSP^2 problems over symmetric signatures, Theorem 2.35 says that there are five tractability classes $\mathscr{P}, \mathscr{A}, \mathscr{A}^{\dagger}, \widehat{\mathscr{M}}$, and $\widehat{\mathscr{M}}^{\dagger}$. But now we will apply these on the "dual side" $\widehat{\mathcal{EQ}} \cup \widehat{\mathcal{F}}$, instead of the "primal side" $\mathcal{F}$. The "cognitive dissonance" is that the transformation from $(\mathcal{EQ}, \mathcal{F}) \mapsto (\widehat{\mathcal{EQ}}, \widehat{\mathcal{F}})$ is precisely for the purpose of transforming $\mathcal{EQ}$ to be a subset $\widehat{\mathcal{EQ}}$ of $\mathscr{M}$, but now we will subject $\widehat{\mathcal{EQ}}$ to tractability tests including $\widehat{\mathscr{M}}$ and $\widehat{\mathscr{M}}^{\dagger}$. But clearly $\widehat{\mathcal{EQ}}$ contains both $[1,0] \notin \widehat{\mathscr{M}} \cup \widehat{\mathscr{M}}^{\dagger}$ and $[1,0,1,0] \notin \mathscr{P} \cup \mathscr{A}^{\dagger}$; therefore, the only remaining possibility for tractability is $\mathscr{A}$.

Of course, if $\widehat{\mathcal{F}} \subseteq \mathscr{A}$, then Pl-Holant$(\widehat{\mathcal{EQ}}, \widehat{\mathcal{F}})$ is tractable. Suppose $\widehat{\mathcal{F}} \not\subseteq \mathscr{A}$, and we want to construct a *symmetric* nonaffine signature. We produce such a signature of arity 2 by arity reduction. From any $f \in \widehat{\mathcal{F}} \setminus \mathscr{A}$, which satisfies the Parity Condition, we can first get a nonaffine signature satisfying the even Parity Condition. Then every signature constructible from that has even parity, as $\widehat{\mathcal{EQ}}$ also has even parity. Any nonaffine binary signature satisfying the even Parity Condition is automatically symmetric. This part of the proof is the content of section 4 (Theorem 4.9).

A technical difficulty is that when $\widehat{\mathcal{F}}$ satisfies the even Parity Condition, it is impossible to construct $[0,1]$. Instead we find that we can try to construct $[0,1]^{\otimes 2}$ and prove that $[0,1]^{\otimes 2}$ is almost as good as $[0,1]$ with the help of $[1,0,1,0] \in \widehat{\mathcal{EQ}}$.

Then there are three cases. (1) If some function in $\widehat{F}$ does not take values in $\{0, 1, -1\}$ up to a constant, then we can construct $[0, 1]^{\otimes 2}$ and $[1, 0, -1]$ and complete the proof. (2) If every function in $\widehat{\mathcal{F}}$ takes values in $\{0, 1, -1\}$ up to a constant but not every function in $\widehat{\mathcal{F}}$ takes values in $\{0, 1\}$ up to a constant, then we can construct $[1, 0, -1]$ and complete the proof. (3) If every function in $\widehat{\mathcal{F}}$ takes values in $\{0, 1\}$ up to a constant, then we prove it separately. In all cases, we use the Tableau Calculus.

This completes an outline of the proof guided by an overall vision that (A) there is a dichotomy, and (B) the right form of this dichotomy is as stated in Theorem 6.1.

Of course, as a proof strategy, logically this is a bit self-serving. Essentially we want the validity of the very statement we want to prove to provide its own guarantee of success in every step in its proof. Given the fact that there are other tractable classes for Pl-Holant problems [6] not encompassed in the list given in Theorem 6.1, the validity of this vision for Pl-#CSP problems is at least not obvious. Luckily, this vision is correct. And therefore, the self-serving plan becomes a reliable guide to the proof—a bit self-fulfilling. Sometimes the statement of a theorem helps its own proof.

We include a dependency graph of the proof of Theorem 6.1 as a very high level summary of its logical flow chart (see Figure 5).

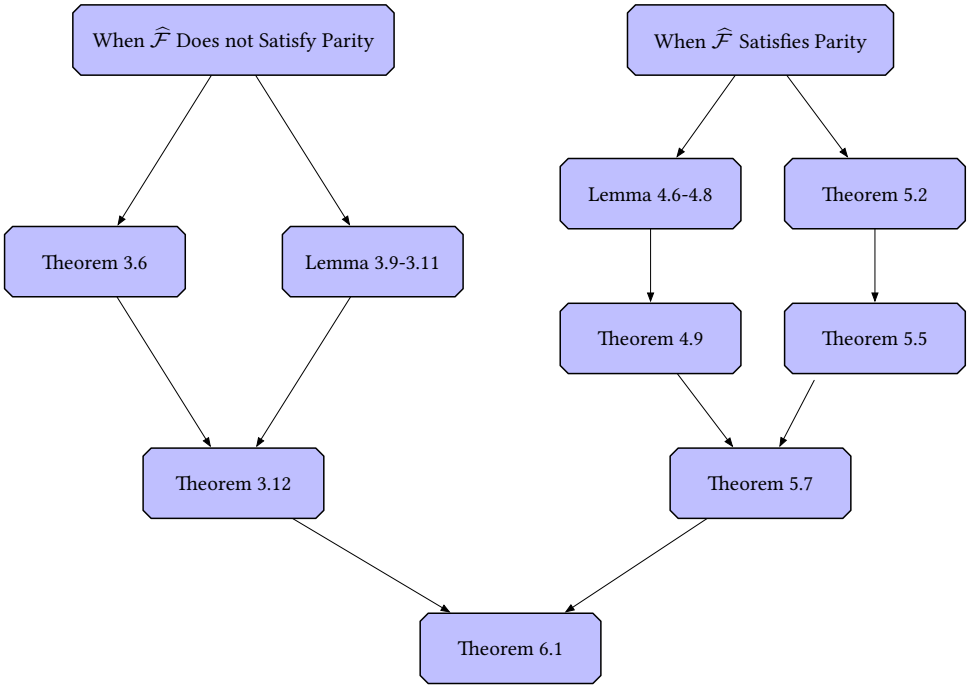


FIG. 5. *Dependency graph of the proof of Theorem* 6.1*: Theorem* 3.6 *and Lemmas* 3.9–3.11 *deal with the nonproduct arity reduction and nonaffine arity reduction, respectively, when* $\widehat{\mathcal{F}}$ *does not satisfy the Parity Condition. Theorem* 3.12 *is the dichotomy theorem for Pl-#CSP(*$\mathcal{F}$*) when* $\widehat{\mathcal{F}}$ *does not satisfy the Parity Condition. Lemmas* 4.6–4.8 *and Theorem* 5.2 *deal with the nonaffine arity reduction and nonmatchgate arity reduction, respectively, when* $\widehat{\mathcal{F}}$ *satisfies the Parity Condition. Theorems* 4.9 *and* 5.5 *are dichotomy theorems for Pl-Holant(*$\widehat{\mathcal{EQ}}, \widehat{\mathcal{F}}$*) when we have* $=_4$ *and* $[1, 0, x]$ *with* $x^4 \neq 0, 1$*, respectively. Theorem* 5.7 *is the dichotomy theorem for Pl-#CSP(*$\mathcal{F}$*) when* $\widehat{\mathcal{F}}$ *satisfies the Parity Condition.*

2.11. A sample of problems. We illustrate the scope of Theorem 1.1 by several problems.

Problem : Pl-λ-Ising
Input : An undirected planar graph G.
Output : $\sum_\sigma \lambda^{e_{\neq}(\sigma)}$, where σ is an assignment of $\{0,1\}$ to each vertex of G, and $e_{\neq}(\sigma)$ is the number of edges whose two endpoints have different values.

This problem can be expressed as Pl-#CSP($[1,\lambda,1]$). The binary signature $[1,\lambda,1]$ has the signature matrix $\left[\begin{smallmatrix}1 & \lambda\\ \lambda & 1\end{smallmatrix}\right]$. Since $[1,\lambda,1] \in \widehat{\mathscr{M}}$, Pl-$\lambda$-Ising is tractable by Theorem 1.1. Similarly, an asymmetric problem Pl-λ-AsymIsing Pl-#CSP($(1,\lambda,-\lambda,-1)$) is also tractable, since $(1,\lambda,-\lambda,-1) \in \widehat{\mathscr{M}}$. The asymmetric binary signature $(1,\lambda,-\lambda,-1)$ has signature matrix $\left[\begin{smallmatrix}1 & \lambda\\ -\lambda & -1\end{smallmatrix}\right]$.

Kasteleyn [34] and Fisher [24] gave a polynomial-time algorithm for Pl-λ-Ising in the 1960s. This was a breakthrough for statistical physics. The result for Pl-λ-AsymIsing is new to the best of our knowledge. A more general model is the so-called 2-spin systems. They can be expressed as #CSP(f), where f is a binary signature with the matrix form $\left[\begin{smallmatrix}a & b\\ c & d\end{smallmatrix}\right]$. Now we can fully generalize the result of Kasteleyn and Fisher by Theorem 1.1, which gives a complete characterization in terms of (a,b,c,d); e.g., the spin system Pl-#CSP(f) is tractable but #CSP(f) is #P-hard iff $a = \epsilon d$ and $b = \epsilon c$ for $\epsilon = \pm 1$, and $ab \neq 0$ and $a^4 \neq b^4$. (If $a = \epsilon d$, $b = \epsilon c$, but $ab = 0$ or $a^4 = b^4$, #CSP(f) is tractable even for nonplanar graphs.) To quote from the classical paper by Jerrum and Sinclair [32], "The search for efficient computational solutions to these problems has proved extremely hard and has generated a vast body of literature. A major breakthrough was achieved in the early 1960s by Kasteleyn [34] and Fisher [24], ... This must rank as one of the highlights in the field of combinatorial algorithms. It remains the state of the art as far as exact solutions are concerned; in particular, it does not appear to generalise to nonplanar systems." Theorem 1.1 (when applied to the special case of a single binary constraint function) gives a complete answer to this question.

Problem : Pl-(λ,μ)-VC
Input : A directed planar graph G.
Output : $\sum_{C \in \mathcal{C}(G)} \lambda^{e_{\rightarrow}(C)} \mu^{e_{\leftarrow}(C)}$, where $\mathcal{C}(G)$ denotes the set of all vertex covers of G, and $e_{\leftarrow}(C)$ is the number of directed edges (u,v) with source $u \notin C$ and sink $v \in C$, and $e_{\rightarrow}(C)$ is the opposite, $u \in C$ and $v \notin C$.

This problem can be expressed as Pl-#CSP($(0,\lambda,\mu,1)$), with signature matrix $\left[\begin{smallmatrix}0 & \lambda\\ \mu & 1\end{smallmatrix}\right]$. When $\lambda = \mu = 1$, it is the classical counting problem of vertex covers over planar graphs. By Theorem 1.1, we can easily show that this problem is #P-hard for $\lambda\mu \neq 0$, and tractable otherwise.

Relatedly, we can prove the #P-hardness for the hardcore gas model, which can be defined as Pl-#CSP($\left[\begin{smallmatrix}0 & 1\\ 1 & 1\end{smallmatrix}\right], [1,\lambda]$), with one binary function and one unary function.

Problem : Pl-AntiChains
Input : A finite partially ordered set $(P;\leq)$ represented by a planar directed acyclic graph G.
Output : The number of antichains of the poset $(P;\leq)$.

A directed acyclic graph G represents a partial order by transitive closure of the directed edge relation; thus $u \leq v$ iff there is a directed path from u to v. Provan and Ball [40] proved that this problem is #P-hard for general graphs. Bulatov and Dalmau [2] proved that AntiChains is equivalent to #CSP(f) for general graphs,

where f is a binary signature with $f_{00} = f_{01} = f_{11} = 1$ and $f_{10} = 0$. This is the Boolean IMPLICATION function. The signature matrix of f is $\left[\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right]$. This #CSP problem counts the number of upward closed subsets in the partial order. The equivalence is also valid for planar graphs. So Pl-ANTICHAINS is equivalent to Pl-#CSP(f). The problem Pl-#CSP(f) is #P-hard even for planar graphs by Theorem 1.1. Hence the corresponding problem Pl-ANTICHAINS is also #P-hard. To the best of our knowledge, this is the first proof that this problem is #P-hard for planar graphs.

Theorem 1.1 gives a precise demarcation of what #P-hard #CSP problems on general graphs become tractable on planar graphs. This is precisely captured by holographic algorithms with matchgates (i.e., the class $\widehat{\mathscr{M}}$). This class gives us some highly nontrivial problems which can be computed in polynomial time. Furthermore the boundary is delicate.

(a) $x_1 \neq x_3$ and $x_2 \neq x_4$ (b) $x_1 \neq x_4$ and $x_2 \neq x_3$

FIG. 6. *The assignments to the variables on the dashed lines (solid lines) have to be different.*

Consider the following pair of similar-looking problems.

Problem : Pl-WEIGHTEDCROSSCOMPLEMENTARY
Input : A planar signature grid $\Omega = (G, \pi)$, where $G = (V, E)$ is an undirected bipartite graph with $V = V_1 \dot{\cup} V_2$, and all vertices in V_2 have degree 4. Each vertex $v \in V_1$ of $\deg(v) = d$ is assigned $=_d$ by π, and each vertex in V_2 is assigned $f = f_{x_1 \neq x_3, x_2 \neq x_4}$, where $M_{x_1x_2,x_4x_3}(f) = \left[\begin{smallmatrix} f_{0000} & f_{0010} & f_{0001} & f_{0011} \\ f_{0100} & f_{0110} & f_{0101} & f_{0111} \\ f_{1000} & f_{1010} & f_{1001} & f_{1011} \\ f_{1100} & f_{1110} & f_{1101} & f_{1111} \end{smallmatrix}\right] = \left[\begin{smallmatrix} 0 & 0 & 0 & a \\ 0 & b & 0 & 0 \\ 0 & 0 & c & 0 \\ d & 0 & 0 & 0 \end{smallmatrix}\right]$ with $abcd \neq 0$ (Figure 6a).

Output : $\text{Holant}(\Omega; \mathcal{EQ} \mid f) = \sum_{\sigma: E \to \{0,1\}} \prod_{v \in V} f_v(\sigma|_{E(v)})$, where $f_v \in \{f\} \cup \mathcal{EQ}$ is the function assigned to the vertex v by the mapping π.

Problem : Pl-WEIGHTEDNEARBYCOMPLEMENTARY
Input : A planar signature grid $\Omega = (G, \pi)$, the same as in Pl-WEIGHTEDCROSSCOMPLEMENTARY, but each vertex in V_2 is assigned to $g = g_{x_1 \neq x_4, x_2 \neq x_3}$, where $M_{x_1x_2,x_4x_3}(g) = \left[\begin{smallmatrix} 0 & 0 & 0 & a' \\ 0 & 0 & b' & 0 \\ 0 & c' & 0 & 0 \\ d' & 0 & 0 & 0 \end{smallmatrix}\right]$ with $a'b'c'd' \neq 0$. (See Figure 6b.)
Output : $\text{Holant}(\Omega; \mathcal{EQ} \mid g) = \sum_{\sigma: E \to \{0,1\}} \prod_{v \in V} f_v(\sigma|_{E(v)})$, where $f_v \in \{g\} \cup \mathcal{EQ}$ is the function assigned to the vertex v by the mapping π.

The necessary and sufficient conditions on (a, b, c, d) and (a', b', c', d') for $f_{x_1 \neq x_3, x_2 \neq x_4}$ and $g_{x_1 \neq x_4, x_2 \neq x_3}$ to belong to $\mathscr{P}$ are that there are at least two zero elements in $\{a, b, c, d\}$ and in $\{a', b', c', d'\}$, respectively. Since $abcd \neq 0$ and $a'b'c'd' \neq 0$ in our problems, neither $f_{x_1 \neq x_3, x_2 \neq x_4}$ nor $g_{x_1 \neq x_4, x_2 \neq x_3}$ is in $\mathscr{P}$. Similarly, the necessary and sufficient conditions for membership in $\mathscr{A}$ are that $abcd = \pm 1$ and $a'b'c'd' = \pm 1$, respectively. Therefore, the conditions for $f_{x_1 \neq x_3, x_2 \neq x_4}$ and for $g_{x_1 \neq x_4, x_2 \neq x_3}$ to belong to $\mathscr{P} \cup \mathscr{A}$, which are the tractability conditions for nonplanar #CSP, have exactly the same expression. But for $\widehat{\mathscr{M}}$, the situation is very different. For example, consider

the following two very similar cases:

- Pl-WeightedCrossComplementary with

$$M_{x_1x_2,x_4x_3}(f) = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & c & 0 & 0 \\ 0 & 0 & c & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix},$$

  $c^4 \neq 0, 1$, and
- Pl-WeightedNearbyComplementary with

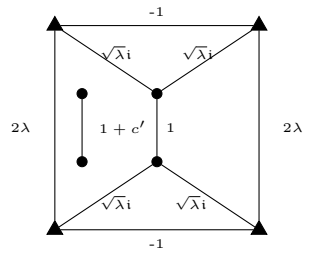
$$M_{x_1x_2,x_4x_3}(g) = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & c' & 0 \\ 0 & c' & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix},$$

  $(c')^4 \neq 0, 1$.

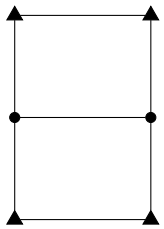
We have $f \notin \mathscr{P} \cup \mathscr{A} \cup \widehat{\mathscr{M}}$, and so the former problem is #P-hard by Theorem 1.1. But the latter problem is tractable. Let $\widehat{g} = H_2^{\otimes 4} g$. Then the signature matrix of $\widehat{g}$ is

$$M_{x_1x_2,x_4x_3}(\widehat{g}) = \begin{bmatrix} 1+c' & 0 & 0 & 1-c' \\ 0 & -(1+c') & -(1-c') & 0 \\ 0 & -(1-c') & -(1+c') & 0 \\ 1-c' & 0 & 0 & 1+c' \end{bmatrix}.$$

One can verify that $\widehat{g} \in \mathscr{M}$ by MGIs [7], or by a direct construction of a matchgate in Figure 7a, which realizes the signature. Thus the problem is tractable by FKT. Note that this problem is #P-hard without the planar restriction by Theorem 3.1 in [20] since $g \notin \mathscr{P} \cup \mathscr{A}$.



(a) The matchgate realizing $\widehat{g}$ with $\lambda = \frac{1-c'}{1+c'}$



(b) The matchgate realizing $\widehat{h}$

FIG. 7. *In both matchgates, the triangles are external nodes and all other nodes are internal nodes. In Figure* 7b*, all the edges have weight* 1.

In fact, a simple matchgate can give us a highly nontrivial problem that is tractable for planar graphs but is #P-hard for general graphs. For example, consider the signature h with the signature matrix

$$M_{x_1x_2,x_4x_3}(h) = \begin{bmatrix} 5 & 1 & 1 & -1 \\ 1 & 3 & 1 & 1 \\ 1 & 1 & 3 & 1 \\ -1 & 1 & 1 & 4 \end{bmatrix}.$$

Let $\widehat{h} = H_2^{\otimes 4} h$; then

$$M_{x_1x_2,x_4x_3}(\widehat{h}) = \begin{bmatrix} 3 & 0 & 0 & 0 \\ 0 & 2 & 1 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

The signature $\widehat{h}$ can be realized by the matchgate in Figure 7b. Thus $\widehat{h} \in \mathscr{M}$ and Pl-#CSP(h) is tractable.

Some tractable problems can appear rather unintuitive. Let us consider the following problem Pl-CrazyPell.

Problem: Pl-CrazyPell
Input : A planar signature grid $\Omega = (G, \pi)$, where $G = (V, E)$ is an undirected bipartite graph with $V = V_1 \dot\cup V_2$, and all vertices in V_2 have degree 4. Each vertex $v \in V_1$ of $\deg(v) = d$ is assigned $=_d$ by π, and each vertex in V_2 is assigned f whose signature matrix is $M(f)$, where

$$M(f) = \begin{bmatrix} 669669112435114949 & -598015350142588611 & 598015350142588607 & -669669112435114945 \\ 533639108484318913 & -476540387460305851 & 476540387460305855 & -533639108484318909 \\ -533639108484318909 & 476540387460305855 & -476540387460305851 & 533639108484318913 \\ -669669112435114945 & 598015350142588607 & -598015350142588611 & 669669112435114949 \end{bmatrix}.$$

Output : $\text{Holant}(\Omega; \mathcal{EQ} \mid f) = \sum_{\sigma: E \to \{0,1\}} \prod_{v \in V} f_v(\sigma|_{E(v)})$, where $f_v \in \{f\} \cup \mathcal{EQ}$ is the function assigned to the vertex v by the mapping π.

Let $\widehat{f} = H_2^{\otimes 4} f$, then $\widehat{f}$ has the signature matrix

$$4 \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 64376241658269698 & 3638760317128320 & 0 \\ 0 & 569465989630582080 & 32188120829134849 & 0 \\ -1 & 0 & 0 & 1 \end{bmatrix}.$$

One can verify that $\widehat{f} \in \mathscr{M}$ by MGIs [7]. Thus $f \in \widehat{\mathscr{M}}$ and #CSP(f) is tractable.

Of course a natural reaction one may have when faced with such an incomprehensible looking counting problem is "This looks rather dull, and why would anyone be interested in it?" We take a rather different view. As reminded by the Hardy–Ramanujan "taxicab number" 1729, who is to say which number is dull and which number is interesting? We posit that complexity classes such as #P or #CSP are interesting mathematical objects, and as such the internal structures of these classes and the inner connections among *all* problems within a class are a Platonic reality. This is no longer subject to our personal taste (and, to our taste, they are also beautiful). The strength of a complexity classification theorem such as Theorem 1.1 is precisely that it applies to *every* problem in the class, regardless of whether it looks interesting or dull to anyone. (The underlying reason for the problem Pl-CrazyPell to be tractable over planar graphs is that $(32188120829134849, 1819380158564160)$ is the smallest integer solution to the Pell's equation $x^2 - 313y^2 = 1$. This enables a suitable matchgate to be constructed. And there are infinitely many such problems.)

3. When $\widehat{\mathcal{F}}$ does not satisfy parity. The following lemma shows that if there is a signature in $\widehat{\mathcal{F}}$ that does not satisfy the Parity Condition, then in Pl-Holant$(\widehat{\mathcal{EQ}}, \widehat{\mathcal{F}})$, we can construct a unary signature which does not satisfy the Parity Condition.

Lemma 3.1. *If $\widehat{\mathcal{F}}$ contains a signature f that does not satisfy the Parity Condition, then we can construct a unary signature $[1, w]$ with $w \neq 0$ in* Pl-Holant$(\widehat{\mathcal{EQ}}, \widehat{\mathcal{F}})$, *such that*

$$\text{Pl-Holant}(\widehat{\mathcal{EQ}}, [1, w], \widehat{\mathcal{F}}) \leq_{\text{T}} \text{Pl-Holant}(\widehat{\mathcal{EQ}}, \widehat{\mathcal{F}}).$$

Proof. Let f have arity $n \geq 1$. Since f does not satisfy the Parity Condition, there is a nonzero entry f_α with minimum odd Hamming weight $\text{wt}(\alpha)$ among all nonzero entries f_η of odd Hamming weight $\text{wt}(\eta)$. Since $[1, 0] \in \widehat{\mathcal{EQ}}$, we can construct the signature $\partial^S_{[1,0]}(f) = (f_{00\cdots 0}, \ldots, f_\alpha)$, where $S = \{k \mid \text{the } k\text{th bit of } \alpha \text{ is } 0\}$. Note that the signature $\partial^S_{[1,0]}(f)$ has an odd arity $\text{wt}(\alpha) = n - |S|$, where every entry having an odd weight index is 0 except for f_α, by the minimality of $\text{wt}(\alpha)$ among all nonzero

entries of f of odd weight. Then by connecting all variables $x_2, \ldots, x_{\mathrm{wt}(\alpha)}$ of $\partial^S_{[1,0]}(f)$ in adjacent pairs using $=_2$ in a planar way, we get the unary signature $[a, f_\alpha]$, where a is the sum of entries of $\partial^S_{[1,0]}(f)$ with even index. Note that the entry f_α remains unchanged, since when connecting adjacent $\frac{\mathrm{wt}(\alpha)-1}{2}$ pairs of variables of $\partial^S_{[1,0]}(f)$ using $=_2$, only entries of f with lower odd indices, which are all zero, are combined with f_α. If $a \neq 0$, then we already have the desired $[1, w]$ by normalization, where $w = \frac{f_\alpha}{a}$.

If $a = 0$, then we have $[0, 1]$ up to the nonzero scalar f_α. Since f does not satisfy the parity condition, there exist β and $\gamma \in \{0,1\}^n$, satisfying the following: $f_\beta \neq 0, f_\gamma \neq 0$, $\mathrm{wt}(\beta)$ and $\mathrm{wt}(\gamma)$ have opposite parity and

$$d = \mathrm{wt}(\beta \oplus \gamma) = \min_{\zeta,\eta \in \{0,1\}^n} \{\mathrm{wt}(\zeta \oplus \eta) \mid \mathrm{wt}(\zeta) \text{ and } \mathrm{wt}(\eta) \text{ have opposite parity, } f_\zeta \neq 0,\ f_\eta \neq 0\}.$$

Then we have $g = \partial^{S_0}_{[1,0]}[\partial^{S_1}_{[0,1]}(f)]$, where, for $b \in \{0,1\}$,

$$S_b = \{k \mid \text{the } k\text{th bits of both } \beta \text{ and } \gamma \text{ are } b\}.$$

Note that the arity of g is d.

By deleting all bits in $S_0 \cup S_1$ from β and γ, we get two bit strings $\beta', \gamma' \in \{0,1\}^d$, respectively. We have $g_{\beta'} = f_\beta$ and $g_{\gamma'} = f_\gamma$ and all other entries of g are 0. Note that β' and γ' have opposite parity since β and γ have opposite parity. Without loss of generality, assume that $\mathrm{wt}(\beta')$ is odd and $\mathrm{wt}(\gamma')$ is even. Then $\partial_g([1,0,1,\ldots,0,1]) = [g_{\beta'}, g_{\gamma'}] = [f_\beta, f_\gamma]$ and we are done, where $[1,0,1,\ldots,0,1] = \frac{1}{2}\{[1,1]^{\otimes d+1} + [1,-1]^{\otimes d+1}\} \in \widehat{\mathcal{EQ}}$. ∎

The next lemma is a simple fact from linear algebra [29]. It will be used in the proof of Lemma 3.3.

LEMMA 3.2. *Let* $\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d} \in \mathbb{C}^2$*, and suppose* $\mathbf{c}, \mathbf{d}$ *are linearly independent. Suppose for some* $n \geq 3$ *we have* $\mathbf{a}^{\otimes n} + \mathbf{b}^{\otimes n} = \mathbf{c}^{\otimes n} + \mathbf{d}^{\otimes n}$*. Then* $\mathbf{a} = \xi\mathbf{c}$, $\mathbf{b} = \eta\mathbf{d}$ *or* $\mathbf{a} = \xi\mathbf{d}$, $\mathbf{b} = \eta\mathbf{c}$ *for some* $\xi^n = \eta^n = 1$*.*

3.1. Arity reduction for nonproduct-type signatures. Our plan is to use the dichotomy theorems for symmetric signatures. For that we have to construct symmetric signatures from asymmetric signatures. For example, starting from a signature not in $\mathscr{P}$, we want to construct a symmetric signature not in $\mathscr{P}$. It is generally difficult to construct symmetric signatures from asymmetric signatures in a planar construction, especially when the arity is high. So one of our main techniques is arity reduction. We want to reduce the arity of a signature while keeping it outside $\mathscr{P}$. Every unary signature is in $\mathscr{P}$. So the lowest arity outside $\mathscr{P}$ is two. If we obtain a binary signature $f = (f_{00},\ f_{01},\ f_{10},\ f_{11}) = (a,\ b,\ c,\ d) \notin \mathscr{P}$, we can take 3 copies of f and connect the first input of each f to an edge of $(=_3)$ and leave the second input of the 3 copies of f as 3 dangling edges. This planar gadget has the symmetric signature $[a,b]^{\otimes 3} + [c,d]^{\otimes 3}$. The following lemma says that this ternary symmetric signature does not belong to $\mathscr{P}$. Our main construction for a symmetric signature not in $\mathscr{P}$ will be an induction on arity n with a base case at $n = 3$. The reason we start at $n = 3$ is that certain steps for $n \geq 3$ will not work for $n = 2$. However, Lemma 3.3 implies that if we have a binary signature that is not in $\mathscr{P}$, then we can construct in Pl-#CSP(f) a symmetric signature that is not in $\mathscr{P}$.

LEMMA 3.3. *For any binary signature* $f = (a,\ b,\ c,\ d)$*,* $f \in \mathscr{P}$ *iff* $g = [a,b]^{\otimes 3} + [c,d]^{\otimes 3} \in \mathscr{P}$*.*

Proof. If $f \in \mathscr{P}$, then either f is degenerate or $a = d = 0$ or $b = c = 0$ by definition. If f is degenerate, then $[a,b], [c,d]$ are linearly dependent. Then g is

degenerate and $g \in \mathscr{P}$. If $a = d = 0$ or $b = c = 0$, then g is a GENERALIZED EQUALITY and $g \in \mathscr{P}$.

Conversely, if $g \in \mathscr{P}$, then either g is degenerate or g is a GENERALIZED EQUALITY since g is symmetric.

If g is degenerate, then there exists a vector $[e, f]$ such that

$$g = [a, b]^{\otimes 3} + [c, d]^{\otimes 3} = [e, f]^{\otimes 3}.$$

To use Lemma 3.2, we rewrite it as

$$g = [a, b]^{\otimes 3} + [c, d]^{\otimes 3} = [e, f]^{\otimes 3} + [0, 0]^{\otimes 3}.$$

If $[a, b], [c, d]$ are linearly independent, then $[a, b] = [0, 0]$ or $[c, d] = [0, 0]$. This contradicts that $[a, b], [c, d]$ are linearly independent. Thus $[a, b], [c, d]$ are linearly dependent. This implies that f is degenerate. Thus $f \in \mathscr{P}$.

If g is a GENERALIZED EQUALITY, then there exist x, y such that

$$g = [a, b]^{\otimes 3} + [c, d]^{\otimes 3} = [x, 0]^{\otimes 3} + [0, y]^{\otimes 3}.$$

If $[a, b], [c, d]$ are linearly dependent, then f is degenerate. Thus $f \in \mathscr{P}$. If $[a, b], [c, d]$ are linearly independent, by Lemma 3.2, there exist ω_1, ω_2, where $\omega_1^3 = \omega_2^3 = 1$, such that $[a, b] = \omega_1[x, 0], [c, d] = \omega_2[0, y]$ or $[a, b] = \omega_1[0, y], [c, d] = \omega_2[x, 0]$. This implies that $b = c = 0$ or $a = d = 0$. Thus $f \in \mathscr{P}$. ∎

We will use the next two lemmas in the proof of Theorem 3.6. Recall that in Definition 2.25 we defined the situation in which two signatures are said to have compatible type.

LEMMA 3.4. *Suppose f is a signature of arity $n \geq 3$, and $[1, a]$ and $[1, b]$ are two unary signatures. Let $P = \partial^{\{1\}}_{[1,a]}(f)$ and $Q = \partial^{\{1\}}_{[1,b]}(f)$. Suppose both $P \in \mathscr{P}$ and $Q \in \mathscr{P}$ and neither is identically zero. If P and Q do not have compatible type, then $f \notin \mathscr{P}$.*

Proof. For a contradiction suppose $f \in \mathscr{P}$. Then there exists a primitive decomposition of $f = \prod_{i=1}^{k} F_i(X|_{I_i})$ with partition $\mathcal{I} = \{I_1, I_2, \ldots, I_k\}$ of $[n]$ and signatures $F_1, F_2, \ldots, F_k$. Without loss of generality, suppose $1 \in I_1$. If $|I_1| = 1$, then $\partial^{\{1\}}_{[1,a]}(F_1)$ is a nonzero constant; it is nonzero because $P = \partial^{\{1\}}_{[1,a]}(F_1) \cdot \prod_{i=2}^{k} F_i(X|_{I_i})$ is not identically zero. Similarly $Q = \partial^{\{1\}}_{[1,b]}(F_1) \cdot \prod_{i=2}^{k} F_i(X|_{I_i})$ is also a nonzero constant multiplied by the same decomposition. Hence P and Q have the same primitive decomposition. Thus they have compatible type. This is a contradiction.

So we may assume $|I_1| \geq 2$. Let $F_1' = \partial^{\{1\}}_{[1,a]} F_1$ and $F_1'' = \partial^{\{1\}}_{[1,b]} F_1$. Being a factor in a primitive decomposition of arity at least 2, F_1 is nondegenerate, and $\mathrm{supp}(F_1)$ consists of two antipodal points $\{0\alpha, 1\bar{\alpha}\}$ for some $\alpha \in \{0, 1\}^{|I_1|-1}$. If $ab \neq 0$, then both F_1' and F_1'' still have support consisting of two antipodal points $\{\alpha, \bar{\alpha}\}$, e.g., $F_1'(\bar{\alpha}) = 1F_1(0\bar{\alpha}) + aF_1(1\bar{\alpha}) = aF_1(1\bar{\alpha}) \neq 0$. In particular, both $F_1' \in \mathcal{E}$ and $F_1'' \in \mathcal{E}$. Thus $F_1' \prod_{i=2}^{k} F_i(X|_{I_i})$ and $F_1'' \prod_{i=2}^{k} F_i(X|_{I_i})$ are the primitive decompositions of P and Q, respectively. Thus P and Q have compatible type. This is a contradiction.

Now suppose $ab = 0$. Since P and Q do not have compatible type, certainly $a \neq b$. Without loss of generality, we may assume that $a = 0$ and $b \neq 0$. In this case, F_1' is further decomposed as a product of unary signatures, $F_1' = \prod_{j=1}^{|I_1|-1} G_j$, where each G_j is a nonzero scalar multiple of $[1, 0]$ or $[0, 1]$. The primitive decomposition

of P is $\prod_{j=1}^{|I_1|-1} G_j \prod_{i=2}^{k} F_i(X|_{I_i})$. The support of $\prod_{j=1}^{|I_1|-1} G_j$ is a singleton point $\{\alpha\}$, a proper subset of the support of F_1'', which is $\{\alpha, \bar{\alpha}\}$. Thus P and Q still have compatible type. This is a contradiction. ∎

LEMMA 3.5. *Suppose $f \notin \mathscr{P}$ is a signature of arity $n \geq 4$, and $[1,a],[1,b],[1,c]$ are three unary signatures that are pairwise linearly independent. Suppose further that both $P = \partial^{\{1\}}_{[1,a]}(f)$ and $Q = \partial^{\{1\}}_{[1,b]}(f)$ belong to $\mathscr{P}$ and are not identically 0, and*

$$P(X) = \prod_{i=1}^{k} P_i(X|_{I_i}) \quad \text{and} \quad Q(X) = \prod_{j=1}^{\ell} Q_j(X|_{J_j})$$

are their primitive decompositions, where the signatures $P_1, P_2, \ldots, P_k$ and $Q_1, Q_2, \ldots, Q_\ell$ are on two respective partitions $\mathcal{I} = \{I_1, I_2, \ldots, I_k\}$ and $\mathcal{J} = \{J_1, J_2, \ldots, J_\ell\}$ of $[n] \setminus \{1\}$.

If the partitions $\mathcal{I}$ and $\mathcal{J}$ are not compatible, then there exists $f' \notin \mathscr{P}$ of arity $n-1$ such that

$$\text{Pl-\#CSP}(f', [1,a],[1,b],[1,c]) \leq_{\mathrm{T}} \text{Pl-\#CSP}(f, [1,a],[1,b],[1,c]). \tag{3.1}$$

Proof. Since the partitions $\mathcal{I}$ and $\mathcal{J}$ are not compatible, by (2.5) there exist I_i and J_j such that

$$|I_i| \geq 2, \quad |J_j| \geq 2, \quad I_i \cap J_j \neq \emptyset, \quad \text{and} \quad I_i \neq J_j.$$

Without loss of generality, we assume that there exist $s, t \in [n] \setminus \{1\}$ such that $\{s,t\} \subseteq I_i$ and $s \in J_j$ but $t \notin J_j$. (Note that after this choice the roles of I_i and J_j are no longer symmetric.) Since $n \geq 4$, we have at least one other variable x_r, where $r \neq 1, s, t$.

Suppose $r \in I_p$ and $r \in J_q$ for some $p \in [k]$ and $q \in [\ell]$. We will connect x_r to some unary signature from $\{[1,a],[1,b],[1,c]\}$ in the following way. There are at least two unary signatures $u_1, u_2 \in \{[1,a],[1,b],[1,c]\}$ that are not $[1,0]$, i.e., with both entries nonzero.

- If $|I_p| \geq 2$ and $|J_q| \geq 2$, then we connect u_1 to x_r. Note that P_i (or $\partial^{\{r\}}_{u_1}(P_i)$ if $i = p$) still has at least two variables x_s and x_t and two antipodal points in its support, regardless of whether $p = i$. The function $\partial^{\{r\}}_{u_1}(Q_q)$ still has two antipodal points in its support set. Thus, regardless of whether $q = j$, the functions $\partial^{\{r\}}_{u_1}(P)$ and $\partial^{\{r\}}_{u_1}(Q)$ do not have compatible type.
- Suppose $|I_p| = 1$ and $|J_q| \geq 2$. Then P_p is a nonzero unary signature $[P_p(0), P_p(1)]$. For $1 \leq k \leq 2$, $\partial^{\{r\}}_{u_k}(P_p)$ are constants, and at most one of them can be zero. We choose one u_k such that $\partial^{\{r\}}_{u_k}(P_p) \neq 0$ and connect that u_k to x_r. In the new partition $\mathcal{I}'$ obtained from $\mathcal{I}$ by removing I_p, I_i is unchanged, still containing both s and t. P_i still has at least two variables x_s and x_t and two antipodal points in its support. The function $\partial^{\{r\}}_{u_k}(Q_q)$ still has two antipodal points in its support set because both entries of u_k are nonzero. Thus, regardless of whether $q = j$, the functions $\partial^{\{r\}}_{u_k}(P)$ and $\partial^{\{r\}}_{u_k}(Q)$ do not have compatible type.
- Suppose $|I_p| \geq 2$ and $|J_q| = 1$. Then Q_q is a nonzero unary signature $[Q_q(0), Q_q(1)]$. For $1 \leq k \leq 2$, $\partial^{\{r\}}_{u_k}(Q_q)$ are constants, and at most one of them can be zero. We choose one u_k such that $\partial^{\{r\}}_{u_k}(Q_q) \neq 0$ and connect

that u_k to x_r. In the new partition $\mathcal{I}'$, obtained by removing r from I_p, I_i still contains both s and t. P_i still has at least two variables x_s and x_t, and two antipodal points in its support, regardless of whether $i = p$. The function Q_j is unchanged and still has two antipodal points in its support set. Thus the functions $\partial_{u_k}^{\{r\}}(P)$ and $\partial_{u_k}^{\{r\}}(Q)$ do not have compatible type.

- For $|I_p| = 1$ and $|J_q| = 1$, we have $I_p = J_q = \{r\}$. Note that there exists at least one unary signature $u \in \{[1,a],[1,b],[1,c]\}$ such that $\partial_u^{\{r\}}(P_p) \neq 0$ and $\partial_u^{\{r\}}(Q_q) \neq 0$. Then we connect this u to x_r. Again, the functions $\partial_u^{\{r\}}(P)$ and $\partial_u^{\{r\}}(Q)$ do not have compatible type, as both P_i and Q_j are unchanged.

Thus after connecting x_r to a suitable unary signature u in this way, we get $P' = \partial_u^{\{r\}}(P) \in \mathscr{P}$ and $Q' = \partial_u^{\{r\}}(Q) \in \mathscr{P}$, both not identically zero, and not having compatible type. If we connect x_r in f to the unary signature u, we get $f' = \partial_u^{\{r\}}(f)$. Note that $\partial_{[1,a]}^{\{1\}}(f') = P'$ and $\partial_{[1,b]}^{\{1\}}(f') = Q'$.

$$\begin{array}{ccc} f & \xrightarrow{\partial_u^{\{r\}}} & f' \\ {\scriptstyle \partial_{[1,a]}^{\{1\}}}\downarrow & & \downarrow{\scriptstyle \partial_{[1,a]}^{\{1\}}} \\ P & \xrightarrow[\partial_u^{\{r\}}]{} & P' \end{array} \qquad \begin{array}{ccc} f & \xrightarrow{\partial_u^{\{r\}}} & f' \\ {\scriptstyle \partial_{[1,b]}^{\{1\}}}\downarrow & & \downarrow{\scriptstyle \partial_{[1,b]}^{\{1\}}} \\ Q & \xrightarrow[\partial_u^{\{r\}}]{} & Q' \end{array}$$

We use such diagrams to indicate commutativity of operations; e.g., the first diagram indicates that from f one arrives at the same P' in both alternative ways, via f' or via P.

By Lemma 3.4, $f' \notin \mathscr{P}$ and this f' satisfies (3.1). ∎

THEOREM 3.6. *Suppose $\mathcal{F}$ contains a signature $f \notin \mathscr{P}$ of arity $n \geq 3$. Let $[1,a],[1,b],[1,c]$ be three unary signatures that are pairwise linearly independent. Then there exists a symmetric signature $g \notin \mathscr{P}$ such that*

$$\text{Pl-\#CSP}(g,[1,a],[1,b],[1,c],\mathcal{F}) \leq_{\mathrm{T}} \text{Pl-\#CSP}([1,a],[1,b],[1,c],\mathcal{F}). \tag{3.2}$$

Proof. We prove the theorem by induction on n. The base case is $n = 3$ and it is done in section 7. By Theorem 7.11 we can produce $g \notin \mathscr{P}$ satisfying (3.2) such that either g has arity 2 or g is symmetric and has arity 3. If g has arity 2, we use Lemma 3.3 to produce a symmetric $g' \notin \mathscr{P}$ of arity 3.

Now assume $n \geq 4$, and the theorem is true for $n-1$. We show how to construct some $g \notin \mathscr{P}$ of arity $n-1$ satisfying (3.2). Define

$$P = \partial_{[1,a]}^{\{1\}}(f), \quad Q = \partial_{[1,b]}^{\{1\}}(f), \quad R = \partial_{[1,c]}^{\{1\}}(f).$$

If any of P, Q, or $R \notin \mathscr{P}$, then we are done by induction. So we may assume P, Q, and R all belong to $\mathscr{P}$.

Claim. P, Q, and R are pairwise linearly independent.

For a contradiction, without loss of generality, suppose P and Q are linearly dependent. Note that each of P, Q, and R is a linear combination of $f^{x_1=0}$ and $f^{x_1=1}$. From

$$\begin{bmatrix} P \\ Q \end{bmatrix} = \begin{bmatrix} 1 & a \\ 1 & b \end{bmatrix} \begin{bmatrix} f^{x_1=0} \\ f^{x_1=1} \end{bmatrix}, \tag{3.3}$$

since $a \neq b$, we have

$$\begin{bmatrix} f^{x_1=0} \\ f^{x_1=1} \end{bmatrix} = \frac{1}{b-a} \begin{bmatrix} b & -a \\ -1 & 1 \end{bmatrix} \begin{bmatrix} P \\ Q \end{bmatrix}. \tag{3.4}$$

If both P and Q are identically zero, then both $f^{x_1=0}$ and $f^{x_1=1}$ are identically zero, and so f is identically zero. This contradicts that $f \notin \mathscr{P}$. So we may assume that $P \neq 0$. Then there exists a constant λ such that $Q = \lambda P$. This implies that

$$f^{x_1=0} = \frac{b - a\lambda}{b-a} P, \quad f^{x_1=1} = \frac{\lambda - 1}{b-a} P.$$

So $f = \frac{1}{b-a}[b - a\lambda, \lambda - 1] \otimes P$. This implies that $f \in \mathscr{P}$. This is a contradiction and finishes the proof of the claim.

In the following, P, Q, and R all belong to $\mathscr{P}$, and they are pairwise linearly independent. In particular, none of them is identically zero.

- If all three pairs from $\{P, Q, R\}$ (pairwise) have compatible types, then by Lemma 2.26, there exist a common partition $\mathcal{I} = \{I_1, I_2, \ldots, I_k\}$ of $[n] \setminus \{1\}$ and signatures $P_1, P_2, \ldots, P_k$, $Q_1, Q_2, \ldots, Q_k$, and $R_1, R_2, \ldots, R_k$ such that

$$P(X) = \prod_{i=1}^{k} P_i(X|_{I_i}), \quad Q(X) = \prod_{i=1}^{k} Q_i(X|_{I_i}), \quad R(X) = \prod_{i=1}^{k} R_i(X|_{I_i}),$$

where $P_i, Q_i, R_i \in \mathcal{E}$ and there exists $\alpha_i \in \{0,1\}^{|I_i|}$ such that $\mathrm{supp}(P_i)$, $\mathrm{supp}(Q_i)$, $\mathrm{supp}(R_i) \subseteq \{\alpha_i, \overline{\alpha}_i\}$ for $1 \leq i \leq k$.
Since P and Q are linearly independent, there is at least one $1 \leq i \leq k$ such that P_i and Q_i are linearly independent.
Claim. There exists exactly one i such that P_i and Q_i are linearly independent.
Otherwise, without loss of generality, we can assume that both pairs P_1, Q_1 and P_2, Q_2 are linearly independent, respectively. Because P and Q are linearly independent, P_i and Q_i are not the zero signature for any $i \in [k]$. Choose any unary signature $u \in \{[1,a],[1,b],[1,c]\}$ that is not $[1,0]$. This is clearly possible because at most one of them can be $[1,0]$.
For any $i \in [k] \setminus \{1,2\}$, we shrink both P_i and Q_i to a nonzero constant in two steps as follows:
Step 1. If the arity $|I_i|$ of both P_i and Q_i is greater than 1 (skip Step 1 if $|I_i| = 1$), we combine $|I_i| - 1$ copies of u to both P_i and Q_i. Since P_i (resp., Q_i) is not identically zero and has either a single point in $\mathrm{supp}(P_i)$ (resp., $\mathrm{supp}(Q_i)$) or a pair of antipodal points, and both entries of the unary signature u are nonzero, this operation shrinks P_i (resp., Q_i) to a nonzero unary signature $[c_1, d_1]$ (resp., $[c_2, d_2]$), where either c_1 or $d_1 \neq 0$ (resp., c_2 or $d_2 \neq 0$).
Step 2. Since we have three unary signatures that are pairwise linearly independent, there exists at least one $u' \in \{[1,a],[1,b],[1,c]\}$ such that $\partial_{u'}([c_1, d_1])$ and $\partial_{u'}([c_2, d_2])$ are both nonzero constants. So we combine u' to $[c_1, d_1]$ and $[c_2, d_2]$ and we have shrunken both P_i and Q_i to nonzero constants.
By (3.4) we have

$$R = f^{x_1=0} + c f^{x_1=1} = \frac{1}{b-a}[(b-c)P + (c-a)Q].$$

After shrinking P_i and Q_i for every $i \in [k] \setminus \{1,2\}$, there exist constants c_P, c_Q, and c_R such that

$$c_R R_1 \otimes R_2 = c_P P_1 \otimes P_2 + c_Q Q_1 \otimes Q_2,$$

where $c_P \neq 0$ and $c_Q \neq 0$. If we write $R_1 \otimes R_2$ in its matrix form as a matrix in $\mathbb{C}^{2^{|I_1|} \times 2^{|I_2|}}$ where the rows and columns are indexed by assignments to the variables in I_1 and I_2, respectively, it has rank at most 1, because it is expressible as the product of a column vector times a row vector $R_1^{\mathsf{T}} R_2$, where R_i is a row vector in $\mathbb{C}^{2^{|I_i|}}$. But the signature matrix of $c_P P_1 \otimes P_2 + c_Q Q_1 \otimes Q_2$ is

$$\begin{bmatrix} P_1^{\mathsf{T}} & Q_1^{\mathsf{T}} \end{bmatrix} \begin{bmatrix} c_P & 0 \\ 0 & c_Q \end{bmatrix} \begin{bmatrix} P_2 \\ Q_2 \end{bmatrix},$$

which has rank 2 because P_i, Q_i are linearly independent for $i = 1, 2$, and $\begin{bmatrix} P_1^{\mathsf{T}} & Q_1^{\mathsf{T}} \end{bmatrix} \in \mathbb{C}^{2^{|I_1|} \times 2}$ and $\begin{bmatrix} P_2 \\ Q_2 \end{bmatrix} \in \mathbb{C}^{2 \times 2^{|I_2|}}$ both have rank 2. This is a contradiction. This proves the claim.

Now, without loss of generality, we may assume that P_1, Q_1 are linearly independent and P_i, Q_i are linearly dependent for $i = 2, \ldots, k$. Thus each Q_i is a nonzero multiple of P_i for $i = 2, \ldots, k$. By replacing Q_1 with a nonzero multiple of Q_1, we may assume $Q_i = P_i$ for $i = 2, \ldots, k$. We have three unary signatures that are pairwise linearly independent, so by a similar argument, we can connect the variables of f in $I_2, \ldots, I_k$ to some unary signatures such that each of $P_2, \ldots, P_k$ contributes a nonzero constant factor. Let the resulting signature be h on variables from $\{x_s \mid s \in \{1\} \cup I_1\}$. Note that $\partial^{\{1\}}_{[1,a]}(h) = \lambda P_1$, $\partial^{\{1\}}_{[1,b]}(h) = \lambda Q_1$, where λ is a nonzero constant, as the following diagrams commute.

$$\begin{array}{ccc} f & \xrightarrow{\partial^{I_2} \cdots \partial^{I_k}} & h \\ {\scriptstyle \partial^{\{1\}}_{[1,a]}}\downarrow & & \downarrow{\scriptstyle \partial^{\{1\}}_{[1,a]}} \\ P & \xrightarrow[\partial^{I_2} \cdots \partial^{I_k}]{} & \lambda P_1 \end{array} \qquad \begin{array}{ccc} f & \xrightarrow{\partial^{I_2} \cdots \partial^{I_k}} & h \\ {\scriptstyle \partial^{\{1\}}_{[1,b]}}\downarrow & & \downarrow{\scriptstyle \partial^{\{1\}}_{[1,b]}} \\ Q & \xrightarrow[\partial^{I_2} \cdots \partial^{I_k}]{} & \lambda Q_1 \end{array}$$

Then we have

$$\begin{bmatrix} h^{x_1=0} \\ h^{x_1=1} \end{bmatrix} = \frac{\lambda}{b-a} \begin{bmatrix} b & -a \\ -1 & 1 \end{bmatrix} \begin{bmatrix} P_1 \\ Q_1 \end{bmatrix}. \tag{3.5}$$

Note that

$$\begin{bmatrix} P_1 \\ Q_1 \end{bmatrix} = \begin{bmatrix} 0 & \ldots & 0 & P_1^{\alpha_1} & 0 & \ldots & 0 & P_1^{\overline{\alpha_1}} & 0 & \ldots & 0 \\ 0 & \ldots & 0 & Q_1^{\alpha_1} & 0 & \ldots & 0 & Q_1^{\overline{\alpha_1}} & 0 & \ldots & 0 \end{bmatrix}. \tag{3.6}$$

Let

$$\begin{bmatrix} \check{P}_1^{\alpha_1} & \check{P}_1^{\overline{\alpha_1}} \\ \check{Q}_1^{\alpha_1} & \check{Q}_1^{\overline{\alpha_1}} \end{bmatrix} = \frac{\lambda}{b-a} \begin{bmatrix} b & -a \\ -1 & 1 \end{bmatrix} \begin{bmatrix} P_1^{\alpha_1} & P_1^{\overline{\alpha_1}} \\ Q_1^{\alpha_1} & Q_1^{\overline{\alpha_1}} \end{bmatrix}. \tag{3.7}$$

Then

$$\begin{bmatrix} h^{x_1=0} \\ h^{x_1=1} \end{bmatrix} = \begin{bmatrix} 0 & \ldots & 0 & \check{P}_1^{\alpha_1} & 0 & \ldots & 0 & \check{P}_1^{\overline{\alpha_1}} & 0 & \ldots & 0 \\ 0 & \ldots & 0 & \check{Q}_1^{\alpha_1} & 0 & \ldots & 0 & \check{Q}_1^{\overline{\alpha_1}} & 0 & \ldots & 0 \end{bmatrix}. \tag{3.8}$$

Let $\check{h}$ be the binary signature $(\check{h}^{00}, \check{h}^{01}, \check{h}^{10}, \check{h}^{11}) = (\check{P}_1^{\alpha_1}, \check{P}_1^{\overline{\alpha_1}}, \check{Q}_1^{\alpha_1}, \check{Q}_1^{\overline{\alpha_1}})$. If $|I_1| = 1$, then $\check{h}$ is h. If $|I_1| > 1$, by combining all but one variable in I_1 using $[1, 1]$, which is just $(=_1)$ present in Pl-#CSP, we can get $\check{h}$. If $\check{h} \notin \mathscr{P}$, then we are done by Lemma 3.3.

If $\check{h} \in \mathscr{P}$, then $h \in \mathscr{P}$. Since P_1 and Q_1 are linearly independent,

$$\det \begin{bmatrix} \check{P}_1^{\alpha_1} & \check{P}_1^{\overline{\alpha_1}} \\ \check{Q}_1^{\alpha_1} & \check{Q}_1^{\overline{\alpha_1}} \end{bmatrix} \neq 0.$$

Hence either $\check{P}_1^{\alpha_1} = \check{Q}_1^{\overline{\alpha_1}} = 0$ or $\check{P}_1^{\overline{\alpha_1}} = \check{Q}_1^{\alpha_1} = 0$. In either case, compare (3.5) to (3.8) with

$$\begin{bmatrix} f^{x_1=0} \\ f^{x_1=1} \end{bmatrix} = \frac{1}{b-a} \begin{bmatrix} b & -a \\ -1 & 1 \end{bmatrix} \begin{bmatrix} (0 & \dots & 0 & P_1^{\alpha_1} & 0 & \dots & 0 & P_1^{\overline{\alpha_1}} & 0 & \dots & 0) \otimes P_2 \otimes \dots \otimes P_k \\ (0 & \dots & 0 & Q_1^{\alpha_1} & 0 & \dots & 0 & Q_1^{\overline{\alpha_1}} & 0 & \dots & 0) \otimes P_2 \otimes \dots \otimes P_k \end{bmatrix};$$

we have

$$\begin{bmatrix} f^{x_1=0} \\ f^{x_1=1} \end{bmatrix} = \frac{1}{\lambda} \begin{bmatrix} (0 & \dots & 0 & \check{P}_1^{\alpha_1} & 0 & \dots & 0 & \check{P}_1^{\overline{\alpha_1}} & 0 & \dots & 0) \otimes P_2 \otimes \dots \otimes P_k \\ (0 & \dots & 0 & \check{Q}_1^{\alpha_1} & 0 & \dots & 0 & \check{Q}_1^{\overline{\alpha_1}} & 0 & \dots & 0) \otimes P_2 \otimes \dots \otimes P_k \end{bmatrix}.$$

We conclude that $f \in \mathscr{P}$. But this is a contradiction.

- If there are two functions among $\{P, Q, R\}$ that do not have compatible type, without loss of generality, we assume that P and Q do not have compatible type. There exist two partitions $\mathcal{I} = \{I_1, I_2, \dots, I_k\}$ and $\mathcal{J} = \{J_1, J_2, \dots, J_\ell\}$ of $[n] \setminus \{1\}$ and signatures $P_1, P_2, \dots, P_k$ and $Q_1, Q_2, \dots, Q_\ell$ such that

$$P(X) = \prod_{i=1}^{k} P_i(X|_{I_i}) \quad \text{and} \quad Q(X) = \prod_{j=1}^{\ell} Q_j(X|_{J_j})$$

are the primitive decompositions of P and Q, respectively.

If the partitions $\mathcal{I}$ and $\mathcal{J}$ are not compatible, then we are done by Lemma 3.5. So we may assume $\mathcal{I}$ and $\mathcal{J}$ are compatible, and yet P and Q still do not have compatible type. Then, without loss of generality, one of the following holds:

1. There exist $I_i \in \mathcal{I}$, $J_j \in \mathcal{J}$, such that $I_i = J_j$ and $|I_i| \geq 2$ but $\operatorname{supp}(P_i) \neq \operatorname{supp}(Q_j)$.
2. There exist $I_i \in \mathcal{I}$ with $|I_i| \geq 2$ and $J_{j_1}, J_{j_2}, \dots, J_{j_{|I_i|}} \in \mathcal{J}$ with $|J_{j_k}| = 1$ for $1 \leq k \leq |I_i|$, such that $I_i = \bigcup_{k=1}^{|I_i|} J_{j_k}$ but supp $(\prod_{k=1}^{|I_i|} Q_{j_k})$ is not a singleton subset of $\operatorname{supp}(P_i)$.

In case 1, $\operatorname{supp}(P_i) = \{\alpha, \overline{\alpha}\}$ and $\operatorname{supp}(Q_j) = \{\beta, \overline{\beta}\}$ for some $\alpha, \beta \in \{0, 1\}^{|I_i|}$, because both are factors in a primitive decomposition and $|I_i| = |J_j| \geq 2$. Being that both are antipodal pairs, and $\alpha \neq \beta$ and $\alpha \neq \overline{\beta}$, it follows that $\operatorname{supp}(P_i) \cap \operatorname{supp}(Q_j) = \emptyset$. In case 2, $\operatorname{supp}(P_i) = \{\alpha, \overline{\alpha}\}$, and we have supp $(\prod_{k=1}^{|I_i|} Q_{j_k}) \not\subseteq \operatorname{supp}(P_i)$. To see that, if any $Q_{j_k} \neq \lambda[1, 0]$ or $\lambda[0, 1]$ ($\lambda \in \mathbb{C}$), then supp $(\prod_{k=1}^{|I_i|} Q_{j_k})$ is clearly not a subset of any set of the form $\{\alpha, \overline{\alpha}\}$. If all Q_{j_k} are of this form, then supp $(\prod_{k=1}^{|I_i|} Q_{j_k})$ is a singleton set, but not a subset of $\operatorname{supp}(P_i)$. Hence, there exists some $\beta \in \{0, 1\}^{|I_i|}$ such that $\beta \in$ supp $(\prod_{k=1}^{|I_i|} Q_{j_k}) \setminus \operatorname{supp}(P_i)$. Thus we have $\alpha \neq \beta$ and $\alpha \neq \overline{\beta}$ as

well. This is equivalent to the existence of some $s, t \in I_i$, $s \neq t$, such that $\alpha_s = \beta_s$ and $\alpha_t = \overline{\beta_t}$.

Aside from x_1, x_s, and x_t, there exists another variable x_r, since $n \geq 4$.

Suppose $r \in I_i$. In case 1, we take any $u \in \{[1,a],[1,b],[1,c]\}$ with two nonzero entries and connect it to x_r. We get $\partial_u^{\{r\}}(P_i)$ and $\partial_u^{\{r\}}(Q_j)$ with support $\{\alpha', \overline{\alpha'}\}$ and $\{\beta', \overline{\beta'}\}$, where α' and $\overline{\alpha'}$ are obtained from α and $\overline{\alpha}$ by removing the rth bit, and similarly for β' and $\overline{\beta'}$. Since $r \neq s, t$, we still have $\alpha'_s = \beta'_s$ and $\alpha'_t = \overline{\beta'_t}$, and thus $\alpha' \neq \beta'$ and $\alpha' \neq \overline{\beta'}$. Also $|I_i \setminus \{r\}| \geq 2$. Hence $\partial_u^{\{r\}}(P)$ and $\partial_u^{\{r\}}(Q)$ do not have compatible type. The proof for case 2 is similar; we pick $u \in \{[1,a],[1,b],[1,c]\}$ with two nonzero entries as well as satisfying $\partial_u^{\{r\}}(Q_{j_k}) \neq 0$ for that (nonzero) unary signature Q_{j_k}, where $J_{j_k} = \{r\}$.

If $r \notin I_i$, then there are some i' and j' such that $r \in I_{i'}$ and $r \in J_{j'}$. If $|I_{i'}| = 1$, then $P_{i'}$ is a nonzero unary function, and at most one $u \in \{[1,a],[1,b],[1,c]\}$ satisfies $\partial_u^{\{r\}}(P_{i'}) = 0$; if so, we exclude this u. If $|I_{i'}| \geq 2$, then there are two antipodal support points in $\mathrm{supp}(\partial_u^{\{r\}}(P_{i'}))$ for any $u \in \{[1,a],[1,b],[1,c]\}$ with two nonzero entries, which again excludes at most one unary, namely $[1,0]$. Thus in either case we exclude at most one $u \in \{[1,a],[1,b],[1,c]\}$ on account of $I_{i'}$. Similarly we exclude at most one u on account of $J_{j'}$. Pick one $u \in \{[1,a],[1,b],[1,c]\}$ not excluded, and form $P' = \partial_u^{\{r\}}(P)$ and $Q' = \partial_u^{\{r\}}(Q)$. These do not have compatible type.

By connecting u to x_r in f, we get $f' = \partial_u^{\{r\}}(f)$ with arity $n-1$, and

$$\partial_{[1,a]}^{\{1\}}(f') = \partial_{[1,a]}^{\{1\}}(\partial_u^{\{r\}}(f)) = \partial_u^{\{r\}}(\partial_{[1,a]}^{\{1\}}(f)) = \partial_u^{\{r\}}(P) = P',$$

and similarly

$$\partial_{[1,b]}^{\{1\}}(f') = \partial_{[1,b]}^{\{1\}}(\partial_u^{\{r\}}(f)) = \partial_u^{\{r\}}(\partial_{[1,b]}^{\{1\}}(f)) = \partial_u^{\{r\}}(Q) = Q'.$$

$$\begin{array}{ccc} f & \xrightarrow{\partial_u^{\{r\}}} & f' \\ {\scriptstyle \partial_{[1,a]}^{\{1\}}}\downarrow & & \downarrow{\scriptstyle \partial_{[1,a]}^{\{1\}}} \\ P & \xrightarrow[\partial_u^{\{r\}}]{} & P' \end{array} \qquad \begin{array}{ccc} f & \xrightarrow{\partial_u^{\{r\}}} & f' \\ {\scriptstyle \partial_{[1,b]}^{\{1\}}}\downarrow & & \downarrow{\scriptstyle \partial_{[1,b]}^{\{1\}}} \\ Q & \xrightarrow[\partial_u^{\{r\}}]{} & Q' \end{array}$$

Here again the diagrams indicate commutativity of operations.

This implies that $f' \notin \mathscr{P}$ by Lemma 3.4. Thus we are done by induction. ∎

3.2. Arity reduction for nonaffine signatures.

LEMMA 3.7. *Let f be a signature of arity n with affine support of dimension $k < n$, and let $S = \{i_1, i_2, \ldots, i_k\}$ be the indices of a set of k free variables. Let $f' = \partial_{[1,a]}^{[n]\setminus S}(f)$, where $a^4 = 1$; then $f \in \mathscr{A}$ iff $f' \in \mathscr{A}$.*

Proof. Define

$$\widetilde{f}(x_1, x_2, \ldots, x_n) = a^{\sum_{i \in [n]\setminus S} x_i} f(x_1, x_2, \ldots, x_n).$$

Then $f \in \mathscr{A}$ iff $\widetilde{f} \in \mathscr{A}$, as the modifier $a^{\sum_{i \in [n]\setminus S} x_i}$ is a power of $\mathbf{i}$ raised to a linear sum, and the inverse transformation is of the same kind.

Note that $f' = \partial_{[1,1]}^{[n]\setminus S}(\widetilde{f})$. It follows that $\widetilde{f} \in \mathscr{A}$ iff $f' \in \mathscr{A}$ by Corollary 2.14 since f' is just the compressed signature of $\widetilde{f}$ for X. ∎

LEMMA 3.8. *Suppose $S \subseteq \mathbb{Z}_2^n$ is not a linear subspace but for some $i \in [n]$, $S^{x_i=0}$ is a linear subspace of $\mathbb{Z}_2^{n-1}$, where $S^{x_i=0}$ is the subset of S with the ith bit $x_i = 0$. Then there exist $\mathbf{a} = a_1 a_2 \cdots a_n \in S, \mathbf{b} = b_1 b_2 \cdots b_n \in S$, $\mathbf{a} \oplus \mathbf{b} \notin S$, and for some $j \in [n]$, $a_j \neq b_j$.*

Proof. Since $S^{x_i=0}$ is a linear subspace of $\mathbb{Z}_2^{n-1}$, $0^{n-1} \in S^{x_i=0}$; thus $0^n \in S$. Since S is not a linear space, there exist $\mathbf{a} = a_1 a_2 \cdots a_n \in S$, $\mathbf{b} = b_1 b_2 \cdots b_n \in S$, and $\mathbf{c} = \mathbf{a} \oplus \mathbf{b} \notin S$. In particular, $\mathbf{a} \neq \mathbf{b}$ or else $\mathbf{c} = 0^n$. So there exists $j \in [n]$ such that $a_j \neq b_j$. ∎

For the proof of the following lemma, we use the Tableau Calculus. It is basically a schematic tabulation from known assertions about a constraint function to derive additional assertions about the function based on some closure properties. Often it starts off with some assumptions, and after successive "tables" we ultimately find a contradiction, thereby proving the opposite of the initial assumption must hold. The closure properties generally come in two varieties. One variety deals with support sets. Here it is essentially an adaptation of the Mal'tsev polymorphisms used by Dyer and Richerby in [23], with one additional twist. We often manage to shift the underlying affine subspace so that the operations take place on a linear subspace, rather than on an affine subspace. This way we can use just two known vectors in the support set to arrive at a third such vector. This tactical maneuver seems to be very beneficial in a practical sense, without which the proof would have been quite unwieldy. The second variety of closure properties deals with combining function values. Often we end up getting values of a constraint function that are linear combinations of known values of the constraint function. In particular, they can be sums of ± 1 and $\pm \mathrm{i}$, arriving at a value of norm $\sqrt{2}$. However, this is "incompatible" with values of norm 1 or 2 at other points, because to stay within the tractable class $\mathscr{A}$ the nonzero values must have the same norm. As a practical observation in carrying out this proof, this tabulated search has helped us find some tortuous routes to achieve a proof on multiple occasions. And so we call it a Tableau Calculus; the reader will find its appearance many times in this paper.

LEMMA 3.9. *Fix any complex number $x \neq 0$. If $\widehat{\mathcal{F}}$ contains a signature $f \notin \mathscr{A}$, then there exists a unary signature $u \notin \mathscr{A}$, such that*

$$\text{Pl-Holant}(u, \widehat{\mathcal{EQ}}, [0,1], [1,x], \widehat{\mathcal{F}}) \leq_{\mathrm{T}} \text{Pl-Holant}(\widehat{\mathcal{EQ}}, [0,1], [1,x], \widehat{\mathcal{F}}).$$

Proof. Let f have arity $n \geq 1$. If $n = 1$, we can choose $u = f$. If $x^4 \neq 1$, we can choose $u = [1, x] \notin \mathscr{A}$, as $x \neq 0$ by assumption. So we may assume $n \geq 2$ and $x^4 = 1$. We prove the lemma by constructing some signature $g \notin \mathscr{A}$ of arity less than n, such that

$$\text{Pl-Holant}(\widehat{\mathcal{EQ}}, [0,1], [1,x], \widehat{\mathcal{F}}, g) \leq_{\mathrm{T}} \text{Pl-Holant}(\widehat{\mathcal{EQ}}, [0,1], [1,x], \widehat{\mathcal{F}}).$$

By Lemma 2.38, we can assume that $f_{00\ldots0} = 1$. We have $[0,1]$ explicitly given, as well as $[1,0] \in \widehat{\mathcal{EQ}}$. If there exists $i \in [n]$ such that $f^{x_i=0} \notin \mathscr{A}$ or $f^{x_i=1} \notin \mathscr{A}$, then we can choose g to be one of these which has arity $n-1$. So we may assume that both $f^{x_i=0} \in \mathscr{A}$ and $f^{x_i=1} \in \mathscr{A}$ for all $i \in [n]$.

We first prove that if $\mathrm{supp}(f)$ is not an affine subspace, then we can construct some signature $g \notin \mathscr{A}$ of arity less than n in Pl-Holant$(\widehat{\mathcal{EQ}}, [0,1], [1,x], \widehat{\mathcal{F}})$. Suppose $\mathrm{supp}(f)$ is not affine. Note that a subset of $\mathbb{Z}_2^n$ containing $(0,0,\ldots,0)$ is affine iff

it is a linear subspace. Since $(0,0,\ldots,0) \in \operatorname{supp}(f)$, by Lemma 3.8, there exist $\mathbf{a} = a_1a_2\ldots a_n, \mathbf{b} = b_1b_2\ldots b_n \in \operatorname{supp}(f)$, such that $\mathbf{c} = \mathbf{a} \oplus \mathbf{b} = c_1c_2\ldots c_n \notin \operatorname{supp}(f)$, and there exists $i \in [n]$ such that $a_i \neq b_i$. Without loss of generality, we assume that $i = 1$, $a_1 = 0$, $b_1 = 1$, and $c_1 = 1$. We denote $\mathbf{a}' = a_2\ldots a_n, \mathbf{b}' = b_2\ldots b_n$, and $\mathbf{c}' = c_2\ldots c_n$.

$$\begin{array}{rccccc} & \mathbf{a} & = & a_1\mathbf{a}' & = & 0\ a_2\ldots a_n \quad \in \operatorname{supp}(f) \\ \oplus & \mathbf{b} & = & b_1\mathbf{b}' & = & 1\ b_2\ldots b_n \quad \in \operatorname{supp}(f) \\ \hline & \mathbf{c} & = & c_1\mathbf{c}' & = & 1\ c_2\ldots c_n \quad \notin \operatorname{supp}(f) \end{array}$$

By connecting the unary signature $[1, x]$ to the first variable of f we get $h = \partial^{\{1\}}_{[1,x]}(f)$, which has arity $n-1$. If $h \notin \mathscr{A}$, then we are done. Therefore, we may assume $h \in \mathscr{A}$. Note that $h_\alpha = f_{0\alpha} + xf_{1\alpha}$ for all $\alpha \in \{0,1\}^{n-1}$. The next claim will be used several times in the following proof.

Claim. If there exists $\alpha \in \{0,1\}^{n-1}$ such that $h_\alpha = 0$ and $f_{0\alpha} \neq 0$, then we can construct $[1, -x]$.

To prove this claim, we can first obtain the unary signature $[f_{0\alpha}, f_{1\alpha}]$ from f by pinning on all variables $x_2, \ldots, x_n$ according to α, using $[1, 0] \in \widehat{\mathcal{EQ}}$ and the explicitly given $[0, 1]$. We have $x \in \{\pm 1, \pm \mathfrak{i}\}$. If $x = \pm 1$, then $f_{1\alpha} = -xf_{0\alpha}$ from $h_\alpha = 0$, and so $[f_{0\alpha}, f_{1\alpha}] = f_{0\alpha}[1, -x]$. Thus we have $[1, -x]$ up to the nonzero scalar $f_{0\alpha}$. If $x = \pm\mathfrak{i}$, then from $[1, 0, 1, 0] \in \widehat{\mathcal{EQ}}$ we have $\partial_{[0,1]}([1, 0, 1, 0]) = [0, 1, 0]$ and $\partial_{[1,x]}([0, 1, 0]) = [x, 1] = x[1, -x]$.

Once we have $[1, -x]$, we can construct another signature $\widetilde{h} = \partial_{[1,-x]}(f)$ in addition to h. The analysis below will use both h and $\widetilde{h}$:

$$\begin{aligned} h_\alpha &= f_{0\alpha} + xf_{1\alpha}, \\ \widetilde{h}_\alpha &= f_{0\alpha} - xf_{1\alpha}. \end{aligned}$$

In the following we consider various cases according to the membership of $\bar{a}_1\mathbf{a}'$ and $\bar{b}_1\mathbf{b}'$ in $\operatorname{supp}(f)$.

- Suppose $\bar{a}_1\mathbf{a}' \in \operatorname{supp}(f)$ and $\bar{b}_1\mathbf{b}' \in \operatorname{supp}(f)$.
  Note that $\operatorname{supp}(f^{x_1=0})$ is a linear subspace since $f^{x_1=0}$ is affine and $f_{00\ldots0} \neq 0$. By $\mathbf{a} = a_1\mathbf{a}' = 0\mathbf{a}' \in \operatorname{supp}(f)$, we have $\mathbf{a}' \in \operatorname{supp}(f^{x_1=0})$. By $\bar{b}_1\mathbf{b}' = 0\mathbf{b}' \in \operatorname{supp}(f)$, we have $\mathbf{b}' \in \operatorname{supp}(f^{x_1=0})$. By definition, $\mathbf{a}' \oplus \mathbf{b}' = \mathbf{c}'$, and thus $\mathbf{c}' \in \operatorname{supp}(f^{x_1=0})$. This implies that $f_{\bar{c}_1\mathbf{c}'} \neq 0$.
  Since all of $f_{00\ldots0}, f_{a_1\mathbf{a}'}, f_{\bar{b}_1\mathbf{b}'}, f_{\bar{c}_1\mathbf{c}'}$ are nonzero entries of $f^{x_1=0} \in \mathscr{A}$, they are all powers of $\mathfrak{i}$ as $f_{00\ldots0} = 1$. We have $h_{\mathbf{c}'} = xf_{c_1\mathbf{c}'} + f_{\bar{c}_1\mathbf{c}'} = f_{\bar{c}_1\mathbf{c}'}$ since $c_1 = 1$ and $f_{c_1\mathbf{c}'} = 0$. Hence $|h_{\mathbf{c}'}| = 1$ since it is a power of $\mathfrak{i}$. Moreover, since both $f_{a_1\mathbf{a}'}$ and $f_{\bar{a}_1\mathbf{a}'}$ are nonzero entries of $f^{x_2=a_2} \in \mathscr{A}$ and $f_{a_1\mathbf{a}'}$ is a power of $\mathfrak{i}$, so is $f_{\bar{a}_1\mathbf{a}'}$. Any nonzero sum of two quantities that are both powers of $\mathfrak{i}$ must have norm either 2 or $\sqrt{2}$. This implies that if $h_{\mathbf{a}'} = f_{a_1\mathbf{a}'} + xf_{\bar{a}_1\mathbf{a}'}$ is nonzero, then $|h_{\mathbf{a}'}| = 2$ or $\sqrt{2}$. This implies that $|h_{\mathbf{a}'}| \neq |h_{\mathbf{c}'}|$ and both are nonzero. This contradicts that $h \in \mathscr{A}$, by Proposition 2.17.
  Therefore, $h_{\mathbf{a}'} = f_{a_1\mathbf{a}'} + xf_{\bar{a}_1\mathbf{a}'} = 0$. Then we have $[1, -x]$ by the claim, and we obtain $\widetilde{h}$. If $\widetilde{h} \notin \mathscr{A}$, then we are done since the arity of $\widetilde{h}$ is $n-1$. Therefore, we may assume $\widetilde{h} \in \mathscr{A}$. We have $|\widetilde{h}_{\mathbf{c}'}| = |f_{\bar{c}_1\mathbf{c}'} - xf_{c_1\mathbf{c}'}| = 1$ since $c_1 = 1$, $f_{c_1\mathbf{c}'} = 0$, and $f_{\bar{c}_1\mathbf{c}'}$ is a power of $\mathfrak{i}$.
  We already have $h_{\mathbf{a}'} = 0$. If additionally $\widetilde{h}_{\mathbf{a}'} = f_{a_1\mathbf{a}'} - xf_{\bar{a}_1\mathbf{a}'} = 0$, then we

have

$$f_{a_1\mathbf{a}'} + xf_{\bar{a}_1\mathbf{a}'} = 0,$$
$$f_{a_1\mathbf{a}'} - xf_{\bar{a}_1\mathbf{a}'} = 0.$$

This implies that $f_{a_1\mathbf{a}'} = 0$, and it is a contradiction to $\mathbf{a} = a_1\mathbf{a}' \in \mathrm{supp}(f)$. Therefore, $\widetilde{h}_{\mathbf{a}'} = f_{a_1\mathbf{a}'} - xf_{\bar{a}_1\mathbf{a}'} \neq 0$. Since both $f_{a_1\mathbf{a}'}$ and $xf_{\bar{a}_1\mathbf{a}'}$ are powers of $\mathbf{i}$, the norm $|\widetilde{h}_{\mathbf{a}'}|$ is either 2 or $\sqrt{2}$. This implies that $|\widetilde{h}_{\mathbf{a}'}| \neq |\widetilde{h}_{\mathbf{c}'}|$, and both are nonzero. This contradicts that $\widetilde{h} \in \mathscr{A}$, by Proposition 2.17.

- Suppose $\bar{a}_1\mathbf{a}' \notin \mathrm{supp}(f)$ and $\bar{b}_1\mathbf{b}' \notin \mathrm{supp}(f)$.
We have $h_{\mathbf{a}'} = f_{a_1\mathbf{a}'} + xf_{\bar{a}_1\mathbf{a}'} \neq 0$ and $h_{\mathbf{b}'} = f_{\bar{b}_1\mathbf{b}'} + xf_{b_1\mathbf{b}'} \neq 0$ by $f_{\bar{a}_1\mathbf{a}'} = f_{\bar{b}_1\mathbf{b}'} = 0$ and $f_{a_1\mathbf{a}'} \neq 0$, $f_{b_1\mathbf{b}'} \neq 0$, and also $x \neq 0$. We show next that $h_{0\ldots0} = 0$. Suppose for a contradiction that $h_{0\ldots0} \neq 0$. Since $h \in \mathscr{A}$, and $0\ldots0 \in \mathrm{supp}(h)$, $\mathrm{supp}(h)$ is a linear subspace. As $\mathbf{a}', \mathbf{b}' \in \mathrm{supp}(h)$ we have $\mathbf{c}' \in \mathrm{supp}(h)$.

$$\begin{array}{rll} & \mathbf{a}' & \in \mathrm{supp}(h) \\ \oplus & \mathbf{b}' & \in \mathrm{supp}(h) \\ \hline & \mathbf{c}' & \in \mathrm{supp}(h) \end{array}$$

By $h_{\mathbf{c}'} = xf_{c_1\mathbf{c}'} + f_{\bar{c}_1\mathbf{c}'} \neq 0$, we have $f_{\bar{c}_1\mathbf{c}'} \neq 0$ since $f_{c_1\mathbf{c}'} = 0$. Thus $\mathbf{c}' \in \mathrm{supp}(f^{x_1=0})$ as $\bar{c}_1 = 0$. Since $\mathbf{a}' \in \mathrm{supp}(f^{x_1=0})$, and the support of $f^{x_1=0}$ is a linear subspace, we have $\mathbf{b}' \in \mathrm{supp}(f^{x_1=0})$.

$$\begin{array}{rll} & \mathbf{a}' & \in \mathrm{supp}(f^{x_1=0}) \\ \oplus & \mathbf{c}' & \in \mathrm{supp}(f^{x_1=0}) \\ \hline & \mathbf{b}' & \in \mathrm{supp}(f^{x_1=0}) \end{array}$$

This contradicts that $f_{\bar{b}_1\mathbf{b}'} = 0$.

Therefore, $h_{0\ldots0} = f_{00\ldots0} + xf_{10\cdots0} = 0$. Then we can obtain $[1, -x]$ and $\widetilde{h}$ by the claim. If $\widetilde{h} \notin \mathscr{A}$, then we are done. Therefore, we may assume $\widetilde{h} \in \mathscr{A}$. Moreover, $\widetilde{h}_{0\ldots0} = f_{00\ldots0} - xf_{10\cdots0} \neq 0$ by $f_{00\ldots0} + xf_{10\ldots0} = 0$ and $f_{00\ldots0} \neq 0$. Thus $\mathrm{supp}(\widetilde{h})$ is a linear subspace.

Note that $\widetilde{h}_{\mathbf{a}'} = f_{a_1\mathbf{a}'} - xf_{\bar{a}_1\mathbf{a}'} \neq 0$ and $\widetilde{h}_{\mathbf{b}'} = f_{\bar{b}_1\mathbf{b}'} - xf_{b_1\mathbf{b}'} \neq 0$ by $f_{\bar{a}_1\mathbf{a}'} = f_{\bar{b}_1\mathbf{b}'} = 0$ and $f_{a_1\mathbf{a}'} \neq 0$, $f_{b_1\mathbf{b}'} \neq 0$, and $x \neq 0$. Thus $\mathbf{a}' \in \mathrm{supp}(\widetilde{h})$ and $\mathbf{b}' \in \mathrm{supp}(\widetilde{h})$. It follows that $\mathbf{c}' = \mathbf{a}' \oplus \mathbf{b}' \in \mathrm{supp}(\widetilde{h})$. This implies that $\widetilde{h}_{\mathbf{c}'} = f_{\bar{c}_1\mathbf{c}'} - xf_{c_1\mathbf{c}'} \neq 0$. So $f_{\bar{c}_1\mathbf{c}'} \neq 0$ since $f_{c_1\mathbf{c}'} = 0$. We have $\mathbf{a}' \in \mathrm{supp}(f^{x_1=0})$ and $\mathbf{c}' \in \mathrm{supp}(f^{x_1=0})$. Because the support of $f^{x_1=0}$ is a linear subspace, it follows that $\mathbf{b}' \in \mathrm{supp}(f^{x_1=0})$.

$$\begin{array}{rll} & \mathbf{a}' & \in \mathrm{supp}(f^{x_1=0}) \\ \oplus & \mathbf{c}' & \in \mathrm{supp}(f^{x_1=0}) \\ \hline & \mathbf{b}' & \in \mathrm{supp}(f^{x_1=0}) \end{array}$$

This contradicts that $f_{\bar{b}_1\mathbf{b}'} = 0$.

- Suppose $\bar{a}_1\mathbf{a}' \in \mathrm{supp}(f)$ and $\bar{b}_1\mathbf{b}' \notin \mathrm{supp}(f)$.
Note that $h_{\mathbf{b}'} = f_{\bar{b}_1\mathbf{b}'} + xf_{b_1\mathbf{b}'} = xf_{b_1\mathbf{b}'}$, since $f_{\bar{b}_1\mathbf{b}'} = 0$. We have $h_{\mathbf{b}'} \neq 0$ since $b_1\mathbf{b}' \in \mathrm{supp}(f)$, and $x \neq 0$. As $f_{00\ldots0} = 1$, and $f_{a_1\mathbf{a}'} \neq 0$, by $f^{x_1=0} \in \mathscr{A}$, and both $0\ldots0$ and $\mathbf{a}' \in \mathrm{supp}(f^{x_1=0})$, we have that $f_{a_1\mathbf{a}'}$ is a power of $\mathbf{i}$. By hypothesis $f_{\bar{a}_1\mathbf{a}'} \neq 0$. By pinning x_2 to the same value $a_2 \in \{0, 1\}$ in both $f_{a_1\mathbf{a}'} \neq 0$ and $f_{\bar{a}_1\mathbf{a}'} \neq 0$, and by $f^{x_2=a_2} \in \mathscr{A}$, we conclude that the value

$f_{\bar{a}_1\mathbf{a}'}$ is a power of i. As $\bar{a}_1 = b_1 = 1$, by pinning x_1 to 1, and $f^{x_1=1} \in \mathscr{A}$, we conclude that the nonzero value $f_{b_1\mathbf{b}'}$ is a power of $\mathfrak{i}$. By $x^4 = 1$, $h_{\mathbf{b}'} = xf_{b_1\mathbf{b}'}$ is also a power of $\mathfrak{i}$.

To recap, we have that $f_{a_1\mathbf{a}'}$, $f_{\bar{a}_1\mathbf{a}'}$, $f_{b_1\mathbf{b}'}$, and $h_{\mathbf{b}'}$ are all powers of $\mathfrak{i}$. In particular, $|h_{\mathbf{b}'}| = 1$. Moreover, $h_{\mathbf{a}'} = f_{a_1\mathbf{a}'} + xf_{\bar{a}_1\mathbf{a}'}$ is a sum of two quantities that are both powers of $\mathfrak{i}$. If $h_{\mathbf{a}'} \neq 0$, then its norm is 2 or $\sqrt{2}$. This is a contradiction to $h \in \mathscr{A}$ by Proposition 2.17.

Thus $h_{\mathbf{a}'} = 0$. Then we can construct $[1, -x]$ and $\widetilde{h}$ by the claim. If $\widetilde{h} \notin \mathscr{A}$, then we are done. Otherwise, we have

$$h_{\mathbf{a}'} = f_{a_1\mathbf{a}'} + xf_{\bar{a}_1\mathbf{a}'} = 0,$$
$$\widetilde{h}_{\mathbf{a}'} = f_{a_1\mathbf{a}'} - xf_{\bar{a}_1\mathbf{a}'}.$$

If $\widetilde{h}_{\mathbf{a}'} = 0$, then we would have $f_{a_1\mathbf{a}'} = 0$, a contradiction. Hence $\widetilde{h}_{\mathbf{a}'} \neq 0$ and is the sum of two quantities that are both powers of $\mathfrak{i}$. Hence $|\widetilde{h}_{\mathbf{a}'}|$ is 2 or $\sqrt{2}$. Yet, $\widetilde{h}_{\mathbf{b}'} = f_{\bar{b}_1\mathbf{b}'} - xf_{b_1\mathbf{b}'} = -xf_{b_1\mathbf{b}'}$ is a power of $\mathfrak{i}$, as $f_{\bar{b}_1\mathbf{b}'} = 0$ by hypothesis. Thus $|\widetilde{h}_{\mathbf{b}'}| = 1$. This is a contradiction to $\widetilde{h} \in \mathscr{A}$ by Proposition 2.17.

- Suppose $\bar{a}_1\mathbf{a}' \notin \mathrm{supp}(f)$ and $\bar{b}_1\mathbf{b}' \in \mathrm{supp}(f)$.

  Consider $h_{\mathbf{a}'} = f_{a_1\mathbf{a}'} + xf_{\bar{a}_1\mathbf{a}'}$. Since $f_{\bar{a}_1\mathbf{a}'} = 0$ and $\mathbf{a} = a_1\mathbf{a}' \in \mathrm{supp}(f)$, $h_{\mathbf{a}'} = f_{a_1\mathbf{a}'} \neq 0$. As $f_{00\ldots0} = 1$, $a_1 = 0$, and $f_{a_1\mathbf{a}'} \neq 0$, by $f^{x_1=0} \in \mathscr{A}$, $f_{a_1\mathbf{a}'}$ is a power of $\mathfrak{i}$, and so is $h_{\mathbf{a}'}$. In particular, $|h_{\mathbf{a}'}| = 1$.

  Also by hypothesis, $f_{\bar{b}_1\mathbf{b}'} \neq 0$. As $\bar{b}_1 = 0$ and $f_{00\ldots0} = 1$, by pinning x_1 to 0, and $f^{x_1=0} \in \mathscr{A}$, we have that $f_{\bar{b}_1\mathbf{b}'}$ is a power of $\mathfrak{i}$. Then pinning x_2 to b_2 in $f_{\bar{b}_1\mathbf{b}'}$ and the nonzero value $f_{b_1\mathbf{b}'}$, we have that $f_{b_1\mathbf{b}'}$ is also a power of $\mathfrak{i}$.

  We have $h_{\mathbf{b}'} = f_{\bar{b}_1\mathbf{b}'} + xf_{b_1\mathbf{b}'}$, which is a sum of two quantities both a power of $\mathfrak{i}$. If $h_{\mathbf{b}'} \neq 0$, it would have norm 2 or $\sqrt{2}$. As $|h_{\mathbf{a}'}| = 1$, and $h \in \mathscr{A}$, this is a contradiction. Hence $h_{\mathbf{b}'} = 0$.

  Then we can construct $[1, -x]$ and $\widetilde{h}$ by the claim. If $\widetilde{h} \notin \mathscr{A}$, then we are done. Otherwise, we have

$$h_{\mathbf{b}'} = f_{\bar{b}_1\mathbf{b}'} + xf_{b_1\mathbf{b}'} = 0,$$
$$\widetilde{h}_{\mathbf{b}'} = f_{\bar{b}_1\mathbf{b}'} - xf_{b_1\mathbf{b}'}.$$

  If $\widetilde{h}_{\mathbf{b}'} = 0$, then we would have $f_{\bar{b}_1\mathbf{b}'} = 0$, a contradiction. Hence $|\widetilde{h}_{\mathbf{b}'}| = 2$ or $\sqrt{2}$. Yet, $\widetilde{h}_{\mathbf{a}'} = f_{a_1\mathbf{a}'} - xf_{\bar{a}_1\mathbf{a}'} = f_{a_1\mathbf{a}'}$ is a power of $\mathfrak{i}$ and is hence of norm 1. This is a contradiction to $\widetilde{h} \in \mathscr{A}$, by Proposition 2.17.

The above argument is what we call the Tableau Calculus.

Now we can assume that $\mathrm{supp}(f)$ is an affine subspace, indeed a linear subspace since $f_{00\ldots0} = 1$. Suppose it has dimension k. If $k = 0$, then $f \in \mathscr{A}$. This is a contradiction. For $k \geq 1$, let $S = \{x_{i_1}, x_{i_2}, \ldots, x_{i_k}\}$ be a set of free variables defining the linear subspace $\mathrm{supp}(f)$. Let $\check{f}$ be obtained from f by connecting $[1, x]$ to every variable outside S (if there is any), $\check{f} = \partial^{[n]\setminus S}_{[1,x]}(f)$. Note that for every assignment to S, the sum in the expression defining $\check{f}$ has exactly one nonzero entry of f, multiplied by a suitable power of $x \neq 0$. Hence all entries of $\check{f}$ are nonzero. By Lemma 3.7, $\check{f} \notin \mathscr{A}$. If there exists $j \in [k]$ such that $\check{f}^{x_{i_j}=0}$ or $\check{f}^{x_{i_j}=1}$ is not affine, then we get a signature not in $\mathscr{A}$ with arity $k - 1 < n$. This completes the proof. Therefore, we may assume both $\check{f}^{x_{i_j}=0}$ and $\check{f}^{x_{i_j}=1}$ are affine for all $j \in [k]$. In the following we will rename the variables of $\check{f}$ as $x_1, \ldots, x_k$.

We claim that if we have $[1, \mathfrak{i}^r]$, where $r \in \{0,1,2,3\}$, then we can construct $[1, \mathfrak{i}^{-r}]$. If $r = 0$ or 2, then $\mathfrak{i}^r = \mathfrak{i}^{-r}$ and we are done. If $r = 1$ or 3, we have $\partial_{[0,1]}([1,0,1,0]) = [0,1,0]$, where $[1,0,1,0] \in \widehat{\mathcal{EQ}}$ and $[0,1]$ is given in the hypothesis of the lemma. Then we can obtain $\partial_{[1,\mathfrak{i}^r]}([0,1,0]) = [\mathfrak{i}^r, 1] = \mathfrak{i}^r[1, \mathfrak{i}^{-r}]$, a nonzero multiple of $[1, \mathfrak{i}^{-r}]$.

Now we finish the proof by constructing a signature not in $\mathscr{A}$ with arity less than n.

1. If $k = 1$, then $\check{f}$ is the desired signature not in $\mathscr{A}$.
2. If $k = 2$, then all four values $\check{f}_{00}$, $\check{f}_{01}$, $\check{f}_{10}$, and $\check{f}_{11}$ are powers of i. This can be seen by noting that $\check{f}_{00} = 1$, and the signatures $\check{f}^{x_1=0} = [\check{f}_{00}, \check{f}_{01}]$, $\check{f}^{x_2=0} = [\check{f}_{00}, \check{f}_{10}]$, and $\check{f}^{x_1=1} = [\check{f}_{10}, \check{f}_{11}]$ all belong to $\mathscr{A}$. So there exist $r, s, t \in \{0,1,2,3\}$ such that
$$\check{f} = (\check{f}_{00}, \check{f}_{01}, \check{f}_{10}, \check{f}_{11}) = (1, \mathfrak{i}^r, \mathfrak{i}^s, \mathfrak{i}^t).$$
If $r + s \equiv t \mod 2$, then $\check{f} \in \mathscr{A}$ by Lemma 2.15. This is a contradiction. Therefore, we have $\mathfrak{i}^t = \pm\mathfrak{i}^{r+s+1}$. Note that we have $\check{f}^{x_1=0} = [1, \mathfrak{i}^r]$. Thus we can construct $[1, \mathfrak{i}^{-r}]$ by the claim, and we can obtain $\partial^{\{2\}}_{[1,\mathfrak{i}^{-r}]}(\check{f}) = [2, \mathfrak{i}^s(1 \pm \mathfrak{i})]$. However, $|\mathfrak{i}^s(1 \pm \mathfrak{i})| = \sqrt{2}$. Thus $[2, \mathfrak{i}^s(1 \pm \mathfrak{i})] \notin \mathscr{A}$ by Proposition 2.17.
3. If $k = 3$, then there exist $r, s, t \in \{0,1,2,3\}$ and $\epsilon_j \in \{1, -1\}$ for $1 \le j \le 4$ such that
$$M_{x_1,x_2x_3}(\check{f}) = \begin{bmatrix} f^{000} & f^{001} & f^{010} & f^{011} \\ f^{100} & f^{101} & f^{110} & f^{111} \end{bmatrix} = \begin{bmatrix} 1 & \mathfrak{i}^r & \mathfrak{i}^s & \epsilon_1\mathfrak{i}^{r+s} \\ \mathfrak{i}^t & \epsilon_2\mathfrak{i}^{r+t} & \epsilon_3\mathfrak{i}^{s+t} & \epsilon_4\mathfrak{i}^{r+s+t} \end{bmatrix}.$$
This can be seen by observing that all signatures $\check{f}^{x_k=0}$ for $k = 1,2,3$ and $\check{f}^{x_1=1}$ are affine.
If $\epsilon_1\epsilon_2\epsilon_3\epsilon_4 = 1$, then $\check{f} \in \mathscr{A}$ by Lemma 2.16. This is a contradiction. Therefore, $\epsilon_4 = -\epsilon_1\epsilon_2\epsilon_3$.
Since we have $\partial^{\{2,3\}}_{[1,0]} = [f^{000}, f^{100}] = [1, \mathfrak{i}^t]$ and $\partial^{\{1,2\}}_{[1,0]}(\check{f}) = [f^{000}, f^{001}] = [1, \mathfrak{i}^r]$, by the claim we also have $[1, \mathfrak{i}^{-t}]$ and $[1, \mathfrak{i}^{-r}]$.
We have
$$\partial^{\{1\}}_{[1,\mathfrak{i}^{-t}]}(\check{f}) = (2, \quad (1+\epsilon_2)\mathfrak{i}^r, \quad (1+\epsilon_3)\mathfrak{i}^s, \quad \epsilon_1(1-\epsilon_2\epsilon_3)\mathfrak{i}^{r+s}).$$

- If $\epsilon_2 = -\epsilon_3$ or $\epsilon_2 = \epsilon_3 = 1$, then $\partial^{\{1\}}_{[1,\mathfrak{i}^{-t}]}(\check{f})$ is not affine since its support is not affine. Thus we are done. So we may assume in the following that $\epsilon_2 = \epsilon_3 = -1$ and
$$M_{x_1,x_2x_3}(\check{f}) = \begin{bmatrix} f^{000} & f^{001} & f^{010} & f^{011} \\ f^{100} & f^{101} & f^{110} & f^{111} \end{bmatrix} = \begin{bmatrix} 1 & \mathfrak{i}^r & \mathfrak{i}^s & \epsilon_1\mathfrak{i}^{r+s} \\ \mathfrak{i}^t & -\mathfrak{i}^{r+t} & -\mathfrak{i}^{s+t} & -\epsilon_1\mathfrak{i}^{r+s+t} \end{bmatrix}.$$
- If $\epsilon_1 = 1$ and $\epsilon_2 = \epsilon_3 = -1$, then
$$\partial^{\{3\}}_{[1,\mathfrak{i}^{-r}]}(\check{f}) = (2, \quad 2\mathfrak{i}^s, \quad 0, \quad -2\mathfrak{i}^{r+s})$$
is not affine since its support is not affine. Thus we are done. So we may assume in the following that $\epsilon_1 = \epsilon_2 = \epsilon_3 = -1$:
$$M_{x_1,x_2x_3}(\check{f}) = \begin{bmatrix} f^{000} & f^{001} & f^{010} & f^{011} \\ f^{100} & f^{101} & f^{110} & f^{111} \end{bmatrix} = \begin{bmatrix} 1 & \mathfrak{i}^r & \mathfrak{i}^s & -\mathfrak{i}^{r+s} \\ \mathfrak{i}^t & -\mathfrak{i}^{r+t} & -\mathfrak{i}^{s+t} & \mathfrak{i}^{r+s+t} \end{bmatrix}.$$

- For $\epsilon_1 = \epsilon_2 = \epsilon_3 = -1$, we take two copies of $\check{f}$ and connect the variables x_2 and x_3 of one copy with the variables x_3 and x_2 of the other copy, creating a planar binary gadget with a symmetric signature $g(y, z) = \sum_{x_2,x_3\in\{0,1\}} \check{f}(y, x_2, x_3)\check{f}(z, x_3, x_2)$. Notice the reversal of the order of x_2 and x_3 in the second copy of $\check{f}$; this amounts to a cyclic permutation of its inputs and is necessary in order to make a planar connection in the gadget construction. The signature of g can be computed as a matrix product:

$$\begin{bmatrix} f^{000} & f^{001} & f^{010} & f^{011} \\ f^{100} & f^{101} & f^{110} & f^{111} \end{bmatrix} \begin{bmatrix} f^{000} & f^{100} \\ f^{010} & f^{110} \\ f^{001} & f^{101} \\ f^{011} & f^{111} \end{bmatrix}.$$

In symmetric signature notation

$$g = [1{+}2\mathrm{i}^{r+s}{+}(-1)^{r+s},\ \ \mathrm{i}^t(1{-}2\mathrm{i}^{r+s}{-}(-1)^{r+s}),\ \ (-1)^t(1{+}2\mathrm{i}^{r+s}{+}(-1)^{r+s})].$$

Further, we have the unary signature $\partial_{[1,0]}(g) = [1{+}2\mathrm{i}^{r+s}{+}(-1)^{r+s},\ \mathrm{i}^t(1{-}2\mathrm{i}^{r+s} - (-1)^{r+s})]$.
If $r + s$ is odd, then the norm $|1 + 2\mathrm{i}^{r+s} + (-1)^{r+s}| = 2$ and the norm $|\mathrm{i}^t(1 - 2\mathrm{i}^{r+s} - (-1)^{r+s})| = |2 \pm 2\mathrm{i}| = 2\sqrt{2}$, and hence $\partial_{[1,0]}(g) \notin \mathscr{A}$, by Proposition 2.17, and we are done. For even $r + s = 2k$, the norms of the entries of $\partial_{[1,0]}(g)$ are $2 + 2(-1)^k$ and 2, respectively. Hence if k is even, then $\partial_{[1,0]}(g) \notin \mathscr{A}$, by Proposition 2.17, and we are done. Hence we may assume k is odd, and

$$r + s \equiv 2 \bmod 4. \tag{3.9}$$

By symmetry of argument, we have

$$s + t \equiv 2 \bmod 4 \tag{3.10}$$

and

$$t + r \equiv 2 \bmod 4. \tag{3.11}$$

From (3.9) and (3.10) we get $r \equiv t \bmod 4$, and by symmetry,

$$r \equiv s \equiv t \bmod 4. \tag{3.12}$$

Also by (3.12) and (3.9) we have

$$r \equiv s \equiv t \equiv 1 \bmod 4 \quad \text{or} \quad r \equiv s \equiv t \equiv 3 \bmod 4. \tag{3.13}$$

If $r \equiv s \equiv t \equiv 1 \bmod 4$, then

$$M_{x_1,x_2x_3}(\check{f}) = \begin{bmatrix} f^{000} & f^{001} & f^{010} & f^{011} \\ f^{100} & f^{101} & f^{110} & f^{111} \end{bmatrix} = \begin{bmatrix} 1 & \mathrm{i} & \mathrm{i} & 1 \\ \mathrm{i} & 1 & 1 & -\mathrm{i} \end{bmatrix}.$$

This is the symmetric ternary signature $[1, \mathrm{i}, 1, -\mathrm{i}] \notin \mathscr{A}$. Having $[1, 0] \in \widehat{\mathcal{EQ}}$, we can get $[1, \mathrm{i}]$ from $[1, \mathrm{i}, 1, -\mathrm{i}]$. Then $\partial_{[1,\mathrm{i}]}([1, \mathrm{i}, 1, -\mathrm{i}]) = [0, 2\mathrm{i}, 2]$. Once again $\partial_{[1,\mathrm{i}]}([0, 2\mathrm{i}, 2]) = [-2, 4\mathrm{i}] \notin \mathscr{A}$.

Similarly, if $r \equiv s \equiv t \equiv 3 \bmod 4$, then

$$M_{x_1,x_2x_3}(\check{f}) = \begin{bmatrix} f^{000} & f^{001} & f^{010} & f^{011} \\ f^{100} & f^{101} & f^{110} & f^{111} \end{bmatrix} = \begin{bmatrix} 1 & -\mathrm{i} & -\mathrm{i} & 1 \\ -\mathrm{i} & 1 & 1 & \mathrm{i} \end{bmatrix}.$$

This is the symmetric ternary signature $[1, -\mathrm{i}, 1, \mathrm{i}] \notin \mathscr{A}$, and we can get $[-2, -4\mathrm{i}] \notin \mathscr{A}$ in a similar way.

4. If $k \geq 4$, then $\check{f}$ is affine by Lemma 2.18. This is a contradiction.

This completes the proof of Lemma 3.9. ∎

The next lemma says that generally we can construct a unary signature $[1, a]$, with $a \neq 0, 1$, in Pl-#CSP$([1, 0], \mathcal{F})$. The condition on $\mathcal{F}$ is satisfied as long as not every signature in $\mathcal{F}$ is $\{0, 1\}$-valued up to a constant.

LEMMA 3.10. *Suppose $\mathcal{F}$ contains a signature f of arity $n \geq 1$ that has two distinct nonzero values: $f_\alpha \neq 0, f_\beta \neq 0$, and $f_\alpha \neq f_\beta$ for some $\alpha, \beta \in \{0, 1\}^n$. Then there is a unary signature $[1, a]$, where $a \neq 0, 1$, such that*

$$\text{Pl-\#CSP}([1, 0], [1, a], \mathcal{F}) \leq_{\mathrm{T}} \text{Pl-\#CSP}([1, 0], \mathcal{F}).$$

The statement is also valid if we replace $[1, 0]$ by $[0, 1]$.

Proof. We prove the lemma for Pl-#CSP$([1, 0], \mathcal{F})$. The proof for Pl-#CSP$([0, 1], \mathcal{F})$ is symmetric.

Since $f_\alpha \neq f_\beta$, at least one value of f_α or f_β is not equal to $f_{00\cdots0}$. Without loss of generality, we assume that $f_\alpha \neq f_{00\cdots0}$. Then we have $f' = \partial^S_{[1,0]}(f) = (f_{00\cdots0}, \ldots, f_\alpha)$, which is a signature of arity $\mathrm{wt}(\alpha)$, where $S = \{k \mid \text{the } k\text{th bit of } \alpha \text{ is } 0\}$. We only care about the value of f' at $0^{\mathrm{wt}(\alpha)}$ and $1^{\mathrm{wt}(\alpha)}$, as specified. By connecting f' to an EQUALITY $(=_{\mathrm{wt}(\alpha)+1})$ of arity $\mathrm{wt}(\alpha) + 1$ in a planar fashion, we get the unary signature $\partial_{f'}(=_{\mathrm{wt}(\alpha)+1}) = [f_{00\cdots0}, f_\alpha]$. If $f_{00\cdots0} \neq 0$, then we have $f_{00\cdots0}[1, a]$, where $a = \frac{f_\alpha}{f_{00\cdots0}}$. Then we get $[1, a]$ up to the nonzero scalar $f_{00\cdots0}$. This finishes the proof.

Otherwise, $f_{00\cdots0} = 0$. This implies that $[f_{00\cdots0}, f_\alpha] = f_\alpha[0, 1]$. Then we have $[0, 1]$ up to the nonzero scalar f_α, and can now use both pinning signatures $[1, 0]$ and $[0, 1]$.

Since the set $T = \{(\xi, \eta) \mid f_\xi \neq 0, f_\eta \neq 0, f_\xi \neq f_\eta\}$ is nonempty by $(\alpha, \beta) \in T$, there exists some $(\xi, \eta) \in T$ with minimum Hamming distance, i.e., $f_\xi \neq 0, f_\eta \neq 0$, $f_\xi \neq f_\eta$, and

$$\mathrm{wt}(\xi \oplus \eta) = \min_{\xi', \eta' \in \{0,1\}^n} \{\mathrm{wt}(\xi' \oplus \eta') \mid f_{\xi'} \neq 0, f_{\eta'} \neq 0, f_{\xi'} \neq f_{\eta'}\}.$$

For $b \in \{0, 1\}$, let $S_b = \{k \mid \text{the } k\text{th bits of both } \xi \text{ and } \eta \text{ are } b\}$. Then we can construct $f'' = \partial^{S_0}_{[1,0]}[\partial^{S_1}_{[0,1]}(f)]$.

Denote $\mathrm{wt}(\xi \oplus \eta)$ by d. Note that f'' has arity d. Let $\check{\xi} \in \{0, 1\}^d$ denote the d-bit string obtained from ξ by deleting all bits in $S_0 \cup S_1$. Similarly define $\check{\eta} \in \{0, 1\}^d$. Clearly $f''_{\check{\xi}} = f_\xi$ and $f''_{\check{\eta}} = f_\eta$. If $d = 1$, i.e., $f'' = [f_\xi, f_\eta]$ or $f'' = [f_\eta, f_\xi]$, then we are done by normalizing. Therefore, we may assume that $d \geq 2$.

All entries of f'' are zero except for $f''_{\check{\xi}}, f''_{\check{\eta}}$. To see this, if there is another nonzero entry of f'', it has the form $f''_{\check{\gamma}} = f_\gamma$ for some $\gamma \in \{0, 1\}^n$ and $\check{\gamma} \in \{0, 1\}^d$, where γ has the same bits as ξ on $S_0 \cup S_1$, and $\check{\gamma}$ is obtained from γ by deleting all bits in $S_0 \cup S_1$. Then $f''_{\check{\gamma}} \neq 0$ implies that $f_\gamma \neq 0$. Both $\mathrm{wt}(\xi \oplus \gamma) < \mathrm{wt}(\xi \oplus \eta)$ and $\mathrm{wt}(\gamma \oplus \eta) < \mathrm{wt}(\xi \oplus \eta)$, and either $f_\gamma \neq f_\xi$ or $f_\gamma \neq f_\eta$. This contradicts the minimality of d.

By using $[1,1]$, we have $\partial_{[1,1]}^{\{2,\dots,d\}}(f'')$ that is $[f_\xi, f_\eta]$ or $[f_\eta, f_\xi]$, since all entries of f'' are zero except for $f''_\xi = f_\xi$ and $f''_{\breve{\eta}} = f_\eta$. Thus we are done by normalizing $\partial_{[1,1]}^{\{2,\dots,d\}}(f'')$. ∎

The next lemma handles $\{0,1\}$-valued function sets. More generally, it applies to any $\mathcal{F}$ where every $f \in \mathcal{F}$ has at most one nonzero output value, since factoring out a nonzero constant factor does not change the problem complexity. Notice that in this $\{0,1\}$-valued case, any function $f \in \mathcal{F} \cap \mathscr{P}$ is also in $\mathscr{A}$. (And since $\widehat{\mathcal{F}}$ does not satisfy the Parity Condition, $\mathcal{F} \subseteq \widehat{\mathscr{M}}$ is not feasible. Therefore, the statement of Lemma 3.11 is in accordance with Theorem 6.1.) When Lemma 3.11 is used in Theorem 3.12, the auxiliary unary signature $[1,0]$ or $[0,1]$ will be provided by $[1,\omega]$ (for some $\omega \neq 0$) from the Pl-Holant side, as constructed in Lemma 3.1. Note that $[1,\pm 1]$ is transformed to $[1,0]$ or $[0,1]$ by the holographic transformation H_2. The proof of the following lemma will again use the Tableau Calculus.

LEMMA 3.11. *If each $f \in \mathcal{F}$ takes values in $\{0,1\}$, then either* Pl-#CSP$(\mathcal{F}, [1,0])$ *is #P-hard or $\mathcal{F} \subseteq \mathscr{A}$. The statement is also true if we replace $[1,0]$ by $[0,1]$.*

Proof. Suppose $\mathcal{F} \nsubseteq \mathscr{A}$. We show that Pl-#CSP$(\mathcal{F}, [1,0])$ is #P-hard. The statement for $[0,1]$ is symmetric.

As $\mathcal{F} \nsubseteq \mathscr{A}$, there exists $f \in \mathcal{F}$ such that $f \notin \mathscr{A}$. First, we claim that $f \notin \mathscr{P}$. By definition, $\mathscr{P} = \langle \mathcal{E} \rangle$. Note that all signatures in $\mathcal{E}$ have affine support. Thus all signatures in $\mathscr{P}$ have affine support. On the other hand, a $\{0,1\}$-valued signature is in $\mathscr{A}$ iff its support is affine. So $\mathrm{supp}(f)$ is not affine since f takes values in $\{0,1\}$ and $f \notin \mathscr{A}$. This implies that $f \notin \mathscr{P}$.

We prove the lemma by induction on the arity n of f. Note that $n \geq 2$ since $\mathrm{supp}(f)$ is not affine.

For $n = 2$, there is exactly one entry of f that is 0 since $\mathrm{supp}(f)$ is not affine. So $f = (1,1,1,0)$, or $f = (1,1,0,1)$, or $f = (1,0,1,1)$, or $f = (0,1,1,1)$. In each case, we take three copies of f and connect the first input of each f to an edge of $=_3$ and leave the second input as dangling edges. The resulting signature g is either $[1,1]^{\otimes 3} + [1,0]^{\otimes 3}$ or $[1,1]^{\otimes 3} + [0,1]^{\otimes 3}$. Both of these signatures are symmetric but not in $\mathscr{P} \cup \mathscr{A} \cup \widehat{\mathscr{M}}$. By Theorem 2.34, Pl-#CSP(g) is #P-hard.

To see that $g = [1,1]^{\otimes 3} + [1,0]^{\otimes 3} \notin \mathscr{P} \cup \mathscr{A} \cup \widehat{\mathscr{M}}$, note the following:

- The fact $g \notin \mathscr{P}$ follows from Proposition 2.20 for symmetric signatures in $\mathscr{P}$.
- Also $g \notin \mathscr{A}$ follows from the fact that the norms of the nonzero entries of $[1,1]^{\otimes 3} + [1,0]^{\otimes 3}$ are different, violating Definition 2.10.
- Since $\widehat{g} = gH_2^{\otimes 3} = \frac{1}{2\sqrt{2}}([2,0]^{\otimes 3} + [1,1]^{\otimes 3})$ does not satisfy the Parity Condition, it follows that $\widehat{g} \notin \mathscr{M}$, and so $g \notin \widehat{\mathscr{M}}$.

Similarly, we have $[1,1]^{\otimes 3} + [0,1]^{\otimes 3} \notin \mathscr{P} \cup \mathscr{A} \cup \widehat{\mathscr{M}}$. By the reduction

$$\text{Pl-\#CSP}(g) \leq_{\mathrm{T}} \text{Pl-\#CSP}([1,0], \mathcal{F}),$$

we conclude that Pl-#CSP$([1,0], \mathcal{F})$ is #P-hard.

In the following, by induction we assume that the lemma is true for $n-1$, where $n \geq 3$.

We will prove that we can construct a signature $f' \notin \mathscr{A}$ with arity $< n$ by a gadget construction in Pl-#CSP$([1,0], \mathcal{F})$. If f' takes values in $\{0,1\}$ up to a scalar, then the induction is finished. Otherwise, there exist two distinct nonzero entries f'_α and f'_β of f'. By Lemma 3.10, we can construct some $[1,a]$ with $a \neq 0, 1$ from f' by gadget construction. Because signatures in $\mathcal{F}$ take values in $\{0,1\}$, all nonzero

entries of the signature of any gadget construction in Pl-#CSP$([1,0],\mathcal{F})$ take positive rational values, in $\mathbb{Q}^+$, after normalization. Since $a \neq 1$ and is positive, we have $|a| \neq 1$. So we can get the unary signatures $[1,2],[1,3],[1,4]$ (in fact, we can get any constant number of unary signatures) using interpolation, by Lemma 2.39. Then by Theorem 3.6 and $f \notin \mathscr{P}$, we can get a symmetric signature f'' that is not in $\mathscr{P}$. Note that the symmetric signature set $\{[1,2], f''\}$ satisfies

$$\{[1,2], f''\} \nsubseteq \mathscr{P}, \quad \{[1,2], f''\} \nsubseteq \mathscr{A}, \quad \{[1,2], f''\} \nsubseteq \widehat{\mathscr{M}}.$$

By Theorem 2.34, Pl-#CSP$([1,2], f'')$ is #P-hard. By

$$\text{Pl-\#CSP}([1,2], f'') \leq_{\mathrm{T}} \text{Pl-\#CSP}([1,0], \mathcal{F}),$$

Pl-#CSP$([1,0],\mathcal{F})$ is #P-hard.

Thus we only need to construct a signature $f' \notin \mathscr{A}$ with arity $< n$ by gadget construction to finish the proof. If there exists $i \in [n]$ such that $f^{x_i=0} \notin \mathscr{A}$, then we are done since we have $[1,0]$. Therefore, we may assume $f^{x_i=0} \in \mathscr{A}$ for all $i \in [n]$. By connecting the unary signature $[1,1]$ to the ith variable of f, we get a signature $h = \partial^{\{i\}}_{[1,1]}(f)$, which has arity $n-1$. Note that

$$h(\alpha) = f^{x_i=0}(\alpha) + f^{x_i=1}(\alpha)$$

for all $\alpha \in \{0,1\}^{n-1}$. If $h \notin \mathscr{A}$, then we are done. Therefore, we may assume $h \in \mathscr{A}$ for every $i \in [n]$.

- Suppose $f_{00\cdots0} = 1$.
  Since $\mathrm{supp}(f)$ is not affine, it is also not a linear subspace. Thus, there exist $\mathbf{a} = a_1 a_2 \cdots a_n \in \mathrm{supp}(f)$ and $\mathbf{b} = b_1 b_2 \cdots b_n \in \mathrm{supp}(f)$ such that $\mathbf{c} = \mathbf{a} \oplus \mathbf{b} \notin \mathrm{supp}(f)$. Here $\mathbf{a} \neq \mathbf{b}$, since $\mathbf{a} \oplus \mathbf{a} = 0^n \in \mathrm{supp}(f)$. Thus for some $i \in [n]$, $a_i \neq b_i$. So, by rotating the variables in a cyclic fashion (thus maintaining planarity), and switching $\mathbf{a}$ and $\mathbf{b}$, we may without loss of generality assume that $a_1 = 0, b_1 = 1$. Then we choose our $h = \partial^{\{1\}}_{[1,1]}(f)$, and

  $$h_\alpha = f_{0\alpha} + f_{1\alpha}$$

  for all $\alpha \in \{0,1\}^{n-1}$.

$$\begin{array}{rcccccl} & \mathbf{a} & = & a_1\mathbf{a}' & = & 0\ a_2 \ldots a_n & \in \mathrm{supp}(f) \\ \oplus & \mathbf{b} & = & b_1\mathbf{b}' & = & 1\ b_2 \ldots b_n & \in \mathrm{supp}(f) \\ \hline & \mathbf{c} & = & c_1\mathbf{c}' & = & 1\ c_2 \ldots c_n & \notin \mathrm{supp}(f) \end{array}$$

  We have

  $$h_{\mathbf{a}'} = f_{0\mathbf{a}'} + f_{1\mathbf{a}'} \neq 0,$$
  $$h_{\mathbf{b}'} = f_{0\mathbf{b}'} + f_{1\mathbf{b}'} \neq 0$$

  since $f_{0\mathbf{a}'} = f_{1\mathbf{b}'} = 1$ and $f_{1\mathbf{a}'} \geq 0, f_{0\mathbf{b}'} \geq 0$. Note that $h_{00\cdots0} = f_{00\cdots0} + f_{10\cdots0} \neq 0$ since $f_{00\cdots0} = 1, f_{10\cdots0} \geq 0$. So $\mathrm{supp}(h)$ is a linear space since h is affine. As $\mathbf{a}', \mathbf{b}' \in \mathrm{supp}(h)$ we have $\mathbf{c}' \in \mathrm{supp}(h)$.

$$\begin{array}{rcl} & \mathbf{a}' & \in \mathrm{supp}(h) \\ \oplus & \mathbf{b}' & \in \mathrm{supp}(h) \\ \hline & \mathbf{c}' & \in \mathrm{supp}(h) \end{array}$$

This implies that $h_{\mathbf{c}'} = f_{c_1\mathbf{c}'} + f_{\bar{c}_1\mathbf{c}'} \neq 0$. So we have $f_{\bar{c}_1\mathbf{c}'} \neq 0$ since $f_{c_1\mathbf{c}'} = 0$. Thus $\mathbf{c}' \in \mathrm{supp}(f^{x_1=0})$, as $\bar{c}_1 = 0$. As f takes values in $\{0,1\}$, $f_{\bar{c}_1\mathbf{c}'} = 1$. Since $\mathbf{a}' \in \mathrm{supp}(f^{x_1=0})$, and the support of $f^{x_1=0}$ is a linear subspace, we have $\mathbf{b}' \in \mathrm{supp}(f^{x_1=0})$, since $\mathbf{b}' = \mathbf{a}' \oplus \mathbf{c}'$.

$$\begin{array}{rll} & \mathbf{a}' & \in \mathrm{supp}(f^{x_1=0}) \\ \oplus & \mathbf{c}' & \in \mathrm{supp}(f^{x_1=0}) \\ \hline & \mathbf{b}' & \in \mathrm{supp}(f^{x_1=0}) \end{array}$$

This implies that $\bar{b}_1\mathbf{b}' \in \mathrm{supp}(f)$. Thus we have $f_{\bar{b}_1\mathbf{b}'} = 1$, as f takes values in $\{0,1\}$. But we also have $f_{b_1\mathbf{b}'} = 1$; thus $h_{\mathbf{b}'} = f_{b_1\mathbf{b}'} + f_{\bar{b}_1\mathbf{b}'} = 2$. However, $h_{\mathbf{c}'} = f_{c_1\mathbf{c}'} + f_{\bar{c}_1\mathbf{c}'} = 1$ since $f_{c_1\mathbf{c}'} = 0$ and $f_{\bar{c}_1\mathbf{c}'} = 1$. Thus $h \notin \mathscr{A}$. This is a contradiction.

- Suppose $f_{00\cdots0} = 0$.
  Since f is not identically 0, there exists $\beta \in \mathrm{supp}(f)$ with $\mathrm{wt}(\beta)$ minimum among all nonzero entries. Then we have $\partial^S_{[1,0]}(f)$, where $S = \{k \mid \text{the } k\text{th bit of } \beta \text{ is } 0\}$. $\partial^S_{[1,0]}(f)$ is the symmetric signature $[0,\ldots,0,1]$ of arity $\mathrm{wt}(\beta)$. This gives $[0,1] = \partial^{\{2,\ldots,\mathrm{wt}(\beta)\}}_{[1,1]}([0,\cdots,0,1])$. If there exists i such that $f^{x_i=1} \notin \mathscr{A}$, then we are done since we have $[0,1]$ now. Therefore, we may assume $f^{x_i=1} \in \mathscr{A}$ as well as $f^{x_i=0} \in \mathscr{A}$ for all $i \in [n]$, since we have $[1,0]$ explicitly.
  Since $\mathrm{supp}(f)$ is not affine, there exist $\mathbf{a},\mathbf{b},\mathbf{c} \in \mathrm{supp}(f)$ such that $\mathbf{a}\oplus\mathbf{b}\oplus\mathbf{c} \notin \mathrm{supp}(f)$. Let $\mathbf{a} = a_1a_2\cdots a_n$, $\mathbf{b} = b_1b_2\cdots b_n$, $\mathbf{c} = c_1c_2\cdots c_n$, and $\mathbf{d} = \mathbf{a}\oplus\mathbf{b}\oplus\mathbf{c} = d_1d_2\cdots d_n$, and we denote $\mathbf{a}' = a_2\cdots a_n$, $\mathbf{b}' = b_2\cdots b_n$, $\mathbf{c}' = c_2\cdots c_n$, and $\mathbf{d}' = d_2\cdots d_n$.

$$\begin{array}{rllllll} & \mathbf{a} & = & a_1\mathbf{a}' & = & a_1a_2\ldots a_n & \in \mathrm{supp}(f) \\ & \mathbf{b} & = & b_1\mathbf{b}' & = & b_1b_2\ldots b_n & \in \mathrm{supp}(f) \\ \oplus & \mathbf{c} & = & c_1\mathbf{c}' & = & c_1c_2\ldots c_n & \in \mathrm{supp}(f) \\ \hline & \mathbf{d} & = & d_1\mathbf{d}' & = & d_1d_2\ldots d_n & \notin \mathrm{supp}(f) \end{array}$$

If $a_1 = b_1 = c_1$, then it follows that $a_1 = b_1 = c_1 = d_1$. This implies that $\mathrm{supp}(f^{x_1=a_1})$ is not an affine subspace of $\mathbb{Z}_2^{n-1}$. This contradicts that $f^{x_1=a_1} \in \mathscr{A}$. Hence, without loss of generality, we can assume that $a_1 = b_1 = \bar{c}_1$, and it follows that $a_1 = b_1 = \bar{c}_1 = \bar{d}_1$.
Now we choose our $h = \partial^{\{1\}}_{[1,1]}(f)$, and

$$h_\alpha = f_{0\alpha} + f_{0\alpha}$$

for all $\alpha \in \{0,1\}^{n-1}$.
We have

$$h_{\mathbf{a}'} = f_{a_1\mathbf{a}'} + f_{\bar{a}_1\mathbf{a}'} \neq 0,$$
$$h_{\mathbf{b}'} = f_{b_1\mathbf{b}'} + f_{\bar{b}_1\mathbf{b}'} \neq 0,$$
$$h_{\mathbf{c}'} = f_{c_1\mathbf{c}'} + f_{\bar{c}_1\mathbf{c}'} \neq 0,$$

since $f_{a_1\mathbf{a}'} = f_{b_1\mathbf{b}'} = f_{c_1\mathbf{c}'} = 1$ and $f_{\bar{a}_1\mathbf{a}'} \geq 0, f_{\bar{b}_1\mathbf{b}'} \geq 0, f_{\bar{c}_1\mathbf{c}'} \geq 0$. Since $\mathrm{supp}(h)$ is affine and $\mathbf{a}',\mathbf{b}',\mathbf{c}' \in \mathrm{supp}(h)$, we have $\mathbf{d}' \in \mathrm{supp}(h)$, as $\mathbf{d}' = \mathbf{a}'\oplus\mathbf{b}'\oplus\mathbf{c}'$.

$$\begin{array}{rll} & \mathbf{a}' & \in \mathrm{supp}(h) \\ & \mathbf{b}' & \in \mathrm{supp}(h) \\ \oplus & \mathbf{c}' & \in \mathrm{supp}(h) \\ \hline & \mathbf{d}' & \in \mathrm{supp}(h) \end{array}$$

By $h_{\mathbf{d}'} = f_{d_1\mathbf{d}'} + f_{\bar{d}_1\mathbf{d}'} \neq 0$, we have $f_{\bar{d}_1\mathbf{d}'} \neq 0$ since $f_{d_1\mathbf{d}'} = 0$. Thus $\mathbf{d}' \in \mathrm{supp}(f^{x_1=\bar{d}_1})$. As f takes values in $\{0,1\}$, $f_{\bar{d}_1\mathbf{d}'} = 1$. Recall that $a_1 = b_1 = \bar{c}_1 = \bar{d}_1$; we have $\mathbf{d}' \in \mathrm{supp}(f^{x_1=\bar{c}_1})$, and also $\mathbf{a}', \mathbf{b}' \in \mathrm{supp}(f^{x_1=\bar{c}_1})$. The support of $f^{x_1=\bar{c}_1}$ is an affine subspace, and so we have $\mathbf{c}' \in \mathrm{supp}(f^{x_1=\bar{c}_1})$, as $\mathbf{c}' = \mathbf{a}' \oplus \mathbf{b}' \oplus \mathbf{d}'$.

$$\begin{array}{cll} & \mathbf{a}' & \in \mathrm{supp}(f^{x_1=\bar{c}_1}) \\ & \mathbf{b}' & \in \mathrm{supp}(f^{x_1=\bar{c}_1}) \\ \oplus & \mathbf{d}' & \in \mathrm{supp}(f^{x_1=\bar{c}_1}) \\ \hline & \mathbf{c}' & \in \mathrm{supp}(f^{x_1=\bar{c}_1}) \end{array}$$

Thus $\bar{c}_1\mathbf{c}' \in \mathrm{supp}(f)$. So $h_{\mathbf{c}'} = f_{c_1\mathbf{c}'} + f_{\bar{c}_1\mathbf{c}'} = 2$. But $h_{\mathbf{d}'} = f_{d_1\mathbf{d}'} + f_{\bar{d}_1\mathbf{d}'} = 1$ since $f_{d_1\mathbf{d}'} = 0$ and $f_{\bar{d}_1\mathbf{d}'} = 1$. Thus $h \notin \mathscr{A}$. This is a contradiction. ∎

3.3. Dichotomy when $\widehat{\mathcal{F}}$ does not satisfy parity.

THEOREM 3.12. *If the signature set $\widehat{\mathcal{F}}$ contains a signature that does not satisfy the parity condition, then either* Pl-Holant$(\widehat{\mathcal{EQ}}, \widehat{\mathcal{F}})$ *is #P-hard, or $\widehat{\mathcal{F}} \subseteq \widehat{\mathscr{P}}$, or $\widehat{\mathcal{F}} \subseteq \mathscr{A}$, in which case the problem is tractable.*

Proof. Let $\mathcal{F} = H_2\widehat{\mathcal{F}}$, where $H_2 = \left[\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}\right]$. For any $\widehat{f} \in \widehat{\mathcal{F}}$ of arity n, let $H_2^{\otimes n}\widehat{f} = f$. One can translate the theorem statement to an equivalent statement in the Pl-#CSP setting, i.e., either Pl-#CSP$(\mathcal{F})$ is #P-hard, or $\mathcal{F} \subseteq \mathscr{P}$, or $\mathcal{F} \subseteq \mathscr{A}$. Recall that $\widehat{\mathscr{A}} = H_2\mathscr{A} = \mathscr{A}$.

If $\widehat{\mathcal{F}} \subseteq \widehat{\mathscr{P}}$ or $\widehat{\mathcal{F}} \subseteq \mathscr{A}$, equivalently if $\mathcal{F} \subseteq \mathscr{P}$ or $\mathcal{F} \subseteq \mathscr{A}$, then the problem Pl-#CSP$(\mathcal{F})$ is tractable by Theorem 2.33. Otherwise, there exist $\widehat{f}, \widehat{g} \in \widehat{\mathcal{F}}$ such that $\widehat{f} \notin \widehat{\mathscr{P}}$ and $\widehat{g} \notin \mathscr{A}$. Translating to the Pl-#CSP setting, there exist $f, g \in \mathcal{F}$ such that $f \notin \mathscr{P}$ and $g \notin \mathscr{A}$.

Moreover, by Lemma 3.1, we can construct $[1, w]$ in Pl-Holant$(\widehat{\mathcal{EQ}}, \widehat{\mathcal{F}})$, where $w \neq 0$, such that

$$\text{Pl-Holant}(\widehat{\mathcal{EQ}}, [1, w], \widehat{\mathcal{F}}) \leq_{\mathrm{T}} \text{Pl-Holant}(\widehat{\mathcal{EQ}}, \widehat{\mathcal{F}}).$$

This implies that

$$\text{Pl-\#CSP}(H_2[1, w], \mathcal{F}) \leq_{\mathrm{T}} \text{Pl-\#CSP}(\mathcal{F}).$$

Depending on the value of w, we will be able to finish the proof by one of the following two alternatives:

(A) Suppose we are able to construct a unary signature $[1, b]$ with $b^4 \neq 0, 1$ in Pl-#CSP$(\mathcal{F})$, i.e.,

$$\text{Pl-\#CSP}([1, b], \mathcal{F}) \leq_{\mathrm{T}} \text{Pl-\#CSP}(\mathcal{F}).$$

In this case we have unary signatures $\partial_{[1,b]}^{\{1,2\}}(=_3) = [1, b^2]$ and $\partial_{[1,b]}^{\{1,2,3\}}(=_4) = [1, b^3]$. So

$$\text{Pl-\#CSP}([1, b], [1, b^2], [1, b^3], \mathcal{F}) \leq_{\mathrm{T}} \text{Pl-\#CSP}(\mathcal{F}).$$

Note that $[1, b], [1, b^2], [1, b^3]$ are pairwise linearly independent since $b^4 \neq 0, 1$. We have $[1, b] \notin \mathscr{A} \cup \widehat{\mathscr{M}}$ by Propositions 2.17 and 2.32. Then by Theorem 3.6 and $f \in \mathcal{F} \setminus \mathscr{P}$, there exists a symmetric signature $f' \notin \mathscr{P}$ such that

$$\text{Pl-\#CSP}(f', [1, b], [1, b^2], [1, b^3], \mathcal{F}) \leq_{\mathrm{T}} \text{Pl-\#CSP}([1, b], [1, b^2], [1, b^3], \mathcal{F}).$$

Note that the symmetric signature set $\{[1,b], f'\}$ satisfies

$$\{[1,b], f'\} \not\subseteq \mathscr{P}, \quad \{[1,b], f'\} \not\subseteq \mathscr{A}, \quad \{[1,b], f'\} \not\subseteq \widehat{\mathscr{M}}.$$

So Pl-#CSP$(f', [1,b])$ is #P-hard by Theorem 2.34. Thus Pl-#CSP$(\mathcal{F})$ is #P-hard.

(B) Next, suppose we can construct $[0,1]$ and $[1,x]$ with $x^4 = 1$ in Pl-Holant$(\widehat{\mathcal{EQ}}, \widehat{\mathcal{F}})$, i.e.,

$$\text{Pl-Holant}(\widehat{\mathcal{EQ}}, [0,1], [1,x], \widehat{\mathcal{F}}) \leq_{\text{T}} \text{Pl-Holant}(\widehat{\mathcal{EQ}}, \widehat{\mathcal{F}}).$$

Then by $\widehat{g} \in \widehat{\mathcal{F}} \setminus \mathscr{A}$ and Lemma 3.9, we have a unary signature $[y,z]$ in Pl-Holant$(\widehat{\mathcal{EQ}}, \widehat{\mathcal{F}})$ that is not in $\mathscr{A}$. Translating into Pl-#CSP$(\mathcal{F})$, this means that we have a unary signature $H_2[y,z] \notin \mathscr{A}$, because $\mathscr{A}$ is invariant under H_2. Then we are done by the previous case.

Now we prove the theorem according to the value of w.

1. Suppose $w^4 \neq 1$. As $w \neq 0$ is given, by Proposition 2.17 we have $[1,w] \notin \mathscr{A}$, and thus we have $H_2[1,w] \notin \mathscr{A}$ in Pl-#CSP$(\mathcal{F})$. Thus we are done by alternative (A).
2. Suppose $w = \pm 1$.
   In this case, in Pl-#CSP$(\mathcal{F})$ we have $H_2[1,w] = [1+w, 1-w] = 2[1,0]$ if $w = 1$, or $2[0,1]$ if $w = -1$. If, up to a scalar, each signature in $\mathcal{F}$ takes value in $\{0,1\}$, then we are done by Lemma 3.11. Otherwise, we can get a unary signature $[1,c]$ with $c \neq 0, 1$ by Lemma 3.10 in Pl-#CSP$(\mathcal{F})$.
   - If $c^4 \neq 1$, then we are done by alternative (A), as $c \neq 0$ is given by Lemma 3.10.
   - If $c = \pm \mathrm{i}$, we translate into Pl-Holant$(\widehat{\mathcal{EQ}}, \widehat{\mathcal{F}})$ by $H_2^{-1} = \frac{1}{2} H_2$. So we have $H_2^{-1}[1,c] = \frac{1+c}{2}[1,-c]$, and therefore also $\partial^{\{1,2\}}_{[1,-c]}([1,0,1,0]) = -2c[0,1]$ in Pl-Holant$(\widehat{\mathcal{EQ}}, \widehat{\mathcal{F}})$, where the signature $[1,0,1,0] \in \widehat{\mathcal{EQ}}$. Thus we are done by alternative (B).
   - If $c = -1$, again we translate into Pl-Holant$(\widehat{\mathcal{EQ}}, \widehat{\mathcal{F}})$ by H_2^{-1}. In addition to $[1,w]$ where $w = \pm 1$, and so $w^4 = 1$, we also have $H_2^{-1}[1,-1] = [0,1]$ in Pl-Holant$(\widehat{\mathcal{EQ}}, \widehat{\mathcal{F}})$. Thus we are done by alternative (B).
3. For $w = \pm \mathrm{i}$, in addition to $[1,w]$ with $w^4 = 1$, we also have $\partial^{\{1,2\}}_{[1,w]}([1,0,1,0]) = 2w[0,1]$ in Pl-Holant$(\widehat{\mathcal{EQ}}, \widehat{\mathcal{F}})$. Thus we are done by alternative (B). ∎

Theorem 3.12 is a dichotomy for Pl-#CSP$(\mathcal{F})$ in the case when $\widehat{\mathcal{F}}$ does not satisfy the Parity Condition. It conforms to the final form of Theorem 6.1. Note that since some signature in $\widehat{\mathcal{F}}$ violates the Parity Condition, the potential tractability condition $\widehat{\mathcal{F}} \subseteq \mathscr{M}$ is impossible so it does not appear in the statement of Theorem 3.12.

4. A dichotomy theorem for Pl-CSP$^2(\widehat{\mathcal{EQ}}, \widehat{\mathcal{F}})$. In this section we prove a dichotomy theorem for Pl-CSP$^2(\widehat{\mathcal{EQ}}, \widehat{\mathcal{F}})$, Theorem 4.9, where all signatures in $\widehat{\mathcal{F}}$ satisfy the Parity Condition. By (2.2), we have

$$\text{Pl-CSP}^2(\widehat{\mathcal{EQ}}, \widehat{\mathcal{F}}) \equiv_{\text{T}} \text{Pl-Holant}(\mathcal{EQ}_2, \widehat{\mathcal{EQ}}, \widehat{\mathcal{F}}).$$

Theorem 4.9 will be used later in section 5 in the situation when we can construct $(=_4)$ in Pl-Holant$(\widehat{\mathcal{EQ}}, \widehat{\mathcal{F}})$. Then we have the following chain of equivalent problems:

$$\begin{aligned}\text{Pl-CSP}(\mathcal{F}) &\equiv_{\mathrm{T}} \text{Pl-Holant}(\widehat{\mathcal{EQ}}, \widehat{\mathcal{F}})\\ &\equiv_{\mathrm{T}} \text{Pl-Holant}(\widehat{\mathcal{EQ}}, (=_4), \widehat{\mathcal{F}}) \quad \text{(when we can construct } (=_4))\\ &\equiv_{\mathrm{T}} \text{Pl-Holant}(\widehat{\mathcal{EQ}}, \mathcal{EQ}_2, \widehat{\mathcal{F}}) \quad \text{(by Lemma 2.3)}\\ &\equiv_{\mathrm{T}} \text{Pl-CSP}^2(\widehat{\mathcal{EQ}}, \widehat{\mathcal{F}}).\end{aligned}$$

By Proposition 2.17, a binary signature $[1, 0, x]$ is not in $\mathscr{A}$ iff $x^4 \neq 0, 1$. Suppose we have some $[1, 0, x]$. The following lemma says that if $[1, 0, x] \notin \mathscr{A}$, then we can get $[1, 0, z]$ for any $z \in \mathbb{C}$, as well as $[0, 1]^{\otimes 2}$, in Pl-Holant$(\widehat{\mathcal{EQ}}, \widehat{\mathcal{F}})$. Moreover, even if $x = \pm\mathrm{i}$, we still can get $[0, 1]^{\otimes 2}$ and $[1, 0, -1]$ from $[1, 0, x]$ in Pl-Holant$(\widehat{\mathcal{EQ}}, \widehat{\mathcal{F}})$. The proof of the lemma is the same as the proof of Lemma 8.3 in [27].

LEMMA 4.1. *If $x \in \mathbb{C}$ and $[1, 0, x] \notin \mathscr{A}$, then for any $z \in \mathbb{C}$,*

$$\text{Pl-Holant}([1, 0, z], [0, 1]^{\otimes 2}, \widehat{\mathcal{EQ}}, \widehat{\mathcal{F}}) \leq_{\mathrm{T}} \text{Pl-Holant}([1, 0, x], \widehat{\mathcal{EQ}}, \widehat{\mathcal{F}}).$$

Moreover, if $x = \pm\mathrm{i}$, we have

$$\text{Pl-Holant}([1, 0, -1], [0, 1]^{\otimes 2}, \widehat{\mathcal{EQ}}, \widehat{\mathcal{F}}) \leq_{\mathrm{T}} \text{Pl-Holant}([1, 0, x], \widehat{\mathcal{EQ}}, \widehat{\mathcal{F}}).$$

Proof. We will use the following gadget (Figure 8) from [27], where circle vertices are assigned $[1, 0, 1, 0] \in \widehat{\mathcal{EQ}}$ and square vertices are assigned $[1, 0, x]$. It has signature $[1 + x^2, 0, 2x]$.

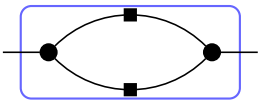


FIG. 8. *A gadget with signature* $[1 + x^2, 0, 2x]$.

- For $[1, 0, x] \notin \mathscr{A}$ we have $x^4 \neq 0, 1$ by Proposition 2.17. If $|x| \neq 1$, by combining k copies of $[1, 0, x]$, we have $[1, 0, x^k]$. Then we can use polynomial interpolation to get a reduction
$$\text{Pl-Holant}([y, 0, z], \widehat{\mathcal{EQ}}, \widehat{\mathcal{F}}) \leq_{\mathrm{T}} \text{Pl-Holant}([1, 0, x], \widehat{\mathcal{EQ}}, \widehat{\mathcal{F}})$$
for any $y, z \in \mathbb{C}$. In particular, we can get $[1, 0, z]$ for any $z \in \mathbb{C}$ and $[0, 0, 1] = [0, 1]^{\otimes 2}$.
Otherwise, $|x| = 1$. In $[1 + x^2, 0, 2x]$, $0 < |1 + x^2| < 2$ by $x \neq \pm\mathrm{i}$ and $x \neq \pm 1$. However, $|2x| = 2$. Therefore, after normalizing, the signature $[1, 0, \frac{2x}{1+x^2}]$ can interpolate $[1, 0, z]$ for any $z \in \mathbb{C}$ and $[0, 0, 1] = [0, 1]^{\otimes 2}$.
- For $x = \pm\mathrm{i}$, $[1 + x^2, 0, 2x] = \pm 2\mathrm{i}[0, 0, 1] = \pm 2\mathrm{i}[0, 1]^{\otimes 2}$, and by combining two copies of $[1, 0, x]$, we have $[1, 0, -1]$. ∎

A very desirable tool is to pin a variable to 0 or 1. This means we would like to have $[1, 0]$ and $[0, 1]$. We do have $[1, 0] \in \widehat{\mathcal{EQ}}$. However, if all signatures in $\widehat{\mathcal{F}}$ satisfy the even Parity Condition, namely $f_\alpha = 0$ for all α of odd weight, then every signature constructed in Pl-Holant$(\widehat{\mathcal{EQ}}, \widehat{\mathcal{F}})$ satisfies the even Parity Condition. Therefore, it is impossible to construct $[0, 1]$. But it is possible to construct $[0, 1]^{\otimes 2}$. The next lemma shows that with $\widehat{\mathcal{EQ}}$, getting $[0, 1]^{\otimes 2}$ is almost as good as getting $[0, 1]$.

LEMMA 4.2. *For $\mathscr{C} = \mathscr{A}$ or $\mathscr{M}$, if there exists $f \in \widehat{\mathcal{F}}$ of arity $n \geq 2$ such that $f^{x_i=1} \notin \mathscr{C}$ for some $i \in [n]$, then there exists a signature $g \notin \mathscr{C}$ with* $\text{arity}(g) = n - 1$ *such that*

$$\text{Pl-Holant}(\widehat{\mathcal{EQ}}, [0,1]^{\otimes 2}, g, \widehat{\mathcal{F}}) \leq_{\text{T}} \text{Pl-Holant}(\widehat{\mathcal{EQ}}, [0,1]^{\otimes 2}, \widehat{\mathcal{F}}).$$

Furthermore, if f satisfies the even Parity Condition, then so does g.

Proof. We have that $[1,0,1,0] \in \widehat{\mathcal{EQ}}$. By having $[0,1]^{\otimes 2}$, and the fact that $\partial_{[0,1]}([1,0,1,0]) = [0,1,0]$, we get $[0,1,0]^{\otimes 2}$ by applying $\partial_{[0,1]^{\otimes 2}}$ on $[1,0,1,0]^{\otimes 2}$. Then by having $[1,0] \in \widehat{\mathcal{EQ}}$, we get $h(x_1,x_2,x_3) = [0,1,0] \otimes [0,1]$, where the binary disequality is on x_1, x_2 and the unary $[0,1]$ is on x_3. By connecting the variable x_2 of h to the variable x_{i+1} (if $i = n$, then let $x_{i+1} = x_1$) of f and connecting the variable x_3 of h to the variable x_i of f (see Figure 9), the gadget gives an $(n-1)$-ary signature g, such that

$$g(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n) = \sum_{x'_i, x'_{i+1} \in \{0,1\}} f(x_1, \ldots, x'_i, x'_{i+1}, \ldots, x_n) h(x_{i+1}, x'_{i+1}, x'_i).$$

Notice that the variables of f and h are ordered counterclockwise, and connections respect this order in a planar fashion.

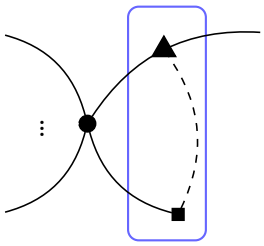


FIG. 9. *The circle vertex is assigned f, the square vertex denotes $[0,1]$, and the triangle vertex denotes $[0,1,0]$. The two nodes connected by the dashed line are a single signature $[0,1,0] \otimes [0,1]$.*

We have

$$g(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n) = f^{x_i=1}(x_1, \ldots, x_{i-1}, \overline{x_{i+1}}, \ldots, x_n).$$

Thus $g \in \mathscr{A}$ iff $f^{x_i=1} \in \mathscr{A}$, and $g \in \mathscr{M}$ iff $f^{x_i=1} \in \mathscr{M}$, by Lemma 2.36. ∎

The next lemma shows that for an n-ary signature with affine support and a set of free variables $X = \{x_{i_1}, x_{i_2}, \ldots, x_{i_k}\}$, if two consecutive variables $x_s, x_{s+1} \notin X$, then we can combine the two variables to one new variable using $[1,0,1,0]$, without changing the compressed signature.

LEMMA 4.3. *Let f be an n-ary signature with affine support of dimension k, and let $X = \{x_{i_1}, x_{i_2}, \ldots, x_{i_k}\}$ be a set of free variables. If there exists $s \in [n]$ such that X does not include x_s, x_{s+1} (if $s = n$, then $x_{s+1} = x_1$), letting*

$$\begin{aligned} &g(x_1 \ldots, x_{s-1}, x', x_{s+2}, \ldots, x_n) \\ &= \sum_{x_s, x_{s+1} \in \{0,1\}} f(x_1 \ldots, x_{s-1}, x_s, x_{s+1}, \ldots, x_n)[1,0,1,0](x_{s+1}, x_s, x'), \end{aligned}$$

then g has affine support, X is a set of free variables of g, and $\underline{f_X} = \underline{g_X}$. On its support, $x' = x_s \oplus x_{s+1}$. Furthermore, if f satisfies the even Parity Condition, then so does g.

Proof. We first note that g as defined is the signature of a planar gadget using f and $[1,0,1,0]$; the order of variable connections respects planarity. Since X is a set of free variables for f, if we fix an assignment on X, then there exist unique $x_s = x_s(X)$, $x_{s+1} = x_{s+1}(X)$ as affine linear functions such that f is nonzero. Moreover, in

$$\sum_{x_s, x_{s+1} \in \{0,1\}} f(x_1 \ldots, x_{s-1}, x_s, x_{s+1}, \ldots, x_n)[1,0,1,0](x_{s+1}, x_s, x'),$$

if $[1,0,1,0]$ takes value 1, then x' must be $x_{s+1} \oplus x_s$. This implies that g has affine support, and X is a set of free variables. It follows that $\underline{f_X} = \underline{g_X}$. ∎

The following two lemmas describe how to find a set of free variables that includes an adjacent pair of variables.

LEMMA 4.4. *Let f be a signature with affine support of dimension $k \geq 2$, and suppose there are no variables that take a constant value in the support. Then there exists a set of free variables $X = \{x_{i_1}, x_{i_2}, \ldots, x_{i_k}\}$ where some two variables are adjacent.*

Proof. Let $X = \{x_{i_1}, x_{i_2}, \ldots, x_{i_k}\}$ be the set of free variables which is minimum in the lexicographic order. If $i_2 = i_1 + 1$, then we are done. Otherwise, x_{i_2-1} is not in X. If its dependency $x_{i_2-1} = \sum_{j=1}^{k} a_j x_{i_j} + b$ involves any variable other than x_{i_1}, namely if $a_j \neq 0$ for some $1 < j \leq k$, then by switching x_{i_2-1} with the variable x_{i_j}, we get another set of free variables which is lexicographically smaller than X, a contradiction. Thus we have $x_{i_2-1} = a x_{i_1} + b$. Since there are no variables that take a constant value in the support, we have $a \neq 0$. So $X' = \{x_{i_2-1}, x_{i_2}, \ldots, x_{i_k}\}$ is a set of free variables that includes x_{i_2-1}, x_{i_2}. ∎

LEMMA 4.5. *Let f be a 5-ary signature with affine support of dimension 3. Suppose there are no variables that take a constant value in the support. Then there exists a set of free variables X such that the variables in X are consecutive in a cyclic sense.*

Proof. By Lemma 4.4, without loss of generality, we can assume that there exists a set of free variables including x_1, x_2. If the other free variable is x_3 or x_5, then we are done. Otherwise, $\{x_1, x_2, x_4\}$ is a set of free variables and $x_3 = a_1 x_1 + a_2 x_2 + a_4 x_4 + c$ and $x_5 = b_1 x_1 + b_2 x_2 + b_4 x_4 + d$, where $a_i, b_i, c, d \in \mathbb{Z}_2$ for $i \in \{1, 2, 4\}$.

- If $a_1 \neq 0$, then $\{x_2, x_3, x_4\}$ is a set of free variables.
- If $a_4 \neq 0$, then $\{x_1, x_2, x_3\}$ is a set of free variables.
- If $b_2 \neq 0$, then $\{x_1, x_4, x_5\}$ is a set of free variables.
- If $b_4 \neq 0$, then $\{x_1, x_2, x_5\}$ is a set of free variables.
- If $a_1 = a_4 = 0$ and $b_2 = b_4 = 0$, then $a_2 \neq 0$ and $b_1 \neq 0$ since there are no variables that take a constant value in the support. So $\{x_3, x_4, x_5\}$ is a set of free variables.

This finishes the proof. ∎

If $\widehat{\mathcal{F}} \subseteq \mathscr{A}$, then Pl-Holant$(\widehat{\mathcal{EQ}}, \widehat{\mathcal{F}})$ is tractable. Otherwise, there exists $f \in \widehat{\mathcal{F}} \setminus \mathscr{A}$. The following three lemmas are about reducing the arity of f. Since all signatures in $\widehat{\mathcal{EQ}} \cup \widehat{\mathcal{F}}$ satisfy the Parity Condition, any constructible unary signature in Pl-Holant$(\widehat{\mathcal{EQ}}, \widehat{\mathcal{F}})$ also satisfies the Parity Condition and therefore is in $\mathscr{A}$. So the lowest arity a constructible nonaffine signature can have is 2. Furthermore, a binary signature satisfying the even Parity Condition is symmetric and takes the form $[a, 0, b]$. By Proposition 2.17, $[1, 0, x] \notin \mathscr{A}$ iff $x^4 \neq 0, 1$. The next lemma implies that in Pl-Holant$(\widehat{\mathcal{EQ}}, [0,1]^{\otimes 2}, [1, 0, -1], \widehat{\mathcal{F}})$, from any $f \in \widehat{\mathcal{F}} \setminus \mathscr{A}$ we can construct some $[1, 0, x] \notin \mathscr{A}$.

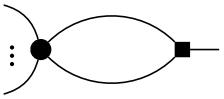


FIG. 10. *The circle vertex is assigned f and the square vertex is assigned* $[1,0,1,0]$.

LEMMA 4.6. *If all signatures in $\widehat{\mathcal{F}}$ satisfy the Parity Condition and $\widehat{\mathcal{F}} \not\subseteq \mathscr{A}$, then there exists $[1,0,x] \notin \mathscr{A}$ such that*

$$\text{Pl-Holant}([1,0,x], \widehat{\mathcal{EQ}}, [0,1]^{\otimes 2}, [1,0,-1], \widehat{\mathcal{F}}) \leq_{\text{T}} \text{Pl-Holant}(\widehat{\mathcal{EQ}}, [0,1]^{\otimes 2}, [1,0,-1], \widehat{\mathcal{F}}).$$

Proof. Since $\widehat{\mathcal{F}} \not\subseteq \mathscr{A}$, there exists $f \in \widehat{\mathcal{F}} \setminus \mathscr{A}$. By Lemma 2.38, we can assume that $f_{00\cdots0} = 1$ and f satisfies the even Parity Condition. If f has arity 1, then $f = [1,0] \in \mathscr{A}$. This is a contradiction. If f has arity 2, then $f = [1,0,x]$ with $x^4 \neq 0, 1$ and we are done. In the following, we assume that f has arity ≥ 3. If we can construct a nonaffine signature with arity $\leq n-1$, then we are done by induction.

If there exists $i \in [n]$ such that $f^{x_i=0} \notin \mathscr{A}$, then we are done by induction since we have $[1,0] \in \widehat{\mathcal{EQ}}$. If there exists $i \in [n]$ such that $f^{x_i=1} \notin \mathscr{A}$, then we are done by induction and Lemma 4.2. So in the following we assume that both $f^{x_i=0}$ and $f^{x_i=1}$ are affine signatures for any $i \in [n]$.

Claim. If $\text{supp}(f)$ is not affine, then we can construct a signature that is not in $\mathscr{A}$ with arity $\leq n-1$.

Certainly $\text{supp}(f)$ is not a linear subspace. Note that $(0,0,\ldots,0) \in \text{supp}(f)$. A subset of $\mathbb{Z}_2^n$ containing $(0,0,\ldots,0)$ is affine iff it is a linear subspace. So $\text{supp}(f^{x_i=0})$ is a linear subspace of $\mathbb{Z}_2^{n-1}$ since $f^{x_i=0}$ is affine and $f^{x_i=0}_{0\cdots0} = 1$. By Lemma 3.8, there exist $\mathbf{a} = a_1a_2\cdots a_n, \mathbf{b} = b_1b_2\cdots b_n$, such that $\mathbf{a}, \mathbf{b} \in \text{supp}(f)$, $\mathbf{c} = \mathbf{a} \oplus \mathbf{b} = c_1c_2\cdots c_n \notin \text{supp}(f)$, and there exists $i \in [n]$ such that $a_i \neq b_i$. Without loss of generality, we assume that $a_1 = 0, b_1 = 1$. It follows that $c_1 = 1$. Let $\mathbf{a}' = a_3\cdots a_n, \mathbf{b}' = b_3\cdots b_n, \mathbf{c}' = c_3\cdots c_n$.

By connecting one variable of $[1,0,-1]$ to the first variable of f, we get a gadget that gives

$$\begin{aligned}\bar{f}(x_1,x_2,\ldots,x_n) &= \sum_{x_1' \in \{0,1\}} [1,0,-1](x_1,x_1')f(x_1',x_2,\ldots,x_n)\\ &= (-1)^{x_1} f(x_1,x_2,\ldots,x_n).\end{aligned}$$

Moreover, by connecting the variables x_2, x_1 of $[1,0,1,0]$ to the variables x_1, x_2 of f, respectively, the planar gadget in Figure 10 gives the signature $h(x', x_3, x_4, \ldots, x_n)$ satisfying

$$h(x', x_3, x_4, \ldots, x_n) = \sum_{x_1, x_2 \in \{0,1\}} [1,0,1,0](x_2, x_1, x')f(x_1, x_2, \ldots, x_n).$$

This way of connecting the variables satisfies planarity. Note that

$$h(x_1 \oplus x_2, x_3, \ldots, x_n) = f(x_1, x_2, x_3, \ldots, x_n) + f(\overline{x_1}, \overline{x_2}, x_3, \ldots, x_n).$$

Similarly, by connecting the variables x_2, x_1 of $[1,0,1,0]$ to the variables x_1, x_2 of $\bar{f}$, respectively, the planar gadget gives the signature $\bar{h}(x', x_3, x_4, \ldots, x_n)$ satisfying

$$\bar{h}(x', x_3, x_4, \ldots, x_n) = \sum_{x_1, x_2 \in \{0,1\}} [1,0,1,0](x_2, x_1, x')\bar{f}(x_1, x_2, \ldots, x_n),$$

and we have

$$\bar{h}(x_1 \oplus x_2, x_3, \ldots, x_n) = (-1)^{x_1} f(x_1, x_2, x_3, \ldots, x_n) + (-1)^{\overline{x_1}} f(\overline{x_1}, \overline{x_2}, x_3, \ldots, x_n).$$

If $h \notin \mathscr{A}$ or $\bar{h} \notin \mathscr{A}$, then we are done since both h and $\bar{h}$ have arity $n-1$. In the following, we assume that both of h and $\bar{h}$ are affine.

Now we use the Tableau Calculus.

- If both $\bar{a}_1\bar{a}_2\mathbf{a}'$ and $\bar{b}_1\bar{b}_2\mathbf{b}'$ are not in $\mathrm{supp}(f)$, then

$$\begin{aligned} h_{(a_1\oplus a_2)\mathbf{a}'} &= f_{a_1a_2\mathbf{a}'} + f_{\bar{a}_1\bar{a}_2\mathbf{a}'} \neq 0, \\ h_{(b_1\oplus b_2)\mathbf{b}'} &= f_{b_1b_2\mathbf{b}'} + f_{\bar{b}_1\bar{b}_2\mathbf{b}'} \neq 0 \end{aligned} \tag{4.1}$$

since $f_{a_1a_2\mathbf{a}'} \neq 0, f_{b_1b_2\mathbf{b}'} \neq 0$, and $f_{\bar{a}_1\bar{a}_2\mathbf{a}'} = f_{\bar{b}_1\bar{b}_2\mathbf{b}'} = 0$. Similarly, we have

$$\begin{aligned} \bar{h}_{(a_1\oplus a_2)\mathbf{a}'} &= f_{a_1a_2\mathbf{a}'} - f_{\bar{a}_1\bar{a}_2\mathbf{a}'} \neq 0, \\ \bar{h}_{(b_1\oplus b_2)\mathbf{b}'} &= -f_{b_1b_2\mathbf{b}'} + f_{\bar{b}_1\bar{b}_2\mathbf{b}'} \neq 0. \end{aligned} \tag{4.2}$$

We have

$$\begin{aligned} h_{00\cdots0} &= f_{000\cdots0} + f_{110\cdots0}, \\ \bar{h}_{00\cdots0} &= f_{000\cdots0} - f_{110\cdots0}. \end{aligned}$$

Then by $f_{000\cdots0} \neq 0$, either $h_{00\cdots0}$ or $\bar{h}_{00\cdots0}$ is nonzero.

1. Suppose $h_{00\cdots0} \neq 0$. Note that h is affine, so $\mathrm{supp}(h)$ is a linear subspace. Then, by (4.1),
$$\begin{array}{rll} & (a_1 \oplus a_2)\mathbf{a}' & \in \mathrm{supp}(h) \\ \oplus & (b_1 \oplus b_2)\mathbf{b}' & \in \mathrm{supp}(h) \\ \hline & (c_1 \oplus c_2)\mathbf{c}' & \end{array}$$
and we have $(c_1 \oplus c_2)\mathbf{c}' \in \mathrm{supp}(h)$. In other words, $h_{(c_1\oplus c_2)\mathbf{c}'} \neq 0$, which is just $f_{c_1c_2\mathbf{c}'} + f_{\bar{c}_1\bar{c}_2\mathbf{c}'} \neq 0$. Since $f_{c_1c_2\mathbf{c}'} = 0$, we have $f_{\bar{c}_1\bar{c}_2\mathbf{c}'} \neq 0$.
2. Suppose $\bar{h}_{00\cdots0} \neq 0$. Now we use $\bar{h}$ in place of h and use (4.2). With the same proof as above we can derive the same conclusion that $f_{\bar{c}_1\bar{c}_2\mathbf{c}'} \neq 0$.

Hence, in either case, we have $\bar{c}_1\bar{c}_2\mathbf{c}' \in \mathrm{supp}(f)$. Then $\bar{c}_2\mathbf{c}' \in \mathrm{supp}(f^{x_1=0})$ since $\bar{c}_1 = 0$. Note that $\mathrm{supp}(f^{x_1=0})$ is a linear subspace and

$$\begin{array}{rll} & a_2\mathbf{a}' & \in \mathrm{supp}(f^{x_1=0}) \\ \oplus & \bar{c}_2\mathbf{c}' & \in \mathrm{supp}(f^{x_1=0}) \\ \hline & \bar{b}_2\mathbf{b}' & \end{array}$$

and we have $\bar{b}_2\mathbf{b}' \in \mathrm{supp}(f^{x_1=0})$. This implies that $\bar{b}_1\bar{b}_2\mathbf{b}' \in \mathrm{supp}(f)$. This contradicts the hypothesis $\bar{b}_1\bar{b}_2\mathbf{b}' \notin \mathrm{supp}(f)$.

- If both $\bar{a}_1\bar{a}_2\mathbf{a}'$ and $\bar{b}_1\bar{b}_2\mathbf{b}'$ are in $\mathrm{supp}(f)$, then by

$$\begin{array}{rll} & a_2\mathbf{a}' & \in \mathrm{supp}(f^{x_1=0}) \\ \oplus & \bar{b}_2\mathbf{b}' & \in \mathrm{supp}(f^{x_1=0}) \\ \hline & \bar{c}_2\mathbf{c}' & \end{array}$$

we have $\bar{c}_2\mathbf{c}' \in \mathrm{supp}(f^{x_1=0})$. Thus $\bar{c}_1\bar{c}_2\mathbf{c}' \in \mathrm{supp}(f)$ since $\bar{c}_1 = 0$.

We claim that all of $f_{a_1a_2\mathbf{a}'}, f_{\bar{a}_1\bar{a}_2\mathbf{a}'}, f_{b_1b_2\mathbf{b}'}, f_{\bar{b}_1\bar{b}_2\mathbf{b}'}, f_{\bar{c}_1\bar{c}_2\mathbf{c}'}$ are powers of $\mathbf{i}$. First, since $f^{x_1=0}$ is affine and all of $f_{a_1a_2\mathbf{a}'}, f_{\bar{b}_1\bar{b}_2\mathbf{b}'}, f_{\bar{c}_1\bar{c}_2\mathbf{c}'}$ are nonzero entries

of $f^{x_1=0}$, we derive that all of $f_{a_1a_2\mathbf{a}'}$, $f_{\bar{b}_1\bar{b}_2\mathbf{b}'}$, $f_{\bar{c}_1\bar{c}_2\mathbf{c}'}$ are powers of $\mathfrak{i}$ by $f_{00\cdots0}=1$. Second, since $f^{x_3=a_3}$ is affine, both $f_{a_1a_2\mathbf{a}'}$ and $f_{\bar{a}_1\bar{a}_2\mathbf{a}'}$ are nonzero entries of $f^{x_3=a_3}$, and $f_{a_1a_2\mathbf{a}'}$ is a power of $\mathfrak{i}$, we derive that $f_{\bar{a}_1\bar{a}_2\mathbf{a}'}$ is a power of $\mathfrak{i}$. Finally, since $f^{x_3=b_3}$ is affine, both $f_{b_1b_2\mathbf{b}'}$ and $f_{\bar{b}_1\bar{b}_2\mathbf{b}'}$ are nonzero entries of $f^{x_3=b_3}$, and $f_{\bar{b}_1\bar{b}_2\mathbf{b}'}$ is a power of $\mathfrak{i}$, we derive that $f_{b_1b_2\mathbf{b}'}$ is a power of $\mathfrak{i}$.
By

$$h_{(c_1\oplus c_2)\mathbf{c}'}=f_{c_1c_2\mathbf{c}'}+f_{\bar{c}_1\bar{c}_2\mathbf{c}'},$$
$$\bar{h}_{(c_1\oplus c_2)\mathbf{c}'}=-f_{c_1c_2\mathbf{c}'}+f_{\bar{c}_1\bar{c}_2\mathbf{c}'},$$

we have $|h_{(c_1\oplus c_2)\mathbf{c}'}|=|\bar{h}_{(c_1\oplus c_2)\mathbf{c}'}|=1$ since $f_{\bar{c}_1\bar{c}_2\mathbf{c}'}$ is a power of $\mathfrak{i}$ and $f_{c_1c_2\mathbf{c}'}=0$. On the other hand,

$$\begin{aligned} h_{(a_1\oplus a_2)\mathbf{a}'}&=f_{a_1a_2\mathbf{a}'}+f_{\bar{a}_1\bar{a}_2\mathbf{a}'},\\ \bar{h}_{(a_1\oplus a_2)\mathbf{a}'}&=f_{a_1a_2\mathbf{a}'}-f_{\bar{a}_1\bar{a}_2\mathbf{a}'},\end{aligned} \tag{4.3}$$

and both $h_{(a_1\oplus a_2)\mathbf{a}'}$ and $\bar{h}_{(a_1\oplus a_2)\mathbf{a}'}$ are sums of two quantities, each a power of $\mathfrak{i}$. If at least one of them is not zero, then it has norm 2 or $\sqrt{2}$. This implies that h or $\bar{h}$ is not affine. This is a contradiction.
On the other hand, if both $h_{(a_1\oplus a_2)\mathbf{a}'}$ and $\bar{h}_{(a_1\oplus a_2)\mathbf{a}'}$ are zero, then $f_{a_1a_2\mathbf{a}'}=0$, by treating (4.3) as a linear system. This contradicts that $a_1a_2\mathbf{a}'\in\operatorname{supp}(f)$.

- If $\bar{a}_1\bar{a}_2\mathbf{a}'\in\operatorname{supp}(f)$ and $\bar{b}_1\bar{b}_2\mathbf{b}'\notin\operatorname{supp}(f)$, we claim that all of $f_{a_1a_2\mathbf{a}'}$, $f_{\bar{a}_1\bar{a}_2\mathbf{a}'}$, $f_{b_1b_2\mathbf{b}'}$ are powers of $\mathfrak{i}$. First, since $f^{x_1=0}$ is affine and both $f_{a_1a_2\mathbf{a}'}$ and $f_{00\cdots0}$ are nonzero entries of $f^{x_1=0}$, $f_{a_1a_2\mathbf{a}'}$ is a power of $\mathfrak{i}$ by $f_{00\cdots0}=1$. Second, since $f^{x_3=a_3}$ is affine and both $f_{a_1a_2\mathbf{a}'}$ and $f_{\bar{a}_1\bar{a}_2\mathbf{a}'}$ are nonzero entries of $f^{x_3=a_3}$, $f_{\bar{a}_1\bar{a}_2\mathbf{a}'}$ is a power of $\mathfrak{i}$. Finally, since $f^{x_1=1}$ is affine and both $f_{b_1b_2\mathbf{b}'}$ and $f_{\bar{a}_1\bar{a}_2\mathbf{a}'}$ are nonzero entries of $f^{x_1=1}$, $f_{b_1b_2\mathbf{b}'}$ is a power of $\mathfrak{i}$.
  By

$$h_{(b_1\oplus b_2)\mathbf{b}'}=f_{b_1b_2\mathbf{b}'}+f_{\bar{b}_1\bar{b}_2\mathbf{b}'},$$
$$\bar{h}_{(b_1\oplus b_2)\mathbf{b}'}=-f_{b_1b_2\mathbf{b}'}+f_{\bar{b}_1\bar{b}_2\mathbf{b}'},$$

  we have $|h_{(b_1\oplus b_2)\mathbf{b}'}|=|\bar{h}_{(b_1\oplus b_2)\mathbf{b}'}|=1$ since $f_{b_1b_2\mathbf{b}'}$ is a power of $\mathfrak{i}$ and $f_{\bar{b}_1\bar{b}_2\mathbf{b}'}=0$. Moreover, by

$$h_{(a_1\oplus a_2)\mathbf{a}'}=f_{a_1a_2\mathbf{a}'}+f_{\bar{a}_1\bar{a}_2\mathbf{a}'},$$
$$\bar{h}_{(a_1\oplus a_2)\mathbf{a}'}=f_{a_1a_2\mathbf{a}'}-f_{\bar{a}_1\bar{a}_2\mathbf{a}'},$$

  and an argument similar to that of the previous case, at least one of $h_{(a_1\oplus a_2)\mathbf{a}'}$ and $\bar{h}_{(\bar{a}_1\oplus\bar{a}_2)\mathbf{a}'}$ has norm 2 or $\sqrt{2}$. This implies that h or $\bar{h}$ is not affine. This is a contradiction.
- If $\bar{a}_1\bar{a}_2\mathbf{a}'\notin\operatorname{supp}(f)$ and $\bar{b}_1\bar{b}_2\mathbf{b}'\in\operatorname{supp}(f)$, the proof is symmetric by reversing the order of $\mathbf{a}$ and $\mathbf{b}$ in the previous item.

This completes the proof of the claim.

Now we can assume that $\operatorname{supp}(f)$ is affine and has dimension k.

If $k=0$, then $f\in\mathscr{A}$. This is a contradiction.

If $k=1$, then there exists exactly one $\alpha\in\{0,1\}^n$ such that $f_\alpha\neq0$ other than $f_{00\cdots0}=1$. Note that $\operatorname{wt}(\alpha)$ is even since f satisfies the even Parity Condition. The signature $\partial^S_{[1,0]}(f)$, where $S=\{k|$ the kth bit of α is $0\}$, has arity $\operatorname{wt}(\alpha)$ and has

exactly two nonzero values at $0^{\mathrm{wt}(\alpha)}$ and $1^{\mathrm{wt}(\alpha)}$. By connecting $\frac{\mathrm{wt}(\alpha)}{2}-1$ many copies of $=_2$ to $\partial^S_{[1,0]}(f)$ we get the binary signature $[1,0,f_\alpha]$. If $f_\alpha^4 \neq 1$, then we are done as $[1,0,f_\alpha] \notin \mathscr{A}$ by Proposition 2.17. Otherwise, f is affine. This is a contradiction.

If $k \geq 4$, then since both $f^{x_i=0}$ and $f^{x_i=1}$ are affine for all $i \in [n]$, we get $f \in \mathscr{A}$ by Lemma 2.18. This is a contradiction.

Thus we only need to consider $k = 2$ or $k = 3$. If on its support some variable x_i is a constant c, then $f = f^{x_i=c} \otimes [1,0](x_i)$ or $f^{x_i=c} \otimes [0,1](x_i)$ depending on whether $c = 0$ or 1, respectively. Then f would be affine, a contradiction. So no variable of f takes a constant value on its support.

- For $k = 2$, by Lemma 4.4, without loss of generality, we can assume that $\{x_1, x_2\}$ is a set of free variables. Then applying Lemma 4.3 repeatedly, we can get a ternary signature $\hat{f}$ such that

  $$\text{Pl-Holant}(\widehat{\mathcal{EQ}}, [0,1]^{\otimes 2}, [1,0,-1], \hat{f}, \widehat{\mathcal{F}}) \leq_{\mathrm{T}} \text{Pl-Holant}(\widehat{\mathcal{EQ}}, [0,1]^{\otimes 2}, [1,0,-1], \widehat{\mathcal{F}}),$$

  where the compressed signatures of $\hat{f}$ and f for $\{x_1, x_2\}$ are the same. By Lemma 4.3, $\hat{f}$ satisfies the even Parity Condition and $\hat{f}_{000} = f_{00\ldots0} = 1$. By Corollary 2.14, the compressed signature of f is not affine. Thus the compressed signature of $\hat{f}$ is not affine. So $\hat{f}$ is not affine.
  If there exists $i \in [3]$ such that $\hat{f}^{x_i=0} \notin \mathscr{A}$, then we are done by $[1,0] \in \widehat{\mathcal{EQ}}$. If there exists $i \in [3]$ such that $\hat{f}^{x_i=1} \notin \mathscr{A}$, then we are done by Lemma 4.2 and $[0,1]^{\otimes 2}$. Therefore, we may assume that $\hat{f}^{x_i=0}$, $\hat{f}^{x_i=1}$ are affine for all $i \in [3]$. Then there exist $r, s, t \in \{0,1,2,3\}$ such that $\hat{f}^{x_1=0} = [1,0,\mathrm{i}^r]$, $\hat{f}^{x_2=0} = [1,0,\mathrm{i}^s]$, $\hat{f}^{x_3=0} = [1,0,\mathrm{i}^t]$. So

  $$M_{x_1,x_2x_3}(\hat{f}) = \begin{bmatrix} \hat{f}_{000} & \hat{f}_{001} & \hat{f}_{010} & \hat{f}_{011} \\ \hat{f}_{100} & \hat{f}_{101} & \hat{f}_{110} & \hat{f}_{111} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & \mathrm{i}^r \\ 0 & \mathrm{i}^s & \mathrm{i}^t & 0 \end{bmatrix}.$$

  Note that the compressed signature of $\hat{f}$ for the free variable set $\{x_1, x_2\}$ is $[1, \mathrm{i}^r, \mathrm{i}^s, \mathrm{i}^t]$. It is affine iff $\mathrm{i}^t = \pm\mathrm{i}^{r+s}$ by Lemma 2.15. Since $\hat{f} \notin \mathscr{A}$, we have $\mathrm{i}^t = \pm\mathrm{i}^{r+s+1}$, i.e.,

  $$M_{x_1,x_2x_3}(\hat{f}) = \begin{bmatrix} 1 & 0 & 0 & \mathrm{i}^r \\ 0 & \mathrm{i}^s & \pm\mathrm{i}^{r+s+1} & 0 \end{bmatrix}.$$

  By connecting three copies of $\hat{f}^{x_2=0} = [1,0,\mathrm{i}^s]$ consecutively, the gadget gives $[1,0,\mathrm{i}^{3s}] = [1,0,\mathrm{i}^{-s}]$. Then we have (see (7.15))

  $$\hat{g}(x_1,x_2,x_3) = \sum_{x_1' \in \{0,1\}} [1,0,\mathrm{i}^{-s}](x_1,x_1')\hat{f}(x_1',x_2,x_3),$$

  with signature matrix

  $$M_{x_1,x_2x_3}(\hat{g}) = \begin{bmatrix} 1 & 0 & 0 & \mathrm{i}^r \\ 0 & 1 & \pm\mathrm{i}^{r+1} & 0 \end{bmatrix}, \quad \text{with} \quad M_{x_2,x_1x_3}(\hat{g}) = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & \mathrm{i}^r & \pm\mathrm{i}^{r+1} & 0 \end{bmatrix}.$$

  Connecting variables x_1, x_3 of $\hat{g}$ to variables x_3', x_1' of $[1,0,1,0]$, respectively, creates a planar gadget with signature

  $$\sum_{x_1,x_3 \in \{0,1\}} \hat{g}(x_1,x_2,x_3)[1,0,1,0](x_3,x_2',x_1) = [2,0,(1 \pm \mathrm{i})\mathrm{i}^r](x_1,x_2').$$

  This is not affine since the norms of 2 and $(1 \pm \mathrm{i})\mathrm{i}^r$ are different. Thus we are done.

- If the dimension k of the support is 3, then $n \geq 4$ since f satisfies the even Parity Condition.
  Claim. We can get a signature $\tilde{f}$ of arity 4 that has the same compressed signature as f. Furthermore, $\tilde{f}$ also satisfies the even Parity Condition.
  To prove this, if $n = 4$, then we set $\tilde{f} = f$. Otherwise, $n \geq 5$. By Lemma 4.4, without loss of generality, we can assume that there exists a set of free variables including $\{x_1, x_2\}$. If $\{x_1, x_2, x_3\}$ or $\{x_n, x_1, x_2\}$ is a set of free variables, then by applying Lemma 4.3 repeatedly, we can shrink the variables that are not in the free variable set to one variable while keeping planarity. Then we get a signature $\tilde{f}$ that has arity 4, and it has the same compressed signature as f.
  Otherwise, there exists k such that $\{x_1, x_2, x_k\}$ is a set of free variables, where $4 \leq k \leq n-1$. By Lemma 4.3, we can shrink the variables indexed by $3 \leq i \leq k-1$ to one variable x' and shrink the variables indexed by $k+1 \leq i \leq n$ to one variable x'' while keeping planarity. Then we get a signature $g(x_1, x_2, x', x_k, x'')$ with arity 5 that has the same compressed signature as f. Then, by Lemma 4.5, there exists a set of free variables which are consecutive. Then, by Lemma 4.3, we can get a signature $\tilde{f}$ with arity 4 that has the same compressed signature as f.
  In either case, we used Lemma 4.3 to derive the arity 4 signature $\tilde{f}$ from f; therefore, $\tilde{f}$ also satisfies the even Parity Condition.
  This completes the proof of the claim.
  Since f is not affine, $\tilde{f}$ is not affine. By Lemma 2.38 we can assume that

$$M_{x_1x_2,x_4x_3}(\tilde{f}) = \begin{bmatrix} \tilde{f}_{0000} & \tilde{f}_{0010} & \tilde{f}_{0001} & \tilde{f}_{0011} \\ \tilde{f}_{0100} & \tilde{f}_{0110} & \tilde{f}_{0101} & \tilde{f}_{0111} \\ \tilde{f}_{1000} & \tilde{f}_{1010} & \tilde{f}_{1001} & \tilde{f}_{1011} \\ \tilde{f}_{1100} & \tilde{f}_{1110} & \tilde{f}_{1101} & \tilde{f}_{1111} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & b \\ 0 & \alpha & \beta & 0 \\ 0 & \gamma & \delta & 0 \\ c & 0 & 0 & d \end{bmatrix}.$$

  Let $\underline{\tilde{f}}$ be the compressed signature of $\tilde{f}$ for $\{x_1, x_2, x_3\}$; then

$$M_{x_1,x_2x_3}(\underline{\tilde{f}}) = \begin{bmatrix} \tilde{f}_{0000} & \tilde{f}_{0011} & \tilde{f}_{0101} & \tilde{f}_{0110} \\ \tilde{f}_{1001} & \tilde{f}_{1010} & \tilde{f}_{1100} & \tilde{f}_{1111} \end{bmatrix} = \begin{bmatrix} 1 & b & \beta & \alpha \\ \delta & \gamma & c & d \end{bmatrix}. \tag{4.4}$$

  If there exists $i \in [4]$ such that $\tilde{f}^{x_i=0}$ is not affine, then we are done by $[1, 0] \in \widehat{\mathcal{EQ}}$. If there exists $i \in [4]$ such that $\tilde{f}^{x_i=1}$ is not affine, then we are done by Lemma 4.2 and $[0, 1]^{\otimes 2}$. Thus we may assume that both $\tilde{f}^{x_i=0}$ and $\tilde{f}^{x_i=1}$ are affine for $i \in [4]$. Since $\tilde{f}^{x_i=0}$ is affine for $i = 1, 2, 3$, by Lemma 2.15 there exist $r, s, t \in \{0, 1, 2, 3\}$ and $\epsilon_1, \epsilon_2, \epsilon_3 \in \{1, -1\}$ such that

$$M_{x_2,x_4x_3}(\tilde{f}^{x_1=0}) = \begin{bmatrix} 1 & 0 & 0 & b \\ 0 & \alpha & \beta & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & \mathfrak{i}^r \\ 0 & \epsilon_1\mathfrak{i}^{r+s} & \mathfrak{i}^s & 0 \end{bmatrix},$$

$$M_{x_1,x_4x_3}(\tilde{f}^{x_2=0}) = \begin{bmatrix} 1 & 0 & 0 & b \\ 0 & \gamma & \delta & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & \mathfrak{i}^r \\ 0 & \epsilon_2\mathfrak{i}^{r+t} & \mathfrak{i}^t & 0 \end{bmatrix},$$

$$M_{x_1x_2,x_4}(\tilde{f}^{x_3=0}) = \begin{bmatrix} 1 & 0 \\ 0 & \beta \\ 0 & \delta \\ c & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & \mathfrak{i}^s \\ 0 & \mathfrak{i}^t \\ \epsilon_3\mathfrak{i}^{s+t} & 0 \end{bmatrix}.$$

Moreover, since $\tilde{f}^{x_4=1}$ is affine, there exists $\epsilon_4 \in \{1,-1\}$ such that

$$M_{x_1x_2,x_3}(\tilde{f}^{x_4=1}) = \begin{bmatrix} 0 & b \\ \beta & 0 \\ \delta & 0 \\ 0 & d \end{bmatrix} = \begin{bmatrix} 0 & \mathfrak{i}^r \\ \mathfrak{i}^s & 0 \\ \mathfrak{i}^t & 0 \\ 0 & \epsilon_4\mathfrak{i}^{r+s+t} \end{bmatrix}.$$

So we have

$$M_{x_1x_2,x_4x_3}(\tilde{f}) = \begin{bmatrix} 1 & 0 & 0 & b \\ 0 & \alpha & \beta & 0 \\ 0 & \gamma & \delta & 0 \\ c & 0 & 0 & d \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & \mathfrak{i}^r \\ 0 & \epsilon_1\mathfrak{i}^{r+s} & \mathfrak{i}^s & 0 \\ 0 & \epsilon_2\mathfrak{i}^{r+t} & \mathfrak{i}^t & 0 \\ \epsilon_3\mathfrak{i}^{s+t} & 0 & 0 & \epsilon_4\mathfrak{i}^{r+s+t} \end{bmatrix}. \tag{4.5}$$

Note that we have $\partial^{\{3,4\}}_{[1,0]}(\tilde{f}) = [1,0,\epsilon_3\mathfrak{i}^{s+t}]$, $\partial^{\{1,2\}}_{[1,0]}(\tilde{f}) = [1,0,\mathfrak{i}^r]$. By connecting consecutively three copies of $[1,0,\epsilon_3\mathfrak{i}^{s+t}]$ (resp., of $[1,0,\mathfrak{i}^r]$), we have $[1,0,\epsilon_3^3\mathfrak{i}^{3s+3t}] = [1,0,\epsilon_3(\mathfrak{i}^{s+t})^{-1}]$ (resp., $[1,0,\mathfrak{i}^{3r}] = [1,0,\mathfrak{i}^{-r}]$). Let

$$h(x_1,x_2,x_3,x_4) = \sum_{x_2',x_4' \in \{0,1\}} \tilde{f}(x_1,x_2',x_3,x_4')[1,0,\epsilon_3(\mathfrak{i}^{s+t})^{-1}](x_2',x_2)[1,0,\mathfrak{i}^{-r}](x_4',x_4).$$

Then (see (7.19) and (7.21))

$$M_{x_1x_2,x_4x_3}(h) = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & \epsilon_1\epsilon_3\mathfrak{i}^{r-t} & \epsilon_3\mathfrak{i}^{-r-t} & 0 \\ 0 & \epsilon_2\mathfrak{i}^{r+t} & \mathfrak{i}^{t-r} & 0 \\ 1 & 0 & 0 & \epsilon_3\epsilon_4 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & \epsilon_1\epsilon_3(-1)^t a & \epsilon_3(-1)^{r+t}a & 0 \\ 0 & \epsilon_2 a & (-1)^r a & 0 \\ 1 & 0 & 0 & \epsilon_3\epsilon_4 \end{bmatrix},$$

where $a = \mathfrak{i}^{r+t}$.

Take two copies of h, and connect the third and fourth variables of one copy to the fourth and third variables of another copy to give the planar gadget (see Figure 11) with the signature

$$h'(x_1,x_2,x_3,x_4) = \sum_{x_3',x_4' \in \{0,1\}} h(x_1,x_2,x_3',x_4')h(x_3,x_4,x_4',x_3').$$

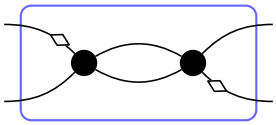


FIG. 11. *The two vertices are assigned h. The edges with a diamond indicate the first variable. Other variables are ordered counterclockwise.*

Note that (see Figure 2; for the second copy of h we rotate 180°)

$$M_{x_1x_2,x_4x_3}(h') = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & \epsilon_1\epsilon_3(-1)^t a & \epsilon_3(-1)^{r+t}a & 0 \\ 0 & \epsilon_2 a & (-1)^r a & 0 \\ 1 & 0 & 0 & \epsilon_3\epsilon_4 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & (-1)^r a & \epsilon_3(-1)^{r+t}a & 0 \\ 0 & \epsilon_2 a & \epsilon_1\epsilon_3(-1)^t a & 0 \\ 1 & 0 & 0 & \epsilon_3\epsilon_4 \end{bmatrix}$$

$$= \begin{bmatrix} 2 & 0 & 0 & 1+\epsilon_3\epsilon_4 \\ 0 & \epsilon_3(\epsilon_1+\epsilon_2) & 2\epsilon_1(-1)^t & 0 \\ 0 & 2\epsilon_2(-1)^t & \epsilon_3(\epsilon_1+\epsilon_2) & 0 \\ 1+\epsilon_3\epsilon_4 & 0 & 0 & 2 \end{bmatrix}.$$

Let $h'' = \partial^{\{1\}}_{[1,0]}(h')$; then

$$M_{x_2,x_4x_3}(h'') = \begin{bmatrix} 2 & 0 & 0 & 1+\epsilon_3\epsilon_4 \\ 0 & \epsilon_3(\epsilon_1+\epsilon_2) & 2\epsilon_1(-1)^t & 0 \end{bmatrix}.$$

Thus we have the following:

- If $\epsilon_1\epsilon_2\epsilon_3\epsilon_4 = -1$, then exactly one of $\epsilon_1 = -\epsilon_2$ or $\epsilon_3 = -\epsilon_4$ holds. This implies that h'' does not have affine support. Thus h'' is not affine. Then we are done by induction.
- If $\epsilon_1\epsilon_2\epsilon_3\epsilon_4 = 1$, by (4.4) and (4.5),

$$M_{x_1,x_2x_3}(\underline{\tilde{f}}) = \begin{bmatrix} 1 & \mathrm{i}^r & \mathrm{i}^s & \epsilon_1\mathrm{i}^{r+s} \\ \mathrm{i}^t & \epsilon_2\mathrm{i}^{r+t} & \epsilon_3\mathrm{i}^{s+t} & \epsilon_4\mathrm{i}^{r+s+t} \end{bmatrix},$$

where $\underline{\tilde{f}}$ is the compressed signature of f for $\{x_1, x_2, x_3\}$. Note that $\underline{\tilde{f}}$ is affine by Lemma 2.16. Thus f is affine. This is a contradiction. □

The next lemma shows how to reduce the arity of a nonaffine signature in Pl-Holant$(\widehat{\mathcal{EQ}}, \widehat{\mathcal{F}})$ when all signatures in $\widehat{\mathcal{F}}$ take values in $\{0,1\}$.

LEMMA 4.7. *Suppose all signatures in $\widehat{\mathcal{F}}$ take values in $\{0,1\}$ and satisfy the Parity Condition. If $\widehat{\mathcal{F}}$ contains a signature $f \notin \mathscr{A}$ of arity $n \geq 3$, then there exists a signature $g \notin \mathscr{A}$ of arity $< n$ such that*

$$\text{Pl-Holant}(\widehat{\mathcal{EQ}}, g, \widehat{\mathcal{F}}) \leq_{\mathrm{T}} \text{Pl-Holant}(\widehat{\mathcal{EQ}}, \widehat{\mathcal{F}}).$$

Furthermore, if f satisfies the even Parity Condition, then so does g.

Proof. The proof is by induction on n and uses the Tableau Calculus.

If there exists $i \in [n]$ such that $f^{x_i=0} \notin \mathscr{A}$, then we are done since we have $[1,0] \in \widehat{\mathcal{EQ}}$. In the following, we assume that $f^{x_i=0} \in \mathscr{A}$ for $1 \leq i \leq n$. By Lemma 2.38, we can assume that $f_{00\cdots0} = 1$ and f satisfies the even Parity Condition.

For a $\{0,1\}$-valued signature f, $f \in \mathscr{A}$ iff $\mathrm{supp}(f)$ is an affine subspace. Thus $\mathrm{supp}(f)$ is not an affine subspace and in particular not a linear subspace. A subset of $\mathbb{Z}_2^n$ containing $(0,0,\ldots,0)$ is affine iff it is a linear subspace. By $(0,0,\ldots,0) \in \mathrm{supp}(f)$ and $f^{x_i=0} \in \mathscr{A}$, $\mathrm{supp}(f^{x_i=0})$ is a linear subspace of $\mathbb{Z}_2^{n-1}$. By Lemma 3.8, there exist $\mathbf{a} = a_1a_2\cdots a_n, \mathbf{b} = b_1b_2\cdots b_n$ such that $\mathbf{a}, \mathbf{b} \in \mathrm{supp}(f)$, $\mathbf{c} = \mathbf{a} \oplus \mathbf{b} = c_1c_2\cdots c_n \notin \mathrm{supp}(f)$, and there exists $i \in [n]$ such that $a_i \neq b_i$. Without loss of generality, we assume that $a_1 = 0, b_1 = 1$. It follows that $c_1 = 1$. Let $\mathbf{a}' = a_3\cdots a_n, \mathbf{b}' = b_3\cdots b_n, \mathbf{c}' = c_3\cdots c_n$.

By connecting the variables x_2, x_1 of $[1,0,1,0]$ to the variables x_1, x_2 of f, respectively, the planar gadget gives the signature

$$h(x', x_3, x_4, \ldots, x_n) = \sum_{x_1,x_2\in\{0,1\}} [1,0,1,0](x_2, x_1, x')f(x_1, x_2, \ldots, x_n).$$

Note that

$$h(x_1 \oplus x_2, x_3, \ldots, x_n) = f(x_1, x_2, x_3, \ldots, x_n) + f(\bar{x}_1, \bar{x}_2, x_3, \ldots, x_n).$$

If $h \notin \mathscr{A}$, then we are done since h has arity $n-1$. So, in the following, we assume that $h \in \mathscr{A}$. Now comes the Tableau Calculus.

- For $\bar{a}_1\bar{a}_2\mathbf{a}' \in \mathrm{supp}(f)$, $\bar{b}_1\bar{b}_2\mathbf{b}' \in \mathrm{supp}(f)$, since $\mathrm{supp}(f^{x_1=0})$ is a linear subspace and
$$\begin{array}{rll} & a_2\mathbf{a}' & \in \mathrm{supp}(f^{x_i=0}) \\ \oplus & \bar{b}_2\mathbf{b}' & \in \mathrm{supp}(f^{x_i=0}) \\ \hline & \bar{c}_2\mathbf{c}' & \end{array}$$
we have $\bar{c}_2\mathbf{c}' \in \mathrm{supp}(f^{x_i=0})$. This means that $f_{\bar{c}_1\bar{c}_2\mathbf{c}'} \neq 0$ since $\bar{c}_1 = 0$. Note that
$$h_{(c_1\oplus c_2)\mathbf{c}'} = f_{c_1c_2\mathbf{c}'} + f_{\bar{c}_1\bar{c}_2\mathbf{c}'}.$$
Thus $|h_{(c_1\oplus c_2)\mathbf{c}'}| = 1$ since $f_{c_1c_2\mathbf{c}'} = 0$ and $f_{\bar{c}_1\bar{c}_2\mathbf{c}'} = 1$ (as f is $\{0,1\}$-valued). Moreover, by
$$h_{(a_1\oplus a_2)\mathbf{a}'} = f_{a_1a_2\mathbf{a}'} + f_{\bar{a}_1\bar{a}_2\mathbf{a}'},$$
we have $|h_{(a_1\oplus a_2)\mathbf{a}'}| = 2$, since $f_{a_1a_2\mathbf{a}'} = f_{\bar{a}_1\bar{a}_2\mathbf{a}'} = 1$. This implies that h is not affine, as nonzero values of h have different norms (Proposition 2.17). This is a contradiction.
- For $\bar{a}_1\bar{a}_2\mathbf{a}' \notin \mathrm{supp}(f)$, $\bar{b}_1\bar{b}_2\mathbf{b}' \notin \mathrm{supp}(f)$, by
$$h_{00\cdots0} = f_{000\cdots0} + f_{110\cdots0},$$
we have $h_{00\cdots0} \neq 0$ since $f_{000\cdots0} = 1$ and $f_{110\cdots0} \in \{0,1\}$. This implies that $\mathrm{supp}(h)$ is a linear subspace. By
$$h_{(a_1\oplus a_2)\mathbf{a}'} = f_{a_1a_2\mathbf{a}'} + f_{\bar{a}_1\bar{a}_2\mathbf{a}'},$$
$$h_{(b_1\oplus b_2)\mathbf{b}'} = f_{b_1b_2\mathbf{b}'} + f_{\bar{b}_1\bar{b}_2\mathbf{b}'},$$
we have $h_{(a_1\oplus a_2)\mathrm{a}'} = 1$, $h_{(b_1\oplus b_2)\mathbf{b}'} = 1$ since $f_{a_1a_2\mathbf{a}'} = f_{b_1b_2\mathbf{b}'} = 1$ and $f_{\bar{a}_1\bar{a}_2\mathbf{a}'} = f_{\bar{b}_1\bar{b}_2\mathbf{b}'} = 0$. This implies that $(a_1 \oplus a_2)\mathbf{a}', (b_1 \oplus b_2)\mathbf{b}' \in \mathrm{supp}(h)$. Then by
$$\begin{array}{rll} & (a_1 \oplus a_2)\mathbf{a}' & \in \mathrm{supp}(h) \\ \oplus & (b_1 \oplus b_2)\mathbf{b}' & \in \mathrm{supp}(h) \\ \hline & (c_1 \oplus c_2)\mathbf{c}' & \end{array}$$
we have $(c_1 \oplus c_2)\mathbf{c}' \in \mathrm{supp}(h)$. This implies that
$$h_{(c_1\oplus c_2)\mathbf{c}'} = f_{c_1c_2\mathbf{c}'} + f_{\bar{c}_1\bar{c}_2\mathbf{c}'} \neq 0.$$
Thus $f_{\bar{c}_1\bar{c}_2\mathbf{c}'} \neq 0$ since $f_{c_1c_2\mathbf{c}'} = 0$. So $\bar{c}_1\bar{c}_2\mathbf{c}' \in \mathrm{supp}(f)$. Hence, $\bar{c}_2\mathbf{c}' \in \mathrm{supp}(f^{x_1=0})$ since $\bar{c}_1 = 0$. By
$$\begin{array}{rll} & a_2\mathbf{a}' & \in \mathrm{supp}(f^{x_1=0}) \\ \oplus & \bar{c}_2\mathbf{c}' & \in \mathrm{supp}(f^{x_1=0}) \\ \hline & \bar{b}_2\mathbf{b}' & \end{array}$$
we have $\bar{b}_2\mathbf{b}' \in \mathrm{supp}(f^{x_1=0})$. Thus $\bar{b}_1\bar{b}_2\mathbf{b}' \in \mathrm{supp}(f)$ as $\bar{b}_1 = 0$. This is a contradiction.
- If $\bar{a}_1\bar{a}_2\mathbf{a}' \in \mathrm{supp}(f)$ and $\bar{b}_1\bar{b}_2\mathbf{b}' \notin \mathrm{supp}(f)$, then we consider
$$\begin{aligned} h_{(a_1\oplus a_2)\mathbf{a}'} &= f_{a_1a_2\mathbf{a}'} + f_{\bar{a}_1\bar{a}_2\mathbf{a}'}, \\ h_{(b_1\oplus b_2)\mathbf{b}'} &= f_{b_1b_2\mathbf{b}'} + f_{\bar{b}_1\bar{b}_2\mathbf{b}'}. \end{aligned} \tag{4.6}$$
This implies that $h_{(a_1\oplus a_2)\mathbf{a}'} = 2$ and $h_{(b_1\oplus b_2)\mathbf{b}'} = 1$. Thus $h \notin \mathscr{A}$. This is a contradiction.

- If $\bar{a}_1\bar{a}_2\mathbf{a}' \notin \mathrm{supp}(f)$ and $\bar{b}_1\bar{b}_2\mathbf{b}' \in \mathrm{supp}(f)$, then (4.6) implies that $h_{(a_1\oplus a_2)\mathbf{a}'} = 1$ and $h_{(b_1\oplus b_2)\mathbf{b}'} = 2$. Thus $h \notin \mathscr{A}$. This is a contradiction. ∎

The next lemma shows how to reduce the arity of a nonaffine signature with the help of an additional binary signature $[1,0,-1]$ in Pl-Holant$(\widehat{\mathcal{EQ}}, [1,0,-1], \widehat{\mathcal{F}})$, when all signatures in $\widehat{\mathcal{F}}$ take $\{0,1,-1\}$ values.

LEMMA 4.8. *Suppose all signatures in $\widehat{\mathcal{F}}$ take values in $\{0,1,-1\}$ and satisfy the Parity Condition. If $\widehat{\mathcal{F}}$ contains a signature $f \notin \mathscr{A}$ of arity $n \geq 3$, then there exists a signature $h \notin \mathscr{A}$ of arity $< n$ such that*

$$\text{Pl-Holant}(\widehat{\mathcal{EQ}}, h, [1,0,-1], \widehat{\mathcal{F}}) \leq_{\mathrm{T}} \text{Pl-Holant}(\widehat{\mathcal{EQ}}, [1,0,-1], \widehat{\mathcal{F}}).$$

Furthermore, if f satisfies the even Parity Condition, then so does h.

Proof. This proof is a bit more involved; it is also by induction on n and uses the Tableau Calculus.

If there exists $i \in [n]$ such that $f^{x_i=0} \notin \mathscr{A}$, then we are done since we have $[1,0] \in \widehat{\mathcal{EQ}}$. In the following, we assume that $f^{x_i=0} \in \mathscr{A}$ for $1 \leq i \leq n$. By Lemma 2.38, we may assume that $f_{00\cdots0} = 1$ and f satisfies the even Parity Condition.

Claim. If $\mathrm{supp}(f)$ is not an affine subspace, then we can construct a signature that is not in $\mathscr{A}$ and has arity $\leq n-1$.

Suppose $\mathrm{supp}(f)$ is not an affine subspace; then it is not a linear subspace. But $\mathrm{supp}(f^{x_i=0})$ is a linear subspace of $\mathbb{Z}_2^{n-1}$ since $f^{x_i=0} \in \mathscr{A}$ and $f^{x_i=0}_{0\cdots0} = 1$. By Lemma 3.8, there exist $\mathbf{a} = a_1a_2\cdots a_n, \mathbf{b} = b_1b_2\cdots b_n$, such that $\mathbf{a}, \mathbf{b} \in \mathrm{supp}(f)$, $\mathbf{c} = \mathbf{a} \oplus \mathbf{b} = c_1c_2\cdots c_n \notin \mathrm{supp}(f)$, and there exists $i \in [n]$ such that $a_i \neq b_i$. Without loss of generality, we assume that $a_1 = 0, b_1 = 1$. It follows that $c_1 = 1$. Let $\mathbf{a}' = a_3\cdots a_n, \mathbf{b}' = b_3\cdots b_n, \mathbf{c}' = c_3\cdots c_n$.

Connecting the first variable of $[1,0,-1]$ to the first variable of f, the gadget gives the signature

$$\begin{aligned}\bar{f}(x_1, x_2, \ldots, x_n) &= \sum_{x_1' \in \{0,1\}} [1,0,-1](x_1', x_1) f(x_1', x_2, \ldots, x_n) \\ &= (-1)^{x_1} f(x_1, x_2, \ldots, x_n).\end{aligned}$$

Moreover, by connecting the variables x_2, x_1 of $[1,0,1,0]$ to the variables x_1, x_2 of f, respectively, the planar gadget gives the signature $h(x', x_3, x_4, \ldots, x_n)$ satisfying

$$h(x_1 \oplus x_2, x_3, \ldots, x_n) = f(x_1, x_2, x_3, \ldots, x_n) + f(\overline{x_1}, \overline{x_2}, x_3, \ldots, x_n).$$

Similarly, by connecting the variables x_2, x_1 of $[1,0,1,0]$ to the variables x_1, x_2 of $\bar{f}$, respectively, the planar gadget gives the signature $\bar{h}(x', x_3, x_4, \ldots, x_n)$ satisfying

$$\bar{h}(x_1 \oplus x_2, x_3, \ldots, x_n) = (-1)^{x_1} f(x_1, x_2, x_3, \ldots, x_n) + (-1)^{\overline{x_1}} f(\overline{x_1}, \overline{x_2}, x_3, \ldots, x_n).$$

If at least one of $\{h, \bar{h}\}$ is not affine, then we are done since both h and $\bar{h}$ have arity $n-1$. In the following, we assume that both h and $\bar{h}$ are affine.

Next comes the Tableau Calculus.

- For $\bar{a}_1\bar{a}_2\mathbf{a}' \in \mathrm{supp}(f)$, $\bar{b}_1\bar{b}_2\mathbf{b}' \in \mathrm{supp}(f)$, note that $(0,0,\ldots,0) \in \mathrm{supp}(f)$ and $f^{x_1=0} \in \mathscr{A}$. Thus $\mathrm{supp}(f^{x_1=0})$ is a linear subspace of $\mathbb{Z}_2^{n-1}$. By

$$\begin{array}{rll} & a_2\mathbf{a}' & \in \mathrm{supp}(f^{x_i=0}) \\ \oplus & \bar{b}_2\mathbf{b}' & \in \mathrm{supp}(f^{x_i=0}) \\ \hline & \bar{c}_2\mathbf{c}' & \end{array}$$

we have $\bar{c}_2\mathbf{c}' \in \mathrm{supp}(f^{x_i=0})$. This means that $f_{\bar{c}_1\bar{c}_2\mathbf{c}'} \neq 0$ since $\bar{c}_1 = 0$. Note that

$$\begin{aligned} h_{(c_1\oplus c_2)\mathbf{c}'} &= \quad f_{c_1c_2\mathbf{c}'} + f_{\bar{c}_1\bar{c}_2\mathbf{c}'}, \\ \bar{h}_{(c_1\oplus c_2)\mathbf{c}'} &= -f_{c_1c_2\mathbf{c}'} + f_{\bar{c}_1\bar{c}_2\mathbf{c}'}. \end{aligned}$$

Thus $|h_{(c_1\oplus c_2)\mathbf{c}'}| = |\bar{h}_{(c_1\oplus c_2)\mathbf{c}'}| = 1$ since $f_{c_1c_2\mathbf{c}'} = 0$ and $f_{\bar{c}_1\bar{c}_2\mathbf{c}'} = \pm 1$. Moreover, by

$$\begin{aligned} h_{(a_1\oplus a_2)\mathbf{a}'} &= f_{a_1a_2\mathbf{a}'} + f_{\bar{a}_1\bar{a}_2\mathbf{a}'}, \\ \bar{h}_{(a_1\oplus a_2)\mathbf{a}'} &= f_{a_1a_2\mathbf{a}'} - f_{\bar{a}_1\bar{a}_2\mathbf{a}'}, \end{aligned} \tag{4.7}$$

if both of $h_{(a_1\oplus a_2)\mathbf{a}'}$ and $\bar{h}_{(a_1\oplus a_2)\mathbf{a}'}$ are zero, then $f_{a_1a_2\mathbf{a}'} = 0$, by treating (4.7) as a system of linear equations. This contradicts that $\mathbf{a} \in \mathrm{supp}(f)$. Thus we have $|h_{(a_1\oplus a_2)\mathbf{a}'}| \neq 0$ or $|\bar{h}_{(a_1\oplus a_2)\mathbf{a}'}| \neq 0$. This implies that one of $|h_{(a_1\oplus a_2)\mathbf{a}'}|$ or $|\bar{h}_{(a_1\oplus a_2)\mathbf{a}'}|$ is 2 since $f_{a_1a_2\mathbf{a}'}, f_{\bar{a}_1\bar{a}_2\mathbf{a}'} \in \{1,-1\}$. So h or $\bar{h}$ is not affine, because at least one of them has nonzero values of unequal norms. This is a contradiction.

- For $\bar{a}_1\bar{a}_2\mathbf{a}' \notin \mathrm{supp}(f)$, $\bar{b}_1\bar{b}_2\mathbf{b}' \notin \mathrm{supp}(f)$, by treating the following as a linear system:

$$\begin{aligned} h_{00\cdots0} &= f_{000\cdots0} + f_{110\cdots0}, \\ \bar{h}_{00\cdots0} &= f_{000\cdots0} - f_{110\cdots0}, \end{aligned} \tag{4.8}$$

we have $h_{00\cdots0} \neq 0$ or $\bar{h}_{00\cdots0} \neq 0$ since $f_{000\cdots0} \neq 0$. Without loss of generality, we assume that $h_{00\cdots0} \neq 0$. The same argument can be applied to $\bar{h}$ if $\bar{h}_{00\cdots0} \neq 0$. Then $\mathrm{supp}(h)$ is a linear subspace. By

$$h_{(a_1\oplus a_2)\mathbf{a}'} = f_{a_1a_2\mathbf{a}'} + f_{\bar{a}_1\bar{a}_2\mathbf{a}'},$$

$$h_{(b_1\oplus b_2)\mathbf{b}'} = f_{b_1b_2\mathbf{b}'} + f_{\bar{b}_1\bar{b}_2\mathbf{b}'},$$

we have $|h_{(a_1\oplus a_2)\mathbf{a}'}| = 1$, $|h_{(b_1\oplus b_2)\mathbf{b}'}| = 1$ since $f_{\bar{a}_1\bar{a}_2\mathbf{a}'} = f_{\bar{b}_1\bar{b}_2\mathbf{b}'} = 0$ and $f_{a_1a_2\mathbf{a}'}, f_{b_1b_2\mathbf{b}'} \in \{1,-1\}$. This implies that $(a_1\oplus a_2)\mathbf{a}', (b_1\oplus b_2)\mathbf{b}' \in \mathrm{supp}(h)$. Then by

$$\begin{array}{rll} & (a_1\oplus a_2)\mathbf{a}' & \in \mathrm{supp}(h) \\ \oplus & (b_1\oplus b_2)\mathbf{b}' & \in \mathrm{supp}(h) \\ \hline & (c_1\oplus c_2)\mathbf{c}' & \end{array}$$

we have $(c_1\oplus c_2)\mathbf{c}' \in \mathrm{supp}(h)$. This implies that

$$h_{(c_1\oplus c_2)\mathbf{c}'} = f_{c_1c_2\mathbf{c}'} + f_{\bar{c}_1\bar{c}_2\mathbf{c}'} \neq 0.$$

Thus $f_{\bar{c}_1\bar{c}_2\mathbf{c}'} \neq 0$ since $f_{(c_1c_2)\mathbf{c}'} = 0$. So $\bar{c}_1\bar{c}_2\mathbf{c}' \in \mathrm{supp}(f)$. Therefore, $\bar{c}_2\mathbf{c}' \in \mathrm{supp}(f^{x_1=0})$ since $\bar{c}_1 = 0$. By

$$\begin{array}{rll} & a_2\mathbf{a}' & \in \mathrm{supp}(f^{x_1=0}) \\ \oplus & \bar{c}_2\mathbf{c}' & \in \mathrm{supp}(f^{x_1=0}) \\ \hline & \bar{b}_2\mathbf{b}' & \end{array}$$

we have $\bar{b}_2\mathbf{b}' \in \mathrm{supp}(f^{x_1=0})$. Thus $\bar{b}_1\bar{b}_2\mathbf{b}' \in \mathrm{supp}(f)$ as $\bar{b}_1 = 0$. This is a contradiction.

- If $\bar{a}_1\bar{a}_2\mathbf{a}' \in \mathrm{supp}(f)$, $\bar{b}_1\bar{b}_2\mathbf{b}' \notin \mathrm{supp}(f)$, by
$$h_{(b_1\oplus b_2)\mathbf{b}'} = \quad f_{b_1b_2\mathbf{b}'} + f_{\bar{b}_1\bar{b}_2\mathbf{b}'},$$
$$\bar{h}_{(b_1\oplus b_2)\mathbf{b}'} = -f_{b_1b_2\mathbf{b}'} + f_{\bar{b}_1\bar{b}_2\mathbf{b}'},$$
we have $|h_{(b_1\oplus b_2)\mathrm{b}'}| = |\bar{h}_{(b_1\oplus b_2)\mathbf{b}'}| = 1$ since $f_{b_1b_2\mathbf{b}'} = \pm 1$ and $f_{\bar{b}_1\bar{b}_2\mathbf{b}'} = 0$. Then by
$$h_{(a_1\oplus a_2)\mathbf{a}'} = f_{a_1a_2\mathbf{a}'} + f_{\bar{a}_1\bar{a}_2\mathbf{a}'},$$
$$\bar{h}_{(a_1\oplus a_2)\mathbf{a}'} = f_{a_1a_2\mathbf{a}'} - f_{\bar{a}_1\bar{a}_2\mathbf{a}'},$$
and $f_{a_1a_2\mathbf{a}'}, f_{\bar{a}_1\bar{a}_2\mathbf{a}'} \in \{1,-1\}$, we have $|h_{(a_1\oplus a_2)\mathbf{a}'}| = 2$ or $|\bar{h}_{(a_1\oplus a_2)\mathbf{a}'}| = 2$. This implies that h or $\bar{h}$ is not affine. This is a contradiction.
- If $\bar{a}_1\bar{a}_2\mathbf{a}' \notin \mathrm{supp}(f)$, $\bar{b}_1\bar{b}_2\mathbf{b}' \in \mathrm{supp}(f)$, the proof is symmetric by reversing the order of $\mathbf{a}$ and $\mathbf{b}$ in the previous item.

This completes the proof of the claim.

Now we can assume that $\mathrm{supp}(f)$ is affine with dimension k. Let $Y = \{y_1, y_2, \ldots, y_k\}$ be a set of free variables, where $Y \subseteq \{x_1, x_2, \ldots, x_n\}$, and let $\underline{f}$ be the compressed signature of f for Y. Since the variable names $\{x_1, x_2, \ldots, x_n\}$ of f can be cyclically permuted (without violating planarity), we may assume that $y_1 = x_1$. If $k \le 2$, then $\underline{f}$ is affine by Lemma 2.15. So f is affine by Corollary 2.14. This is a contradiction.

In the following, we assume that $k \ge 3$. By Lemma 2.12, and since $\underline{f}$ takes values in $\{1,-1\}$, there exists a unique multilinear polynomial $Q(y_1, y_2, \ldots, y_k) \in \mathbb{Z}_2[Y]$ such that $\underline{f}(y_1, y_2, \ldots, y_k) = (-1)^{Q(y_1,y_2,\ldots,y_k)}$.

Note that if $Q(y_1, y_2, \ldots, y_k)$ is a quadratic multilinear polynomial, then $\underline{f}$ is affine, and so is f. This is a contradiction. Thus we can assume that there exists at least one term of degree greater than 2 in Q in the following. (Recall that for $\{1,-1\}$-valued functions, there is no requirement on cross-term coefficients being even for the quadratic multilinear polynomial.)

If $k \ge 4$ and there exists a term $y_{i_1}y_{i_2}\cdots y_{i_s}$ with nonzero coefficient, where $3 \le s < k$, then there exists some $y_j \notin \{y_{i_1}, y_{i_2}, \ldots, y_{i_s}\}$, such that
$$\underline{f^{y_j=0}}(y_1, \ldots, y_{j-1}, y_{j+1}, \ldots, y_k) = (-1)^{Q'(y_1,\ldots,y_{j-1},y_{j+1},\ldots,y_k)},$$
where Q' is a polynomial on $k-1$ variables, where the term $y_{i_1}y_{i_2}\cdots y_{i_s}$ of degree $s > 2$ still appears. This implies that $f^{y_j=0}$ is not affine. This is a contradiction. Thus we may assume that $Q(y_1, y_2, \ldots, y_k) = P(y_1, y_2, \ldots, y_k) + ay_1y_2\cdots y_k$, where $P(y_1, y_2, \ldots, y_k) \in \mathbb{Z}_2[Y]$ is a multilinear polynomial of total degree at most 2, and $a \in \mathbb{Z}_2$. Note that this statement is also vacuously true if $k = 3$. If $a = 0$, then f is affine. This is a contradiction. Otherwise, $Q(y_1, y_2, \ldots, y_k) = P(y_1, y_2, \ldots, y_k) + y_1y_2\cdots y_k$. Moreover, by connecting the first variable of $[1, 0, -1]$ to y_i of f, the gadget gives the signature f' such that
$$f'(x_1, x_2, \ldots, x_n) = (-1)^{y_i} f(x_1, x_2, \ldots, x_n).$$
This implies that f' has the same support of f and
$$\underline{f'}(y_1, y_2, \ldots, y_k) = (-1)^{y_i + P(y_1,y_2,\ldots,y_k) + y_1y_2\cdots y_k},$$
where $\underline{f'}$ is the compressed signature of f' for Y. Thus $f' \notin \mathscr{A}$. This implies that we can add a linear term to $P(y_1, y_2, \ldots, y_k)$ freely.

In the following, we connect all variables of f except for y_1 to $n-1$ variables of the signature $\frac{1}{2}\{[1,1]^{\otimes n}+[1,-1]^{\otimes n}\} = [1,0,1,\ldots,0\ (\text{or}\ 1)] \in \widehat{\mathcal{EQ}}$ to get a binary signature f^*. If the input to f^* is 00, we get a sum of all values of $f(0,x_2,\ldots,x_n)$ where $\mathrm{wt}(x_2,\ldots,x_n)$ is even. By the even Parity Condition of f, this is the sum of all values of $f^{x_1=0}$, which is the sum of all values of $\underline{f}^{y_1=0}$. If the input to f^* is 11, we get a sum of all values of $f(1,x_2,\ldots,x_n)$ where $\mathrm{wt}(x_2,\ldots,x_n)$ is odd. Again by the even Parity Condition of f, this is the sum of all values of $f^{x_1=1}$, which is the sum of all values of $\underline{f}^{y_1=1}$. Finally, if the input to f^* is 01 or 10, we get a sum of all values of $f(0,x_2,\ldots,x_n)$ where $\mathrm{wt}(x_2,\ldots,x_n)$ is odd, or all values of $f(1,x_2,\ldots,x_n)$ where $\mathrm{wt}(x_2,\ldots,x_n)$ is even, which is 0 in either case. Thus, the binary f^* is symmetric and has the signature $[f^*_{00},0,f^*_{11}]$, where

$$
\begin{aligned}
f^*_{00} &= \sum_{y_2,y_3,\ldots,y_k\in\{0,1\}} \underline{f}^{y_1=0}(y_2,y_3,\ldots,y_k),\\
f^*_{11} &= \sum_{y_2,y_3,\ldots,y_k\in\{0,1\}} \underline{f}^{y_1=1}(y_2,y_3,\ldots,y_k).
\end{aligned}
\tag{4.9}
$$

First, we consider the special case that the coefficient of y_iy_j in $P(y_1,y_2,\ldots,y_k)$ is nonzero for all $1\le i<j\le k$.

- If $k=3$, we may assume that $P(y_1,y_2,y_3)=y_1+y_2+y_3+y_1y_2+y_1y_3+y_2y_3$ since we can add linear terms to $P(y_1,y_2,\ldots,y_k)$ at will. Then we have $\underline{f}(y_1,y_2,y_3)=(-1)^{P(y_1,y_2,y_3)+y_1y_2y_3}$. The polynomial $P(y_1,y_2,y_3)+y_1y_2y_3 = \overline{1}+(1+y_1)(1+y_2)(1+y_3)\in\mathbb{Z}_2[y_1,y_2,y_3]$ corresponds to the OR function on 3 bits, $y_1\vee y_2\vee y_3$. Thus $M_{y_1,y_2y_3}(\underline{f}) = \left[\begin{smallmatrix}1 & -1 & -1 & -1\\ -1 & -1 & -1 & -1\end{smallmatrix}\right]$. Thus $f^*=[-2,0,-4]$, which has nonzero terms of unequal norms, thus not in $\mathscr{A}$, and we are done.
- For $k\ge 4$, we may assume that $P(y_1,y_2,\ldots,y_k)$ has no linear terms since we can add linear terms to $P(y_1,y_2,\ldots,y_k)$ freely. Since P has all terms y_iy_j, both $\underline{f}^{y_1=0}$ and $\underline{f}^{y_1=1}$ are symmetric signatures. For $\underline{f}^{y_1=0}$, the entry of Hamming weight ℓ is $(\underline{f}^{y_1=0})_\ell=(-1)^{\frac{\ell(\ell-1)}{2}}$ for $0\le\ell\le k-1$. For $\underline{f}^{y_1=1}$, we have $(\underline{f}^{y_1=1})_\ell=(-1)^{\frac{\ell(\ell+1)}{2}}$ for $0\le\ell\le k-2$ and $(\underline{f}^{y_1=1})_{k-1}=(-1)^{\frac{k(k-1)}{2}+1}$. This implies that

$$
\begin{aligned}
\underline{f}^{y_1=0} &= [1,1,-1,-1,\ldots,(-1)^{\frac{(k-1)(k-2)}{2}}]\\
&= \frac{1}{1+\mathrm{i}}\left\{[1,\mathrm{i}]^{\otimes k-1}+\mathrm{i}[1,-\mathrm{i}]^{\otimes k-1}\right\},\\
\underline{f}^{y_1=1} &= [1,-1,-1,1,\ldots,(-1)^{\frac{k(k-1)}{2}}]-2(-1)^{\frac{k(k-1)}{2}}[0,1]^{\otimes k-1}\\
&= \frac{1}{1-\mathrm{i}}\left\{[1,\mathrm{i}]^{\otimes k-1}-\mathrm{i}[1,-\mathrm{i}]^{\otimes k-1}\right\}-2(-1)^{\frac{k(k-1)}{2}}[0,1]^{\otimes k-1}.
\end{aligned}
$$

Thus

$$\begin{aligned}
f^*_{00} &= \sum_{\beta\in\{0,1\}^{k-1}} (\underline{f}^{y_1=0})_\beta \\
&= \frac{1}{1+\mathtt{i}} \sum_{w=0}^{k-1} \binom{k-1}{w} [\mathtt{i}^w + \mathtt{i}(-\mathtt{i})^w] \\
&= \frac{1}{1+\mathtt{i}} [(1+\mathtt{i})^{k-1} + \mathtt{i}(1-\mathtt{i})^{k-1}] \\
&= (1+\mathtt{i})^{k-2} + (1-\mathtt{i})^{k-2} \\
&= 2^{\frac{k}{2}} \cos((k-2)\pi/4), \\
f^*_{11} &= \sum_{\beta\in\{0,1\}^{k-1}} (\underline{f}^{y_1=1})_\beta \\
&= \frac{1}{1-\mathtt{i}} \sum_{w=0}^{k-1} \binom{k-1}{w} [\mathtt{i}^w - \mathtt{i}(-\mathtt{i})^w] - 2(-1)^{\frac{k(k-1)}{2}} \\
&= \frac{1}{1-\mathtt{i}} [(1+\mathtt{i})^{k-1} - \mathtt{i}(1-\mathtt{i})^{k-1}] - 2(-1)^{\frac{k(k-1)}{2}} \\
&= -2^{\frac{k}{2}} \sin((k-2)\pi/4) - 2(-1)^{\frac{k(k-1)}{2}}.
\end{aligned}$$

For $k \equiv 1 \mod 2$, $|f^*_{00}| = 2^{\frac{k-1}{2}}$, and $|f^*_{11}| = 2^{\frac{k-1}{2}} \pm 2$ (since $k \geq 4$), we have $f^*_{11} f^*_{00} \neq 0$ and $|f^*_{11}| \neq |f^*_{00}|$. Thus $f^* \notin \mathscr{A}$ and we are done.
For $k \equiv 0 \mod 4$, $f^*_{00} = 0$, $|f^*_{11}| = 2^{\frac{k}{2}} \pm 2 \neq 0$. This implies that we have $f^* = f^*_{11}[0,1]^{\otimes 2}$. By $[1,0,-1]$, $[0,1]^{\otimes 2}$ and $f \notin \mathscr{A}$, we can get a binary signature that is not in $\mathscr{A}$ by Lemma 4.6. Thus we are done.
For $k \equiv 2 \mod 4$, $|f^*_{00}| = 2^{\frac{k}{2}} \geq 4$ since $k \geq 4$, and $|f^*_{11}| = 2$, and so $f^* \notin \mathscr{A}$ and we are done.

Now we assume that there exist $i \neq j \in [k]$ such that the coefficient of $y_i y_j$ is 0 in $P(y_1, y_2, \ldots, y_k)$. For notational simplicity, without loss of generality we assume that $i = k-1, j = k$. Then we can assume that (with the linear term y_{k-1} and y_k removed if needed)

$$P(y_1, y_2, \ldots, y_k) = y_1(L_1 + \epsilon_1) + y_2(L_2 + \epsilon_2) + \cdots + y_{k-2}(L_{k-2} + \varepsilon_{k-2}),$$

where

$$L_1 = \sum_{i=2}^{k} a_{1i} y_i, \quad L_2 = \sum_{i=3}^{k} a_{2i} y_i, \quad \ldots, \quad L_{k-2} = \sum_{i=k-1}^{k} a_{(k-2)i} y_i,$$

and $a_{ji} \in \mathbb{Z}_2$ are fixed, but we can choose $\epsilon_i \in \mathbb{Z}_2$ freely since we can add linear terms freely. Let $F_{(0)} = \underline{f}$, $F_{(i)} = \underline{f}^{y_1=0, y_2=0, \ldots, y_i=0}$ for $i \in [k-2]$.

Claim. There exist $\epsilon_1, \epsilon_2, \ldots, \epsilon_{k-2} \in \mathbb{Z}_2$ such that

$$\sum_{y_{i+1}, y_{i+2}, \ldots, y_k \in \{0,1\}} F_{(i)}(y_{i+1}, y_{i+2}, \ldots, y_k) \geq 4$$

for all $1 \leq i \leq k-2$.

We prove this claim by induction. The base case is for $F_{(k-2)}$. Note that $P(0,\dots,0,y_{k-1},y_k)$ is identically 0. Thus

$$\sum_{y_{k-1},y_k\in\{0,1\}} F_{(k-2)}(y_{k-1},y_k)=4.$$

By induction, we may assume that

$$\sum_{y_{i+1},y_{i+2},\dots,y_k\in\{0,1\}} F_{(i)}(y_{i+1},y_{i+2},\dots,y_k)\geq 4$$

and prove that

$$\sum_{y_i,y_{i+1},\dots,y_k\in\{0,1\}} F_{(i-1)}(y_i,y_{i+1},\dots,y_k)\geq 4$$

for $i\leq k-2$. Note that

$$\begin{aligned} F_{(i-1)}(1,y_{i+1},y_{i+2},\dots,y_k)&=(-1)^{(L_i+\epsilon_i)+y_{i+1}(L_{i+1}+\epsilon_{i+1})+\cdots+y_{k-2}(L_{k-2}+\epsilon_{k-2})},\\ F_{(i-1)}(0,y_{i+1},y_{i+2},\dots,y_k)&=(-1)^{y_{i+1}(L_{i+1}+\epsilon_{i+1})+\cdots+y_{k-2}(L_{k-2}+\epsilon_{k-2})}. \end{aligned} \tag{4.10}$$

By the inductive hypothesis,

$$\sum_{y_{i+1},\dots,y_k\in\{0,1\}} F_{(i-1)}(0,y_{i+1},\dots,y_k)=\sum_{y_{i+1},\dots,y_k\in\{0,1\}} F_{(i)}(y_{i+1},\dots,y_k)\geq 4.$$

If L_i is identically 0, then we set $\epsilon_i=0$. It follows that

$$\sum_{y_{i+1},\dots,y_k\in\{0,1\}} F_{(i-1)}(1,y_{i+1},\dots,y_k)=\sum_{y_{i+1},\dots,y_k\in\{0,1\}} F_{(i-1)}(0,y_{i+1},\dots,y_k).$$

Thus we have

$$\sum_{y_i,y_{i+1},\dots,y_k\in\{0,1\}} F_{(i-1)}(y_i,y_{i+1},\dots,y_k)=2\sum_{y_{i+1},\dots,y_k\in\{0,1\}} F_{(i-1)}(0,y_{i+1},\dots,y_k)\geq 8,$$

and we are done. Otherwise, $L_i=0$ defines a subspace V of $\mathbb{Z}_2^{k-i}$ that has dimension $k-i-1$. Let $V'=\mathbb{Z}_2^{k-i}\setminus V$; then V' is an affine space defined by $L_i=1$ which has dimension $k-i-1$. Let

$$a=\sum_{y_{i+1}y_{i+2}\cdots y_k\in V} F_{(i-1)}(0,y_{i+1},y_{i+2},\dots,y_k),$$

$$a'=\sum_{y_{i+1}y_{i+2}\cdots y_k\in V'} F_{(i-1)}(0,y_{i+1},y_{i+2},\dots,y_k),$$

$$b=\sum_{y_{i+1}y_{i+2}\cdots y_k\in V} F_{(i-1)}(1,y_{i+1},y_{i+2},\dots,y_k),$$

and

$$b'=\sum_{y_{i+1}y_{i+2}\cdots y_k\in V'} F_{(i-1)}(1,y_{i+1},y_{i+2},\dots,y_k).$$

Then

$$\sum_{y_i,y_{i+1},\dots,y_k\in\{0,1\}} F_{(i-1)}(y_i,y_{i+1},\dots,y_k)=a+a'+b+b'.$$

By induction, we have $a + a' \geq 4$. Thus $a \geq 2$ or $a' \geq 2$. If $a \geq 2$, we choose $\epsilon_i = 0$; then $a = b$, $a' = -b'$ by (4.10). Thus $a + b + a' + b' = 2a \geq 4$. If $a' \geq 2$, we choose $\epsilon_i = 1$; then $a = -b$, $a' = b'$ by (4.10). Thus $a + b + a' + b' = 2a' \geq 4$.

This completes the proof of the claim.

The claim shows that

$$f^*_{00} = \sum_{y_2,y_3,\ldots,y_k \in \{0,1\}} \underline{f}^{y_1=0}(y_2, y_3, \ldots, y_k) = \sum_{y_2,y_3,\ldots,y_k \in \{0,1\}} F_{(1)}(y_2, y_3, \ldots, y_k) \geq 4.$$

Let g be the n-ary signature with the same support as f (thus it satisfies the even Parity Condition), and on its support

$$g(x_1, x_2, \ldots, x_n) = (-1)^{P(y_1,y_2,\ldots,y_k)};$$

then $\underline{f}_\beta = \underline{g}_\beta$ for any $\beta \in \{0,1\}^k$ other than $\beta = 11\cdots1$. For $\beta = 11\cdots1$,

$$\underline{f}_\beta = (-1)^{P(1,1,\ldots,1)+1\cdot1\cdots1} = -(-1)^{P(1,1,\ldots,1)} = -\underline{g}_\beta.$$

This implies that

$$\underline{f} = \underline{g} \pm 2[0,1]^{\otimes k}. \tag{4.11}$$

(Note that we do not really construct g. We just use g to argue that $f^* \notin \mathscr{A}$.) Since both $\frac{1}{2}\{[1,1]^{\otimes n} + [1,-1]^{\otimes n}\}$ and g are affine signatures, the following construction would produce an affine signature: Connect all variables of g other than y_1 to $n-1$ variables of $\frac{1}{2}\{[1,1]^{\otimes n} + [1,-1]^{\otimes n}\} = [1,0,1,\ldots,0\ (\text{or}\ 1)]$. This construction gives a binary signature $g^* = [g^*_{00}, 0, g^*_{11}]$. (By the even Parity Condition, the weight 1 entry must be 0.) Note that

$$g^*_{00} = \sum_{y_2,y_3,\ldots,y_k \in \{0,1\}} \underline{g}^{y_1=0}(y_2, y_3, \ldots, y_k),$$

$$g^*_{11} = \sum_{y_2,y_3,\ldots,y_k \in \{0,1\}} \underline{g}^{y_1=1}(y_2, y_3, \ldots, y_k).$$

Thus by (4.9) and (4.11), we have

$$f^*_{00} = g^*_{00}, \quad f^*_{11} = g^*_{11} \pm 2.$$

Since g^* is an affine signature, we must have either $g^*_{00} = 0$ or $g^*_{11} = 0$ or $(g^*_{00})^4 = (g^*_{11})^4$. Since we have $g^*_{00} = f^*_{00} \neq 0$ and both g^*_{00} and g^*_{11} are real numbers, we must have $g^*_{11} = 0$ or $g^*_{11} = \pm g^*_{00}$. Recall that $f^*_{00} \geq 4$. If $g^*_{11} = 0$, then $f^*_{11} = \pm 2$ has a different nonzero norm than f^*_{00}. If $g^*_{11} = \pm g^*_{00}$, then $g^*_{11} = \pm f^*_{00}$ has norm at least 4, and thus $f^*_{11} = g^*_{11} \pm 2$ has norm $|g^*_{11}| \pm 2 = |f^*_{00}| \pm 2$. And so in this case f^*_{11} also has a different nonzero norm than f^*_{00}. In each case, $|f^*_{00}| \neq |f^*_{11}|$ and $f^*_{00} f^*_{11} \neq 0$. This implies that $f^* \notin \mathscr{A}$, and we are done. ∎

Now we give the main dichotomy theorem of this section. By (2.2), we have

$$\text{Pl-\#CSP}^2(\widehat{\mathcal{EQ}}, \widehat{\mathcal{F}}) \equiv_T \text{Pl-Holant}(\mathcal{EQ}_2, \widehat{\mathcal{EQ}}, \widehat{\mathcal{F}}).$$

We will prove a dichotomy theorem for Pl-\#CSP$^2(\widehat{\mathcal{EQ}}, \widehat{\mathcal{F}})$ when every signature in $\widehat{\mathcal{F}}$ satisfies the Parity Condition.

THEOREM 4.9. *If all signatures in $\widehat{\mathcal{F}}$ satisfy the Parity Condition, then* Pl-Holant$(\mathcal{EQ}_2, \widehat{\mathcal{EQ}}, \widehat{\mathcal{F}})$ *(equivalently* Pl-#CSP$^2(\widehat{\mathcal{EQ}}, \widehat{\mathcal{F}})$*) is* #P*-hard, or $\widehat{\mathcal{F}} \subseteq \mathscr{A}$, in which case the problem is in P.*

Proof. If $\widehat{\mathcal{F}} \subseteq \mathscr{A}$, then Pl-Holant$(\mathcal{EQ}_2, \widehat{\mathcal{EQ}}, \widehat{\mathcal{F}})$ is tractable by Theorem 2.33′, since $\widehat{\mathcal{EQ}} \subset \mathscr{A}$ as well.

If $\widehat{\mathcal{F}} \not\subseteq \mathscr{A}$, then there exists an n-ary signature $f \in \widehat{\mathcal{F}} \setminus \mathscr{A}$. By Lemma 2.38, we can assume that $f_{00\cdots0} = 1$ and f satisfies the even Parity Condition. If f is a unary signature, then $f = [1, 0] \in \mathscr{A}$, a contradiction. If f has arity 2, then f must be symmetric and has the form $f = [1, 0, x]$, where $x^4 \neq 0, 1$, by Proposition 2.17. Note that $\widehat{\mathcal{EQ}}$ contains the symmetric signatures $[1, 0]$ and $[1, 0, 1, 0]$, and $[1, 0] \notin \widehat{\mathscr{M}} \cup \widehat{\mathscr{M}^\dagger}$, $[1, 0, 1, 0] \notin \mathscr{P} \cup \mathscr{A}^\dagger$ by Proposition 2.32 and Corollary 2.21. By Theorem 2.35, Pl-Holant$(\mathcal{EQ}_2, [1, 0], [1, 0, 1, 0], [1, 0, x])$ is #P-hard. Thus Pl-Holant$(\mathcal{EQ}_2, \widehat{\mathcal{EQ}}, \widehat{\mathcal{F}})$ is #P-hard since

$$\text{Pl-Holant}(\mathcal{EQ}_2, [1, 0], [1, 0, x], [1, 0, 1, 0]) \leq_{\mathrm{T}} \text{Pl-Holant}(\mathcal{EQ}_2, \widehat{\mathcal{EQ}}, \widehat{\mathcal{F}}).$$

So in the following, we assume that the arity of f is $n \geq 3$.

(A) If there exists $\alpha \in \{0, 1\}^n$ such that $f_\alpha^4 \neq 0, 1$, then we can get $[1, 0, f_\alpha]$ in the following way: first, using $[1, 0] \in \widehat{\mathcal{EQ}}$, we can get $\partial^S_{[1,0]}(f) = (1, \ldots, f_\alpha)$, where $S = \{k \mid \text{the } k\text{th bit of } \alpha \text{ is } 0\}$. Note that the arity of $(1, \ldots, f_\alpha)$ is $\mathrm{wt}(\alpha)$, which is even by the even Parity Condition, and we have $(=_{\mathrm{wt}(\alpha)+2}) \in \mathcal{EQ}_2$. So we have

$$\partial_{(1,\ldots,f_\alpha)}(=_{\mathrm{wt}(\alpha)+2}) = [1, 0, f_\alpha] \notin \mathscr{A}.$$

Then by Theorem 2.35, Pl-Holant$(\mathcal{EQ}_2, [1, 0], [1, 0, 1, 0], [1, 0, f_\alpha])$ is #P-hard. It follows that Pl-Holant$(\mathcal{EQ}_2, \widehat{\mathcal{EQ}}, \widehat{\mathcal{F}})$ is #P-hard.

Now we may assume that all nonzero entries of f are powers of i.

(B) If there exists $\alpha \in \{0, 1\}^n$ such that $f_\alpha = \pm\mathrm{i}$, then by the same argument as item (A), we have $[1, 0, \mathrm{i}]$ or $[1, 0, -\mathrm{i}]$. In each case, we have $[1, 0, -1]$ and $[0, 1]^{\otimes 2}$ by Lemma 4.1. Then by f and Lemma 4.6, we can get $[1, 0, x] \notin \mathscr{A}$. Then we are done by Theorem 2.35.

Now we may assume that all nonzero entries of f are 1 or -1.

(C) If f takes values in $\{0, 1, -1\}$ and there exists at least one $\alpha \in \{0, 1\}^n$ such that $f_\alpha = -1$, then we can get $[1, 0, -1]$ in the same way as (A). Now we prove the lemma by induction on the arity $n \geq 3$ of f.

For $n = 3$, by Lemma 4.8, we can get a signature $g \notin \mathscr{A}$ with arity < 3. Note that g also satisfies the even Parity Condition. If g has arity 1, then $g \in \mathscr{A}$. This is a contradiction. If g has arity 2, then it must be of the form $[x, 0, y]$, and $xy \neq 0, (x/y)^4 \neq 1$ lest $g \in \mathscr{A}$. In particular, g is symmetric. Then, by Theorem 2.35, Pl-Holant$(\mathcal{EQ}_2, [1, 0], [1, 0, 1, 0], g)$ is #P-hard. Thus Pl-Holant$(\mathcal{EQ}_2, \widehat{\mathcal{EQ}}, \widehat{\mathcal{F}})$ is #P-hard.

For $n \geq 4$, by Lemma 4.8, we can get a signature $g \notin \mathscr{A}$ with arity $< n$. If g takes values in $\{0, 1, -1\}$, up to a nonzero factor, then we are done by induction. Otherwise, we are done by items (A) and (B).

Now we may assume that f takes values in $\{0, 1\}$.

(D) For a $\{0, 1\}$-valued signature satisfying the even Parity Condition, if it has arity ≤ 2, then it is affine. Hence f has arity≥ 3 since $f \notin \mathscr{A}$. We induct on the arity $n \geq 3$ of f in this case.

For $n = 3$, by Lemma 4.7, we can get a nonaffine signature h with arity < 3. Note that h satisfies the even Parity Condition. If h has arity

1, then $h \in \mathscr{A}$. This is a contradiction. If h has arity 2, then h has the form $[1, 0, x] \notin \mathscr{A}$ up to a nonzero factor. Then by Theorem 2.35, Pl-Holant$(\mathcal{EQ}_2, [1, 0], [1, 0, 1, 0], [1, 0, x])$ is #P-hard. Thus Pl-Holant$(\mathcal{EQ}_2, \widehat{\mathcal{EQ}}, \widehat{\mathcal{F}})$ is #P-hard.

For $n \geq 4$, by Lemma 4.7, we can get a nonaffine signature h with arity $< n$. If h takes values in $\{0, 1\}$, then we are done by induction. Otherwise, we are done by items (A), (B), and (C). ∎

5. When $\widehat{\mathcal{F}}$ satisfies parity. In this section, we give a dichotomy for Pl-Holant$(\widehat{\mathcal{EQ}}, \widehat{\mathcal{F}})$, where all signatures in $\widehat{\mathcal{F}}$ satisfy the Parity Condition. In this case, $\widehat{\mathcal{F}}$ will involve matchgate signatures. If $\widehat{\mathcal{F}} \subseteq \mathcal{M}$, then the problem is tractable. Assume $\widehat{\mathcal{F}} \not\subseteq \mathcal{M}$. General matchgate signatures are governed by the matchgate identities (MGIs). For asymmetric signatures of high arities, these are intricate and difficult to handle. So we first try to reduce the arity of a nonmatchgate signature.

5.1. Arity reduction of nonmatchgate signatures. The following lemma follows from the general theory of matchgates [7] (see also [4]). Recall that for length n, we use e_i to denote the string whose ith bit is 1, and all other bits are 0.

LEMMA 5.1. *For any signature f of arity $n \geq 2$ with $f_{00\cdots0} = 1$, there exists a matchgate signature g of arity n such that $g_{00\cdots0} = f_{00\cdots0} = 1$ and $g_{00\cdots0\oplus e_i\oplus e_j} = f_{00\cdots0\oplus e_i\oplus e_j}$ for all $i, j \in [n]$ and $i < j$.*

Proof. Let $m_{ij} = f_{00\cdots0\oplus e_i\oplus e_j}$ for all $i, j \in [n]$ and $i < j$. Flipping all bits from 0 to 1, we construct a planar matchgate Γ with n external nodes such that its signature g satisfies $g_{11\cdots1} = 1$ and $g_{11\cdots1\oplus e_i\oplus e_j} = m_{ij}$ for all $i, j \in [n]$ and $i < j$. To obtain the desired matchgate stated in the lemma, for every external node v_i of Γ, we append a weight 1 edge (u_i, v_i) and make u_i the new external node (for $1 \leq i \leq n$).

If $n = 2$, we can just take a single edge (v_1, v_2) with weight m_{12}. This matchgate has signature $g_{11} = 1$ and $g_{00} = m_{12}$. (If one prefers not to consider PerfMatch(G) for an empty graph G, which this single edge matchgate with both v_1 and v_2 removed would be, we can add an extra isolated edge with weight 1 to the matchgate.) Below we assume $n \geq 4$.

Let K_n be the complete graph on n nodes. We place the nodes of K_n, labeled $1 < 2 < \cdots < n$, clockwise on a lower semicircle, as illustrated in Figure 12. The n nodes are placed in a general position, so that any pair of crossing edges intersect at a unique point. There are exactly $\binom{n}{4}$ such intersection points. We assign weight m_{ij} to the edge $\{i, j\}$ for all $1 \leq i < j \leq n$. Note that if we remove all nodes except i and j, there is a single edge left with weight m_{ij}. Thus this weighted K_n would satisfy the lemma, except that it is not planar.

Now we construct a planar matchgate Γ by the use of a *crossover gadget* in Figure 13. The crossover gadget is itself a matchgate X with the following signature:

$$X^{0000} = 1, \quad X^{0101} = 1, \quad X^{1010} = 1, \quad X^{1111} = -1,$$

and for all other $\beta \in \{0, 1\}^4$, $X^\beta = 0$. We note that even though geometrically this gadget is only symmetric under a rotation of π (but not $\pi/2$), its signature is invariant under a cyclic permutation, and thus functionally it is symmetric under a rotation of $\pi/2$. The support of this matchgate X requires that the alternate pairs of four inputs are equal (i.e., $x_1 = x_3$ and $x_2 = x_4$), and each pair can be independently both 0 or both 1 (when all four inputs are 1, the value $X^{1111} = -1$; but this last property will not be used in the proof here). Our matchgate Γ is obtained by replacing every

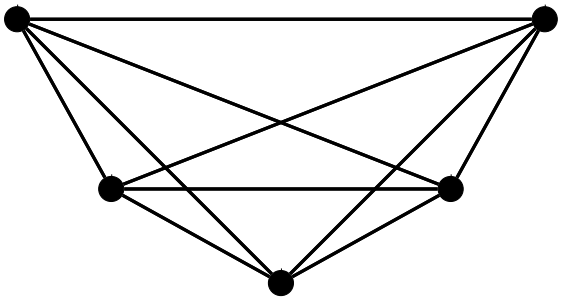


FIG. 12. *The embedding for* K_5.

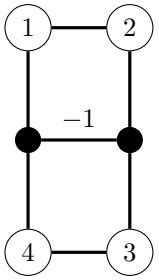


FIG. 13. *The crossover gadget. The external nodes are those labeled, and all edge weights are* 1 *except the edge labeled* -1.

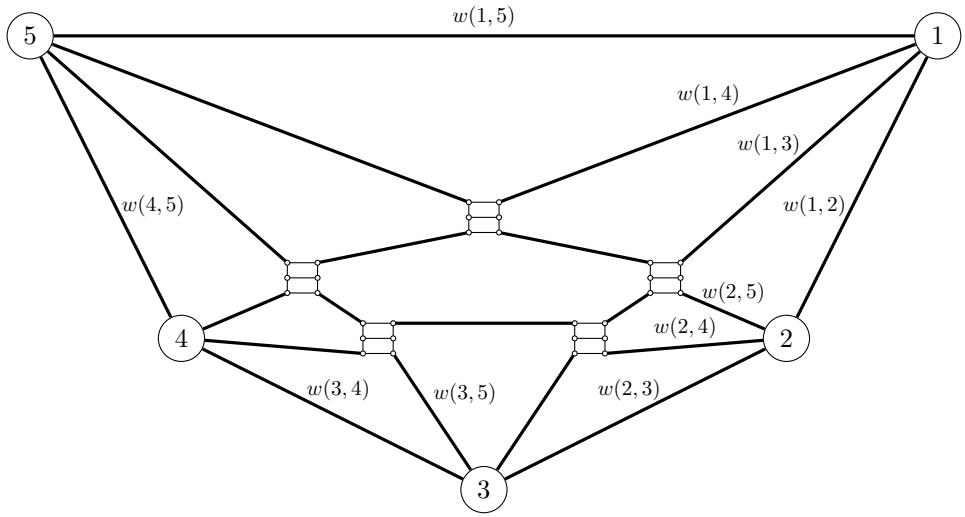


FIG. 14. *The "planarized"* K_5 *with edge weights. The unlabeled edges have weight* 1. *For notational simplicity, in the figure we use the notation* $w(i,j)$ *for* $w(\{i,j\})$.

crossing of a pair of edges of K_n by a copy of X. If $\{i,j\}$ in K_n crosses $t \geq 0$ other edges, then there are t copies of X which break $\{i,j\}$ into $t+1$ segment edges outside of these crossover gadgets. All these segment edges have weight 1 except one which has weight m_{ij}. This defines the planar matchgate Γ. For $n = 5$, this is illustrated in Figure 14.

Suppose an external node i is removed in Γ (i.e., the ith bit is set to 1 in the signature of Γ). Consider any other external node j and the replacement part in Γ for the edge $\{i,j\}$ in K_n. Suppose t copies of X are used along that. By properties of the signature of X ($x_1 = x_3$ and $x_2 = x_4$), we may assume that all $t+1$ segment edges are not contained in perfect matchings, the sum over which defines the signature entry for Γ. In particular, all t copies of X contribute a factor 1, regardless of the bit assignment for any bit other than i. (So, in effect, when the ith bit is set to 1 in Γ, we can imagine that all edges in K_n incident to node i are removed.)

Suppose we remove all external nodes of Γ except $\{i,j\}$. Then we only need to consider perfect matchings involving the replacement part in Γ for the edge $\{i,j\}$ in K_n. Again suppose t copies of X are used in Γ for the edge $\{i,j\}$ in K_n. Then it is easy to see that, by the property of X, we may assume all $t+1$ segment edges outside of these X are included in the perfect matchings. These edges contribute a value m_{ij} (one edge has weight m_{ij}, and the other t edges have weight 1), and each X also contributes a factor 1. This completes the proof. ∎

THEOREM 5.2. *If all signatures in $\widehat{\mathcal{F}}$ satisfy the Parity Condition, and $\widehat{\mathcal{F}} \not\subseteq \mathscr{M}$, then there exists $h \notin \mathscr{M}$ of arity 4 such that*

$$\text{Pl-Holant}(h, \widehat{\mathcal{EQ}}, \widehat{\mathcal{F}}) \leq_{\text{T}} \text{Pl-Holant}(\widehat{\mathcal{EQ}}, \widehat{\mathcal{F}}).$$

Proof. Since $\widehat{\mathcal{F}} \not\subseteq \mathscr{M}$, there exists $f \in \widehat{\mathcal{F}} \setminus \mathscr{M}$. By Lemma 2.29, f has arity $n \geq 4$ since f satisfies the Parity Condition. Moreover, by Lemma 2.38, we can assume that $f_{00\cdots0} = 1$ and f satisfies the even Parity Condition. By Lemma 5.1, there exists $g \in \mathscr{M}$ such that $g_{00\cdots0} = 1$ and $g_{00\cdots0\oplus e_i\oplus e_j} = f_{00\cdots0\oplus e_i\oplus e_j}$ for any $i,j \in [n]$ and $i < j$.

We will prove the theorem by induction on n. If $n = 4$, then we are done. Now we assume that the theorem is true for arity $\leq n-1$ and prove the theorem for $n \geq 5$. If there exists $i \in [n]$ such that $f^{x_i=0} \notin \mathscr{M}$, then we are done by induction since we have $[1,0] \in \widehat{\mathcal{EQ}}$. Therefore, we may assume that $f^{x_i=0} \in \mathscr{M}$ for $1 \leq i \leq n$.

Claim. We have $f_\alpha = g_\alpha$ for all $\alpha \in \{0,1\}^n$ with $\text{wt}(\alpha) < n$.

If $\text{wt}(\alpha)$ is odd, then $f_\alpha = g_\alpha = 0$ since both f and g satisfy the even Parity Condition. If $\text{wt}(\alpha)$ is even, we prove the claim by induction on $k = \text{wt}(\alpha)$. For $k = 0, 2$, $f_\alpha = g_\alpha$ by the definition of g. By induction, we may assume that $f_\beta = g_\beta$ for any $\text{wt}(\beta) < k$. For $\text{wt}(\alpha) = k \geq 4$, let $P = \{p_1, p_2, \ldots, p_{\text{wt}(\alpha)}\} \subseteq [n]$ be such that $\alpha_{p_i} = 1$ and all other bits of α are 0. Then there exists $\ell \in [n]$ such that $\ell \notin P$ since $k < n$. Since $f^{x_\ell=0}$ is a matchgate signature, by the MGIs we have

$$\sum_{j=1}^{k} (-1)^j f_{e_{p_1}\oplus e_{p_j}} f_{\alpha\oplus e_{p_1}\oplus e_{p_j}} = 0, \tag{5.1}$$

where the position vector is P and the pattern is e_{p_1}. Note that all entries of f that appear in (5.1) are indeed entries of $f^{x_\ell=0}$.

The first term of (5.1) is $-f_\alpha$ since $f_{00\cdots0} = 1$. Thus

$$f_\alpha = \sum_{j=2}^{k} (-1)^j f_{e_{p_1}\oplus e_{p_j}} f_{\alpha\oplus e_{p_1}\oplus e_{p_j}}. \tag{5.2}$$

Similarly, since g is a matchgate signature and $g_{00\cdots0} = f_{00\cdots0} = 1$, we have

$$g_\alpha = \sum_{j=2}^{k} (-1)^j g_{e_{p_1}\oplus e_{p_j}} g_{\alpha\oplus e_{p_1}\oplus e_{p_j}}. \tag{5.3}$$

Since $\mathrm{wt}(e_{p_1} \oplus e_{p_j}) = 2$, $\mathrm{wt}(\alpha \oplus e_{p_1} \oplus e_{p_j}) = k-2$ for $2 \le j \le k$, the RHS expressions of (5.2) and (5.3) are equal by induction, and thus $f_\alpha = g_\alpha$.

This completes the proof of the claim.

If n is odd, then since f satisfies the even Parity Condition and so does $g \in \mathscr{M}$ with $g_{00\cdots0} = 1$, both $f_{11\cdots1} = g_{11\cdots1} = 0$. Then by the claim, f is identically equal to g; thus $f \in \mathscr{M}$. This is a contradiction.

If n is even, then $n \ge 6$. Since $f_\alpha = g_\alpha$ for $\mathrm{wt}(\alpha) < n$, there exists $x \in \mathbb{C}$ such that

$$f = g + x[0,1]^{\otimes n}. \tag{5.4}$$

If $x = 0$, then $f \in \mathscr{M}$, a contradiction. Thus $x \neq 0$. Since we have $(=_2) \in \widehat{\mathcal{EQ}}$, we can construct $f' = \partial^{\{1,2\}}_{(=_2)}(f)$. Let $g' = \partial^{\{1,2\}}_{(=_2)}(g)$. Note that $g' \in \mathscr{M}$. A matchgate for g' is obtained from a matchgate for g by joining the two adjacent external dangling edges corresponding to x_1 and x_2. On the other hand, since the operator ∂ is linear, by (5.4) we have

$$f' = g' + x[0,1]^{\otimes(n-2)}.$$

This implies that

$$f'_\beta = g'_\beta \tag{5.5}$$

for any $\beta \in \{0,1\}^{n-2}$ with $\mathrm{wt}(\beta) < n-2$. If $f' \notin \mathscr{M}$, then we are done by induction. Otherwise, $f' \in \mathscr{M}$.

If $f'_{00\cdots0} \neq 0$, consider the MGIs for f' and g' determined by the position vector $P' = \{1, 2, \ldots, n-2\}$ and the pattern $e_1 = 10\cdots0 \in \{0,1\}^{n-2}$. Then

$$f'_{00\cdots0} f'_{11\cdots1} = \sum_{j=2}^{n-2} (-1)^j f'_{e_1 \oplus e_j} f'_{11\cdots1 \oplus e_1 \oplus e_j},$$

$$g'_{00\cdots0} g'_{11\cdots1} = \sum_{j=2}^{n-2} (-1)^j g'_{e_1 \oplus e_j} g'_{11\cdots1 \oplus e_1 \oplus e_j}.$$

Note that $\mathrm{wt}(e_1 \oplus e_j) = 2$ and $\mathrm{wt}(11\cdots1 \oplus e_1 \oplus e_j) = n-4$ for $2 \le j \le n-2$ in the above expressions. Thus by (5.5) we have

$$f'_{00\cdots0} f'_{11\cdots1} = g'_{00\cdots0} g'_{11\cdots1}.$$

By $f'_{00\cdots0} = g'_{00\cdots0} \neq 0$, we have

$$f'_{11\cdots1} = g'_{11\cdots1}.$$

This contradicts that $x \neq 0$.

If $f'_{00\cdots0} = 0$, i.e., $f_{000\cdots0} + f_{110\cdots0} = 0$, then $f_{110\cdots0} = -1$, and we can construct $\partial^{\{3,4,\ldots,n\}}_{[1,0]}(f) = [1,0,-1]$, since $[1,0] \in \widehat{\mathcal{EQ}}$. Then we can construct $f'' = \partial^{\{1,2\}}_{[1,0,-1]}(f)$ and define $g'' = \partial^{\{1,2\}}_{[1,0,-1]}(g)$. It follows from (5.4) that

$$f'' = g'' - x[0,1]^{\otimes(n-2)}. \tag{5.6}$$

Also $f''_{0\cdots0} = f_{000\cdots0} - f_{110\cdots0} = 2$. Note that g'' is a matchgate signature because $[1,0,-1] \in \mathscr{M}$. If $f'' \notin \mathscr{M}$, then we are done by induction. Otherwise, $f'' \in \mathscr{M}$. Then we can get a contradiction the same way using MGIs as in the case for $f'_{00\cdots0} \neq 0$. ∎

5.2. A dichotomy theorem for Pl-Holant($\widehat{\mathcal{EQ}}, [1,0,x], \widehat{\mathcal{F}}$) with $[1,0,x] \notin \mathscr{A}$. A signature f of arity 4 satisfying the even Parity Condition has signature matrix of the form

$$M_{x_1x_2,x_4x_3}(f) = \begin{bmatrix} a & 0 & 0 & b \\ 0 & \alpha & \beta & 0 \\ 0 & \gamma & \delta & 0 \\ c & 0 & 0 & d \end{bmatrix}.$$

For such signatures we call $\left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right]$ the outer matrix and $\left[\begin{smallmatrix} \alpha & \beta \\ \gamma & \delta \end{smallmatrix}\right]$ the inner matrix. The following lemma implies that we can switch the outer matrix and the inner matrix, and also reverse the order of the columns, when we have $\widehat{\mathcal{EQ}}$ and $[0,1]^{\otimes 2}$.

LEMMA 5.3. *If $\widehat{\mathcal{F}}$ contains an f with signature matrix*

$$M_{x_1x_2,x_4x_3}(f) = \begin{bmatrix} f_{0000} & f_{0010} & f_{0001} & f_{0011} \\ f_{0100} & f_{0110} & f_{0101} & f_{0111} \\ f_{1000} & f_{1010} & f_{1001} & f_{1011} \\ f_{1100} & f_{1110} & f_{1101} & f_{1111} \end{bmatrix} = \begin{bmatrix} a & 0 & 0 & b \\ 0 & \alpha & \beta & 0 \\ 0 & \gamma & \delta & 0 \\ c & 0 & 0 & d \end{bmatrix}, \tag{5.7}$$

then we can construct g and h, where

$$M_{x_1x_2,x_4x_3}(g) = \begin{bmatrix} \alpha & 0 & 0 & \beta \\ 0 & a & b & 0 \\ 0 & c & d & 0 \\ \gamma & 0 & 0 & \delta \end{bmatrix} \quad \text{and} \quad M_{x_1x_2,x_4x_3}(h) = \begin{bmatrix} \beta & 0 & 0 & \alpha \\ 0 & b & a & 0 \\ 0 & d & c & 0 \\ \delta & 0 & 0 & \gamma \end{bmatrix}$$

such that

$$\text{Pl-Holant}(\widehat{\mathcal{EQ}}, g, h, [0,1]^{\otimes 2}, \widehat{\mathcal{F}}) \leq_{\text{T}} \text{Pl-Holant}(\widehat{\mathcal{EQ}}, [0,1]^{\otimes 2}, \widehat{\mathcal{F}}).$$

Proof. We have $[0,1], [1,0,1,0] \in \widehat{\mathcal{EQ}}$, and $[0,1]^{\otimes 2}$. Lemma 2.37 shows that we can flip any two variables in f. If we flip variables x_2, x_3 of f, we get g. If we flip variables x_2, x_4 of f, we get h. ∎

If $f_{0000} = a \neq 0$, we can normalize it to 1. The next lemma deals with signatures of arity 4 that "just miss" being matchgate signatures. Note that a signature of the form (5.7) is a matchgate signature iff the determinants of the inner matrix and the outer matrix are equal (Lemma 2.29). Lemma 5.4 shows how to clear some entries of (5.7).

LEMMA 5.4. *Suppose $[1,0,x] \notin \mathscr{A}$, and $\widehat{\mathcal{F}}$ contains a signature f of arity 4 such that*

$$M_{x_1x_2,x_4x_3}(f) = \begin{bmatrix} 1 & 0 & 0 & b \\ 0 & \alpha & \beta & 0 \\ 0 & \gamma & \delta & 0 \\ c & 0 & 0 & d \end{bmatrix},$$

satisfying $\det\left[\begin{smallmatrix} 1 & b \\ c & d \end{smallmatrix}\right] = -\det\left[\begin{smallmatrix} \alpha & \beta \\ \gamma & \delta \end{smallmatrix}\right] \neq 0$. *Then we can construct f' such that*

$$\text{Pl-Holant}(f', \widehat{\mathcal{EQ}}, [1,0,x], \widehat{\mathcal{F}}) \leq_{\text{T}} \text{Pl-Holant}(\widehat{\mathcal{EQ}}, [1,0,x], \widehat{\mathcal{F}}),$$

where f' has the form

$$M_{x_1x_2,x_4x_3}(f') = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \alpha' & \beta' & 0 \\ 0 & \gamma' & \delta' & 0 \\ 0 & 0 & 0 & d' \end{bmatrix},$$

and f' satisfies the following conditions:

- $\det\left[\begin{smallmatrix}1 & 0\\ 0 & d'\end{smallmatrix}\right] = d' = -\det\left[\begin{smallmatrix}\alpha' & \beta'\\ \gamma' & \delta'\end{smallmatrix}\right] \neq 0$.
- *If* $\left[\begin{smallmatrix}\alpha & \beta\\ \gamma & \delta\end{smallmatrix}\right]$ *is a diagonal (resp., antidiagonal) matrix, then* $\left[\begin{smallmatrix}\alpha' & \beta'\\ \gamma' & \delta'\end{smallmatrix}\right]$ *is also a diagonal (resp., antidiagonal) matrix.*

Proof. If $b = c = 0$, then we are done by letting $f' = f$. In the following, we assume that there is at most one zero in $\{b, c\}$.

By Lemma 4.1, we can get $[1, 0, z]$ for all $z \in \mathbb{C}$ from the given $[1, 0, x] \notin \mathscr{A}$. We use two binary signatures $[1, 0, u], [1, 0, v]$, where $u, v \in \mathbb{C}$ and $uv \neq 0$. By connecting the first variables of $[1, 0, u]$ and $[1, 0, v]$ to the variables x_2, x_4 of f, respectively (see (7.19) and (7.21)), we get a gadget with signature

$$h(x_1, x_2, x_3, x_4) = \sum_{x_2', x_4' \in \{0,1\}} f(x_1, x_2', x_3, x_4')[1, 0, u](x_2', x_2)[1, 0, v][x_4', x_4],$$

and

$$M_{x_1x_2, x_4x_3}(h) = \begin{bmatrix} 1 & 0 & 0 & bv \\ 0 & \alpha u & \beta uv & 0 \\ 0 & \gamma & \delta v & 0 \\ cu & 0 & 0 & duv \end{bmatrix}.$$

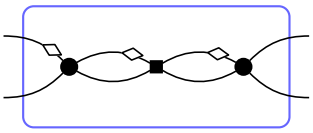


FIG. 15. *The two circle vertices are assigned f, and the square vertex is assigned h.*

In Figure 15, by assigning f to the circle vertices and assigning h to the square vertex, we get a gadget with the signature

$$f'(x_1, x_2, x_3, x_4) = \sum_{y_1', y_2', x_3', x_4' \in \{0,1\}} f(x_1, x_2, x_3', x_4')h(x_4', x_3', y_2', y_1')f(y_1', y_2', x_3, x_4).$$

We have

$$M_{x_1x_2, x_4x_3}(f') = \begin{bmatrix} 1 & 0 & 0 & b \\ 0 & \alpha & \beta & 0 \\ 0 & \gamma & \delta & 0 \\ c & 0 & 0 & d \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & bv \\ 0 & \alpha u & \beta uv & 0 \\ 0 & \gamma & \delta v & 0 \\ cu & 0 & 0 & duv \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & b \\ 0 & \alpha & \beta & 0 \\ 0 & \gamma & \delta & 0 \\ c & 0 & 0 & d \end{bmatrix}.$$

Thus

$$\begin{aligned} \begin{bmatrix} f'_{0000} & f'_{0011} \\ f'_{1100} & f'_{1111} \end{bmatrix} &= \begin{bmatrix} 1 & b \\ c & d \end{bmatrix} \begin{bmatrix} 1 & bv \\ cu & duv \end{bmatrix} \begin{bmatrix} 1 & b \\ c & d \end{bmatrix} \\ &= \begin{bmatrix} 1 + bc(u + v + duv) & b[1 + bcu + dv(1 + du)] \\ c[1 + bcv + du(1 + dv)] & d^3uv + bc(1 + du + dv) \end{bmatrix}, \end{aligned} \tag{5.8}$$

and

$$\begin{bmatrix} f'_{0110} & f'_{0101} \\ f'_{1010} & f'_{1001} \end{bmatrix} = \begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix} \begin{bmatrix} \alpha u & \beta uv \\ \gamma & \delta v \end{bmatrix} \begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix}. \tag{5.9}$$

By (5.8), (5.9), we have

$$\det\begin{bmatrix} f'_{0000} & f'_{0011} \\ f'_{1100} & f'_{1111} \end{bmatrix} = (d-bc)^3uv \quad \text{and} \quad \det\begin{bmatrix} f'_{0110} & f'_{0101} \\ f'_{1010} & f'_{1001} \end{bmatrix} = (\alpha\delta-\beta\gamma)^3uv.$$

Then, by $\alpha\delta-\beta\gamma = -(d-bc) \neq 0$, we have

$$\det\begin{bmatrix} f'_{0000} & f'_{0011} \\ f'_{1100} & f'_{1111} \end{bmatrix} = -\det\begin{bmatrix} f'_{0110} & f'_{0101} \\ f'_{1010} & f'_{1001} \end{bmatrix} \neq 0.$$

Moreover, if $\left[\begin{smallmatrix} \alpha & \beta \\ \gamma & \delta \end{smallmatrix}\right]$ is diagonal (resp., antidiagonal), then $\left[\begin{smallmatrix} f'_{0110} & f'_{0101} \\ f'_{1010} & f'_{1001} \end{smallmatrix}\right]$ is also diagonal (resp., antidiagonal).

In the following proof, first, in item (A), we will prove the lemma for a special case that $b = 0$ or $c = 0$. Then we finish the proof in item (B) by reducing the general case to the special case in item (A).

(A) Suppose $bc = 0$. If $c = 0$, we can rotate the signature by 180° to get $b = 0$ (Figure 2). So we assume that $b = 0$ and $c \neq 0$. Note that $d \neq 0$ since $\det\left[\begin{smallmatrix} 1 & 0 \\ c & d \end{smallmatrix}\right] \neq 0$. Thus we may assume that $cd \neq 0$. Note that $f'_{0011} = 0$ by $b = 0$ and $f'_{1100} = c[1 + du(1 + dv)]$. Let $u = \frac{1}{d}, v = -\frac{2}{d}$; then $1 + du(1 + dv) = 0$ and therefore $f'_{1100} = c[1 + du(1 + dv)] = 0$. This implies that

$$M_{x_1x_2,x_4x_3}(f') = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \alpha' & \beta' & 0 \\ 0 & \gamma' & \delta' & 0 \\ 0 & 0 & 0 & d' \end{bmatrix},$$

which satisfies the requirements of the lemma.

(B) For $bc \neq 0$, we reduce the proof to item (A) by choosing suitable u, v such that $f'_{0000} \neq 0, f'_{0011} = 0$ and replacing f by f'. Note that $f'_{0000} \neq 0$ follows from $f'_{0011} = 0$ since $\left[\begin{smallmatrix} f'_{0000} & f'_{0011} \\ f'_{1100} & f'_{1111} \end{smallmatrix}\right]$ has full rank.
Let $\Delta = u + v + duv$; then

$$f'_{0011} = b[1 + (bc - d)u + d\Delta].$$

- For $bc \neq 2d$, let $u = \frac{1}{d-bc}$ and $v = \frac{1}{bc-2d}$; then $\Delta = 0$ and $1+(bc-d)u = 0$. Thus $f'_{0011} = 0$.
- For $bc = 2d$, let $u = \frac{1}{\sqrt{2}d}$ and $v = -\frac{\sqrt{2}}{d}$; then $\Delta = -\frac{1+\sqrt{2}}{\sqrt{2}d}, f'_{0011} = b[1 + du + d\Delta] = 0$. ∎

Now we prove that if all signatures in $\widehat{\mathcal{F}}$ satisfy the Parity Condition and $\widehat{\mathcal{F}}$ contains a binary nonaffine signature $[1, 0, x]$, then either $\widehat{\mathcal{F}} \subseteq \mathscr{M}$ or Pl-Holant$(\widehat{\mathcal{EQ}}, [1, 0, x], \widehat{\mathcal{F}})$ is #P-hard. Note that this is consistent with the final dichotomy Theorem 6.1. If $\widehat{\mathcal{F}}$ satisfies the Parity Condition, then $\mathcal{F} \subseteq \mathscr{P}$ (equivalently $\widehat{\mathcal{F}} \subseteq \widehat{\mathscr{P}}$) would imply $\mathcal{F} \subseteq \mathscr{A}$ (equivalently $\widehat{\mathcal{F}} \subseteq \mathscr{A}$) (see Proposition 7.12). But it contains $[1, 0, x] \notin \mathscr{A}$, and also $[1, 0, x] \in \mathscr{M}$ and $\widehat{\mathcal{EQ}} \subset \mathscr{M}$; therefore, the only tractable case in Theorem 5.5 is $\widehat{\mathcal{F}} \subseteq \mathscr{M}$.

THEOREM 5.5. *Suppose all signatures in $\widehat{\mathcal{F}}$ satisfy the Parity Condition, and suppose $[1, 0, x] \notin \mathscr{A}$, where x is a complex number. Then either $\widehat{\mathcal{F}} \subseteq \mathscr{M}$ or* Pl-Holant$(\widehat{\mathcal{EQ}}, [1, 0, x], \widehat{\mathcal{F}})$ *is* #P*-hard.*

Proof. By Lemma 4.1, we can construct $[0,1]^{\otimes 2}$ and $[1,0,z]$ from $[1,0,x]$ for all $z \in \mathbb{C}$.

Suppose $\widehat{\mathcal{F}} \not\subseteq \mathscr{M}$. By Theorem 5.2, we can construct $f \notin \mathscr{M}$ and f has arity 4, such that

$$\text{Pl-Holant}(f, \widehat{\mathcal{EQ}}, [1,0,x], \widehat{\mathcal{F}}) \leq_{\text{T}} \text{Pl-Holant}(\widehat{\mathcal{EQ}}, [1,0,x], \widehat{\mathcal{F}}).$$

Moreover, by Lemma 2.38, we can assume that f satisfies the even Parity Condition and $f_{0000} = 1$, i.e.,

$$M_{x_1x_2,x_4x_3}(f) = \begin{bmatrix} 1 & 0 & 0 & b \\ 0 & \alpha & \beta & 0 \\ 0 & \gamma & \delta & 0 \\ c & 0 & 0 & d \end{bmatrix}.$$

Let $A = \left[\begin{smallmatrix} 1 & b \\ c & d \end{smallmatrix}\right]$, $B = \left[\begin{smallmatrix} \alpha & \beta \\ \gamma & \delta \end{smallmatrix}\right]$; then $f \notin \mathscr{M}$ iff $\det A \neq \det B$ by Lemma 2.29. We may assume that $\det A \neq 0$. If $\det A = 0$, then $\det B \neq 0$, which implies that $\alpha \neq 0$ or $\beta \neq 0$. By Lemma 5.3, we may switch the inner and outer matrices and reverse the order of the columns if necessary. Hence we may assume that $\det A \neq 0$.

Claim. We can construct some $f' \notin \mathscr{M}$ that has the form

$$M_{x_1x_2,x_4x_1}(f') = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \alpha' & \beta' & 0 \\ 0 & \gamma' & \delta' & 0 \\ 0 & 0 & 0 & d' \end{bmatrix},$$

such that $d' \neq 0$ and

$$\text{Pl-Holant}(f', \widehat{\mathcal{EQ}}, [1,0,x], \widehat{\mathcal{F}}) \leq_{\text{T}} \text{Pl-Holant}(\widehat{\mathcal{EQ}}, [1,0,x], \widehat{\mathcal{F}}).$$

To prove this claim, if $\det A = -\det B$, then since $\det A \neq 0$, we may apply Lemma 5.4, and the claim is proved. So we may assume $\det A \neq -\det B$. Together with the nonmatchgate condition, we have $\det A \neq \pm \det B$. If $d = 0$, then we have $bc \neq 0$ by $\det A \neq 0$. Use a binary $[1, 0, -2/(bc)]$ to modify the second variable of f (as in the proof of Lemma 5.4, see (7.19)) and then connect a copy of f with the modified f. We get

$$h(x_1,x_2,x_3,x_4) = \sum_{x_3', x_3'', x_4' \in \{0,1\}} f(x_1,x_2,x_3',x_4') f(x_4',x_3'',x_3,x_4) \left[1, 0, -\frac{2}{bc}\right](x_3', x_3'').$$

Then

$$M_{x_1x_2,x_4x_3}(h) = \begin{bmatrix} 1 & 0 & 0 & b \\ 0 & \alpha & \beta & 0 \\ 0 & \gamma & \delta & 0 \\ c & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & b \\ 0 & -\frac{2\alpha}{bc} & -\frac{2\beta}{bc} & 0 \\ 0 & \gamma & \delta & 0 \\ -\frac{2}{b} & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} -1 & 0 & 0 & b \\ 0 & \alpha'' & \beta'' & 0 \\ 0 & \gamma'' & \delta'' & 0 \\ c & 0 & 0 & bc \end{bmatrix},$$

where

$$\begin{bmatrix} \alpha'' & \beta'' \\ \gamma'' & \delta'' \end{bmatrix} = \begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix} \begin{bmatrix} -\frac{2\alpha}{bc} & -\frac{2\beta}{bc} \\ \gamma & \delta \end{bmatrix}.$$

Thus $\det \left[\begin{smallmatrix} \alpha'' & \beta'' \\ \gamma'' & \delta'' \end{smallmatrix}\right] = -\frac{2(\alpha\delta - \beta\gamma)^2}{bc}$. So

$$\det \begin{bmatrix} \alpha'' & \beta'' \\ \gamma'' & \delta'' \end{bmatrix} \neq \det \begin{bmatrix} -1 & b \\ c & bc \end{bmatrix}$$

by $b^2c^2 \neq (\alpha\delta - \beta\gamma)^2$, which is the same as $\det A \neq \pm \det B$. This implies that $h \notin \mathscr{M}$. If

$$\det \begin{bmatrix} \alpha'' & \beta'' \\ \gamma'' & \delta'' \end{bmatrix} = -\det \begin{bmatrix} -1 & b \\ c & bc \end{bmatrix}, \tag{5.10}$$

then this quantity in (5.10) is also nonzero because $bc \neq 0$. Then we can finish the proof of the claim by Lemma 5.4. Therefore, we may assume

$$\det \begin{bmatrix} \alpha'' & \beta'' \\ \gamma'' & \delta'' \end{bmatrix} \neq \pm \det \begin{bmatrix} -1 & b \\ c & bc \end{bmatrix}.$$

To summarize on h when $d = 0$ in f, we have $h \notin \mathscr{M}$ satisfying the even Parity Condition, its outer determinant is nonzero, $h_{0000} \neq 0, h_{1111} \neq 0$, and the squares of the determinants of the outer matrix and inner matrix are not equal. We may substitute h in place of f. To simplify the notation, we may assume that in the expression for f, we have $\det A \neq 0$, $\det A \neq \pm \det B$, and $d \neq 0$.

Using the same construction as above but with the binary $[1, 0, -1/d]$ instead, we get

$$f'(x_1, x_2, x_3, x_4) = \sum_{x_3', x_3'', x_4' \in \{0,1\}} f(x_1, x_2, x_3', x_4') f(x_4', x_3'', x_3, x_4) \left[1, 0, -\frac{1}{d}\right](x_3', x_3'').$$

Then

$$M_{x_1x_2, x_4x_3}(f') = \begin{bmatrix} 1 & 0 & 0 & b \\ 0 & \alpha & \beta & 0 \\ 0 & \gamma & \delta & 0 \\ c & 0 & 0 & d \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & b \\ 0 & -\frac{\alpha}{d} & -\frac{\beta}{d} & 0 \\ 0 & \gamma & \delta & 0 \\ -\frac{c}{d} & 0 & 0 & -1 \end{bmatrix} = \begin{bmatrix} 1 - \frac{bc}{d} & 0 & 0 & 0 \\ 0 & \alpha' & \beta' & 0 \\ 0 & \gamma' & \delta' & 0 \\ 0 & 0 & 0 & bc - d \end{bmatrix},$$

where

$$\begin{bmatrix} \alpha' & \beta' \\ \gamma' & \delta' \end{bmatrix} = \begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix} \begin{bmatrix} -\frac{\alpha}{d} & -\frac{\beta}{d} \\ \gamma & \delta \end{bmatrix}.$$

Thus $\det \left[\begin{smallmatrix} \alpha' & \beta' \\ \gamma' & \delta' \end{smallmatrix}\right] = -\frac{(\alpha\delta - \beta\gamma)^2}{d}$. So

$$\det \begin{bmatrix} \alpha' & \beta' \\ \gamma' & \delta' \end{bmatrix} \neq \det \begin{bmatrix} 1 - \frac{bc}{d} & 0 \\ 0 & bc - d \end{bmatrix}$$

by $\det A \neq \pm \det B$. This implies that $f' \notin \mathscr{M}$. Note that $f'_{0000} = 1 - \frac{bc}{d} = \frac{\det A}{d} \neq 0$, so we can renormalize f'_{0000} to 1. Also $f'_{1111} = bc - d = -\det A \neq 0$. Thus f' satisfies the requirement of the claim.

This completes the proof of the claim.

The claim shows that we may assume that $b = c = 0$ and $d \neq 0$ in $M_{x_1x_2, x_4x_3}(f)$ and have the form

$$M_{x_1x_2, x_4x_3}(f) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \alpha & \beta & 0 \\ 0 & \gamma & \delta & 0 \\ 0 & 0 & 0 & d \end{bmatrix}.$$

In the following, we can finish the proof of the theorem by two alternatives:

(A) If we can construct the crossover function $\mathfrak{X}$ such that

$$\text{Pl-Holant}(\mathfrak{X}, \widehat{\mathcal{EQ}}, [1,0,x], \widehat{\mathcal{F}}) \le_{\text{T}} \text{Pl-Holant}(\widehat{\mathcal{EQ}}, [1,0,x], \widehat{\mathcal{F}}),$$

then the presence of $\mathfrak{X}$ reduces a general (nonplanar) Holant problem to a Pl-Holant problem

$$\text{Holant}(\widehat{\mathcal{EQ}}, [1,0,x]) \le_{\text{T}} \text{Pl-Holant}(\mathfrak{X}, \widehat{\mathcal{EQ}}, [1,0,x], \widehat{\mathcal{F}}).$$

So we have

$$\text{Holant}(\widehat{\mathcal{EQ}}, [1,0,x]) \le_{\text{T}} \text{Pl-Holant}(\widehat{\mathcal{EQ}}, [1,0,x], \widehat{\mathcal{F}}),$$

and we can apply Theorem 2.33′. We have $[1,0,x] \notin \widehat{\mathscr{P}}$ (this can be directly verified, or use Proposition 7.12) and $[1,0,x] \notin \mathscr{A}$. Thus by Theorem 2.33′, Holant$(\widehat{\mathcal{EQ}}, [1,0,x])$ is #P-hard. Hence Pl-Holant$(\widehat{\mathcal{EQ}}, [1,0,x], \widehat{\mathcal{F}})$ is #P-hard.

(B) If we can construct $(=_4)$ such that

$$\text{Pl-Holant}((=_4), \widehat{\mathcal{EQ}}, [1,0,x], \widehat{\mathcal{F}}) \le_{\text{T}} \text{Pl-Holant}(\widehat{\mathcal{EQ}}, [1,0,x], \widehat{\mathcal{F}}),$$

then, by Lemma 2.3,

$$\text{Pl-Holant}(\widehat{\mathcal{EQ}}, \mathcal{EQ}_2, [1,0,x], \widehat{\mathcal{F}}) \le_{\text{T}} \text{Pl-Holant}(\widehat{\mathcal{EQ}}, [1,0,x], \widehat{\mathcal{F}}).$$

It follows that

$$\text{Pl-Holant}(\mathcal{EQ}_2, [1,0], [1,0,1,0], [1,0,x]) \le_{\text{T}} \text{Pl-Holant}(\widehat{\mathcal{EQ}}, [1,0,x], \widehat{\mathcal{F}})$$

since $[1,0], [1,0,1,0] \in \widehat{\mathcal{EQ}}$. Note that $[1,0] \notin \widehat{\mathscr{M}} \cup \widehat{\mathscr{M}}^{\dagger}$ (Proposition 2.32), $[1,0,1,0] \notin \mathscr{P} \cup \mathscr{A}^{\dagger}$ (Corollary 2.21 and Proposition 2.32), and we are given $[1,0,x] \notin \mathscr{A}$; thus, for the symmetric signature set $\mathcal{G} = \{[1,0], [1,0,1,0], [1,0,x]\}$

$$\mathcal{G} \not\subseteq \mathscr{P}, \quad \mathcal{G} \not\subseteq \mathscr{A}, \quad \mathcal{G} \not\subseteq \mathscr{A}^{\dagger}, \quad \mathcal{G} \not\subseteq \widehat{\mathscr{M}}, \quad \mathcal{G} \not\subseteq \widehat{\mathscr{M}}^{\dagger}.$$

By Theorem 2.35 on Pl-CSP^2 problems for symmetric signatures, the problem Pl-Holant$(\mathcal{EQ}_2, \mathcal{G})$ is #P-hard. It follows that Pl-Holant$(\widehat{\mathcal{EQ}}, [1,0,x], \widehat{\mathcal{F}})$ is #P-hard.

In the following, in case 1, we prove the theorem when $\det A = -\det B$. Then in case 2, we prove the theorem when $\det A \neq -\det B$. Since $f \notin \mathscr{M}$, we are given $\det A \neq \det B$. So case 2 is equivalent to $\det A \neq \pm \det B$.

1. Suppose $\det A = -\det B$. Since we have $\det A \neq 0$, both $\det A$ and $\det B \neq 0$. At least one of α or β is nonzero by $\det B \neq 0$.
   - Suppose $\alpha \neq 0$. By Lemma 5.3, we also have the 4-ary signature g such that
   $$M_{x_1x_2,x_4x_3}(g) = \begin{bmatrix} \alpha & 0 & 0 & \beta \\ 0 & 1 & 0 & 0 \\ 0 & 0 & d & 0 \\ \gamma & 0 & 0 & \delta \end{bmatrix}.$$
   Since $\alpha \neq 0$, we can assume that $\alpha = 1$ by normalizing. Then we may write
   $$M_{x_1x_2,x_4x_3}(g) = \begin{bmatrix} 1 & 0 & 0 & \beta \\ 0 & a & 0 & 0 \\ 0 & 0 & d & 0 \\ \gamma & 0 & 0 & \delta \end{bmatrix},$$

where $\det\left[\begin{smallmatrix} a & 0 \\ 0 & d \end{smallmatrix}\right] = -\det\left[\begin{smallmatrix} 1 & \beta \\ \gamma & \delta \end{smallmatrix}\right] \neq 0$. Then, by Lemma 5.4, we have g' such that

$$M_{x_1x_2,x_4x_3}(g') = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & a' & 0 & 0 \\ 0 & 0 & d' & 0 \\ 0 & 0 & 0 & \delta' \end{bmatrix},$$

where $\det\left[\begin{smallmatrix} a' & 0 \\ 0 & d' \end{smallmatrix}\right] = -\det\left[\begin{smallmatrix} 1 & 0 \\ 0 & \delta' \end{smallmatrix}\right] \neq 0$. Thus $\left[\begin{smallmatrix} 1 & d' \\ a' & \delta' \end{smallmatrix}\right]$ has full rank. Then, by Lemma 2.40, we have $(=_4)$ by interpolation, and we are done by alternative (B).

- Suppose $\beta \neq 0$; by Lemma 5.3, we have h such that

$$M_{x_1x_2,x_4x_3}(h) = \begin{bmatrix} \beta & 0 & 0 & \alpha \\ 0 & 0 & 1 & 0 \\ 0 & d & 0 & 0 \\ \delta & 0 & 0 & \gamma \end{bmatrix}.$$

Since $\beta \neq 0$, we can assume that $\beta = 1$ by normalizing. Then we may write

$$M_{x_1x_2,x_4x_3}(h) = \begin{bmatrix} 1 & 0 & 0 & \alpha \\ 0 & 0 & a & 0 \\ 0 & d & 0 & 0 \\ \delta & 0 & 0 & \gamma \end{bmatrix},$$

where $\det\left[\begin{smallmatrix} 0 & a \\ d & 0 \end{smallmatrix}\right] = -\det\left[\begin{smallmatrix} 1 & \alpha \\ \delta & \gamma \end{smallmatrix}\right] \neq 0$. Then, by Lemma 5.4, we have h' such that

$$M_{x_1x_2,x_4x_3}(h') = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & a' & 0 \\ 0 & d' & 0 & 0 \\ 0 & 0 & 0 & \gamma' \end{bmatrix},$$

where $\det\left[\begin{smallmatrix} 0 & a' \\ d' & 0 \end{smallmatrix}\right] = -\det\left[\begin{smallmatrix} 1 & 0 \\ 0 & \gamma' \end{smallmatrix}\right] \neq 0$. By Lemma 4.1 and using $[1, 0, x]$, we have $[1, 0, (d')^{-1}]$ and $[1, 0, (a')^{-1}]$. Modifying h' on the first and second variables by $[1, 0, (d')^{-1}]$ and $[1, 0, (a')^{-1}]$, respectively, (see (7.18) and (7.19)) gives the crossover function since $\gamma' = a'd'$:

$$\mathfrak{X} = \sum_{x_1', x_2' \in \{0,1\}} h'(x_1', x_2', x_3, x_4)[1, 0, (d')^{-1}](x_1', x_1)[1, 0, (a')^{-1}](x_2', x_2).$$

Then we are done by alternative (A).

2. Suppose $\det A \neq -\det B$. Since $f \notin \mathscr{A}$, we have $\det A \neq \pm \det B$. So $d^2 \neq (\alpha\delta - \beta\gamma)^2$.
   - If $\alpha = \delta = 0$, then $d^2 \neq \beta^2\gamma^2$. We can construct

$$\tilde{f}(x_1, x_2, x_3, x_4) = \sum_{u,v \in \{0,1\}} f(x_1, x_2, u, v) f(v, u, x_3, x_4).$$

Then

$$M_{x_1x_2,x_4x_3}(\tilde{f}) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & \beta & 0 \\ 0 & \gamma & 0 & 0 \\ 0 & 0 & 0 & d \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & \beta & 0 \\ 0 & \gamma & 0 & 0 \\ 0 & 0 & 0 & d \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \beta\gamma & 0 & 0 \\ 0 & 0 & \beta\gamma & 0 \\ 0 & 0 & 0 & d^2 \end{bmatrix}.$$

Note that $\left[\begin{smallmatrix} 1 & \beta\gamma \\ \beta\gamma & d^2 \end{smallmatrix}\right]$ has full rank. Then by Lemma 2.40, we have $(=_4)$ by interpolation. So we are done by alternative (B).

- If $\alpha \neq 0$ and $\delta \neq 0$, for any $u \in \mathbb{C}$ we can construct

$$\hat{h}(x_1, x_2, x_3, x_4) = \sum_{x_2' \in \{0,1\}} f(x_1, x_2', x_3, x_4)[1, 0, u](x_2', x_2).$$

Then, by (7.19),

$$M_{x_1x_2, x_4x_3}(\hat{h}) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \alpha u & \beta u & 0 \\ 0 & \gamma & \delta & 0 \\ 0 & 0 & 0 & du \end{bmatrix}.$$

Then we can construct

$$\hat{f}(x_1, x_2, x_3, x_4) = \sum_{u,v \in \{0,1\}} f(x_1, x_2, u, v)\hat{h}(v, u, x_3, x_4).$$

Then

$$M_{x_1x_2, x_4x_3}(\hat{f}) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \alpha & \beta & 0 \\ 0 & \gamma & \delta & 0 \\ 0 & 0 & 0 & d \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \alpha u & \beta u & 0 \\ 0 & \gamma & \delta & 0 \\ 0 & 0 & 0 & du \end{bmatrix}$$
$$= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \alpha^2 u + \beta\gamma & \beta(\delta + \alpha u) & 0 \\ 0 & \gamma(\delta + \alpha u) & \beta\gamma u + \delta^2 & 0 \\ 0 & 0 & 0 & d^2 u \end{bmatrix}.$$

Choose $u = -\frac{\delta}{\alpha} \neq 0$; then $\beta(\delta + \alpha u) = \gamma(\delta + \alpha u) = 0$ and

$$M_{x_1x_2, x_4x_3}(\hat{f}) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -(\alpha\delta - \beta\gamma) & 0 & 0 \\ 0 & 0 & \frac{\delta}{\alpha}(\alpha\delta - \beta\gamma) & 0 \\ 0 & 0 & 0 & -\frac{d^2\delta}{\alpha} \end{bmatrix}.$$

Note that $\left[\begin{smallmatrix} 1 & \frac{\delta}{\alpha}(\alpha\delta - \beta\gamma) \\ -(\alpha\delta - \beta\gamma) & -\frac{d^2\delta}{\alpha} \end{smallmatrix}\right]$ has full rank since $d^2 \neq (\alpha\delta - \beta\gamma)^2$ by $(\det A)^2 \neq (\det B)^2$. Then by Lemma 2.40 we have $(=_4)$ by interpolation and we are done by alternative (B).

- If $\alpha \neq 0$ and $\delta = 0$, then after a rotation (Figure 2) clockwise by 90° we have

$$M_{x_2x_3, x_1x_4}(f) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & \gamma & 0 \\ 0 & \beta & 0 & 0 \\ \alpha & 0 & 0 & d \end{bmatrix}.$$

We can construct

$$\bar{h}(x_1, x_2, x_3, x_4) = \sum_{x_3' \in \{0,1\}} f(x_1, x_2, x_3', x_4)\left[1, 0, -\frac{1}{d}\right](x_3', x_3).$$

Then (see (7.19))

$$M_{x_2x_3,x_1x_4}(\bar{h}) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & -\frac{\gamma}{d} & 0 \\ 0 & \beta & 0 & 0 \\ -\frac{\alpha}{d} & 0 & 0 & -1 \end{bmatrix}.$$

With this we can further construct

$$\bar{f}(x_1, x_2, x_3, x_4) = \sum_{x_1', x_4' \in \{0,1\}} f(x_1', x_2, x_3, x_4')\bar{h}(x_1, x_1', x_4', x_4),$$

with

$$M_{x_2x_3,x_1x_4}(\bar{f}) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & \gamma & 0 \\ 0 & \beta & 0 & 0 \\ \alpha & 0 & 0 & d \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & -\frac{\gamma}{d} & 0 \\ 0 & \beta & 0 & 0 \\ -\frac{\alpha}{d} & 0 & 0 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \beta\gamma & 0 & 0 \\ 0 & 0 & -\frac{\beta\gamma}{d} & 0 \\ 0 & 0 & 0 & -d \end{bmatrix}.$$

Note that $\begin{bmatrix} 1 & \beta\gamma \\ -\frac{\beta\gamma}{d} & -d \end{bmatrix}$ has full rank since $d^2 \neq \beta^2\gamma^2$ by $(\det A)^2 \neq (\det B)^2$. By Lemma 2.40, we have $(=_4)$ by interpolation, and we are done by alternative (B).

- If $\alpha = 0$ and $\delta \neq 0$, then the proof is symmetric by first rotating f counterclockwise by 90° (Figure 2) and then switching the roles of α and δ in the previous item. ∎

5.3. A dichotomy when $\widehat{\mathcal{F}}$ satisfies parity.

LEMMA 5.6. *Suppose $\mathcal{F}$ contains a 4-ary signature $f \notin \mathscr{M}$ of the form*

$$M_{x_1x_2,x_4x_3}(f) = \begin{bmatrix} 1 & 0 & 0 & b \\ 0 & \alpha & \beta & 0 \\ 0 & \gamma & \delta & 0 \\ c & 0 & 0 & d \end{bmatrix},$$

where at least one of $\{b, c, \alpha, \delta\}$ is nonzero; then we can construct $[0,1]^{\otimes 2}$ such that

$$\text{Pl-Holant}([0,1]^{\otimes 2}, \widehat{\mathcal{EQ}}, [1,0,-1], \widehat{\mathcal{F}}) \leq_{\text{T}} \text{Pl-Holant}(\widehat{\mathcal{EQ}}, [1,0,-1], \widehat{\mathcal{F}}).$$

Proof. By a rotation (Figure 2), without loss of generality we can assume that $b \neq 0$. We have $\partial^{\{1,2\}}_{[1,0]}(f) = [1,0,b]$. If $b^2 \neq 1$, then we have $[0,1]^{\otimes 2}$ by Lemma 4.1. Therefore, we may assume that $b = \pm 1$.

By a planar gadget we can construct the signature

$$h(x_1, x_2, x_3, x_4) = \sum_{x_3', x_3'', x_4' \in \{0,1\}} f(x_1, x_2, x_3', x_4')f(x_3, x_4, x_4', x_3'')[1,0,-1](x_3'', x_3').$$

Then (see Figure 2 and (7.19))

$$M_{x_1x_2,x_4x_3}(h) = \begin{bmatrix} 1 & 0 & 0 & b \\ 0 & \alpha & \beta & 0 \\ 0 & \gamma & \delta & 0 \\ c & 0 & 0 & d \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & c \\ 0 & -\delta & -\beta & 0 \\ 0 & \gamma & \alpha & 0 \\ -b & 0 & 0 & -d \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & c - bd \\ 0 & \alpha' & \beta' & 0 \\ 0 & \gamma' & \delta' & 0 \\ c - bd & 0 & 0 & c^2 - d^2 \end{bmatrix}.$$

If $c - bd \neq 0$, then we have $\partial^{\{1,2\}}_{[1,0]}(h) = (c - bd)[0,1]^{\otimes 2}$, a nonzero multiple of $[0,1]^{\otimes 2}$. Otherwise, we have

$$d - bc = d - b^2 d = 0,$$

by substituting $c = bd$ and $b^2 = 1$. We also have $c^2 = d^2$ in this case. This implies that $\det \left[\begin{smallmatrix} 1 & b \\ c & d \end{smallmatrix}\right] = 0$. Since $f \notin \mathscr{M}$, we have $\det \left[\begin{smallmatrix} \alpha & \beta \\ \gamma & \delta \end{smallmatrix}\right] \neq 0$ by Lemma 2.29. Thus $\det \left[\begin{smallmatrix} -\delta & -\beta \\ \gamma & \alpha \end{smallmatrix}\right] \neq 0$. So $\left[\begin{smallmatrix} \alpha' & \beta' \\ \gamma' & \delta' \end{smallmatrix}\right] = \left[\begin{smallmatrix} \alpha & \beta \\ \gamma & \delta \end{smallmatrix}\right] \left[\begin{smallmatrix} -\delta & -\beta \\ \gamma & \alpha \end{smallmatrix}\right]$ has full rank. This implies that at least one of $\alpha', \beta', \gamma', \delta'$ is nonzero.

Now h takes the form

$$M_{x_1x_2,x_4x_3}(h) = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \alpha' & \beta' & 0 \\ 0 & \gamma' & \delta' & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Then we can obtain $[0,1]^{\otimes 2}$ from $[1,0] \in \widehat{\mathcal{EQ}}$ and h. For example, if $\alpha' \neq 0$, then $\partial^{\{1,4\}}_{[1,0]}(h) = [0,0,\alpha'] = \alpha'[0,1]^{\otimes 2}$. ∎

THEOREM 5.7. *If all signatures in $\widehat{\mathcal{F}}$ satisfy the Parity Condition, then the following dichotomy holds: If $\widehat{\mathcal{F}} \subseteq \mathscr{A}$, or $\widehat{\mathcal{F}} \subseteq \widehat{\mathscr{P}}$, or $\widehat{\mathcal{F}} \subseteq \mathscr{M}$, then* Pl-Holant$(\widehat{\mathcal{EQ}}, \widehat{\mathcal{F}})$ *is tractable; otherwise it is* #P*-hard.*

Proof. Clearly if $\widehat{\mathcal{F}} \subseteq \mathscr{A}$, or $\widehat{\mathcal{F}} \subseteq \widehat{\mathscr{P}}$, or $\widehat{\mathcal{F}} \subseteq \mathscr{M}$, then Pl-Holant$(\widehat{\mathcal{EQ}}, \widehat{\mathcal{F}})$ is tractable. (Since all signatures in $\widehat{\mathcal{F}}$ satisfy the Parity Condition, Proposition 7.12 implies that if $\widehat{\mathcal{F}} \subseteq \widehat{\mathscr{P}}$, then in fact $\widehat{\mathcal{F}} \subseteq \mathscr{A}$. But the proof below will not use this fact.)

Now suppose $\widehat{\mathcal{F}} \not\subseteq \mathscr{A}$, and $\widehat{\mathcal{F}} \not\subseteq \widehat{\mathscr{P}}$, and $\widehat{\mathcal{F}} \not\subseteq \mathscr{M}$. Since $\widehat{\mathcal{F}} \not\subseteq \mathscr{M}$, by Theorem 5.2, we can construct a 4-ary signature $f \notin \mathscr{M}$ from $\widehat{\mathcal{F}}$. By Lemma 2.38, we can assume that

$$M_{x_1x_2,x_4x_3}(f) = \begin{bmatrix} 1 & 0 & 0 & b \\ 0 & \alpha & \beta & 0 \\ 0 & \gamma & \delta & 0 \\ c & 0 & 0 & d \end{bmatrix}.$$

We can finish the proof by the following four alternatives:

(A) If we can get $[1,0,x] \notin \mathscr{A}$ such that

$$\text{Pl-Holant}(\widehat{\mathcal{EQ}}, [1,0,x], \widehat{\mathcal{F}}) \leq_{\text{T}} \text{Pl-Holant}(\widehat{\mathcal{EQ}}, \widehat{\mathcal{F}}),$$

then Pl-Holant$(\widehat{\mathcal{EQ}}, \widehat{\mathcal{F}})$ is #P-hard since Pl-Holant$(\widehat{\mathcal{EQ}}, [1,0,x], \widehat{\mathcal{F}})$ is #P-hard by Theorem 5.5.

(B) If we can get $[1,0,-1]$ and $[0,1]^{\otimes 2}$ such that

$$\text{Pl-Holant}(\widehat{\mathcal{EQ}}, [1,0,-1], [0,1]^{\otimes 2}, \widehat{\mathcal{F}}) \leq_{\text{T}} \text{Pl-Holant}(\widehat{\mathcal{EQ}}, \widehat{\mathcal{F}}),$$

then since $\widehat{\mathcal{F}} \not\subseteq \mathscr{A}$, we can get $[1,0,x] \notin \mathscr{A}$ by Lemma 4.6, such that

$$\text{Pl-Holant}(\widehat{\mathcal{EQ}}, [1,0,x], [1,0,-1], [0,1]^{\otimes 2}, \widehat{\mathcal{F}}) \leq_{\text{T}} \text{Pl-Holant}(\widehat{\mathcal{EQ}}, [1,0,-1], [0,1]^{\otimes 2}, \widehat{\mathcal{F}}).$$

Pl-Holant$(\widehat{\mathcal{EQ}}, [1,0,x], [1,0,-1], [0,1]^{\otimes 2}, \widehat{\mathcal{F}})$ is #P-hard by Theorem 5.5. Thus Pl-Holant$(\widehat{\mathcal{EQ}}, \widehat{\mathcal{F}})$ is #P-hard.

(C) If we can construct the crossover function $\mathfrak{X}$ (Definition 2.4) such that

$$\text{Pl-Holant}(\widehat{\mathcal{EQ}}, \mathfrak{X}, \widehat{\mathcal{F}}) \leq_{\text{T}} \text{Pl-Holant}(\widehat{\mathcal{EQ}}, \widehat{\mathcal{F}}),$$

then note that the Holant problem (on general, not necessarily planar, instances) can be reduced to the planar one

$$\text{Holant}(\widehat{\mathcal{EQ}}, \widehat{\mathcal{F}}) \equiv_{\text{T}} \text{Pl-Holant}(\widehat{\mathcal{EQ}}, \mathfrak{X}, \widehat{\mathcal{F}}).$$

Since $\widehat{\mathcal{F}} \not\subseteq \widehat{\mathscr{P}}$ and $\widehat{\mathcal{F}} \not\subseteq \mathscr{A}$, we have that Pl-Holant$(\widehat{\mathcal{EQ}}, \widehat{\mathcal{F}})$ is #P-hard by Theorem 2.33′.

(D) If we have $(=_4)$, by Lemma 2.3, we have

$$\text{Pl-Holant}(\widehat{\mathcal{EQ}}, \mathcal{EQ}_2, \widehat{\mathcal{F}}) \leq_{\text{T}} \text{Pl-Holant}(\widehat{\mathcal{EQ}}, \widehat{\mathcal{F}}).$$

Then by Theorem 4.9, Pl-Holant$(\widehat{\mathcal{EQ}}, \mathcal{EQ}_2, \widehat{\mathcal{F}})$ is #P-hard, since $\widehat{\mathcal{F}} \not\subseteq \mathscr{A}$. Thus Pl-Holant$(\widehat{\mathcal{EQ}}, \widehat{\mathcal{F}})$ is #P-hard.

Note that for any $x \in \{b, c, \alpha, \beta, \gamma, \delta\}$, we have $[1, 0, x]$ by taking $\partial^{\{i,j\}}_{[1,0]}(f)$ on some two variables x_i and x_j. If there exists $x \in \{b, c, \alpha, \beta, \gamma, \delta\}$ such that $x^4 \neq 0, 1$, then $[1, 0, x] \notin \mathscr{A}$ by Proposition 2.17, and we are done by alternative (A). If there exists $x \in \{b, c, \alpha, \beta, \gamma, \delta\}$ such that $x^2 = -1$, then we have $[1, 0, -1]$ and $[0, 1]^{\otimes 2}$ by Lemma 4.1 such that

$$\text{Pl-Holant}(\widehat{\mathcal{EQ}}, [1, 0, -1], [0, 1]^{\otimes 2}, \widehat{\mathcal{F}}) \leq_{\text{T}} \text{Pl-Holant}(\widehat{\mathcal{EQ}}, \widehat{\mathcal{F}}).$$

Thus we are done by alternative (B). So in the following, we may assume that

$$\{b, c, \alpha, \beta, \gamma, \delta\} \subseteq \{0, 1, -1\}.$$

Now we finish the proof by a case analysis of $\{b, c, \alpha, \beta, \gamma, \delta\}$.

If $b = c = \alpha = \delta = 0$, then

$$M_{x_1x_2, x_4x_3}(f) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & \beta & 0 \\ 0 & \gamma & 0 & 0 \\ 0 & 0 & 0 & d \end{bmatrix}.$$

In this case,

- if $\beta\gamma \neq 0$, then we have (see (7.18) and (7.19), and note that $\beta = \beta^{-1}$ and $\gamma = \gamma^{-1}$)

$$h(x_1, x_2, x_3, x_4) = \sum_{x'_1, x'_2 \in \{0,1\}} f(x'_1, x'_2, x_3, x_4)[1, 0, \gamma](x'_1, x_1)[1, 0, \beta](x'_2, x_2)$$

and

$$M_{x_1x_2, x_4x_3}(h) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & d\beta\gamma \end{bmatrix}.$$

Note that we have $\partial^{\{1,2\}}_{(=_2)}(f) = [1, 0, d\beta\gamma]$.

If $(d\beta\gamma)^4 \neq 0, 1$, then we are done by alternative (A).

If $(d\beta\gamma)^2 = -1$, then we have $[1, 0, -1]$ and $[0, 1]^{\otimes 2}$ by Lemma 4.1. Then we are done by alternative (B).

If $d\beta\gamma = -1$, then $f \in \mathscr{M}$ by Lemma 2.29. This is a contradiction.

If $d\beta\gamma = 1$, then h is the crossover function $\mathfrak{X}$ (Definition 2.4). Thus we are done by alternative (C).

If $d\beta\gamma = 0$, then

$$M_{x_1x_2,x_4x_3}(h) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Take two copies of h, and connect x_2, x_3, x_4 of the first copy with x_4, x_3, x_2 of the second copy; the planar gadget has the signature

$$\tilde{h}(x_1, x_2) = \sum_{x_2', x_3', x_4' \in \{0,1\}} h(x_1, x_2', x_3', x_4') h(x_2, x_4', x_3', x_2').$$

Then $\tilde{h} = [2, 0, 1] = 2[1, 0, \frac{1}{2}]$. Thus we are done by alternative (A).

- If $\beta\gamma = 0$, then $d \neq 0$ by Lemma 2.29, since $f \notin \mathscr{A}$. By connecting two copies of f, we get

$$h'(x_1, x_2, x_3, x_4) = \sum_{x_3', x_4' \in \{0,1\}} f(x_1, x_2, x_3', x_4') f(x_4', x_3', x_3, x_4),$$

where

$$M_{x_1x_2,x_4x_3}(h') = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & \beta & 0 \\ 0 & \gamma & 0 & 0 \\ 0 & 0 & 0 & d \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & \beta & 0 \\ 0 & \gamma & 0 & 0 \\ 0 & 0 & 0 & d \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & d^2 \end{bmatrix}.$$

Then by Lemma 2.40 we can get $(=_4)$ by interpolation. Thus we are done by alternative (D).

Now we may assume that at least one of $\{b, c, \alpha, \delta\}$ is nonzero. By a rotation, without loss of generality, we assume that $b \neq 0$. In this case, if any one of $\{b, c, \alpha, \beta, \gamma, \delta\}$ is -1, then we have $[1, 0, -1]$ and then also have $[0, 1]^{\otimes 2}$ by Lemma 5.6. Then we are done by alternative (B).

Now we may assume that

$$\{b, c, \alpha, \beta, \gamma, \delta\} \subseteq \{0, 1\} \quad \text{and} \quad b = 1.$$

In this case, we have the following:

- For $c = 1$, note that we have

$$M_{x_2,x_4x_3}(f^{x_1=0}) = \begin{bmatrix} f_{0000} & f_{0010} & f_{0001} & f_{0011} \\ f_{0100} & f_{0110} & f_{0101} & f_{0111} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & \alpha & \beta & 0 \end{bmatrix},$$
$$M_{x_1,x_4x_3}(f^{x_2=0}) = \begin{bmatrix} f_{0000} & f_{0010} & f_{0001} & f_{0011} \\ f_{1000} & f_{1010} & f_{1001} & f_{1011} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & \gamma & \delta & 0 \end{bmatrix},$$
$$M_{x_1x_2,x_4}(f^{x_3=0}) = \begin{bmatrix} f_{0000} & f_{0001} \\ f_{0100} & f_{0101} \\ f_{1000} & f_{1001} \\ f_{1100} & f_{1101} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & \beta \\ 0 & \delta \\ 1 & 0 \end{bmatrix},$$
$$M_{x_1x_2,x_3}(f^{x_4=0}) = \begin{bmatrix} f_{0000} & f_{0010} \\ f_{0100} & f_{0110} \\ f_{1000} & f_{1010} \\ f_{1100} & f_{1110} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & \alpha \\ 0 & \gamma \\ 1 & 0 \end{bmatrix}.$$

  If $\{\alpha, \beta, \gamma, \delta\}$ has at least one 0 and one 1, then there exists $i \in [4]$ such that the support of $f^{x_i=0}$ is not affine. Then by Lemma 4.7, we can construct $g \notin \mathscr{A}$ and arity$(g) <$ arity$(f^{x_i=0}) = 3$, and g satisfies the even Parity Condition. Thus g has arity 2. This implies that, up to a nonzero factor, g has the form $[1, 0, x] \notin \mathscr{A}$. Thus we are done by alternative (A).
  So we may assume that $\alpha = \beta = \gamma = \delta = 0$ or $\alpha = \beta = \gamma = \delta = 1$. For $\alpha = \beta = \gamma = \delta = 0$,

$$M_{x_1x_2,x_4x_3}(f) = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & d \end{bmatrix}.$$

  Since $f \notin \mathscr{M}$, by Lemma 2.29 $\det \left[\begin{smallmatrix} 1 & 1 \\ 1 & d \end{smallmatrix}\right] \neq 0$. By Lemma 2.40, we have $(=_4)$, and we are done by alternative (D).
  For $\alpha = \beta = \gamma = \delta = 1$, f is symmetric, namely $f = [1, 0, 1, 0, d]$. If $d = 1$, then $f \in \mathscr{M}$. This is a contradiction. Otherwise, $d \neq 1$; then $f \notin \mathscr{A} \cup \mathscr{M}$. Moreover, $f \notin \widehat{\mathscr{P}}$ by Proposition 7.12. Thus Pl-Holant$(\widehat{\mathcal{EQ}}, f)$ is #P-hard by Theorem 2.34′.
- If $c = 0, d = 0$, then the outer matrix of f is degenerate. Thus the inner matrix has full rank by $f \notin \mathscr{M}$ (Lemma 2.29). This implies that either $\alpha \neq \beta$ or $\gamma \neq \delta$. Because these signature entries α, β, γ, and δ are all 0-1 valued, this implies that either supp$(f^{x_1=0})$ is not affine or supp$(f^{x_2=0})$ is not affine. By Lemma 4.7, we can construct $[1, 0, x] \notin \mathscr{A}$, and we are done by alternative (A).
- If $c = 0, d \neq 0$, then we have $\partial^{\{1,2\}}_{(=_2)} = [1, 0, 1 + d]$ and $\partial^{\{3,4\}}_{=_2} = 2[1, 0, \frac{d}{2}]$.
  If $(\frac{d}{2})^4 \neq 0, 1$, then we are done by alternative (A) and $[1, 0, \frac{d}{2}]$.
  Otherwise, $d = \pm 2$ or $d = \pm 2\mathrm{i}$. If $d \neq -2$, then $(1 + d)^4 \neq 0, 1$, and we are done by alternative (A) and $[1, 0, 1+d]$. If $d = -2$, then $[1, 0, 1+d] = [1, 0, -1]$. By Lemma 5.6, we have $[1, 0, -1]$ and $[0, 1]^{\otimes 2}$, and we are done by alternative (B). ∎

6. Main theorem. By Theorems 3.12 and 5.7, we have the following dichotomy theorem for Pl-Holant$(\widehat{\mathcal{EQ}}, \widehat{\mathcal{F}})$.

THEOREM 6.1. *Let $\widehat{\mathcal{F}}$ be any set of complex-valued signatures in Boolean variables. Then* Pl-Holant$(\widehat{\mathcal{EQ}}, \widehat{\mathcal{F}})$ *is* #P*-hard unless $\widehat{\mathcal{F}} \subseteq \widehat{\mathscr{A}}$, $\widehat{\mathcal{F}} \subseteq \widehat{\mathscr{P}}$, or $\widehat{\mathcal{F}} \subseteq \widehat{\mathscr{M}}$, in which case the problem is computable in polynomial time.*

After the holographic transformation by $\left[\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}\right]$, we have the following dichotomy theorem for planar #CSP over the Boolean domain.

THEOREM 6.1′. *Let $\mathcal{F}$ be any set of complex-valued signatures in Boolean variables. Then* Pl-#CSP$(\mathcal{F})$ *is* #P*-hard unless $\mathcal{F} \subseteq \mathscr{A}$, $\mathcal{F} \subseteq \mathscr{P}$, or $\mathcal{F} \subseteq \widehat{\mathscr{M}}$, in which case the problem is computable in polynomial time.*

Now we prove Theorem 1.1.

Proof. By Theorem 2.33, if $\mathcal{F} \subseteq \mathscr{A}$ or $\mathcal{F} \subseteq \mathscr{P}$, then #CSP$(\mathcal{F})$ is tractable over general graphs. This is category (1).

If $\mathcal{F} \subseteq \mathscr{A}$ or $\mathcal{F} \subseteq \mathscr{P}$ or $\mathcal{F} \subseteq \widehat{\mathscr{M}}$, then by Theorem 6.1′, over planar graphs the #CSP problem #CSP$(\mathcal{F})$, i.e., Pl-#CSP$(\mathcal{F})$, is tractable.

Suppose $\mathcal{F} \not\subseteq \mathscr{A}$ and $\mathcal{F} \not\subseteq \mathscr{P}$; then #CSP$(\mathcal{F})$ is #P-hard by Theorem 2.33. If further $\mathcal{F} \not\subseteq \widehat{\mathscr{M}}$, then by Theorem 6.1′, Pl-#CSP$(\mathcal{F})$ is #P-hard. This is category (3). It also implies that $\mathcal{F} \subseteq \widehat{\mathscr{M}}$ precisely captures all problems of the form #CSP$(\mathcal{F})$ in category (2), and that holographic algorithms with matchgates constitute a universal method to obtain polynomial-time algorithms for problems in the class #CSP$(\mathcal{F})$ that are #P-hard on general graphs but solvable in polynomial time on planar graphs. ∎

7. Appendix.

7.1. Ternary nonproduct type under unary actions. In this subsection, we will show how to construct a binary nonproduct signature or a symmetric nonproduct signature from a nonproduct signature of arity 3 with some unary signatures. This is the base case of the induction in the proof of Theorem 3.6.

Throughout this subsection, we are given a ternary signature f and a finite set of pairwise linearly independent unary signatures $[a_j, b_j]$ $(1 \le j \le m)$. Let $\partial^{\{i\}}_{[a_j,b_j]}(f)$ denote the binary signature obtained by connecting $[a_j, b_j]$ to the ith variable of f. For example, in matrix form, the binary signature $\partial^{\{1\}}_{[a_j,b_j]}(f)$ takes the form $\left[\begin{smallmatrix} a_j f_{000}+b_j f_{100} & a_j f_{001}+b_j f_{101} \\ a_j f_{010}+b_j f_{110} & a_j f_{011}+b_j f_{111} \end{smallmatrix}\right]$, where x_2 is the row index and x_3 is the column index. It is clear that a necessary and sufficient condition for $\partial^{\{1\}}_{[a_j,b_j]}(f) \in \mathscr{P}$ is

$$a_j f_{000} + b_j f_{100} = 0, \quad a_j f_{011} + b_j f_{111} = 0, \tag{$\partial^{\{1\}}\mathrm{D}_j$}$$

or
$$a_j f_{001} + b_j f_{101} = 0, \quad a_j f_{010} + b_j f_{110} = 0, \tag{$\partial^{\{1\}}\mathrm{E}_j$}$$

or
$$\begin{vmatrix} a_j f_{000} + b_j f_{100} & a_j f_{001} + b_j f_{101} \\ a_j f_{010} + b_j f_{110} & a_j f_{011} + b_j f_{111} \end{vmatrix} = 0. \tag{$\partial^{\{1\}} \det_j$}$$

Note the new notation $(\partial^{\{i\}}\mathrm{D}_j)$, $(\partial^{\{i\}}\mathrm{E}_j)$, $(\partial^{\{i\}} \det_j)$ we are introducing above; these are not to be confused with derivatives of signatures.

We can similarly define the conditions $(\partial^{\{i\}}\mathrm{D}_j)$, $(\partial^{\{i\}}\mathrm{E}_j)$, $(\partial^{\{i\}} \det_j)$ for $1 \le i \le 3$, $1 \le j \le m$.

LEMMA 7.1. *Let unary signatures $[a_j, b_j]$ $(1 \le j \le 2)$ be linearly independent. Suppose $\partial^{\{i\}}_{[a_j,b_j]}(f) \in \mathscr{P}$ for all $1 \le i \le 3$ and $1 \le j \le 2$. If $f(x_1, x_2, x_3) = g(x_r, x_s)h(x_t)$ where $\{r, s, t\} = \{1, 2, 3\}$, then $f \in \mathscr{P}$.*

Proof. If h is identically zero, then so is f, and the claim is trivial. Otherwise, by linear independence there exists $1 \le j \le 2$ such that $\partial^{\{t\}}_{[a_j,b_j]}(h)$ is a nonzero constant c. Then $g(x_r, x_s) = c^{-1}\partial^{\{t\}}_{[a_j,b_j]}(f) \in \mathscr{P}$. Hence $f(x_1, x_2, x_3) = g(x_r, x_s)h(x_t) \in \mathscr{P}$. ∎

LEMMA 7.2. *Let $m \ge 3$, and let unary signatures $[a_j, b_j]$ $(1 \le j \le m)$ be pairwise linearly independent. Suppose $\partial^{\{i\}}_{[a_j,b_j]}(f) \in \mathscr{P}$ for all $1 \le i \le 3$ and $1 \le j \le m$. If for some $1 \le i \le 3$,*

- *there are two distinct j such that $(\partial^{\{i\}}\mathrm{D}_j)$ hold, or*
- *there are two distinct j such that $(\partial^{\{i\}}\mathrm{E}_j)$ hold,*

then $f \in \mathscr{P}$.

Proof. By symmetry of the 3 variables we may assume $i = 1$. By pairwise linear independence of $[a_j, b_j]$ $(1 \le j \le m)$ we have either

1. $f_{000} = f_{100} = f_{011} = f_{111} = 0$ (by $(\partial^{\{1\}}\mathrm{D}_j)$ for two distinct j) or
2. $f_{001} = f_{101} = f_{010} = f_{110} = 0$ (by $(\partial^{\{1\}}\mathrm{E}_j)$ for two distinct j).

Suppose it is the first case.

By pairwise linear independence, there exists some $1 \le k \le 3$ such that $a_k \ne 0$ and $b_k \ne 0$. Consider $(\partial^{\{2\}}\mathrm{D}_k)$, $(\partial^{\{2\}}\mathrm{E}_k)$, and $(\partial^{\{2\}}\det_k)$, namely

$$\begin{aligned}
&\cancel{a_k f_{000}} + b_k f_{010} = 0, \quad a_k f_{101} + \cancel{b_k f_{111}} = 0, && (\partial^{\{2\}}\mathrm{D}_k)\\
\text{or } &a_k f_{001} + \cancel{b_k f_{011}} = 0, \quad \cancel{a_k f_{100}} + b_k f_{110} = 0, && (\partial^{\{2\}}\mathrm{E}_k)\\
\text{or } &\begin{vmatrix} \cancel{a_k f_{000}} + b_k f_{010} & a_k f_{001} + \cancel{b_k f_{011}} \\ \cancel{a_k f_{100}} + b_k f_{110} & a_k f_{101} + \cancel{b_k f_{111}} \end{vmatrix} = 0. && (\partial^{\{2\}}\det_k)
\end{aligned}$$

We have $f_{010} = f_{101} = 0$ from $(\partial^{\{2\}}\mathrm{D}_k)$ or $f_{001} = f_{110} = 0$ from $(\partial^{\{2\}}\mathrm{E}_k)$ or $\begin{vmatrix} f_{010} & f_{001} \\ f_{110} & f_{101} \end{vmatrix} = 0$ from $(\partial^{\{2\}}\det_k)$. When $f_{010} = f_{101} = 0$, together with the four vanishing entries $f_{000} = f_{100} = f_{011} = f_{111} = 0$ from $(\partial^{\{1\}}\mathrm{D}_j)$ for two distinct j, the support of f is contained in the two diagonal points $\{001, 110\}$, and hence $f \in \mathscr{P}$. Similarly, when $f_{001} = f_{110} = 0$, the support of f is contained in the two diagonal points $\{010, 101\}$, and again $f \in \mathscr{P}$. Suppose $\begin{vmatrix} f_{010} & f_{001} \\ f_{110} & f_{101} \end{vmatrix} = 0$. Then f is the product of the functions $(x_2 \ne x_3)$ and the degenerate function $g(x_1, x_3)$ with the signature in matrix form $\begin{bmatrix} f_{010} & f_{001} \\ f_{110} & f_{101} \end{bmatrix}$, where x_1 is the row index and x_3 is the column index.

The second case $f_{001} = f_{101} = f_{010} = f_{110} = 0$ is similar. We exchange all the crossed-out terms in $(\partial^{\{2\}}\mathrm{D}_k)$, $(\partial^{\{2\}}\mathrm{E}_k)$, and $(\partial^{\{2\}}\det_k)$ with the uncrossed-out terms. The conclusions from $(\partial^{\{2\}}\mathrm{D}_k)$ or from $(\partial^{\{2\}}\mathrm{E}_k)$ are still that the support of f is contained in two diagonal points. From $(\partial^{\{2\}}\det_k)$ we get $\begin{vmatrix} f_{000} & f_{011} \\ f_{100} & f_{111} \end{vmatrix} = 0$, and we conclude that f is the product of the functions $(x_2 = x_3)$ and the degenerate function $g(x_1, x_2)$ with the signature in matrix form $\begin{bmatrix} f_{000} & f_{011} \\ f_{100} & f_{111} \end{bmatrix}$, where x_1 is the row index and x_2 is the column index. ∎

LEMMA 7.3. *Let $m \ge 3$, and let unary signatures $[a_j, b_j]$ $(1 \le j \le m)$ be pairwise linearly independent. Suppose $\partial^{\{i\}}_{[a_j,b_j]}(f) \in \mathscr{P}$ for all $1 \le i \le 3$ and $1 \le j \le m$. If, for some $1 \le i \le 3$, there are three distinct j such that $(\partial^{\{i\}}\det_j)$ hold, then $f \in \mathscr{P}$.*

Proof. By symmetry of the 3 variables we may assume $i = 1$. Each $(\partial^{\{1\}}\det_j)$ is a quadratic form in a_j and b_j:

$$\begin{vmatrix} f_{000} & f_{001} \\ f_{010} & f_{011} \end{vmatrix} a_j^2 + \left(\begin{vmatrix} f_{000} & f_{101} \\ f_{010} & f_{111} \end{vmatrix} + \begin{vmatrix} f_{100} & f_{001} \\ f_{110} & f_{011} \end{vmatrix} \right) a_j b_j + \begin{vmatrix} f_{100} & f_{101} \\ f_{110} & f_{111} \end{vmatrix} b_j^2 = 0. \tag{7.1}$$

Assume $(\partial^{\{1\}} \det_j)$ holds for 3 distinct values j, k, ℓ. By pairwise linear independence, the 3×3 determinant

$$\begin{vmatrix} a_j^2 & a_j b_j & b_j^2 \\ a_k^2 & a_k b_k & b_k^2 \\ a_\ell^2 & a_\ell b_\ell & b_\ell^2 \end{vmatrix} \neq 0.$$

Indeed, if all $a_j, a_k, a_\ell \neq 0$, then the determinant is

$$a_j^2 a_k^2 a_\ell^2 \begin{vmatrix} 1 & b_j/a_j & (b_j/a_j)^2 \\ 1 & b_k/a_k & (b_k/a_k)^2 \\ 1 & b_\ell/a_\ell & (b_\ell/a_\ell)^2 \end{vmatrix},$$

where the Vandermonde determinant is nonzero because b_j/a_j, b_k/a_k, and b_ℓ/a_ℓ are pairwise distinct. If any $a_j, a_k, a_\ell = 0$, say $a_\ell = 0$, then by pairwise linear independence $a_j, a_k \neq 0$, and $b_\ell \neq 0$, and

$$\begin{vmatrix} a_j^2 & a_j b_j & b_j^2 \\ a_k^2 & a_k b_k & b_k^2 \\ a_\ell^2 & a_\ell b_\ell & b_\ell^2 \end{vmatrix} = b_\ell^2 a_j^2 a_k^2 \begin{vmatrix} 1 & b_j/a_j \\ 1 & b_k/a_k \end{vmatrix} \neq 0.$$

It follows from (7.1) that

$$\begin{vmatrix} f_{000} & f_{001} \\ f_{010} & f_{011} \end{vmatrix} = 0, \quad \begin{vmatrix} f_{000} & f_{101} \\ f_{010} & f_{111} \end{vmatrix} + \begin{vmatrix} f_{100} & f_{001} \\ f_{110} & f_{011} \end{vmatrix} = 0, \quad \begin{vmatrix} f_{100} & f_{101} \\ f_{110} & f_{111} \end{vmatrix} = 0. \tag{7.2}$$

There is a transitive group action on the four vectors

$$f_{0\bullet 0} = \begin{bmatrix} f_{000} \\ f_{010} \end{bmatrix}, \quad f_{0\bullet 1} = \begin{bmatrix} f_{001} \\ f_{011} \end{bmatrix}, \quad f_{1\bullet 0} = \begin{bmatrix} f_{100} \\ f_{110} \end{bmatrix}, \quad f_{1\bullet 1} = \begin{bmatrix} f_{101} \\ f_{111} \end{bmatrix},$$

generated by the permutations σ exchanging $f_{0\bullet 0} \leftrightarrow f_{1\bullet 0}$ and $f_{0\bullet 1} \leftrightarrow f_{1\bullet 1}$, and τ exchanging $f_{0\bullet 0} \leftrightarrow f_{0\bullet 1}$ and $f_{1\bullet 0} \leftrightarrow f_{1\bullet 1}$. This group $\mathbb{Z}_2 \times \mathbb{Z}_2$ preserves the equations (7.2). Thus either all four vectors are zero, in which case $f \in \mathscr{P}$ trivially, or we may assume $f_{0\bullet 0} \neq 0$.

By the first equation in (7.2), there exists $\lambda \in \mathbb{C}$ such that $f_{0\bullet 1} = \lambda f_{0\bullet 0}$. Substituting $f_{0\bullet 1}$ into the second equation in (7.2), we get $\begin{vmatrix} f_{000} & f_{101} - \lambda f_{100} \\ f_{010} & f_{111} - \lambda f_{110} \end{vmatrix} = 0$, and thus there exists $\mu \in \mathbb{C}$ such that $f_{1\bullet 1} - \lambda f_{1\bullet 0} = \mu f_{0\bullet 0}$. If $\mu = 0$, then $f_{1\bullet 1} = \lambda f_{1\bullet 0}$. Then f is the product of the unary function $[1, \lambda]$ on x_3 and the binary function $g(x_1, x_2)$ with the signature matrix $\begin{bmatrix} f_{000} & f_{010} \\ f_{100} & f_{110} \end{bmatrix}$, where $x_1 = 0, 1$ is the row index, and $x_2 = 0, 1$ is the column index. By Lemma 7.1 we are done.

Suppose $\mu \neq 0$. Substituting $f_{1\bullet 1} = \lambda f_{1\bullet 0} + \mu f_{0\bullet 0}$ into the third equation in (7.2), there exists $\nu \in \mathbb{C}$ such that $f_{1\bullet 0} = \nu f_{0\bullet 0}$ and $f_{1\bullet 1} = (\lambda\nu + \mu) f_{0\bullet 0}$. Hence, f is the product of the unary function $[f_{000}, f_{010}]$ on x_2 and the binary function $g(x_1, x_3)$ with the signature matrix $\begin{bmatrix} 1 & \lambda \\ \nu & \lambda\nu + \mu \end{bmatrix}$, where $x_1 = 0, 1$ is the row index, and $x_3 = 0, 1$ is the column index. By Lemma 7.1 we are done. ∎

LEMMA 7.4. *Let $m \geq 5$, and let unary signatures $[a_j, b_j]$ $(1 \leq j \leq m)$ be pairwise linearly independent. Suppose $\partial^{\{i\}}_{[a_j, b_j]}(f) \in \mathscr{P}$ for all $1 \leq i \leq 3$ and $1 \leq j \leq m$. Then $f \in \mathscr{P}$.*

Proof. For all $1 \le j \le m$, either $(\partial^{\{1\}}\mathrm{D}_j)$ or $(\partial^{\{1\}}\mathrm{E}_j)$ or $(\partial^{\{1\}}\det_j)$ holds, since $\partial^{\{1\}}_{[a_j,b_j]}(f) \in \mathscr{P}$. Since $m \ge 5$, either $(\partial^{\{1\}}\mathrm{D}_j)$ is satisfied for at least two distinct j, or $(\partial^{\{1\}}\mathrm{E}_j)$ is satisfied for at least two distinct j, or $(\partial^{\{1\}}\det_j)$ is satisfied for at least three distinct j. Hence by Lemmas 7.2 and 7.3, $f \in \mathscr{P}$. ∎

Consider the Boolean cube $\{0,1\}^3$ with its four diagonal pairs. We will name them a, b, c, d, where

$$a = (000, 111), \quad b = (001, 110), \quad c = (010, 101), \quad d = (011, 100),$$

respectively. A consequence of each statement $(\partial^{\{i\}}\mathrm{D}_j)$ and $(\partial^{\{i\}}\mathrm{E}_j)$ is that for some two diagonal pairs, the values of f at the diagonal pairs have the same product; e.g., for any j, $(\partial^{\{1\}}\mathrm{D}_j)$ implies that the product $f_{000}f_{111} = f_{011}f_{100}$; i.e., the product of the values of f at the diagonal a is the same as that at the diagonal d. Similarly, the statement $(\partial^{\{1\}}\mathrm{E}_j)$ implies that the product $f_{001}f_{110} = f_{010}f_{101}$; i.e., the product of the values of f at the diagonal b is the same as that at the diagonal c. For $(\partial^{\{2\}}\mathrm{D}_j)$ (resp., $(\partial^{\{2\}}\mathrm{E}_j)$), the implications are for the diagonals a and c (resp., b and d). For $(\partial^{\{3\}}\mathrm{D}_j)$ (resp., $(\partial^{\{3\}}\mathrm{E}_j)$), the implications are for the diagonals a and b (resp., c and d).

Define a graph on the vertex set $\{a, b, c, d\}$ where we add an edge whenever the corresponding diagonals have the same product value of f, and we get a spanning subgraph (a subgraph containing all four vertices) of K_4 on $\{a, b, c, d\}$. We remark that since its edge relation is defined by equality of values, any connected component of this spanning subgraph is a clique. If there are at least four edges in this spanning subgraph, then it is connected (which is equivalent to this spanning subgraph being just K_4), with the implication that all diagonals have the same product value of f. If this spanning subgraph is connected, then each statement $(\partial^{\{i\}}\det_j)$ takes the form

$$D_0 a_j^2 + D_2 b_j^2 = 0 \tag{7.3}$$

for some coefficients D_0 and D_2 with a zero coefficient of the cross term $a_j b_j$. Note that D_0 and D_2 are the coefficients of a_j^2 and b_j^2, respectively, in $(\partial^{\{i\}}\det_j)$ and do not depend on j. They depend on i, but D_0 (resp., D_2) is the same for all j. For example, for $(\partial^{\{1\}}\det_j)$ we have

$$D_0 = \begin{vmatrix} f_{000} & f_{001} \\ f_{010} & f_{011} \end{vmatrix} \quad \text{and} \quad D_2 = \begin{vmatrix} f_{100} & f_{101} \\ f_{110} & f_{111} \end{vmatrix}.$$

Moreover, the only disconnected spanning subgraph with three edges in K_4 is a triangle (plus an isolated vertex), meaning that the three statements among all $(\partial^{\{i\}}\mathrm{D}_j)$ and $(\partial^{\{i\}}\mathrm{E}_j)$ that hold must be those with implications among only three letters out of four $\{a, b, c, d\}$. For example, if the triangle is on $\{a, b, c\}$, then the three statements must be among $(\partial^{\{1\}}\mathrm{E}_j)$, $(\partial^{\{2\}}\mathrm{D}_j)$, and $(\partial^{\{3\}}\mathrm{D}_j)$ (but not $(\partial^{\{1\}}\mathrm{D}_j)$, not $(\partial^{\{2\}}\mathrm{E}_j)$, and not $(\partial^{\{3\}}\mathrm{E}_j)$). Similarly if the triangle is on $\{b, c, d\}$, then the three statements must be among $(\partial^{\{1\}}\mathrm{E}_j)$, $(\partial^{\{2\}}\mathrm{E}_j)$, and $(\partial^{\{3\}}\mathrm{E}_j)$ (but not $(\partial^{\{1\}}\mathrm{D}_j)$, not $(\partial^{\{2\}}\mathrm{D}_j)$, and not $(\partial^{\{3\}}\mathrm{D}_j)$).

LEMMA 7.5. *Suppose the ternary function $f \notin \mathscr{P}$. Let $[1, b_j]$ $(1 \le j \le 4)$ be the unary signatures $[1, 1], [1, -1], [1, \mathrm{i}], [1, -\mathrm{i}]$, respectively. Suppose $\partial^{\{i\}}_{[1,b_j]}(f) \in \mathscr{P}$ for all $1 \le i \le 3$ and $1 \le j \le 4$. Then, up to a nonzero constant factor, f is a product of some subset of unary functions $[1, b_j]$ and the symmetric ternary function $[1, 1, -1, -1]$.*

Proof. By Lemmas 7.2 and 7.3, we may assume that for each $1 \le i \le 3$ there can be at most one j such that $(\partial^{\{i\}}\mathrm{D}_j)$ holds, at most one j' such that $(\partial^{\{i\}}\mathrm{E}_{j'})$ holds, and at most two distinct values k and k' such that $(\partial^{\{i\}}\det_k)$ and $(\partial^{\{i\}}\det_{k'})$ hold. Moreover, since there are four such requirements that must be satisfied altogether for the same i, there exist exactly one such j, j', k, and k', respectively, and $\{j, j', k, k'\} = \{1, 2, 3, 4\}$.

The spanning subgraph of K_4 in this case is the full graph K_4, and all diagonals have the same product value. To see this, note that the edge $\{a, d\}$ exists because $(\partial^{\{1\}}\mathrm{D}_j)$ for some j; $\{a, c\}$ exists because $(\partial^{\{2\}}\mathrm{D}_j)$ also for some j; and $\{a, b\}$ exists because $(\partial^{\{3\}}\mathrm{D}_j)$ yet again for some j.

For any $1 \le i \le 3$ if the two valid determinantal statements $(\partial^{\{i\}}\det_k)$ and $(\partial^{\{i\}}\det_{k'})$ are not for the pair $\{[1,1],[1,-1]\}$ or $\{[1,\mathrm{i}],[1,-\mathrm{i}]\}$, then we have $b_k^2 \ne b_{k'}^2$. Note that in this case $a_k = a_{k'} = 1$, and since (7.3) holds for two distinct indices k and k' (and D_0, D_2 are the same for these two k and k'), we conclude that $D_0 = D_2 = 0$ for this i. Then $(\partial^{\{i\}}\det_\ell)$, which has the form (7.3) with the vanished cross term and index ℓ, must hold for all $1 \le \ell \le 4$. By Lemma 7.3 we are done. Hence we may assume that the pair for which $(\partial^{\{i\}}\det_k)$ and $(\partial^{\{i\}}\det_{k'})$ hold is $\{[1,b_k],[1,b_{k'}]\} = \{[1,1],[1,-1]\}$ or $\{[1,\mathrm{i}],[1,-\mathrm{i}]\}$. Then we have $b_k = \pm b_{k'}$. Furthermore, $b_k \ne b_{k'}$ because $\{j, j', k, k'\} = \{1, 2, 3, 4\}$; hence $b_k = -b_{k'} \in \{1, -1\}$ or $b_k = -b_{k'} \in \{\mathrm{i}, -\mathrm{i}\}$. This also implies that $b_j = -b_{j'}$.

For the valid $(\partial^{\{1\}}\mathrm{D}_j)$ let $x = -b_j \in \{1, -1, \mathrm{i}, -\mathrm{i}\}$. This is the multiplier for which $f_{000} = xf_{100}$ and $f_{011} = xf_{111}$. The corresponding multiplier for $(\partial^{\{1\}}\mathrm{E}_{j'})$ is $-x$ such that $f_{001} = -xf_{101}$ and $f_{010} = -xf_{110}$. Similarly, for the valid $(\partial^{\{2\}}\mathrm{D}_j)$ we define $y \in \{1, -1, \mathrm{i}, -\mathrm{i}\}$ such that $f_{000} = yf_{010}$ and $f_{101} = yf_{111}$. Also $f_{001} = -yf_{011}$ and $f_{100} = -yf_{110}$. Finally, for the valid $(\partial^{\{3\}}\mathrm{D}_j)$ we define $z \in \{1, -1, \mathrm{i}, -\mathrm{i}\}$ such that $f_{000} = zf_{001}$ and $f_{110} = zf_{111}$. Also $f_{010} = -zf_{011}$ and $f_{100} = -zf_{101}$. Note that $x, y, z \in \{1, -1, \mathrm{i}, -\mathrm{i}\}$ are nonzero complex numbers.

Let $g(x_1, x_2, x_3)$ be the product of three unary functions $g_1(x_1)g_2(x_2)g_3(x_3)$, where $g_1(x_1) = [x, 1]$ on x_1, $g_2(x_2) = [y, 1]$ on x_2, and $g_3(x_3) = [z, 1]$ on x_3. If we denote $g_1(x_1)$ by $[x, 1]_{x_1}$ to indicate that the unary function $[x, 1]$ is on the variable x_1, and similarly for $g_2(x_2)$ and $g_3(x_3)$, we can write $g = [x, 1]_{x_1} \otimes [y, 1]_{x_2} \otimes [z, 1]_{x_3}$. Note that each of g_1, g_2, g_3 is a nonzero constant multiple of some $[1, b_j]$ $(1 \le j \le 4)$.

Finally, it is easy to verify that $f = gh$, where h is the ternary symmetric function $f_{000}[1, 1, -1, -1]$. ∎

COROLLARY 7.6. *Let $\mathcal{F}$ be a set of signatures containing a ternary signature $f \notin \mathscr{P}$. Suppose $\mathcal{F}$ contains the unary signatures $\{[1, b_j] \mid 1 \le j \le 4\} = \{[1,1],[1,-1],[1,\mathrm{i}],[1,-\mathrm{i}]\}$, and $\partial^{\{i\}}_{[1,b_j]}(f) \in \mathscr{P}$ for all $1 \le i \le 3$ and $1 \le j \le 4$. Then*

$$\text{Pl-\#CSP}(\mathcal{F} \cup \{[1, 1, -1, -1]\}) \le_{\mathrm{T}} \text{Pl-\#CSP}(\mathcal{F}).$$

Proof. By Lemma 7.5, up to a nonzero constant factor, f is a product of $[1, 1, -1, -1]$ and some unary functions from $\{[1,1],[1,-1],[1,\mathrm{i}],[1,-\mathrm{i}]\}$. For any instance I of Pl-#CSP($\mathcal{F} \cup \{[1, 1, -1, -1]\}$), every occurrence of $[1, 1, -1, -1]$ can be replaced by f together with some unary functions from $\{[1,1],[1,-1],[1,\mathrm{i}],[1,-\mathrm{i}]\}$. This is because the two product functions, $[1, \pm 1](x_i) \cdot [1, \pm 1](x_i)$ and $[1, \pm\mathrm{i}](x_i) \cdot [1, \mp\mathrm{i}](x_i)$, each a product of two unary functions on the same variable, are both the constant 1 function on x_i. ∎

We denote by $\omega = e^{\mathrm{i}\frac{2\pi}{3}}$ a primitive third root of unity.

LEMMA 7.7. *Suppose $f \notin \mathscr{P}$. Let $[1, b_j]$ $(1 \le j \le 3)$ be the unary signatures*

$[1,1],[1,\omega],[1,\omega^2]$, *respectively. Suppose* $\partial^{\{i\}}_{[1,b_j]}(f) \in \mathscr{P}$ *for all* $1 \le i \le 3$ *and* $1 \le j \le 3$. *Then* f *is a product of some unary functions* $[1,b_j]$ *with*

1. *the symmetric function* $[1,-1,x,-x]$ *where* $x \in \{\omega,\omega^2\}$, *or*
2. *the symmetric function* $[-2,1,1,-2]$, *or*
3. *after a cyclic permutation of its three variables a ternary function* $g(1-x_1,x_2,x_3)$ *where* $g(x_1,x_2,x_3)$ *is the symmetric function* $[-2,1,1,-2]$.

Proof. By Lemmas 7.2 and 7.3, we may assume that for each $1 \le i \le 3$ there can be at most one j such that $(\partial^{\{i\}}\mathrm{D}_j)$ holds, at most one j' such that $(\partial^{\{i\}}\mathrm{E}_{j'})$ holds, and at most two distinct values k and k' such that $(\partial^{\{i\}}\det_k)$ and $(\partial^{\{i\}}\det_{k'})$ hold.

Let N be the total number of valid statements among $(\partial^{\{i\}}\mathrm{D}_j)$ and $(\partial^{\{i\}}\mathrm{E}_j)$. If $N \le 2$, then for some $1 \le i \le 3$, all three statements $(\partial^{\{i\}}\det_j)$ for $1 \le j \le 3$ must hold. Hence $N \ge 3$.

We first assume that the spanning subgraph of K_4 defined by the valid statements $(\partial^{\{i\}}\mathrm{D}_j)$ and $(\partial^{\{i\}}\mathrm{E}_j)$ is connected. In particular, if $N \ge 4$, then this is the case. In this case, all statements $(\partial^{\{i\}}\det_j)$ are of the form (7.3) with a vanishing cross term. For any two distinct $b_j, b_k \in \{1,\omega,\omega^2\}$, $b_j^2 \ne b_k^2$, therefore for any $1 \le i \le 3$, if there are two distinct valid statements $(\partial^{\{i\}}\det_j)$ and $(\partial^{\{i\}}\det_{j'})$ $(j \ne j')$, then $(\partial^{\{i\}}\det_k)$ is valid for all $1 \le k \le 3$. By Lemma 7.3 we reach a contradiction to $f \notin \mathscr{P}$. Therefore, there cannot be more than one valid $(\partial^{\{i\}}\det_j)$ for each $1 \le i \le 3$. It follows that for every $1 \le i \le 3$, there is exactly one valid $(\partial^{\{i\}}\mathrm{D}_j)$, one valid $(\partial^{\{i\}}\mathrm{E}_k)$, and one valid $(\partial^{\{i\}}\det_\ell)$, such that $\{j,k,\ell\} = \{1,2,3\}$.

For the valid $(\partial^{\{1\}}\mathrm{D}_j)$, let $x = b_j \in \{1,\omega,\omega^2\}$; then $f_{000} = -xf_{100}$ and $f_{011} = -xf_{111}$. The corresponding multiplier for the valid $(\partial^{\{1\}}\mathrm{E}_k)$ is x' where $x' = x\omega$ or $x\omega^2$, such that $f_{001} = -x'f_{101}$ and $f_{010} = -x'f_{110}$. Similarly, for the valid $(\partial^{\{2\}}\mathrm{D}_j)$ we define $y \in \{1,\omega,\omega^2\}$ such that $f_{000} = -yf_{010}$ and $f_{101} = -yf_{111}$. Also $f_{001} = -y'f_{011}$ and $f_{100} = -y'f_{110}$, where $y' = y\omega$ or $y\omega^2$. Finally, for the valid $(\partial^{\{3\}}\mathrm{D}_j)$ we define $z \in \{1,\omega,\omega^2\}$ such that $f_{000} = -zf_{001}$ and $f_{110} = -zf_{111}$. Also $f_{010} = -z'f_{011}$ and $f_{100} = -z'f_{101}$, where $z' = z\omega$ or $z\omega^2$. Clearly, if any value $f_{abc} = 0$, then f is identically 0, a contradiction to $f \notin \mathscr{P}$. So we may assume that f has no zero values. By consistency of values, $f_{001} = -x'f_{101} = x'yf_{111}$ and $f_{001} = -y'f_{011} = xy'f_{111}$; hence $x'/x = y'/y$. Similarly, $f_{000} = -zf_{001} = y'zf_{011}$ and $f_{000} = -yf_{010} = yz'f_{011}$; hence $y'/y = z'/z$.

Let $\rho = x'/x = y'/y = z'/z \in \{\omega,\omega^2\}$. Let $g(x_1,x_2,x_3)$ be the product function $g_1(x_1)g_2(x_2)g_3(x_3)$ where $g_1(x_1) = [-x,1]_{x_1}, g_2(x_2) = [-y,1]_{x_2}, g_3(x_3) = [-z,1]_{x_3}$, i.e., $g = [-x,1]_{x_1} \otimes [-y,1]_{x_2} \otimes [-z,1]_{x_3}$. Then $f = gh$ where h is the ternary symmetric function $f_{111}[-\rho,\rho,-1,1]$. Alternatively we have $h = f_{000}[1,-1,\xi,-\xi]$, where $\xi = 1/\rho \in \{\omega,\omega^2\}$.

Now suppose $N = 3$ and the spanning subgraph with three edges in K_4 is a triangle (together with an isolated vertex). Then for each $1 \le i \le 3$ there are exactly two distinct values k and k' such that $(\partial^{\{i\}}\det_k)$ and $(\partial^{\{i\}}\det_{k'})$ hold. The triangle of the spanning subgraph is either the triangle on $\{b,c,d\}$ or a triangle involving the vertex a, in which case by a cyclic permutation of the three variables, we may assume the triangle is on $\{a,b,c\}$. For the triangle on $\{b,c,d\}$ the three valid statements among all $(\partial^{\{i\}}\mathrm{D}_j)$ and $(\partial^{\{i\}}\mathrm{E}_j)$ must be among $(\partial^{\{1\}}\mathrm{E}_j)$, $(\partial^{\{2\}}\mathrm{E}_j)$, and $(\partial^{\{3\}}\mathrm{E}_j)$. For the triangle on $\{a,b,c\}$ the three valid statements among all $(\partial^{\{i\}}\mathrm{D}_j)$ and $(\partial^{\{i\}}\mathrm{E}_j)$ must be among $(\partial^{\{1\}}\mathrm{E}_j)$, $(\partial^{\{2\}}\mathrm{D}_j)$, and $(\partial^{\{3\}}\mathrm{D}_j)$.

We first consider the triangle $\{b,c,d\}$ case. According to the valid $(\partial^{\{1\}}\mathrm{E}_j)$, we let $x = b_j \in \{1,\omega,\omega^2\}$; then $f_{001} = -xf_{101}$ and $f_{010} = -xf_{110}$. Then the two valid $(\partial^{\{1\}}\det_k)$ and $(\partial^{\{1\}}\det_\ell)$ hold for $b_k = b_j\omega$ and $b_\ell = b_j\omega^2$. Hence the following

equation has two roots $X = x\omega$ and $X = x\omega^2$:

$$\begin{vmatrix} f_{000} & f_{001} \\ f_{010} & f_{011} \end{vmatrix} + \begin{vmatrix} f_{000} & f_{101} \\ f_{010} & f_{111} \end{vmatrix} X + \begin{vmatrix} f_{100} & f_{101} \\ f_{110} & f_{111} \end{vmatrix} X^2 = 0. \tag{7.4}$$

Notice that we have used the fact that one cross term is zero: $\left| \begin{smallmatrix} f_{100} & f_{001} \\ f_{110} & f_{011} \end{smallmatrix} \right| = 0$, because the diagonal b and d have the same product value $f_{001}f_{110} = f_{011}f_{100}$. Subtracting one equation from another in (7.4) with $X = x\omega$ and $X = x\omega^2$, we get

$$\begin{vmatrix} f_{000} & f_{101} \\ f_{010} & f_{111} \end{vmatrix} = x \begin{vmatrix} f_{100} & f_{101} \\ f_{110} & f_{111} \end{vmatrix}. \tag{7.5}$$

Similarly, we have a valid $(\partial^{\{2\}}\mathrm{E}_j)$ for some j according to which we let $y = b_j \in \{1, \omega, \omega^2\}$, and then $f_{001} = -yf_{011}$ and $f_{100} = -yf_{110}$. Also, the two statements $(\partial^{\{2\}} \det_k)$ and $(\partial^{\{2\}} \det_\ell)$ hold for $b_k = b_j\omega$ and $b_\ell = b_j\omega^2$. We have a valid $(\partial^{\{3\}}\mathrm{E}_j)$ for some j according to which we let $z = b_j \in \{1, \omega, \omega^2\}$, and then $f_{100} = -zf_{101}$ and $f_{010} = -zf_{011}$. The two statements $(\partial^{\{3\}} \det_k)$ and $(\partial^{\{2\}} \det_\ell)$ hold for $b_k = b_j\omega$ and $b_\ell = b_j\omega^2$.

It follows that

$$f_{001} = -xf_{101}, \quad f_{100} = -zf_{101}, \quad f_{110} = \frac{z}{y}f_{101}, \quad f_{011} = \frac{x}{y}f_{101}, \quad f_{010} = -\frac{xz}{y}f_{101}.$$

Let $g(x_1, x_2, x_3) = f(x_1, x_2, x_3)/\left([-x, 1]_{x_1} \otimes [-y, 1]_{x_2} \otimes [-z, 1]_{x_3}\right)$; then $g_{000} = f_{000}/(-xyz)$, $g_{111} = f_{111}$, and $g_{001} = f_{001}/(xy) = (-1/y)f_{101} = g_{101}$. Similarly, we can show

$$g_{001} = g_{100} = g_{110} = g_{011} = g_{010} = g_{101}.$$

If $g_{101} = 0$, then $g \in \mathscr{P}$, and so is f, a contradiction to $f \notin \mathscr{P}$. Hence we may normalize by a constant and assume that $g_{101} = 1$.

After some computation, (7.5) simplifies to

$$YZ + Z - 2 = 0, \tag{7.6}$$

where $Y = g_{000}$ and $Z = g_{111}$. Equation (7.4) for the root $X = x\omega$ simplifies to

$$Y - 1 + (1 - YZ)\omega + (Z - 1)\omega^2 = 0, \tag{7.7}$$

where Y and Z are as above. If we substitute $1 - YZ = Z - 1$ from (7.6) into (7.7), we get $Y = Z$. Substituting this back into (7.6), we get $(Y - 1)(Y + 2) = 0$, and $Y = 1$ or $Y = -2$. The solution $Y = 1$ gives a degenerate g and hence f, a contradiction to $f \notin \mathscr{P}$. The solution $Y = Z = -2$ gives $g = [-2, 1, 1, -2]$. This gives

$$f(x_1, x_2, x_3) = \left([-x, 1]_{x_1} \otimes [-y, 1]_{x_2} \otimes [-z, 1]_{x_3}\right) [-2, 1, 1, -2],$$

where $x, y, z \in \{1, \omega, \omega^2\}$.

The last case is that $N = 3$ and the spanning subgraph with three edges in K_4 is the triangle on $\{a, b, c\}$ (with an isolated vertex d) after a cyclic permutation of the three variables. By flipping x_1 with its negation $\overline{x_1}$ we can invoke what has been proved for the triangle $\{b, c, d\}$ case and conclude that $f(x_1, x_2, x_3)$ is a product of some unary functions with the function $g(\overline{x_1}, x_2, x_3)$, where $g(x_1, x_2, x_3)$ is the symmetric function $[-2, 1, 1, -2]$. ∎

We remark that since we are interested in *planar* #CSP problems, we may not use arbitrary permutation of variables. In the proof above, whenever the conclusion is symmetric in all three variables, the argument can apply an arbitrary permutation in the proof without loss of generality. However, if the conclusion is not symmetric in all three variables, we may only apply a cyclic permutation in the proof, as in the last case in Lemma 7.7 with the triangle on $\{a, b, c\}$. Notice that the function $g(\overline{x_1}, x_2, x_3)$ has the signature matrix $\left[\begin{smallmatrix} 1 & 1 & 1 & -2 \\ -2 & 1 & 1 & 1 \end{smallmatrix}\right]$, where $x_1 = 0, 1$ is the row index, and $x_2x_3 = 00, 01, 10, 11$ is the column index. If we connect two copies of $g(\overline{x_1}, x_2, x_3)$ with both x_1 as external edges, and the variable x_2 of one copy connected to the x_3 of the other copy, for both pairs of (x_2, x_3), we obtain a planar gadget with a symmetric signature not in $\mathscr{P}$ with its signature matrix

$$\begin{bmatrix} 1 & 1 & 1 & -2 \\ -2 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & -2 \\ 1 & 1 \\ 1 & 1 \\ -2 & 1 \end{bmatrix} = \begin{bmatrix} 7 & -2 \\ -2 & 7 \end{bmatrix} \notin \mathscr{P}.$$

COROLLARY 7.8. *Let $\mathcal{F}$ be a set of signatures containing a ternary signature $f \notin \mathscr{P}$. Suppose $\mathcal{F}$ contains the unary signatures $\{[1, b_j] \mid 1 \le j \le 3\} = \{[1, 1], [1, \omega], [1, \omega^2]\}$, and $\partial^{\{i\}}_{[1,b_j]}(f) \in \mathscr{P}$ for all $1 \le i \le 3$ and $1 \le j \le 3$. Then there exists $g \in \{[1, -1, \omega, -\omega], [1, -1, \omega^2, -\omega^2], [-2, 1, 1, -2], [7, -2, 7]\}$ such that*

$$\text{Pl-\#CSP}(\mathcal{F} \cup \{g\}) \le_{\text{T}} \text{Pl-\#CSP}(\mathcal{F}).$$

LEMMA 7.9. *Suppose $f \notin \mathscr{P}$. Let $[1, b_j]$ ($1 \le j \le 3$) be the unary signatures $[1, 0], [1, 1], [1, -1]$, respectively. Suppose $\partial^{\{i\}}_{[1,b_j]}(f) \in \mathscr{P}$ for all $1 \le i \le 3$ and $1 \le j \le 3$. Then, after a cyclic permutation of its three variables, f is a product of some unary functions $[1, b_j]$ with the symmetric function $[1, 0, 1, 0]$ or $[0, 1, 0, 1]$.*

Proof. The requirements for $\partial^{\{i\}}_{[1,b_j]}(f) \in \mathscr{P}$ for all $1 \le i \le 3$ and $1 \le j \le 3$ are listed below, where each line is for one (i, j). More specifically, the condition $\partial^{\{i\}}_{[1,b_j]}(f) \in \mathscr{P}$ is expressed by the disjunction $(\partial^{\{i\}}\text{D}_j) \vee (\partial^{\{i\}}\text{E}_j) \vee (\partial^{\{i\}} \det_j)$:

$$\begin{array}{llcllcl}
f_{000} = f_{011} = 0 & (\partial^{\{1\}}\text{D}_1) & \text{or} & f_{001} = f_{010} = 0 & (\partial^{\{1\}}\text{E}_1) & \text{or} & \begin{vmatrix} f_{000} & f_{001} \\ f_{010} & f_{011} \end{vmatrix} = 0 \quad (\partial^{\{1\}} \det_1) \\
f_{000} + f_{100} = f_{011} + f_{111} = 0 & (\partial^{\{1\}}\text{D}_2) & \text{or} & f_{001} + f_{101} = f_{010} + f_{110} = 0 & (\partial^{\{1\}}\text{E}_2) & \text{or} & \begin{vmatrix} f_{000} + f_{100} & f_{001} + f_{101} \\ f_{010} + f_{110} & f_{011} + f_{111} \end{vmatrix} = 0 \quad (\partial^{\{1\}} \det_2) \\
f_{000} - f_{100} = f_{011} - f_{111} = 0 & (\partial^{\{1\}}\text{D}_3) & \text{or} & f_{001} - f_{101} = f_{010} - f_{110} = 0 & (\partial^{\{1\}}\text{E}_3) & \text{or} & \begin{vmatrix} f_{000} - f_{100} & f_{001} - f_{101} \\ f_{010} - f_{110} & f_{011} - f_{111} \end{vmatrix} = 0 \quad (\partial^{\{1\}} \det_3) \\
f_{000} = f_{101} = 0 & (\partial^{\{2\}}\text{D}_1) & \text{or} & f_{001} = f_{100} = 0 & (\partial^{\{2\}}\text{E}_1) & \text{or} & \begin{vmatrix} f_{000} & f_{001} \\ f_{100} & f_{101} \end{vmatrix} = 0 \quad (\partial^{\{2\}} \det_1) \\
f_{000} + f_{010} = f_{101} + f_{111} = 0 & (\partial^{\{2\}}\text{D}_2) & \text{or} & f_{001} + f_{011} = f_{100} + f_{110} = 0 & (\partial^{\{2\}}\text{E}_2) & \text{or} & \begin{vmatrix} f_{000} + f_{010} & f_{001} + f_{011} \\ f_{100} + f_{110} & f_{101} + f_{111} \end{vmatrix} = 0 \quad (\partial^{\{2\}} \det_2) \\
f_{000} - f_{010} = f_{101} - f_{111} = 0 & (\partial^{\{2\}}\text{D}_3) & \text{or} & f_{001} - f_{011} = f_{100} - f_{110} = 0 & (\partial^{\{2\}}\text{E}_3) & \text{or} & \begin{vmatrix} f_{000} - f_{010} & f_{001} - f_{011} \\ f_{100} - f_{110} & f_{101} - f_{111} \end{vmatrix} = 0 \quad (\partial^{\{2\}} \det_3) \\
f_{000} = f_{110} = 0 & (\partial^{\{3\}}\text{D}_1) & \text{or} & f_{010} = f_{100} = 0 & (\partial^{\{3\}}\text{E}_1) & \text{or} & \begin{vmatrix} f_{000} & f_{010} \\ f_{100} & f_{110} \end{vmatrix} = 0 \quad (\partial^{\{3\}} \det_1) \\
f_{000} + f_{001} = f_{110} + f_{111} = 0 & (\partial^{\{3\}}\text{D}_2) & \text{or} & f_{010} + f_{011} = f_{100} + f_{101} = 0 & (\partial^{\{3\}}\text{E}_2) & \text{or} & \begin{vmatrix} f_{000} + f_{001} & f_{010} + f_{011} \\ f_{100} + f_{101} & f_{110} + f_{111} \end{vmatrix} = 0 \quad (\partial^{\{3\}} \det_2) \\
f_{000} - f_{001} = f_{110} - f_{111} = 0 & (\partial^{\{3\}}\text{D}_3) & \text{or} & f_{010} - f_{011} = f_{100} - f_{101} = 0 & (\partial^{\{3\}}\text{E}_3) & \text{or} & \begin{vmatrix} f_{000} - f_{001} & f_{010} - f_{011} \\ f_{100} - f_{101} & f_{110} - f_{111} \end{vmatrix} = 0 \quad (\partial^{\{3\}} \det_3)
\end{array}$$

By Lemma 7.2, we may assume that for each $1 \le i \le 3$ there can be at most one j such that $(\partial^{\{i\}}\text{D}_j)$ holds, and at most one j' such that $(\partial^{\{i\}}\text{E}_{j'})$ holds. This implies that for every i there is at least one j such that $(\partial^{\{i\}} \det_j)$ holds. By Lemma 7.3, we may assume that for each $1 \le i \le 3$ there are at most two distinct values k and k' such that $(\partial^{\{i\}} \det_k)$ and $(\partial^{\{i\}} \det_{k'})$ hold.

We first suppose the spanning subgraph of K_4 is connected. This implies that all diagonal pairs have the same product value. In that case, the statements $(\partial^{\{i\}} \det_2)$

and $(\partial^{\{i\}} \det_3)$ are identical. Thus if $(\partial^{\{i\}} \det_1)$ holds then we may assume $(\partial^{\{i\}} \det_2)$ and $(\partial^{\{i\}} \det_3)$ do not hold. On the other hand, if $(\partial^{\{i\}} \det_1)$ does not hold, then $(\partial^{\{i\}} \det_2)$ and $(\partial^{\{i\}} \det_3)$ must hold by Lemma 7.2.

1. Suppose there exists some $1 \le i \le 3$ such that either $(\partial^{\{i\}}\mathrm{D}_1)$ or $(\partial^{\{i\}}\mathrm{E}_1)$ holds. By cyclically permuting the variables we may assume $i = 1$.
   - Case where $(\partial^{\{1\}}\mathrm{D}_1)$ holds.
     If $(\partial^{\{1\}} \det_2)$, which is equivalent to $(\partial^{\{1\}} \det_3)$, does not hold, then two more valid statements must hold among $\{(\partial^{\{1\}}\mathrm{D}_2), (\partial^{\{1\}}\mathrm{E}_2), (\partial^{\{1\}}\mathrm{D}_3), (\partial^{\{1\}}\mathrm{E}_3)\}$, and together with $(\partial^{\{1\}}\mathrm{D}_1)$, we are done by Lemma 7.2. Thus $(\partial^{\{1\}} \det_2)$ holds.
     We have $f_{000} = f_{011} = 0$ by $(\partial^{\{1\}}\mathrm{D}_1)$. If we have additionally $f_{001} = 0$ and $f_{010} = 0$, then $f^{x_1=0}$ is identically 0, and $f(x_1x_2, x_3) = g(x_2, x_3)[0, 1]_{x_1}$ for some binary function g. By Lemma 7.1 we reach a contradiction to $f \notin \mathscr{P}$. So we assume f_{001} and f_{010} are not both zero.
     If $(\partial^{\{2\}} \det_1)$ holds, then we must have either $(\partial^{\{2\}}\mathrm{D}_2)$ and $(\partial^{\{2\}}\mathrm{E}_3)$ or $(\partial^{\{2\}}\mathrm{D}_3)$ and $(\partial^{\{2\}}\mathrm{E}_2)$. In either case, by $(\partial^{\{2\}}\mathrm{D}_2)$ or $(\partial^{\{2\}}\mathrm{D}_3)$ it easily follows that $f_{010} = 0$ and by $(\partial^{\{2\}}\mathrm{E}_3)$ or $(\partial^{\{2\}}\mathrm{E}_2)$ that $f_{001} = 0$. But this is a contradiction to the statement that f_{001} and f_{010} are not both zero, from the previous paragraph. This contradiction proves that in fact $(\partial^{\{2\}} \det_1)$ does not hold. Similarly, if $(\partial^{\{3\}} \det_1)$ holds, we reach the same contradiction. Thus $(\partial^{\{3\}} \det_1)$ does not hold.
     Therefore, we must have $(\partial^{\{2\}} \det_2)$ which is identical to $(\partial^{\{2\}} \det_3)$, and also $(\partial^{\{3\}} \det_2)$ which is identical to $(\partial^{\{3\}} \det_3)$, in addition to $(\partial^{\{1\}} \det_2)$.
     These statements take the form

$$\begin{aligned}(7.8)\quad \begin{vmatrix} f_{000}+f_{100} & f_{001}+f_{101} \\ f_{010}+f_{110} & f_{011}+f_{111} \end{vmatrix} &= \begin{vmatrix} f_{000} & f_{001} \\ f_{010} & f_{011} \end{vmatrix} + \begin{vmatrix} f_{100} & f_{101} \\ f_{110} & f_{111} \end{vmatrix} \\ &= -f_{001}f_{010} + f_{100}f_{111} - f_{101}f_{110} = 0,\end{aligned}$$

$$\begin{aligned}(7.9)\quad \begin{vmatrix} f_{000}+f_{010} & f_{001}+f_{011} \\ f_{100}+f_{110} & f_{101}+f_{111} \end{vmatrix} &= \begin{vmatrix} f_{000} & f_{001} \\ f_{100} & f_{101} \end{vmatrix} + \begin{vmatrix} f_{010} & f_{011} \\ f_{110} & f_{111} \end{vmatrix} \\ &= -f_{001}f_{100} + f_{010}f_{111} = 0,\end{aligned}$$

$$\begin{aligned}(7.10)\quad \begin{vmatrix} f_{000}+f_{001} & f_{010}+f_{011} \\ f_{100}+f_{101} & f_{110}+f_{111} \end{vmatrix} &= \begin{vmatrix} f_{000} & f_{010} \\ f_{100} & f_{110} \end{vmatrix} + \begin{vmatrix} f_{001} & f_{011} \\ f_{101} & f_{111} \end{vmatrix} \\ &= -f_{010}f_{100} + f_{001}f_{111} = 0.\end{aligned}$$

     If $f_{001} = 0$, then $f_{010}f_{111} = 0$ from (7.9) and $f_{010}f_{100} = 0$ from (7.10). Since f_{001} and f_{010} are not both zero, we have $f_{010} \neq 0$ in this case, and we conclude that $f_{111} = f_{100} = 0$. Then f is the product of $(x_2 \neq x_3)$ with the degenerate binary function $g(x_1, x_3) = \begin{bmatrix} f_{010} & f_{001} \\ f_{110} & f_{101} \end{bmatrix}$ with row index $x_1 = 0, 1$ and column index $x_3 = 0, 1$, and $\begin{vmatrix} f_{010} & f_{001} \\ f_{110} & f_{101} \end{vmatrix} = 0$ from (7.9). This is a contradiction to $f \notin \mathscr{P}$.
     If $f_{010} = 0$, then we also get the same conclusion. So we assume that both $f_{001} \neq 0$ and $f_{010} \neq 0$. Then from (7.9) and (7.10) we get $f_{100} = \frac{f_{010}}{f_{001}} f_{111} = \frac{f_{001}}{f_{010}} f_{111}$. If $f_{111} = 0$, then $f_{100} = 0$, and we have $f = (x_2 \neq x_3)g(x_1, x_3)$ for a degenerate binary function g as before. Therefore, we may assume that $f_{111} \neq 0$; then $f_{100} \neq 0$ as well. Then $(f_{010})^2 = (f_{001})^2$;

thus $f_{010} = \epsilon f_{001}$, where $\epsilon = \pm 1$. Also $f_{100} = \epsilon f_{111}$.
Since all diagonal pairs have the same product value, $f_{001}f_{110} = f_{010}f_{101} = f_{000}f_{111} = 0$. As $f_{001} \neq 0$ and $f_{010} \neq 0$, we have $f_{110} = f_{101} = 0$. Then from (7.8) we have $\left|\begin{smallmatrix} f_{100} & f_{001} \\ f_{010} & f_{111} \end{smallmatrix}\right| = \left|\begin{smallmatrix} \epsilon f_{111} & f_{001} \\ \epsilon f_{001} & f_{111} \end{smallmatrix}\right| = 0$. It follows that $f_{001} = \epsilon^* f_{111}$, where $\epsilon^* = \pm 1$. Thus $f_{010} = \epsilon\epsilon^* f_{111}$.
Hence f is the product of $[\epsilon, 1]_{x_1} \otimes [\epsilon\epsilon^*, 1]_{x_2} \otimes [\epsilon^*, 1]_{x_3}$ with the symmetric ternary function $f_{111}[0, 1, 0, 1]$.

- Case where $(\partial^{\{1\}}\mathrm{E}_1)$ holds.
  This case is similar to the case when $(\partial^{\{1\}}\mathrm{D}_1)$ holds. The conclusion is that if $f \notin \mathscr{P}$, then f is the product of $[1, \epsilon]_{x_1} \otimes [1, \epsilon^*]_{x_2} \otimes [1, \epsilon\epsilon^*]_{x_3}$ and the ternary symmetric function $f_{000}[1, 0, 1, 0]$, where $\epsilon, \epsilon^* = \pm 1$.

2. Otherwise (i.e., suppose instead the condition in item 1 does not hold), we have, for all $1 \leq i \leq 3$, that neither $(\partial^{\{i\}}\mathrm{D}_1)$ nor $(\partial^{\{i\}}\mathrm{E}_1)$ holds. Hence for all $1 \leq i \leq 3$, $(\partial^{\{i\}} \det_1)$ holds. Then for all $1 \leq i \leq 3$, $(\partial^{\{i\}} \det_2)$, which is equivalent to $(\partial^{\{i\}} \det_3)$, does not hold, by Lemma 7.3. Thus for all $1 \leq i \leq 3$, either $(\partial^{\{i\}}\mathrm{D}_2)$ and $(\partial^{\{i\}}\mathrm{E}_3)$ or $(\partial^{\{i\}}\mathrm{D}_3)$ and $(\partial^{\{i\}}\mathrm{E}_2)$ must hold.
   We consider the case when $(\partial^{\{1\}}\mathrm{D}_2)$ and $(\partial^{\{1\}}\mathrm{E}_3)$ hold. The alternative case when $(\partial^{\{1\}}\mathrm{D}_3)$ and $(\partial^{\{1\}}\mathrm{E}_2)$ hold is similar.
   By $(\partial^{\{1\}} \det_1)$ we have $\left|\begin{smallmatrix} f_{000} & f_{001} \\ f_{010} & f_{011} \end{smallmatrix}\right| = 0$. By $(\partial^{\{1\}}\mathrm{D}_2)$ and $(\partial^{\{1\}}\mathrm{E}_3)$ we have

$$f_{000} = -f_{100}, \quad f_{011} = -f_{111}, \quad f_{001} = f_{101}, \quad f_{010} = f_{110}.$$

   (In the case of $(\partial^{\{1\}}\mathrm{D}_3)$ and $(\partial^{\{1\}}\mathrm{E}_2)$, all four RHSs are multiplied by an extra -1.)
   If $f_{000} = 0$, then by $(\partial^{\{1\}} \det_1)$ we have $f_{001}f_{010} = 0$. If $f_{001} = 0$, then $f^{x_2=0}$ is identically 0, and $f = [0, 1]_{x_2} g(x_1, x_3)$ for some binary function g, and so f is a product of the unary function $[0, 1]$ on x_2 with the binary function g on (x_1, x_3). If $f_{010} = 0$, then $f^{x_3=0}$ is identically 0, and $f = [0, 1]_{x_3} g(x_1, x_2)$ for some binary function g, and so f is a product of the unary function $[0, 1]$ on x_3 with the binary function g on (x_1, x_2). In either case, this is a contradiction to $f \notin \mathscr{P}$ by Lemma 7.1.
   Thus $f_{000} \neq 0$. By $(\partial^{\{2\}} \det_1)$ we have $\left|\begin{smallmatrix} f_{000} & f_{001} \\ f_{100} & f_{101} \end{smallmatrix}\right| = \left|\begin{smallmatrix} f_{000} & f_{001} \\ -f_{000} & f_{001} \end{smallmatrix}\right| = 0$, which implies that $f_{001} = 0$. Similarly, by $(\partial^{\{3\}} \det_1)$, we have $\left|\begin{smallmatrix} f_{000} & f_{010} \\ f_{100} & f_{110} \end{smallmatrix}\right| = \left|\begin{smallmatrix} f_{000} & f_{010} \\ -f_{000} & f_{010} \end{smallmatrix}\right| = 0$, which implies that $f_{010} = 0$. This implies that

$$f_{001} = 0, \quad f_{101} = 0, \quad f_{010} = 0, \quad f_{110} = 0. \tag{7.11}$$

   Hence f is the product of $(x_2 = x_3)$ and the degenerate binary function $g(x_1, x_3) = \left[\begin{smallmatrix} f_{000} & f_{011} \\ f_{100} & f_{111} \end{smallmatrix}\right]$. Note that the determinant $\left|\begin{smallmatrix} f_{000} & f_{011} \\ f_{100} & f_{111} \end{smallmatrix}\right| = 0$ by (7.11) and $(\partial^{\{2\}} \det_2)$.

Now we deal with the case when the spanning subgraph of K_4 is disconnected. This implies that the spanning subgraph is a triangle (plus an isolated vertex) and the number N of valid statements among all $(\partial^{\{i\}}\mathrm{D}_j)$ and $(\partial^{\{i\}}\mathrm{E}_j)$ is exactly 3. Furthermore, we may assume that either the triangle is on $\{b, c, d\}$, and then the three valid statements among all $(\partial^{\{i\}}\mathrm{D}_j)$ and $(\partial^{\{i\}}\mathrm{E}_j)$ are among $(\partial^{\{1\}}\mathrm{E}_j)$, $(\partial^{\{2\}}\mathrm{E}_j)$, and $(\partial^{\{3\}}\mathrm{E}_j)$, or, up to a cyclic permutation of the three variables of f, the triangle is on $\{a, b, c\}$, and then the three statements must be among $(\partial^{\{1\}}\mathrm{E}_j)$, $(\partial^{\{2\}}\mathrm{D}_j)$, and $(\partial^{\{3\}}\mathrm{D}_j)$.

1. Suppose the triangle is on $\{b, c, d\}$.
Since $N = 3$, there are at least six valid statements among $(\partial^{\{i\}} \det_j)$, where $1 \le i, j \le 3$. By Lemma 7.3, for every $1 \le i \le 3$, there must be exactly two valid statements among $(\partial^{\{i\}} \det_j)$ for $1 \le j \le 3$. Since the diagonals b, c, and d have the same product value, the statements $(\partial^{\{1\}} \det_j)$ take the form

$$\begin{vmatrix} f_{000} & f_{001} \\ f_{010} & f_{011} \end{vmatrix} = 0,$$
$$\begin{vmatrix} f_{000}+f_{100} & f_{001}+f_{101} \\ f_{010}+f_{110} & f_{011}+f_{111} \end{vmatrix} = \begin{vmatrix} f_{000} & f_{001} \\ f_{010} & f_{011} \end{vmatrix} + \begin{vmatrix} f_{000} & f_{101} \\ f_{010} & f_{111} \end{vmatrix} + \begin{vmatrix} f_{100} & f_{101} \\ f_{110} & f_{111} \end{vmatrix} = 0,$$
$$\begin{vmatrix} f_{000}-f_{100} & f_{001}-f_{101} \\ f_{010}-f_{110} & f_{011}-f_{111} \end{vmatrix} = \begin{vmatrix} f_{000} & f_{001} \\ f_{010} & f_{011} \end{vmatrix} - \begin{vmatrix} f_{000} & f_{101} \\ f_{010} & f_{111} \end{vmatrix} + \begin{vmatrix} f_{100} & f_{101} \\ f_{110} & f_{111} \end{vmatrix} = 0.$$

Notice that we used the fact that $\left|\begin{smallmatrix} f_{100} & f_{001} \\ f_{110} & f_{011} \end{smallmatrix}\right| = 0$, because the diagonals b and c have the same product value.
If the two valid statements among $(\partial^{\{1\}} \det_j)$ are for $j = 2$ and $j = 3$, then $\left|\begin{smallmatrix} f_{000} & f_{101} \\ f_{010} & f_{111} \end{smallmatrix}\right| = 0$, and we would have all four diagonals with an equal product value. As the spanning subgraph of K_4 is disconnected, the two valid statements among $(\partial^{\{1\}} \det_j)$ must include $j = 1$. Thus we have

$$\begin{vmatrix} f_{000} & f_{001} \\ f_{010} & f_{011} \end{vmatrix} = 0, \tag{7.12}$$
$$\begin{vmatrix} f_{000} & f_{101} \\ f_{010} & f_{111} \end{vmatrix} = -\epsilon_1 \begin{vmatrix} f_{100} & f_{101} \\ f_{110} & f_{111} \end{vmatrix}, \tag{7.13}$$

where $\epsilon_1 = +1$ if $(\partial^{\{1\}} \det_2)$ holds, and $\epsilon_1 = -1$ if $(\partial^{\{1\}} \det_3)$ holds.
Notice that if $(\partial^{\{1\}} \det_2)$ holds, we must have $(\partial^{\{1\}} \mathrm{E}_3)$, and then $f_{001} = f_{101}$ and $f_{010} = f_{110}$. On the other hand, if $(\partial^{\{1\}} \det_3)$ holds, then we must have $(\partial^{\{1\}} \mathrm{E}_2)$, and then $f_{001} = -f_{101}$ and $f_{010} = -f_{110}$. Hence $f_{001} = \epsilon_1 f_{101}$ and $f_{010} = \epsilon_1 f_{110}$ hold in either case.
Similarly by $(\partial^{\{2\}} \mathrm{E}_2)$ or $(\partial^{\{2\}} \mathrm{E}_3)$, one of which must hold, we have $f_{001} = \epsilon_2 f_{011}$ and $f_{100} = \epsilon_2 f_{110}$, where $\epsilon_2 = \pm 1$. By $(\partial^{\{3\}} \mathrm{E}_2)$ or $(\partial^{\{3\}} \mathrm{E}_3)$, one of which must hold, we have $f_{010} = \epsilon_3 f_{011}$ and $f_{100} = \epsilon_3 f_{101}$.
If any of $f_{100}, f_{101}, f_{110}, f_{001}, f_{010}, f_{011}$ is 0, then all six quantities are 0. Then the support of f is contained in $\{000, 111\}$, and we have a contradiction to $f \notin \mathscr{P}$. Thus we may normalize $f_{101} = 1$. Then

$$f_{100} = \epsilon_3, \quad f_{110} = \epsilon_2\epsilon_3, \quad f_{001} = \epsilon_1, \quad f_{010} = \epsilon_1\epsilon_2\epsilon_3, \quad f_{011} = \epsilon_1\epsilon_2.$$

By $(\partial^{\{1\}} \det_1)$, $\left|\begin{smallmatrix} f_{000} & f_{001} \\ f_{010} & f_{011} \end{smallmatrix}\right| = \left|\begin{smallmatrix} f_{000} & \epsilon_1 \\ \epsilon_1\epsilon_2\epsilon_3 & \epsilon_1\epsilon_2 \end{smallmatrix}\right| = 0$, which implies that $f_{000} = \epsilon_1\epsilon_3$.
By (7.13) we get $\left|\begin{smallmatrix} \epsilon_1\epsilon_3 & 1 \\ \epsilon_1\epsilon_2\epsilon_3 & f_{111} \end{smallmatrix}\right| = -\epsilon_1 \left|\begin{smallmatrix} \epsilon_3 & 1 \\ \epsilon_2\epsilon_3 & f_{111} \end{smallmatrix}\right|$, which implies that $f_{111} = \epsilon_2$.
It follows that f is simply the function $f_{000}[1, \epsilon_1]_{x_1} \otimes [1, \epsilon_2]_{x_2} \otimes [1, \epsilon_3]_{x_3} \in \mathscr{P}$. This is a contradiction.
2. Suppose the triangle is on $\{a, b, c\}$. We can similarly prove that under this hypothesis $f \in \mathscr{P}$, a contradiction. ∎

COROLLARY 7.10. *Let $\mathcal{F}$ be a set of signatures containing a ternary signature $f \notin \mathscr{P}$. Suppose $\mathcal{F}$ contains the unary signatures $\{[1, b_j] \mid 1 \le j \le 3\} = \{[1, 0], [1, 1], [1, -1]\}$,*

and $\partial^{\{i\}}_{[1,b_j]}(f) \in \mathscr{P}$ *for all* $1 \le i \le 3$ *and* $1 \le j \le 3$. *Then*

$$\text{Pl-\#CSP}(\mathcal{F} \cup \{g\}) \le_{\text{T}} \text{Pl-\#CSP}(\mathcal{F}),$$

where g *is either* $[1,0,1,0]$ *or* $[0,1,0,1]$, *both symmetric ternary functions.*

THEOREM 7.11. *Suppose* $\mathcal{F}$ *contains a signature* $f \notin \mathscr{P}$ *of arity* 3. *Let* $[1,a],[1,b]$, $[1,c]$ *be three unary signatures that are pairwise linearly independent. Then there exists* $g \notin \mathscr{P}$ *such that*

$$\text{Pl-\#CSP}(g,[1,a],[1,b],[1,c],\mathcal{F}) \le_{\text{T}} \text{Pl-\#CSP}([1,a],[1,b],[1,c],\mathcal{F}), \tag{7.14}$$

where g *has arity* 2 *or* g *is a symmetric signature of arity* 3.

Proof. In Pl-#CSP$([1,a],[1,b],[1,c],\mathcal{F})$, for any $[1,x] \in \{[1,a],[1,b],[1,c]\}$, we have $[1,x^k] = \partial^k_{[1,x]}(=_{k+1})$ for any $k \in \mathbb{Z}^+$. Since $[1,a],[1,b],[1,c]$ are pairwise linearly independent, there is at most one of a,b,c that can be zero. Without loss of generality, we can assume that $bc \neq 0$.

For b,c, if one of them is not a root of unity or is a root of unity of primitive order greater than 4, then we can construct five unary signatures that are pairwise linearly independent, and we are done by Lemma 7.4.

If one of b,c is a root of unity of primitive order 4, then we can construct $[1,1],[1,-1],[1,\mathrm{i}],[1,-\mathrm{i}]$, and we are done by Corollary 7.6.

If one of b,c is a root of unity of primitive order 3, then we can construct $[1,1],[1,\omega],[1,\omega^2]$ with $\omega^3 = 1, \omega \neq 1$, and we are done by Corollary 7.8.

If both b,c are roots of unity of order at most 2, then $\{[1,b],[1,c]\} = \{[1,1],[1,-1]\}$ since $[1,b],[1,c]$ are linearly independent. If $a = 0$, then we are done by Corollary 7.10. If $a \neq 0$, then $a \neq \pm 1$ since $[1,a],[1,b],[1,c]$ are pairwise linearly independent. Thus a is not a root of unity or a is a root of unity of primitive order greater than 2. In each case, we are done by Lemma 7.4, Corollary 7.6, or Corollary 7.8. ∎

7.2. $\mathscr{P}$, $\mathscr{A}$, and the parity condition. In this subsection, we give the following proposition, which implies that if a signature is of product type and satisfies the Parity Condition after the holographic transformation by $H = \left[\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}\right]$, then it is of affine type.

PROPOSITION 7.12. *Let* $f \in \mathscr{P}$ *be a signature of arity* n *and* $\hat{f} = H^{\otimes n} f$, *where* $H = \left[\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}\right]$. *If* $\hat{f}$ *satisfies the Parity Condition, then* $f \in \mathscr{A}$.

Proof. Since $f \in \mathscr{P}$, there exist f^i of arity n_i for $1 \le i \le s$ such that $f = f^1 \otimes f^2 \otimes \cdots \otimes f^s$, where $f^i \in \mathcal{E}$. Thus $\hat{f} = \hat{f}^1 \otimes \hat{f}^2 \otimes \cdots \otimes \hat{f}^s$, where $\hat{f}^i = H^{\otimes n_i} f_i$. Since $\hat{f}$ satisfies the Parity Condition, all of $\hat{f}^i$ satisfy the Parity Condition. Note that there exists $\alpha_i \in \{0,1\}^{n_i}$ such that $\text{supp}(f_i) \subseteq \{\alpha_i, \bar{\alpha}_i\}$ for $1 \le i \le s$.

We claim that $f^i \in \mathscr{A}$ for $1 \le i \le s$. Let $f^i_{\alpha_i} = a_i, f^i_{\bar{\alpha}_i} = b_i$. If $a_i = 0$ or $b_i = 0$, then $f^i \in \mathscr{A}$. Otherwise, $a_i b_i \neq 0$. For any $\beta \in \{0,1\}^{n_i}$, if $\text{wt}(\beta)$ is even, then $\hat{f}^i_\beta = \pm(a_i + b_i)$. If $\text{wt}(\beta)$ is odd, then $\hat{f}^i_\beta = \pm(a_i - b_i)$. Since $\hat{f}^i$ satisfies the Parity Condition, we have $a_i = \pm b_i$. Thus $f^i \in \mathscr{A}$. This finishes the proof of the claim. Since $f^i \in \mathscr{A}$ for $1 \le i \le s$, we have $f \in \mathscr{A}$. ∎

7.3. Normalizing signatures by a binary signature. For a ternary signature f, where $M_{x_1,x_2x_3}(f) = \left[\begin{smallmatrix} f_{000} & f_{001} & f_{010} & f_{011} \\ f_{100} & f_{101} & f_{110} & f_{111} \end{smallmatrix}\right]$, and a binary signature $[1,0,\mathbf{a}]$ (note that $\mathbf{a}$ is a scalar, not a vector, and is written in bold to highlight the modification in the following matrices), we often construct new signatures f_i by connecting one variable

of $[1, 0, \mathbf{a}]$ to the variable x_i of f for $1 \le i \le 3$. Then

$$M_{x_1,x_2x_3}(f_1) = \begin{bmatrix} f_{000} & f_{001} & f_{010} & f_{011} \\ \mathbf{a}f_{100} & \mathbf{a}f_{101} & \mathbf{a}f_{110} & \mathbf{a}f_{111} \end{bmatrix}, \tag{7.15}$$

$$M_{x_1,x_2x_3}(f_2) = \begin{bmatrix} f_{000} & f_{001} & \mathbf{a}f_{010} & \mathbf{a}f_{011} \\ f_{100} & f_{101} & \mathbf{a}f_{110} & \mathbf{a}f_{111} \end{bmatrix}, \tag{7.16}$$

$$M_{x_1,x_2x_3}(f_3) = \begin{bmatrix} f_{000} & \mathbf{a}f_{001} & f_{010} & \mathbf{a}f_{011} \\ f_{100} & \mathbf{a}f_{101} & f_{110} & \mathbf{a}f_{111} \end{bmatrix}. \tag{7.17}$$

For signatures of arity 4 we have similar operations. In the following we list the entries for a general signature of arity 4 as well as one satisfying the even Parity Condition. This is to highlight graphically the locations where $\mathbf{a}$ appears. (This operation will actually be performed on signatures of arity 4 satisfying the even Parity Condition.) For

$$M_{x_1x_2,x_4x_3}(f) = \begin{bmatrix} f_{0000} & f_{0010} & f_{0001} & f_{0011} \\ f_{0100} & f_{0110} & f_{0101} & f_{0111} \\ f_{1000} & f_{1010} & f_{1001} & f_{1011} \\ f_{1100} & f_{1110} & f_{1101} & f_{1111} \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} f_{0000} & 0 & 0 & f_{0011} \\ 0 & f_{0110} & f_{0101} & 0 \\ 0 & f_{1010} & f_{1001} & 0 \\ f_{1100} & 0 & 0 & f_{1111} \end{bmatrix},$$

and a binary signature $[1, 0, \mathbf{a}]$, we can construct new signatures f_i or g_i by connecting one variable of $[1, 0, \mathbf{a}]$ to the variable x_i of f or g for $1 \le i \le 4$. Then

$$M_{x_1x_2,x_4x_3}(f_1) = \begin{bmatrix} f_{0000} & f_{0010} & f_{0001} & f_{0011} \\ f_{0100} & f_{0110} & f_{0101} & f_{0111} \\ \mathbf{a}f_{1000} & \mathbf{a}f_{1010} & \mathbf{a}f_{1001} & \mathbf{a}f_{1011} \\ \mathbf{a}f_{1100} & \mathbf{a}f_{1110} & \mathbf{a}f_{1101} & \mathbf{a}f_{1111} \end{bmatrix} \text{or} \begin{bmatrix} f_{0000} & 0 & 0 & f_{0011} \\ 0 & f_{0110} & f_{0101} & 0 \\ 0 & \mathbf{a}f_{1010} & \mathbf{a}f_{1001} & 0 \\ \mathbf{a}f_{1100} & 0 & 0 & \mathbf{a}f_{1111} \end{bmatrix}, \tag{7.18}$$

$$M_{x_1x_2,x_4x_3}(f_2) = \begin{bmatrix} f_{0000} & f_{0010} & f_{0001} & f_{0011} \\ \mathbf{a}f_{0100} & \mathbf{a}f_{0110} & \mathbf{a}f_{0101} & \mathbf{a}f_{0111} \\ f_{1000} & f_{1010} & f_{1001} & f_{1011} \\ \mathbf{a}f_{1100} & \mathbf{a}f_{1110} & \mathbf{a}f_{1101} & \mathbf{a}f_{1111} \end{bmatrix} \text{or} \begin{bmatrix} f_{0000} & 0 & 0 & f_{0011} \\ 0 & \mathbf{a}f_{0110} & \mathbf{a}f_{0101} & 0 \\ 0 & f_{1010} & f_{1001} & 0 \\ \mathbf{a}f_{1100} & 0 & 0 & \mathbf{a}f_{1111} \end{bmatrix}, \tag{7.19}$$

$$M_{x_1x_2,x_4x_3}(f_3) = \begin{bmatrix} f_{0000} & \mathbf{a}f_{0010} & f_{0001} & \mathbf{a}f_{0011} \\ f_{0100} & \mathbf{a}f_{0110} & f_{0101} & \mathbf{a}f_{0111} \\ f_{1000} & \mathbf{a}f_{1010} & f_{1001} & \mathbf{a}f_{1011} \\ f_{1100} & \mathbf{a}f_{1110} & f_{1101} & \mathbf{a}f_{1111} \end{bmatrix} \text{or} \begin{bmatrix} f_{0000} & 0 & 0 & \mathbf{a}f_{0011} \\ 0 & \mathbf{a}f_{0110} & f_{0101} & 0 \\ 0 & \mathbf{a}f_{1010} & f_{1001} & 0 \\ f_{1100} & 0 & 0 & \mathbf{a}f_{1111} \end{bmatrix}, \tag{7.20}$$

$$M_{x_1x_2,x_4x_3}(f_4) = \begin{bmatrix} f_{0000} & f_{0010} & \mathbf{a}f_{0001} & \mathbf{a}f_{0011} \\ f_{0100} & f_{0110} & \mathbf{a}f_{0101} & \mathbf{a}f_{0111} \\ f_{1000} & f_{1010} & \mathbf{a}f_{1001} & \mathbf{a}f_{1011} \\ f_{1100} & f_{1110} & \mathbf{a}f_{1101} & \mathbf{a}f_{1111} \end{bmatrix} \text{or} \begin{bmatrix} f_{0000} & 0 & 0 & \mathbf{a}f_{0011} \\ 0 & f_{0110} & \mathbf{a}f_{0101} & 0 \\ 0 & f_{1010} & \mathbf{a}f_{1001} & 0 \\ f_{1100} & 0 & 0 & \mathbf{a}f_{1111} \end{bmatrix}. \tag{7.21}$$

Acknowledgments. We wish to express our enormous gratitude to the three anonymous referees, who gave us many pages of thoughtful and insightful comments. We benefited greatly from their suggestions and ideas. In particular, the simplified

proof of Lemma 3.8 is due to a referee. The ideas of the referees also helped us simplify the proof of Lemma 5.3.

We also thank many colleagues who listened to our ideas and partial results, whose comments we benefited from: Eric Bach, Xi Chen, Martin Dyer, Alan Frieze, Leslie Goldberg, Heng Guo, Steve Homer, Mark Jerrum, Dick Karp, Dick Lipton, Pinyan Lu, Mike Paterson, Ken Regan, David Richerby, Leslie Valiant, and Mingji Xia. We also thank Tianyu Liu for help with typesetting figures.

REFERENCES

[1] R. J. Baxter, *Exactly Solved Models in Statistical Mechanics*, Academic Press, London, 1982.
[2] A. Bulatov and V. Dalmau, *Towards a dichotomy theorem for the counting constraint satisfaction problem*, Inform. and Comput., 205 (2007), pp. 651–678.
[3] A. Bulatov, M. E. Dyer, L. A. Goldberg, M. Jalsenius, and D. Richerby, *The complexity of weighted Boolean #CSP with mixed signs*, Theoret. Comput. Sci., 410 (2009), pp. 3949–3961.
[4] J.-Y. Cai and X. Chen, *Complexity Dichotomies for Counting Problems*, Cambridge University Press, Cambridge, UK, 2017, https://doi.org/10.1017/9781107477063.
[5] J.-Y. Cai and Z. Fu, *A collapse theorem for holographic algorithms with matchgates on domain size at most* 4, Inform. and Comput., 239 (2014), pp. 149–169.
[6] J.-Y. Cai, Z. Fu, H. Guo, and T. Williams, *A Holant dichotomy: Is the FKT algorithm universal?*, in FOCS 2015, IEEE, Washington, DC, 2015, pp. 1259–1276.
[7] J.-Y. Cai and A. Gorenstein, *Matchgates revisited*, Theory Comput., 10 (2014), pp. 167–197.
[8] J.-Y. Cai, H. Guo, and T. Williams, *A complete dichotomy rises from the capture of vanishing signatures*, SIAM J. Comput., 45 (2016), pp. 1671–1728, https://doi.org/10.1137/15M1049798.
[9] J.-Y. Cai, H. Guo, and T. Williams, *The complexity of counting edge colorings and a dichotomy for some higher domain Holant problems*, Res. Math. Sci., 3 (2016), 18.
[10] J.-Y. Cai and P. Lu, *Holographic algorithms: The power of dimensionality resolved*, Theoret. Comput. Sci., 410 (2009), pp. 1618–1628.
[11] J.-Y. Cai and P. Lu, *Holographic algorithms: From art to science*, J. Comput. System Sci., 77 (2011), pp. 41–61.
[12] J.-Y. Cai, P. Lu, and M. Xia, *Holographic algorithms by Fibonacci gates and holographic reductions for hardness*, in FOCS 2008, IEEE, Washington, DC, 2008, pp. 644–653.
[13] J.-Y. Cai, P. Lu, and M. Xia, *Holant problems and counting CSP*, in STOC 2009, ACM, New York, 2009, pp. 715–724.
[14] J.-Y. Cai and P. Lu, *Signature theory in holographic algorithms*, Algorithmica, 61 (2011), pp. 779–816.
[15] J.-Y. Cai, P. Lu, and M. Xia, *Holographic algorithms with matchgates capture precisely tractable planar #CSP*, SIAM J. Comput., 46 (2017), pp. 853–889, https://doi.org/10.1137/16M1073984.
[16] J.-Y. Cai, P. Lu, and M. Xia, *Dichotomy for Holant* problems of Boolean domain*, in SODA 2011, ACM, New York, SIAM, Philadelphia, 2011, pp. 1714–1728, https://doi.org/10.1137/1.9781611973082.132.
[17] J.-Y. Cai, P. Lu, and M. Xia, *Computational complexity of Holant problems*, SIAM J. Comput., 40 (2011), pp. 1101–1132, https://doi.org/10.1137/100814585.
[18] J.-Y. Cai, P. Lu, and M. Xia, *Holographic algorithms by Fibonacci gates*, Linear Algebra Appl., 438 (2013), pp. 690–707.
[19] J.-Y. Cai, P. Lu, and M. Xia, *Dichotomy for Holant* problems with a function on domain size* 3, in SODA 2013, ACM, New York, SIAM, Philadelphia, 2013, pp. 1278–1295, https://doi.org/10.1137/1.9781611973105.93.
[20] J.-Y. Cai, P. Lu, and M. Xia, *The complexity of complex weighted Boolean #CSP*, J. Comput. System Sci., 80 (2014), pp. 217–236.
[21] N. Creignou and M. Hermann, *Complexity of generalized satisfiability counting problems*, Inform. and Comput., 125 (1996), pp. 1–12.
[22] M. E. Dyer, L. A. Goldberg, and M. Jerrum, *The complexity of weighted Boolean #CSP*, SIAM J. Comput., 38 (2009), pp. 1970–1986, https://doi.org/10.1137/070690201.
[23] M. E. Dyer and D. Richerby, *An effective dichotomy for the counting constraint satisfaction problem*, SIAM J. Comput., 42 (2013), pp. 1245–1274, https://doi.org/10.1137/100811258.

[24] M. E. Fisher, *On the dimer solution of planar Ising models*, J. Math. Phys., 7 (1966), pp. 1776–1781.
[25] M. Freedman, L. Lovász, and A. Schrijver, *Reflection positivity, rank connectivity, and homomorphism of graphs*, J. Amer. Math. Soc., 20 (2007), pp. 37–51.
[26] H. Guo, S. Huang, P. Lu, and M. Xia, *The complexity of weighted Boolean #CSP modulo k*, in STACS 2011, ACM, New York, 2011, pp. 249–260.
[27] H. Guo and T. Williams, *The complexity of planar Boolean #CSP with complex weights*, in ICALP 2013, Elsevier, New York, pp. 516–527.
[28] H. Guo, P. Lu, and L. G. Valiant, *The complexity of symmetric Boolean parity Holant problems*, SIAM J. Comput., 42 (2013), pp. 324–356, https://doi.org/10.1137/100815530.
[29] J. Harris, *Algebraic Geometry: A First Course*, Grad. Texts in Math. 133, Springer, New York, 1992.
[30] S. Huang and P. Lu, *A dichotomy for real weighted Holant problems*, Comput. Complex., 25 (2016), pp. 255–304.
[31] E. Ising, *Beitrag zur Theorie des Ferromagnetismus*, Zeitschrift für Physik, 31 (1925), pp. 253–258.
[32] M. Jerrum and A. Sinclair, *Polynomial-time approximation algorithms for the Ising model*, SIAM J. Comput., 22 (1993), pp. 1087–1116, https://doi.org/10.1137/0222066.
[33] P. W. Kasteleyn, *The statistics of dimers on a lattice*, Physica, 27 (1961), pp. 1209–1225.
[34] P. W. Kasteleyn, *Dimer statistics and phase transitions*, J. Math. Phys., 4 (1963), pp. 287–293.
[35] P. W. Kasteleyn, *Graph theory and crystal physics*, in Graph Theory and Theoretical Physics, F. Harary, ed., Academic Press, London, 1967, pp. 43–110.
[36] J. M. Landsberg, J. Morton, and S. Norine, *Holographic Algorithms without Matchgates*, preprint, https://arxiv.org/abs/0904.0471, 2009.
[37] T. D. Lee and C. N. Yang, *Statistical theory of equations of state and phase transitions.* II. *Lattice gas and Ising model*, Phys. Rev., 87 (1952), pp. 410–419.
[38] E. H. Lieb, *Residual entropy of square ice*, Phys. Rev., 162 (1967), pp. 162–172.
[39] L. Onsager, *Crystal statistics.* I. *A two-dimensional model with an order-disorder transition*, Phys. Rev., 65 (1944), pp. 117–149.
[40] J. S. Provan and M. O. Ball, *The complexity of counting cuts and of computing the probability that a graph is connected*, SIAM J. Comput., 12 (1983), pp. 777–788, https://doi.org/10.1137/0212053.
[41] H. N. V. Temperley and M. E. Fisher, *Dimer problem in statistical mechanics—an exact result*, Philos. Mag., 6 (1961), pp. 1061–1063.
[42] L. G. Valiant, *The complexity of computing the permanent*, Theoret. Comput. Sci., 8 (1979), pp. 189–201.
[43] L. G. Valiant, *The complexity of enumeration and reliability problems*, SIAM J. Comput., 8 (1979), pp. 410–421, https://doi.org/10.1137/0208032.
[44] L. G. Valiant, *Quantum circuits that can be simulated classically in polynomial time*, SIAM J. Comput., 31 (2002), pp. 1229–1254, https://doi.org/10.1137/S0097539700377025.
[45] L. G. Valiant, *Expressiveness of matchgates*, Theoret. Comput. Sci., 289 (2002), pp. 457–471.
[46] L. G. Valiant, *Holographic algorithms*, SIAM J. Comput., 37 (2008), pp. 1565–1594, https://doi.org/10.1137/070682575.
[47] L. G. Valiant, *Accidental algorithms*, in FOCS 2006 (Berkeley, CA), IEEE, Washington, DC, 2006, pp. 509–517.
[48] C. N. Yang, *The spontaneous magnetization of a two-dimensional Ising model*, Phys. Rev., 85 (1952), pp. 808–816.
[49] C. N. Yang and T. D. Lee, *Statistical theory of equations of state and phase transitions.* I. *Theory of condensation*, Phys. Rev., 87 (1952), pp. 404–409.

Paper 3

2015 IEEE 56th Annual Symposium on Foundations of Computer Science

A Holant Dichotomy: Is the FKT Algorithm Universal?

Jin-Yi Cai*, Zhiguo Fu†, Heng Guo*, and Tyson Williams*
* *Computer Sciences Department, University of Wisconsin–Madison, Madison, WI, USA*
{jyc,hguo,tdw}@cs.wisc.edu
† *School of Mathematics, Jilin University, Changchun, Jilin, China*
fuzg@jlu.edu.cn

Abstract

We prove a complexity dichotomy for complex-weighted Holant problems with an arbitrary set of symmetric constraint functions on Boolean variables.

In the study of counting complexity, such as #CSP, there are problems which are #P-hard over general graphs but P-time solvable over planar graphs. A recurring theme has been that a holographic reduction [36] to FKT precisely captures these problems. This dichotomy answers the question: Is this a *universal* strategy? Surprisingly, we discover new planar tractable problems in the Holant framework (which generalizes #CSP) that are not expressible by a holographic reduction to FKT. In particular, the putative form of a dichotomy for planar Holant problems is false. Nevertheless, we prove a dichotomy for #CSP^2, a variant of #CSP where every variable appears even times, that the presumed *universality* holds for #CSP^2. This becomes an important tool in the proof of the full dichotomy, which refutes this universality in general. The full dichotomy says that the new P-time algorithms and the strategy of holographic reductions to FKT together are *universal* for these locally defined counting problems.

As a special case of our new planar tractable problems, counting perfect matchings (#PM) over k-uniform hypergraphs is P-time computable when the incidence graph is planar and $k \geq 5$. The same problem is #P-hard when $k = 3$ or $k = 4$, also a consequence of the dichotomy. More generally, over hypergraphs with specified hyperedge sizes and the same planarity assumption, #PM is P-time computable if the greatest common divisor (gcd) of all hyperedge sizes is at least 5.

Keywords

Computational Complexity; Counting Problems; Dichotomy Theorem; Holographic Algorithms; Holant Problems;

I. Introduction

The Fisher-Kasteleyn-Temperley (FKT) algorithm [32], [21], [22] is a classical gem that counts perfect matchings over planar graphs in polynomial time. This was an important milestone in a decades-long research program by physicists in statistical mechanics to determine what is known as Exactly Solved Models [1], [20], [30], [41], [42], [25], [32], [21], [22], [26], [27], [40].

For four decades, the FKT algorithm stood as *the* polynomial-time algorithm for any counting problem over planar graphs that is #P-hard over general graphs. Then Valiant introduced *matchgates* [34], [33] and *holographic* reductions to the FKT algorithm [36], [35]. These reductions differ from classical ones by introducing quantum-like superpositions. This novel technique extended the reach of the FKT algorithm and produced polynomial-time algorithms for a number of problems for which only exponential-time algorithms were previously known.

Since the new polynomial-time algorithms appear so exotic and unexpected, and the problems appear so close to being #P-hard, they challenge our faith in the well-accepted conjecture that P $\neq$ NP. Quoting Valiant [35]: "The objects enumerated are sets of polynomial systems such that the solvability of any one member would give a polynomial time algorithm for a specific problem. ... the situation with the P

0272-5428/15 $31.00 © 2015 IEEE
DOI 10.1109/FOCS.2015.81

= NP question is not dissimilar to that of other unresolved enumerative conjectures in mathematics. The possibility that accidental or freak objects in the enumeration exist cannot be discounted if the objects in the enumeration have not been studied systematically." Indeed, if any "freak" object exists in this framework, it would collapse #P to P.

Therefore, over the past 10 to 15 years, this technique has been intensely studied in order to gain a systematic understanding to the limit of the trio of holographic reductions, matchgates, and the FKT algorithm [33], [3], [4], [10], [37], [11], [24], [28], [29]. Without settling the P versus #P question, the best hope is to achieve a complexity classification. This program finds its sharpest expression in a complexity dichotomy theorem, which classifies *every* problem expressible in a framework as either solvable in P or #P-hard, with nothing in between.

Out of this work, a strong theme has emerged. For a wide variety of problems, such as those expressible as a #CSP, holographic reductions to the FKT algorithm is a *universal* technique for turning problems that are #P-hard in general to P-time solvable over planar graphs. In fact, a preponderance of evidence suggests the following putative classification of all counting problems defined by local constraints into *exactly* three categories: (1) those that are P-time solvable over general graphs; (2) those that are P-time solvable over planar graphs but #P-hard over general graphs; and (3) those that remain #P-hard over planar graphs. Moreover, category (2) consists precisely of those problems that are holographically reducible to the FKT algorithm. This theme is so strong that it has become an intuitive and trusty guide for us when we investigate unknown problems and plan proof strategies. In fact, many of the results in the present paper were proved in this way. However, one is still left wondering whether a holographic reduction to the FKT algorithm is a *universal* strategy for all such counting problems that are planar tractable but not in general.

We list some of the supporting evidence for this putative classification. These date back to the classification of the complexity of the Tutte polynomial [39], [38]. It has also been an unfailing theme in the classification of spin systems and #CSP [23], [12], [9], [18]. However, these frameworks do not capture all locally specified counting problems. Some natural problems, such as counting perfect matchings (#PM), are not expressible as a point on the Tutte polynomial, and #PM is provably not expressible as a partition function of spin systems (vertex assignment models) [16], [15], [31]. However, this is the problem for which FKT was designed, and is the basis of Valiant's matchgates and holographic reductions.

A refined framework, called Holant problems [13], was proposed to address this issue. It is an edge assignment model. It naturally encodes and expresses #PM as well as Valiant's matchgates and holographic reductions. Thus, Holant is the proper framework in which to study the power of holographic algorithms. It is also more general than #CSP in the sense that a complete complexity classification for Holant problems implies one for #CSP.

In this paper, we classify for the first time the complexity of Holant problems over planar graphs. Our result generalizes both the dichotomy for Holant [19], [6] and the dichotomy for planar #CSP [12], [18]. Although the #CSP dichotomy does not resolve the complexity of #PM, planar tractable classes of #CSP are tractable due to holographic algorithms with matchgates, which essentially relies on counting (weighted) perfect matchings by FKT. On the other hand, #PM, even for d-regular graphs, is shown to be #P-hard under the Holant framework [19], yet its planar tractability is not addressed in either [19] or [6] until the current work.

Surprisingly, we discover new planar tractable problems that are not expressible by a holographic reduction to matchgates and FKT. To the best of our knowledge, this is the first primitive extension since FKT to a counting problem solvable in P over planar instances but #P-hard in general. We consider this a primitive extension because it is provably not based on a (holographic) transformation to the

FKT algorithm. Furthermore, our dichotomy theorem says that this completes the picture: there are no more undiscovered extensions for problems expressible in this framework, unless #P collapses to P. In particular, the putative form of the planar Holant dichotomy is *false*.

Before stating our main theorem, we give a brief description of the Holant framework [13]. Fix a set of local constraint functions $\mathcal{F}$. A *signature grid* $\Omega = (G, \pi)$ is a tuple, where $G = (V, E)$ is a graph, π labels each $v \in V$ with a function $f_v \in \mathcal{F}$ with input variables from the incident edges $E(v)$ at v. Each f_v maps $\{0,1\}^{\deg(v)}$ to $\mathbb{C}$. We consider all 0-1 edge assignments. An assignment σ for every $e \in E$ gives an evaluation $\prod_{v \in V} f_v(\sigma \mid_{E(v)})$, where $\sigma \mid_{E(v)}$ denotes the restriction of σ to $E(v)$. The counting problem on the instance Ω is to compute

$$\text{Holant}_\Omega = \sum_{\sigma: E \to \{0,1\}} \prod_{v \in V} f_v \left(\sigma \mid_{E(v)}\right). \tag{I.1}$$

For example, #PM, the problem of counting perfect matchings in G, corresponds to assigning the EXACTONE function at every vertex of G. The Holant problem parameterized by the set $\mathcal{F}$ is denoted by $\text{Holant}(\mathcal{F})$.

At a high level, we can state our main theorem as follows.

Theorem I.1. *Let $\mathcal{F}$ be a set of complex-valued, symmetric functions on Boolean variables. Then there is an effective classification for all possible $\mathcal{F}$, according to which, $\text{Holant}(\mathcal{F})$ is either (1) P-time computable over general graphs, or (2) P-time computable over planar graphs but #P-hard over general graphs, or (3) #P-hard over planar graphs.*

Note that here we restrict our focus on symmetric functions, which are invariant under permutation of arguments. Most natural combinatorial problems like counting vertex covers or perfect matchings can be encoded by symmetric functions.

The complete statement is given in Theorem III.1. The classification is explicit. The tractability criterion is decidable in polynomial time due to [11], [7]. Tractable problems over general graphs have been previously studied in [6]. The planar tractable class includes both those solvable by holographic reductions to FKT and those newly discovered. Explicit criteria for these are also proved in this paper.

Let us meet some new tractable problems. They can be described as orientation problems, which are Holant problems after a complex-valued holographic transformation. Given a planar (multi)graph, we allow two kinds of vertices. The first kind can be either a sink or a source while the second kind only allow one incoming edge. The goal is to compute the number orientations satisfying these constraints. This problem can be expressed in the Holant framework under a Z-transformation.[1] It can be shown that this is equivalent to the Holant problem on the edge-vertex incidence graph where we assign the DISEQUALITY function to every edge, and to each vertex, we assign either the EQUALITY function or the EXACTONE function. Suppose vertices assigned EQUALITY functions all have degree k. If $k = 2$, then this problem can be solved by FKT. We show that this problem is #P-hard if $k = 3$ or $k = 4$, but is tractable again if $k \geq 5$. The algorithm involves a recursive procedure that simplifies the instance until it can be solved by known algorithms, including FKT. This simplilfication process pins edges to fixed values, yet the final answer is still possible to be non-trivial as the pinning will end when the instance is solvable by known algorithms. The algorithm crucially uses global topological properties of a planar graph, in particular Euler's characteristic formula. If the graph is not planar, then this algorithm does not work, and indeed the problem is #P-hard over general graphs.

[1]This transformation is $Z = \left[\begin{smallmatrix} 1 & 1 \\ i & -i \end{smallmatrix}\right]$. It is common that one problem can be transformed to another over $\mathbb{C}$ while one or both problems are specified by real-valued constraint functions, and provably no transformation exists over $\mathbb{R}$. Thus to study the classification question over complex-valued constraint functions is natural and proper. For example, the integer-valued orientation problem studied here, if expressed as Holant directly, is complex weighted.

More generally, we allow vertices of arbitrary degrees to be assigned EQUALITY. If all the degrees are at most 2, then the problem is tractable by the FKT algorithm. Otherwise, the complexity depends on the greatest common divisor (gcd) of the degrees. The problem is tractable if $\gcd \geq 5$ and #P-hard if $\gcd \leq 4$. It is worth noting that the criterion for tractability is not a degree lower bound. Moreover, the planarity assumption and the degree rigidity pose a formidable challenge in the hardness proofs for $\gcd \leq 4$. We note that these degree restrictions and planarity will not make new tractable instances vacuous, since there are two types of vertices and we do not lower bound degrees of those assigned EXACTONE. In addition, as common in the study of #CSP, we allow multigraphs as valid instances.

If the graph is bipartite with EQUALITY functions assigned on one side and EXACTONE functions on the other, then this is the problem of #PM over hypergraphs with planar incidence graphs. Our results imply that the complexity of this problem depends on the gcd of the hyperedge sizes. The problem is computable in polynomial time when $\gcd \geq 5$ and is #P-hard when $\gcd \leq 4$ (assuming there are hyperedges of size at least 3).

Most reductions in previous Holant dichotomy theorems [19], [6] do not hold for planar graphs, so we are forced to develop new techniques. In particular, an important ingredient in previous proofs is the #CSP^d dichotomy by Huang and Lu [19]. Here #CSP^d denotes #CSP where every variable appears a multiple of d times. The very first step in the #CSP^d dichotomy proof uses the pinning technique. Multiple copies of an instance are created and vertices are connected across different copies. This construction violates planarity. Moreover, this violation is unavoidable, a consequence of the new dichotomy. Due to our newly discovered tractable problems, the putative form of a planar #CSP^d dichotomy is *false* when $d \geq 5$. Nevertheless, we prove a dichotomy for planar #CSP^2 for which the putative form is, luckily for us, true (but not obvious in hindsight). A dichotomy for planar #CSP^2 is essential because it captures a significant fraction of planar Holant problems either directly or through reductions. We manage to prove the planar Holant dichotomy without appealing to planar #CSP^d for $d \geq 3$.

The proof of the planar #CSP^2 dichotomy comprises the entire Part II of the full version [8] starting on page 63. A brief outline of the proof is given in Section IV. Among the concepts and techniques introduced are some special tractable families of constraint functions specific to the #CSP^2 framework. We also introduce a *derivative* ∂ and its inverse operator *integral* $\int$ to streamline the proof argument. There is also an application of the theory of *cyclotomic fields*.

We began this project expecting to prove the putative form of the planar Holant dichotomy. It was determined that a planar #CSP^d dichotomy would be both a more modest and attainable intermediate step as well as a good launch station for the final goal. However after some attempts, even the planar #CSP^d dichotomy appeared too difficult to achieve, and so we scaled back the ambition to solve just $d = 2$. Luckily, a successful #CSP^2 dichotomy can carry most of the weight of a full #CSP^d dichotomy, *and*, as it turned out, the putative form of the planar #CSP^2 dichotomy is *true* while that for planar #CSP^d is not. Ironically, many steps of our proof in this paper were guided by the putative form of the complexity classification. The discovery of the new tractable problems changed the original plan, but also helped complete the picture.

Coming back to the challenge of the P vs. NP question posed by Valiant's holographic algorithms, we venture the opinion that the dichotomy theorem provides a satisfactory answer. Indeed, it would be difficult to conceive a world where #P is in fact equal to P, and yet all this algebraic theory can somehow maintain a consistent, sharp but faux division where there is none. (Consider the following Gedankenexperiment: #P is really equal to P, but the Supreme Fascist keeps scores on how much of #P we have learned to be in P. For every problem in this broad class that is yet unknown to be in P the SF lets we prove it #P-hard—a superfluous notion really. Nevertheless for every problem in the class known to be in P, the SF makes sure our proof method for #P-hardness on that problem fails, thus preventing

one from making the ultimate discovery.)

II. Preliminaries

Fix a set of local constraint functions $\mathcal{F}$. A *signature grid* $\Omega = (G, \pi)$ is a tuple, where $G = (V, E)$ is a graph, π labels each $v \in V$ with a function $f_v \in \mathcal{F}$ with input variables from the incident edges $E(v)$ at v. Each f_v maps $\{0,1\}^{\deg(v)}$ to $\mathbb{C}$. We consider all 0-1 edge assignments. An assignment σ for every $e \in E$ gives an evaluation $\prod_{v \in V} f_v(\sigma \mid_{E(v)})$, where $\sigma \mid_{E(v)}$ denotes the restriction of σ to $E(v)$. The counting problem on the instance Ω is to compute $\text{Holant}_\Omega = \sum_{\sigma: E \to \{0,1\}} \prod_{v \in V} f_v\left(\sigma \mid_{E(v)}\right)$. The Holant problem parameterized by the set $\mathcal{F}$ is denoted by $\text{Holant}(\mathcal{F})$ and Pl-Holant$(\mathcal{F})$ is defined similarly using a signature grid with a planar graph.

A function f_v can be represented by listing its values in lexicographical order as in a truth table, which is a vector in $\mathbb{C}^{2^{\deg(v)}}$, or as a tensor in $(\mathbb{C}^2)^{\otimes \deg(v)}$. A symmetric function f on k Boolean variables can be expressed as $[f_0, f_1, \ldots, f_k]$, where f_w is the value of f on inputs of Hamming weight w. This is called the *signature* of f, and we may use the terms "signature" and "function" interchangeably below. For example, we use $=_k$ to denote the Equality signature $[1, 0, \ldots, 0, 1]$ of arity k.

A symmetric signature f of arity n is *degenerate* if there exist a unary signature $u \in \mathbb{C}^2$ such that $f = u^{\otimes n}$. Replacing such signatures by n copies of the corresponding unary signature does not change the Holant value. Replacing a signature $f \in \mathcal{F}$ by a constant multiple cf, where $c \neq 0$, does not change the complexity of $\text{Holant}(\mathcal{F})$. It introduces a global nonzero factor to $\text{Holant}(\Omega; \mathcal{F})$.

To introduce the idea of holographic reductions, it is convenient to consider bipartite graphs. For a general graph, we can always transform it into a bipartite graph while preserving the Holant value. For each edge in the graph, we replace it by a path of length two. Each new vertex is assigned the binary Equality signature $(=_2) = [1, 0, 1]$.

For a 2-by-2 matrix T and a signature set $\mathcal{F}$, define $T\mathcal{F} = \{g \mid \exists f \in \mathcal{F} \text{ of arity } n,\ g = T^{\otimes n} f\}$, similarly for $\mathcal{F}T$. Whenever we write $T^{\otimes n} f$ or $T\mathcal{F}$, we view the signatures as column vectors; similarly for $fT^{\otimes n}$ or $\mathcal{F}T$ as row vectors. We use $\text{Holant}\ (\mathcal{R} \mid \mathcal{G})$ to denote the Holant problem on bipartite graphs $H = (U, V, E)$, where each vertex in U or V is assigned a signature in $\mathcal{R}$ or $\mathcal{G}$, respectively. Let T be an invertible 2-by-2 matrix. The holographic transformation defined by T is the following operation: given a signature grid Ω with underlying graph H, we create a new grid Ω' such that the graph is still H, and any functions f on the left (or g on the right) is replaced by $fT^{\otimes n}$ (or $\left(T^{-1}\right)^{\otimes m} g$) where n and m are arities of f and g. We frequently apply a holographic transformation defined by the matrix $Z = \left[\begin{smallmatrix} 1 & 1 \\ i & -i \end{smallmatrix}\right]$.

Theorem II.1 (Valiant's Holant Theorem [36])**.** *If there is a holographic transformation mapping signature grid Ω to Ω', then* $\text{Holant}_\Omega = \text{Holant}_{\Omega'}$.

In order to do holographic transformations on a general graph, we can always modify it into an equivalent bipartite graph preserving the Holant value as follows. For each edge in the graph, we replace it by a path of length two. (This operation is called the *2-stretch* of the graph and yields the edge-vertex incidence graph.) Each new vertex is then assigned the binary Equality signature $(=_2) = [1, 0, 1]$.

We say a signature set $\mathcal{F}$ is $\mathscr{C}$-transformable if there exists a transformation $T \in \mathbf{GL}_2(\mathbb{C})$ such that $[1, 0, 1]T^{\otimes 2} \in \mathscr{C}$ (viewed as row vectors) and $\mathcal{F} \subseteq T\mathscr{C}$ (viewed as column vectors). The importance of this definition is that if Pl-Holant$(\mathscr{C})$ is tractable, then Pl-Holant$(\mathcal{F})$ is also tractable for any $\mathscr{C}$-transformable set $\mathcal{F}$.

III. Main Theorem and Proof Outline

In this section, we state the main theorem and give an outline of its proof.

We use $\mathscr{A}$, $\mathscr{P}$, $\mathscr{V}$, and $\mathscr{M}$ to denote four base classes of tractable signatures. The classes $\mathscr{A}$ and $\mathscr{P}$ are identified as tractable for #CSP [14]. Problems defined by $\mathscr{A}$ are tractable essentially by Gauss sums [2].

The signatures in $\mathscr{P}$ are tensor products of signatures whose supports are among two complementery bit vectors. Problems defined by them are tractable by a propagation algorithm. The class $\mathscr{V}$ contains vanishing signatures [17], [6], which means the Holant value is always 0. We split $\mathscr{V}$ into $\mathscr{V}^+$ and $\mathscr{V}^-$. Any subset of $\mathscr{V}^+$ or $\mathscr{V}^-$ vanishes, but mixing these two classes does not necessarily vanish. Valiant [34], [33] introduced matchgates, which we denote by $\mathscr{M}$. They can be locally expressed by weighted perfect matchings, so problems defined by them are tractable by the FKT algorithm over planar graphs. The full version [8] contains complete definitions and characterizations of these four classes. As mentioned at the end of last section, a problem defined by a signature that is transformable to any of these tractable classes is also tractable. In fact, $\mathscr{V}$ is closed under this transformable notion.

We need some more notations. Let $\mathscr{R}_2^{\pm}$ denote the set of all unary signatures plus symmetric signatures $f = [f_0, f_1, \ldots, f_n]$ satisfying $f_i \pm 2if_{i+1} + f_{i+2} = 0$ for all $0 \le i \le n-2$. For a signature set $\mathcal{F}$, let $\mathcal{F}^*$ denote $\mathcal{F}$ with all degenerate signatures $[a, b]^{\otimes m}$ replaced by unary $[a, b]$. We denote by EXACTONE_d the signature $[0, 1, 0, \ldots, 0]$ of arity d. Let $\mathcal{EO} = \{\text{EXACTONE}_d \mid d \ge 3\}$.

Theorem III.1. *Let $\mathcal{F}$ be any set of symmetric, complex-valued signatures in Boolean variables. Then* $\text{Pl-Holant}(\mathcal{F})$ *is #P-hard unless $\mathcal{F}$ satisfies one of the following conditions:*

1) *All non-degenerate signatures in $\mathcal{F}$ are of arity at most 2;*
2) *$\mathcal{F}$ is $\mathscr{A}$- or $\mathscr{P}$-transformable;*
3) *$\mathcal{F} \subseteq \mathscr{V}^\sigma \cup \{f \in \mathscr{R}_2^\sigma \mid \text{arity}(f) = 2\}$ for some $\sigma \in \{+, -\}$;*
4) *All non-degenerate signatures in $\mathcal{F}$ are in $\mathscr{R}_2^\sigma$ for some $\sigma \in \{+, -\}$.*
5) *$\mathcal{F}$ is $\mathscr{M}$-transformable;*
6) *$\mathcal{F} \subseteq Z(\mathscr{P} \cup \mathcal{EO})$ or $Z(\mathscr{P} \cup \left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right]\mathcal{EO})$, and the greatest common divisor of the arities of all signatures in $\mathcal{F}^* \cap \mathscr{P}_2$ is at least 5.*

In each exceptional case, $\text{Pl-Holant}(\mathcal{F})$ *is computable in polynomial time.* $\text{Holant}(\mathcal{F})$ *is computable in polynomial time without planarity if $\mathcal{F}$ satisfies conditions 1, 2, 3, or 4, and is #P-hard otherwise.*

Proof sketch: We first prove a dichotomy theorem when $\mathcal{F}$ contains a single non-degenerate signature f of arity ≥ 3 (cf. Theorem 6.1 in the full version [8]). The proof is by induction on the arity of f. Base cases are when the signature has arity 3 or 4, which have been proved in previous work [12], [18]. The inductive step reduces the arity of f by two each time, and then we apply the induction hypothesis. This essentially yields seven different cases that are $\mathscr{A}$- or $\mathscr{P}$- or $\mathscr{M}$-transformable or in $\mathscr{V}$ (not in 1-to-1 correspondence), plus the possibility of degenerate cases. However, we can roughly split them into two categories: (1) those tractable by orthogonal and related transformations; and (2) those tractable by a $Z = \left[\begin{smallmatrix} 1 & 1 \\ i & -i \end{smallmatrix}\right]$ transformation. We show that any case in category (1) can be solved by the Pl-#CSP^2 dichotomy via reductions. We handle each case in category (2) separately and show hardness using gadget constructions and polynomial interpolations.

Given the single signature dichotomy, we assume that every nontrivial signature in $\mathcal{F}$ falls into one of the two categories above. Again, if any signature is in category (1), then we can apply the Pl-#CSP^2 dichotomy through reductions. Otherwise, all nontrivial signatures are from category (2). Then we rule out any possible mixing of signatures in $\mathscr{V}$ with other signatures in category (2). This leaves two kinds of signatures in category (2), from $Z\mathscr{P}$ or from $Z\mathcal{EO} \cup Z\left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right]\mathcal{EO}$. A putative form of the planar Holant dichotomy would dictate that any mixture from these two sets is intractable. However, we found that there are tractable cases violating the putative dichotomy, which are summarized above as Case 6. Then we finish the proof by showing that there are no other tractable cases. ■

IV. Dichotomy for Pl-#CSP^2

In this section, we state the dichotomy for Pl-#CSP^2, and provide a sketch of the proof here. Let $\mathcal{T}_k = \left\{ \left[\begin{smallmatrix} 1 & 0 \\ 0 & \omega \end{smallmatrix}\right] \in \mathbb{C}^{2\times 2} \mid \omega^k = 1 \right\}$. Let $\widehat{\mathscr{M}} = \left[\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}\right] \mathscr{M}$.

Theorem IV.1. *Let $\mathcal{F}$ be a set of symmetric signatures. Then Pl-#CSP$^2(\mathcal{F})$ is #P-hard unless $\mathcal{F}$ satisfies one of the following conditions:*

1) *there exists $T \in \mathcal{T}_8$ such that $\mathcal{F} \subseteq T\mathscr{A}$;*
2) *$\mathcal{F} \subseteq \mathscr{P}$;*
3) *there exists $T \in \mathcal{T}_4$ such that $\mathcal{F} \subseteq T\widehat{\mathscr{M}}$.*

In each exceptional case, Pl-#CSP$^2(\mathcal{F})$ is computable in polynomial time.

Proof Sketch: We first define some tractable families of signatures specific to the Pl-#CSP^2 framework. Let $\widetilde{\mathscr{A}} = \mathscr{A} \cup \left[\begin{smallmatrix} 1 & 0 \\ 0 & e^{\pi i/4} \end{smallmatrix}\right] \mathscr{A}$ and $\widetilde{\mathscr{M}} = \widehat{\mathscr{M}} \cup \left[\begin{smallmatrix} 1 & 0 \\ 0 & i \end{smallmatrix}\right] \widehat{\mathscr{M}}$. One can show that $\widetilde{\mathscr{A}}$ covers Case 1 above, and $\widetilde{\mathscr{M}}$ covers Case 3. The proof will revolve around these tractable classes.

The overall plan is to break the proof into two main steps.

The first step is to prove the dichotomy theorem for Pl-#CSP$^2(\mathcal{F})$ when there is at least one nonzero signature of *odd* arity in $\mathcal{F}$. In this case we can make use of a lemma that shows that we can simulate Pl-#CSP$(\mathcal{F})$ by Pl-#CSP$^2(\mathcal{F})$ if $\mathcal{F}$ includes a unary signature $[a, b]$ with $ab \neq 0$. Then we can apply the known planar #CSP dichotomy [18] for Pl-#CSP. However this strategy (provably) *cannot* work in the case when every signature in $\mathcal{F}$ satisfies the *parity* constraint. In that case we employ other means. This first step of the proof is relatively uncomplicated.

The second step is to deal with the case when all signatures in $\mathcal{F}$ have even arity. This is where the real difficulty lies. In this case it is impossible to directly construct *any* unary signature. So we cannot use that lemma pertaining to a unary signature. But we prove another lemma which provides a way to simulate Pl-#CSP$(\mathcal{F})$ by Pl-#CSP$^2(\mathcal{F})$ in a *global* fashion, *if* $\mathcal{F}$ includes some tensor power of the form $[a, b]^{\otimes 2}$ where $ab \neq 0$. Moreover, we have a lucky break (for the complexity of the proof) if $\mathcal{F}$ includes a signature that is in $\widehat{\mathscr{M}} \setminus (\mathscr{P} \cup \widetilde{\mathscr{A}})$. In this case, we can construct a special binary signature, and obtain $[1, 1]^{\otimes 2}$ by interpolation. This proof uses the theory of *cyclotomic fields*. This simplifies the proof greatly. For all other cases (when $\mathcal{F}$ has only even arity signatures), the proof gets going in earnest—we will attempt an induction on the arity of signatures.

The lowest arity of this induction will be two. We will try to reduce the arity to two whenever possible; however for many cases an arity reduction to two destroys the #P-hardness at hand. Therefore the true basis of this induction proof of Pl-#CSP^2 starts with arity 4. Consequently we will first prove a dichotomy theorem for Pl-#CSP$^2(f)$, where f is a signature of arity 4. Several tools will be used. These include the rank criterion for redundant signatures, complex weighted k-regular graph homomorphisms [5] for arity two signatures, and a trick we call the *Three Stooges* by domain pairing.

However in the next step we do not attempt a general Pl-#CSP^2 dichotomy for a *single* signature of even arity. This would have been natural at this point, but it would have been too difficult. We will need some additional leverage by proving a conditional No-Mixing Lemma for pairs of signatures of even arity. So, taking a detour, we prove that for two signatures f and g both of even arity, that individually belong to some tractable class, but do not belong to a single tractable class in the conjectured dichotomy (that is yet to be proved), the problem Pl-#CSP$^2(f, g)$ is #P-hard. We prove this No-Mixing Lemma for any pair of signatures f and g both of even arity, not restricted to arity 4. Even though at this point we only have a dichotomy for a single signature of arity 4, we prove this No-Mixing Lemma for higher even arity pairs f and g by simulating two signatures f' and g' of arity 4 that belong to different tractable sets, from that of Pl-#CSP$^2(f, g)$. After this arity reduction (within the No-Mixing Lemma), we prove

that Pl-#CSP$^2(f', g')$ is #P-hard by the dichotomy for a *single* signature of arity 4. After this, we prove a No-Mixing Lemma for a *set* of signatures $\mathcal{F}$ of even arities, which states that *if* $\mathcal{F}$ is contained in the union of all tractable classes, then it is still #P-hard unless it is *entirely* contained in a single tractable class. Note that at this point we still only have a *conditional* No-Mixing Lemma in the sense that we have to assume every signature in $\mathcal{F}$ belongs to some tractable set.

We then attempt the proof of a Pl-#CSP^2 dichotomy for a *single* signature of arbitrary even arity. This uses all the previous lemmas, in particular the (conditional) No-Mixing Lemma for a set of signatures. However, after completing the proof of this Pl-#CSP^2 dichotomy for a single signature of even arity, the No-Mixing Lemma becomes absolute.

Finally we extend the dichotomy for a single signature of even arity to a dichotomy theorem for Pl-#CSP$^2(\mathcal{F})$ where all signatures in $\mathcal{F}$ have even arity. Together with the first main step when $\mathcal{F}$ contains some nonzero signature of odd arity, this completes the proof of Theorem IV.1. ■

V. New Tractable Problems and Related Hardness Results

We are not able to include the whole proof of Theorem III.1. In this last section, we highlight a tractable case and include some related hardness results, summarized as follows.

Theorem V.1. Pl-Holant $(\neq_2 \mid =_k, \mathcal{EO})$ *is #P-hard when* $k \in \{3, 4\}$*, and is computable in polynomial time when* $k \in \{1, 2\}$ *or* $k \geq 5$*.*

Under Z, Pl-Holant $(\neq_2 \mid =_k, \mathcal{EO})$ is Holant$(Z(=_k), Z(\mathcal{EO}))$. When $k \in \{1, 2\}$, the problem is tractable by either Case 4 or Case 5 of Theorem III.1. The interesting tractable case is when $k \geq 5$, belonging to Case 6. The claim about hypergraph #PM in the introduction follows from Theorem V.1, where tractability follows directly and hardness requires a gadget, which we omit here.

A. Tractability when $k \geq 5$

We first prove that Pl-Holant $(\neq_2 \mid =_k, \mathcal{EO})$ is tractable when $k \geq 6$. After that, we consider $k = 5$. A key observation is that a planar (bipartite) graph cannot be simple if its degrees are large enough. The proof is a straightforward application of Euler's characteristic equation for planar graphs.

Lemma V.2. *Let* $G = (L \cup R, E)$ *be a planar bipartite graph with parts* L *and* R*. If every vertex in* L *has degree at least* 6 *and every vertex in* R *has degree at least* 3*, then* G *cannot be simple.*

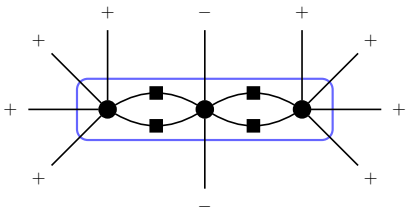


Figure 1: An E_6-block. Circles are $=_6$ and squares are $\neq_2$.

For Pl-Holant $(\neq_2 \mid =_k, \mathcal{EO})$, we may contract edges between $=_k$ and between EXACTONE functions. Note that we want to count the number of satisfying assignments as there is no weight. We call an edge *pinned* if it has the same value in all satisfying assignments, if there is any. Any connection among EXACTONE's either creates pinned edges, or results in a larger EXACTONE. We create components called E_k-blocks composed by $=_k$'s and $\neq_2$'s. An E_k-block is *trivial* if it has no satisfying assignment. A nontrivial E_k-block has exactly two complementary assignments, and we mark edges with signs "+"

and "$-$" such that edges with the same sign (or distinct signs) take the same value (or distinct values). Figure 1 pictures an example. Parallel edges between an E_k-block and an EXACTONE always result in pinned edges. Lemma V.2 does not give us tractability for the case of $k \geq 6$ directly. The reason is that E_k-blocks may have arity less than 6, in which case Lemma V.2 does not apply. However, for $k \geq 6$ and a nontrivial E_k-block of arity n where $n < 6$, we can show that it is either a binary $\neq_2$, or has arity 4, identified in Figure 2a up to a rotation.

(a) Two different arity 4 E_k-blocks. (b) Replace them by parallel $\neq_2$'s.

Figure 2: Arity 4 E_k-blocks.

In the following lemma, we show how to replace an E_k-block of arity 4 by some other signatures while keeping track of, but not preserving, the Holant value.

Lemma V.3. *For any integer $k \geq 6$,* Pl-Holant $(\neq_2 \mid =_k, \mathcal{EO})$ *is computable in polynomial time.*

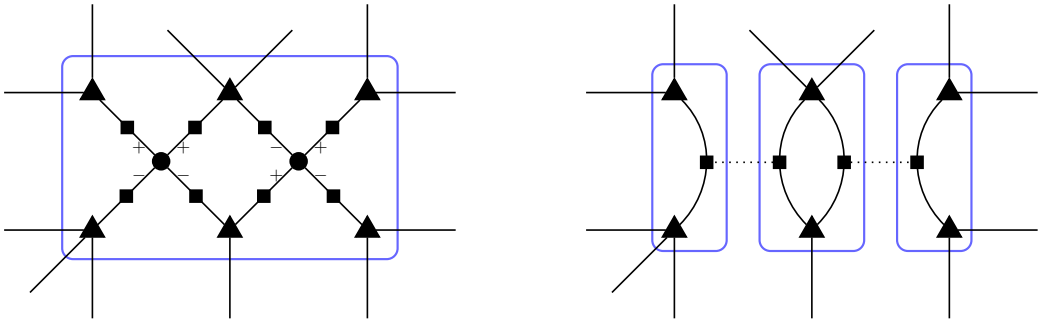
Proof: Let Ω be a connected instance of Pl-Holant $(\neq_2 \mid =_k, \mathcal{EO})$. When an edge is pinned to a known value, we get a smaller instance of Pl-Holant $(\neq_2 \mid =_k, \mathcal{EO})$ without changing the number of satisfying assignments. In our algorithm, we may also find a contradiction and simply return 0.

We claim that there always exists an edge in Ω that is pinned, unless Ω does not contain $=_k$, or does not contain EXACTONE$_d$ functions (for some $d \geq 3$), or there is a contradiction. Furthermore if there are $=_k$ or EXACTONE$_d$ functions (for some $d \geq 3$), in polynomial time we can find a pinned edge with a known value, or return that there is a contradiction. (If there is a contradiction in Ω, we may still return a purported pinned edge with a known value, which we can apply and simplify Ω. The contradiction will eventually be found.) If Ω does not contain $=_k$, or does not contain EXACTONE$_d$ functions (for some $d \geq 3$), then the problem is tractable, since Ω is an instance of $\mathscr{M}$, or an instance of $\mathscr{P}$. The lemma follows from the claim, since we either recurse on a smaller instance or have a tractable instance.

Suppose Ω is an instance where at least one $=_k$ and at least one EXACTONE$_d \in \mathcal{EO}$ appear. If a signature EXACTONE$_d \in \mathcal{EO}$ is connected to itself by a self-loop through a $\neq_2$, then the remaining $d-2 \geq 1$ edges are pinned to 0 with a factor of 2 to the Holant. Suppose two signatures EXACTONE$_d$ and EXACTONE$_\ell$ from $\mathcal{EO}$ are connected by some number of $\neq_2$'s. Depending on the number of connecting edges being 1 or 2 or ≥ 3, we replace all three signatures by EXACTONE$_{d+\ell-2}$, or find pinned edge, or return 0. We hence assume no connection between any pair of EXACTONE's.

Define an E_k-block as a connected component composed of $=_k$ and $\neq_2$. All external connecting edges of each E_k-block are marked with $+$ or $-$ and this can be found by testing bipartiteness of an E_k-block where we treat $\neq_2$'s as edges. If any E_k-block is not bipartite, then it is trivial and we return 0. We contract all E_k-blocks and maintain planarity, one edge at a time, and remove self loops. We may assume all E_k-blocks are nontrivial. If there is a nontrivial E_k-block of arity 2, its signature is $\neq_2$. We replace it with an edge assigned $\neq_2$ to form an instance Ω', maintaining planarity, such that any pinned edge in Ω' corresponds to a pinned edge in Ω. This new edge is between EXACTONE signatures and can be dealt with as described earlier. So we may assume the arity of any E_k-block is at least 4. Since $k \geq 6$, the only possible E_k-blocks of arity 4 are those in Figure 2a up to a rotation. Since there is at least one EXACTONE$_d$ signature with $d \geq 3$, forming E_k-blocks does not consume all of Ω.

After these steps we may consider Ω a bipartite graph, with one side consisting of E_k-blocks and the other side consisting of EXACTONE signatures. They are now connected by edges assigned $=_2$. It is easy to verify that parallel edges between an E_k-block and an EXACTONE$_d$ signature always lead to some pinned edges. Therefore, we may assume there are no parallel edges between any E_k-block and any EXACTONE signature.



(a) An EO-Eq-4-block. Triangles are assigned EXACTONE signatures, squares are assigned $\neq_2$, and circles are E_k-blocks of arity 4.

(b) Break the EO-Eq-4-block into three components, The one in the middle contains a cycle, and hence is degenerate. The other two are EXACTONE functions.

Figure 3: EO-Eq-4-blocks

Now consider E_k-blocks of arity 4 with EXACTONE signatures together. Call a connected component consisting of E_k-blocks of arity 4 and EXACTONE an EO-Eq-4-block. Figure 3a illustrates an example. Notice that the two possibilities of E_k-blocks of arity 4 can be viewed as two parallel $\neq_2$'s but with some correlation between them, namely their satisfying assignments are paired up in a unique way. This is illustrated in Figure 2b. Note that the two dotted lines in Figure 2b represent different correlations.

At this point, we would like to replace every arity 4 E_k-block by two parallel $\neq_2$'s. However this replacement destroys the equivalence of the Holant values, before and after.

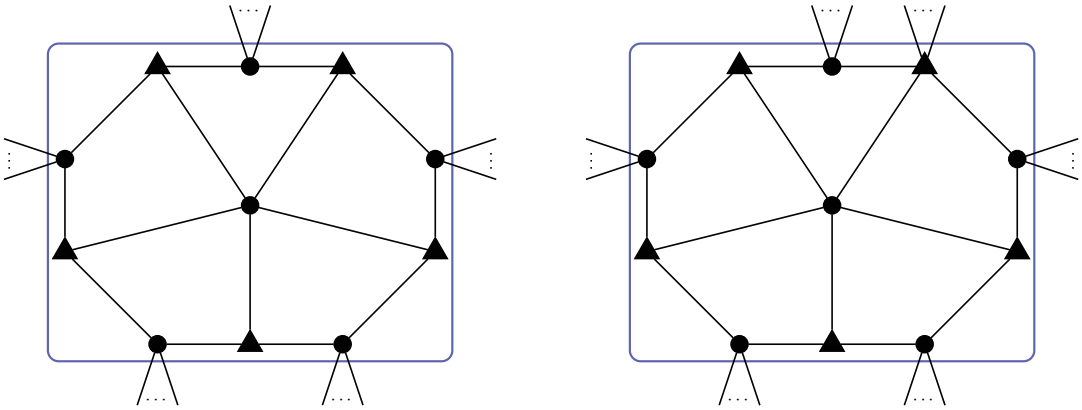
The surprising move of this proof is that we shall do so anyway!

We ignore the correlation for the time being and replace every arity 4 E_k-block by two parallel $\neq_2$'s as in Figure 2b. This replacement produces a *planar* signature grid Ω_1. Every edge in Ω_1 corresponds to a unique edge in Ω. The set of satisfying assignments of Ω_1 is a superset of that of Ω. Moreover, if there is an edge pinned in Ω_1 to a known value, the corresponding edge is also pinned in Ω to the same value. Once we find a pinned edge in Ω_1, we revert back to work in Ω and apply the pinning to the pinned edge.

All that remains to be shown is that pinning always happens in Ω_1. Each EO-Eq-4-block splits into some number of connected components in Ω_1. Figure 3b is an example. We can show that any cycle in such a component creates at least one pinned edge. Hence we may assume there are no cycles in these components, and every such component forms a tree, whose vertices are EXACTONE functions and edges are $\neq_2$'s. Suppose there are $n \geq 2$ vertices and t many leaves in such a tree. One can verify that replacing the whole tree by an EXACTONE$_t$ function of the same arity t maintains the number of satisfying assignments. Since each vertex in the tree has degree at least 3, we have $t \geq 3n - 2(n-1) = n + 2 \geq 4$. We replace these components by EXACTONE$_t$'s.

Thus, each connected component in the graph underlying Ω_1 is a planar bipartite graph with E_k-blocks of arity at least 6 on one side and EXACTONE$_d$ signatures of arity at least 3 on the other. By Lemma V.2, no component is simple, so there are parallel edges between some E_k-block and some

EXACTONE$_d$ signature. Parallel edges between two parts lead to pinned edges, and we can find a pinned edge with a known value in polynomial time. This finishes the proof. ■



(a) Type 1. (b) Type 2.

Figure 4: The wheel structures. Each circle is a E_5-block and triangle a EXACTONE function.

Unlike the situation in Lemma V.2, a planar $(5, 3)$-regular bipartite graph can be simple. However, we show that such graphs have a special structure. We call this structure a "wheel", which is pictured in Figure 4. There is an arity 5 vertex v in the middle, and all faces adjacent to this vertex must be 4-gons (i.e. quadrilaterals). Moreover, at least four neighbors of v are of degree 3. Depending on the degree of the fifth neighbor (whether it is 3 or not), we have two types of wheels, which are pictured in Figure 4a and Figure 4b.

Lemma V.4. *Let $G = (L \cup R, E)$ be a planar bipartite graph with parts L and R. Suppose every vertex in L has degree at least 5 and every vertex in R has degree at least 3. If G is simple, then there exists one of the two wheel structures in Figure 4 in G.*

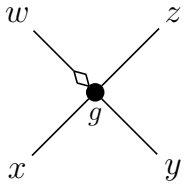
Proof: Let $V = L \cup R$ and F be the set of faces. We assign a "score" s_v on each vertex $v \in V$. We will define s_v so that $\sum_{v \in V} s_v = |V| - |E| + |F| = 2 > 0$. The base score is $+1$ for each vertex, which accounts for $|V|$. For each k-gon face, we assign $\frac{1}{k}$ to each of its vertex. This accounts for $|F|$. Notice that G is bipartite. Hence $k \geq 4$ and a score coming from a face can be at most $\frac{1}{4}$.

For $-|E|$, we need to separate two cases. If one of the two endpoints has degree 3, we give the degree 3 vertex a score of $-\frac{7}{12}$, and the other one $-\frac{5}{12}$. This is well defined because all degree 3 vertices are in R. Otherwise, we give each endpoint $-\frac{1}{2}$. This accounts for $-|E|$.

One can verify that $s_v \leq 0$ unless $v \in L$ has degree 5. Since the total score is positive, there must exist $v \in L$, v has degree 5 and $s_v > 0$. We then claim that there exists such a v so that all its adjacent faces are 4-gons. Suppose otherwise. One can show that a positively scored vertex v is adjacent to exactly one face with more than 4 edges. Call this face F_v.

In F_v, v has two neighbors in R. We match all vertices that have positive scores to their own clockwise next one in F_v. We do this matching in all faces containing at least one positively scored vertex. Suppose a vertex $u \in R$ is matched with ℓ different vertices. This means that u is adjacent to at least ℓ many k-gons with $k \geq 6$. One can verify that $s_u \leq -\frac{\ell}{12}$. It implies that the total score of u and all positively scored vertices matched with u is at most 0. However each positively scored vertex is matched with a vertex in R. Hence the total score cannot be positive. Contradiction.

Therefore there exists $v \in L$ such that $s_v > 0$, $\deg(v) = 5$, and all adjacent faces are 4-gons. We

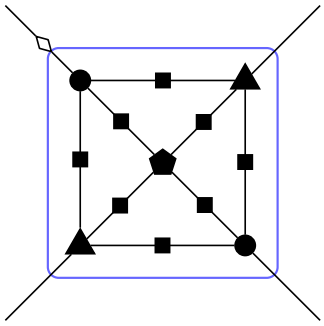


(a)

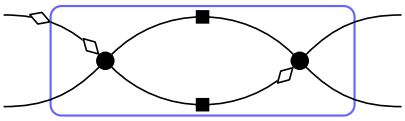
wx \ zy	00	01	10	11
00	g^{0000}	g^{0010}	g^{0001}	g^{0011}
01	g^{0100}	g^{0110}	g^{0101}	g^{0111}
10	g^{1000}	g^{1010}	g^{1001}	g^{1011}
11	g^{1100}	g^{1110}	g^{1101}	g^{1111}

(b)

Figure 5: The quaternary signature g is assigned to the vertex in (a). Its first input corresponds to the edge marked with the diamond, which is w. The order of the remaining inputs is given by traveling counterclockwise. In (b), g^{wxzy} denotes the value $g(w, x, y, z)$.



(a) $(\neq_2 | [0,1,0,0,0], [0,0,0,1,0], \hat{g})$-gate on right side



(b) Gadget with a symmetric signature matrix

Figure 6: Two gadgets used in Lemma V.6.

further note that at most one neighbor of v is of degree ≥ 4, for otherwise, $s_v \leq 0$. It is type 1 as in Figure 4a if all neighbors of v has degree 3, and is type 2 as in Figure 4b otherwise. ■

Either structure in Figure 4 leads to pinned edges. We get the following lemma, which finishes the tractability of Theorem V.1.

Lemma V.5. Pl-Holant $(\neq_2 \mid =_5, \mathcal{EO})$ *is computable in polynomial time.*

B. Hardness when $k \in \{3, 4\}$

We prove the hardness of Theorem V.1. The proofs differ for $k = 3$ and $k = 4$. For $k = 3$, we use the following technical lemma. This lemma is invoked three times in the full proof of Theorem III.1. The main challenge in these proofs is how to build planar gadgets. As discussed in the previous subsection, parallel edges lead to degeneracy in the instance. A simple calculation based on Euler's characteristic implies that a large number of vertices is necessary to avoid parallel edges.

Lemma V.6. *Let* $\hat{g}$ *be the arity* 4 *signature whose support contains only* 0101 *and* 1010 *(invariant under rotations). Then* Pl-Holant $(\neq_2 \mid [0,1,0,0,0], [0,0,0,1,0], \hat{g})$ *is #P-hard.*

Proof: For an arity 4 signature, we can express it as a 4-by-4 matrix, where rows are indexed by the two inputs on the left, and columns by the two inputs on the right in reversed order. This is depicted

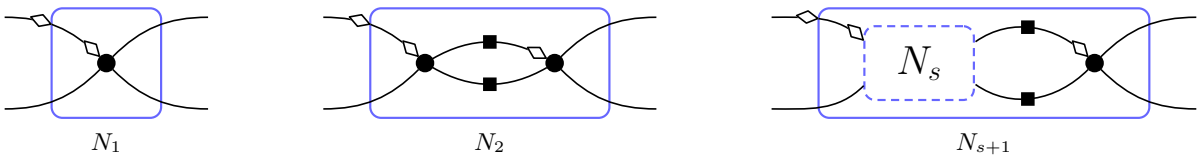
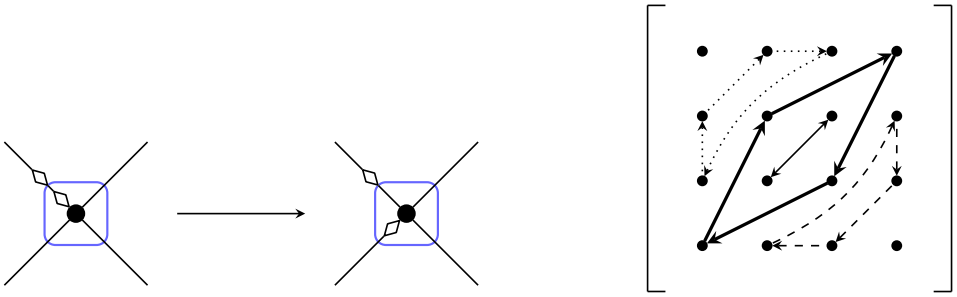


Figure 7: Linear recursive construction used for interpolation in a nonstandard basis.



(a) A counterclockwise rotation

(b) Movement of signature matrix entries

Figure 8: The movement of the entries in the signature matrix of a quaternary signature under a counterclockwise rotation of the input edges. Entires of Hamming weight 1 are in the dotted cycle, entires of Hamming weight 2 are in the two solid cycles (one has length 4 and the other one is a swap), and entries of Hamming weight 3 are in the dashed cycle.

in Figure 5. With this notation, sequential connections correspond to matrix multiplications.

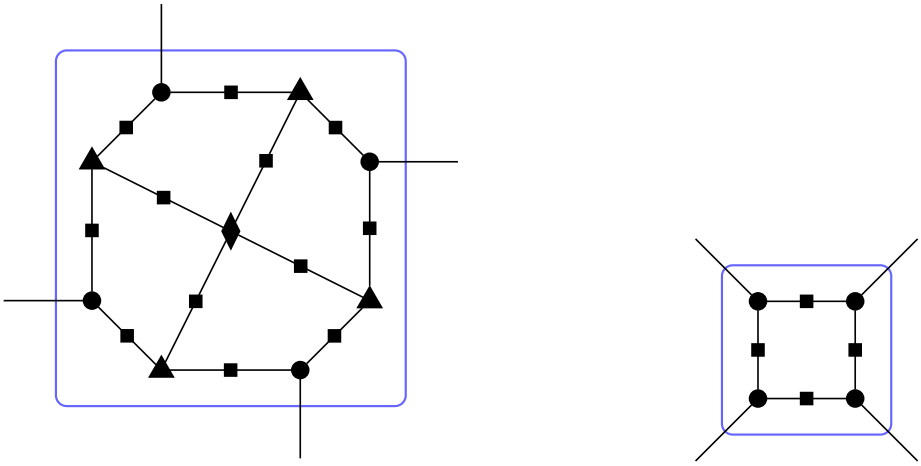
Consider the gadget in Figure 6a. We assign $[0,0,0,1,0]$ to triangles, $[0,1,0,0,0]$ to circles, $\hat{g}$ to the pentagon, and $[0,1,0]$ to squares. The resulting signature is $\hat{h}$ with $M_{\hat{h}} = \begin{bmatrix} 0&0&0&1 \\ 0&1&3&0 \\ 0&1&1&0 \\ 1&0&0&0 \end{bmatrix}$. Consider the gadget in Figure 6b. We assign $\hat{h}$ to circles and $[0,1,0]$ to squares. The resulting signature is $\hat{r}$ with $M_{\hat{r}} = \begin{bmatrix} 0&0&0&1 \\ 0&6&4&0 \\ 0&4&2&0 \\ 1&0&0&0 \end{bmatrix}$. We use $\hat{r}$ to interpolate a signature $\hat{r}'$ with $M_{\hat{r}'} = \begin{bmatrix} 0&0&0&1 \\ 0&3&1&0 \\ 0&1&1&0 \\ 1&0&0&0 \end{bmatrix}$. Consider an instance Ω of Pl-Holant $(\neq_2 \mid \hat{r}')$. Suppose that $\hat{r}'$ appears n times in Ω. We construct from Ω a sequence of instances Ω_s of Pl-Holant $(\neq_2 \mid \hat{r})$ indexed by $s \geq 1$. We obtain Ω_s from Ω by replacing each occurrence of $\hat{r}'$ with the gadget N_s in Figure 7 with $\hat{r}$ assigned to circles and $[0,1,0]$ assigned to squares. In Ω_s, the edge corresponding to the ith significant index bit of N_s connects to the same location as the edge corresponding to the ith bit of $\hat{r}'$ in Ω.

The signature matrix of $\hat{r}'$ is $M_{\hat{r}'} = XPDP^{-1}$ where $X = \begin{bmatrix} 0&0&0&1 \\ 0&0&1&0 \\ 0&1&0&0 \\ 1&0&0&0 \end{bmatrix}$, $P = \begin{bmatrix} 1&0&0&0 \\ 0&1&1&0 \\ 0&\sqrt{3}&-\sqrt{3}&0 \\ 0&0&0&1 \end{bmatrix}$, and D is a diagonal matrix $\operatorname{diag}\left(1, 1+\sqrt{3}, 1-\sqrt{3}, 1\right)$. The signature matrix of N_s is $M_{N_s} = X(XM_{\hat{r}})^s = XPD_1^sP^{-1}$, where $D_1 = \operatorname{diag}\left(1, 4+2\sqrt{3}, 4-2\sqrt{3}, 1\right)$. We can view our construction of Ω_s as first replacing $M_{\hat{r}'}$ with $XPDP^{-1}$ (each matrix corresponds to a vertex), which does not change the Holant value, and then replacing D with D_1^s.

We stratify the assignments in Ω based on the assignments to the n occurrences of the signature corresponding to D. We only need to consider the assignments that assign i many times the bit patterns 0000 or 1111, j many times the bit pattern 0110, and k many times the bit pattern 1001, since any other assignment contributes a factor of 0. Let c_{ijk} be the sum over all such assignments of the products of evaluations of all other signatures (those corresponding to X, P, and P^{-1}) in Ω except for those corresponding to D. Then $\operatorname{Holant}_\Omega = \sum_{i+j+k=n} \left(1+\sqrt{3}\right)^j \left(1-\sqrt{3}\right)^k c_{ijk}$ and the value of the Holant

on Ω_s, for $s \geq 1$, is $\mathrm{Holant}_{\Omega_s} = \sum_{i+j+k=n} \left(\left(4+2\sqrt{3}\right)^{j-k} 4^k \right)^s c_{ijk}$. We view c_{ijk} as unknown variables to be solved, and the Vandermonde system given by $\mathrm{Holant}_{\Omega_s}$ has full rank. To see this, we only need to show that $\left(4+2\sqrt{3}\right)^{j-k} 4^k \neq \left(4+2\sqrt{3}\right)^{j'-k'} 4^{k'}$ unless $(j,k) = (j',k')$. If $\left(4+2\sqrt{3}\right)^{j-k} 4^k = \left(4+2\sqrt{3}\right)^{j'-k'} 4^{k'}$, then we have $\left(4+2\sqrt{3}\right)^{j-k-(j'-k')} 4^{k-k'} = 1$. Since any nonzero integer power of $4+2\sqrt{3}$ is not rational, we have $j-k = j'-k'$, so $k = k'$ and $j = j'$.

Therefore, by polynomially many oracle calls to $\mathrm{Holant}_{\Omega_s}$, we can solve for the unknown c_{ijk}'s and obtain Holant_{Ω}. After a counterclockwise rotation of $\hat{r}'$ (cf. Figure 8), we get a nonsingular redundant matrix. The hardness follows (cf. Corollary 2.31 in the full version [8]). ■



(a) A cycle-like gadget used twice.

(b) A gadget to realize $\hat{g}$.

Figure 9: Two gadgets in the proof of Lemma V.7.

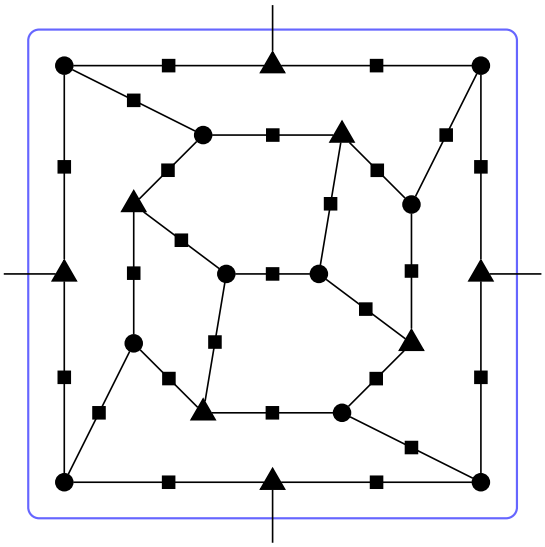


Figure 10: The whole gadget to realize $[0,0,0,1,0]$.

Lemma V.7. Pl-Holant $(\neq_2 \mid =_3, [0,1,0,0])$ *is #P-hard.*

Proof: By connecting two copies of $[0,1,0,0]$ together via $\neq_2$, we have $[0,1,0,0,0]$ on the right. Consider the gadget in Figure 9a. We assign $=_3$ to triangles, $[0,1,0,0]$ to circles, $[0,1,0,0,0]$ to the diamond, and $\neq_2$ to squares. Let f be the signature of this gadget. The support of f is $\{0011, 0110, 1100, 1001\}$. We construct the gadget in Figure 9a again. This time we assign $[0,1,0,0]$ to triangles, $=_3$ to circles, f to the diamond, and $\neq_2$ to squares. The resulting signature has support $\{0111, 1011, 1101, 1110\}$, and therefore is $[0,0,0,1,0]$. The whole gadget is illustrated in Figure 10, where circles are assigned $[0,1,0,0]$, triangles $=_3$, and squares $\neq_2$.

Consider the gadget in Figure 9b. We assign $=_3$ to circles and $\neq_2$ to squares. It follows that the support of the resulting signature is $\{0101, 1010\}$. This is $\hat{g}$ from Lemma V.6. We have constructed $[0,1,0,0,0]$, $[0,0,0,1,0]$, and $\hat{g}$, all on the right, so we are done by Lemma V.6. ■

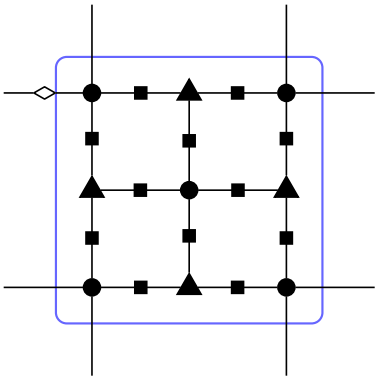


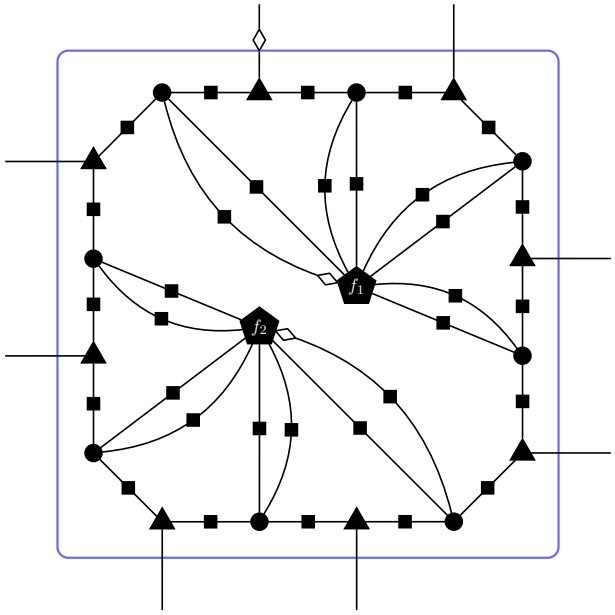
Figure 11: A grid-like gadget used in the proof of Lemma V.8, whose support vectors are 00110011, 11001100, and 11111111.

Lemma V.8. Pl-Holant $(\neq_2 \mid =_4, [0,1,0,0])$ *is #P-hard.*

Proof: Consider the gadget in Figure 11. We assign $\neq_2$ to squares, $=_4$ to circles, and $[0,1,0,0]$ to triangles. The resulting signature has support $\{00110011, 11001100, 11111111\}$, where each vector is the assignment ordered clockwise starting from the diamond. Every two wires at each corner are always of the same value. Further connect each corner to a $=_4$ via two copies of $\neq_2$, resulting in a signature f whose support is $\{11001100, 00110011, 00000000\}$, reversing the original.

Consider the gadget in Figure 12a. We assign $\neq_2$ to squares, $=_4$ to circles, $[0,1,0,0]$ to triangles, and f to pentagons. Each pair of parallel edges coming out of f are at the same corner of f. We call the pentagon above f_1, the one below f_2, and the resulting signature g. We order the inputs to f_1, f_2, and g clockwise starting from the diamond-marked edge. With this notation, we get Table 12b listing the support of g.

The support of g is $\{11111111, 00001111, 0001110, 11110000, 00000000, 11100001\}$, and 00000000 has multiplicity 2. We pair adjacent outputs clockwise, starting from the diamond. We treat g as an arity 4 signature, using $=_4$ to do a domain pairing argument. In the paired domain, $=_4$ becomes $=_2$, which lifts the bipartite restriction. Moreover, 0001110 and 11100001 in the support of g are eliminated as they do not agree on adjacent paired outputs. So in the paired domain, the support of g is $\{1111, 0011, 1100, 0000\}$ with multiplicity 2 for 0000. We rotate g so that the support is $\{1111, 0110, 1001, 0000\}$. The arity 4 signature matrix of g is $\mathrm{diag}(2,1,1,1)$. We can show that Pl-#CSP$([2,1,1]) \leq_T$ Pl-Holant(g) (cf. Lemma



(a) The gadget.

f_1	f_2	g
00000000	00000000	11111111
00110011	00000000	00001111
11001100	00000000	00011110
00000000	00110011	11110000
00110011	00110011	00000000
11001100	00110011	-
00000000	11001100	11100001
00110011	11001100	-
11001100	11001100	00000000

(b) The table of supports.

Figure 12: Another gadget used in the proof of Lemma V.8.

7.2 in the full version [8]), $\text{Pl-Holant}(g)$ is #P-hard by the planar #CSP dichotomy (cf. Theorem 2.27 in the full version [8]), so we are done. ■

ACKNOWLEDGMENT

We are thankful to Pinyan Lu who discussed with us in an early stage of this work. We also thank the anonymous referees for their helpful comments. All authors were supported by NSF CCF-1217549. Heng Guo was also supported by a Simons Award for Graduate Students in Theoretical Computer Science from the Simons Foundation. Tyson Williams was also supported by a Cisco Systems Distinguished Graduate Fellowship.

REFERENCES

[1] Rodney J. Baxter. *Exactly solved models in statistical mechanics*. Academic press London, 1982.

[2] Jin-Yi Cai, Xi Chen, Richard J. Lipton, and Pinyan Lu. On tractable exponential sums. In *FAW*, pages 148–159. Springer Berlin Heidelberg, 2010.

[3] Jin-Yi Cai and Vinay Choudhary. Some results on matchgates and holographic algorithms. *Int. J. Software and Informatics*, 1(1):3–36, 2007.

[4] Jin-Yi Cai, Vinay Choudhary, and Pinyan Lu. On the theory of matchgate computations. *Theory Comput. Syst.*, 45(1):108–132, 2009.

[5] Jin-Yi Cai and Michael Kowalczyk. Spin systems on k-regular graphs with complex edge functions. *Theoretical Computer Science*, 2012.

[6] Jin-Yi Cai, Heng Guo, and Tyson Williams. A complete dichotomy rises from the capture of vanishing signatures (extended abstract). In *STOC*, pages 635–644. ACM, 2013. *CoRR*, abs/1204.6445.

[7] Jin-Yi Cai, Heng Guo, and Tyson Williams. Holographic algorithms beyond matchgates. In *ICALP*, pages 271–282. Springer Berlin Heidelberg, 2014. *CoRR*, abs/1307.7430.

[8] Jin-Yi Cai, Zhiguo Fu, Heng Guo, and Tyson Williams. A Holant Dichotomy: Is the FKT Algorithm Universal? *CoRR*, abs/1505.02993, 2015.

[9] Jin-Yi Cai, Michael Kowalczyk, and Tyson Williams. Gadgets and anti-gadgets leading to a complexity dichotomy. In *ITCS*, pages 452–467. ACM, 2012.

[10] Jin-Yi Cai and Pinyan Lu. On symmetric signatures in holographic algorithms. *Theory Comput. Syst.*, 46(3):398–415, 2010.

[11] Jin-Yi Cai and Pinyan Lu. Holographic algorithms: From art to science. *J. Comput. Syst. Sci.*, 77(1):41–61, 2011.

[12] Jin-Yi Cai, Pinyan Lu, and Mingji Xia. Holographic algorithms with matchgates capture precisely tractable planar #CSP. In *FOCS*, pages 427–436. IEEE Computer Society, 2010.

[13] Jin-Yi Cai, Pinyan Lu, and Mingji Xia. Computational complexity of Holant problems. *SIAM J. Comput.*, 40(4):1101–1132, 2011.

[14] Jin-Yi Cai, Pinyan Lu, and Mingji Xia. The complexity of complex weighted Boolean #CSP. *J. Comput. System Sci.*, 80(1):217–236, 2014.

[15] Jan Draisma, Dion C. Gijswijt, László Lovász, Guus Regts, and Alexander Schrijver. Characterizing partition functions of the vertex model. *J. Algebra*, 350:197–206, 2012.

[16] Michael Freedman, László Lovász, and Alexander Schrijver. Reflection positivity, rank connectivity, and homomorphism of graphs. *J. Amer. Math. Soc.*, 20(1):37–51, 2007.

[17] Heng Guo, Pinyan Lu, and Leslie G. Valiant. The complexity of symmetric Boolean parity Holant problems. *SIAM J. Comput.*, 42(1):324–356, 2013.

[18] Heng Guo and Tyson Williams. The complexity of planar Boolean #CSP with complex weights. In *ICALP*, pages 516–527. Springer Berlin Heidelberg, 2013. *CoRR*, abs/1212.2284.

[19] Sangxia Huang and Pinyan Lu. A dichotomy for real weighted Holant problems. In *CCC*, pages 96–106. IEEE Computer Society, 2012. Full version available at `http://www.csc.kth.se/~sangxia/papers/2012-ccc.pdf`.

[20] Ernst Ising. Beitrag zür theorie des ferromagnetismus. *Zeitschrift für Physik*, 31(1):253–258, 1925.

[21] P. W. Kasteleyn. The statistics of dimers on a lattice. *Physica*, 27:1209–1225, 1961.

[22] P. W. Kasteleyn. Graph theory and crystal physics. In F. Harary, editor, *Graph Theory and Theoretical Physics*, pages 43–110. Academic Press, London, 1967.

[23] Michael Kowalczyk. *Dichotomy theorems for Holant problems*. PhD thesis, University of Wisconsin—Madison, 2010. `http://cs.nmu.edu/~mkowalcz/research/main.pdf`.

[24] J. M. Landsberg, Jason Morton, and Serguei Norine. Holographic algorithms without matchgates. *Linear Algebra Appl.*, 438(2):782–795, 2013.

[25] T. D. Lee and C. N. Yang. Statistical theory of equations of state and phase transitions. II. Lattice gas and Ising model. *Phys. Rev.*, 87(3):410–419, 1952.

[26] Elliott H. Lieb. Residual entropy of square ice. *Phys. Rev.*, 162(1):162–172, 1967.

[27] Elliott H. Lieb and Alan D. Sokal. A general Lee-Yang theorem for one-component and multicomponent ferromagnets. *Comm. Math. Phys.*, 80(2):153–179, 1981.

[28] Jason Morton. Pfaffian circuits. *CoRR*, abs/1101.0129, 2011.

[29] Jason Morton and Susan Margulies. Polynomial-time solvable #CSP problems via algebraic models and Pfaffian circuits. *CoRR*, abs/1311.4066, 2013. To appear in Journal of Symbolic Computation.

[30] Lars Onsager. Crystal statistics. I. A two-dimensional model with an order-disorder transition. *Phys. Rev.*, 65(3-4):117–149, 1944.

[31] Alexander Schrijver. Characterizing partition functions of the spin model by rank growth. *Indag. Math. (N.S.)*, 24(4):1018–1023, 2013.

[32] H. N. V. Temperley and M. E. Fisher. Dimer problem in statistical mechanics—an exact result. *Philosophical Magazine*, 6:1061–1063, 1961.

[33] Leslie G. Valiant. Expressiveness of matchgates. *Theor. Comput. Sci.*, 289(1):457–471, 2002.

[34] Leslie G. Valiant. Quantum circuits that can be simulated classically in polynomial time. *SIAM J. Comput.*, 31(4):1229–1254, 2002.

[35] Leslie G. Valiant. Accidental algorthims. In *FOCS*, pages 509–517. IEEE Computer Society, 2006.

[36] Leslie G. Valiant. Holographic algorithms. *SIAM J. Comput.*, 37(5):1565–1594, 2008.

[37] Leslie G. Valiant. Some observations on holographic algorithms. In *LATIN*, pages 577–590. Springer Berlin Heidelberg, 2010.

[38] Dirk Vertigan. The computational complexity of Tutte invariants for planar graphs. *SIAM Journal on Computing*, 35(3):690–712, 2005.

[39] Dirk Llewellyn Vertigan. *On the computational complexity of Tutte, Jones, Homfly and Kauffman invariants*. PhD thesis, University of Oxford, 1991.

[40] Dominic Welsh. *Complexity: Knots, Colourings and Countings*. London Mathematical Society Lecture Note Series. Cambridge University Press, 1993.

[41] C. N. Yang. The spontaneous magnetization of a two-dimensional Ising model. *Phys. Rev.*, 85(5):808–816, 1952.

[42] C. N. Yang and T. D. Lee. Statistical theory of equations of state and phase transitions. I. Theory of condensation. *Phys. Rev.*, 87(3):404–409, 1952.

Full version of Paper 3 attached

A Holant Dichotomy: Is the FKT Algorithm Universal?

Jin-Yi Cai* jyc@cs.wisc.edu
Zhiguo Fu*† zfu8@wisc.edu
Heng Guo*‡ hguo@cs.wisc.edu
Tyson Williams*§ tdw@cs.wisc.edu

Abstract

We prove a complexity dichotomy for complex-weighted Holant problems with an arbitrary set of symmetric constraint functions on Boolean variables. This dichotomy is specifically to answer the question: Is the FKT algorithm under a holographic transformation [38] a *universal* strategy to obtain polynomial-time algorithms for problems over planar graphs that are intractable in general? This dichotomy is a culmination of previous ones, including those for Spin Systems [25], Holant [21, 6], and #CSP [20].

In the study of counting complexity, such as #CSP, there are problems which are #P-hard over general graphs but polynomial-time solvable over planar graphs. A recurring theme has been that a holographic reduction to FKT precisely captures these problems. Surprisingly, for planar Holant, we discover new planar tractable problems that are not expressible by a holographic reduction to FKT. In particular, a straightforward formulation of a dichotomy for planar Holant problems along the above recurring theme is false.

In previous work, an important tool was a dichotomy for #CSP^d, which denotes #CSP where every variable appears a multiple of d times. However the very first step in the #CSP^d dichotomy proof fundamentally violates planarity. In fact, due to our newly discovered tractable problems, the putative form of a planar #CSP^d dichotomy is false when $d \geq 5$. Nevertheless, we prove a dichotomy for planar #CSP^2. In this case, the putative form of the dichotomy is true. We manage to prove the planar Holant dichotomy without relying on a planar #CSP^d dichotomy for $d \geq 3$, while the dichotomy for planar #CSP^2 plays an essential role.

As a special case of our new planar tractable problems, counting perfect matchings (#PM) over k-uniform hypergraphs is polynomial-time computable when the incidence graph is planar and $k \geq 5$. The same problem is #P-hard when $k = 3$ or $k = 4$, which is also a consequence of our dichotomy. When $k = 2$, it becomes #PM over planar graphs and is tractable again. More generally, over hypergraphs with specified hyperedge sizes and the same planarity assumption, #PM is polynomial-time computable if the greatest common divisor (gcd) of all hyperedge sizes is at least 5. It is worth noting that it is the gcd, and not a bound on hyperedge sizes, that is the criterion for tractability.

*University of Wisconsin–Madison. Supported by NSF CCF-1217549.

†School of Mathematics, Jilin University

‡Also supported by a Simons Award for Graduate Students in Theoretical Computer Science from the Simons Foundation.

§Also supported by a Cisco Systems Distinguished Graduate Fellowship.

1 Introduction

The Fisher-Kasteleyn-Temperley (FKT) algorithm [34, 23, 24] is a classical gem that counts perfect matchings over planar graphs in polynomial time. This was an important milestone in a decades-long research program by physicists in statistical mechanics to determine what is known as Exactly Solved Models [1, 22, 32, 43, 44, 27, 34, 23, 24, 28, 29, 42].

For four decades, the FKT algorithm stood as *the* polynomial-time algorithm for any counting problem over planar graphs that is #P-hard over general graphs. Then Valiant introduced *matchgates* [36, 35] and *holographic* reductions to the FKT algorithm [38, 37]. These reductions differ from classical ones by introducing quantum-like superpositions. This novel technique extended the reach of the FKT algorithm and produced polynomial-time algorithms for a number of problems for which only exponential-time algorithms were previously known.

Since the new polynomial-time algorithms appear so exotic and unexpected, and since they solve problems that appear so close to being #P-hard, they challenge our faith in the well-accepted conjecture that P $\neq$ NP. Quoting Valiant [37]: "The objects enumerated are sets of polynomial systems such that the solvability of any one member would give a polynomial time algorithm for a specific problem. ... the situation with the P = NP question is not dissimilar to that of other unresolved enumerative conjectures in mathematics. The possibility that accidental or freak objects in the enumeration exist cannot be discounted if the objects in the enumeration have not been studied systematically." Indeed, if any "freak" object exists in this framework, it would collapse #P to P.

Therefore, over the past 10 to 15 years, this technique has been intensely studied in order to gain a systematic understanding to the limit of the trio of holographic reductions, matchgates, and the FKT algorithm [35, 3, 4, 10, 39, 11, 26, 30, 31]. Without settling the P versus #P question, the best hope is to achieve a complexity classification. This program finds its sharpest expression in a complexity dichotomy theorem, which classifies *every* problem expressible in a framework as either solvable in P or #P-hard, with nothing in between.

Out of this work, a strong theme has emerged. For a wide variety of problems, such as those expressible as a #CSP, holographic reductions to the FKT algorithm is a *universal* technique for turning problems that are #P-hard in general to P-time solvable over planar graphs. In fact, a preponderance of evidence suggests the following putative classification of all counting problems defined by local constraints into *exactly* three categories: (1) those that are P-time solvable over general graphs; (2) those that are P-time solvable over planar graphs but #P-hard over general graphs; and (3) those that remain #P-hard over planar graphs. Moreover, category (2) consists precisely of those problems that are holographically reducible to the FKT algorithm. This theme is so strong that it has become an intuitive and trusty guide for us when we investigate unknown problems and plan proof strategies. In fact, many of the results in the present paper were proved in this way. However, one is still left wondering whether the FKT algorithm is *universal*, or more precisely, is the combined algorithmic power of the trio sufficient to capture all tractable problems over planar graphs that are intractable in general?

We list some of the supporting evidence for this putative classification. These date back to the classification of the complexity of the Tutte polynomial [41, 40]. It has also been an unfailing theme in the classification of spin systems and #CSP [25, 12, 9, 20]. However, these frameworks do not capture all locally specified counting problems. Some natural problems, such as counting perfect matchings (#PM), are not expressible as a point on the Tutte polynomial or a #CSP, and #PM is provably not expressible within the special case of vertex assignment models [18, 17, 33].

However, this is the problem for which FKT was designed, and is the basis of Valiant's matchgates and holographic reductions.

A refined framework, called Holant problems [13], was proposed to address this issue. It is an edge assignment model. It naturally encodes and expresses #PM as well as Valiant's matchgates and holographic reductions. Thus, Holant is the proper framework in which to study the power of holographic algorithms. It is also more general than #CSP in the sense that a complete complexity classification for Holant problems implies one for #CSP.

In this paper, we classify for the first time the complexity of Holant problems over planar graphs. Our result generalizes both the dichotomy for Holant [21, 6] and the dichotomy for planar #CSP [12, 20]. Surprisingly, we discover new planar tractable problems that are not expressible by a holographic reduction to matchgates and FKT. To the best of our knowledge, this is the first *primitive* extension since FKT to a problem solvable in P over planar instances but #P-hard in general. Furthermore, our dichotomy theorem says that this completes the picture: there are no more undiscovered extensions for problems expressible in this framework, unless #P collapses to P. In particular, the putative form of the planar Holant dichotomy is *false*.

Before stating our main theorem, we give a brief description of the Holant framework [13]. Fix a set of local constraint functions $\mathcal{F}$. A *signature grid* $\Omega = (G, \pi)$ is a tuple, where $G = (V, E)$ is a graph, π labels each $v \in V$ with a function $f_v \in \mathcal{F}$ with input variables from the incident edges $E(v)$ at v. Each f_v maps $\{0,1\}^{\deg(v)}$ to $\mathbb{C}$. We consider all 0-1 edge assignments. An assignment σ for every $e \in E$ gives an evaluation $\prod_{v \in V} f_v(\sigma \mid_{E(v)})$, where $\sigma \mid_{E(v)}$ denotes the restriction of σ to $E(v)$. The counting problem on the instance Ω is to compute

$$\text{Holant}(\Omega; \mathcal{F}) = \sum_{\sigma: E \to \{0,1\}} \prod_{v \in V} f_v \left(\sigma \mid_{E(v)}\right). \tag{1.1}$$

For example, #PM, the problem of counting perfect matchings in G, corresponds to assigning the ExactOne function at every vertex of G. The Holant problem parameterized by the set $\mathcal{F}$ is denoted by Holant($\mathcal{F}$).

At a high level, we can state our main theorem as follows.

Theorem 1.1. *Let $\mathcal{F}$ be a set of complex-valued, symmetric functions on Boolean variables. Then there is an effective classification for all possible $\mathcal{F}$, according to which,* Holant($\mathcal{F}$) *is either (1) P-time computable over general graphs, or (2) P-time computable over planar graphs but #P-hard over general graphs, or (3) #P-hard over planar graphs.*

The complete statement is given in Theorem 8.1. The classification is explicit. The tractability criterion is decidable in polynomial time due to [11, 7]. Tractable problems over general graphs have been previously studied in [6]. The planar tractable class includes both those solvable by holographic reductions to FKT and those newly discovered. Explicit criteria for these are also proved in this paper.

Let us meet some new tractable problems. They can be described as orientation problems, which are Holant problems after a complex-valued holographic transformation.[1] Given a planar graph, we allow two kinds of vertices. The first kind can be either a sink or a source while the second kind

[1]This transformation is $Z = \left[\begin{smallmatrix} 1 & 1 \\ i & -i \end{smallmatrix}\right]$. It is common that one problem can be transformed to another over $\mathbb{C}$ while one or both problems are specified by real-valued constraint functions, and provably no transformation exists over $\mathbb{R}$. Thus to study the classification question over complex-valued constraint functions is natural and proper. For example, the integer-valued orientation problem studied here is complex weighted if expressed directly as Holant.

only allow one incoming edge. The goal is to compute the number of orientations satisfying these constraints. This problem can be expressed in the Holant framework under a Z-transformation. It can be shown that this is equivalent to the Holant problem on the edge-vertex incidence graph where we assign the DISEQUALITY function to every edge, and to each vertex, we assign either the EQUALITY function or the EXACTONE function. Suppose vertices assigned EQUALITY functions all have degree k. If $k = 2$, then this problem can be solved by FKT. We show that this problem is #P-hard if $k = 3$ or $k = 4$, but is tractable again if $k \geq 5$. The algorithm involves a recursive procedure that simplifies the instance until it can be solved by known algorithms, including FKT. The algorithm crucially uses global topological properties of a planar graph, in particular Euler's characteristic formula. If the graph is not planar, then this algorithm does not work, and indeed the problem is #P-hard over general graphs.

More generally, we allow vertices of arbitrary degrees to be assigned EQUALITY. If all the degrees are at most 2, then the problem is tractable by the FKT algorithm. Otherwise, the complexity depends on the greatest common divisor (gcd) of the degrees. The problem is tractable if gcd ≥ 5 and #P-hard if gcd ≤ 4. It is worth noting that the criterion for tractability is not a degree lower bound. Moreover, the planarity assumption and the degree rigidity pose a formidable challenge in the hardness proofs for gcd ≤ 4.

If the graph is bipartite with EQUALITY functions assigned on one side and EXACTONE functions on the other, then this is the problem of #PM over hypergraphs with planar incidence graphs. Our results imply that the complexity of this problem depends on the gcd of the hyperedge sizes. The problem is computable in polynomial time when gcd ≥ 5 and is #P-hard when gcd ≤ 4 (assuming there are hyperedges of size at least 3). For a formal statement, see Theorem 7.15.

Most of the reductions in previous Holant dichotomy theorems [21, 6] do not hold for planar graphs, so we are forced to develop new techniques and formulate new proof strategies. In particular, an important ingredient in previous proofs is the #CSP^d dichotomy by Huang and Lu [21]. Here #CSP^d denotes #CSP where every variable appears a multiple of d times. The very first step in the #CSP^d dichotomy proof uses the popular pinning technique. Multiple copies of an instance are created and vertices are connected across different copies. But this construction fundamentally violates planarity. Moreover, this violation of planarity is unavoidable, a consequence of the new dichotomy. Due to our newly discovered tractable problems, the putative form of a planar #CSP^d dichotomy is *false* when $d \geq 5$. Nevertheless, we prove a dichotomy for planar #CSP^2 for which the putative form is, luckily for us, true (but not obvious in hindsight). Obtaining a dichotomy for planar #CSP^2 is essential because it captures a significant fraction of planar Holant problems either directly or through reductions. We manage to prove the planar Holant dichotomy without appealing to planar #CSP^d for $d \geq 3$.

The proof of the planar #CSP^2 dichotomy comprises the entire Part II of this paper that starts on page 63. A brief outline of the proof is given in Section 5 of Part I. Among the concepts and techniques introduced are some special tractable families of constraint functions specific to the #CSP^2 framework. We also introduce a *derivative* ∂ and its inverse operator *integral* $\int$ to streamline the proof argument. There is also an application of the theory of *cyclotomic fields*.

We began this project expecting to prove the putative form of the planar Holant dichotomy. It was determined that a planar #CSP^d dichotomy in the putative form would be both a more modest, and thus hopefully more attainable, intermediate step as well as a good launch station for the final goal. However after some attempt, even the planar #CSP^d dichotomy appeared too difficult to achieve, and so we scaled back the ambition to prove just a planar #CSP^2 dichotomy.

Luckily, a successful #CSP^2 dichotomy can carry most of the weight of a full #CSP^d dichotomy, *and*, as it turned out, the putative form of the planar #CSP^2 dichotomy is *true* while that for planar #CSP^d is not. Ironically, many steps of our proof in this paper were guided by the putative form of the complexity classification. The discovery of the new tractable problems changed the original plan, but also helped complete the picture.

Coming back to the challenge of the P vs. NP question posed by Valiant's holographic algorithms, we venture the opinion that the dichotomy theorem provides a satisfactory answer. Indeed, it would be difficult to conceive a world where #P is P, and yet all this algebraic theory can somehow maintain a consistent, sharp but faux division where there is none.

2 Preliminaries

2.1 Problems and Definitions

The framework of Holant problems is defined for functions mapping any $[q]^n \to R$ for a finite q and some commutative semiring R. In this paper, we investigate complex-weighted Boolean Holant problems, that is, all functions are of the form $[2]^n \to \mathbb{C}$. For consideration of models of computation, functions take complex algebraic numbers.

Graphs may have self-loops and parallel edges. A graph without self-loops or parallel edges is a *simple* graph. Fix a set of local constraint functions $\mathcal{F}$. A *signature grid* $\Omega = (G, \pi)$ consists of a graph $G = (V, E)$, where π assigns to each vertex $v \in V$ and its incident edges some $f_v \in \mathcal{F}$ and its input variables. We say that Ω is a *planar signature grid* if G is planar, where the variables of f_v are ordered counterclockwise starting from an edge specified by π. The Holant problem on instance Ω is to evaluate $\text{Holant}(\Omega; \mathcal{F}) = \sum_\sigma \prod_{v \in V} f_v(\sigma \mid_{E(v)})$, a sum over all edge assignments $\sigma : E \to \{0, 1\}$, where $E(v)$ denotes the incident edges of v and $\sigma \mid_{E(v)}$ denotes the restriction of σ to $E(v)$. We write G in place of Ω when π is clear from context.

A function f_v can be represented by listing its values in lexicographical order as in a truth table, which is a vector in $\mathbb{C}^{2^{\deg(v)}}$, or as a tensor in $(\mathbb{C}^2)^{\otimes \deg(v)}$. A function $f \in \mathcal{F}$ is also called a *signature*. A symmetric signature f on n Boolean variables can be expressed as $[f_0, f_1, \ldots, f_n]$, where f_w is the value of f on inputs of Hamming weight w. In this paper, we consider symmetric signatures. An example is the EQUALITY signature $=_n$ of arity n.

A Holant problem is parametrized by a set of signatures.

Definition 2.1. *Given a set of signatures $\mathcal{F}$, we define the counting problem* Holant($\mathcal{F}$) *as:*
Input: A signature grid $\Omega = (G, \pi)$*;*
Output: Holant($\Omega; \mathcal{F}$).

The problem Pl-Holant($\mathcal{F}$) is defined similarly using a planar signature grid.

A signature f of arity n is *degenerate* if there exist unary signatures $u_j \in \mathbb{C}^2$ $(1 \le j \le n)$ such that $f = u_1 \otimes \cdots \otimes u_n$. A symmetric degenerate signature has the form $u^{\otimes n}$. Replacing such signatures by n copies of the corresponding unary signature does not change the Holant value. Replacing a signature $f \in \mathcal{F}$ by a constant multiple cf, where $c \neq 0$, does not change the complexity of Holant($\mathcal{F}$). In this paper, we may say we obtain a signature f when in fact we have obtained a signature cf for some $c \neq 0$. It introduces a global nonzero factor to Holant($\Omega; \mathcal{F}$).

We allow $\mathcal{F}$ to be an infinite set. For Pl-Holant($\mathcal{F}$) to be tractable, the problem must be computable in polynomial time even when the description of the signatures in the input Ω are

included in the input size. In contrast, we say Pl-Holant($\mathcal{F}$) is #P-hard if there exists a finite subset of $\mathcal{F}$ for which the problem is #P-hard. We say a signature set $\mathcal{F}$ is tractable (resp. #P-hard) if the corresponding counting problem Pl-Holant($\mathcal{F}$) is tractable (resp. #P-hard). Similarly for a signature f, we say f is tractable (resp. #P-hard) if $\{f\}$ is. We follow the usual conventions about polynomial time Turing reduction $\leq_T$ and polynomial time Turing equivalence $\equiv_T$.

2.2 Holographic Reduction

To introduce the idea of holographic reductions, it is convenient to consider bipartite graphs. For a general graph, we can always transform it into a bipartite graph while preserving the Holant value, as follows. For each edge in the graph, we replace it by a path of length two. (This operation is called the *2-stretch* of the graph and yields the edge-vertex incidence graph.) Each new vertex is assigned the binary EQUALITY signature $(=_2) = [1, 0, 1]$.

We use Holant $(\mathcal{F} \mid \mathcal{G})$ to denote the Holant problem over signature grids with a bipartite graph $H = (U, V, E)$, where each vertex in U or V is assigned a signature in $\mathcal{F}$ or $\mathcal{G}$, respectively. Signatures in $\mathcal{F}$ are considered as row vectors (or covariant tensors); signatures in $\mathcal{G}$ are considered as column vectors (or contravariant tensors) [16]. Similarly, Pl-Holant $(\mathcal{F} \mid \mathcal{G})$ denotes the Holant problem over signature grids with a planar bipartite graph.

For a 2-by-2 matrix T and a signature set $\mathcal{F}$, define $T\mathcal{F} = \{g \mid \exists f \in \mathcal{F} \text{ of arity } n,\ g = T^{\otimes n} f\}$, and similarly for $\mathcal{F}T$. Whenever we write $T^{\otimes n} f$ or $T\mathcal{F}$, we view the signatures as column vectors; similarly for $fT^{\otimes n}$ or $\mathcal{F}T$ as row vectors. In the special case that $T = \left[\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}\right]$, we also define $T\mathcal{F} = \widehat{\mathcal{F}}$.

Let T be an invertible 2-by-2 matrix. The holographic transformation defined by T is the following operation: given a signature grid $\Omega = (H, \pi)$ of Holant $(\mathcal{F} \mid \mathcal{G})$, for the same bipartite graph H, we get a new grid $\Omega' = (H, \pi')$ of Holant $\left(\mathcal{F}T \mid T^{-1}\mathcal{G}\right)$ by replacing each signature in $\mathcal{F}$ or $\mathcal{G}$ with the corresponding signature in $\mathcal{F}T$ or $T^{-1}\mathcal{G}$.

Theorem 2.2 (Valiant's Holant Theorem [38])**.** *If $T \in \mathbb{C}^{2\times 2}$ is an invertible matrix, then we have* $\text{Holant}(\Omega; \mathcal{F} \mid \mathcal{G}) = \text{Holant}(\Omega'; \mathcal{F}T \mid T^{-1}\mathcal{G})$.

Therefore, an invertible holographic transformation does not change the complexity of the Holant problem in the bipartite setting. Furthermore, there is a special kind of holographic transformation, the orthogonal transformation, that preserves the binary equality and thus can be used freely in the standard setting.

Theorem 2.3 (Theorem 2.6 in [13])**.** *If $T \in \mathbf{O}_2(\mathbb{C})$ is an orthogonal matrix (i.e. $TT^T = I_2$), then* $\text{Holant}(\Omega; \mathcal{F}) = \text{Holant}(\Omega'; T\mathcal{F})$.

We frequently apply a holographic transformation defined by the matrix $Z = \frac{1}{\sqrt{2}} \left[\begin{smallmatrix} 1 & 1 \\ i & -i \end{smallmatrix}\right]$ (or sometimes without the nonzero factor of $\frac{1}{\sqrt{2}}$ since this does not affect the complexity). This matrix has the property that the binary EQUALITY signature $(=_2) = [1, 0, 1]$ is transformed to $[1, 0, 1]Z^{\otimes 2} = [0, 1, 0] = (\neq_2)$, the binary DISEQUALITY signature.

An important definition involving a holographic transformation is the notion of a signature set being transformable.

Definition 2.4. *We say a signature set $\mathcal{F}$ is $\mathscr{C}$-transformable if there exists a $T \in \mathbf{GL}_2(\mathbb{C})$ such that $[1, 0, 1]T^{\otimes 2} \in \mathscr{C}$ and $\mathcal{F} \subseteq T\mathscr{C}$.*

This definition is important because if Pl-Holant($\mathscr{C}$) is tractable, then Pl-Holant($\mathcal{F}$) is tractable for any $\mathscr{C}$-transformable set $\mathcal{F}$.

2.3 Counting Constraint Satisfaction Problems

We can define the framework of counting constraint satisfaction problems (#CSP) in terms of the Holant framework. An instance of #CSP($\mathcal{F}$) has the following bipartite view. Create a vertex for each variable and each constraint. Connect a variable vertex to a constraint vertex if the variable appears in the constraint. This bipartite graph is also known as the *constraint graph*. Moreover, each variable can be viewed as an EQUALITY function, as it takes two values. Under this view, we see that #CSP($\mathcal{F}$) $\equiv_T$ Holant $(\mathcal{EQ} \mid \mathcal{F})$, where $\mathcal{EQ} = \{=_1, =_2, =_3, \dots\}$ is the set of EQUALITY signatures of all arities. By restricting to planar constraint graphs, we have the planar #CSP framework, which we denote by Pl-#CSP. The construction above also shows that Pl-#CSP($\mathcal{F}$) $\equiv_T$ Pl-Holant $(\mathcal{EQ} \mid \mathcal{F})$.

For any positive integer d, the problem #CSP$^d(\mathcal{F})$ is the same as #CSP($\mathcal{F}$) except that every variable appears a multiple of d times. Thus, Pl-#CSP$^d(\mathcal{F}) \equiv_T$ Pl-Holant $(\mathcal{EQ}_d \mid \mathcal{F})$, where $\mathcal{EQ}_d = \{=_d, =_{2d}, =_{3d}, \dots\}$ is the set of EQUALITY signatures of arities that are a multiple of d. If $d \in \{1, 2\}$, then we further have

$$\text{Pl-\#CSP}^d(\mathcal{F}) \equiv_T \text{Pl-Holant}\ (\mathcal{EQ}_d \mid \mathcal{F}) \equiv_T \text{Pl-Holant}(\mathcal{EQ}_d \cup \mathcal{F}). \tag{2.2}$$

The reduction from left to right in the second equivalence is trivial. For the other direction, we take a signature grid for the problem on the right and create a bipartite signature grid for the problem on the left such that both signature grids have the same Holant value up to an easily computable factor. If two signatures in $\mathcal{F}$ are assigned to adjacent vertices, then we subdivide all edges between them and assign the binary EQUALITY signature $=_2 \in \mathcal{EQ}_d$ to all new vertices. Suppose EQUALITY signatures $=_n, =_m \in \mathcal{EQ}_d$ are assigned to adjacent vertices connected by k edges. If $n = m = k$, then we simply remove these two vertices. The Holant of the resulting signature grid differs from the original by a factor of 2. Otherwise, we contract all k edges and assign $=_{n+m-2k} \in \mathcal{EQ}_d$ to the new vertex.

2.4 Realization

One basic notion used throughout the paper is realization. We say a signature f is *realizable* or *constructible* from a signature set $\mathcal{F}$ if there is a gadget with some dangling edges such that each vertex is assigned a signature from $\mathcal{F}$, and the resulting graph, when viewed as a black-box signature with inputs on the dangling edges, is exactly f. If f is realizable from a set $\mathcal{F}$, then we can freely add f into $\mathcal{F}$ while preserving the complexity.

Formally, such a notion is defined by an $\mathcal{F}$-gate [12]. An $\mathcal{F}$-gate is similar to a signature grid (G, π) for Holant($\mathcal{F}$) except that $G = (V, E, D)$ is a graph with some dangling edges D. The dangling edges define external variables for the $\mathcal{F}$-gate. (See Figure 1 for an example.) We denote the regular edges in E by $1, 2, \dots, m$ and the dangling edges in D by $m+1, \dots, m+n$. Then we can define a function Γ for this $\mathcal{F}$-gate as

$$\Gamma(y_1, \dots, y_n) = \sum_{x_1, \dots, x_m \in \{0,1\}} H(x_1, \dots, x_m, y_1, \dots, y_n),$$

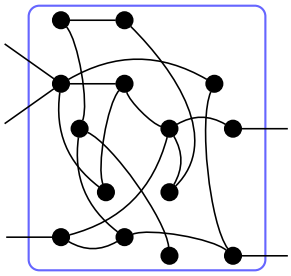


Figure 1: An $\mathcal{F}$-gate with 5 dangling edges.

where $(y_1, \ldots, y_n) \in \{0,1\}^n$ is an assignment on the dangling edges and $H(x_1, \ldots, x_m, y_1, \ldots, y_n)$ is the value of the signature grid on an assignment of all edges in G, which is the product of evaluations at all internal vertices. We also call this function Γ the signature of the $\mathcal{F}$-gate.

An $\mathcal{F}$-gate is planar if the underlying graph G is a planar graph, and the dangling edges, ordered counterclockwise corresponding to the order of the input variables, are in the outer face in a planar embedding. A planar $\mathcal{F}$-gate can be used in a planar signature grid as if it is just a single vertex with the particular signature.

Using the idea of planar $\mathcal{F}$-gates, we can reduce one planar Holant problem to another. Suppose g is the signature of some planar $\mathcal{F}$-gate. Then Pl-Holant$(\mathcal{F} \cup \{g\}) \leq_T$ Pl-Holant$(\mathcal{F})$. The reduction is simple. Given an instance of Pl-Holant$(\mathcal{F} \cup \{g\})$, by replacing every appearance of g by the $\mathcal{F}$-gate, we get an instance of Pl-Holant$(\mathcal{F})$. Since the signature of the $\mathcal{F}$-gate is g, the Holant values for these two signature grids are identical.

Although our main result is about symmetric signatures, some of our proofs utilize asymmetric signatures. When a gadget has an asymmetric signature, we place a diamond on the edge corresponding to the first input. The remaining inputs are ordered counterclockwise around the vertex. (See Figure 8 for two examples.)

We note that even for a very simple signature set $\mathcal{F}$, the signatures for all $\mathcal{F}$-gates can be quite complicated and expressive.

2.5 Tractable Signature Sets

We define the sets of signatures that were previously known to be tractable. All quotations of results and definitions from [6, 20, 7], both in this section and throughout the paper, refer to the full versions of these papers.

Affine Signatures

Definition 2.5 (Definition 3.1 in [15]). *A k-ary function $f(x_1, \ldots, x_k)$ is* affine *if it has the form*

$$\lambda \cdot \chi_{Ax=0} \cdot i^{\sum_{j=1}^{n} \langle \mathbf{v}_j, x \rangle},$$

where $\lambda \in \mathbb{C}$, $x = (x_1, x_2, \ldots, x_k, 1)^T$, A is a matrix over $\mathbb{F}_2$, $\mathbf{v}_j$ is a vector over $\mathbb{F}_2$, and χ is a 0-1 indicator function such that $\chi_{Ax=0}$ is 1 iff $Ax = 0$. Note that the dot product $\langle \mathbf{v}_j, x \rangle$ is calculated over $\mathbb{F}_2$, while the summation $\sum_{j=1}^{n}$ on the exponent of $i = \sqrt{-1}$ is evaluated as a sum mod 4 of 0-1 terms. We use $\mathscr{A}$ to denote the set of all affine functions.

Notice that there is no restriction on the number of rows in the matrix A. It is permissible that A is the zero matrix so that $\chi_{Ax=0} = 1$ holds for all x. An equivalent way to express the exponent of i is as a quadratic polynomial where all cross terms have an even coefficient (cf. [2]).

It is known that the set of non-degenerate symmetric signatures in $\mathscr{A}$ is precisely the nonzero signatures ($\lambda \neq 0$) in $\mathscr{F}_1 \cup \mathscr{F}_2 \cup \mathscr{F}_3$ with arity at least 2, where $\mathscr{F}_1$, $\mathscr{F}_2$, and $\mathscr{F}_3$ are three families of signatures defined as

$$\mathscr{F}_1 = \left\{ \lambda \left([1,0]^{\otimes k} + i^r [0,1]^{\otimes k} \right) \mid \lambda \in \mathbb{C}, k = 1,2,\ldots, r = 0,1,2,3 \right\},$$
$$\mathscr{F}_2 = \left\{ \lambda \left([1,1]^{\otimes k} + i^r [1,-1]^{\otimes k} \right) \mid \lambda \in \mathbb{C}, k = 1,2,\ldots, r = 0,1,2,3 \right\}, \text{ and}$$
$$\mathscr{F}_3 = \left\{ \lambda \left([1,i]^{\otimes k} + i^r [1,-i]^{\otimes k} \right) \mid \lambda \in \mathbb{C}, k = 1,2,\ldots, r = 0,1,2,3 \right\}.$$

We explicitly list these signatures up to an arbitrary constant multiple from $\mathbb{C}$:

1. $[1,0,\ldots,0,\pm 1]$; $(\mathscr{F}_1, r = 0, 2)$
2. $[1,0,\ldots,0,\pm i]$; $(\mathscr{F}_1, r = 1, 3)$
3. $[1,0,1,0,\ldots,0 \text{ or } 1]$; $(\mathscr{F}_2, r = 0)$
4. $[1,-i,1,-i,\ldots,(-i) \text{ or } 1]$; $(\mathscr{F}_2, r = 1)$
5. $[0,1,0,1,\ldots,0 \text{ or } 1]$; $(\mathscr{F}_2, r = 2)$
6. $[1,i,1,i,\ldots,i \text{ or } 1]$; $(\mathscr{F}_2, r = 3)$
7. $[1,0,-1,0,1,0,-1,0,\ldots,0 \text{ or } 1 \text{ or } (-1)]$; $(\mathscr{F}_3, r = 0)$
8. $[1,1,-1,-1,1,1,-1,-1,\ldots,1 \text{ or } (-1)]$; $(\mathscr{F}_3, r = 1)$
9. $[0,1,0,-1,0,1,0,-1,\ldots,0 \text{ or } 1 \text{ or } (-1)]$; $(\mathscr{F}_3, r = 2)$
10. $[1,-1,-1,1,1,-1,-1,1,\ldots,1 \text{ or } (-1)]$. $(\mathscr{F}_3, r = 3)$

Product-Type Signatures

Definition 2.6 (Definition 3.3 in [15])**.** *A function is of* product type *if it can be expressed as a product of unary functions, binary equality functions* ($[1,0,1]$)*, and binary disequality functions* ($[0,1,0]$)*. We use* $\mathscr{P}$ *to denote the set of product-type functions.*

An alternate definition for $\mathscr{P}$, implicit in [14], is the tensor closure of signatures with support on two complementary bit vectors. It can be shown (cf. Lemma A.1 in the full version of [21]) that if f is a symmetric signature in $\mathscr{P}$, then f is either degenerate, binary Disequality $\neq_2$, or $[a,0,\ldots,0,b]$ for some $a, b \in \mathbb{C}$.

Matchgate Signatures Matchgates were introduced by Valiant [36, 35] to give polynomial-time algorithms for a collection of counting problems over planar graphs. As the name suggests, problems expressible by matchgates can be reduced to computing a weighted sum of perfect matchings. The latter problem is tractable over planar graphs by Kasteleyn's algorithm [24], a.k.a. the FKT algorithm [34, 23]. These counting problems are naturally expressed in the Holant framework using *matchgate signatures*. We use $\mathscr{M}$ to denote the set of all matchgate signatures; thus Pl-Holant($\mathscr{M}$) is tractable. Holographic transformations extend the reach of the FKT algorithm even further, as stated below.

Theorem 2.7. *Let $\mathcal{F}$ be any set of symmetric, complex-valued signatures in Boolean variables. If $\mathcal{F}$ is $\mathscr{M}$-transformable, then* Pl-Holant($\mathcal{F}$) *is computable in polynomial time.*

Matchgate signatures are characterized by the matchgate identities (see [5] for the identities and a self-contained proof). The parity of a matchgate signature is even (resp. odd) if its support is on entries of even (resp. odd) Hamming weight. We explicitly list all the symmetric signatures in $\mathscr{M}$ (see [5]).

Proposition 2.8. *Let f be a symmetric signature in $\mathscr{M}$. Then there exists $a, b \in \mathbb{C}$ and $n \in \mathbb{N}$ such that f takes one of the following forms:*

1. $[a^n, 0, a^{n-1}b, 0, \ldots, 0, ab^{n-1}, 0, b^n]$ *(of arity $2n \geq 2$);*
2. $[a^n, 0, a^{n-1}b, 0, \ldots, 0, ab^{n-1}, 0, b^n, 0]$ *(of arity $2n+1 \geq 1$);*
3. $[0, a^n, 0, a^{n-1}b, 0, \ldots, 0, ab^{n-1}, 0, b^n]$ *(of arity $2n+1 \geq 1$);*
4. $[0, a^n, 0, a^{n-1}b, 0, \ldots, 0, ab^{n-1}, 0, b^n, 0]$ *(of arity $2n+2 \geq 2$).*

In the last three cases with $n = 0$, the signatures are $[1, 0]$, $[0, 1]$, and $[0, 1, 0]$. Any multiple of these is also a matchgate signature.

Roughly speaking, the symmetric matchgate signatures have 0 for every other entry (which is called the *parity condition*), and form a geometric progression with the remaining entries.

Another useful way to view the symmetric signature in $\mathscr{M}$ is via a low tensor rank decomposition. To state these low rank decompositions, we use the following definition.

Definition 2.9. *Let S_n be the symmetric group of degree n. Then for positive integers t and n with $t \leq n$ and unary signatures $v, v_1, \ldots, v_{n-t}$, we define*

$$\mathrm{Sym}_n^t(v; v_1, \ldots, v_{n-t}) = \sum_{\pi \in S_n} \bigotimes_{k=1}^{n} u_{\pi(k)},$$

where the ordered sequence $(u_1, u_2, \ldots, u_n) = (\underbrace{v, \ldots, v}_{t \text{ copies}}, v_1, \ldots, v_{n-t})$.

Proposition 2.10. *Let f be a symmetric signature in $\mathscr{M}$ of arity n. Then there exist $a, b, \lambda \in \mathbb{C}$ such that f takes one of the following forms:*

1. $[a, b]^{\otimes n} + [a, -b]^{\otimes n} = \begin{cases} 2[a^n, 0, a^{n-2}b^2, 0, \ldots, 0, b^n] & n \text{ is even,} \\ 2[a^n, 0, a^{n-2}b^2, 0, \ldots, 0, ab^{n-1}, 0] & n \text{ is odd;} \end{cases}$
2. $[a, b]^{\otimes n} - [a, -b]^{\otimes n} = \begin{cases} 2[0, a^{n-1}b, 0, a^{n-3}b^3, 0, \ldots, 0, ab^{n-1}, 0] & n \text{ is even,} \\ 2[0, a^{n-1}b, 0, a^{n-3}b^3, 0, \ldots, 0, b^n] & n \text{ is odd;} \end{cases}$
3. $\lambda\, \mathrm{Sym}_n^{n-1}([1, 0]; [0, 1]) = [0, \lambda, 0, \ldots, 0]$;
4. $\lambda\, \mathrm{Sym}_n^{n-1}([0, 1]; [1, 0]) = [0, \ldots, 0, \lambda, 0]$.

The understanding of matchgates was further developed in [11], which characterized, for every symmetric signature, the set of holographic transformations under which the transformed signature becomes a matchgate signature.

Vanishing Signatures Vanishing signatures were first introduced in [19] in the parity setting to denote signatures for which the Holant value is always 0 modulo 2.

Definition 2.11. *A set of signatures $\mathcal{F}$ is called* vanishing *if the value* $\mathrm{Holant}_\Omega(\mathcal{F})$ *is 0 for every signature grid Ω. A signature f is called* vanishing *if the singleton set $\{f\}$ is vanishing.*

A Holant problem defined only by vanishing signatures is trivially tractable by definition. Question is how to determine which sets of signatures are vanishing? We introduce the following definitions to answer this question.

Definition 2.12 (Definition 4.4 in [6]). *A nonzero symmetric signature f of arity n has* positive vanishing degree $k \geq 1$, *which is denoted by* $\mathrm{vd}^+(f) = k$, *if $k \leq n$ is the largest positive integer such that there exists $n-k$ unary signatures $v_1, \ldots, v_{n-k}$ satisfying*

$$f = \mathrm{Sym}_n^k([1, i]; v_1, \ldots, v_{n-k}).$$

If f cannot be expressed as such a symmetrization form, we define $\mathrm{vd}^+(f) = 0$. *If f is the all zero signature, define* $\mathrm{vd}^+(f) = n + 1$.

We define negative vanishing degree vd^- *similarly, using $-i$ instead of i.*

Definition 2.13 (Definition 4.5 in [6]). *For $\sigma \in \{+, -\}$, we define $\mathscr{V}^\sigma = \{f \mid 2\,\mathrm{vd}^\sigma(f) > \mathrm{arity}(f)\}$.*

Furthermore, we let $\mathscr{V} = \mathscr{V}^+ \cup \mathscr{V}^-$. The fact that $\mathscr{V}$ is closed under orthogonal transformations follows directly from the next lemma.

Lemma 2.14. *For a symmetric signature f of arity n, $\sigma \in \{+, -\}$, and an orthogonal matrix $T \in \mathbb{C}^{2\times 2}$, either* $\mathrm{vd}^\sigma(f) = \mathrm{vd}^\sigma(T^{\otimes n} f)$ *or* $\mathrm{vd}^\sigma(f) = \mathrm{vd}^{-\sigma}(T^{\otimes n} f)$.

The following characterization of vanishing signature sets holds.

Theorem 2.15 (Theorem 4.13 in [6]). *Let $\mathcal{F}$ be a set of symmetric signatures. Then $\mathcal{F}$ is vanishing if and only if $\mathcal{F} \subseteq \mathscr{V}^+$ or $\mathcal{F} \subseteq \mathscr{V}^-$.*

To prove this theorem, two more definitions were made, which complement the previous two definitions because of Corollary 2.18.

Definition 2.16 (Definition 4.7 in [6]). *A symmetric signature $f = [f_0, f_1, \ldots, f_n]$ of arity n is in $\mathscr{R}_t^+$ for a nonnegative integer $t \geq 0$ if $t > n$ or for any $0 \leq k \leq n - t$, $f_k, \ldots, f_{k+t}$ satisfy the recurrence relation*

$$\binom{t}{t} i^t f_{k+t} + \binom{t}{t-1} i^{t-1} f_{k+t-1} + \cdots + \binom{t}{0} i^0 f_k = 0. \tag{2.3}$$

We define $\mathscr{R}_t^-$ similarly but with $-i$ in place of i in (2.3).

Definition 2.17 (Definition 4.8 in [6]). *For a nonzero symmetric signature f of arity n, it is of* positive *(resp.* negative*)* recurrence degree $t \leq n$, *denoted by* $\mathrm{rd}^+(f) = t$ *(resp.* $\mathrm{rd}^-(f) = t$*), if and only if $f \in \mathscr{R}_{t+1}^+ - \mathscr{R}_t^+$ (resp. $f \in \mathscr{R}_{t+1}^- - \mathscr{R}_t^-$). If f is the all zero signature, we define* $\mathrm{rd}^+(f) = \mathrm{rd}^-(f) = -1$.

Corollary 2.18 (Corollary 4.16 in [6]). *If f is a symmetric signature and $\sigma \in \{+, -\}$, then* $\mathrm{vd}^\sigma(f) + \mathrm{rd}^\sigma(f) = \mathrm{arity}(f)$.

An observation was made in Section 4.3 of [6] that we utilize. We state it here as a lemma.

Lemma 2.19. *Suppose f is a symmetric signature of arity n. Let $\hat{f} = (Z^{-1})^{\otimes n} f$. If* $\mathrm{rd}^+(f) = d$, *then $\hat{f} = [\hat{f}_0, \hat{f}_1, \ldots, \hat{f}_d, 0, \ldots, 0]$ and $\hat{f}_d \neq 0$. Also $f \in \mathscr{R}_d^+$ iff all nonzero entries of $\hat{f}$ are among the first d entries in its symmetric signature notation.*

Similarly, if $\mathrm{rd}^-(f) = d$, *then $\hat{f} = [0, \ldots, 0, \hat{f}_{n-d}, \ldots, \hat{f}_n]$ and $\hat{f}_{n-d} \neq 0$. Also $f \in \mathscr{R}_d^-$ iff all nonzero entries of $\hat{f}$ are among the last d entries in its symmetric signature notation.*

The following lemma is a reduction involving binary signatures in the Z basis. It is used in Section 4 to help determine what binary signatures can mix with vanishing signatures. The original statement is for general graphs, but the proof clearly holds for planar graphs as well.

Lemma 2.20 (Lemma A.1 in [6]). *Let $x \in \mathbb{C}$. If $x \neq 0$, then for any set $\mathcal{F}$ containing $[x, 1, 0]$, we have*

$$\text{Pl-Holant}\left(\neq_2 \mid \mathcal{F} \cup \{[v, 1, 0]\}\right) \leq_T \text{Pl-Holant}\left(\neq_2 \mid \mathcal{F}\right)$$

for any $v \in \mathbb{C}$.

2.6 Some Known Dichotomies

Here we list several known dichotomies. The first is the dichotomy for Holant.

Theorem 2.21 (Theorem 5.1 in [6]). *Let $\mathcal{F}$ be any set of symmetric, complex-valued signatures in Boolean variables. Then* Holant$(\mathcal{F})$ *is* #P*-hard unless $\mathcal{F}$ satisfies one of the following conditions, in which case the problem is in* P*:*

1. *All non-degenerate signatures in $\mathcal{F}$ are of arity at most 2;*
2. *$\mathcal{F}$ is $\mathscr{A}$-transformable;*
3. *$\mathcal{F}$ is $\mathscr{P}$-transformable;*
4. *$\mathcal{F} \subseteq \mathscr{V}^{\sigma} \cup \{f \in \mathscr{R}_2^{\sigma} \mid \text{arity}(f) = 2\}$ for $\sigma \in \{+, -\}$;*
5. *All non-degenerate signatures in $\mathcal{F}$ are in $\mathscr{R}_2^{\sigma}$ for $\sigma \in \{+, -\}$.*

We also use several dichotomy theorems for planar Holant problems with additional restrictions. The first of these is a dichotomy theorem for a single signature of small arity. It is a combination of Theorem V.1 in [12] and Theorem 14 in [20] for arity 3 and 4, respectively. This theorem forms the base case of an inductive proof of Theorem 6.1, our single signature dichotomy.

Theorem 2.22. *If f is a non-degenerate, symmetric, complex-valued signature of arity* 3 *or* 4 *in Boolean variables, then* Pl-Holant(f) *is* #P*-hard unless f satisfies one of the following conditions, in which case, the problem is computable in polynomial time:*

1. Holant(f) *is tractable (i.e. f is $\mathscr{A}$-transformable, $\mathscr{P}$-transformable, or vanishing);*
2. *f is $\mathscr{M}$-transformable.*

We also state a corollary of this result, which shows that counting weighted matchings in 4-regular planar graphs is #P-hard. This is easier to apply than Theorem 2.22.

Corollary 2.23 (Lemma 5.5 in [20]). *Let $v \in \mathbb{C}$. If $v \neq 0$, then* Pl-Holant$([v, 1, 0, 0, 0])$ *is* #P*-hard.*

Next is a dichotomy theorem about counting complex weighted graph homomorphisms over degree prescribed graphs.

Theorem 2.24 (Theorem 3 in [8]). *Let $S \subseteq \mathbb{Z}^+$ containing some $r \geq 3$, let $\mathcal{G} = \{=_k \mid k \in S\}$, and let $d = \gcd(S)$. Further suppose that $f_0, f_1, f_2 \in \mathbb{C}$. Then* Pl-Holant $\left([f_0, f_1, f_2] \mid \mathcal{G}\right)$ *is* #P*-hard unless one of the following conditions holds:*

1. *$f_0 f_2 = f_1^2$;*
2. *$f_0 = f_2 = 0$;*
3. *$f_1 = 0$;*
4. *$f_0 f_2 = -f_1^2$ and $f_0^d = -f_2^d \neq 0$;*

5. $f_0^d = f_2^d \neq 0$.

In all exceptional cases, the problem is computable in polynomial time.

Theorem 2.24 is the original statement as in [8]. It is explicit and easy to apply. Conceptually, it can be restated as Theorem 2.24′, which supports the putative form of the Pl-#CSP^d dichotomy.

Theorem 2.24′ (Theorem 3 in [8]). *Let $S \subseteq \mathbb{Z}^+$ contain $k \geq 3$, let $\mathcal{G} = \{=_k \mid k \in S\}$, and let $d = \gcd(S)$. Further suppose that f is a non-degenerate, symmetric, complex-valued binary signature in Boolean variables. Then* Pl-Holant $(f \mid \mathcal{G})$ *is* #P*-hard unless f satisfies one of the following conditions, in which case, the problem is computable in polynomial time:*

1. *there exists $T \in \mathcal{T}_{4d}$ such that $T^{\otimes 2} f \in \mathscr{A}$;*
2. *$f \in \mathscr{P}$;*
3. *there exists $T \in \mathcal{T}_{2d}$ such that $T^{\otimes 2} f \in \widehat{\mathscr{M}}$.*

Lastly, we quote the Pl-#CSP dichotomy. It also supports the putative form of a dichotomy, which states that holographic algorithms using matchgates followed by the FKT algorithm is a universal strategy.

Theorem 2.25 (Theorem 19 in [20]). *Let $\mathcal{F}$ be any set of symmetric, complex-valued signatures in Boolean variables. Then* Pl-#CSP$(\mathcal{F})$ *is* #P*-hard unless $\mathcal{F} \subseteq \mathscr{A}$, $\mathcal{F} \subseteq \mathscr{P}$, or $\mathcal{F} \subseteq \widehat{\mathscr{M}}$, in which case the problem is computable in polynomial time.*

2.7 Redundant Signature Matrices and Related Hardness Results

Definition 2.26 (Definition 6.1 in [6]). *A 4-by-4 matrix is* redundant *if its middle two rows and middle two columns are the same.*

An example of a redundant matrix is the signature matrix of a symmetric arity 4 signature.

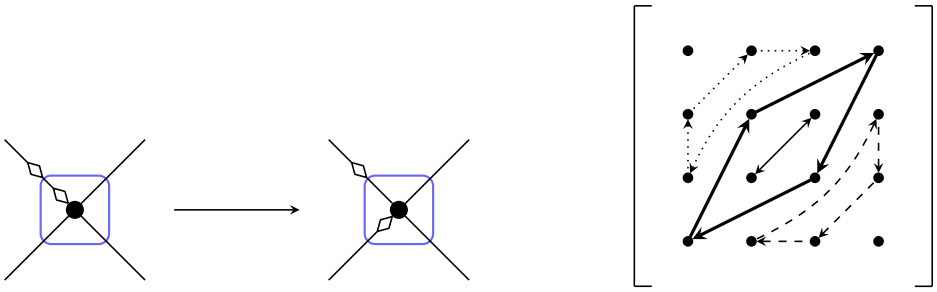
Definition 2.27 (Definition 6.2 in [6]). *The* signature matrix *of a symmetric arity 4 signature $f = [f_0, f_1, f_2, f_3, f_4]$ is*

$$M_f = \begin{bmatrix} f_0 & f_1 & f_1 & f_2 \\ f_1 & f_2 & f_2 & f_3 \\ f_1 & f_2 & f_2 & f_3 \\ f_2 & f_3 & f_3 & f_4 \end{bmatrix}.$$

This definition extends to an asymmetric signature g as

$$M_g = \begin{bmatrix} g^{0000} & g^{0010} & g^{0001} & g^{0011} \\ g^{0100} & g^{0110} & g^{0101} & g^{0111} \\ g^{1000} & g^{1010} & g^{1001} & g^{1011} \\ g^{1100} & g^{1110} & g^{1101} & g^{1111} \end{bmatrix},$$

where g^{wxyz} is the output of g on input $wxyz$. When we present g as an $\mathcal{F}$-gate, we order the four external edges $ABCD$ counterclockwise. In M_g, the row index bits are ordered AB and the column index bits are ordered DC, in reverse order. This is for convenience so that the signature matrix of the linking of two arity 4 $\mathcal{F}$-gates is the matrix product of the signature matrices of the two $\mathcal{F}$-gates.



(a) A counterclockwise rotation (b) Movement of signature matrix entries

Figure 2: The movement of the entries in the signature matrix of a quaternary signature under a counterclockwise rotation of the input edges. Entires of Hamming weight 1 are in the dotted cycle, entires of Hamming weight 2 are in the two solid cycles (one has length 4 and the other one is a swap), and entries of Hamming weight 3 are in the dashed cycle.

If M_g is redundant, we also define the compressed signature matrix *of g as*

$$\widetilde{M_g} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} M_g \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Lemma 2.28 (Corollary 3.8 in [20])**.** *Let f be an arity 4 signature with complex weights. If M_f is redundant and $\widetilde{M_f}$ is nonsingular, then* Pl-Holant(f) *is* #P*-hard.*

Furthermore, by combining Lemma 2.28 with Lemma 6.8 in [6], we obtain the planar version of Corollary 6.9 in [6].

Corollary 2.29. *Let f be an arity 4 signature with complex weights. If there exists a nonsingular matrix $T \in \mathbb{C}^{2\times 2}$ such that $\hat{f} = T^{\otimes 4} f$, where $M_{\hat{f}}$ is redundant and $\widetilde{M_{\hat{f}}}$ is nonsingular, then* Pl-Holant(f) *is* #P*-hard.*

In the course of working with symmetric signature, we sometimes construct gadgets with signatures that are not symmetric. The power of Lemma 2.28 and Corollary 2.29 is that they apply to such signatures provided the corresponding signature matrix is redundant. Sometimes one can apply a rotation to obtain a signature with a redundant signature matrix (see Figure 2).

3 $\mathscr{A}$-, $\mathscr{P}$-, and $\mathscr{M}$-transformable Signatures

In this section, we investigate the properties of $\mathscr{A}$-, $\mathscr{P}$-, and $\mathscr{M}$-transformable signatures. Throughout, we define $\alpha = \frac{1+i}{\sqrt{2}} = \sqrt{i} = e^{\frac{\pi i}{4}}$ and use $\mathbf{O}_2(\mathbb{C})$ to denote the group of 2-by-2 orthogonal matrices over $\mathbb{C}$. While the main results in this section assume that the signatures involved are symmetric, we note that some of the lemmas also hold without this assumption.

3.1 Characterization of $\mathscr{A}$- and $\mathscr{P}$-transformable Signatures

$\mathscr{A}$- and $\mathscr{M}$-transformable signatures have been well studied in previous work [6, 7]. We summarize some useful notions and lemmas here. The three sets $\mathscr{A}_1$, $\mathscr{A}_2$, and $\mathscr{A}_3$ capture all symmetric $\mathscr{A}$-transformable signatures.

Definition 3.1. *A symmetric signature f of arity n is in, respectively, $\mathscr{A}_1$, or $\mathscr{A}_2$, or $\mathscr{A}_3$ if there exist an $H \in \mathbf{O}_2(\mathbb{C})$ and a nonzero constant $c \in \mathbb{C}$ such that f has the form, respectively, $cH^{\otimes n}\left(\left[\begin{smallmatrix}1\\1\end{smallmatrix}\right]^{\otimes n} + \beta \left[\begin{smallmatrix}1\\-1\end{smallmatrix}\right]^{\otimes n}\right)$, or $cH^{\otimes n}\left(\left[\begin{smallmatrix}1\\i\end{smallmatrix}\right]^{\otimes n} + \left[\begin{smallmatrix}1\\-i\end{smallmatrix}\right]^{\otimes n}\right)$, or $cH^{\otimes n}\left(\left[\begin{smallmatrix}1\\\alpha\end{smallmatrix}\right]^{\otimes n} + i^r \left[\begin{smallmatrix}1\\-\alpha\end{smallmatrix}\right]^{\otimes n}\right)$, where $\beta = \alpha^{tn+2r}$, $r \in \{0,1,2,3\}$, and $t \in \{0,1\}$.*

For $k \in \{1,2,3\}$, when such an orthogonal H exists, we say that $f \in \mathscr{A}_k$ with transformation H. If $f \in \mathscr{A}_k$ with I_2, then we say f is in the canonical form of $\mathscr{A}_k$.

The following lemma characterizes the signatures in $\mathscr{A}_2$.

Lemma 3.2 (Lemma 8.8 in [6])**.** *Let f be a symmetric signature of arity n. Then $f \in \mathscr{A}_2$ if and only if $f = c\left(\left[\begin{smallmatrix}1\\i\end{smallmatrix}\right]^{\otimes n} + \beta \left[\begin{smallmatrix}1\\-i\end{smallmatrix}\right]^{\otimes n}\right)$ for some nonzero constants $c, \beta \in \mathbb{C}$.*

Membership in these three sets characterize the $\mathscr{A}$-transformable signatures.

Lemma 3.3 (Lemma 8.10 in [6])**.** *Let f be a non-degenerate symmetric signature. Then f is $\mathscr{A}$-transformable if and only if $f \in \mathscr{A}_1 \cup \mathscr{A}_2 \cup \mathscr{A}_3$.*

There is a similar characterization for $\mathscr{P}$-transformable signatures.

Definition 3.4. *A symmetric signature f of arity n is in $\mathscr{P}_1$ if there exist an $H \in \mathbf{O}_2(\mathbb{C})$ and a nonzero $c \in \mathbb{C}$ such that $f = cH^{\otimes n}\left(\left[\begin{smallmatrix}1\\1\end{smallmatrix}\right]^{\otimes n} + \beta \left[\begin{smallmatrix}1\\-1\end{smallmatrix}\right]^{\otimes n}\right)$, where $\beta \neq 0$.*

We define $\mathscr{P}_2 = \mathscr{A}_2$. For $k \in \{1,2\}$, when such an H exists, we say that $f \in \mathscr{P}_k$ with transformation H. If $f \in \mathscr{P}_k$ with I_2, then we say f is in the canonical form of $\mathscr{P}_k$.

Lemma 3.5 (Lemma 8.13 in [6])**.** *Let f be a non-degenerate symmetric signature. Then f is $\mathscr{P}$-transformable if and only if $f \in \mathscr{P}_1 \cup \mathscr{P}_2$.*

3.2 Characterization of $\mathscr{M}$-transformable Signatures

Now we develop a similar theory for the $\mathscr{M}$-transformable signatures. Recall from Definition 2.4 that for a signature set $\mathcal{F}$ to be $\mathscr{M}$-transformable, it must be that there exists a $T \in \mathbf{GL}_2(\mathbb{C})$ such that $[1,0,1]T^{\otimes 2} \in \mathscr{M}$. Since $[1,0,1]$ is symmetric, $[1,0,1]T^{\otimes 2}$ is also symmetric. However, it is unnecessary to consider all binary signatures in $\mathscr{M}$. We can normalize via *right* multiplication by elements in

$$\mathrm{Stab}(\mathscr{M}) = \{T \in \mathbf{GL}_2(\mathbb{C}) \mid T\mathscr{M} \subseteq \mathscr{M}\},$$

the stabilizer group of $\mathscr{M}$. Technically this set is the left stabilizer group of $\mathscr{M}$, but it is easy to see that the left and right stabilizer groups of $\mathscr{M}$ coincide and that they are generated by nonzero scalar multiples of matrices of the form $\left[\begin{smallmatrix}1 & 0\\ 0 & \nu\end{smallmatrix}\right]$ for any nonzero $\nu \in \mathbb{C}$ and $X = \left[\begin{smallmatrix}0 & 1\\ 1 & 0\end{smallmatrix}\right]$.

After this normalization, it is enough to consider cases 1 and 3 in the following proposition.

Proposition 3.6 (Proposition 8.1 in [6])**.** *Let $T \in \mathbb{C}^{2\times 2}$ be a matrix. Then the following hold:*

1. $[1,0,1]T^{\otimes 2} = [1,0,1]$ *if and only if* $T \in \mathbf{O}_2(\mathbb{C})$;
2. $[1,0,1]T^{\otimes 2} = [1,0,i]$ *if and only if there exists an* $H \in \mathbf{O}_2(\mathbb{C})$ *such that* $T = H\left[\begin{smallmatrix}1 & 0\\ 0 & \alpha\end{smallmatrix}\right]$;
3. $[1,0,1]T^{\otimes 2} = [0,1,0]$ *if and only if there exists an* $H \in \mathbf{O}_2(\mathbb{C})$ *such that* $T = \frac{1}{\sqrt{2}} H \left[\begin{smallmatrix}1 & 1\\ i & -i\end{smallmatrix}\right]$.

Lemma 3.7. *Let* $\mathcal{F}$ *be a set of signatures. Then* $\mathcal{F}$ *is* $\mathscr{M}$*-transformable if and only if* $\mathcal{F} \subseteq \left[\begin{smallmatrix}1 & 1\\ i & -i\end{smallmatrix}\right]\mathscr{M}$ *or there exists an* $H \in \mathbf{SO}_2(\mathbb{C})$ *such that* $\mathcal{F} \subseteq H\mathscr{M}$.

Proof. Sufficiency is easily verified by checking that $=_2$ is transformed into $\mathscr{M}$ in both cases. In particular, H leaves $=_2$ unchanged.

If $\mathcal{F}$ is $\mathscr{M}$-transformable, then by definition, there exists a matrix T such that $(=_2)T^{\otimes 2} \in \mathscr{M}$ and $\mathcal{F} \subseteq T\mathscr{M}$. The non-degenerate binary signatures in $\mathscr{M}$ are either $[0,1,0]$ or of the form $[1,0,\nu]$, up to a scalar. However, notice that $[1,0,1] = [1,0,\nu]\begin{bmatrix}1 & 0\\ 0 & \nu^{-\frac{1}{2}}\end{bmatrix}^{\otimes 2}$ and $\begin{bmatrix}1 & 0\\ 0 & \nu^{-\frac{1}{2}}\end{bmatrix} \in \mathrm{Stab}(\mathscr{M})$. Thus, we only need to consider $[1,0,1]$ and $[0,1,0]$. Now we apply Proposition 3.6.

1. If $(=_2)T^{\otimes 2} = [1,0,1]$, then by case 1 of Proposition 3.6, we have $T \in \mathbf{O}_2(\mathbb{C})$. If $T \in \mathbf{SO}_2(\mathbb{C})$, then we are done with $H = T$. Otherwise, $T \in \mathbf{O}_2(\mathbb{C}) - \mathbf{SO}_2(\mathbb{C})$. We want to find an $H \in \mathbf{SO}_2(\mathbb{C})$ such that $\mathcal{F} \subseteq H\mathscr{M}$. Let $H = T\left[\begin{smallmatrix}1 & 0\\ 0 & -1\end{smallmatrix}\right] \in \mathbf{SO}_2(\mathbb{C})$. Then
$$\mathcal{F} \subseteq T\mathscr{M} = T\begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix}\mathscr{M} = H\mathscr{M}$$
since $\left[\begin{smallmatrix}1 & 0\\ 0 & -1\end{smallmatrix}\right] \in \mathrm{Stab}(\mathscr{M})$.
2. If $(=_2)T^{\otimes 2} = [0,1,0]$, then by case 3 of Proposition 3.6, there exists an $H \in \mathbf{O}_2(\mathbb{C})$ such that $T = \frac{1}{\sqrt{2}} H \left[\begin{smallmatrix}1 & 1\\ i & -i\end{smallmatrix}\right]$. Therefore $\mathcal{F} \subseteq H\left[\begin{smallmatrix}1 & 1\\ i & -i\end{smallmatrix}\right]\mathscr{M}$. Furthermore, if $H = \left[\begin{smallmatrix}a & b\\ -b & a\end{smallmatrix}\right] \in \mathbf{SO}_2(\mathbb{C})$, then $a^2 + b^2 = 1$ and
$$\mathcal{F} \subseteq H\begin{bmatrix}1 & 1\\ i & -i\end{bmatrix}\mathscr{M} = \begin{bmatrix}1 & 1\\ i & -i\end{bmatrix}\begin{bmatrix}a+bi & 0\\ 0 & a-bi\end{bmatrix}\mathscr{M} = \begin{bmatrix}1 & 1\\ i & -i\end{bmatrix}\mathscr{M}$$
since $H\left[\begin{smallmatrix}1 & 1\\ i & -i\end{smallmatrix}\right] = \left[\begin{smallmatrix}1 & 1\\ i & -i\end{smallmatrix}\right]\left[\begin{smallmatrix}a+bi & 0\\ 0 & a-bi\end{smallmatrix}\right]$ and $\left[\begin{smallmatrix}a+bi & 0\\ 0 & a-bi\end{smallmatrix}\right] \in \mathrm{Stab}(\mathscr{M})$. Otherwise, $H = \left[\begin{smallmatrix}a & b\\ b & -a\end{smallmatrix}\right] \in \mathbf{O}_2(\mathbb{C}) - \mathbf{SO}_2(\mathbb{C})$, so $a^2 + b^2 = 1$ and
$$\mathcal{F} \subseteq H\begin{bmatrix}1 & 1\\ i & -i\end{bmatrix}\mathscr{M} = \begin{bmatrix}1 & 1\\ i & -i\end{bmatrix}\begin{bmatrix}0 & a-bi\\ a+bi & 0\end{bmatrix}\mathscr{M} = \begin{bmatrix}1 & 1\\ i & -i\end{bmatrix}\mathscr{M}$$
since $H\left[\begin{smallmatrix}1 & 1\\ i & -i\end{smallmatrix}\right] = \left[\begin{smallmatrix}1 & 1\\ i & -i\end{smallmatrix}\right]\left[\begin{smallmatrix}0 & a-bi\\ a+bi & 0\end{smallmatrix}\right]$ and $\left[\begin{smallmatrix}0 & a-bi\\ a+bi & 0\end{smallmatrix}\right] \in \mathrm{Stab}(\mathscr{M})$. □

We use four sets to characterize the $\mathscr{M}$-transformable signatures. The notation Sym is from Definition 2.9.

Definition 3.8. *A symmetric signature* f *of arity* n *is in, respectively,* $\mathscr{M}_1$*, or* $\mathscr{M}_2$*, or* $\mathscr{M}_3$*, or* $\mathscr{M}_4$ *if there exist an* $H \in \mathbf{O}_2(\mathbb{C})$ *and nonzero constants* $c, \gamma \in \mathbb{C}$ *such that* f *has the form, respectively,* $cH^{\otimes n}\left(\left[\begin{smallmatrix}1\\ 1\end{smallmatrix}\right]^{\otimes n} \pm i^n \left[\begin{smallmatrix}1\\ -1\end{smallmatrix}\right]^{\otimes n}\right)$*, or* $cH^{\otimes n}\left(\left[\begin{smallmatrix}1\\ \gamma\end{smallmatrix}\right]^{\otimes n} \pm \left[\begin{smallmatrix}1\\ -\gamma\end{smallmatrix}\right]^{\otimes n}\right)$*, or* $cH^{\otimes n}\,\mathrm{Sym}_n^{n-1}([\begin{smallmatrix}1\\ 0\end{smallmatrix}];[\begin{smallmatrix}0\\ 1\end{smallmatrix}])$*, or* $cH^{\otimes n}\,\mathrm{Sym}_n^{n-1}([\begin{smallmatrix}1\\ i\end{smallmatrix}];\left[\begin{smallmatrix}1\\ -i\end{smallmatrix}\right])$.

For $k \in \{1,2,3,4\}$, when such an H exists, we say that $f \in \mathscr{M}_k$ with transformation H. If $f \in \mathscr{M}_k$ with I_2, then we say f is in the canonical form of $\mathscr{M}_k$.

Notice that $\{\left[\begin{smallmatrix}1\\ i\end{smallmatrix}\right], \left[\begin{smallmatrix}1\\ -i\end{smallmatrix}\right]\}$ is set-wise invariant under any transformation in $\mathbf{O}_2(\mathbb{C})$ up to nonzero constants. Using this fact, the following lemma gives a characterization of $\mathscr{M}_4$. It says that any signature in $\mathscr{M}_4$ is essentially in canonical form.

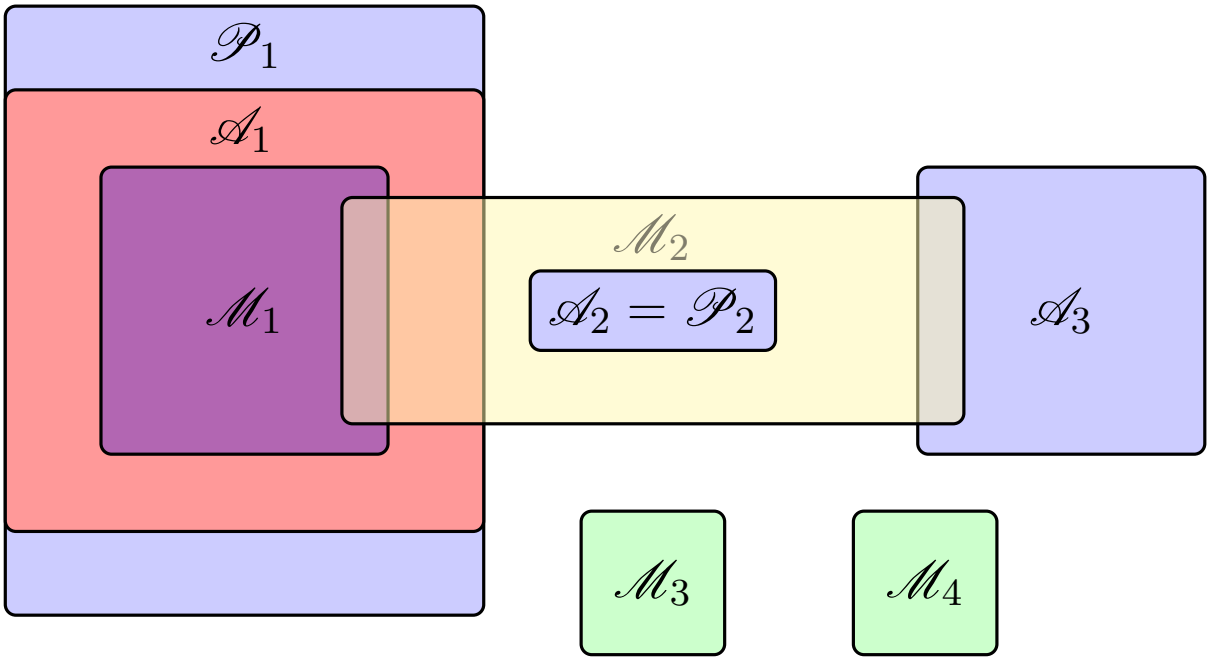


Figure 3: Relationships among $\mathscr{A}_1$, $\mathscr{A}_2$, $\mathscr{A}_3$, $\mathscr{P}_1$, $\mathscr{P}_2$, $\mathscr{M}_1$, $\mathscr{M}_2$, $\mathscr{M}_3$, and $\mathscr{M}_4$. Note that $\mathscr{P}_1 \cap \mathscr{M}_2 \subseteq \mathscr{A}_1$.

Lemma 3.9. *Let f be a symmetric signature of arity n. Then $f \in \mathscr{M}_4$ if and only if $f = c\,\mathrm{Sym}_n^{n-1}([\begin{smallmatrix}1\\ i\end{smallmatrix}]; [\begin{smallmatrix}1\\ -i\end{smallmatrix}])$ or $f = c\,\mathrm{Sym}_n^{n-1}([\begin{smallmatrix}1\\ -i\end{smallmatrix}]; [\begin{smallmatrix}1\\ i\end{smallmatrix}])$ for some nonzero constant $c \in \mathbb{C}$.*

Proof. Suppose $f \in \mathscr{M}_4$, so that $f = cH^{\otimes n}\,\mathrm{Sym}_n^{n-1}([\begin{smallmatrix}1\\ i\end{smallmatrix}]; [\begin{smallmatrix}1\\ -i\end{smallmatrix}])$. If $H \in \mathbf{SO}_2(\mathbb{C})$, then $H = [\begin{smallmatrix}a & b\\ -b & a\end{smallmatrix}]$ for some $a, b \in \mathbb{C}$ such that $a^2 + b^2 = 1$. Since $H[\begin{smallmatrix}1\\ i\end{smallmatrix}] = (a+bi)[\begin{smallmatrix}1\\ i\end{smallmatrix}]$ and $H[\begin{smallmatrix}1\\ -i\end{smallmatrix}] = (a-bi)[\begin{smallmatrix}1\\ -i\end{smallmatrix}]$, it follows that $f = c(a+bi)^{n-1}(a-bi)\,\mathrm{Sym}_n^{n-1}([\begin{smallmatrix}1\\ i\end{smallmatrix}]; [\begin{smallmatrix}1\\ -i\end{smallmatrix}])$. Otherwise, $H \in \mathbf{O}_2(\mathbb{C}) - \mathbf{SO}_2(\mathbb{C})$, so $H = [\begin{smallmatrix}a & b\\ b & -a\end{smallmatrix}]$ for some $a, b \in \mathbb{C}$ such that $a^2+b^2 = 1$. Then $f = c(a+bi)(a-bi)^{n-1}\,\mathrm{Sym}_n^{n-1}([\begin{smallmatrix}1\\ -i\end{smallmatrix}]; [\begin{smallmatrix}1\\ i\end{smallmatrix}])$.

Now suppose $f = c\,\mathrm{Sym}_n^{n-1}([\begin{smallmatrix}1\\ i\end{smallmatrix}]; [\begin{smallmatrix}1\\ -i\end{smallmatrix}])$ or $f = c\,\mathrm{Sym}_n^{n-1}([\begin{smallmatrix}1\\ -i\end{smallmatrix}]; [\begin{smallmatrix}1\\ i\end{smallmatrix}])$. The first case is already in the standard form of $\mathscr{M}_4$. In the second case, we pick $H = [\begin{smallmatrix}1 & 0\\ 0 & -1\end{smallmatrix}] \in \mathbf{O}_2(\mathbb{C})$. Then $H^{\otimes n} f$ is in the standard form of $\mathscr{M}_4$. □

We further split $\mathscr{M}_4$ into $\mathscr{M}_4^{\pm}$ for future use. Define $\mathscr{M}_4^{\pm} = \{f | f = c\,\mathrm{Sym}_n^{n-1}([\begin{smallmatrix}1\\ \pm i\end{smallmatrix}]; [\begin{smallmatrix}1\\ \mp i\end{smallmatrix}])\}$. In other words, $\mathscr{M}_4^{+}$ contains signatures of the form $Z^{\otimes n}[0,1,0,\ldots,0]$ and $\mathscr{M}_4^{-}$ contains signatures of the form $Z^{\otimes n}[0,\ldots,0,1,0]$ up to a scalar, where $Z = [\begin{smallmatrix}1 & 1\\ i & -i\end{smallmatrix}]$. We will denote $[0,1,0,\ldots,0]$ of arity k by $\textsc{ExactOne}_k$, and $[0,\ldots,0,1,0]$ of arity k by $\textsc{AllButOne}_k$. Note that these are precisely the Perfect Matching signatures and corresponding reversals.

Notice that $\mathscr{M}_1 \subset \mathscr{A}_1 \subset \mathscr{P}_1$ and $\mathscr{A}_2 = \mathscr{P}_2 \subset \mathscr{M}_2$. See Figure 3 for a visual description of the relationships among sets.

Next we show that $\mathscr{M}_k$ for $k = 1, 2, 3, 4$ captures all $\mathscr{M}$-transformable signatures.

Lemma 3.10. *Let f be a non-degenerate symmetric signature. Then f is $\mathscr{M}$-transformable if and only if $f \in \mathscr{M}_1 \cup \mathscr{M}_2 \cup \mathscr{M}_3 \cup \mathscr{M}_4$.*

Proof. Assume that f is $\mathscr{M}$-transformable of arity n. By applying Lemma 3.7 to $\{f\}$, we have $f \in [\begin{smallmatrix}1 & 1\\ i & -i\end{smallmatrix}]\mathscr{M}$ or there exists an $H \in \mathbf{SO}_2(\mathbb{C})$ such that $f \in H\mathscr{M}$. Proposition 2.10 lists the

symmetric signatures in $\mathscr{M}$. Since we are only interested in non-degenerate signatures, we only consider a, b, and λ that are nonzero. Now we consider the possible cases.

1. Suppose $f \in \left[\begin{smallmatrix} 1 & 1 \\ i & -i \end{smallmatrix}\right] \mathscr{M}$.
   - Further suppose $f = \left[\begin{smallmatrix} 1 & 1 \\ i & -i \end{smallmatrix}\right]^{\otimes n} \left(\left[\begin{smallmatrix} a \\ b \end{smallmatrix}\right]^{\otimes n} \pm \left[\begin{smallmatrix} a \\ -b \end{smallmatrix}\right]^{\otimes n} \right)$ for some nonzero $a, b \in \mathbb{C}$. Let $T = \frac{1-i}{2} \left[\begin{smallmatrix} u & v \\ v & -u \end{smallmatrix}\right]$, where $u = a + bi$ and $v = i(a - bi)$. Then $f = T^{\otimes n} \left(\left[\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right]^{\otimes n} \pm i^n \left[\begin{smallmatrix} 1 \\ -1 \end{smallmatrix}\right]^{\otimes n} \right)$. Since $T \in \mathbf{O}_2(\mathbb{C})$ up to a nonzero factor of $\sqrt{2ab}$, we have $f \in \mathscr{M}_1$.
   - Further suppose $f = \lambda \left[\begin{smallmatrix} 1 & 1 \\ i & -i \end{smallmatrix}\right]^{\otimes n} \operatorname{Sym}_n^{n-1}(\left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right]; \left[\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right])$ for some nonzero $\lambda \in \mathbb{C}$. Then we have $f = \lambda \operatorname{Sym}_n^{n-1}(\left[\begin{smallmatrix} 1 \\ i \end{smallmatrix}\right]; \left[\begin{smallmatrix} 1 \\ -i \end{smallmatrix}\right])$, so $f \in \mathscr{M}_4$.
   - Further suppose $f = \lambda \left[\begin{smallmatrix} 1 & 1 \\ i & -i \end{smallmatrix}\right]^{\otimes n} \operatorname{Sym}_n^{n-1}(\left[\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right]; \left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right])$ for some nonzero $\lambda \in \mathbb{C}$. Then we have $f = \lambda \operatorname{Sym}_n^{n-1}(\left[\begin{smallmatrix} 1 \\ -i \end{smallmatrix}\right]; \left[\begin{smallmatrix} 1 \\ i \end{smallmatrix}\right])$, so $f \in \mathscr{M}_4$ by Lemma 3.9.
2. Suppose $f \in H\mathscr{M}$.
   - Further suppose $f = H^{\otimes n} \left(\left[\begin{smallmatrix} a \\ b \end{smallmatrix}\right]^{\otimes n} \pm \left[\begin{smallmatrix} a \\ -b \end{smallmatrix}\right]^{\otimes n} \right)$ for some nonzero $a, b \in \mathbb{C}$. Then we have $f = a^n H^{\otimes n} \left(\left[\begin{smallmatrix} 1 \\ \gamma \end{smallmatrix}\right]^{\otimes n} \pm \left[\begin{smallmatrix} 1 \\ -\gamma \end{smallmatrix}\right]^{\otimes n} \right)$, where $\gamma = \frac{b}{a}$, so $f \in \mathscr{M}_2$.
   - Further suppose $f = \lambda H^{\otimes n} \operatorname{Sym}_n^{n-1}(\left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right]; \left[\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right])$ for some nonzero $\lambda \in \mathbb{C}$. Then $f \in \mathscr{M}_3$.
   - Further suppose $f = \lambda H^{\otimes n} \operatorname{Sym}_n^{n-1}(\left[\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right]; \left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right])$ for some nonzero $\lambda \in \mathbb{C}$. Let $H' = H \left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right] \in \mathbf{O}_2(\mathbb{C})$. Then we have $f = \lambda H'^{\otimes n} \operatorname{Sym}_n^{n-1}(\left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right]; \left[\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right])$, so $f \in \mathscr{M}_3$.

Conversely, if there exists a matrix $H \in \mathbf{O}_2(\mathbb{C})$ such that $H^{\otimes n} f$ is in one of the canonical forms of $\mathscr{M}_1$, $\mathscr{M}_2$, $\mathscr{M}_3$, or $\mathscr{M}_4$, then one can directly check that f is $\mathscr{M}$-transformable by Definition 2.4. In fact, the transformations that we applied above are all invertible, except for $\mathscr{M}_1$, if the given orthogonal transformation is of the form $\left[\begin{smallmatrix} u & -v \\ v & u \end{smallmatrix}\right]$, we do $\left[\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right]$ first followed by $\left[\begin{smallmatrix} u & v \\ v & -u \end{smallmatrix}\right]$. □

Furthermore, we show that a nontrivial signature f in the set $\mathscr{M}_3$ is not $\mathscr{A}$- or $\mathscr{P}$-transformable. Moreover, the only transformation to make f in $\mathscr{M}$ is very restricted. This is for future use.

Lemma 3.11. *Let $f \in \mathscr{M}_3$ be a non-degenerate signature of arity $n \geq 3$ with $H \in \mathbf{O}_2(\mathbb{C})$. Then f is not $\mathscr{A}$- or $\mathscr{P}$-transformable. Moreover, f is $\mathscr{M}$-transformable with only HD or $H \left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right] D$ for some diagonal matrix D.*

Proof. Suppose $f = [f_0, f_1, \ldots, f_n]$. If f is $\mathscr{A}$- or $\mathscr{P}$-transformable, then f has to satisfy a second order recurrence relation that $af_i + bf_{i+1} + cf_{i+2} = 0$, for $a, b, c \in \mathbb{C}$ such that not all a, b, c are 0 and $b^2 - 4ac \neq 0$. In other words, the second order recurrence relation has to have distinct eigenvalues. This is due to Lemma 6.15 or Lemma 7.2 in [7]. Moreover, this property is preserved by holographic transformations (cf. Lemma 6.2 in [7]). However, f is in $\mathscr{M}_3$. Hence $f = H^{\otimes n}\textsc{ExactOne}_n$ for some $H \in \mathbf{O}_2(\mathbb{C})$ up to a nonzer factor. On the other hand, $\textsc{ExactOne}_n$ does not satisfy a second recurrence with distinct eigenvalues if $n \geq 3$, a contradiction.

Moreover, notice that the only signatures in $\mathscr{M}$ that do not satisfy such second order recurrence relations are $\textsc{ExactOne}_k$ and $\textsc{AllButOne}_k$ functions. If f is $\mathscr{M}$-transformable, then there exists a transformation T such that $f = T^{\otimes n} g$ for some $g \in \mathscr{M}$ and $[1, 0, 1]T^{\otimes 2} \in \mathscr{M}$. Hence $g = \textsc{ExactOne}_n$ or $\textsc{AllButOne}_n$. On the other hand $f = H^{\otimes n}\textsc{ExactOne}_n$ up to a nonzer factor. Therefore $(T^{-1}H)^{\otimes n}\textsc{ExactOne}_n = \textsc{ExactOne}_n$ or $\textsc{AllButOne}_n$ up to a nonzer factor.

Let $J = T^{-1}H = \left[\begin{smallmatrix} x & y \\ z & w \end{smallmatrix}\right]$ and let $h = J^{\otimes n}\textsc{ExactOne}_n$. As $\textsc{ExactOne}_n = \operatorname{Sym}_n^{n-1}(\left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right]; \left[\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right])$, $h = (\left[\begin{smallmatrix} x & y \\ z & w \end{smallmatrix}\right])^{\otimes n}\textsc{ExactOne}_n = \operatorname{Sym}_n^{n-1}(\left[\begin{smallmatrix} x \\ z \end{smallmatrix}\right]; \left[\begin{smallmatrix} y \\ w \end{smallmatrix}\right])$. The first and last entries of h are $x^{n-1}y$ and $z^{n-1}w$. As $h = \textsc{ExactOne}_n$ or $\textsc{AllButOne}_n$, we have that $x^{n-1}y = z^{n-1}w = 0$. It is easy to see that x and z, or y and w cannot be both 0. Then $x = w = 0$ or $y = z = 0$. This implies that $J = D$ or $J = D \left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right]$ for some diagonal matrix D. Thus $T = HJ^{-1} = HD^{-1}$ or $H \left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right] D^{-1}$. □

Let $g = [x, y, 0, \ldots, 0, z]$ have arity $n \geq 3$, where $xyz \neq 0$. As an example of the theory developed in this section, we discuss the signature $Z^{\otimes n}g$ in the following lemma, which will be used in Lemma 6.6 in the proof of the single signature dichotomy Theorem 6.1.

Lemma 3.12. *Let $n \geq 3$, $g = [x, y, 0, \ldots, 0, z]$ have arity n and $xyz \neq 0$. Then the signature $Z^{\otimes n}g$ is neither $\mathscr{A}$-, $\mathscr{P}$-, $\mathscr{M}$-transformable, nor vanishing.*

Remark 1. By Theorem 2.22, for arity $n = 3$ or 4, Lemma 3.12 implies that Pl-Holant$(Z^{\otimes n}g)$ is #P-hard. After we have proved Theorem 6.1, this lemma will imply that Pl-Holant$(Z^{\otimes n}g)$ is #P-hard for all $n \geq 3$.

Proof. That $Z^{\otimes n}g$ is not vanishing follows from Lemma 2.19 combined with Corollary 2.18 and Theorem 2.15. To show that $Z^{\otimes n}g$ is not $\mathscr{A}$-, $\mathscr{P}$-, $\mathscr{M}$-transformable, we only need to show that $Z^{\otimes n}g \notin \mathscr{P}_1 \cup \mathscr{M}_2 \cup \mathscr{A}_3 \cup \mathscr{M}_3 \cup \mathscr{M}_4$ by Lemma 3.3, 3.5 and 3.10, and the fact that $\mathscr{M}_1 \subset \mathscr{A}_1 \subset \mathscr{P}_1$ and $\mathscr{A}_2 = \mathscr{P}_2 \subset \mathscr{M}_2$. See Figure 3.

We first show that $Z^{\otimes n}g \notin \mathscr{P}_1 \cup \mathscr{M}_2 \cup \mathscr{A}_3$. We say a signature $f = [f_0, f_1, \ldots, f_n]$ satisfies a second order recurrence of type $\langle a, b, c\rangle$ if $af_k - bf_{k+1} + cf_{k+2} = 0$ for $1 \leq k \leq n-2$, for some a, b and c not all zero. Suppose $Z^{\otimes n}g$ is a nonzer constant multiple of $Hf \in \mathscr{P}_1 \cup \mathscr{M}_2 \cup \mathscr{A}_3$ in the forms given in Definitions 3.4, 3.8 and 3.1, then f, and hence also $(Z^{-1})^{\otimes n}f$, satisfies a second order recurrence. We have $H^{-1}Z = ZD$ or $ZD\left[\begin{smallmatrix}0 & 1\\ 1 & 0\end{smallmatrix}\right]$ for some non-singular diagonal D since $H \in \mathbf{O}_2(\mathbb{C})$. Thus $f = Z^{\otimes n}g'$ for some $g' = [x', y', 0, \ldots, 0, z']$ or $[x', 0, \ldots, 0, y', z']$, with $x'y'z' \neq 0$. We assume the former; the proof is similar for the latter.

However, for $n \geq 4$, g' does not satisfy any second order recurrence. For a contradiction suppose g' does. By $x'y'z' \neq 0$, $ay' - b0 + c0 = 0$ gives $a = 0$, $ax' - by' + c0 = 0$ gives $b = 0$, and $a0 - b0 + cz' = 0$ gives $c = 0$; but a, b, c cannot be all zero.

Next suppose $n = 3$, and we show that $g' = (Z^{-1})^{\otimes n}f$ is still impossible. For $\mathscr{P}_1$, $f = \left[\begin{smallmatrix}1\\ 1\end{smallmatrix}\right]^{\otimes 3} + \beta\left[\begin{smallmatrix}1\\ -1\end{smallmatrix}\right]^{\otimes 3}$. It is easy to check that $(Z^{-1})^{\otimes n}f$ satisfies a second order recurrence with its two eigenvalues sum to zero. However $g' = [x', y', 0, z']$ has type $\langle y'z', x'z', -y'^2\rangle$, the sum of its two eigenvalues is $-x'z'/y'^2 \neq 0$.

For $\mathscr{M}_2$, $f = \left[\begin{smallmatrix}1\\ \gamma\end{smallmatrix}\right]^{\otimes 3} \pm \left[\begin{smallmatrix}1\\ -\gamma\end{smallmatrix}\right]^{\otimes 3}$. In $(Z^{-1})^{\otimes n}f$, $Z^{-1}\left[\begin{smallmatrix}1 & 1\\ \gamma & -\gamma\end{smallmatrix}\right]$ has the form $\left[\begin{smallmatrix}u & v\\ v & u\end{smallmatrix}\right]$, and $(Z^{-1})^{\otimes n}f = \left[\begin{smallmatrix}u\\ v\end{smallmatrix}\right]^{\otimes 3} \pm \left[\begin{smallmatrix}v\\ u\end{smallmatrix}\right]^{\otimes 3}$. Thus the weight 1 and weight 2 entries of $(Z^{-1})^{\otimes n}f$ are either equal or negative of each other. If $g' = (Z^{-1})^{\otimes n}f$ this would imply $y' = 0$, a contradiction.

For $\mathscr{A}_3$, $f = \left[\begin{smallmatrix}1\\ \alpha\end{smallmatrix}\right]^{\otimes n} + i^r\left[\begin{smallmatrix}1\\ -\alpha\end{smallmatrix}\right]^{\otimes n}$. $Z^{-1}\left[\begin{smallmatrix}1 & 1\\ \alpha & -\alpha\end{smallmatrix}\right] = \left[\begin{smallmatrix}u & v\\ v & u\end{smallmatrix}\right]$, with $u = 1 - \alpha i$ and $v = 1 + \alpha i$. The weight 2 entry of $(Z^{-1})^{\otimes n}f$ is $uv^2 + i^r vu^2 = (uv)(v + i^r u)$. This is nonzer for all r. However $g' = [x', y', 0, z']$ has this property.

It remains to show that $Z^{\otimes n}g \notin \mathscr{M}_3 \cup \mathscr{M}_4$. If $Z^{\otimes n}g \in \mathscr{M}_3$, then $Z^{\otimes n}g = cHf$ for some $H \in \mathbf{O}_2(\mathbb{C})$ and $f = \mathrm{Sym}_n^{n-1}(\left[\begin{smallmatrix}1\\ 0\end{smallmatrix}\right]; \left[\begin{smallmatrix}0\\ 1\end{smallmatrix}\right])$. Again $f = (cH)^{-1}Z^{\otimes n}g = Z^{\otimes n}g'$ for some g' having the same or its reversal form as g. Then $g' = (Z^{-1})^{\otimes n}f$ is the signature $[n, n-2, \ldots, -(n-2), -n]$. The weight 1 entry and weight $n-1$ entry have the same absolute value. By the form of g' this is a contradiction.

Finally if $Z^{\otimes n}g \in \mathscr{M}_4$, then by Lemma 3.9, $Z^{\otimes n}g = cZ^{\otimes n}f$, for some nonzero constant $c \in \mathbb{C}$, and $f = \mathrm{Sym}_n^{n-1}(\left[\begin{smallmatrix}1\\ 0\end{smallmatrix}\right]; \left[\begin{smallmatrix}0\\ 1\end{smallmatrix}\right])$ or its reversal $\mathrm{Sym}_n^{n-1}(\left[\begin{smallmatrix}0\\ 1\end{smallmatrix}\right]; \left[\begin{smallmatrix}1\\ 0\end{smallmatrix}\right])$. In either case, after canceling out Z, the weight 0 entry is 0 in the expression but not so in g; a contradiction. ∎

4 Mixing with Vanishing Signatures

In this section, we prove some hardness results for vanishing signature sets when augmented by other signatures. We first consider the mixing of vanishing signatures with unary and binary signatures. Over general graphs, these cases are handled by Lemma 7.1 and Lemma 7.2 in [6]. One can check that the hardness in Lemma 7.1 in [6] holds for planar graphs. We state the planar version of Lemma 7.1 in [6] and provide a proof for completeness. Specifically, the reduction to obtain the signature f'' is planar and Pl-Holant(f'') is #P-hard by Theorem 2.22.

Lemma 4.1. *Let $f \in \mathscr{V}^{\sigma}$ be a symmetric signature of arity n with $\mathrm{rd}^{\sigma}(f) = d \geq 2$ where $\sigma \in \{+, -\}$. Suppose $v = u^{\otimes m}$ is a symmetric degenerate signature for some unary signature u and some integer $m \geq 1$. If u is not a multiple of $[1, \sigma i]$, then* Pl-Holant(f, v) *is* #P*-hard.*

Proof. We consider $\sigma = +$ since the other case is similar. Since $f \in \mathscr{V}^{+}$, we have $n > 2d \geq 4$. Under a holographic transformation by Z, we have

$$\text{Pl-Holant}(f, v) \equiv \text{Pl-Holant}\left(\neq_2 \mid \hat{f}, [a, b]^{\otimes m}\right),$$

where $\hat{f} = \left(Z^{-1}\right)^{\otimes n} f$ and $[a, b]^{\otimes m} = \left(Z^{-1}\right)^{\otimes m} v$ with $b \neq 0$ since u is not a multiple of $[1, i]$. Moreover, $\hat{f} = [\hat{f}_0, \hat{f}_1, \ldots, \hat{f}_d, 0, \ldots, 0]$ with $\hat{f}_d \neq 0$ by Lemma 2.19.

We get $\widehat{f'} = [\hat{f}_{d-2}, \hat{f}_{d-1}, \hat{f}_d, 0, \ldots, 0]$ of arity $n - 2d + 4$ by $d - 2$ self-loops via $\neq_2$ on $\hat{f}$. This is on the right side. With two more self-loops, we get $[1, 0]^{\otimes n-2d}$, also on the right.

We claim that we can use $[1, 0]^{\otimes n-2d}$ and $[a, b]^{\otimes m}$ to create $[a, b]^{\otimes n-2d}$. Let $t = \gcd(m, n - 2d)$. If $n - 2d > m$, then we connect $[a, b]^{\otimes m}$ to $[1, 0]^{\otimes n-2d}$ via $\neq_2$ to get $[1, 0]^{\otimes n-2d-m}$ up to a nonzero factor $b \neq 0$. We repeat this process until we get a tensor power $[1, 0]^{\otimes \ell}$ for some $\ell \leq m$. We can do a similar construction if $m > n - 2d$. Repeat this process, which is a subtractive Euclidean algorithm. Halt upon getting both $[1, 0]^{\otimes t}$ and $[a, b]^{\otimes t}$. Then we combine $\frac{n-2d}{t}$ copies of $[a, b]^{\otimes t}$ to get $[a, b]^{\otimes n-2d}$.

Now connecting $[a, b]^{\otimes n-2d}$ back to $\widehat{f'}$ via $\neq_2$, gives $\widehat{f''} = [\widehat{f''}_0, \widehat{f''}_1, \widehat{f''}_2, 0, 0]$ of arity 4. Moreover, $\widehat{f''}_2 = b^{n-2d} \hat{f}_d \neq 0$. Notice that Pl-Holant$(\neq_2 \mid [\widehat{f''}_0, \widehat{f''}_1, \widehat{f''}_2, 0, 0]) \equiv$ Pl-Holant $(\neq_2 \mid [0, 0, 1, 0, 0])$, the Eulerian Orientation problem over planar 4-regular graphs, which is #P-hard by Corollary 2.29 (or more directly by [20, Theorem 3.7]). Thus, Pl-Holant(f, v) is #P-hard. □

Next come binary signatures. The statement of Lemma 7.2 in [6] must be modified to rule out a planar tractable case (which is proved #P-hard for general graphs in Lemma 7.2 in [6]). Excluding this planar tractable case, there is one more nonplanar reduction in the proof of Lemma 7.2 in [6]. This reduction is used to show that Holant $(\neq_2 \mid \{[t, 1, 0, 0, 0], [c, 0, 1]\})$ is #P-hard when $c \neq 0$ (since the gadget in Figure 12a of [6] is nonplanar). In the following lemma, we first show that this problem Holant $(\neq_2 \mid \{[t, 1, 0, 0, 0], [c, 0, 1]\})$ remains #P-hard even restricted to planar graphs provided $t \neq 0$. If $t = 0$, then all signatures belong to $\mathscr{M}$ and the problem is tractable.

Lemma 4.2. *Let $c, t \in \mathbb{C}$. If $ct \neq 0$, then* Pl-Holant $(\neq_2 \mid [t, 1, 0, 0, 0], [c, 0, 1])$ *is* #P*-hard.*

Proof. By connecting two copies of $\neq_2$ to either side of $[c, 0, 1]$, we get the signature $[1, 0, c]$ on the left. Clearly Pl-Holant $([1, 0, c] \mid [t, 1, 0, 0, 0]) \leq_T$ Pl-Holant $(\neq_2 \mid [t, 1, 0, 0, 0], [c, 0, 1])$. Then under a

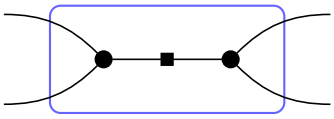


Figure 4: Circle vertices are assigned $[t, 1, 0, 0]$ and the square vertex is assigned $\neq_2$.

holographic transformation by T^{-1}, where $T = \begin{bmatrix} 1 & 0 \\ 0 & \sqrt{c} \end{bmatrix}$, we have

$$\begin{aligned} \text{Pl-Holant}\left([1,0,c] \mid [t,1,0,0,0]\right) &\equiv \text{Pl-Holant}\left([1,0,c](T^{-1})^{\otimes 2} \mid T^{\otimes 4}[t,1,0,0,0]\right) \\ &\equiv \text{Pl-Holant}\left([1,0,1] \mid [t,\sqrt{c},0,0,0]\right) \\ &\equiv \text{Pl-Holant}([t,\sqrt{c},0,0,0]). \end{aligned}$$

The last problem is #P-hard by Corollary 2.23 after dividing by $\sqrt{c}$. □

Next we prove the planar version of Lemma 7.2 in [6] using Lemma 4.2. We have to rule out the planar tractable case $f \in \mathscr{M}_4^{\pm}$. Also note that if $f \in \mathscr{V}^{\pm}$ is a symmetric non-degenerate signature, then f has arity at least 3. This is because a unary signature is degenerate, and if a binary symmetric signature f is vanishing, then its vanishing degree is greater than 1, hence at least 2, and therefore f is also degenerate. In the following lemma, we explicitly state this condition $\text{arity}(f) \geq 3$.

Lemma 4.3. *Let $f \in \mathscr{V}^{\sigma}$ be a symmetric non-degenerate signature of arity $n \geq 3$ for some $\sigma \in \{+, -\}$. Suppose h is a non-degenerate binary signature. If $f \notin \mathscr{M}_4^{\sigma}$ and $h \notin \mathscr{R}_2^{\sigma}$, then* Pl-Holant$(f, h)$ *is* #P*-hard.*

Proof. We consider $\sigma = +$ since the other case is similar. Under a Z transformation,

$$\text{Pl-Holant}(f, h) \equiv \text{Pl-Holant}\left(\neq_2 \mid \hat{f}, \hat{h}\right),$$

where $\hat{f} = \left(Z^{-1}\right)^{\otimes n} f$ and $\hat{h} = \left(Z^{-1}\right)^{\otimes 2} h$. Since $h \notin \mathscr{R}_2^{+}$, we may assume that $\hat{h} = [a, b, 1]$ by Lemma 2.19 with a nonzero entry $\hat{h}_2$. Moreover since h is non-degenerate, so is $\hat{h}$, and $b^2 \neq a$.

We prove the lemma by induction on the arity of f (or equivalently $\hat{f}$). There are two base cases, $n = 3$ and $n = 4$. However, the arity 3 case is easily reduced to the arity 4 case. We show this first, and then show that the lemma holds in the arity 4 case.

Assume $n = 3$. Since $f \in \mathscr{V}^{+}$, we have $\hat{f} = [t, 1, 0, 0]$ for some $t \neq 0$, by Lemma 2.19 and $f \notin \mathscr{M}_4^{+}$. Consider the gadget in Figure 4. We assign $\hat{f}$ to the circle vertices and $\neq_2$ to the square vertex. Let $\hat{f}'$ be the signature of the resulting gadget. The signature $\hat{f}'$ may not seem symmetric by construction, but it is not hard to verify that indeed $\hat{f}' = [2t, 1, 0, 0, 0]$. The crucial observation is that it takes the same value 0 on inputs 1010 and 1100, where bits are ordered counterclockwise, starting from an arbitrary edge. This finishes our reduction to $n = 4$.

Now we consider the base case of $n = 4$. Since $f \in \mathscr{V}^{+}$, we have $\text{vd}^{+}(f) > 2$ and $\text{rd}^{+}(f) < 2$. As f is not degenerate, $\text{rd}^{+}(f) \notin \{-1, 0\}$. It implies that $\text{rd}^{+}(f) = 1$ and by Lemma 2.19, $\hat{f} = [t, 1, 0, 0, 0]$.

Our next goal is to show that we can realize a signature of the form $[c, 0, 1]$ with $c \neq 0$. Then Pl-Holant $(\neq_2 \mid [t, 1, 0, 0, 0], [c, 0, 1]) \leq$ Pl-Holant(f, h). Moreover, $t \neq 0$ since $f \notin \mathscr{M}_4^{+}$. Then by Lemma 4.2, Pl-Holant $(\neq_2 \mid [t, 1, 0, 0, 0], [c, 0, 1])$ is #P-hard.

Figure 5: A sequence of binary gadgets that forms another binary gadget. The circles are assigned $[v, 1, 0]$, the squares are assigned $\neq_2$, and the triangle is assigned $[a, b, 1]$.

If $b = 0$, then $\hat{h}$ is what we want since in this case $a = a - b^2 \neq 0$.

Otherwise $b \neq 0$. By connecting $\hat{h}$ to $\hat{f}$ via $\neq_2$, we get $[t + 2b, 1, 0]$. If $t \neq -2b$, then by Lemma 2.20, we can interpolate any binary signature of the form $[v, 1, 0]$. Otherwise $t = -2b$. Then we connect two copies of $\hat{h}$ via $\neq_2$, and get $\widehat{h}' = [2ab, a + b^2, 2b]$. By connecting this $\widehat{h}'$ to $\hat{f}$ via $\neq_2$, we get $[2(a - b^2), 2b, 0]$, using $t = -2b$. Since $a \neq b^2$ and $b \neq 0$, we can once again interpolate any $[v, 1, 0]$ by Lemma 2.20.

Hence, we have the signature $[v, 1, 0]$, where $v \in \mathbb{C}$ is for us to choose. We construct the gadget in Figure 5 with the circles assigned $[v, 1, 0]$, the squares assigned $\neq_2$, and the triangle assigned $[a, b, 1]$. The resulting gadget has signature $[a + 2bv + v^2, b + v, 1]$, which can be verified by the matrix product

$$\begin{bmatrix} v & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} a & b \\ b & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} v & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} a + 2bv + v^2 & b + v \\ b + v & 1 \end{bmatrix}.$$

By setting $v = -b$, we get $[c, 0, 1]$, where $c = a - b^2 \neq 0$.

Now we do the induction step. Assume $n \geq 5$. Since f is non-degenerate, $\mathrm{rd}^+(f) \geq 1$. If $\mathrm{rd}^+(f) = 1$, then $\hat{f} = [t, 1, 0, \ldots, 0]$ for some $t \neq 0$. We connect $\hat{h}$ to $\hat{f}$ via $\neq_2$, getting $[t + 2b, 1, 0, \ldots, 0]$ of arity $n - 2 \geq 3$. If $t + 2b \neq 0$, then we are done by induction hypothesis. Otherwise $t = -2b$, and we connect two $\hat{h}$ together via $\neq_2$. The signature is $\hat{h}' := [2ab, b^2 + a, 2b]$. Connect $\hat{h}'$ to $\hat{f}$ via $\neq_2$. We get $[-4b^2 + 2(b^2 + a), 2b, 0, \ldots, 0] = [2(a - b^2), 2b, 0, \ldots, 0]$. If $b = 0$, then $t = 0$. Contradiction. Hence $b \neq 0$, and $a - b^2 \neq 0$ for b is not degenerate. Then we can apply induction hypothesis on $[2(a - b^2), 2b, 0, \ldots, 0]$.

The case left is that $\mathrm{rd}^+(f) = d \geq 2$. Then $\hat{f} = [\hat{f}_0, \hat{f}_1, \ldots, \hat{f}_d, 0, \ldots, 0]$ with $\hat{f}_d \neq 0$ by Lemma 2.19. We do a self-loop of $\hat{f}$ via $\neq_2$, getting $\hat{f}'' := [\hat{f}_1, \ldots, \hat{f}_d, 0, \ldots, 0]$ of arity $n - 2 \geq 3$. Since $d \geq 2$, $\hat{f}''$ is non-degenerate and $f'' = Z^{\otimes(n-2)} \hat{f}'' \in \mathscr{V}^+$. If $f'' \notin \mathscr{M}_4^+$, then apply the induction hypothesis and we are done. Otherwise $d = 2$ and we may assume $\hat{f} = [\hat{f}_0, 0, 1, 0, \ldots, 0]$ since $\hat{f}_2 \neq 0$.

In this case, we connect $\hat{h}$ to $\hat{f}$ via $\neq_2$, getting $\hat{f}''' := [a + \hat{f}_0, 2b, 1, 0, \ldots, 0]$ of arity $n - 2 \geq 3$. If $n \geq 7$, then we can apply the induction hypothesis. If $n = 6$, then $\hat{f}''' = [a + \hat{f}_0, 2b, 1, 0, 0]$ of arity 4. Notice that Pl-Holant $\left([0, 1, 0] \mid [a + \hat{f}_0, 2b, 1, 0, 0]\right)$ is equivalent to Pl-Holant $([0, 1, 0] \mid [0, 0, 1, 0, 0])$, which is counting Eulerian orientations in 4-regular planar graphs. Then Pl-Holant $\left(\neq_2 \mid \hat{f}'''\right)$ is #P-hard by Corollary 2.29.

The only case left now is when $n = 5$ and $\hat{f} = [\hat{f}_0, 0, 1, 0, 0, 0]$. We do two self-loops on $\hat{f}$ via $\neq_2$ to get $[1, 0]$. Then connect $[1, 0]$ to $\hat{h}$ via $\neq_2$ and get $[b, 1]$. At last, connect $[b, 1]$ to $\hat{f}$ via $\neq_2$, resulting in $[\hat{f}_0, b, 1, 0, 0]$. Similar to the case above, this is counting Eulerian orientations in 4-regular planar graphs, and is #P-hard by Corollary 2.29. □

If $f \in \mathscr{M}_4^{\pm}$, there is an additional case for the binary signature.

Lemma 4.4. *Let $f \in \mathscr{M}_4^\sigma$ be a symmetric non-degenerate signature with $\sigma \in \{+,-\}$ of arity $k \geq 3$. Suppose h is a non-degenerate binary signature such that $h \notin \mathscr{R}_2^\sigma$ and h is not a multiple of $Z^{\otimes 2}[a,0,1]$ for any $a \neq 0$. Then* Pl-Holant(f,h) *is* #P*-hard.*

Proof. We assume $f \in \mathscr{M}_4^+$ since the other case is similar. Suppose $h = Z^{\otimes 2}[a,b,c]$ for some $a,b,c \in \mathbb{C}$. Since $h \notin \mathscr{R}_2^+$, we have $c \neq 0$, so we assume $c = 1$. Moreover $b \neq 0$. This is because, if $b = 0$ then either h is degenerate or is a multiple of $Z^{\otimes 2}[a,0,1]$ for some $a \neq 0$. Either case is a contradiction. Then under a holographic transformation by Z, the problem becomes Pl-Holant $(\neq_2 \mid \text{EXACTONE}_k, [a,b,1])$. If we connect two copies of EXACTONE_k via $\neq_2$, we get EXACTONE_{2k-2}. Hence we may assume that $k \geq 5$. Then we connect $[a,b,1]$ to EXACTONE_k via $\neq_2$, and get $[2b,1,0,\ldots,0]$ of arity $k-2 \geq 3$. Since $b \neq 0$, Pl-Holant(f,h) is #P-hard by Lemma 4.3. □

Next we consider mixing signatures from $\mathscr{V}^+$ and $\mathscr{V}^-$. This is a planar version of Lemma 7.3 in [6]. However, for planar graphs, there is a tractable case when one signature is in $\mathscr{M}_4^+$ and the other is in $\mathscr{M}_4^-$. This case was shown to be #P-hard over general graphs by Lemma 6.12 in [6] using a nonplanar reduction. One can check that the rest of the proof of Lemma 7.3 in [6] holds for planar graphs. For completeness we include a proof.

Lemma 4.5. *Let $f \in \mathscr{V}^+$ and $g \in \mathscr{V}^-$ be symmetric non-degenerate signatures of arities ≥ 3 respectively. If $f \notin \mathscr{M}_4^+$ or $g \notin \mathscr{M}_4^-$ then* Pl-Holant(f,g) *is* #P*-hard.*

Proof. Let $\text{rd}^+(f) = d$, $\text{rd}^-(g) = d'$, $\text{arity}(f) = n$ and $\text{arity}(g) = n'$, then $2d < n$ and $2d' < n'$. Under a holographic transformation by $Z = \left[\begin{smallmatrix} 1 & 1 \\ i & -i \end{smallmatrix}\right]$, we have

$$\text{Pl-Holant}\,(=_2 \mid f,g) \equiv_T \text{Pl-Holant}\left(\neq_2 \mid \hat{f},\hat{g}\right),$$

where $\hat{f} := (Z^{-1})^{\otimes n} f = [\hat{f}_0,\ldots,\hat{f}_d,0,\ldots,0]$ and $\hat{g} := (Z^{-1})^{\otimes n'} g = [0,\ldots,0,\hat{g}_{d'},\ldots,\hat{g}_0]$ due to Lemma 2.19. Moreover $\hat{f}_d \neq 0$ and $\hat{g}_{d'} \neq 0$.

If $d \geq 2$, we can do d' many self-loops of $\neq_2$ on $\hat{g}$, getting $\hat{g}' := [0,\ldots,0,\hat{g}_{d'}]$ of arity $n'-2d' \geq 1$. Thus $g' := Z^{\otimes(n'-2d')}\hat{g}' = [1,-i]^{\otimes(n'-2d')}$ up to a nonzero constant. We apply Lemma 4.1 to derive that Pl-Holant(f,g) is #P-hard. If $d' \geq 2$, we can similarly get $[1,i]^{\otimes(n-2d)}$ and apply Lemma 4.1. Thus we can assume that $d = d' = 1$.

So up to nonzero constants, we have $\hat{f} = [a,1,0,\ldots,0]$ and $\hat{g} = [0,\ldots,0,1,b]$ for some $a,b \in \mathbb{C}$. We can assume that $f \notin \mathscr{M}_4^+$ and $a \neq 0$. The case of $b \neq 0$ is similar. We show that it is always possible to get two such signatures of the same arity $\min\{n,n'\}$. Suppose $n > n'$. We form a loop from $\hat{f}$ via $\neq_2$. It is easy to see that this signature is the degenerate signature $2[1,0]^{\otimes(n-2)}$. Similarly, we can form a loop from $\hat{g}$ and can get $2[0,1]^{\otimes(n'-2)}$. Thus we have both $[1,0]^{\otimes(n-2)}$ and $[0,1]^{\otimes(n'-2)}$. We can connect all $n'-2$ edges of the second to the first, connected by $\neq_2$. This gives $[1,0]^{\otimes(n-n')}$. We can continue subtracting the smaller arity from the larger one. We continue this process in a subtractive version of the Euclidean algorithm, and end up with both $[1,0]^{\otimes t}$ and $[0,1]^{\otimes t}$, where $t = \gcd(n-2,n'-2) = \gcd(n-n',n'-2)$. In particular, $t \mid n-n'$ and by taking $\frac{n-n'}{t}$ copies of $[0,1]^{\otimes t}$, we can get $[0,1]^{\otimes(n-n')}$. Connecting this back to $\hat{f}$ via $\neq_2$, we get a symmetric signature of arity n' consisting of the first $n'+1$ entries of $\hat{f}$. A similar proof works when $n' > n$.

Thus we may assume $n = n'$. Connecting $[0,1]^{\otimes(n-2)}$ to $\hat{f} = [a,1,0,\ldots,0]$ via $\neq_2$ we get $\hat{h} = [a,1,0]$. Recall that $a \neq 0$. Translating this back by Z, we have a binary signature $h \notin \mathscr{R}_2^-$ and h is not a multiple of $Z^{\otimes 2}[c,0,1]$ for any $c \neq 0$. Since $g \in \mathscr{V}^-$, by Lemma 4.3 or Lemma 4.4, Pl-Holant(g,h) is #P-hard. Hence Pl-Holant(f,g) is also #P-hard. □

When signatures in both $\mathscr{M}_4^+$ and $\mathscr{M}_4^-$ appear, we show that the only degenerate signatures that mix must also be vanishing.

Lemma 4.6. *Let $f \in \mathscr{M}_4^+$ and $g \in \mathscr{M}_4^-$ be two non-degenerate signatures of arity ≥ 3. Let $v = u^{\otimes m}$ be a degenerate signature for some unary signature u and some integer $m \geq 1$. If u is not a multiple of $[1, \pm i]$, then* Pl-Holant(f, g, v) *is* #P*-hard.*

Proof. Suppose f is of arity n and g of arity ℓ. Under a holographic transformation by Z, we have

$$\text{Pl-Holant}(f, g, v) \equiv \text{Pl-Holant}\left(\neq_2 \mid \textsc{ExactOne}_n, \textsc{AllButOne}_\ell, [a, b]^{\otimes m}\right),$$

where $ab \neq 0$. Notice that v is transformed to $(Z^{-1}u)^{\otimes m} = [a, b]^{\otimes m}$. We have $ab \neq 0$ since u is not a multiple of $[1, \pm i]$. First we get $[1, 0]^{\otimes n-2}$ by a self-loop via $\neq_2$ on $\textsc{ExactOne}_n$. By the same subtractive Euclidean argument as in Lemma 4.1, we can realize $[a, b]^{\otimes n-2}$ by $[1, 0]^{\otimes n-2}$ and $[a, b]^{\otimes m}$. Connecting $[a, b]^{\otimes n-2}$ to $\textsc{ExactOne}_n$ via $\neq_2$ we get a binary signature $h = [(n-2)ab^{n-3}, b^{n-2}, 0]$. After transforming back, we have

$$\text{Pl-Holant}(g, Z^{\otimes 2}h) \leq_T \text{Pl-Holant}(f, g, v).$$

However $Z^{\otimes 2}h \notin \mathscr{R}_2^-$ by Lemma 2.19 and it is not a multiple of $Z^{\otimes 2}[c, 0, 1]$ for any $c \neq 0$. Hence Pl-Holant(f, g, v) is #P-hard by Lemma 4.4, where $(g, Z^{\otimes 2}h)$ plays the role of "(f, h)" in Lemma 4.4 and $\sigma = -$. ☐

We also consider the mixing of vanishing signatures with those in $\mathscr{P}_2$.

Lemma 4.7. *Let $f \in \mathscr{V} \setminus \mathscr{M}_4$ and $g \in \mathscr{P}_2$ be two non-degenerate signatures with arities m and n respectively. If $m, n \geq 3$, then* Pl-Holant(f, g) *is* #P*-hard.*

Proof. We claim that it suffices to consider $f \in \mathscr{V}^+ \setminus \mathscr{M}_4$ and $g = \left[\begin{smallmatrix}1\\ i\end{smallmatrix}\right]^{\otimes n} + \left[\begin{smallmatrix}1\\ -i\end{smallmatrix}\right]^{\otimes n}$. By Lemma 3.2, we know that $g = \left[\begin{smallmatrix}1\\ i\end{smallmatrix}\right]^{\otimes n} + \beta \left[\begin{smallmatrix}1\\ -i\end{smallmatrix}\right]^{\otimes n}$ for some $\beta \neq 0$ up to a nonzero scalar. Under a holographic transformation by $T = Z \left[\begin{smallmatrix}1 & 0\\ 0 & \beta^{\frac{1}{n}}\end{smallmatrix}\right] Z^{-1}$, which is orthogonal up to a nonzero factor of $\beta^{\frac{1}{n}}$, we have $\hat{g} = (T^{-1})^{\otimes n} g = \left[\begin{smallmatrix}1\\ i\end{smallmatrix}\right]^{\otimes n} + \left[\begin{smallmatrix}1\\ -i\end{smallmatrix}\right]^{\otimes n}$. Now $\mathscr{M}_4$ is closed under orthogonal transformations by definition, and $\mathscr{V}$ is closed under orthogonal transformations by Lemma 2.14. Thus, we still have a signature $\hat{f} = (T^{-1})^{\otimes n} f$ such that $\hat{f} \in \mathscr{V} \setminus \mathscr{M}_4$. If $\hat{f} \in \mathscr{V}^-$, then under a holographic transformation by $D = \left[\begin{smallmatrix}1 & 0\\ 0 & -1\end{smallmatrix}\right]$, we have $\hat{f} \in \mathscr{V}^+$. Furthermore, $\hat{g}$ is invariant under D. This proves the claim.

Now we assume that $f \in \mathscr{V}^+ \setminus \mathscr{M}_4$ and $g = \left[\begin{smallmatrix}1\\ i\end{smallmatrix}\right]^{\otimes n} + \left[\begin{smallmatrix}1\\ -i\end{smallmatrix}\right]^{\otimes n}$. By Corollary 2.18, we have $\text{rd}^+(f) = d < \frac{m}{2}$. Under a holographic transformation by Z, we have

$$\begin{aligned}\text{Pl-Holant}\left(=_2 \mid f, g\right) &\equiv \text{Pl-Holant}\left([1, 0, 1]Z^{\otimes 2} \mid Z^{-1}\{f, g\}\right)\\ &\equiv \text{Pl-Holant}\left(\neq_2 \mid \hat{f}, =_n\right),\end{aligned}$$

where $\hat{f} = (Z^{-1})^{\otimes m} f$. By Lemma 2.19, the support of $\hat{f}$ is on entries with Hamming weight at most d and includes the entry of Hamming weight exactly d. Now $f \notin \mathscr{M}_4$, so by Lemma 3.9, we either have $d = 1$ and $\hat{f} = [\hat{f}_0, 1, 0, \ldots, 0]$ with $\hat{f}_0 \neq 0$ or $d \geq 2$ and $\hat{f} = [\hat{f}_0, \hat{f}_1, \ldots, \hat{f}_{d-1}, 1, 0, \ldots, 0]$ (and up to a nonzero scalar in either case).

In the first case, a self-loop on $\hat{f}$ via $\neq_2$ gives $[1, 0]^{\otimes m-2}$ on the right side. Let $r = \gcd(n, m-2)$, and let ℓ_1, ℓ_2 be two positive integers such that $\ell_1 n - \ell_2(m-2) = r$. We connect ℓ_1 copies of

$=_n$ with ℓ_2 copies of $[1,0]^{\otimes m-2}$ via $\neq_2$'s to get $[0,1]^{\otimes r}$. Since $r \mid m-2$, we can also realize $[0,1]^{\otimes m-2}$ by putting $\frac{m-2}{r}$ copies of $[0,1]^{\otimes r}$ together. Now connect $[0,1]^{\otimes m-2}$ to $\hat{f}$ via $\neq_2$. The resulting signature is $[\hat{f}_0, 1, 0]$. We can also move $=_n$ to the left using n copies of $\neq_2$. Hence, we have Pl-Holant$(=_n \mid [\hat{f}_0, 1, 0]) \leq_T$ Pl-Holant$(\neq_2 \mid \hat{f}, =_n)$. The former problem is #P-hard by Theorem 2.24 since $\hat{f}_0 \neq 0$, so the latter problem is #P-hard as well.

In the second case, we have $m \geq 5$ since $2 \leq d < \frac{m}{2}$. Furthermore, we may assume that $d = 2$, since otherwise can we do $d-2$ self-loops on $\hat{f}$ via $\neq_2$. With this assumption, we do two self-loops on $\hat{f}$ via $\neq_2$ to get $[1,0]^{\otimes m-4}$ on the right side. By a similar argument as in the previous case, we can construct $[0,1]^{\otimes m-4}$ by using $[1,0]^{\otimes m-4}$ and $=_n$ via $\neq_2$. Now connect $[0,1]^{\otimes m-4}$ back to $\hat{f}$ via $\neq_2$. We get the arity 4 signature $[\hat{f}_0, \hat{f}_1, 1, 0, 0]$. Hence, we have Pl-Holant$(\neq_2 \mid [\hat{f}_0, \hat{f}_1, 1, 0, 0]) \leq_T$ Pl-Holant$(\neq_2 \mid \hat{f}, =_n)$. Note that Pl-Holant$(\neq_2 \mid [\hat{f}_0, \hat{f}_1, 1, 0, 0])$ is equivalent to Pl-Holant$(\neq_2 \mid [0, 0, 1, 0, 0])$, counting Eulerian Orientations in planar 4-regular graphs, which is #P-hard by Corollary 2.29. Thus Pl-Holant$(\neq_2 \mid \hat{f}, =_n)$ is #P-hard as well. □

5 Dichotomy for Pl-#CSP^2 and Related Lemmas

In this section, we state the dichotomy for Pl-#CSP^2. We defer the proof to Part II of this paper starting on page 63. We provide a sketch of the proof here. Afterwards, we discuss several related lemmas, which are used for the full dichotomy of Pl-Holant. Let $\mathcal{T}_k = \left\{ \left[\begin{smallmatrix} 1 & 0 \\ 0 & \omega \end{smallmatrix}\right] \in \mathbb{C}^{2\times 2} \mid \omega^k = 1 \right\}$.

Theorem 5.1. *Let $\mathcal{F}$ be a set of symmetric signatures. Then* Pl-#CSP$^2(\mathcal{F})$ *is* #P*-hard unless $\mathcal{F}$ satisfies one of the following conditions:*

1. *there exists $T \in \mathcal{T}_8$ such that $\mathcal{F} \subseteq T\mathscr{A}$;*
2. *$\mathcal{F} \subseteq \mathscr{P}$;*
3. *there exists $T \in \mathcal{T}_4$ such that $\mathcal{F} \subseteq T\widehat{\mathscr{M}}$.*

In each exceptional case, Pl-#CSP$^2(\mathcal{F})$ *is computable in polynomial time.*

Proof Sketch. We first define some tractable families of signatures specific to the Pl-#CSP^2 framework. Let $\widetilde{\mathscr{A}} = \mathscr{A} \cup \left[\begin{smallmatrix} 1 & 0 \\ 0 & e^{\pi i/4} \end{smallmatrix}\right] \mathscr{A}$ and $\widetilde{\mathscr{M}} = \widehat{\mathscr{M}} \cup \left[\begin{smallmatrix} 1 & 0 \\ 0 & i \end{smallmatrix}\right] \widehat{\mathscr{M}}$. One can show that $\widetilde{\mathscr{A}}$ covers Case 1 above, and $\widetilde{\mathscr{M}}$ covers Case 3. The proof will revolve around these tractable classes.

The overall plan is to break the proof into two main steps.

The first step is to prove the dichotomy theorem for Pl-#CSP$^2(\mathcal{F})$ when there is at least one nonzero signature of *odd* arity in $\mathcal{F}$. In this case, we can make use of a lemma showing that we can simulate Pl-#CSP$(\mathcal{F})$ by Pl-#CSP$^2(\mathcal{F})$ if $\mathcal{F}$ includes a unary signature $[a, b]$ with $ab \neq 0$. Then we can apply the known dichotomy Theorem 2.25 for Pl-#CSP. However this strategy (provably) *cannot* work when every signature in $\mathcal{F}$ satisfies the *parity* constraint. In that case we employ other means. This first step of the proof is relatively uncomplicated.

The second step is to deal with the case when all nonzero signatures in $\mathcal{F}$ have even arity. This is where the real difficulties lie. In this case it is impossible to directly construct *any* unary signature. So we cannot use that lemma pertaining to a unary signature. But we prove another lemma which provides a way to simulate Pl-#CSP$(\mathcal{F})$ by Pl-#CSP$^2(\mathcal{F})$ in a *global* fashion, *if* $\mathcal{F}$ includes some tensor power of the form $[a, b]^{\otimes 2}$ where $ab \neq 0$. Moreover, we have a lucky break (for the complexity of the proof) if $\mathcal{F}$ includes a signature that is in $\widehat{\mathscr{M}} \setminus (\mathscr{P} \cup \widetilde{\mathscr{A}})$. In this case, we can construct a special binary signature, and obtain $[1, 1]^{\otimes 2}$ by interpolation. This proof uses the theory of *cyclotomic fields*. This simplifies the proof greatly. For all other cases (when $\mathcal{F}$ has only

even arity signatures), the proof gets going in earnest—we will attempt an induction on the arity of signatures.

The lowest arity of this induction will be 2. We will try to reduce the arity to 2 whenever possible; however for many cases an arity reduction to 2 destroys the #P-hardness at hand. Therefore the true basis of this induction proof of Pl-#CSP^2 starts with arity 4. Consequently we will first prove a dichotomy theorem for Pl-#CSP$^2(f)$, where f is a signature of arity 4. Several tools will be used. These include the rank criterion for redundant signatures, Theorem 2.24 for arity 2 signatures, and a trick we call the *Three Stooges* by domain pairing.

However, in the next step we do not attempt a general Pl-#CSP^2 dichotomy for a *single* signature of even arity. This would have been natural at this point, but it would have been too difficult. We will need some additional leverage by proving a conditional "No-Mixing" Lemma for pairs of signatures of even arity. So, seemingly taking a detour, we prove that for two signatures f and g both of even arity, that individually belong to some tractable class, but do not belong to a single tractable class in the conjectured dichotomy (that is yet to be proved), the problem Pl-#CSP$^2(f, g)$ is #P-hard. We prove this No-Mixing Lemma for any pair of signatures f and g both of even arity, not restricted to arity 4. Even though at this point we only have a dichotomy for a single signature of arity 4, we prove this No-Mixing Lemma for higher even arity pairs f and g by simulating two signatures f' and g' of arity 4 that belong to different tractable sets, from that of Pl-#CSP$^2(f, g)$. After this arity reduction (within the No-Mixing Lemma), we prove that Pl-#CSP$^2(f', g')$ is #P-hard by the dichotomy for a *single* signature of arity 4. After this, we prove a No-Mixing Lemma for a *set* of signatures $\mathcal{F}$ of even arities, which states that *if* $\mathcal{F}$ is contained in the union of all tractable classes, then it is still #P-hard unless it is *entirely* contained in one single tractable class. Note that at this point we still only have a *conditional* No-Mixing Lemma in the sense that we have to assume every signature in $\mathcal{F}$ belongs to some tractable set.

We then attempt the proof of a Pl-#CSP^2 dichotomy for a *single* signature of arbitrary even arity. This uses all the previous lemmas, in particular the (conditional) No-Mixing Lemma for a set of signatures. However, after completing the proof of this Pl-#CSP^2 dichotomy for a single signature of even arity, the No-Mixing Lemma becomes absolute.

Finally the dichotomy for a single signature of even arity is logically extended to a dichotomy theorem for Pl-#CSP$^2(\mathcal{F})$ where all signatures in $\mathcal{F}$ have even arity. Together with the first main step when $\mathcal{F}$ contains some nonzero signature of odd arity, this completes the proof of Theorem 5.1. □

5.1 Related Lemmas

Now we give some consequences of Theorem 5.1. These are cases that can be reduced to Pl-#CSP^2. We consider signatures in $\mathscr{P}_1$, $\mathscr{M}_2 \setminus \mathscr{P}_2$, $\mathscr{A}_3$, or $\mathscr{M}_3$.

We begin with the cases of $\mathscr{P}_1$ and $\mathscr{A}_3$. The following two lemmas are rephrased from [6]. One can check that the reductions in these proofs are planar.

Lemma 5.2 (Lemma 8.15 in [6]). *Let $f \in \mathscr{P}_1$ be a non-degenerate signature of arity $n \geq 3$ with an orthogonal transformation H and $\mathcal{F}$ be a set of signatures containing f. Let H_2 be the 2-by-2 matrix $\frac{1}{\sqrt{2}}\left[\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}\right]$. Then* Pl-#CSP$^2(H_2 H\mathcal{F}) \leq_T$ Pl-Holant$(\mathcal{F})$.

Lemma 5.3 (Lemma 8.17 in [6]). *Let $f \in \mathscr{A}_3$ be a non-degenerate signature of arity $n \geq 3$ with an orthogonal transformation H and $\mathcal{F}$ be a set of signatures containing f. Let $\alpha = e^{\pi i/4}$ and Y be the 2-by-2 matrix $\left[\begin{smallmatrix} \alpha & 1 \\ -\alpha & 1 \end{smallmatrix}\right]$. Then* Pl-#CSP$^2(Y H\mathcal{F} \cup \{[1, -i, 1]\}) \leq_T$ Pl-Holant$(\mathcal{F})$.

With these reductions, we can apply Theorem 5.1 to get the following corollaries. The next one follows directly from Lemma 5.2 and Theorem 5.1 as H_2 is orthogonal and every Pl-#CSP^2 tractable case is also tractable for Pl-Holant.

Corollary 5.4. *Let $\mathcal{F}$ be a set of signatures. Suppose there exists $f \in \mathcal{F}$ which is a non-degenerate signature of arity $n \geq 3$ in $\mathscr{P}_1$ with $H \in \mathbf{O}_2(\mathbb{C})$. Then* Pl-Holant$(\mathcal{F})$ *is* #P*-hard unless $\mathcal{F}$ is $\mathscr{A}$-, $\mathscr{P}$-, or $\mathscr{M}$-transformable, in which case* Pl-Holant$(\mathcal{F})$ *is tractable.*

The proof of this corollary is straightforward. To illustrate the power of Theorem 5.1, we give a short proof here.

Proof. Let $H' = (H_2 H)^{-1} \in \mathbf{O}_2(\mathbb{C})$. By Lemma 5.2 and Theorem 5.1, Pl-Holant$(\mathcal{F})$ is #P-hard unless either (1) $\mathcal{F} \subseteq H'\mathscr{P}$, or (2) $\mathcal{F} \subseteq H'T\mathscr{A}$, or (3) $\mathcal{F} \subseteq H'T' \left[\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}\right] \mathscr{M}$, where $T \in \mathcal{T}_8$ and $T' \in \mathcal{T}_4$. In case (1), $\mathcal{F}$ is $\mathscr{P}$-transformable since $(=_2)H'^{\otimes 2} = (=_2) \in \mathscr{P}$. In case (2), $\mathcal{F}$ is $\mathscr{A}$-transformable since $(=_2)(H'T)^{\otimes 2} = (=_2)T^{\otimes 2} \in \mathscr{A}$. In case (3), $\mathcal{F}$ is $\mathscr{M}$-transformable. If $T' = \left[\begin{smallmatrix} 1 & 0 \\ 0 & \pm 1 \end{smallmatrix}\right]$, then $T' \in \mathbf{O}_2(\mathbb{C})$. So $(=_2)(H'T' \left[\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}\right])^{\otimes 2} = (=_2) \in \mathscr{M}$. If $T' = \left[\begin{smallmatrix} 1 & 0 \\ 0 & \pm i \end{smallmatrix}\right]$, then $T' \left[\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}\right] = \left[\begin{smallmatrix} 1 & 1 \\ i & -i \end{smallmatrix}\right]$, and $(=_2)(H'T' \left[\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}\right])^{\otimes 2} = 2[0,1,0] \in \mathscr{M}$. □

Corollary 5.4 is useful in Section 8. In Section 6, we need the following further specialization.

Corollary 5.5. *Suppose f is a non-degenerate signature of arity $n \geq 5$. Let f' be f with a self loop, and assume that $f' \in \mathscr{P}_1$ is non-degenerate. Then* Pl-Holant(f) *is* #P*-hard unless f is $\mathscr{A}$-, $\mathscr{P}$-, or $\mathscr{M}$-transformable, in which case* Pl-Holant(f) *is tractable.*

For the other case of $\mathscr{A}_3$, some case analysis is required.

Corollary 5.6. *Let $\mathcal{F}$ be a set of signatures. Suppose there exists $f \in \mathcal{F}$ which is a non-degenerate signature of arity $n \geq 3$ in $\mathscr{A}_3$. Then* Pl-Holant$(\mathcal{F})$ *is* #P*-hard unless $\mathcal{F}$ is $\mathscr{A}$- or $\mathscr{M}$-transformable, in which case* Pl-Holant$(\mathcal{F})$ *is tractable.*

Proof. Assume that $f \in \mathscr{A}_3$ with an orthogonal transformation H. By Lemma 5.3, we have Pl-#CSP$^2(YH\mathcal{F} \cup \{[1,-i,1]\}) \leq_T$ Pl-Holant$(\mathcal{F})$, where $Y = \left[\begin{smallmatrix} \alpha & 1 \\ -\alpha & 1 \end{smallmatrix}\right]$ and $\alpha = e^{\pi i/4}$. Let $g = [1,-i,1]$ and $\mathcal{F}' = YH\mathcal{F} \cup \{g\}$.

We apply Theorem 5.1 to Pl-#CSP$^2(\mathcal{F}')$. The consequence is that Pl-#CSP$^2(\mathcal{F}')$ (and hence Pl-Holant (F)) is #P-hard unless $\mathcal{F}' \subseteq \mathscr{P}$, $\mathcal{F}' \subseteq \left[\begin{smallmatrix} 1 & 0 \\ 0 & i^r \end{smallmatrix}\right] \widehat{\mathscr{M}}$ for some integer $0 \leq r \leq 3$, or $\mathcal{F}' \subseteq \left[\begin{smallmatrix} 1 & 0 \\ 0 & \alpha^r \end{smallmatrix}\right] \mathscr{A}$ for some integer $0 \leq r \leq 7$ where $\alpha = e^{i\pi/4}$. Notice that $g \notin \mathscr{P}$ and hence the first case is impossible.

Suppose $\mathcal{F}' \subseteq \left[\begin{smallmatrix} 1 & 0 \\ 0 & i^r \end{smallmatrix}\right] \widehat{\mathscr{M}}$ for some integer $0 \leq r \leq 3$. Then as $g \notin \left[\begin{smallmatrix} 1 & 0 \\ 0 & i^r \end{smallmatrix}\right] \widehat{\mathscr{M}}$ for $r = 1, 3$, we have that $YH\mathcal{F} \subseteq \left[\begin{smallmatrix} 1 & 0 \\ 0 & \pm 1 \end{smallmatrix}\right] \widehat{\mathscr{M}}$. Moreover, notice that $\left[\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right] \widehat{\mathscr{M}} = \left[\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}\right] \left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right] \mathscr{M} = \left[\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}\right] \mathscr{M} = \widehat{\mathscr{M}}$. Hence $YH\mathcal{F} \subseteq \widehat{\mathscr{M}}$. Rewrite Y as $Y = \left[\begin{smallmatrix} 1 & 1 \\ -1 & 1 \end{smallmatrix}\right] \left[\begin{smallmatrix} \alpha & 0 \\ 0 & 1 \end{smallmatrix}\right]$. We deduce that

$$\begin{aligned} H\mathcal{F} &\subseteq \tfrac{1}{2} \left[\begin{smallmatrix} \alpha^{-1} & 0 \\ 0 & 1 \end{smallmatrix}\right] \left[\begin{smallmatrix} 1 & -1 \\ 1 & 1 \end{smallmatrix}\right] \widehat{\mathscr{M}} = \tfrac{1}{2} \left[\begin{smallmatrix} \alpha^{-1} & 0 \\ 0 & 1 \end{smallmatrix}\right] \left[\begin{smallmatrix} 1 & -1 \\ 1 & 1 \end{smallmatrix}\right] \left[\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}\right] \mathscr{M} \\ &= \left[\begin{smallmatrix} \alpha^{-1} & 0 \\ 0 & 1 \end{smallmatrix}\right] \left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right] \mathscr{M} = \mathscr{M}. \end{aligned}$$

Hence $\mathcal{F}$ is $\mathscr{M}$-transformable in this case.

The last case is when $\mathcal{F}' \subseteq \left[\begin{smallmatrix} 1 & 0 \\ 0 & \alpha^r \end{smallmatrix}\right] \mathscr{A}$ for some integer $0 \leq r \leq 7$. It implies that $r = 0, 2, 4, 6$ as $g \in \left[\begin{smallmatrix} 1 & 0 \\ 0 & \alpha^r \end{smallmatrix}\right] \mathscr{A}$ and $g \notin \left[\begin{smallmatrix} 1 & 0 \\ 0 & \alpha \end{smallmatrix}\right] \mathscr{A}$. That is, $\mathcal{F}' \subseteq \left[\begin{smallmatrix} 1 & 0 \\ 0 & i^l \end{smallmatrix}\right] \mathscr{A}$ for some integer $0 \leq l \leq 3$. Notice that $\left[\begin{smallmatrix} 1 & 0 \\ 0 & i^l \end{smallmatrix}\right] \in \mathrm{Stab}(\mathscr{A})$. It implies that $YH\mathcal{F} \subseteq \mathscr{A}$. Again, rewriting Y as $Y = \left[\begin{smallmatrix} 1 & 1 \\ -1 & 1 \end{smallmatrix}\right] \left[\begin{smallmatrix} \alpha & 0 \\ 0 & 1 \end{smallmatrix}\right]$, we have

$$H\mathcal{F} \subseteq \tfrac{1}{2} \left[\begin{smallmatrix} \alpha^{-1} & 0 \\ 0 & 1 \end{smallmatrix}\right] \left[\begin{smallmatrix} 1 & -1 \\ 1 & 1 \end{smallmatrix}\right] \mathscr{A} = \tfrac{1}{2} \left[\begin{smallmatrix} \alpha^{-1} & 0 \\ 0 & 1 \end{smallmatrix}\right] \mathscr{A}.$$

Therefore $\mathcal{F}$ is $\mathscr{A}$-transformable. This finishes the proof. □

Again, we specialize Corollary 5.6 to our need.

Corollary 5.7. *Let f be a non-degenerate signature of arity $n \geq 5$. Let f' be f with a self loop, and f' is non-degenerate and $f' \in \mathscr{A}_3$ with an orthogonal transformation H. Then* Pl-Holant(f) *is* #P*-hard unless f is $\mathscr{A}$- or $\mathscr{M}$-transformable, in which case* Pl-Holant(f) *is tractable.*

The next case is when f is in $\mathscr{M}_2$ but not $\mathscr{P}_2$.

Lemma 5.8. *Let $\mathcal{F}$ be a set of signatures. Suppose there exists $f \in \mathcal{F}$ which is a non-degenerate signature of arity $n \geq 3$ in $\mathscr{M}_2 \setminus \mathscr{P}_2$. Then* Pl-Holant $(\mathcal{F})$ *is* #P*-hard unless $\mathcal{F}$ is $\mathscr{A}$-, $\mathscr{P}$-, or $\mathscr{M}$-transformable, in which case* Pl-Holant $(\mathcal{F})$ *is tractable.*

Proof. As $f \in \mathscr{M}_2 \setminus \mathscr{P}_2$, assume $f = H^{\otimes n}\left(\left[\begin{smallmatrix}1\\ \gamma\end{smallmatrix}\right]^{\otimes n} \pm \left[\begin{smallmatrix}1\\ -\gamma\end{smallmatrix}\right]^{\otimes n}\right)$, where H is an orthogonal 2-by-2 matrix and $\gamma \neq 0, \pm i$.

We first show that

$$\text{Pl-\#CSP}^2(T^{-1}\mathcal{F}, g) \leq_T \text{Pl-Holant}\left(\{f\} \cup \mathcal{F}\right), \tag{5.4}$$

where $T = H\left[\begin{smallmatrix}1 & 1\\ \gamma & -\gamma\end{smallmatrix}\right]$ and $g = (=_2)T^{\otimes 2} = [1+\gamma^2, 1-\gamma^2, 1+\gamma^2]$

Assume that $f = H^{\otimes n}\left(\left[\begin{smallmatrix}1\\ \gamma\end{smallmatrix}\right]^{\otimes n} + \left[\begin{smallmatrix}1\\ -\gamma\end{smallmatrix}\right]^{\otimes n}\right)$ with the $+$ sign. In this case, we do the transformation T:

$$\begin{aligned}\text{Pl-Holant}\left(=_2 \mid f, \mathcal{F}\right) &\equiv_T \text{Pl-Holant}\left([1,0,1]H^{\otimes 2}\left[\begin{smallmatrix}1 & 1\\ \gamma & -\gamma\end{smallmatrix}\right]^{\otimes 2} \;\middle|\; \left(\left[\begin{smallmatrix}1 & 1\\ \gamma & -\gamma\end{smallmatrix}\right]^{-1}\right)^{\otimes n}\left(H^{-1}\right)^{\otimes n} f, T^{-1}\mathcal{F}\right)\\ &\equiv_T \text{Pl-Holant}\left(g \mid =_n, T^{-1}\mathcal{F}\right).\end{aligned}$$

By connecting g to $=_n$, we get $=_{n-2}$ up to a constant factor of $1+\gamma^2 \neq 0$ as $\gamma \neq \pm i$. We repeat this process. If n is even, then we get $=_2$ eventually, which is on the right hand side. If n is odd, then eventually we get $=_3$ and $(=_1) = [1,1]$ on the right. Connecting $[1,1]$ to g we get $2[1,1]$ on the left. Then connecting $[1,1]$ to $=_3$ we get $=_2$ on the right. To summarize, we get that

$$\begin{aligned}\text{Pl-Holant}\left(g \mid =_2, =_n, T^{-1}\mathcal{F}\right) &\leq_T \text{Pl-Holant}\left(g \mid =_n, T^{-1}\mathcal{F}\right)\\ &\leq_T \text{Pl-Holant}\left(f, \mathcal{F}\right).\end{aligned} \tag{5.5}$$

Next we show that

$$\text{Pl-Holant}\left(=_2, g \mid =_2, =_n, T^{-1}\mathcal{F}\right) \leq_T \text{Pl-Holant}\left(g \mid =_2, =_n, T^{-1}\mathcal{F}\right). \tag{5.6}$$

Let $N = \left[\begin{smallmatrix}1+\gamma^2 & 1-\gamma^2\\ 1-\gamma^2 & 1+\gamma^2\end{smallmatrix}\right]$ be the signature matrix of g. If there is a positive integer k and a nonzero constant c such that $N^k = cI_2$, where I_2 is the 2-by-2 identity matrix, then we may directly implement $=_2$ on the left by connecting k copies of $[1+\gamma^2, 1-\gamma^2, 1+\gamma^2]$ via $=_2$ on the right. It implies (5.6) holds.

Otherwise such k and c do not exist. The two eigenvalues of N are $\lambda_1 = 2$ and $\lambda_2 = 2\gamma^2$. If $\lambda_1 = \lambda_2$, then $\gamma^2 = 1$ and $N = [\begin{smallmatrix}2 & 0\\ 0 & 2\end{smallmatrix}]$. Contradiction. Hence $\lambda_1 \neq \lambda_2$, and N is diagonalizable. Let $N = P\left[\begin{smallmatrix}\lambda_1 & 0\\ 0 & \lambda_2\end{smallmatrix}\right]P^{-1}$, for some non-singular matrix P. By connecting l many copies of N on the left

via $=_2$ on the right, where l is a positive integer, we can implement $N^l = P \begin{bmatrix} \lambda_1^l & 0 \\ 0 & \lambda_2^l \end{bmatrix} P^{-1}$. Since N does not have finite order up to a scalar, for any positive integer l, $(\lambda_1/\lambda_2)^l \neq 1$.

Consider an instance Ω of Pl-Holant $\left(=_2, g \mid =_2, =_n, T^{-1}\mathcal{F}\right)$. Suppose that the left $=_2$ appears t times. Let l be a positive integer. We obtain Ω_l from Ω by replacing each occurrence of $=_2$ on the left with N^l.

Since $N^l = P \begin{bmatrix} \lambda_1^l & 0 \\ 0 & \lambda_2^l \end{bmatrix} P^{-1}$, we can view our construction of Ω_l as replacing N^l by 3 signatures, with matrix P, $\begin{bmatrix} \lambda_1^l & 0 \\ 0 & \lambda_2^l \end{bmatrix}$, and P^{-1}, respectively. This does not change the Holant value,

We stratify the assignments in Ω_l based on the assignments to the t occurrences of the signature whose matrix is the diagonal matrix $\begin{bmatrix} \lambda_1^l & 0 \\ 0 & \lambda_2^l \end{bmatrix}$. Suppose there are i many times it was assigned 00 with function value λ_1^l, and j times 11 with function value λ_2^l. Clearly $i+j = t$ if the assignment has a nonzero evaluation. Let c_{ij} be the sum over all such assignments of the products of evaluations of all signatures (including the signatures corresponding to matrices P and P^{-1}) in Ω_l except for this diagonal one. Then

$$\begin{aligned} \text{Holant}_{\Omega_l} &= \sum_{i+j=t} \left(\lambda_1^l\right)^i \left(\lambda_2^l\right)^j c_{ij} \\ &= \lambda_2^{lt} \sum_{0 \leq i \leq t} \left(\left(\frac{\lambda_1}{\lambda_2}\right)^l\right)^i c_{i,t-i}. \end{aligned}$$

By an oracle of Pl-Holant $\left(g \mid =_2, =_n, T^{-1}\mathcal{F}\right)$, we can get Holant_{Ω_l} for any $1 \leq l \leq t+1$. Recall that for any positive integer l, $(\lambda_1/\lambda_2)^l \neq 1$. This implies that for any two distinct integers $i, j \geq 0$, $(\lambda_1/\lambda_2)^i \neq (\lambda_1/\lambda_2)^j$. Therefore we get a non-singular Vandermonde system. We can solve all c_{ij} for $i + j = t$ given Holant_{Ω_l} for all $1 \leq l \leq t+1$. Then notice that $\sum_{i+j=t} c_{ij}$ is the Holant value of Ω_l by replacing both λ_1^l and λ_2^l with 1, which is the instance Ω as $PI_2P^{-1} = I_2$. Therefore we may compute Holant_Ω via $t+1$ many oracle calls to Pl-Holant $\left(g \mid =_2, =_n, T^{-1}\mathcal{F}\right)$. This finishes the reduction in (5.6).

In the left hand side of (5.6) we have $=_2$ on both sides. Therefore we may lift the bipartite restriction. Combining it with (5.5), we get

$$\text{Pl-Holant}\left(=_n, g, T^{-1}\mathcal{F}\right) \leq_T \text{Pl-Holant}\left(f, \mathcal{F}\right).$$

Notice that given an equality of arity $n \geq 3$, we can always construct all equalities of even arity, regardless of the parity of n, in the Pl-Holant setting. Therefore, we have Pl-$\#\text{CSP}^2(T^{-1}\mathcal{F}, g) \leq_T$ Pl-Holant $(f, \mathcal{F})$.

To prove (5.4), there is another case that $f = H^{\otimes n}\left(\left[\begin{smallmatrix} 1 \\ \gamma \end{smallmatrix}\right]^{\otimes n} - \left[\begin{smallmatrix} 1 \\ -\gamma \end{smallmatrix}\right]^{\otimes n}\right)$, with the $-$ sign. Again we do a T transformation, where $(T^{-1})^{\otimes} f = [1, 0, \ldots, 0, -1]$ has arity n:

$$\text{Pl-Holant}\left(=_2 \mid f, \mathcal{F}\right) \equiv_T \text{Pl-Holant}\left(g \mid [1, 0, \ldots, 0, -1], T^{-1}\mathcal{F}\right).$$

We then do the same construction as in the previous case of connecting g to $[1, 0, \ldots, 0, -1]$ repeatedly. Depending on the parity of n, we have two cases.

1. If n is odd, then eventually we get $[1, 0, 0, -1]$ and $[1, -1]$ on the right as $\gamma \neq \pm i$, and therefore $2\gamma^2[1, -1]$, i.e., $[1, -1]$ on the left as $\gamma \neq 0$. Then connecting $[1, -1]$ to $[1, 0, 0, -1]$ we get $=_2$

on the right. Thus, for odd n,

$$\begin{aligned}\text{Pl-Holant}\left(g \mid =_2, [1,0,\ldots,0,-1], T^{-1}\mathcal{F}\right) &\leq_T \text{Pl-Holant}\left(g \mid [1,0,\ldots,0,-1], T^{-1}\mathcal{F}\right)\\ &\leq_T \text{Pl-Holant}\left(f, \mathcal{F}\right).\end{aligned}$$

Notice that our previous binary interpolation proof only relies on g and $=_2$. Hence we get

$$\begin{aligned}\text{Pl-Holant}\left(g \mid =_2, [1,0,\ldots,0,-1], T^{-1}\mathcal{F}\right) &\geq_T \text{Pl-Holant}\left(=_2, g \mid =_2, [1,0,\ldots,0,-1], T^{-1}\mathcal{F}\right)\\ &\equiv_T \text{Pl-Holant}([1,0,\ldots,0,-1], g, T^{-1}\mathcal{F}).\end{aligned}$$

Moreover it is straightforward to construct all even equalities from $[1,0,\ldots,0,-1]$ in the normal Pl-Holant setting as $n \geq 5$. Combining everything together gives us

$$\text{Pl-\#CSP}^2(g, T^{-1}\mathcal{F}) \leq_T \text{Pl-Holant}\left(f, \mathcal{F}\right).$$

2. Otherwise n is even. By the same construction of connecting g to $[1,0,\ldots,0,-1]$ repeatedly, we get $[1,0,0,0,-1]$ and $[1,0,-1]$ on the right eventually. Then we connect two copies of g via $[1,0,-1]$, resulting in $\begin{bmatrix}1+\gamma^2 & 1-\gamma^2\\ 1-\gamma^2 & 1+\gamma^2\end{bmatrix}\begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix}\begin{bmatrix}1+\gamma^2 & 1-\gamma^2\\ 1-\gamma^2 & 1+\gamma^2\end{bmatrix} = 4\gamma^2\begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix}$ on the left. Then connect $[1,0,-1]$ to $[1,0,0,0,-1]$ to get $[1,0,1]$ on the right. At last we connect two $[1,0,-1]$'s on the left via $[1,0,1]$ on the right to get $[1,0,1]$ on the left. Then it reduces to the previous case.

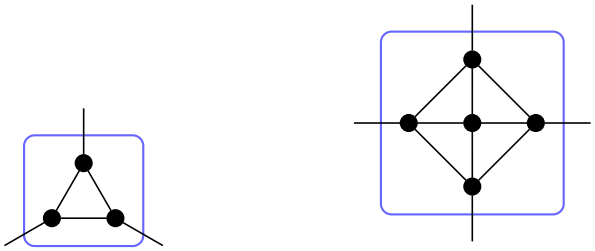
This concludes the proof of (5.4).

We apply Theorem 5.1 to Pl-\#CSP$^2(T^{-1}\mathcal{F}, g)$. Then we have that Pl-\#CSP$^2(T^{-1}\mathcal{F}, g)$ (and hence Pl-Holant$(f, \mathcal{F})$) is \#P-hard unless $T^{-1}\mathcal{F} \cup \{g\} \subseteq \mathscr{P}$, or $T^{-1}\mathcal{F} \cup \{g\} \subseteq \begin{bmatrix}1 & 0\\ 0 & i^r\end{bmatrix}\widehat{\mathscr{M}}$ for some integer $0 \leq r \leq 3$, or $T^{-1}\mathcal{F} \cup \{g\} \subseteq \begin{bmatrix}1 & 0\\ 0 & \alpha^r\end{bmatrix}\mathscr{A}$ for some integer $0 \leq r \leq 7$ where $\alpha = e^{i\pi/4}$. We have three cases.

1. The first case is that $T^{-1}\mathcal{F} \cup \{g\} \subseteq \mathscr{P}$. Recall that $\gamma \neq 0$ or $\pm i$, it can be verified that $g \notin \mathscr{P}$ unless $\gamma^2 = 1$. Hence $\gamma = \pm 1$. In either case we have that $\begin{bmatrix}1 & 1\\ \gamma & -\gamma\end{bmatrix}$ is an orthogonal matrix up to a nonzero scalar, and hence so is T. It implies that $\mathcal{F}$ is $\mathscr{P}$-transformable.
2. Next suppose $T^{-1}\mathcal{F} \cup \{g\} \subseteq \begin{bmatrix}1 & 0\\ 0 & i^r\end{bmatrix}\widehat{\mathscr{M}}$ for some integer $0 \leq r \leq 3$. If $\gamma = \pm 1$, then T is an orthogonal matrix as $\begin{bmatrix}1 & 1\\ \gamma & -\gamma\end{bmatrix}$ is, up to a factor of $\frac{1}{\sqrt{2}}$. Hence $\mathcal{F}$ is $\mathscr{M}$-transformable, as $\mathcal{F} \subseteq T\begin{bmatrix}1 & 0\\ 0 & i^r\end{bmatrix}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}\mathscr{M}$ and $(=_2)\left(T\begin{bmatrix}1 & 0\\ 0 & i^r\end{bmatrix}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}\right)^{\otimes 2}$ is either $[1,0,1]$ when $r = 0, 2$, or $[0,1,0]$ when $r = 1, 3$, up to a nonzero factor.
Otherwise $\gamma^2 \neq 1$ and it is straightforward to verify that $g \notin \begin{bmatrix}1 & 0\\ 0 & i^r\end{bmatrix}\widehat{\mathscr{M}}$ for $r = 1, 3$. Hence we may assume that $T^{-1}\mathcal{F} \subseteq \begin{bmatrix}1 & 0\\ 0 & \pm 1\end{bmatrix}\widehat{\mathscr{M}}$. Moreover, $\begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix}\widehat{\mathscr{M}} = \begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}\mathscr{M} = \begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}\mathscr{M} = \widehat{\mathscr{M}}$. Then $T^{-1}\mathcal{F} \subseteq \widehat{\mathscr{M}}$. As $T^{-1} = \begin{bmatrix}1 & 1\\ \gamma & -\gamma\end{bmatrix}^{-1}H^{-1}$, it implies that

$$\begin{aligned}H^{-1}\mathcal{F} \subseteq \begin{bmatrix}1 & 1\\ \gamma & -\gamma\end{bmatrix}\widehat{\mathscr{M}} &= \begin{bmatrix}1 & 0\\ 0 & \gamma\end{bmatrix}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}\mathscr{M}\\ &= \begin{bmatrix}1 & 0\\ 0 & \gamma\end{bmatrix}\mathscr{M} = \mathscr{M}.\end{aligned}$$

Hence $\mathcal{F} \subseteq H\mathscr{M}$ and $\mathcal{F}$ is $\mathscr{M}$-transformable.
3. In the last case, $T^{-1}\mathcal{F} \cup \{g\} \subseteq \begin{bmatrix}1 & 0\\ 0 & \alpha^r\end{bmatrix}\mathscr{A}$ for some integer $0 \leq r \leq 7$. If $\gamma = \pm 1$, then T is an orthogonal matrix as $\begin{bmatrix}1 & 1\\ \gamma & -\gamma\end{bmatrix}$ is, up to a factor of $\frac{1}{\sqrt{2}}$. Hence $\mathcal{F}$ is $\mathscr{A}$-transformable, as $\mathcal{F} \subseteq T\begin{bmatrix}1 & 0\\ 0 & \alpha^r\end{bmatrix}\mathscr{A}$ and $(=_2)\left(T\begin{bmatrix}1 & 0\\ 0 & \alpha^r\end{bmatrix}\right)^{\otimes 2}$ is $[1,0,i^r] \in \mathscr{A}$, up to a nonzero factor.
Otherwise $\gamma^2 \neq 1$ and $g \notin \begin{bmatrix}1 & 0\\ 0 & \alpha^r\end{bmatrix}\mathscr{A}$ for any integer $r = 1, 3, 5, 7$. Hence $T^{-1}\mathcal{F} \cup \{g\} \subseteq \mathscr{A}$ as $\begin{bmatrix}1 & 0\\ 0 & i^r\end{bmatrix}\mathscr{A} = \mathscr{A}$ for any integer $0 \leq r \leq 3$. If $\frac{1+\gamma^2}{1-\gamma^2} \neq \pm i$, then one can check that $g \notin \mathscr{A}$. A



(a) Triangle gadget (b) Planar tetrahedron gadget

Figure 6: Two gadgets used to create a signature in $\mathscr{M}_2 \setminus \mathscr{P}_2$.

contradiction. Otherwise $\frac{1+\gamma^2}{1-\gamma^2} = \pm i$. It implies that $\gamma = \alpha^l$ for some integer $l = 1, 3, 5, 7$. We may assume $l = 1$ as other cases are similar. In this case it is possible that $T^{-1}\mathcal{F} \cup \{g\} \subseteq \mathscr{A}$. As $T^{-1} = \left[\begin{smallmatrix} 1 & 1 \\ \gamma & -\gamma \end{smallmatrix}\right]^{-1} H^{-1} = \left[\begin{smallmatrix} 1 & 1 \\ \alpha & -\alpha \end{smallmatrix}\right]^{-1} H^{-1}$, it implies that

$$H^{-1}\mathcal{F} \subseteq \left[\begin{smallmatrix} 1 & 1 \\ \alpha & -\alpha \end{smallmatrix}\right] \mathscr{A} = \left[\begin{smallmatrix} 1 & 0 \\ 0 & \alpha \end{smallmatrix}\right] \left[\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}\right] \mathscr{A} = \left[\begin{smallmatrix} 1 & 0 \\ 0 & \alpha \end{smallmatrix}\right] \mathscr{A}.$$

Hence, $\mathcal{F}$ is $\mathscr{A}$-transformable, so Pl-Holant$(\mathcal{F})$ is tractable. This finishes the proof. □

Lemma 5.8 leads to the following specialization.

Corollary 5.9. *Let f be a non-degenerate signature of arity $n \geq 5$. Let f' be f with a self loop, and f' is non-degenerate and $f' \in \mathscr{M}_2 \setminus \mathscr{P}_2$. Then* Pl-Holant$(f)$ *is* #P*-hard unless f is $\mathscr{A}$-, $\mathscr{P}$-, or $\mathscr{M}$-transformable, in which case* Pl-Holant(f) *is tractable.*

We can reduce the case of $f \in \mathscr{M}_3$ to the previous case.

Lemma 5.10. *Let $\mathcal{F}$ be a set of signatures. Suppose there exists $f \in \mathcal{F}$ which is a non-degenerate signature of arity $n \geq 3$ in $\mathscr{M}_3$ with $H \in \mathbf{O}_2(\mathbb{C})$. Then* Pl-Holant$(\mathcal{F})$ *is* #P*-hard unless $\mathcal{F} \subseteq H\mathscr{M}$, in which case $\mathcal{F}$ is $\mathscr{M}$-transformable and* Pl-Holant$(\mathcal{F})$ *is tractable.*

Proof. We first claim that Pl-Holant$(\mathcal{F})$ is #P-hard unless $\mathcal{F}$ is $\mathscr{A}$-, $\mathscr{P}$-, or $\mathscr{M}$-transformable.

By the definition of $\mathscr{M}_3$, we may assume that $f = \textsc{ExactOne}_n$ is of arity n after an orthogonal transformation H. After zero or more self loops, we can further assume that either $f = \textsc{ExactOne}_3$ or $f = \textsc{ExactOne}_4$ depending on the parity of n.

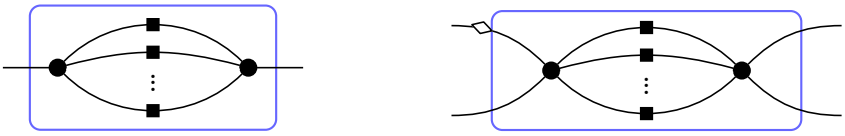
Suppose $f = \textsc{ExactOne}_3$. Consider the gadget in Figure 6a. We assign f to all vertices. The signature of the resulting gadget is $g = [0, 1, 0, 1]$, which is in $\mathscr{M}_2$ and not in $\mathscr{P}_2 = \mathscr{A}_2$ by Lemma 3.2. Thus, the claim follows from Lemma 5.8.

Otherwise, $f = \textsc{ExactOne}_4$. Consider the gadget in Figure 6b. We assign f to all vertices. Note that this is a matchgate. The signature of the resulting gadget is $[0, 2, 0, 1, 0]$, which is in $\mathscr{M}_2$ and not in $\mathscr{P}_2 = \mathscr{A}_2$ by Lemma 3.2. Thus, the claim follows from Lemma 5.8.

However, as $f \in \mathcal{F}$ and $f \in \mathscr{M}_3$, $\mathcal{F}$ cannot be $\mathscr{A}$- or $\mathscr{P}$-transformable by Lemma 3.11. Also by Lemma 3.11, if $\mathcal{F}$ is $\mathscr{M}$-transformable, then $\mathcal{F} \subseteq HD\mathscr{M}$ or $H\left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right] D\mathscr{M}$ for some diagonal matrix D. Notice that $D \in \text{Stab}(\mathscr{M})$ and $\left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right] D \in \text{Stab}(\mathscr{M})$. It implies that $\mathcal{F} \subseteq H\mathscr{M}$. □

Once again, we specialize Lemma 5.10 to our needs.

Corollary 5.11. *Let f be a non-degenerate signature of arity $n \geq 5$. Let f' be f with a self loop, and f' is non-degenerate and $f' \in \mathscr{M}_3$. Then* Pl-Holant(f) *is* #P*-hard unless f is $\mathscr{M}$-transformable, in which case* Pl-Holant(f) *is tractable.*



(a) A binary construction (b) An arity-4 construction

Figure 7: Two gadgets used. In the normal basis, circles are assigned f and squares are assigned $=_2$. In the Z basis, circles are assigned $\hat{f}$ and squares are assigned $\neq_2$.

6 Single Signature Dichotomy

Theorem 6.1 is the single signature dichotomy for Pl-Holant problems.

Theorem 6.1. *If f is a non-degenerate symmetric signature of arity $n \geq 3$ with complex weights in Boolean variables, then* Pl-Holant(f) *is* #P*-hard unless $f \in \mathscr{P}_1 \cup \mathscr{M}_2 \cup \mathscr{A}_3 \cup \mathscr{M}_3 \cup \mathscr{M}_4 \cup \mathscr{V}$, in which case the problem is computable in polynomial time.*

We prove Theorem 6.1 by induction on the arity. Before proceeding to the proof, we first introduce several lemmas involved in the inductive step.

6.1 Lemmas applied to Non-Degenerate Signatures in the Inductive Step

The single signature dichotomy relies on the following key lemma. The important assumption here is that f' is non-degenerate.

Lemma 6.2. *Suppose f is a non-degenerate signature of arity $n \geq 5$. Let f' be f with a self loop. If $f' \in \mathscr{P}_1 \cup \mathscr{M}_2 \cup \mathscr{A}_3 \cup \mathscr{M}_3 \cup \mathscr{V}$ is non-degenerate, then* Pl-Holant(f) *is* #P*-hard unless $f \in \mathscr{P}_1 \cup \mathscr{M}_2 \cup \mathscr{A}_3 \cup \mathscr{M}_3 \cup \mathscr{V}$.*

Lemma 6.2 depends on several results, each of which handles a different case. In fact, the proof of Lemma 6.2 is a straightforward combination of Corollary 5.5 (for $\mathscr{P}_1$), Corollary 5.7 (for $\mathscr{A}_3$), Corollary 5.9 (for $\mathscr{M}_2 \setminus \mathscr{P}_2$), and Corollary 5.11 (for $\mathscr{M}_3$) from Section 5, as well as Corollary 6.4 (for $\mathscr{P}_2$) and Lemma 6.5 (for $\mathscr{V}$), which we will prove shortly. These last two results handle the cases $f' \in \mathscr{P}_2$ and $f' \in \mathscr{V}$ respectively. First we consider the case of $f' \in \mathscr{P}_2$ and show the following lemma.

Lemma 6.3. *Let f be a non-degenerate signature of arity $n \geq 5$. If $f = Z^{\otimes n}[a, 1, 0, \ldots, 0, 1, b]$ for some $a, b \in \mathbb{C}$, where the number of 0's is $n-3$. Then* Pl-Holant(f) *is* #P*-hard.*

Proof. First we use the gadget in Figure 7b, where we put f on both vertices. Let the resulting signature be $h = Z^{\otimes 4}\hat{h}$. It is easier to calculate $\hat{h}$, that is, h in the Z basis. Indeed, $\hat{h}$ is not symmetric, but $\hat{h}$ has the following matrix representation as $n \geq 5$:

$$M_{\hat{h}} = \begin{bmatrix} 0 & a & a & ab+(n-2) \\ a & 2 & 2 & b \\ a & 2 & 2 & b \\ ab+(n-2) & b & b & 0 \end{bmatrix}.$$

Notice that this matrix is redundant, and $\det(\widetilde{M_{\hat{h}}}) = -4(n-2)(ab+n-2)$. If $ab \neq 2-n$, then by Corollary 2.29 Pl-Holant(h) is #P-hard, and so is Pl-Holant(f). Hence in the following we assume $ab = 2-n$.

Let f' be f with a self loop. Then apply the Z transformation as follows:

$$\begin{aligned}\text{Pl-Holant}\left(=_2 \mid f, f'\right) &\equiv_T \text{Pl-Holant}\left([0,1,0] \mid \hat{f}, \widehat{f'}\right)\\ &\equiv_T \text{Pl-Holant}\left([0,1,0] \mid \hat{f}, [1,0,\ldots,0,1]\right),\end{aligned}$$

where $\widehat{f'} = [1,0,\ldots,0,1]$ and $\hat{f} = [a,1,0,\ldots,0,1,b]$ for some $a, b \in \mathbb{C}$. We get this expression of $\widehat{f'}$ because doing a self loop commutes with the operation of holographic transformations.

We connect $\widehat{f'}$ to $\hat{f}$ via $[0,1,0]$, getting $[a,2,b]$. Then we connect $[a,2,b]$ to $\hat{f}$ via $[0,1,0]$ again, getting $\hat{g} = [ab+4, b, 0, \ldots, 0, a, ab+4]$ of arity $n-2$.

If $n \geq 7$, then we use the gadget in Figure 7b again, where we put g on both vertices this time. We get some signature h', which in Z basis has the following matrix representation as $n-2 \geq 5$:

$$M_{\widehat{h'}} = \begin{bmatrix} 0 & a(ab+4) & a(ab+4) & (n-4)ab + (ab+4)^2 \\ a(ab+4) & 2ab & 2ab & b(ab+4) \\ a(ab+4) & 2ab & 2ab & b(ab+4) \\ (n-4)ab + (ab+4)^2 & b(ab+4) & b(ab+4) & 0 \end{bmatrix}.$$

Once again this matrix is redundant. It can be simplified as $ab = 2-n$. The compressed matrix is

$$\widetilde{M_{\widehat{h'}}} = \begin{bmatrix} 0 & -2(n-6)a & -6n+28 \\ -(n-6)a & 8-4n & -(n-6)b \\ -6n+28 & -2(n-6)b & 0 \end{bmatrix}.$$

It is easy to compute that $\det(\widetilde{M_{\widehat{h'}}}) = -8(3n-14)(ab(n-6)^2 - 6n^2 + 40n - 56) = 8(n-4)(n-2)^2(3n-14)$. Since $n \geq 7$, $\det(\widetilde{M_{\widehat{h'}}}) > 0$. Then by Corollary 2.29 Pl-Holant(h') is #P-hard, and so is Pl-Holant(f).

The remaining cases are $n = 6$ and $n = 5$. When $n = 6$, $ab = 2-n = -4$. Moreover, $\hat{g}$ is of arity 4 and $\hat{g} = [ab+4, b, 0, a, ab+4] = [0, b, 0, a, 0]$. We do one more self loop on g via $[0,1,0]$ in the Z basis, resulting in $\widehat{g'} = [b, 0, a]$. Connecting $\widehat{g'}$ to $\hat{f}$ via $[0,1,0]$, we get $\widehat{g_1} = [a^2, a, 0, b, b^2]$. Hence $\det(\widetilde{M_{\widehat{g_1}}}) = -4a^2b^2 = -64 \neq 0$. Then by Corollary 2.29 Pl-Holant(g_1) is #P-hard, and so is Pl-Holant(f).

At last, $n = 5$ and $ab = 2-n = -3$. We also have $\hat{g} = [ab+4, b, a, ab+4] = [1, b, a, 1]$. One more self loop on g via $[0,1,0]$ in the Z basis results in $\widehat{g''} = [b, a]$. Connecting $\widehat{g''}$ to $\hat{f}$ via $[0,1,0]$, we get $\widehat{g_2} = [a^2+b, a, 0, b, b^2+a]$. Hence $\det(\widetilde{M_{\widehat{g_2}}}) = -2(a^3 + 2a^2b^2 + b^3) = -2(a^3+b^3+18)$. If $a^3+b^3+18 \neq 0$, then we are done by Corollary 2.29. Otherwise $a^3+b^3 = -18$, and we construct a binary signature $[a,0,b]$ by doing a self-loop on $\widehat{g_2}$ in Z basis. Then we construct another unary signature by connecting $\widehat{g''} = [b,a]$ to $[a,0,b]$ via $[0,1,0]$, which gives $[a^2, b^2]$. Connecting $[a^2, b^2]$ to $\hat{f}$ via $[0,1,0]$, we have another arity-4 signature $\widehat{g_3} = [ab^2+a^2, b^2, 0, a^2, a^2b+b^2]$. We compute $\det(\widetilde{M_{\widehat{g_3}}}) = -2(a^6 + a^5b^2 + a^2b^5 + b^6) = -2(a^6+b^6-162)$. If $a^6+b^6-162 \neq 0$, again we are done by Corollary 2.29. Otherwise $a^6+b^6 = 162$. Together with $a^3+b^3 = -18$ and $ab = -3$, there is no solution of a and b. This finishes the proof. □

This lemma essentially handles the case of $f' \in \mathscr{P}_2$ due to the following corollary.

Corollary 6.4. *Suppose f be a non-degenerate signature of arity $n \geq 5$. Let f' be f with a self loop. If $f' \in \mathscr{P}_2$ is non-degenerate, then* Pl-Holant(f) *is* #P*-hard.*

Proof. Since $f' \in \mathscr{P}_2$, we have that $f' = Z^{\otimes n-2}[1, 0, \ldots, 0, 1]$ up to an orthogonal transformation H. Since H does not change the complexity, we may assume we are under this transformation. Then f is of the form $Z^{\otimes n}[a, 1, 0, \ldots, 0, 1, b]$. The claim follows by Lemma 6.3. □

The next lemma handles the case when f' is a non-degenerate vanishing signature. Its proof is partly contained in the proof of Theorem 9.1 in [6]. We include this part here for completeness. As we shall see, the case of $f' \in \mathscr{M}_4$ is a special case of this result.

Lemma 6.5. *Suppose f is a non-degenerate signature of arity $n \geq 5$. Let f' be f with a self loop. If f' is non-degenerate and vanishing, then* Pl-Holant(f) *is* #P*-hard unless $\{f, f'\}$ is vanishing, in which case* Pl-Holant(f) *is tractable.*

Proof. Since f' is vanishing, $f' \in \mathscr{V}^{\sigma}$ for some $\sigma \in \{+, -\}$ by Theorem 2.15. For simplicity, assume that $f' \in \mathscr{V}^{+}$. The other case is similar. Let $\mathrm{rd}^{+}(f') = d - 1$, where $2d < n$ and $d \geq 2$ since f' is non-degenerate. Then the entries of f' can be expressed as

$$f'_k = i^k q(k),$$

where $q(x)$ is a polynomial of degree exactly $d - 1$. However, notice that if f' satisfies some recurrence relation with characteristic polynomial $t(x)$, then f satisfies a recurrence relation with characteristic polynomial $(x^2 + 1)t(x)$. In this case, $t(x) = (x - i)^d$. Then the corresponding characteristic polynomial of f is $(x - i)^{d+1}(x + i)$, and thus the entries of f are

$$f_k = i^k p(k) + c(-i)^k$$

for some constant c and a polynomial $p(x)$ of degree at most d. However, the degree of $p(x)$ is exactly d, otherwise the polynomial $q(x)$ for f' would have degree less than $d - 1$. If $c = 0$, then $\{f, f'\}$ is vanishing, the tractable case. Now assume $c \neq 0$, and we want to show that Pl-Holant(f) is #P-hard.

Thus, under the transformation $Z = \frac{1}{\sqrt{2}} \left[\begin{smallmatrix} 1 & 1 \\ i & -i \end{smallmatrix}\right]$, we have

$$\begin{aligned} \text{Pl-Holant}\ (=_2 \mid f) &\equiv_T \text{Pl-Holant}\left([1, 0, 1]Z^{\otimes 2} \mid (Z^{-1})^{\otimes n} f\right) \\ &\equiv_T \text{Pl-Holant}\left([0, 1, 0] \mid \hat{f}\right), \end{aligned}$$

where $\hat{f} = [\hat{f}_0, \hat{f}_1, \ldots, \hat{f}_d, 0, \ldots, 0, c]$, with $\hat{f}_d \neq 0$. Taking a self loop in the original setting is equivalent to connecting $[0, 1, 0]$ to a signature after this transformation. Thus, doing this once on $\hat{f}$, we get $\widehat{f'} = [\hat{f}_1, \ldots, \hat{f}_d, 0, \ldots, 0]$, corresponding to f' transformed, and doing this $d - 2$ times on $\hat{f}$, we get a signature $\hat{h} = [\hat{f}_{d-2}, \hat{f}_{d-1}, \hat{f}_d, 0, \ldots, 0, 0/c]$ of arity $n - 2(d - 2) = n - 2d + 4$. The last entry is c when $d = 2$ and is 0 when $d > 2$. As $n > 2d$, we may do two more self loops and get $[\hat{f}_d, 0, \ldots, 0]$ of arity $k = n - 2d$. Now connect this signature back to $\hat{f}$ via $[0, 1, 0]$. It is the same as getting the last $n - k + 1 = 2d + 1$ signature entries of $\hat{f}$ up to a nonzero scalar. We may repeat this operation zero or more times until the arity k' of the resulting signature is less than or equal to k. We claim that this signature has the form $\hat{g} = [0, \ldots, 0, c]$. In other words, the $k' + 1$ entries of $\hat{g}$ consist of the last c and k' many 0's from the signature $\hat{f}$, all appearing after $\hat{f}_d$. This is because there are $n - d - 1$ many 0 entries in the signature $\hat{f}$ after $\hat{f}_d$, and $n - d - 1 \geq k \geq k'$.

Having both $[\hat{f}_d, 0, \ldots, 0]$ of arity k and $\hat{g} = [0, \ldots, 0, c]$ of arity k' in the Z basis is equivalent to having both $[1, i]^{\otimes k}$ and $[1, -i]^{\otimes k'}$ in the standard basis. If $k > k'$, then we can connect $[1, -i]^{\otimes k'}$ to $[1, i]^{\otimes k}$ and get $[1, i]^{\otimes (k-k')}$. Replacing k by $k - k'$, we can repeat this process until the new $k \leq k'$. If the new $k < k'$, then we can continue as in the subtractive Euclid algorithm. We continue this procedure and eventually we get $[1, i]^{\otimes t}$ and $[1, -i]^{\otimes t}$, where $t = \gcd(k, k')$ Now putting k/t many copies of $[1, -i]^{\otimes t}$ together, we get $[1, -i]^{\otimes k}$.

In the transformed setting, $[1, -i]^{\otimes k}$ is $[0, \ldots, 0, 1]$ of arity k. Then we connect this back to $\hat{h}$ via $[0, 1, 0]$. Doing this is the same as forcing k connected edges of $\hat{h}$ be assigned 0, because $[0, 1, 0]$ flips $[0, \ldots, 0, 1]$. Thus we get a signature of arity $n - 2d + 4 - k = 4$, which is $[\hat{f}_{d-2}, \hat{f}_{d-1}, \hat{f}_d, 0, 0]$. Note that the last entry is 0 (and not c), because $k \geq 1$ and $\text{arity}(\hat{h}) \geq 5$.

However, Pl-Holant$([0, 1, 0] \,|\, [\hat{f}_{d-2}, \hat{f}_{d-1}, \hat{f}_d, 0, 0])$ is equivalent to Pl-Holant$([0, 1, 0] \,|\, [0, 0, 1, 0, 0])$ when $\hat{f}_d \neq 0$, which is transformed back by Z to Pl-Holant$([3, 0, 1, 0, 3])$. This is the Eulerian Orientation problem on planar 4-regular graphs and is #P-hard by Theorem 2.22. □

6.2 Lemmas applied to Degenerate Signatures in the Inductive Step

Lemma 6.2 does not solve the case when f' is degenerate. In general, when f' is degenerate, the inductive step is straightforward unless f' is also vanishing. Lemma 6.6 and 6.8 are the two missing pieces to this end.

Lemma 6.6. *Let $a, b \in \mathbb{C}$. Suppose f is a signature of the form $\left[\begin{smallmatrix} 1 & 1 \\ i & -i \end{smallmatrix}\right]^{\otimes n} [a, 1, 0, \ldots, 0, b]$ with arity $n \geq 3$. If $ab \neq 0$, then* Pl-Holant(f) *is #P-hard.*

Proof. We prove by induction on n. For $n = 3$ or 4, it follows from Lemma 3.12 and Theorem 2.22 that Pl-Holant(f) is #P-hard.

Now assume $n \geq 5$. Under a holographic transformation by $Z = \left[\begin{smallmatrix} 1 & 1 \\ i & -i \end{smallmatrix}\right]$, we have

$$\begin{aligned} \text{Pl-Holant}\,(=_2 \mid f) &\equiv_T \text{Pl-Holant}\left([1, 0, 1]Z^{\otimes 2} \mid (Z^{-1})^{\otimes n} f\right) \\ &\equiv_T \text{Pl-Holant}\left([0, 1, 0] \mid \hat{f}\right), \end{aligned}$$

where $\hat{f} = [a, 1, 0, \ldots, 0, b]$. Now consider the gadget in Figure 7a with $\hat{f}$ assigned to both vertices. This gadget has the binary signature $\widehat{g_1} = [0, ab, 2b]$, which is equivalent to $[0, a, 2]$ since $b \neq 0$. Translating back by Z to the original setting, this signature is $g_1 = [a + 1, -i, a - 1]$. This can be verified as

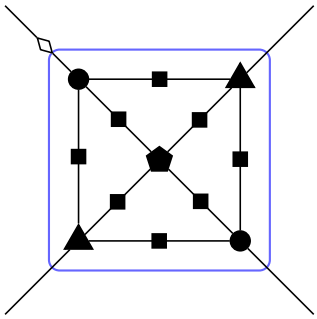
$$\begin{bmatrix} 1 & 1 \\ i & -i \end{bmatrix} \begin{bmatrix} 0 & a \\ a & 2 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ i & -i \end{bmatrix}^{\mathsf{T}} = 2 \begin{bmatrix} a+1 & -i \\ -i & a-1 \end{bmatrix}.$$

By the form of $\widehat{g_1} = [0, ab, 2b]$ and $b \neq 0$, it follows from Lemma 2.19 that $g_1 \notin \mathscr{R}_2^+$. Moreover, since $a \neq 0$, g_1 is non-degenerate.

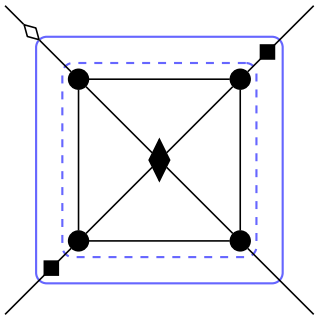
Doing a self loop on f yields $f' = Z^{\otimes n-2}[1, 0, \ldots, 0]$. Connecting f' back to f, we get a binary signature $g_2 = Z^{\otimes 2}[0, 0, b]$. Once again we connect g_2 to f, the resulting signature is $h = Z^{\otimes n-2}[a, 1, 0, \ldots, 0]$ of arity $n - 2 \geq 3$ up to the constant factor of $b \neq 0$.

Notice that h is non-degenerate and $h \in \mathscr{V}^+$. By Lemma 4.3, Pl-Holant(h, g_1) is #P-hard, hence Pl-Holant(f) is also #P-hard. □

The next case uses the following technical lemma. It is also applied more than once in Section 7.

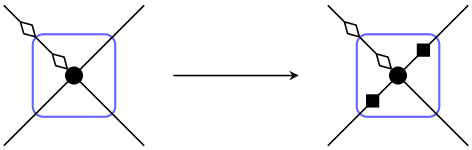


(a) $(\neq_2 \mid [0,1,0,0,0],[0,0,0,1,0],\hat{g})$-gate on right side

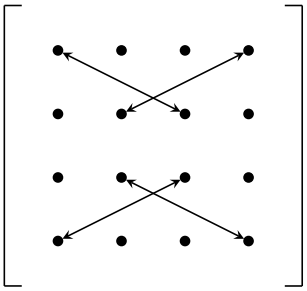


(b) Simpler construction with the same signature

Figure 8: Two gadgets with the same signature used in Lemma 6.7.

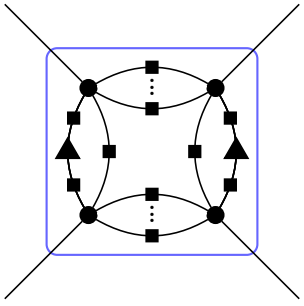


(a) Negating the second and fourth inputs

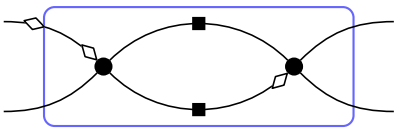


(b) Movement of even Hamming weight entries

Figure 9: The movement of the even Hamming weight entries in the signature matrix of a quaternary signature under the negation of the second and fourth inputs (i.e. the square vertices are assigned $[0,1,0]$).



(a) Gadget that realizes a partial crossover



(b) Gadget with a useful signature matrix

Figure 10: Two quaternary gadgets used in the proof of Lemma 6.7 and 6.8.

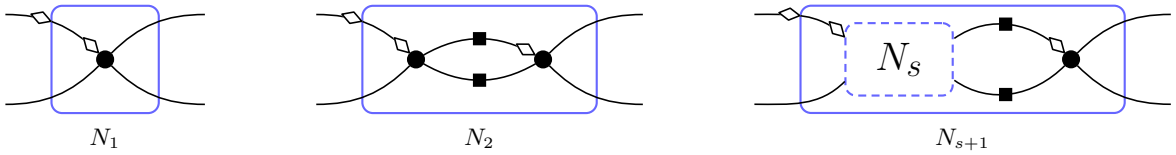


Figure 11: Linear recursive construction used for interpolation in a nonstandard basis.

Lemma 6.7. *Let $\hat{g}$ be the arity 4 signature whose matrix is*

$$M_{\hat{g}} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}. \tag{6.7}$$

Then Pl-Holant $(\neq_2 \mid [0,1,0,0,0], [0,0,0,1,0], \hat{g})$ *is* #P*-hard.*

Proof. Consider the gadget in Figure 8a. We assign $[0,0,0,1,0]$ to the triangle vertices, $[0,1,0,0,0]$ to the circle vertices, $\hat{g}$ to the pentagon vertex, and $[0,1,0]$ to the square vertices. Let $\hat{h}$ be the signature of this gadget. By adding two more disequality signatures and then grouping appropriately, it is clear that the gadget in Figure 8b has the same signature of the gadget in Figure 8a, where the circle vertices are still assigned $[0,1,0,0,0]$, the square vertices are still assigned $[0,1,0]$, and the diamond vertex is assigned the quaternary equality signature. To compute the signature $\hat{h}$, first compute the signature $\hat{h}'$ of the inner gadget enclosed by the dashed line, which has signature matrix

$$M_{\hat{h}'} = \begin{bmatrix} 3 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}. \quad \text{Then by Figure 9, the signature matrix of } \hat{h} \text{ is } M_{\hat{h}} = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 3 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}.$$

One more gadget before we finish the proof using interpolation. Consider the gadget in Figure 10b. We assign $\hat{h}$ to the circle vertices and $[0,1,0]$ to the square vertices. The signature of the resulting gadget is $\hat{r}$ with signature matrix $M_{\hat{r}}$ (see Figure 2 for the signature of a rotated copy of $\hat{h}$ that appears as the second circle vertex in Figure 10b), where

$$M_{\hat{r}} = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 3 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \left(\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \otimes \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \right) \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 3 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 6 & 4 & 0 \\ 0 & 4 & 2 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}.$$

Consider an instance Ω of Pl-Holant $(\neq_2 \mid \mathcal{F} \cup \{\hat{r}'\})$ with $\hat{r} \in \mathcal{F}$, where the signature matrix of $\hat{r}'$ is

$$M_{\hat{r}'} = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 3 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}.$$

Suppose that $\hat{r}'$ appears n times in Ω. We construct from Ω a sequence of instances Ω_s of Pl-Holant $(\neq_2 \mid \mathcal{F})$ indexed by $s \geq 1$. We obtain Ω_s from Ω by replacing each occurrence of $\hat{r}'$

with the gadget N_s in Figure 11 with $\hat{r}$ assigned to the circle vertices and $[0,1,0]$ assigned to the square vertices. In Ω_s, the edge corresponding to the ith significant index bit of N_s connects to the same location as the edge corresponding to the ith significant index bit of $\hat{r}'$ in Ω.

We can express the signature matrix of N_s as

$$M_{N_s} = X(XM_{\hat{r}})^s = XP\,\mathrm{diag}\left(1, 4+2\sqrt{3}, 4-2\sqrt{3}, 1\right)^s P^{-1},$$

where

$$X = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \quad \text{and} \quad P = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & \sqrt{3} & -\sqrt{3} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Since $M_{\hat{r}'} = XP\,\mathrm{diag}\left(1, 1+\sqrt{3}, 1-\sqrt{3}, 1\right)P^{-1}$, we can view our construction of Ω_s as first replacing $M_{\hat{r}'}$ with $XP\,\mathrm{diag}\left(1, 1+\sqrt{3}, 1-\sqrt{3}, 1\right)P^{-1}$, which does not change the Holant value, and then replacing the diagonal matrix with the diagonal matrix $\mathrm{diag}\left(1, 4+2\sqrt{3}, 4-2\sqrt{3}, 1\right)^s$.

We stratify the assignments in Ω based on the assignments to the n occurrences of the signature whose signature matrix is the diagonal matrix

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1+\sqrt{3} & 0 & 0 \\ 0 & 0 & 1-\sqrt{3} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{6.8}$$

We only need to consider the assignments that assign

- i many times the bit patterns 0000 or 1111,
- j many times the bit pattern 0110, and
- k many times the bit pattern 1001,

since any other assignment contributes a factor of 0. Let c_{ijk} be the sum over all such assignments of the products of evaluations of all signatures (including the signatures corresponding to the signature matrices X, P, and P^{-1}) in Ω except for signature corresponding to the signature matrix in (6.8). Then

$$\mathrm{Holant}_{\Omega} = \sum_{i+j+k=n} \left(1+\sqrt{3}\right)^j \left(1-\sqrt{3}\right)^k c_{ijk}$$

and the value of the Holant on Ω_s, for $s \geq 1$, is

$$\mathrm{Holant}_{\Omega_s} = \sum_{i+j+k=n} \left(\left(4+2\sqrt{3}\right)^j \left(4-2\sqrt{3}\right)^k\right)^s c_{ijk} = \sum_{i+j+k=n} \left(\left(4+2\sqrt{3}\right)^{j-k} 4^k\right)^s c_{ijk}.$$

We argue that this Vandermonde system has full rank, which is to say that $\left(4+2\sqrt{3}\right)^{j-k} 4^k \neq \left(4+2\sqrt{3}\right)^{j'-k'} 4^{k'}$ unless $(j,k) = (j',k')$. If $\left(4+2\sqrt{3}\right)^{j-k} 4^k = \left(4+2\sqrt{3}\right)^{j'-k'} 4^{k'}$, then we have $\left(4+2\sqrt{3}\right)^{j-k-(j'-k')} 4^{k-k'} = 1$. Since any nonzero integer power of $4+2\sqrt{3}$ is not rational, we must have $j-k = j'-k'$. And in this case, $4^{k-k'} = 1$, and hence $k = k'$ and $j = j'$.

Therefore, we can solve for the unknown c_{ijk}'s and obtain the value of Holant_{Ω}. Then after a counterclockwise rotation of $\hat{r}'$ (c.f. Figure 2), we are done by Corollary 2.29. □

With Lemma 6.7 at hand, we continue to prove Lemma 6.8.

Lemma 6.8. *Let $b \in \mathbb{C}$. Suppose f is a signature of the form $\left[\begin{smallmatrix} 1 & 1 \\ i & -i \end{smallmatrix}\right]^{\otimes n} [0, 1, 0, \ldots, 0, b]$ with arity $n \geq 4$. If $b \neq 0$, then* Pl-Holant(f) *is* #P*-hard.*

Remark 2. For $n = 3$, $Z^{\otimes 3}[0, 1, 0, b]$ is tractable, as it is $\mathscr{M}$-transformable.

Proof. If $n = 4$, then we are done by Corollary 2.29. Thus, assume that $n \geq 5$.

Under a holographic transformation by $Z = \left[\begin{smallmatrix} 1 & 1 \\ i & -i \end{smallmatrix}\right]$, we have

$$\begin{aligned} \text{Pl-Holant}\left(=_2 \mid f\right) &\equiv_T \text{Pl-Holant}\left([1, 0, 1]Z^{\otimes 2} \mid (Z^{-1})^{\otimes n} f\right) \\ &\equiv_T \text{Pl-Holant}\left([0, 1, 0] \mid \hat{f}\right), \end{aligned}$$

where $\hat{f} = [0, 1, 0, \ldots, 0, b]$. We show how to construct the following three signatures: $[0, 0, 0, 1, 0]$, $[0, 1, 0, 0, 0]$, and $\hat{g}$, where $\hat{g}$ is defined by (6.7). Then we are done by Lemma 6.7.

Consider the gadget in Figure 7b. We assign $\hat{f}$ to the circle vertices and $[0, 1, 0]$ to the square vertices. The signature of the resulting gadget is $[0, 0, 0, 1, 0]$ up to a nonzero factor of b.

Taking a $[0, 1, 0]$ self loop on $[0, 0, 0, 1, 0]$ gives $[0, 0, 1] = [0, 1]^{\otimes 2}$. We connect this back to $\hat{f}$ through $[0, 1, 0]$ until the arity of the resulting signature is either 4 or 5, depending on the parity of n. If n is even, then we have $[0, 1, 0, 0, 0]$ as desired. Otherwise, n is odd and we have $[0, 1, 0, 0, 0, b/0]$, where the last entry is b if $n = 5$ and 0 if $n > 5$. Connection $[0, 1]^{\otimes 2}$ through $[0, 1, 0]$ to $\hat{f}$ twice more gives $[0, 1]$. We connect this through $[0, 1, 0]$ to $[0, 1, 0, 0, 0, b/0]$ to get $[0, 1, 0, 0, 0]$ as desired.

Taking a $[0, 1, 0]$ self loop on $[0, 1, 0, 0, 0]$ gives $[1, 0, 0] = [1, 0]^{\otimes 2}$. Now consider the gadget in Figure 10a. We assign $\hat{f}$ to the circle vertices, $[1, 0]^{\otimes 2}$ to the triangle vertices, and $[0, 1, 0]$ to the square vertices. Up to a factor of b^2, the signature of the resulting gadget is $\hat{g}$ with signature matrix $M_{\hat{g}}$ given in (6.7). To see this, first replace the two copies of the signatures $[1, 0]^{\otimes 2}$ assigned to the triangle vertices with two copies of $[1, 0]$ each. Then notice that $\hat{f}$ simplifies to a weighted equality signature when connected to $[1, 0]$ through $[0, 1, 0]$. □

6.3 Proof of the Single Signature Dichotomy

Now we are ready to prove the dichotomy for a single signature. Recall that $\mathscr{M}_1 \subset \mathscr{A}_1 \subset \mathscr{P}_1$ and $\mathscr{A}_2 = \mathscr{P}_2 \subset \mathscr{M}_2$. Thus $f \in \mathscr{P}_1 \cup \mathscr{M}_2 \cup \mathscr{A}_3 \cup \mathscr{M}_3 \cup \mathscr{M}_4$ if and only if f is $\mathscr{A}$-, $\mathscr{P}$-, or $\mathscr{M}$-transformable by Lemma 3.3, Lemma 3.5, or Lemma 3.10.

Proof of Theorem 6.1. The proof is by induction on n. The base cases of $n = 3$ and $n = 4$ are proved in Theorem 2.22. Now assume $n \geq 5$.

With the signature f, we form a self loop to get a signature f' of arity at least 3. In general we use prime to denote the signature with a self loop. We consider separately whether or not f' is degenerate.

- Suppose $f' = [a, b]^{\otimes (n-2)}$ is degenerate. Then there are three cases to consider.
  1. If $a = b = 0$, then f' is the all zero signature. For f, this means $f_{k+2} = -f_k$ for $0 \leq k \leq n - 2$, so $f \in \mathscr{P}_2$ by Lemma 3.2, and therefore Pl-Holant(f) is tractable.
  2. If $a^2 + b^2 \neq 0$, then f' is nonzero and $[a, b]$ is not a constant multiple of either $[1, i]$ or $[1, -i]$. We may normalize so that $a^2 + b^2 = 1$. Then the orthogonal transformation $\left[\begin{smallmatrix} a & b \\ -b & a \end{smallmatrix}\right]$ transforms the column vector $[a, b]$ to $[1, 0]$. Let $\hat{f}$ be the transformed signature from f, and $\widehat{f'} = [1, 0]^{\otimes (n-2)}$ the transformed signature from f'.

Since an orthogonal transformation keeps $=_2$ invariant, this transformation commutes with the operation of taking a self loop, i.e., $\widehat{f'} = (\hat{f})'$. Here $(\hat{f})'$ is the function obtained from $\hat{f}$ by taking a self loop. As $(\hat{f})' = [1,0]^{\otimes(n-2)}$, we have $\hat{f}_0 + \hat{f}_2 = 1$ and for every integer $1 \le k \le n-2$, we have $\hat{f}_k = -\hat{f}_{k+2}$. With one or more self loops on $(\hat{f})'$, we eventually obtain either $[1,0]$ when n is odd or $[1,0,0]$ when n is even. In either case, we connect $[1,0]$ or $[1,0,0]$ to $\hat{f}$ until we get an arity 4 signature, which is $\hat{g} = [\hat{f}_0, \hat{f}_1, \hat{f}_2, -\hat{f}_1, -\hat{f}_2]$. This is possible because that the parity matches and the arity of $\hat{f}$ is at least 5. We show that Pl-Holant$(\hat{g})$ is #P-hard. To see this, we first compute $\det(\widetilde{M_g}) = -2(\hat{f}_0 + \hat{f}_2)(\hat{f}_1^2 + \hat{f}_2^2) = -2(\hat{f}_1^2 + \hat{f}_2^2)$, since $\hat{f}_0 + \hat{f}_2 = 1$. Therefore if $\hat{f}_1^2 + \hat{f}_2^2 \neq 0$, Pl-Holant$(\hat{g})$ is #P-hard by Lemma 2.28. Otherwise $\hat{f}_1^2 + \hat{f}_2^2 = 0$, and we assume $\hat{f}_2 = i\hat{f}_1$ since the other case is similar. Since f is non-degenerate, $\hat{f}$ is non-degenerate, which implies $\hat{f}_2 \neq 0$. We can rewrite $\hat{g}$ as $[1,0]^{\otimes 4} - \hat{f}_2[1,i]^{\otimes 4}$. Under the holographic transformation by $T = \begin{bmatrix} 1 & (-\hat{f}_2)^{1/4} \\ 0 & i(-\hat{f}_2)^{1/4} \end{bmatrix}$, we have

$$\begin{aligned} \text{Pl-Holant}\,(=_2 \mid \hat{g}) &\equiv_T \text{Pl-Holant}\left([1,0,1]T^{\otimes 2} \mid (T^{-1})^{\otimes 4}\hat{g}\right) \\ &\equiv_T \text{Pl-Holant}\left(\hat{h} \mid =_4\right), \end{aligned}$$

where

$$\hat{h} = [1,0,1]T^{\otimes 2} = [1, (-\hat{f}_2)^{1/4}, 0]$$

and $\hat{g}$ is transformed by T^{-1} into the arity 4 equality $=_4$, since

$$T^{\otimes 4}\left(\begin{bmatrix}1\\0\end{bmatrix}^{\otimes 4} + \begin{bmatrix}0\\1\end{bmatrix}^{\otimes 4}\right) = \begin{bmatrix}1\\0\end{bmatrix}^{\otimes 4} - \hat{f}_2\begin{bmatrix}1\\i\end{bmatrix}^{\otimes 4} = \hat{g}.$$

By Theorem 2.24, Pl-Holant$\left(\hat{h} \mid =_4\right)$ is #P-hard as $\hat{f}_2 \neq 0$.

3. If $a^2 + b^2 = 0$ but $(a,b) \neq (0,0)$, then $[a,b]$ is a nonzero multiple of $[1, \pm i]$. Ignoring the constant multiple, we have $f' = [1,i]^{\otimes(n-2)}$ or $[1,-i]^{\otimes(n-2)}$. We consider the first case since the other case is similar.
In the first case, the characteristic polynomial of the recurrence relation of f' is $x - i$, so that of f is $(x-i)(x^2+1) = (x-i)^2(x+i)$. Hence there exist a_0, a_1, and c such that

$$f_k = (a_0 + a_1 k)i^k + c(-i)^k$$

for every integer $0 \le k \le n$. Let f^+ and f^- be two signatures of arity n such that $f_k^+ = (a_0 + a_1 k)i^k$ and $f_k^- = c(-i)^k$ for every $0 \le k \le n$. Hence $f_k = f_k^+ + f_k^-$ and we write $f = f^+ + f^-$. If $a_1 = 0$, then f' is the all zero signature, a contradiction. If $c = 0$, then f is vanishing, one of the tractable cases. Now we assume $a_1 c \neq 0$ and show that Pl-Holant(f) is #P-hard. Hence $\mathrm{rd}^+(f^+) = 1$ and $\mathrm{rd}^-(f^-) = 0$. Under the holographic transformation $Z = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ i & -i \end{bmatrix}$, we have

$$\begin{aligned} \text{Pl-Holant}\,(=_2 \mid f) &\equiv_T \text{Pl-Holant}\left([1,0,1]Z^{\otimes 2} \mid (Z^{-1})^{\otimes n} f\right) \\ &\equiv_T \text{Pl-Holant}\left([0,1,0] \mid \hat{f}\right), \end{aligned}$$

where $\hat{f}$ takes the form $[\hat{f}_0, \hat{f}_1, 0, \ldots, 0, c']$ with $c' = 2^{n/2}c \neq 0$ and $\hat{f}_1 \neq 0$, since $\hat{f}$ is the Z^{-1}-transformation of the sum of f^+ and f^-, with $\mathrm{rd}^+(f^+) = 1$ and $\mathrm{rd}^-(f^-) = 0$

respectively. On the other side, $(=_2) = [1, 0, 1]$ is transformed into $(\neq_2) = [0, 1, 0]$. Depending on whether $\hat{f}_0 = 0$ or not, we apply Lemma 6.8 or Lemma 6.6 and Pl-Holant(f) is #P-hard.

- Suppose f' is non-degenerate. By inductive hypothesis, Pl-Holant(f) is #P-hard, unless $f' \in \mathscr{P}_1 \cup \mathscr{M}_2 \cup \mathscr{A}_3 \cup \mathscr{M}_3 \cup \mathscr{M}_4 \cup \mathscr{V}$. Note that f' has arity $n - 2 \geq 3$, and every signature in $\mathscr{M}_4$ of arity at least 3 is also in $\mathscr{V}$. Hence the exceptional case is equivalent to $f' \in \mathscr{P}_1 \cup \mathscr{M}_2 \cup \mathscr{A}_3 \cup \mathscr{M}_3 \cup \mathscr{V}$. In this case, we apply Lemma 6.2 to f' and f. Hence Pl-Holant(f) is #P-hard, unless $f \in \mathscr{P}_1 \cup \mathscr{M}_2 \cup \mathscr{A}_3 \cup \mathscr{M}_3 \cup \mathscr{V}$. The exceptional cases imply that f is $\mathscr{A}$- or $\mathscr{P}$- or $\mathscr{M}$-transformable or vanishing, and Pl-Holant(f) is tractable. □

7 Mixing $\mathscr{P}_2$ and $\mathscr{M}_4$—Equalities and Matchgates in the Z Basis

Given a set $\mathcal{F}$ of symmetric signatures, by Theorem 6.1, Pl-Holant$(\mathcal{F})$ is #P-hard unless every single non-degenerate signature f of arity at least 3 in $\mathcal{F}$ is in $\mathscr{P}_1 \cup \mathscr{M}_2 \cup \mathscr{A}_3 \cup \mathscr{M}_3 \cup \mathscr{M}_4 \cup \mathscr{V}$. We have already proved that the desired full dichotomy holds if $\mathcal{F}$ contains such an f in $\mathscr{P}_1$, $\mathscr{A}_3$, $\mathscr{M}_2 \setminus \mathscr{P}_2$, or $\mathscr{M}_3$ due to Corollary 5.4, Corollary 5.6, Lemma 5.8, or Lemma 5.10, respectively.

The remaining cases are when all non-degenerate signatures of arity at least 3 in $\mathcal{F}$ are contained in $\mathscr{P}_2 \cup \mathscr{M}_4 \cup \mathscr{V}$. In this section, we consider the mixing of $\mathscr{P}_2$ and $\mathscr{M}_4$. For this, we do a holographic transformation by Z. Then the problem becomes Pl-Holant $(\neq_2 \mid =_k, \text{EXACTONE}_d)$ with various arities k and d. Recall that EXACTONE_d denotes the exact one function $[0, 1, 0, \ldots, 0]$ of arity d. These are the signatures for PERFECT MATCHING and they are the basic components of *Matchgates*.

A big surprise, against the putative form of a complexity classification for planar counting problems, is that we found the complexity of Pl-Holant $(\neq_2 \mid =_k, \text{EXACTONE}_d)$ depends on the values of d and k, and the problem is tractable for all large k. This result has the consequence that, for the first time since Kasteleyn's algorithm, we have discovered some new *primitive* tractable family of counting problems on planar graphs. These problems *cannot* be captured by a holographic reduction to Kasteleyn's algorithm, or any other known algorithm. Thus for planar problems the paradigm of holographic algorithms using matchgates (i.e., being $\mathscr{M}$-transformable) *is not universal*.

Let $\mathcal{EO} = \{\text{EXACTONE}_d \mid d \geq 3\}$.

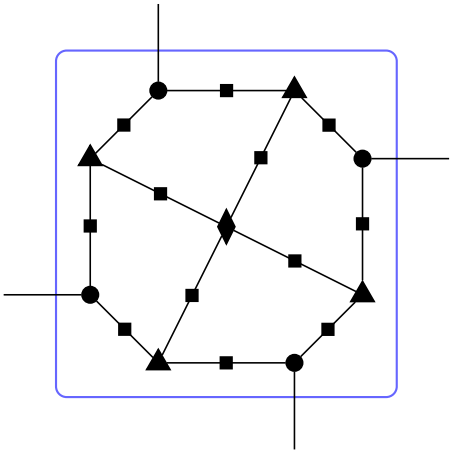
7.1 Hardness when $k = 3$ or 4

We begin with some hardness results.

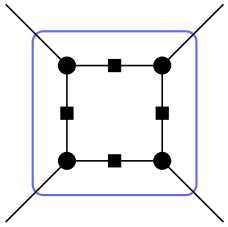
Lemma 7.1. Pl-Holant $(\neq_2 \mid =_3, [0, 1, 0, 0])$ *is* #P*-hard.*

Proof. By connecting two copies of $[0, 1, 0, 0]$ together via $\neq_2$, we have $[0, 1, 0, 0, 0]$ on the right. Consider the gadget in Figure 12a. We assign $=_3$ to the triangle vertices, $[0, 1, 0, 0]$ to the circle vertices, $\neq_2$ to the square vertices, and $[0, 1, 0, 0, 0]$ on the diamond vertex in the middle. Let f be the signature of this gadget.

We claim that the support of f is $\{0011, 0110, 1100, 1001\}$. To see this, notice that $[0, 1, 0, 0, 0]$ in the middle must match exactly one of the half edges, which forces the corresponding equality signature to take the value 0 and all other equality signatures to take value 1. The two $[0, 1, 0, 0]$'s adjacent to the equality assigned 0 must have 0 going out, and the other two $[0, 1, 0, 0]$'s have 1 going out.



(a) Cycle-like gadget used twice



(b) Gadget to realize $\hat{g}$

Figure 12: Two gadgets used in the proof of Lemma 7.1.

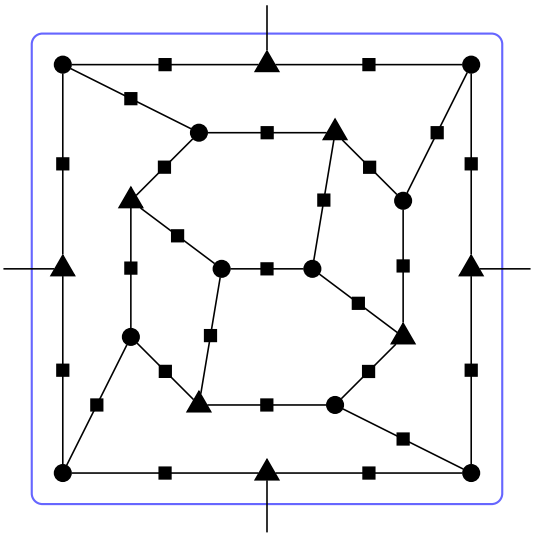
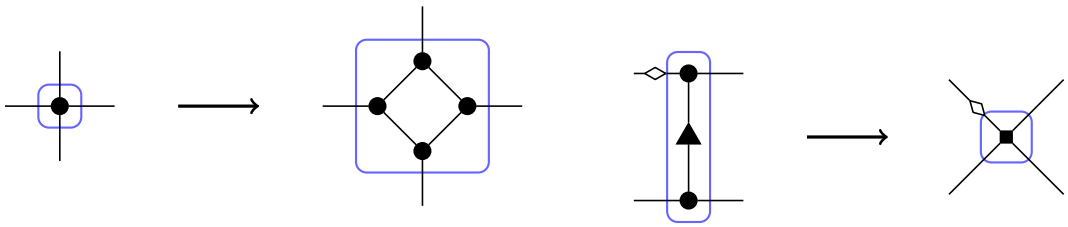


Figure 13: The whole gadget to realize $[0, 0, 0, 1, 0]$.



(a) Step one: Degree 4 vertex example

(b) Step two: Contract edges

Figure 14: A reduction from Pl-Holant $(\mathcal{EQ} \mid h)$ to Pl-Holant(g) for any binary signature h and a quaternary signature g that depends on h. The circle vertices are assigned $=_4$ or $=_3$ respectively, the triangle vertex is assigned h, and the square vertex is assigned the signature of the gadget to its left.

Now we consider the gadget in Figure 12a again. This time we place $[0,1,0,0]$ on each triangle, $=_3$ on each circle, f on the middle diamond, and again $\neq_2$ on each square. Now notice that each support of f makes two $[0,1,0,0]$'s that are cyclically adjacent on the outer cycle to become $[0,1,0]$ and the other two $[1,0,0]$. It is easy to see that the support of the resulting signature is $\{0111, 1011, 1101, 1110\}$. Therefore it is the reversed ExactOne$_4$ signature $[0,0,0,1,0]$ (namely AllButOne$_4$). The whole gadget is illustrated in Figure 13, where each circle is assigned $[0,1,0,0]$, triangle $=_3$, and square $\neq_2$.

Finally, we build the gadget in Figure 12b. We place $=_3$ on each circle and $\neq_2$ on each square. It is easy to see that there are only two support vectors of the resulting signature, which are 0101 and 1010. Recall the definition (6.7) of the partial crossover $\hat{g}$. This gadget realizes exactly $\hat{g}$.

By Lemma 6.7, Pl-Holant $(\neq_2 \mid [0,1,0,0,0], [0,0,0,1,0], \hat{g})$ is #P-hard. We have constructed $[0,1,0,0,0]$, $[0,0,0,1,0]$, and $\hat{g}$ on the right side. Therefore Pl-Holant $(\neq_2 \mid =_3, [0,1,0,0])$ is #P-hard. □

For $k = 4$, we need the following lemma.

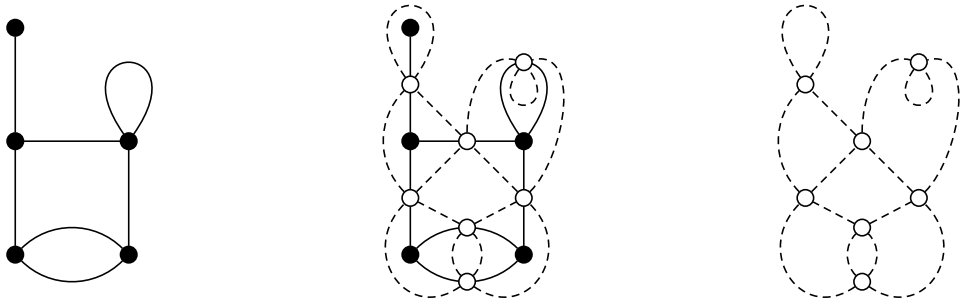
Lemma 7.2. *Let g be the arity 4 signature whose matrix is*

$$M_g = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Then Pl-Holant(g) *is* #P*-hard.*

Proof. Let $h = [2,1,1]$. We show that Pl-#CSP$(h) \leq_T$ Pl-Holant(g) in two steps. In each step, we begin with a signature grid and end with a new signature grid such that the Holants of both signature grids are the same. Then we are done by Theorem 2.25. Or more explicitly, since Pl-#CSP$(h) \equiv$ Pl-Holant $(\mathcal{EQ} \mid h)$ by (2.2), we are done by Theorem 2.24.

For step one, let $G = (U, V, E)$ be an instance of Pl-Holant $(\mathcal{EQ} \mid h)$. Fix an embedding of G in the plane. This defines a cyclic ordering of the edges incident to each vertex. Consider a vertex $u \in U$ of degree k. It is assigned the signature $=_k$. We decompose u into k vertices. Then we connect the k edges originally incident to u to these k new vertices so that each vertex is incident to exactly one edge. We also connect these k new vertices in a cycle according to the cyclic ordering



(a) (b) (c)

Figure 15: A plane graph (a), its medial graph (c), and both graphs superimposed (b).

induced on them by their incident edges. Each of these vertices has degree 3, and we assign them $=_3$. Clearly the Holant value is unchanged. This completes step one. An example of this step applied to a vertex of degree 4 is given in Figure 14a. The resulting graph has the following properties: (1) it is planar; (2) every vertex is either degree 2 (in V and assigned h) or degree 3 (newly created and assigned $=_3$); (3) each degree 2 vertex is connected to two degree 3 vertices; and (4) each degree 3 vertex is connected to one degree 2 vertex and two other degree 3 vertices.

Now step two. For every $v \in V$, v has degree 2. We contract the two edges incident to v, or equivalently, we replace the two circle vertices and one triangle vertex boxed in Figure 14b with a single (square) vertex of degree 4. The resulting graph $G' = (V', E')$ is planar and 4-regular.

Next we determine what is the signature on $v' \in V'$ after this contraction. Clearly the two inputs to each original circle have to be the same. Therefore its support is $0000, 0110, 1001, 1111$, listed starting from the diamond and going counterclockwise. Moreover, due to the triangle assigned h in the middle, the weight on 0000 is 2, and every other weight is 1. Hence it is exactly the signature g, with the diamond in Figure 14b marking the first input bit. This finishes the proof. □

Remark 3. From the planar embedding of the graph G, treating h vertices as edges, the resulting graph G' is known as the medial graph of G. The (constructive) definition is usually phrased in the following way. The medial graph G_m of plane graph G has a vertex on each edge of G and two vertices in G_m are joined by an edge for each face of G in which their corresponding edges occur consecutively. See Figure 15 for an example. However, our construction described in the proof clearly extends to nonplanar graphs as well.

Lemma 7.3. Pl-Holant $(\neq_2 \mid =_4, [0, 1, 0, 0])$ *is* #P*-hard.*

Proof. Consider the gadget in Figure 16. We assign binary disequality $\neq_2$ to the square vertices, $=_4$ to the circle vertices, and $[0, 1, 0, 0]$ to the triangle vertices. We show that the support of the resulting signature is the set $\{00110011, 11001100, 11111111\}$, where each vector is the assignment ordered counterclockwise starting from the diamond point.

We call the equality signature $=_4$ in the middle the origin. There are two possible assignments at the origin. If it is assigned 0, then every adjacent perfect matching signature $[0, 1, 0, 0]$ is matched to the half edge towards the origin, and every equality $=_4$ is forced to be 1. This gives the support vector 11111111.

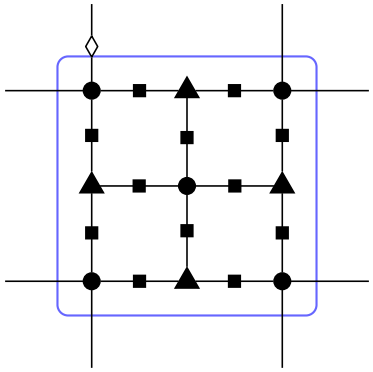
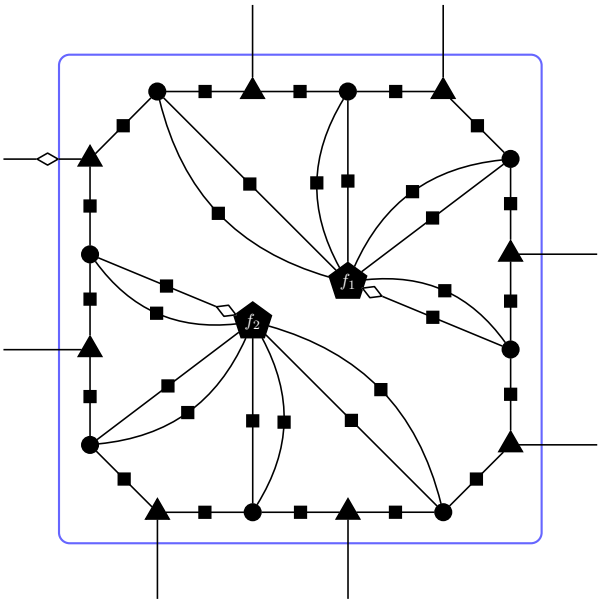


Figure 16: Grid-like gadget used in the proof of Lemma 7.3, whose support vectors are 00110011, 11001100, and 11111111. Each square is assigned a binary disequality $\neq_2$, circle $=_4$, and triangle $[0, 1, 0, 0]$.



(a) Gadget with signature g. Each square is assigned a binary disequality $\neq_2$, circle $=_4$, triangle $[0, 1, 0, 0]$, and pentagon f.

f_1	f_2	g
00000000	00000000	11111111
00110011	00000000	01111000
11001100	00000000	11110000
00000000	00110011	10000111
00110011	00110011	00000000
11001100	00110011	-
00000000	11001100	00001111
00110011	11001100	-
11001100	11001100	00000000

(b) Support of g. Each vector is an assignment ordered counterclockwise from the diamond.

Figure 17: Another gadget used in the proof of Lemma 7.3 and a Table listing the support of its signature.

The other possibility is that the origin is 1. In this case, we can remove the origin leaving the outer cycle, with every $[0,1,0,0]$ becoming $[0,1,0]$. This is effectively a cycle of four equalities connected by $\neq_2$. It is easy to see that there are only two support vectors, which are exactly 00110011 and 11001100.

Every pair of half edges at each corner always take the same value. We further connect each pair of these edges to different copy of $=_4$ via two copies of $\neq_2$. This results in a gadget with signature f whose support is the complement of the original support, that is, $\{11001100, 00110011, 00000000\}$.

Now consider the gadget in Figure 17a. We assign $\neq_2$ to the square vertices, $=_4$ to the circle vertices, $[0,1,0,0]$ to the triangle vertices, and f to the pentagon vertex. Notice that each pair of edges coming out of the pentagon vertex are from the same corner of the gadget in Figure 16 used to realize f. We now study the signature of this gadget.

Notice that if a $=_4$ on the outer cycle is assigned 0, then the two adjacent perfect matchings must match half edges toward that $=_4$, and their outgoing edges must be 0. Furthermore, the two $=_4$ one more step away must be 1. A further observation is that any pair of consecutive $=_4$'s cannot be both 0, and if a pair of consecutive $=_4$'s are both 1, then the $[0,1,0,0]$ in the middle must have a 1 going out. In Figure 17a, we call the pentagon connecting to four equalities $=_4$ on the upper right f_1 and the other one f_2. Let g be the signature of resulting gadget. We further order the external wires of f_1, f_2, and g counterclockwise, each starting from edge marked with a diamond. With this notation and these observations, we get Table 17b listing the support of g. The support of g is $\{11111111, 01111000, 11110000, 10000111, 00000000, 00001111, 00000000\}$, and 00000000 has multiplicity 2.

Next we use a domain pairing argument. First we move $=_4$ to the left hand side, by contracting four $\neq_2$ into it. We apply the domain pairing on the problem Pl-Holant $(=_4 \mid g)$. Specifically, we use $=_4$ as $=_2$, by pairing each pair of edges together. We also pair adjacent two outputs of g clockwise, starting from the diamond point. Each pair of output wires of g are connected to a pair of wires from $=_4$ on the left hand side. Note that $=_4$ enforces that each pair of edges always takes the same value. We re-interpret 00 or 11 as 0 or 1 in the Boolean domain. In this way, we can treat g as an arity 4 signature g' in the Boolean domain. So the reduction is

$$\text{Pl-Holant}\left(=_2 \mid g'\right) \leq_T \text{Pl-Holant}\left(=_4 \mid g\right).$$

We get the expression of g' next. The two support bit strings 01111000 and 10000111 of g are eliminated as they do not agree on adjacent paired outputs. So in the paired (Boolean) domain, the support of g' becomes $\{1111, 1100, 0011, 0000\}$ where 0000 has multiplicity 2. We further rotate g' as a Boolean domain signature such that the support is $\{1111, 0110, 1001, 0000\}$. Now it is easy to see that the matrix of g', an arity 4 signature in the Boolean domain, is

$$M_{g'} = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

By Lemma 7.2 Pl-Holant(g') is #P-hard. Hence Pl-Holant $(\neq_2 \mid =_4, [0,1,0,0])$ is #P-hard. □

To extend Lemma 7.1 and Lemma 7.3 to general ExactOne_d functions, we show that we can always realize constant functions $[1,0]$ and $[0,1]$ in this setting.

Lemma 7.4. *For any integer $k \geq 3$ and $d \geq 3$ and any signature set $\mathcal{F}$,*

$$\text{Pl-Holant}\left(\neq_2 \mid =_k, \text{EXACTONE}_d, [0,1], [1,0], \mathcal{F}\right) \leq_T \text{Pl-Holant}\left(\neq_2 \mid =_k, \text{EXACTONE}_d, \mathcal{F}\right).$$

Proof. Given an instance Ω of Pl-Holant $(\neq_2 \mid =_k, \text{EXACTONE}_d, [0,1], [1,0], \mathcal{F})$ with underlying planar graph G, if there is any $[1,0]$ on the right hand side, then it can be combined with $\neq_2$ as a $[0,1]$ on the left hand side, and then contracted into whatever function it is attached to. If it is connected to $[1,0]$ or $[0,1]$, we either know the Holant is 0 or remove the two vertices. If it is connected to EXACTONE_d, then the contraction gives us $d-1$ many $[1,0]$ pinnings. Similarly, if it is connected to $=_k$, the whole function decomposes into $k-1$ many $[0,1]$'s. These additional pinnings by $[1,0]$'s or $[0,1]$'s can be recursively applied.

By a similar analysis, it is easy to show that the only nontrivial occurrences of $[0,1]$'s are those attached to EXACTONE_d via $\neq_2$. We may therefore assume there is no $[1,0]$ in Ω, and the only appearances of $[0,1]$ and $[1,0]$'s are those of $[0,1]$'s applied to EXACTONE_d via $\neq_2$.

We can construct $=_{\ell k}$ for any integer $\ell \geq 1$, by $\neq_2$ on the left and $=_k$ on the right. In fact if we connect two copies of $=_k$ via $\neq_2$ we get a signature of arity $2k-2$ with $k-1$ consecutive external wires labeled $+$ and the others labeled $-$. As $k \geq 3$, we can take 2 wires of the $k-1$ wires labeled $-$ and attach to two copies of $=_k$ via two $\neq_2$. This creates a signature of arity $3(k-1)+(k-3)$ with $3(k-1)$ consecutive wires labeled $+$ and the other $k-3$ wires labeled $-$. Finally connect $k-3$ pairs of adjacent $+/-$ labeled wires by $\neq_2$ recursively. This creates a planar gadget with an equality signature of arity $3(k-1)-(k-3) = 2k$. This can be extended to any $=_{\ell k}$ by applying the same process on any consecutive k wires.

Next we construct $[0,1]^{\otimes r}$ for some integer $r \geq 1$. We get $[1,0]^{\otimes d-2}$ by a self-loop of EXACTONE_d via $\neq_2$, ignoring the factor 2. We pick an integer ℓ large enough so that $d-2 < \ell k$. Then we connect $[1,0]^{\otimes d-2}$ to $=_{\ell k}$ via $\neq_2$ to get $[0,1]^{\otimes(\ell k - d + 2)}$. This is what we claim with $r = \ell k - d + 2$.

One more construction we will use is $\text{EXACTONE}_{2+\ell(d-2)}$ for any integer $\ell \geq 1$. This is realizable by connecting ℓ many copies of EXACTONE_d sequentially via $\neq_2$.

Consider the dual graph G^* of G. Take a spanning tree T of G^*, with the external face as the root. In each face F, let c_F be the number of $[0,1]$'s in the face. We start from the leaves to recursively move all the pinnings of $[0,1]$ to the external face. Suppose we are working on the face F as a leaf of T. If $c_F = 0$ then we just remove the leaf from T and recurse on another leaf. Otherwise we remove all $[0,1]$'s in F. Let s be the smallest integer such that $sr \geq c_F$. We replace the $\neq_2$ edge bordering between F and its parent F' by a sequence of three signatures: $\neq_2$, $\text{EXACTONE}_{2+\ell(d-2)}$ and $\neq_2$, where ℓ is a sufficiently large integer such that $\ell(d-2) \geq sr - c_F$. From $\text{EXACTONE}_{2+\ell(d-2)}$ there are two edges connected to the two adjacent copies of $\neq_2$. Of the other $\ell(d-2)$ edges we will put $sr - c_F$ many dangling edges in F, and the remaining $\ell(d-2) - (sr - c_F)$ dangling edges in F'. Hence there are sr dangling edges in F, including those c_F many that were connected to $[0,1]$'s before we removed the $[0,1]$'s. We put s copies of $[0,1]^{\otimes r}$ inside the face F to pin all of them in a planar way. We add $\ell(d-2) - (sr - c_F)$ to $c_{F'}$. Remove the leaf F from T, and recurse.

After the process, all $[0,1]$'s are in the external face of G. Suppose the number is p. We put r disjoint copies of G together to form a planar signature grid. Apply a total of pr many $[0,1]$'s by p copies of $[0,1]^{\otimes r}$ in a planar way. This is now an instance of Pl-Holant $(\neq_2 \mid =_k, \text{EXACTONE}_d, \mathcal{F})$ and the Holant value is the rth power of that of Ω. Since the Holant value of Ω is a nonnegative integer, we can take the rth root and finish the reduction. □

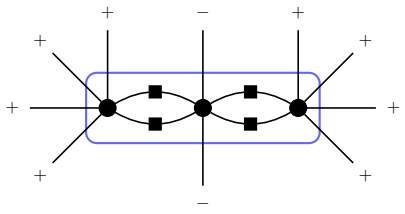


Figure 18: Example E_6-block. Circle vertices are assigned $=_6$ and square vertices are assigned $\neq_2$.

Once we have constant functions $[0,1]$ and $[1,0]$, it is easy to construct EXACTONE_3 from EXACTONE_d. Therefore combining Lemma 7.4 with Lemma 7.1 and Lemma 7.3 we get the following corollary.

Corollary 7.5. *If $d \geq 3$ and $k \in \{3,4\}$, then* Pl-Holant $(\neq_2 \mid =_k, \text{EXACTONE}_d)$ *is* #P*-hard.*

7.2 Tractability when $k \geq 5$

On the other hand, if the arity of the equality signature is at least 5, then the problem is tractable. In this subsection we will first prove that the problem Pl-Holant $(\neq_2 \mid =_k, \mathcal{EO})$ is tractable for $k \geq 6$. After that we will return to $=_5$.

To prove this, we first do some preprocessing. Let G be the underlying graph of an instance of Pl-Holant $(\neq_2 \mid =_k, \mathcal{EO})$. Any self loop on an EXACTONE_d by a $\neq_2$ changes it to a $[1,0]^{\otimes(d-2)}$ with factor 2. These pinning signatures can be applied recursively. Any $[1,0]$ is first transformed to $[0,1]$ via $\neq_2$ on LHS and then applied either to $=_k$ producing $[0,1]^{\otimes(k-1)}$, or to EXACTONE_d (for some d) producing $[1,0]^{\otimes(d-1)}$. Similarly, any $[0,1]$ is first transformed to $[1,0]$ via $\neq_2$ on LHS and then applied either to $=_k$ producing $[1,0]^{\otimes(k-1)}$, or to EXACTONE_d (for some d) producing EXACTONE_{d-1}. Note that if $d = 3$ then EXACTONE_{d-1} is just $\neq_2$ on RHS, which combined with its adjacent two copies of $\neq_2$ of LHS, is equivalent to a single $\neq_2$ of LHS. Moreover, whenever an EXACTONE_d and another EXACTONE_ℓ are connected by a $\neq_2$, we replace it by a single $\text{EXACTONE}_{d+\ell-2}$, shrinking the edge between (and remove the connecting $\neq_2$). On the other hand, consider a connected component made of $=_k$ and $\neq_2$. We call such a component an E_k-block. Notice that each E_k-block has either exactly two or zero support vectors. This depends on whether or not there exists a contradiction, which is formed by an odd cycle of $=_k$ connected by $\neq_2$. We say an E_k-block is trivial if it has no support. This is easy to check. The two support vectors of a nontrivial E_k-block are complements of each other. We mark dangling edges of a nontrivial E_k-block by "+" or "−" signs. Dangling edges marked with the same sign take the same value on both support vectors while dangling edges marked with different signs take opposite values on both support vectors. Let $n_\pm$ be the number of dangling edges marked $\pm$. Then it is easy to see that

$$n_+ \equiv n_- \mod k. \tag{7.9}$$

An example of E_6-block is illustrated in Figure 18, with 8 + signs and 2 − signs.

After contracting all edges between EXACTONE_d's and forming E_k-block's we obtain a bipartite graph connected between EXACTONE_d's and E_k-block's by edges labeled by $=_2$.

A key observation is that a planar (bipartite) graph cannot be simple, i.e., it must have parallel edges, if its degrees are large.

(a) Two different arity 4 E_k-blocks. (b) Replace them by parallel $\neq_2$'s.

Figure 19: Arity 4 E_k-blocks.

Lemma 7.6. *Let $G = (L \cup R, E)$ be a planar bipartite graph with parts L and R. If every vertex in L has degree at least 6 and every vertex in R has degree at least 3, then G is not simple.*

Proof. Suppose G is simple. Let v, e and f be the total number of vertices, edges, and faces, respectively. Let v_i be the number of vertices of degree i in L, where $i \geq 6$, and u_j be the number of vertices of degree j in R, where $j \geq 3$. Since G is simple and bipartite, each face has at least 4 edges. Thus,

$$2e \geq 4f. \tag{7.10}$$

Furthermore, it is easy to see that

$$v = \sum_{i \geq 6} v_i + \sum_{j \geq 3} u_j \qquad \text{and} \qquad e = \sum_{i \geq 6} i v_i = \sum_{j \geq 3} j u_j. \tag{7.11}$$

Then starting from Euler's characteristic equation for planar graphs, we have

$$\begin{aligned} 2 &= v - e + f \\ &\leq v - \frac{e}{2} && \text{(By (7.10))} \\ &= \sum_{i \geq 6} v_i + \sum_{j \geq 3} u_j - \frac{1}{6} \sum_{i \geq 6} i v_i - \frac{1}{3} \sum_{j \geq 3} j u_j && \text{(By (7.11))} \\ &= \sum_{i \geq 6} \frac{6 - i}{6} v_i + \sum_{j \geq 3} \frac{3 - j}{3} u_j \leq 0, \end{aligned}$$

a contradiction. □

Lemma 7.6 does not give us tractability for the case of $k \geq 6$ yet. The reason is that given an instance of Pl-Holant $(\neq_2 \mid =_k, \mathcal{EO})$, after the preprocessing and forming E_k-blocks to make the graph bipartite, it is possible to have E_k-blocks of arity less than 6, in which case Lemma 7.6 does not apply. However, for $k \geq 6$ and a nontrivial E_k-block of arity n where $n < 6$, by (7.9) and the fact that $0 \leq n_+, n_- \leq n < k$, we see that $n_+ = n_-$, and $n = n_+ + n_-$ must be even. Moreover, if $n = 2$, then this means that the E_k-block is just $\neq_2$, in which case we can replace it by a single $\neq_2$ connecting signatures from $\mathcal{EO}$ to produce a new ExactOne signature. The only problematic case is when $n = 4$. We identify two possibilities of such E_k-blocks up to a rotation in Figure 19a.

Formally we define a *contraction* process on the connected graph of E_k-block with dangling edges. Recursively, for any non-dangling non-loop edge e, we shrink it to a point, maintaining planarity. The local cyclic orders of incident edges of the two vertices of e are spliced along e to form the cyclic order of the new vertex. For any loop we simply remove it. This contraction

process ends in a single point with a cyclic order of the dangling edges. Figure 19a depicts the two possibilities of E_k-blocks of arity 4 up to a rotation. An E_k-block of arity 4 can be viewed as a pair of $\neq_2$ in parallel, but there is a correlation between them, namely their support vectors are paired up in a unique way. If we replace the contracted E_k-block of arity 4 by two parallel edges as indicated in Fig 19b, one can revert back to a planar realization in the E_k-block as it connects to the rest of the graph. This can be seen by reversing the contraction process step by step.

We will show in the following lemma how to replace E_k-block of arity 4 by some other signatures while keeping track of the Holant value. We also observe that this tractable set is compatible with binary $\neq_2$ and unary $[1,0]$ or $[0,1]$ signatures.

Lemma 7.7. *For any integer $k \geq 6$,* Pl-Holant $(\neq_2 \mid =_k, \mathcal{EO}, \neq_2, [1,0], [0,1])$ *is tractable.*

Proof. Let Ω be an instance of Pl-Holant $(\neq_2 \mid =_k, \mathcal{EO}, \neq_2, [1,0], [0,1])$. Without loss of generality, we assume that Ω is connected. Any occurrence of $\neq_2$ of the right hand side can be removed as follows: It is connected to two adjacent copies of $\neq_2$ of the left hand side. We replace these 3 copies of $\neq_2$ by a single $\neq_2$ from the left hand side.

The given signatures have no weight, however the proof below can be adapted to the weighted case. For the unweighted case, we only need to count the number of satisfying assignments. We call an edge pinned if it has the same value in all satisfying assignments, *if there is any.* Clearly any edge incident to a vertex assigned $[1,0]$ or $[0,1]$ is pinned.

When an edge is pinned to a known value, we can get a smaller instance of the problem Pl-Holant$(\neq_2 \mid =_k, \mathcal{EO}, \neq_2, [1,0], [0,1])$ without changing the number of satisfying assignments. In our algorithm we may also find a contradiction and simply return 0. If e is a pinned edge, then it is adjacent to another edge e' via $\neq_2$ on the left hand side, and both e and e' are pinned. We remove e, e', and $\neq_2$, and perform the following on e (and on e' as well). If the other endpoint of e is $u = [1,0]$ or $[0,1]$ we either remove that u if the pinned value on e is consistent with u, or return 0 otherwise. If the other endpoint of e is $=_k$, then all edges of this $=_k$ are pinned to the same value which we can recursively apply. If the other endpoint of e is $\textsc{ExactOne}_d \in \mathcal{EO}$, then we replace this signature by $\textsc{ExactOne}_{d-1}$ when the pinned value is 0; or if the pinned value is 1 then the remaining $d-1$ edges of this $\textsc{ExactOne}_d$ are pinned to 0 which we recursively apply. Notice that we may create an $\textsc{ExactOne}_2$ (i.e. $\neq_2$) on the right hand side when we pin 0 on $\textsc{ExactOne}_3$. Such $\neq_2$'s are replaced as described at the beginning. It is easy to see that all these procedures do not change the number of satisfying assignments, and work in polynomial time.

We claim that there always exists an edge in Ω that is pinned, unless Ω does not contain $=_k$, or does not contain $\textsc{ExactOne}_d$ functions (for some $d \geq 3$), or there is a contradiction. Furthermore if there are $=_k$ or $\textsc{ExactOne}_d$ functions (for some $d \geq 3$), in polynomial time we can find a pinned edge with a known value, or return that there is a contradiction. (If there is a contradiction in Ω, we may still return a purported pinned edge with a known value, which we can apply and simplify Ω. The contradiction will eventually be found.) If Ω does not contain $=_k$, or does not contain $\textsc{ExactOne}_d$ functions (for some $d \geq 3$), then the problem is tractable, since Ω is an instance of $\mathscr{M}$, or an instance of $\mathscr{P}$. The lemma follows from the claim, for we either recurse on a smaller instance or have a tractable instance.

Suppose Ω is an instance where at least one $=_k$ and at least one $\textsc{ExactOne}_d \in \mathcal{EO}$ appear. We assume no $\neq_2$ appears on the right hand side. If any $[1,0]$ or $[1,0]$ appear, then we have found a pinned edge with a known value. Hence we may assume neither $[1,0]$ nor $[1,0]$ appears in Ω.

If a signature $\textsc{ExactOne}_d \in \mathcal{EO}$ is connected to itself by a self-loop through a $\neq_2$, then there are two choices for the assignment on this pair of edges through the $\neq_2$, but the remaining $d-2 \geq 1$

edges are pinned to 0. We can keep track of the factor 2 and have found a pinned edge with a known value. Thus we may assume there are no self-loops via $\neq_2$ on ExactOne signatures.

Next we consider the case that two separate signatures ExactOne_d and ExactOne_ℓ from $\mathcal{EO}$ are connected by some number of $\neq_2$'s. Depending on the number of connecting edges, there are three cases:

1. The connection is by a single $\neq_2$. We contract the connecting edge, maintaining planarity, and replace these three signatures by an $\text{ExactOne}_{d+\ell-2}$ to get a new instance Ω'. If an edge is pinned in Ω' then it is also pinned in Ω to the same value. We continue with Ω'.
2. The connection is by two $\neq_2$'s. There are two choices for the assignment on these two pairs of edges through $\neq_2$, but the remaining $d + \ell - 4 \geq 2$ edges are pinned to 0.
3. The connection is by at least three $\neq_2$'s. The three $\neq_2$'s cannot be all satisfied, so there is no satisfying assignment, a contradiction. We return the value 0.

Hence, we may assume there is no connection via any number of $\neq_2$'s among ExactOne signatures.

Define an E_k-block as a connected component composed of $=_k$ and $\neq_2$. All external connecting edges of each E_k-block are marked with $+$ or $-$ and this can be found by testing bipartiteness of a E_k-block where we treat $\neq_2$'s as edges. If any E_k-block is not bipartite, we return 0. We contract all E_k-blocks and maintain planarity. For each E_k-block we contract two vertices that are connected by an edge, one edge at a time, and remove self loops in this contraction process. If a trivial E_k-block appears, then there is no satisfying assignment, we return 0. Thus we may assume all E_k-blocks are nontrivial. If there is a nontrivial E_k-block of arity 2, as discussed earlier, its signature is $\neq_2$. We replace it with an edge labeled by $\neq_2$ to form an instance Ω', maintaining planarity, such that any pinned edge in Ω' corresponds to a pinned edge in Ω. This new edge is between ExactOne signatures and can be dealt with as described earlier. So we may assume the arity of any E_k-block is at least 4. Since $k \geq 6$, the only possible E_k-blocks of arity 4 are those in Figure 19a up to a rotation. Since there is at least one ExactOne_d signature with $d \geq 3$, forming E_k-blocks does not consume all of Ω.

After these steps we may consider Ω a bipartite graph, with one side consisting of E_k-blocks and the other side ExactOne signatures. And they are now connected by edges labeled by $=_2$.

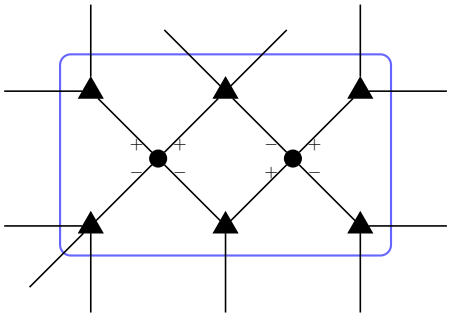
Suppose there are parallel edges between an E_k-block and an ExactOne_d signature. We show that this always leads to some pinned edges. If two parallel edges are marked by the same sign in the E_k-block, then they must be pinned to 0. If they are marked by different signs, then the remaining $d - 2 \geq 1$ edges of the ExactOne_d signature must be pinned to 0. Therefore, we may assume there are no parallel edges between any E_k-block and any ExactOne signature.

The next thing we do is to consider E_k-blocks of arity 4 with ExactOne signatures together. Call a connected component consisting of E_k-blocks of arity 4 and ExactOne an *EO-Eq*-4-block. Figure 20a illustrates an example. Notice that the two possibilities of E_k-blocks of arity 4 can be viewed as two parallel $\neq_2$'s but with some correlation between them. This is illustrated in Figure 19b. Note that the two dotted lines in Figure 19b represent different correlations.

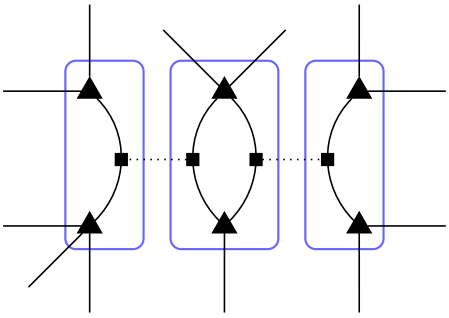
At this point we would like to replace every arity 4 E_k-block by two parallel $\neq_2$'s. However this replacement destroys the equivalence of the Holant values, before and after.

The surprising move of this proof is that we shall do so anyway!

Suppose we ignore the correlation for the time being and replace every arity 4 E_k-block by two parallel $\neq_2$'s as in Figure 19b. This replacement produces a *planar* signature grid Ω_1. Every edge in Ω_1 corresponds to a unique edge in Ω. The set of satisfying assignments of Ω_1 is a superset of that of Ω. Moreover, if there is an edge pinned in Ω_1 to a known value, the corresponding edge is



(a) An *EO-Eq*-4-block. Triangles are assigned ExactOne signatures and circles are E_k-blocks of arity 4.



(b) Break the *EO-Eq*-4-block into three components. Squares are assigned $\neq_2$. The component in the middle contains a cycle, and hence is degenerate. The other two are equivalent to ExactOne signatures.

Figure 20: *EO-Eq*-4-blocks

also pinned in Ω to the same value. Once we find that in Ω_1 we revert back to work in Ω and apply the pinning to the pinned edge.

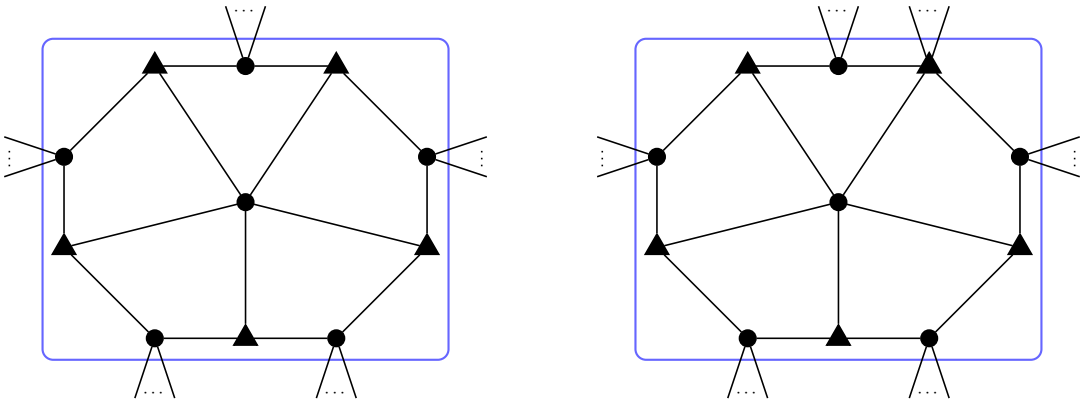
All that remains to be shown is that pinning always happens in Ω_1. Each *EO-Eq*-4-block splits into some number of connected components in Ω_1. If any component contains a cycle (which must alternate between $\neq_2$, which are the newly created ones from the E_k-blocks of arity 4, and ExactOne_d signatures for $d \geq 3$), then any edges not in the cycle but incident to some vertex in the cycle is pinned to 0. Moreover such edges must exist, for ExactOne_d signatures in the cycle are of arity at least 3. Note that the cycle has even length, and there are exactly two satisfying assignments, which assign exactly one 0 and one 1 to the two cycle edges incident to each ExactOne_d signature. This produces pinned edges.

Hence we may assume there are no cycles in these components, and every such component forms a tree, whose vertices are ExactOne functions and edges are $\neq_2$'s. Suppose there are $n \geq 2$ vertices in such a tree. As discussed in item 1 above, the whole tree is an ExactOne_t function for some arity t. Since each vertex in the tree has degree at least 3, $t \geq 3n - 2(n-1) = n + 2 \geq 4$. We replace these components by ExactOne_t's.

Thus, each connected component in the graph underlying Ω_1 is a planar bipartite graph with E_k-blocks of arity at least 6 on one side and ExactOne_d signatures of arity at least 3 on the other. By Lemma 7.6, no component is simple, which means that there are parallel edges between some E_k-block and some ExactOne_d signature. As discussed earlier, there must exist some pinned edge, and we can find a pinned edge with a known value in polynomial time. This finishes the proof. □

Unlike the situation in Lemma 7.6, a planar $(5, 3)$-regular bipartite graph *can* be simple. However, we show that such graphs must have a special induced subgraph. We call this structure a "wheel", which is pictured in Figure 21. There is a vertex v of degree 5 in the middle, and all faces adjacent to this vertex are 4-gons (i.e. quadrilaterals). Moreover, at least four neighbors of v have degree 3. Depending on the degree of the fifth neighbor (whether it is 3 or not), we have two types of wheel, which are pictured in Figure 21a and Figure 21b.

Lemma 7.8. *Let $G = (L \cup R, E)$ be a planar bipartite graph with parts L and R. Every vertex in*



(a) Type 1 (b) Type 2

Figure 21: Two types of wheels. Each circle is an E_5-block and triangle an ExactOne signature.

L has degree at least 5 and every vertex in R has degree at least 3. If G is simple, then there exists one of the two wheel structures in Figure 21 in G.

Proof. Let $V = L \cup R$ be the set of vertices and let F be the set of faces. We assign a *score* s_v to each vertex $v \in V$. We will define s_v so that $\sum_{v \in V} s_v = |V| - |E| + |F| = 2 > 0$. The base score is $+1$ for each vertex, which accounts for $|V|$. For each k-gon face, we assign $\frac{1}{k}$ to each of its vertex. This accounts for $|F|$. As G is a bipartite and a simple graph, $k \geq 4$ and a score coming from a face to a vertex is at most $\frac{1}{4}$.

For $-|E|$, we separate two cases. For any edge if one of the two endpoints has degree 3, we give the degree 3 vertex a score of $-\frac{7}{12}$, and the other one $-\frac{5}{12}$. This is well defined because all degree 3 vertices are in R. If the endpoints are not of degree 3, we give each endpoint $-\frac{1}{2}$. This accounts for $-|E|$.

Now we claim that $s_v \leq 0$ unless $v \in L$ and has degree 5. Suppose $v \in L$ and has degree $d \geq 6$, then

$$s_v \leq 1 + \frac{d}{4} - \frac{5}{12}d = 1 - \frac{d}{6} \leq 0.$$

Now suppose $v \in R$ and v has degree $d \geq 4$. Then every edge adjacent to v gives a score $-\frac{1}{2}$. Hence,

$$s_v \leq 1 + \frac{d}{4} - \frac{1}{2}d = 1 - \frac{d}{4} \leq 0.$$

The remaining case is that $v \in R$ and v has degree 3. Then,

$$s_v \leq 1 + \frac{d}{4} - \frac{7}{12}d = 1 - \frac{d}{3} \leq 0.$$

The claim is proved.

Since the total score is positive, there must exist $v \in L$, v has degree 5 and $s_v > 0$. We then claim that there must exist such a v so that all adjacent faces are 4-gons. Suppose otherwise. Then

any such v is adjacent to at least one k-gon with $k \geq 6$. In this case,

$$s_v \leq 1 + \frac{1}{4} \cdot 4 + \frac{1}{6} - \frac{5}{12} \cdot 5 = \frac{1}{12}.$$

Moreover, if v is adjacent to more than one k-gon with $k \geq 6$, Then

$$s_v \leq 1 + \frac{1}{4} \cdot 3 + \frac{1}{6} \cdot 2 - \frac{5}{12} \cdot 5 = 0,$$

contrary to the assumption that $s_v > 0$. Hence v is adjacent to exactly one k-gon with $k \geq 6$. Call this face F_v.

In F_v, v has two neighbors in R. We match each vertex v that has a positive score to the vertex on F_v that is the next one in clockwise order from v. By bipartiteness, every such v is matched to a vertex in R. We do this matching in all faces containing at least one positively scored vertex. It is possible that more than one such v are matched to the same $u \in R$. Suppose a vertex $u \in R$ is matched to from ℓ different such vertices of positive score. This means that u is adjacent to at least ℓ many k-gons with $k \geq 6$. Then, if u has degree 3 then u has score

$$s_u \leq 1 + \frac{1}{4} \cdot (3 - \ell) + \frac{1}{6} \cdot \ell - \frac{7}{12} \cdot 3 = -\frac{\ell}{12}.$$

If u has degree $d \geq 4$ then u has score

$$s_u \leq 1 + \frac{1}{4} \cdot (d - \ell) + \frac{1}{6} \cdot \ell - \frac{1}{2} \cdot d \leq -\frac{\ell}{12}.$$

Hence in any case, we have $s_u \leq -\frac{\ell}{12}$. It implies that the total score of u and all positively scored vertices matched to u is at most 0. However each positively scored vertex is matched to a vertex in R. Hence the total score cannot be positive. This is a contradiction.

Therefore there exists $v \in L$ such that $s_v > 0$, and has degree 5, and all adjacent faces are 4-gons. We further note that at most one neighbor of v can have degree ≥ 4, for otherwise,

$$s_v \leq 1 + \frac{5}{4} - \frac{1}{2} \cdot 2 - \frac{5}{12} \cdot 3 = 0.$$

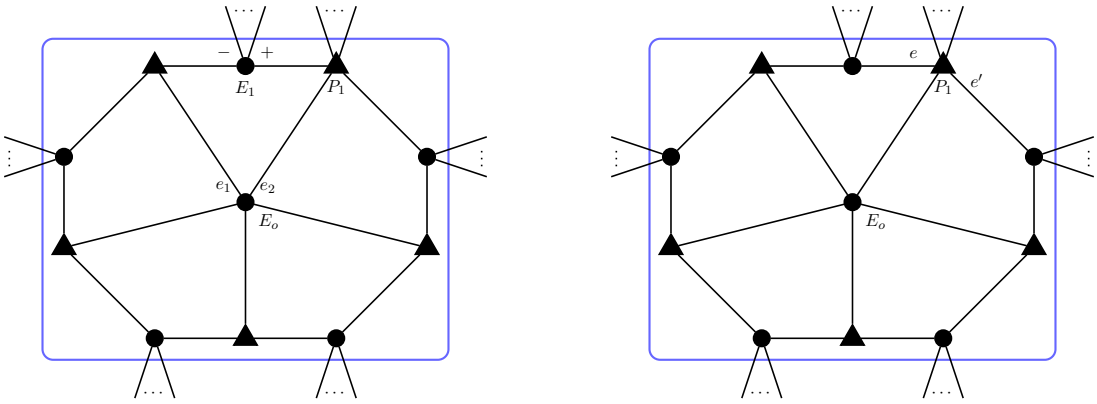
If all neighbors of v have degree 3, that is a wheel of type 1 as in Figure 21a. If one neighbor of v has degree ≥ 4, that is a wheel of type 2 as in Figure 21b. □

As we shall see, either structure in Figure 21 leads to pinned edges.

Lemma 7.9. Pl-Holant $(\neq_2 \mid =_5, \mathcal{EO}, \neq_2, [1, 0], [0, 1])$ *is tractable.*

Proof. We proceed as in Lemma 7.7 up until the point of getting Ω_1. Note that due to (7.9) the only nontrivial E_5-blocks of arity ≤ 4 are $\neq_2$ and those in Figure 19a. Moreover, each connected component of Ω_1 is planar and bipartite with vertices on one side having degree at least 5 and those on the other at least 3. We only need to show that there are edges pinned in Ω_1.

Unlike in Lemma 7.7, these components do not satisfy the condition of Lemma 7.6 but that of Lemma 7.8. If any such component is not simple, then there are pinned edges similar to Lemma 7.7. Otherwise by Lemma 7.8, the wheel structure in Figure 21 appears. All we need to show is that wheel structures of either type contain pinned edges.



(a) Different signs of an E_5-block along the cycle lead to pinning

(b) Edges e and e' are pinned in wheels of type 2

Figure 22: Degeneracies in the wheel structure.

First we claim that if a wheel of either type has a E_5-block, call it E_1, on the outer cycle which has different signs on the two edges incident to it along the cycle, then the middle $=_5$, denoted by E_o, is pinned. This is pictured in Figure 22a. It does not matter whether the wheel is type 1 or 2, or the position of E_1 relative to the special triangle P_1 in type 2. Because E_o is an equality, both e_1 and e_2, the two edges incident to E_o that are connected to the two ExactOne signatures flanking E_1, must take the same value. If both e_1 and e_2 are assigned 1, then the two incoming wires of E_1 along the cycle have to be both assigned 0, whereas they are marked by different signs. This is a contradiction. Hence both e_1 and e_2 are pinned to 0 as well as all edges of E_o.

We may therefore assume that each E_5-block has same signs along the outer cycle, either $++$ or $--$. If the wheel is of type 1, then there is no valid assignment such that E_o is assigned 0 because the cycle has odd length. In fact if E_o is assigned 0, then we can remove E_o and its incident edges, and effectively the five ExactOne signatures are now $\neq_2$'s forming a 5-cycle linked by binary equalities. Hence E_o and all its edges are pinned to 1.

Otherwise the wheel is of type 2, and each E_5-block has signs $++$ or $--$ along the outer cycle. We denote by P_1 the special ExactOne_d function that has arity $d > 3$. We claim that the two edges e and e' incident to P_1 along the cycle are both pinned to 0. This is illustrated in Figure 22b. As P_1 is ExactOne_d, at most one of e and e' is 1. If one of e and e' is 1, the other is 0, and as P_1 is an ExactOne_d function its edge to E_o is also 0, and thus all edges incident to E_o are 0. As all five neighbors of E_o are ExactOne functions, the four ExactOne_3 functions effectively become $(\neq_2)$ functions along the wheel, and we can remove E_o and its incident edges. This becomes the same situation as in the previous case of type 1, where effectively a cycle of five binary equalities are linked by five binary disequalities, which has no valid assignment. It implies that both e and e' are pinned to 0. This finishes the proof. □

7.3 Lemmas related to $\mathscr{M}_4$ and $\mathscr{P}_2$

Now we prove some lemmas relating to $\mathscr{M}_4$ and $\mathscr{P}_2$ that are used in the proof of the full dichotomy.

Recall that AllButOne_d is the signature $[0, \ldots, 0, 1, 0]$ of arity d, which is the reverse of ExactOne_d. After a Z transformation, $\mathscr{M}_4$ contains both AllButOne_d and ExactOne_d. How-

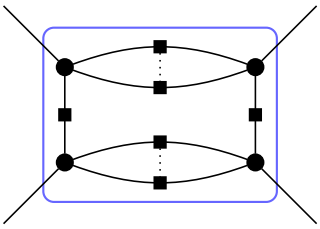


Figure 23: Gadget to realize $\hat{g}$ in Lemma 7.10. Circle vertices are assigned $=_k$ and square vertices are assigned $\neq_2$. The number of parallel edges is $k-2$.

ever, if both appear, then with any $=_k$ the problem is hard.

Lemma 7.10. *If integers $d_1, d_2, k \geq 3$, then* Pl-Holant $(\neq_2 \mid =_k, \text{EXACTONE}_{d_1}, \text{ALLBUTONE}_{d_2})$ *is* #P*-hard.*

Proof. We apply Lemma 7.4 to create constant functions $[1,0]$ and $[0,1]$ first. Then we construct EXACTONE_4 and ALLBUTONE_4. With both $[1,0]$ and $[0,1]$ in hand, we may reduce d_1 or d_2 to 4 if $d_1 > 4$ or $d_2 > 4$. If either of the two arities is 3, then we connect two copies together via $\neq_2$ to realize an arity 4 copy.

Moreover, we use the gadget illustrated in Figure 23 to create the function $\hat{g}$ in Lemma 6.7 as an E_k-block. Then by Lemma 6.7, Pl-Holant $(\neq_2 \mid =_k, \text{EXACTONE}_{d_1}, \text{ALLBUTONE}_{d_2})$ is #P-hard. □

In general signatures in $\mathscr{P}_2$ are non-degenerate weighted equalities under the Z transformation. The next several lemmas show that the hardness criterion is the same regardless of the weight.

Lemma 7.11. *Let $f \in \mathscr{P}_2$, $g_1 \in \mathscr{M}_4^+$, $g_2 \in \mathscr{M}_4^-$ be non-degenerate signatures with arity ≥ 3. Then* Pl-Holant(f, g_1, g_2) *is* #P*-hard.*

Proof. Suppose the arities of f, g_1, and g_2 are n, m_1, and m_2 respectively. Under a holographic transformation by Z, we have

$$\begin{aligned}\text{Pl-Holant}(f, g_1, g_2) &\equiv \text{Pl-Holant}\left(\neq_2 \mid \left(Z^{-1}\right)^{\otimes n} f, \left(Z^{-1}\right)^{\otimes m_1} g_1, \left(Z^{-1}\right)^{\otimes m_2} g_2\right) \\ &\equiv \text{Pl-Holant}\left(\neq_2 \mid \hat{f}, \text{EXACTONE}_{m_1}, \text{ALLBUTONE}_{m_2}\right),\end{aligned}$$

where $\hat{f} = (Z^{-1})^{\otimes n} f$ which has the form $[1, 0, \ldots, 0, c]$ up to a nonzero constant, with $c \neq 0$, as $f \in \mathscr{P}_2$. We do another diagonal transformation by $D = \left[\begin{smallmatrix} 1 & 0 \\ 0 & c^{1/n} \end{smallmatrix}\right]$. Then

$$\begin{aligned}&\text{Pl-Holant}(f, g_1, g_2) \\ &\quad\equiv \text{Pl-Holant}\left((\neq_2) D^{\otimes 2} \;\middle|\; (D^{-1})^{\otimes n} \hat{f}, (D^{-1})^{\otimes m_1} \text{EXACTONE}_{m_1}, (D^{-1})^{\otimes m_2} \text{ALLBUTONE}_{m_2}\right) \\ &\quad\equiv \text{Pl-Holant}\left(\neq_2 \mid =_n, \text{EXACTONE}_{m_1}, \text{ALLBUTONE}_{m_2}\right),\end{aligned}$$

where in the last line we ignored several nonzero factors. The lemma follows from Lemma 7.10. □

We also need to consider the mixture of $\mathscr{P}_2$ and binary signatures.

Lemma 7.12. *Let $\mathcal{F}$ be a set of symmetric signatures. Suppose $\mathcal{F}$ contains a non-degenerate signature $f \in \mathscr{P}_2$ of arity $n \geq 3$ and a binary signature h. Then* Pl-Holant$(\mathcal{F})$ *is* #P*-hard unless $h \in Z\mathscr{P}$, or* Pl-#CSP$^2(DZ^{-1}\mathcal{F}) \leq_T$ Pl-Holant$(\mathcal{F})$ *for some diagonal transformation D.*

Proof. We do a Z transformation and get

$$\begin{aligned}\text{Pl-Holant}(\mathcal{F}) &\equiv \text{Pl-Holant}(\mathcal{F}, h, f)\\ &\equiv \text{Pl-Holant}\left(\neq_2 \mid Z^{-1}\mathcal{F}, \left(Z^{-1}\right)^{\otimes 2} h, \hat{f}\right),\end{aligned}$$

where $\hat{f} = (Z^{-1})^{\otimes n} f = [1, 0, \ldots, 0, t]$ up to a nonzero constant with $t \neq 0$. We further do another diagonal transformation of $D_1 = \left[\begin{smallmatrix} 1 & 0 \\ 0 & t^{1/n} \end{smallmatrix}\right]$. Then

$$\begin{aligned}\text{Pl-Holant}(\mathcal{F}) &\equiv \text{Pl-Holant}\left((\neq_2) D_1^{\otimes 2} \mid (D_1^{-1})^{\otimes n} \hat{f}, (ZD_1)^{-1}\mathcal{F}, \left((ZD_1)^{-1}\right)^{\otimes 2} h\right)\\ &\equiv \text{Pl-Holant}\left(\neq_2 \mid =_n, (ZD_1)^{-1}\mathcal{F}, \left((ZD_1)^{-1}\right)^{\otimes 2} h\right)\\ &\geq_T \text{Pl-Holant}\left(=_n \mid (ZD_1)^{-1}\mathcal{F}, \left((ZD_1)^{-1}\right)^{\otimes 2} h\right),\end{aligned}$$

where in the second line we ignore a nonzero factor on $\neq_2$. Hence by Theorem 2.24, Pl-Holant$(\mathcal{F})$ is #P-hard unless $\left((ZD_1)^{-1}\right)^{\otimes 2} h \in \mathscr{P}$ (cases 1, 2 or 3 in Theorem 2.24) or $\left((ZD_1)^{-1}\right)^{\otimes 2} h = [a, b, c]$ for some $a, b, c \in \mathbb{C}$ such that $ac \neq 0$ and $(a/c)^{2n} = 1$ (cases 4 or 5 in Theorem 2.24).

If $\left((ZD_1)^{-1}\right)^{\otimes 2} h \in \mathscr{P}$, then $h \in ZD_1\mathscr{P} = Z\mathscr{P}$ as $D_1 \in \text{Stab}(\mathscr{P})$. In the latter case, we construct $=_{2n}$ on the right by connecting three copies of $=_n$ to one copy of $=_n$ via $\neq_2$. We do the same construction again to realize $=_{4n}$ using $=_{2n}$. We connect $n-1$ many $[a, b, c]$'s to $=_{2n}$ via $\neq_2$ to realize a binary weighted equality $[1, 0, r]$ with $r = (a/c)^{n-1} \neq 0$ ignoring a factor of c^{n-1}. Note that $r^{2n} = (a/c)^{2n(n-1)} = 1$. Then we do another diagonal transformation of $D_2 = \left[\begin{smallmatrix} 1 & 0 \\ 0 & r^{1/2} \end{smallmatrix}\right]$ to get Pl-Holant $\left(\neq_2 \mid (ZD_1D_2)^{-1}\mathcal{F}, =_2, \left(D_2^{-1}\right)^{\otimes 4n} (=_{4n})\right)$. Notice that

$$\left(D_2^{-1}\right)^{\otimes 4n} (=_{4n}) = [1, 0, \ldots, 0, r^{-2n}] = (=_{4n}),$$

as $r^{2n} = 1$.

Hence we have $=_2$ and $=_{4n}$ on the right. With $\neq_2$ on the left, we get $=_2$ on the left and therefore equalities of all even arities on the right. Let $D = (D_1D_2)^{-1}$. Then we have the reduction chain:

$$\begin{aligned}\text{Pl-Holant}(\mathcal{F}) &\geq_T \text{Pl-Holant}\left(\neq_2 \mid DZ^{-1}\mathcal{F} \cup \{=_2, =_{4n}\}\right)\\ &\geq_T \text{Pl-Holant}\left(\neq_2 \mid DZ^{-1}\mathcal{F} \cup \mathcal{EQ}_2\right)\\ &\geq_T \text{Pl-Holant}\left(\mathcal{EQ}_2 \mid DZ^{-1}\mathcal{F}\right).\end{aligned}$$

The last problem is Pl-#CSP$^2(DZ^{-1}\mathcal{F})$. Thus Pl-#CSP$^2(DZ^{-1}\mathcal{F}) \leq_T$ Pl-Holant$(\mathcal{F})$. □

At last, we strengthen Corollary 7.5, Lemma 7.7, and Lemma 7.9 to weighted equalities. We split the hardness and tractability cases. For a set $\mathcal{F}$ of signatures, denote by $\mathcal{F}_{nd}^{\geq 3}$ the set of non-degenerate signatures in $\mathcal{F}$ of arity at least 3. Moreover denote by $\mathcal{F}^*$ the signature set that is the same as $\mathcal{F}$ but with each degenerate signature $[a, b]^{\otimes m}$ in $\mathcal{F}$ replaced by the unary $[a, b]$.

Notice that $\mathcal{F} \cap \mathscr{P}_2$ and $\mathcal{F}^* \cap \mathscr{P}_2$ agree on signatures of arity at least 2, since signatures in $\mathscr{P}_2$ of arity at least 2 are non-degenerate. So $\mathcal{F} \cap \mathscr{P}_2 \subseteq \mathcal{F}^* \cap \mathscr{P}_2$, and the only possible extra elements

are some unary $[x, y]$'s from $[x, y]^{\otimes m} \in \mathcal{F}$ for some integer $m \geq 2$ and $[x, y]$ is not a multiple of $[1, \pm i]$. Equivalently the only possible extra elements are unary signatures of the form $Z[a, b]$ for $ab \neq 0$, i.e., *not* of the form a multiple of $Z[1, 0]$ or $Z[0, 1]$, when $Z^{-1}\mathcal{F}$ contains some degenerate signatures of the form $[a, b]^{\otimes m}$ for some integer $m \geq 2$ and $ab \neq 0$.

Lemma 7.13. *Let $\mathcal{F}$ be a set of symmetric signatures. Let $\mathcal{F}_{nd}^{\geq 3}$ be the set of non-degenerate signatures in $\mathcal{F}$ of arity at least 3. Suppose $\mathcal{F}_{nd}^{\geq 3}$ contains $f \in \mathcal{M}_4$ of arity $d \geq 3$. Moreover, suppose $\mathcal{F}_{nd}^{\geq 3} \cap \mathscr{P}_2$ is nonempty, and let k be the greatest common divisor of the arities of signatures in $\mathcal{F}^* \cap \mathscr{P}_2$. If $k \leq 4$, then* Pl-Holant$(\mathcal{F})$ *is* #P*-hard.*

Proof. We may assume that $f \in \mathcal{M}_4^+$. Since $\mathcal{F}_{nd}^{\geq 3} \cap \mathscr{P}_2$ is nonempty, there exists $g \in \mathcal{F}_{nd}^{\geq 3} \cap \mathscr{P}_2$. By the definition of $\mathcal{F}_{nd}^{\geq 3}$, g has arity $n \geq 3$. We do a Z transformation,

$$\text{Pl-Holant}(\mathcal{F}) \equiv \text{Pl-Holant}\left(\neq_2 \mid \hat{g}, \textsc{ExactOne}_d, Z^{-1}\mathcal{F}\right),$$

where $\hat{g} = (Z^{-1})^{\otimes n} g$ has the form $[1, 0, \ldots, 0, c]$ of arity n for some $c \neq 0$ up to a nonzero factor. We further do a diagonal transformation $D = \left[\begin{smallmatrix} 1 & 0 \\ 0 & c^{1/n} \end{smallmatrix}\right]$ and get

$$\text{Pl-Holant}(\mathcal{F}) \equiv \text{Pl-Holant}\left(\neq_2 \mid =_n, \textsc{ExactOne}_d, (ZD)^{-1}\mathcal{F}\right),$$

where we ignore nonzero factors on $\neq_2$ and $\textsc{ExactOne}_d$. Then by Lemma 7.4,

$$\text{Pl-Holant}(\mathcal{F}) \geq_T \text{Pl-Holant}\left(\neq_2 \mid =_n, \textsc{ExactOne}_d, [0, 1], [1, 0], (ZD)^{-1}\mathcal{F}\right).$$

By a weighted equality we mean a signature of the form $[a, 0, \ldots, 0, b]$ of some arity ≥ 1, where $ab \neq 0$. Recall that $\mathscr{P}_2$ consists of the Z transformation of all weighted equalities. Let $\mathcal{G}$ be the set of weighted equalities in $(ZD)^{-1}\mathcal{F}$. In other words, $\mathcal{G} = (ZD)^{-1}(\mathcal{F} \cap \mathscr{P}_2)$ as $(ZD)^{-1}\mathscr{P}_2$ contains all weighted equalities. Moreover, up to a nonzero factor, $(=_n) \in \mathcal{G}$.

Let k' be the gcd of all arities of signatures in $\mathcal{G}$, or equivalently the gcd of all arities of signatures in $\mathcal{F} \cap \mathscr{P}_2$. If $k' \neq k$, then the only possibility is that $(ZD)^{-1}\mathcal{F}$ contains a degenerate signature $[a, b]^{\otimes m}$ for some $m \geq 2$ with $ab \neq 0$. In this case we use pinnings $[1, 0]$ or $[0, 1]$ to realize $[a, b]$ from $[a, b]^{\otimes m}$ and put $[a, b]$ in $\mathcal{G}$. Hence we may assume that $k' = k$.

Pick any $g_1, g_2 \in \mathcal{G}$ of arities ℓ_1 and ℓ_2. Let $r = \gcd(\ell_1, \ell_2)$. Let t_1, t_2 be two positive integers such that $t_1\ell_1 - t_2\ell_2 = r$. Then connecting t_1 copies of g_1 and t_2 copies of g_2 via $\neq_2$ in a bipartite and planar way, we get a weighted equality signature of arity r.

Apply the same argument repeatedly. Eventually we construct a weighted equality h of arity k. We further do a diagonal transformation D_1 to make it $=_k$, that is,

$$\begin{aligned}
\text{Pl-Holant}(\mathcal{F}) &\geq_T \text{Pl-Holant}\left(\neq_2 \mid \mathcal{G}, \textsc{ExactOne}_d\right) \\
&\geq_T \text{Pl-Holant}\left(\neq_2 \mid h, \textsc{ExactOne}_d, \mathcal{G}\right) \\
&\geq_T \text{Pl-Holant}\left((\neq_2)D_1^{\otimes 2} \mid =_k, \left(D_1^{-1}\right)^{\otimes d} \textsc{ExactOne}_d, D_1^{-1}\mathcal{G}\right) \\
&\geq_T \text{Pl-Holant}\left(\neq_2 \mid =_k, \textsc{ExactOne}_d, D_1^{-1}\mathcal{G}\right),
\end{aligned}$$

where in the last line we ignored nonzero factors of $\textsc{ExactOne}_d$ and $\neq_2$. If $k = 3$ or 4, then the hardness follows from Corollary 7.5.

If $k = 1$ or 2, then on the right hand side we have $=_k$, which is $=_1$ or $=_2$, and a weighted equality $\left(D_1^{-1}\right)^{\otimes n}(=_n) \in D_1^{-1}\mathcal{G}$. Call it $\hat{g}'$. We move the $=_k$ to the left hand side via $\neq_2$. Then

we connect zero or more copies of this $=_k$, which is $=_1$ or $=_2$, to $\hat{g}'$ such that its arity is 3 or 4. It is possible that $n = 3$ or 4 to begin with, and if so we do nothing. We are done by yet another diagonal transformation and Corollary 7.5. □

Lemma 7.14. *Let $\mathcal{F}$ be a set of symmetric signatures. Suppose $\mathcal{F} \subseteq Z\mathscr{P} \cup \mathscr{M}_4^{\sigma}$ for some $\sigma \in \{+, -\}$ and the greatest common divisor of the arities of all signatures in $\mathcal{F}^* \cap \mathscr{P}_2$ is $k \geq 5$. Then* Pl-Holant$(\mathcal{F})$ *can be computed in polynomial time.*

Proof. We may assume that $\sigma = +$ and the case of $\sigma = -$ is similar. We do a Z transformation on Pl-Holant$(\mathcal{F})$, and get a problem of Pl-Holant $\left(\neq_2 \mid Z^{-1}\mathcal{F}\right)$.

In this bipartite setting, given $=_n$ on the right hand side, we can realize $=_{\ell n}$ for any integer $\ell \geq 1$ as an E_n-block on the right. The problem Pl-Holant $(\neq_2 \mid \mathcal{EQ}_n, \mathcal{EO}, \neq_2, [1, 0], [0, 1])$ is tractable for any $n \geq 5$ by Lemma 7.7 and Lemma 7.9, where $\mathcal{EQ}_n$ denotes the set of all equalities of arity ℓn for all integers $\ell \geq 1$.

The symmetric signatures in the set $Z\mathscr{P}$ consist of $\mathscr{P}_2$, $Z^{\otimes 2}(\neq_2)$, and degenerate signatures. If there is any degenerate signature of the form $(Z[a, b])^{\otimes m} \in \mathcal{F}$ with $ab \neq 0$, then $Z[a, b] \in \mathcal{F}^* \cap \mathscr{P}_2$. This contradicts $k \geq 5$. Hence all degenerate signatures in $\mathcal{F}$ are of the form $(Z[1, 0])^{\otimes m}$ or $(Z[0, 1])^{\otimes m}$, if any. Since $\mathcal{F} \subseteq Z\mathscr{P} \cup \mathscr{M}_4^{+}$, after a Z transformation, Pl-Holant$(\mathcal{F})$ is an instance of Pl-Holant $(\neq_2 \mid \mathcal{EQ}_k, \mathcal{EO}, \neq_2, [1, 0], [0, 1])$ except for the weights on the equalities. It can be checked that the tractability results of Lemma 7.7 and Lemma 7.9 also apply to weighted equalities. The lemma follows. □

Let $\mathcal{G} = \{=_k \mid k \in S\}$ be a set of Equality signatures, where S is a set of positive integers containing at least one $r \geq 3$. Moreover let $\mathcal{EO}^+ := \{\text{ExactOne}_d \mid d \in \mathbb{Z}^+\} = \mathcal{EO} \cup \{\neq_2, [0, 1]\}$. Then Pl-Holant $\left(\mathcal{G} \mid \mathcal{EO}^+\right)$ is the problem of counting perfect matchings over hypergraphs with planar incidence graphs, where the hyperedge sizes are prescribed by S. In the incidence graph, vertices assigned signatures in $\mathcal{G}$ on the left represent hyperedges, and vertices assigned signatures in $\mathcal{EO}^+$ on the right represent vertices of the hypergraph. Let $t = \gcd(S)$. It is stated in the introduction that this problem is tractable if $t \geq 5$ and #P-hard if $t \leq 4$. The tractability when $t \geq 5$ follows from Lemma 7.7 and 7.9, since we can reduce Pl-Holant $\left(\mathcal{G} \mid \mathcal{EO}^+\right)$ to Pl-Holant $(\neq_2 \mid =_t, \mathcal{EO}, \neq_2, [0, 1])$. The reduction goes as follows. With $\neq_2$ on the left hand side and $=_t$ on the right hand side, we can construct all E_t-blocks and hence all of $\mathcal{EQ}_t$ on the right. Note that $\mathcal{G} \subseteq \mathcal{EQ}_t$. Then we move all signatures in $\mathcal{G}$ to the left via $\neq_2$.

The hardness of Pl-Holant $\left(\mathcal{G} \mid \mathcal{EO}^+\right)$ for $t \leq 4$ follows from Corollary 7.5. The reason is as follows. We construct $\neq_2$ on the left using the gadget pictured in Figure 7a with $(=_r) \in \mathcal{G}$ on the left side assigned to circle vertices and $\neq_2$ on the right side assigned to square vertices. Then we move $\mathcal{G}$ to the right side via $\neq_2$ on the right side. We construct $=_t$ on the right side in the same Euclidean process using $\mathcal{G}$ of the right side and $\neq_2$ of the left side. This gives us a reduction from Pl-Holant $(\neq_2 \mid =_t, \mathcal{EO})$, which is #P-hard by Corollary 7.5 if $t = 3, 4$. Otherwise $t = 1, 2$. Recall that $(=_r) \in \mathcal{G}$ for some $r \geq 3$. We use $=_t$ to reduce the arity of $=_r$ to 3 or 4, if necessary. Again we are done by Corollary 7.5.

If we do not assume there is at least one hyperedge of size ≥ 3 in Pl-Holant $\left(\mathcal{G} \mid \mathcal{EO}^+\right)$, and $t = \gcd(S) \leq 2$, then the problem is tractable if and only if $S \subseteq \{1, 2\}$. The tractability is due to Kasteleyn's algorithm, as there is no hyperedge. In summary, we have the following theorem.

Theorem 7.15. *The problem* Pl-Holant $(\mathcal{G} \mid \mathcal{EO}^+)$ *counts perfect matchings over hypergraphs with planar incidence graphs, where the hyperedge sizes are prescribed by a set S of positive integers. Let $t = \gcd(S)$. If $t \geq 5$ or $S \subseteq \{1, 2\}$, then the problem is computable in polynomial time. Otherwise $t \leq 4$, $S \not\subseteq \{1, 2\}$, and the problem is* #P*-hard.*

8 Full Dichotomy

We are finally ready to prove our main dichotomy theorem. Recall that for a set $\mathcal{F}$ of signatures, $\mathcal{F}_{nd}^{\geq 3}$ denotes the set of non-degenerate signatures in $\mathcal{F}$ of arity at least 3, and $\mathcal{F}^*$ denotes $\mathcal{F}$ with all degenerate signatures $[a, b]^{\otimes m}$ replaced by unary $[a, b]$.

Theorem 8.1. *Let $\mathcal{F}$ be any set of symmetric, complex-valued signatures in Boolean variables. Then* Pl-Holant$(\mathcal{F})$ *is* #P*-hard unless $\mathcal{F}$ satisfies one of the following conditions:*

1. *All non-degenerate signatures in $\mathcal{F}$ are of arity at most 2;*
2. *$\mathcal{F}$ is $\mathscr{A}$-transformable;*
3. *$\mathcal{F}$ is $\mathscr{P}$-transformable;*
4. *$\mathcal{F} \subseteq \mathscr{V}^\sigma \cup \{f \in \mathscr{R}_2^\sigma \mid \text{arity}(f) = 2\}$ for some $\sigma \in \{+, -\}$;*
5. *All non-degenerate signatures in $\mathcal{F}$ are in $\mathscr{R}_2^\sigma$ for some $\sigma \in \{+, -\}$.*
6. *$\mathcal{F}$ is $\mathscr{M}$-transformable;*
7. *$\mathcal{F} \subseteq Z\mathscr{P} \cup \mathscr{M}_4^\sigma$ for some $\sigma \in \{+, -\}$, and the greatest common divisor of the arities of the signatures in $\mathcal{F}^* \cap \mathscr{P}_2$ is at least 5.*

In each exceptional case, Pl-Holant$(\mathcal{F})$ *is computable in polynomial time. If $\mathcal{F}$ satisfies condition 1, 2, 3, 4, or 5, then* Holant$(\mathcal{F})$ *is computable in polynomial time without planarity; otherwise* Holant$(\mathcal{F})$ *is* #P*-hard.*

Proof. We may assume that $\mathcal{F}$ contains no identically 0 signatures. We note that removing any identically 0 signature from a set does not affect its complexity, being either tractable or #P-hard, and it does not affect the set $\mathcal{F}$ satisfying any of the listed conditions in Case 1 to 7.

If all non-degenerate signatures in $\mathcal{F}$ are of arity at most 2, then the problem is tractable case 1. Otherwise, there is a non-degenerate signature $f \in \mathcal{F}$ of arity at least 3. By Theorem 6.1, Pl-Holant$(\mathcal{F})$ is #P-hard unless $f \in \mathscr{P}_1 \cup \mathscr{M}_2 \cup \mathscr{A}_3 \cup \mathscr{M}_3 \cup \mathscr{M}_4$ or f is vanishing. If $f \in \mathscr{P}_1$ or $f \in \mathscr{M}_2 \setminus \mathscr{P}_2$ or $f \in \mathscr{A}_3$ or $f \in \mathscr{M}_3$, then we are done by Corollary 5.4 or Lemma 5.8 or Corollary 5.6 or Lemma 5.10 respectively. Therefore, we assume that none of these is the case. This implies that $\mathcal{F}_{nd}^{\geq 3}$ is nonempty and that each of its signatures is in $\mathscr{P}_2$ or in $\mathscr{M}_4$ or vanishing. That is,

$$\emptyset \neq \mathcal{F}_{nd}^{\geq 3} \subseteq \mathscr{P}_2 \cup \mathscr{M}_4 \cup \mathscr{V}.$$

Suppose there exists some $f \in \mathcal{F}_{nd}^{\geq 3}$ which is in $\mathscr{V} \setminus \mathscr{M}_4$. We assume $f \in \mathscr{V}^+$ since the other case $\mathscr{V}^-$ is similar. In this case, we show that Pl-Holant$(\mathcal{F})$ is #P-hard, unless $\mathcal{F}$ is in Case 4 or Case 5. Assume that Pl-Holant$(\mathcal{F})$ is not #P-hard. We will discuss non-degenerate signatures of arity ≥ 3, of arity 2, and degenerate signatures separately.

1. For any $g \in \mathcal{F}_{nd}^{\geq 3}$, we claim that $g \in \mathscr{V}^+$. Suppose otherwise, then $g \in \mathscr{P}_2$ or $g \in \mathscr{V}^-$. Notice that the latter covers the case where $g \in \mathscr{M}_4$ but $g \notin \mathscr{V}^+$ (namely $g \in \mathscr{M}_4^-$). If $g \in \mathscr{P}_2$, then Pl-Holant(f, g) is #P-hard by Lemma 4.7. If $g \in \mathscr{V}^-$, then Pl-Holant(f, g) is #P-hard by Lemma 4.5 as $f \notin \mathscr{M}_4$.
2. For any non-degenerate binary signature $h \in \mathcal{F}$, it must be that $h \in \mathscr{R}_2^+$ as otherwise Pl-Holant(f, h) is #P-hard by Lemma 4.3.

3. If $\mathrm{rd}^+(g) = 1$ for all $g \in \mathcal{F}_{nd}^{\geq 3}$, then $\mathcal{F}_{nd}^{\geq 3} \subseteq \mathscr{R}_2^+$ by Lemma 2.19. Together with the fact just proved that all non-degenerate binary in $\mathcal{F}$ are in $\mathscr{R}_2^+$, Case 5 is satisfied.
Otherwise there exists $g \in \mathcal{F}_{nd}^{\geq 3}$ such that $\mathrm{rd}^+(g) \geq 2$. Then $g \in \mathscr{V}^+$ by the first item above. If $\mathcal{F}$ contains any degenerate signature $v = u^{\otimes m}$ for $m \geq 1$ and some unary u that is not a multiple of $[1, i]$, then by Lemma 4.1, Pl-Holant(g, v) is #P-hard. Hence all degenerate signatures are multiples of tensor powers of $[1, i]$, which are in $\mathscr{V}^+$. It implies that $\mathcal{F}$ is in Case 4.

Now we have that $\emptyset \neq \mathcal{F}_{nd}^{\geq 3} \subseteq \mathscr{P}_2 \cup \mathscr{M}_4$. We handle this in three cases.

1. Suppose $\mathcal{F}_{nd}^{\geq 3} \subseteq \mathscr{M}_4$. First suppose $\mathcal{F}_{nd}^{\geq 3} \subseteq \mathscr{M}_4^\sigma$ for some $\sigma \in \{+, -\}$. Assume $\sigma = +$ as $\sigma = -$ is similar. Then $\mathcal{F}_{nd}^{\geq 3} \subseteq \mathscr{R}_2^+$ by Lemma 3.9 and 2.19. If all non-degenerate binary signatures are in $\mathscr{R}_2^+$ as well, then this is Case 5 and tractable. Let h be a non-degenerate binary signature in $\mathcal{F}$ that is not in $\mathscr{R}_2^+$. We apply Lemma 4.4, and Pl-Holant$(\mathcal{F})$ is #P-hard unless $h = Z^{\otimes 2}[a, 0, 1]$ up to a nonzero factor, where $a \neq 0$. In this case we apply a Z transformation, and get Pl-Holant $\left(\neq_2 \mid [a, 0, 1], Z^{-1}\mathcal{F}\right)$. Then we do a diagonal transformation $D = \left[\begin{smallmatrix} a^{1/2} & 0 \\ 0 & 1 \end{smallmatrix}\right]$. Note that this only changes $\neq_2$ on the left hand side to a nonzero multiple of $\neq_2$. Hence we have the reduction chain:

$$\begin{aligned} \text{Pl-Holant}(\mathcal{F}) &\equiv \text{Pl-Holant}\left(\neq_2 \mid [a, 0, 1], Z^{-1}\mathcal{F}\right) \\ &\equiv \text{Pl-Holant}\left(\neq_2 \mid [1, 0, 1], D^{-1}Z^{-1}\mathcal{F}\right) \\ &\geq_T \text{Pl-Holant}(D^{-1}Z^{-1}\mathcal{F}) \end{aligned}$$

Notice that $D^{-1}Z^{-1}\mathcal{F}$ contains $\textsc{ExactOne}_k$ with $k \geq 3$ that is in $\mathscr{M}_3$ with I_2. Then by Lemma 5.10, Pl-Holant$(\mathcal{F})$ is #P-hard unless $D^{-1}Z^{-1}\mathcal{F} \subseteq I_2\mathscr{M} = \mathscr{M}$, i.e., $\mathcal{F} \subseteq ZD\mathscr{M} = Z\mathscr{M}$. The exceptional case implies that $\mathcal{F}$ is $\mathscr{M}$-transformable via Z, and we are in the tractable Case 6.
Otherwise $\mathcal{F}_{nd}^{\geq 3}$ contains both $f \in \mathscr{M}_4^+$ and $g \in \mathscr{M}_4^-$. Similarly as above, by Lemma 4.4, any non-degenerate binary signature in $\mathcal{F}$ has to be in $\mathscr{R}_2^+ \cap \mathscr{R}_2^- = \{Z^{\otimes 2}(\neq_2)\}$ (cf. Lemma 2.19), or is a nonzero constant multiple of $Z^{\otimes 2}[a, 0, 1]$ where $a \neq 0$, as otherwise Pl-Holant$(\mathcal{F})$ is #P-hard. Moreover, by Lemma 4.6, Pl-Holant$(\mathcal{F})$ is #P-hard, unless all degenerate signatures in $\mathcal{F}$ are of the form $[1, \pm i]^{\otimes m}$. Note that $[1, i] = Z[1, 0]$ and $[1, -i] = Z[0, 1]$. When this is the case, $\mathcal{F}$ is $\mathscr{M}$-transformable via Z.
2. Suppose $\mathcal{F}_{nd}^{\geq 3} \subseteq \mathscr{P}_2$. If $\mathcal{F}$ contains a non-degenerate binary signature h, then we apply Lemma 7.12 and Pl-Holant$(\mathcal{F})$ is #P-hard unless $h \in Z\mathscr{P}$, or Pl-#CSP$^2(DZ^{-1}\mathcal{F}) \leq_T$ Pl-Holant$(\mathcal{F})$ for some diagonal transformation D. If it is the latter case, then by Theorem 5.1, either Pl-Holant$(\mathcal{F})$ is #P-hard, or $DZ^{-1}\mathcal{F}$ is a subset of $T\mathscr{A}$, $\mathscr{P}$, or $T\left[\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}\right]\mathscr{M}$, for some diagonal matrix T. We claim that in any of these cases Pl-Holant$(\mathcal{F})$ is tractable. In fact,
   (a) if $DZ^{-1}\mathcal{F} \subseteq T\mathscr{A}$, then $\mathcal{F}$ is $\mathscr{A}$-transformable as $\mathcal{F} \subseteq ZD^{-1}T\mathscr{A}$ and $[1, 0, 1]$ (as a row vector) is transformed into $[1, 0, 1](ZD^{-1}T)^{\otimes 2}$, which is $[0, 1, 0] \in \mathscr{A}$ up to a nonzero constant;
   (b) if $DZ^{-1}\mathcal{F} \subseteq \mathscr{P}$, then $\mathcal{F}$ is $\mathscr{P}$-transformable as $\mathcal{F} \subseteq ZD^{-1}\mathscr{P}$ and $[1, 0, 1](ZD^{-1})^{\otimes 2}$ is $[0, 1, 0] \in \mathscr{P}$ up to a nonzero constant;
   (c) if $DZ^{-1}\mathcal{F} \subseteq T\left[\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}\right]\mathscr{M}$, then $\mathcal{F}$ is $\mathscr{M}$-transformable as $\mathcal{F} \subseteq ZD^{-1}T\left[\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}\right]\mathscr{M}$ and $[1, 0, 1]$ is transformed to $[1, 0, 1](ZD^{-1}T\left[\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}\right])^{\otimes 2}$, which is $[1, 0, -1] \in \mathscr{M}$ up to a nonzero constant.

Hence we may assume that every non-degenerate binary in $\mathcal{F}$ is in $Z\mathscr{P}$. Notice that degenerate signatures are always in $\mathscr{P}$ under any transformation. Also $\mathcal{F}_{nd}^{\geq 3}$ is a subset of $Z\mathscr{P}$ because $\mathcal{F}_{nd}^{\geq 3} \subseteq \mathscr{P}_2$ and $\mathscr{P}_2$ is just weighted equalities under Z-transformation. It implies that $\mathcal{F}$ is $\mathscr{P}$-transformable under the Z transformation. Hence we are in Case 3.

3. Finally, suppose neither of the above is the case. Then there are $f, g \in \mathcal{F}_{nd}^{\geq 3}$ with $f \in \mathscr{M}_4$ and $g \in \mathscr{P}_2$. If $\mathcal{F}_{nd}^{\geq 3}$ contains both $f \in \mathscr{M}_4^+$ and $f' \in \mathscr{M}_4^-$, then Pl-Holant($\mathcal{F}$) is #P-hard by Lemma 7.11. Otherwise $\mathcal{F}_{nd}^{\geq 3} \cap \mathscr{M}_4 \subseteq \mathscr{M}_4^+$ or $\mathscr{M}_4^-$. Let $\mathcal{G} = \mathcal{F}^* \cap \mathscr{P}_2$, and let d be the gcd of the arities of the signatures in $\mathcal{G}$. Then $\mathcal{G}$ contains at least one non-degenerate signature g of arity ≥ 3. If $d \leq 4$, then Pl-Holant($\mathcal{F}$) is #P-hard by Lemma 7.13. Otherwise $d \geq 5$. If $\mathcal{F}$ contains a non-degenerate binary signature h, then we apply Lemma 7.12 and by a similar analysis as in the case of "$\mathcal{F}_{nd}^{\geq 3} \subseteq \mathscr{P}_2$" above, we are done unless every such h is in $Z\mathscr{P}$. Ignoring a nonzero factor, it implies that either $h = Z^{\otimes 2}[1, 0, a]$ where $a \neq 0$ or $h = Z^{\otimes 2}(\neq_2)$. If $h = Z^{\otimes 2}[1, 0, a]$, then $h \in \mathcal{F}^* \cap \mathscr{P}_2$, and it contradicts $d \geq 5$. Hence $h = Z^{\otimes 2}(\neq_2)$. If there is any degenerate $v = (Z[a, b])^{\otimes m}$ in $\mathcal{F}$ with $ab \neq 0$, then $Z[a, b] \in \mathcal{F}^* \cap \mathscr{P}_2$ and it also contradicts $d \geq 5$.
   In summary, Pl-Holant($\mathcal{F}$) is #P-hard unless $\mathcal{F}_{nd}^{\geq 3} \subseteq \mathscr{P}_2 \cup \mathscr{M}_4$, $\mathcal{F}_{nd}^{\geq 3} \cap \mathscr{M}_4 \subseteq \mathscr{M}_4^\sigma$ for some $\sigma \in \{+, -\}$, the greatest common divisor of the arities of the signatures in $\mathcal{F}^* \cap \mathscr{P}_2$ is at least 5. Every non-degenerate binary in $\mathcal{F}$ is of the form $Z^{\otimes 2}(\neq_2)$, and every degenerate in $\mathcal{F}$ is of the form $(Z[1, 0])^{\otimes m}$ or $(Z[0, 1])^{\otimes m}$. Notice that $\mathscr{P}_2$, $Z^{\otimes 2}(\neq_2)$, $(Z[1, 0])^{\otimes m}$, and $(Z[0, 1])^{\otimes m}$ are all in $Z\mathscr{P}$. Hence the exceptional case implies that $\mathcal{F} \subseteq Z\mathscr{P} \cup \mathscr{M}_4^\sigma$ for some $\sigma \in \{+, -\}$ and the greatest common divisor of the arities of the signatures in $\mathcal{F}^* \cap \mathscr{P}_2$ is at least 5. This is tractable Case 7.

The tractability of Holant($\mathcal{F}$) in Case 1, Case 2, Case 3, Case 4, and Case 5 follows from the Holant dichotomy Theorem 2.21, which also implies that Holant($\mathcal{F}$) is #P-hard otherwise. The tractability of Pl-Holant($\mathcal{F}$) in Case 6 follows from Theorem 2.7. The tractability of Pl-Holant($\mathcal{F}$) in Case 7 follows from Lemma 7.14. This completes the proof. □

A Holant Dichotomy: Is the FKT Algorithm Universal? Part II: Planar #CSP^2 Dichotomy

In Part II of this paper, we prove Theorem A.2, which is the complexity dichotomy theorem of Pl-#CSP$^2(\mathcal{F})$, where $\mathcal{F}$ is a set of complex-valued symmetric signatures on Boolean variables. After we define some relevant notions, we give an outline of the proof of Theorem A.2. Throughout Part II, we denote by α (respectively ρ) any quantity that satisfies $\alpha^4 = -1$ (respectively $\rho^4 = 1$).

A Preliminaries

We will first define some tractable families of signatures that are expressible under a holographic transformation, specific to the Pl-#CSP^2 framework.

Definition A.1. *Let $\mathcal{T}_k = \left\{ \left[\begin{smallmatrix} 1 & 0 \\ 0 & \omega \end{smallmatrix}\right] \mid \omega^k = 1 \right\}$ be a set of diagonal matrices of order dividing k and $\mathcal{T}_{\bar{k}} = \mathcal{T}_{2k} \setminus \mathcal{T}_k = \left\{ \left[\begin{smallmatrix} 1 & 0 \\ 0 & \omega \end{smallmatrix}\right] \mid \omega^k = -1 \right\}$. Let $\mathscr{A}^\dagger = \mathcal{T}_4 \mathscr{A}$ and $\widehat{\mathscr{M}}^\dagger = \mathcal{T}_2 \widehat{\mathscr{M}}$ be the sets of signatures transformed by $\mathcal{T}_4$ from the Affine family $\mathscr{A}$ and transformed by $\mathcal{T}_2$ from $\widehat{\mathscr{M}}$, respectively, where for a class of signatures $\mathscr{C}$, we denote*

$$\mathcal{T}_k \mathscr{C} = \{ T^{\otimes \operatorname{arity}(f)} f \mid T \in \mathcal{T}_k \text{ and } f \in \mathscr{C} \}.$$

Let

$$\widetilde{\mathscr{A}} = \mathscr{A} \cup \mathscr{A}^\dagger \qquad \text{and} \qquad \widetilde{\mathscr{M}} = \widehat{\mathscr{M}} \cup \widehat{\mathscr{M}}^\dagger$$

be the $\mathscr{A}$-transformable and $\mathscr{M}$-transformable signatures for Pl-#CSP^2.

Recall that $\widehat{\mathscr{M}} = H\mathscr{M}$ is the set of Matchgate signatures $\mathscr{M}$ transformed by the Hadamard basis $H = \left[\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}\right]$. Note that $\mathscr{A}$ is unchanged under the transformation by H, and thus there is no need to define $\widehat{\mathscr{A}}$. Also note that $\mathscr{P}$ is unchanged under any diagonal matrix. Thus there is no need to define $\mathscr{P}^\dagger$. For $T = \left[\begin{smallmatrix} 1 & 0 \\ 0 & \omega \end{smallmatrix}\right] \in \mathcal{T}_4$ with $\omega^4 = 1$, $T\mathscr{A} = \mathscr{A}$. Thus $\widetilde{\mathscr{A}}$ is $\mathscr{A}$ under transformations by $T = \left[\begin{smallmatrix} 1 & 0 \\ 0 & \omega \end{smallmatrix}\right] \in \mathcal{T}_8$. For such T, we have $(=_{2n})T^{\otimes 2n} \in \mathscr{A}$. Hence $\widetilde{\mathscr{A}}$ is $\mathscr{A}$-transformable for Pl-#CSP^2. Similarly, for $T = \left[\begin{smallmatrix} 1 & 0 \\ 0 & \pm 1 \end{smallmatrix}\right]$, $TH = \left[\begin{smallmatrix} 1 & 1 \\ \pm 1 & \mp 1 \end{smallmatrix}\right]$ = either H or $H\left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right]$, and $\left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right]\mathscr{M} = \mathscr{M}$. Thus $T\widehat{\mathscr{M}} = \widehat{\mathscr{M}}$, and $\widetilde{\mathscr{M}}$ is $\mathscr{M}$ transformed under TH for all $T \in \mathcal{T}_4$. Also note that for all such T, we have $(=_{2n})(TH)^{\otimes 2n} \in \mathscr{M}$. Hence $\widetilde{\mathscr{M}}$ is $\mathscr{M}$-transformable for Pl-#CSP^2.

In the proof of No-Mixing of different tractable sets, because of a particular order in which we carry out the proof, to make an overall logical structure more apparent we introduce the following notations

$$S_1 = \widehat{\mathscr{M}}, \qquad S_2 = \widehat{\mathscr{M}}^\dagger, \qquad S_3 = \mathscr{A}^\dagger, \qquad S_4 = \mathscr{A}, \qquad \text{and} \qquad S_5 = \mathscr{P}.$$

We will prove the following Main Theorem of Part II. It is not hard to see that this is a rephrase of Theorem 5.1 from Part I. It follows from Theorem C.13, Theorem H.5 and Theorem G.4, which will be shown in later sections. It follows from the definition of $\mathscr{P}$-transformability, $\mathscr{A}$-transformability and $\mathscr{M}$-transformability that if $\mathcal{F} \subseteq S_k$ for any $1 \leq k \leq 5$, then Pl-#CSP$^2(\mathcal{F})$ is tractable.

Theorem A.2. *For any set of complex-valued symmetric signatures $\mathcal{F}$ on Boolean variables, if $\mathcal{F} \subseteq \mathscr{P}$, or $\mathscr{A}$, or $\mathscr{A}^\dagger$, or $\widehat{\mathscr{M}}$, or $\widehat{\mathscr{M}}^\dagger$, then* Pl-#CSP$^2(\mathcal{F})$ *is tractable. Otherwise,* Pl-#CSP$^2(\mathcal{F})$ *is #P-hard.*

Proof Outline. We now give an outline of the proof of Theorem A.2. The overall plan is to break the proof into two main steps.

The first step is to prove the dichotomy theorem for Pl-#CSP$^2(\mathcal{F})$ when there is at least one nonzero signature of *odd* arity in $\mathcal{F}$. In this case we can make use of Lemma B.2 that shows that we can simulate Pl-#CSP$(\mathcal{F})$ by Pl-#CSP$^2(\mathcal{F})$ if $\mathcal{F}$ includes a unary signature $[a, b]$ with $ab \neq 0$. Then we can apply the known dichotomy Theorem A.22 for Pl-#CSP^1. However this strategy (provably) *cannot* work in the case when every signature in $\mathcal{F}$ satisfies the *parity* constraint. In that case we employ other means. This first step of the proof is relatively uncomplicated.

The second step is to deal with the case when all nonzero signatures in $\mathcal{F}$ have even arity. This is where the real difficulties lie. In this case it is impossible to directly construct *any* unary signature. So we cannot use Lemma B.2 in this case. But Lemma B.3 provides a way to simulate Pl-#CSP$(\mathcal{F})$ by Pl-#CSP$^2(\mathcal{F})$ in a *global* fashion, *if* $\mathcal{F}$ includes some tensor power of the form $[a, b]^{\otimes 2}$ where $ab \neq 0$. Moreover, we have a lucky break (for the complexity of the proof) if $\mathcal{F}$ includes a signature that is in $\widehat{\mathscr{M}} \setminus (\mathscr{P} \cup \widetilde{\mathscr{A}})$. In this case, we can construct a special binary signature, and then use Lemma E.2 to obtain $[1, 1]^{\otimes 2}$ by interpolation. This proof uses the theory of *cyclotomic fields*. This simplifies the proof greatly. For all other cases (when $\mathcal{F}$ has only even arity signatures), the proof gets going in earnest—we will attempt an induction on the arity of signatures.

The lowest arity of this induction will be 2. We will try to reduce the arity to 2 whenever possible; however for many cases an arity reduction to 2 destroys the #P-hardness at hand. Therefore the true basis of this induction proof of Pl-#CSP^2 starts with arity 4. Consequently we will first prove a dichotomy theorem for Pl-#CSP$^2(f)$, where f is a signature of arity 4. This proof is presented in Section D. Several tools will be used. These include the rank criterion for redundant signatures, Theorem A.21 for arity 2 signatures, and a trick we call the *Three Stooges* by domain pairing.

However in the next step we do not attempt a general Pl-#CSP^2 dichotomy for a *single* signature of even arity. This would have been natural at this point, but it would have been too difficult. We will need some additional leverage by proving a conditional No-Mixing Lemma for pairs of signatures of even arity. So, seemingly taking a detour, we prove that for two signatures f and g both of even arity, that individually belong to some tractable class, but do not belong to a single tractable class in the conjectured Pl-#CSP^2 dichotomy (that is yet to be proved), the problem Pl-#CSP$^2(f, g)$ is #P-hard. We prove this No-Mixing Lemma for any pair of signatures f and g both of even arity, not restricted to arity 4. Even though at this point we only have a dichotomy for a single signature of arity 4, we prove this No-Mixing Lemma for higher even arity pairs f and g by simulating two signatures f' and g' of arity 4 that belong to different tractable sets, from that of Pl-#CSP$^2(f, g)$. After this arity reduction (within the No-Mixing Lemma), we prove that Pl-#CSP$^2(f', g')$ is #P-hard by the dichotomy for a *single* signature of arity 4. After this, we prove a No-Mixing Lemma for a *set* of signatures $\mathcal{F}$ of even arities, which states that *if* $\mathcal{F}$ is contained in

the union of all tractable classes, then it is still #P-hard unless it is *entirely* contained in a single tractable class. Note that at this point we still only have a *conditional* No-Mixing Lemma in the sense that we have to assume every signature in $\mathcal{F}$ belongs to some tractable set.

We then attempt the proof of a Pl-#CSP^2 dichotomy for a *single* signature of arbitrary even arity. This uses all the previous lemmas, in particular the (conditional) No-Mixing Lemma for a set of signatures. However, after completing the proof of this Pl-#CSP^2 dichotomy for a single signature of even arity, the No-Mixing Lemma becomes absolute.

Finally the dichotomy for a single signature of even arity is logically extended to a dichotomy theorem for Pl-#CSP$^2(\mathcal{F})$ where all signatures in $\mathcal{F}$ have even arity. Together with the first main step when $\mathcal{F}$ contains some nonzero signature of odd arity, this completes the proof of Theorem A.2. □

In the rest of this Section A, we will introduce the operators ∂ and $\int$, and give some characterizations of the tractable classes. We will also introduce some preliminary lemmas, including one using the domain pairing technique, and list some known dichotomies. In Section B, we discuss a technique to simulate Pl-#CSP by Pl-#CSP^2. Section C proves Theorem A.2 in the case when $\mathcal{F}$ contains at least one nonzero signature of odd arity. Section D proves the base case of the even arity case of Theorem A.2 when $\mathcal{F}$ consists of a single signature of arity 4. Section E gives an application of cyclotomic field which simplifies the proof of Theorem A.2 when $\mathcal{F}$ contains a signature in $\widetilde{\mathscr{M}} \setminus (\mathscr{P} \cup \widetilde{\mathscr{A}})$. Section F proves the conditional No Mixing lemmas for a pair of signatures of even arity. Section G generalizes the No Mixing lemmas to a set of signatures of even arity. Section H finishes the proof of Theorem A.2.

Remark 4. We occasionally make some remarks (such as Remark 5 and Remark 6 in Subsection E.2) to explain the complications forced upon the proof by various reasons, and why another more straightforward approach would not succeed. These remarks are not logically necessary to the proof, but hopefully they provide some insight and point out pitfalls in the proof.

The next lemma is a simple fact that is used many times. It essentially says that the set $\{0, 1, i, -1, -i, \infty\}$ is closed set-wise under the mapping $z \mapsto \frac{z+1}{z-1}$. The proof is straightforward, so we omit it.

Lemma A.3. *Let $x \neq y$ and $\lambda = \frac{x+y}{x-y}$. Then $\lambda^4 \notin \{0, 1\}$ iff $x^4 \neq y^4$ and $xy \neq 0$.*

Definition A.4 (Derivative)**.** *Let f and g be two symmetric signatures of arities n and m respectively, and $n > m$. By connecting all m input edges of g to f, we get a planar $\{f, g\}$-gate with a signature of arity $n - m$. This* derivative signature *will be denoted by $\partial_g(f)$. If $kn < m$ and we connect k copies of g to f, which is the same as forming $\partial_g(f)$ sequentially k times, the resulting* repeated derivative signature *is denoted by $\partial_g^k(f)$. If $g = [1, 0, 1]$, we denote $\partial_g(f)$ simply by $\partial(f)$.*

Calculus: Our proof will make substantial use of a *calculus* using this notion of derivatives. This calculus is essentially a systematic way to calculate the signatures of some gadget constructions. In a Pl-Holant problem Pl-Holant$(\mathcal{G} \mid \mathcal{F})$, if $g \in \mathcal{G}$ and $f \in \mathcal{F}$, then we say that g is from the LHS and f is from the RHS. If f has arity n and g has arity m, and $n > m$, then we can form the signature $\partial_g(f)$ and Pl-Holant$(\mathcal{G} \mid \mathcal{F} \cup \{\partial_g(f)\}) \leq_T$ Pl-Holant$(\mathcal{G} \mid \mathcal{F})$. If $m > n$ we can form $\partial_f(g)$ and Pl-Holant$(\mathcal{G} \cup \{\partial_f(g)\} \mid \mathcal{F}) \leq_T$ Pl-Holant$(\mathcal{G} \mid \mathcal{F})$. In particular, for Pl-#CSP$^2(\mathcal{F}) \equiv$ Pl-Holant$(\mathcal{EQ}_2 \mid \mathcal{F})$ we consider all $(=_{2k})$ as from the LHS. In this case if $h \in \mathcal{F}$

with arity $< n$ then we can also form $\partial_h(f)$, by first moving h to LHS via $(=_2) \in \mathcal{EQ}_2$, and then Pl-#CSP$^2(\mathcal{F} \cup \{\partial_h(f)\}) \leq_T$ Pl-#CSP$^2(\mathcal{F})$. Note that if we discuss Pl-#CSP$^4(\mathcal{F})$ then this operation $\partial_h(f)$ is in general not permissible, and has to be justified in each individual case, e.g. when h has even arity and one can construct $[1, 0, 1]^{\otimes 2}$ in the LHS.

To familiarize the readers with this calculus, we list some simple calculations below, which we will use often in our proofs freely without comment.

For any g, the operator $\partial_g(\cdot)$ is a linear operator. It also depends on g linearly.

By definition $\partial([f_0, f_1, \ldots, f_n]) = [f_0 + f_2, f_1 + f_3, \ldots, f_{n-2} + f_n]$ has arity $n - 2$.

1. If $f = [s, t]^{\otimes n}$, then
   - $\partial^k_{[a,b]}(f) = (as + bt)^k [s, t]^{\otimes n-k}$ if $n > k$.
   - $\partial^k_{[a,b,c]}(f) = (as^2 + 2bst + ct^2)^k [s, t]^{\otimes n-2k}$ if $n > 2k$;
     in particular, $\partial^k(f) = (s^2 + t^2)^k [s, t]^{\otimes n-2k}$.
   - $\partial^k_{=_4}(f) = (s^4 + t^4)^k [s, t]^{\otimes n-4k}$, if $n > 4k$.
   - For $g = [g_0, g_1, \ldots, g_m]$, we have $\partial_g(=_n) = [g_0, 0, \ldots, 0, g_m]$ of arity $n - m$, where $n > m$.
2. Let f be of arity n and $f_k = (\pm 1)^k (n - 2k)$ $(0 \leq k \leq n)$, then
   - $\partial(f)$ has arity $n' = n - 2$ and $(\partial(f))_k = 2(\pm 1)^k (n' - 2k)$. If n is odd, then $\partial^{\frac{n-1}{2}}(f) = 2^{\frac{n-1}{2}} [1, \mp 1]$.
   - $\partial_{=_4}(f)$ has arity $n'' = n - 4$ and $(\partial_{=_4}(f))_k = 2(\pm 1)^k (n'' - 2k)$.
     If $n \equiv 1 \pmod 4$, then $\partial^{\frac{n-1}{4}}_{=_4}(f) = 2^{\frac{n-1}{4}} [1, \mp 1]$.
     If $n \equiv 3 \pmod 4$, then $\partial(\partial^{\frac{n-3}{4}}_{=_4}(f)) = 2^{\frac{n+1}{4}} [1, \mp 1]$.
3. Let f be of arity n and $f_k = (\pm i)^k (n - 2k)$ $(0 \leq k \leq n)$, then
   - $\partial(f) = 4[1, \pm i]^{\otimes n-2}$.
   - $\partial_{=_4}(f)$ has arity $m = n - 4$ and $(\partial_{=_4}(f))_k = 2(\pm i)^k (m - 2k)$.
     If $n \equiv 1 \pmod 4$, then $\partial^{\frac{n-1}{4}}_{=_4}(f) = 2^{\frac{n-1}{4}} [1, \mp i]$.
     If $n \equiv 3 \pmod 4$, then $\partial(\partial^{\frac{n-3}{4}}_{=_4}(f)) = 2^{\frac{n+5}{4}} [1, \pm i]$.

Now we define an inverse operator $\int(\cdot)$ to ∂. Just like the usual calculus there is a certain non-uniqueness in the expression in an *indefinite* integral; this non-uniqueness is addressed in Lemma A.5. One reasonable definition for $\int([f_0, f_1, \ldots, f_n])$ is $F = [F_0, F_1, \ldots, F_{n+2}]$ such that

$$F_k = \sum_{s \geq 0} (-1)^s f_{k+2s} = f_k - f_{k+2} + f_{k+4} - \cdots$$

where we define $f_k = 0$ for all $k > n$. Clearly $\partial(F) = f$.

Lemma A.5. *Let F and G be symmetric signatures of arity $n \geq 3$ and suppose $\partial(F) = \partial(G)$. Then $F - G = x[1, i]^{\otimes n} + y[1, -i]^{\otimes n}$, for some constants x and y.*

Proof. The signature $H = F - G$ satisfies $\partial(H) = 0$, and thus satisfies the second order recurrence relation $H_k + H_{k+2} = 0$ for $0 \leq k \leq n - 2$. Hence there exist constants x and y such that $H = x[1, i]^{\otimes n} + y[1, -i]^{\otimes n}$. □

Thus $\int(\cdot)$ is well-defined up to an additive term $x[1, i]^{\otimes n} + y[1, -i]^{\otimes n}$. In this paper, we choose to write the expression $\int(f)$ by the following definition when a certain special expression of f exists. This is more convenient for our proofs.

Definition A.6. *For $n \geq 3$,*

- $\int(0) = 0$.
- *For $a^2 + b^2 \neq 0$, $\int([a, b]^{\otimes n-2}) = \frac{1}{a^2+b^2}[a, b]^{\otimes n}$.*
- $\int([1, \pm i]^{\otimes n-2})$ *has arity n and* $[\int([1, \pm i]^{\otimes n-2})]_k = \frac{1}{4}(\pm i)^k(n - 2k)$.
- *If the signature g has arity $n - 2$ and $g_k = (\pm 1)^k(n - 2 - 2k)$, then $\int(g)$ has arity n and* $[\int(g)]_k = \frac{1}{2}(\pm 1)^k(n - 2k)$.
- *If the signature g has arity $n - 2$ and $g_k = (\pm i)^k(n - 2 - 2k)$, then $\int(g)$ has arity n and* $[\int(g)]_k = (-\frac{n}{2}k + \frac{1}{2}k^2)(\pm i)^k$.

Clearly for all f where $\int(f)$ is given in the above definition, $\partial[\int(f)] = f$.

When we prove the dichotomy theorem for Pl-#CSP$^2(f)$, where f has arity n, we can get a signature f' of arity $n - 2$ by taking a self loop with f, i.e., $f' = \partial(f)$. Clearly Pl-#CSP$^2(f') \leq_T$ Pl-#CSP$^2(f)$. If $f' \notin \mathscr{P} \cup \widetilde{\mathscr{A}} \cup \widetilde{\mathscr{M}}$, then by induction Pl-#CSP$^2(f')$ is #P-hard. Thus Pl-#CSP$^2(f)$ is also #P-hard. Definition A.6 allows us to write down an explicit expression for $\int(f')$ for all cases when $f' \in \mathscr{P} \cup \widetilde{\mathscr{A}} \cup \widetilde{\mathscr{M}}$.

The following is an explicit list of $\int(f')$ for $f' = \partial(f) \in \mathscr{P} \cup \widetilde{\mathscr{A}} \cup \widetilde{\mathscr{M}}$. We can recover f up to the constants x, y from $\partial(f)$ by Lemma A.5. This list is for the convenience of the readers.

Proposition A.7 (Explicit List for $\int(f')$)**.**

- $\int(f') \equiv 0$ *if* $f' \equiv 0$.
- $\int([1, 0]^{\otimes n-2} + a[0, 1]^{\otimes n-2}) = [1, 0]^{\otimes n} + a[0, 1]^{\otimes n}$.
- $\int([1, \gamma]^{\otimes n-2} + i^r[1, -\gamma]^{\otimes n-2}) = \frac{1}{1+\gamma^2}[1, \gamma]^{\otimes n} + \frac{i^r}{1+\gamma^2}[1, -\gamma]^{\otimes n}$ *where* $\gamma^2 \neq -1, \gamma^8 = 1$.
- $\int([s, t\rho]^{\otimes n-2} \pm [t, s\rho]^{\otimes n-2}) = \frac{1}{s^2+\rho^2t^2}[s, \rho t]^{\otimes n} \pm \frac{1}{\rho^2s^2+t^2}[t, \rho s]^{\otimes n}$, *where* $\rho^4 = 1, st \neq 0, s^4 \neq t^4$.
- $[\int(f')]_k = \frac{1}{2}(\pm 1)^k(n - 2k)$ *if f' has arity $n - 2$ and* $f'_k = (\pm 1)^k(n - 2 - 2k)$.
- $[\int(f')]_k = \frac{1}{4}(\pm i)^k(n - 2k)$ *if f' has arity $n - 2$ and* $f' = [1, \pm i]^{\otimes n-2}$.
- $[\int(f')]_k = \frac{1}{4}[i^k + i^r(-i)^k](n - 2k)$ *if f' has arity $n - 2$ and* $f' = [1, i]^{\otimes n-2} + i^r[1, -i]^{\otimes n-2}$.
- $[\int(f')]_k = (-\frac{n}{2}k + \frac{1}{2}k^2)(\pm i)^k$ *if f' has arity $n - 2$ and* $f'_k = (\pm i)^k(n - 2 - 2k)$.

The following lemma is used to determine whether a binary signature belongs to various tractable sets $\mathscr{P}$, $\mathscr{A}$, $\mathscr{A}^\dagger$, $\widehat{\mathscr{M}}$, and $\widehat{\mathscr{M}^\dagger}$. It can be proved directly by the definition.

Lemma A.8. *For any binary symmetric signature f,*

- $f \in \mathscr{P}$ *iff* $f = [a, 0, c]$ *or* $f = [0, b, 0]$ *or* $f = [a, b]^{\otimes 2}$.
- $f \in \mathscr{A}$ *iff up to a scalar,* $f = [1, \rho, -\rho^2]$ *where* $\rho^4 = 1$, *or* $[0, 1, 0]$, *or* $[1, 0, \rho]$ *where* $\rho^4 = 1$, *or* $[x, y]^{\otimes 2}$ *where* $(x^4 = y^4 \neq 0$ *or* $xy = 0)$.
- $f \in \mathscr{A}^\dagger$ *iff up to a scalar,* $f = [1, \alpha, -\alpha^2]$ *where* $\alpha^4 = -1$, *or* $[0, 1, 0]$, *or* $[1, 0, \rho]$ *where* $\rho^4 = 1$, *or* $[x, y]^{\otimes 2}$ *where* $(x^4 = -y^4 \neq 0$ *or* $xy = 0)$.
- $f \in \widehat{\mathscr{M}}$ *iff* $f = [a, b, a]$ *or* $[a, 0, -a]$.
- $f \in \widehat{\mathscr{M}^\dagger}$ *iff* $f = [a, b, -a]$ *or* $[a, 0, a]$.

Corollary A.9 gives some necessary conditions for a binary signature to belong to a tractable set.

Corollary A.9. *For any binary signature $f = [a, b, c]$,*

- $f \in \mathscr{P} \implies f$ *satisfies either the parity constraint or* $b^2 = ac$.
- $f \in \mathscr{A} \implies a^2 = c^2$ *or* $b = 0$. *If* $f \in \mathscr{A} \setminus \mathscr{P}$, *then* $f = [1, \rho, -\rho^2]$, $\rho^4 = 1$.
- $f \in \mathscr{A}^\dagger \implies a^2 = -c^2$ *or* $b = 0$. *If* $f \in \mathscr{A}^\dagger \setminus \mathscr{P}$, *then* $f = [1, \alpha, -\alpha^2]$, $\alpha^4 = -1$.
- $f \in \widetilde{\mathscr{A}} \implies$ *the norms of all nonzero entries are equal.*

- $f \in \widetilde{\mathscr{M}} \implies a^2 = c^2$.

Furthermore, all signatures in each tractable set satisfy a second order recurrence relation.

Definition A.10. *Let $f = [f_0, f_1, \ldots, f_n]$. If there exist constants a, b and c, not all zero, such that $af_k - bf_{k+1} + cf_{k+2} = 0$ for $1 \le k \le n-2$, then we say f has type $\langle a, b, c\rangle$, and it is denoted by $f \in \langle a, b, c\rangle$.*

For a non-degenerate symmetric signature f of arity at least 3, if f has type $\langle a, b, c\rangle$, its type is uniquely determined up to a nonzero multiple. The next lemma states this type information for the various tractable sets. We can use the lemma to check whether a symmetric signature can possibly be in a tractable set.

Lemma A.11. *Let $f \in \mathscr{P} \cup \widetilde{\mathscr{A}} \cup \widetilde{\mathscr{M}}$ be non-degenerate and have arity ≥ 3.*

- *If $f \in \mathscr{P}$ then $f \in \langle 0, 1, 0\rangle$.*
- *If $f \in \mathscr{A}$ then $f \in \langle 0, 1, 0\rangle$ or $f \in \langle 1, 0, \pm 1\rangle$. If $f \in \mathscr{A} \setminus \mathscr{P}$ then $f \in \langle 1, 0, \pm 1\rangle$.*
- *If $f \in \mathscr{A}^\dagger$ then $f \in \langle 0, 1, 0\rangle$ or $f \in \langle 1, 0, \pm i\rangle$. If $f \in \mathscr{A}^\dagger \setminus \mathscr{P}$ then $f \in \langle 1, 0, \pm i\rangle$.*
- *If $f \in \widehat{\mathscr{M}}$ then $f \in \langle 0, 1, 0\rangle$ or $f \in \langle 1, c, 1\rangle$. If $f \in \widehat{\mathscr{M}} \setminus (\mathscr{P} \cup \widetilde{\mathscr{A}})$ then $f \in \langle 1, c, 1\rangle$ with $c \ne 0$.*
- *If $f \in \widehat{\mathscr{M}}^\dagger$ then $f \in \langle 0, 1, 0\rangle$ or $f \in \langle 1, c, -1\rangle$. If $f \in \widehat{\mathscr{M}}^\dagger \setminus (\mathscr{P} \cup \widetilde{\mathscr{A}})$ then $f \in \langle 1, c, -1\rangle$ with $c \ne 0$.*

The following two corollaries follow from Lemma A.8 for the binary case, and Lemma A.11 for arity $n \ge 3$.

Corollary A.12. *If $f \in \mathscr{A} \setminus \mathscr{P}$, then $f \notin \mathscr{A}^\dagger$. Similarly, If $f \in \mathscr{A}^\dagger \setminus \mathscr{P}$, then $f \notin \mathscr{A}$.*

Corollary A.13. *If $f \in \widehat{\mathscr{M}} \setminus (\mathscr{P} \cup \widetilde{\mathscr{A}})$, then $f \notin \widehat{\mathscr{M}}^\dagger$. Similarly, if $f \in \widehat{\mathscr{M}}^\dagger \setminus (\mathscr{P} \cup \widetilde{\mathscr{A}})$, then $f \notin \widehat{\mathscr{M}}$.*

The following lemma gives a characterization for $\widetilde{\mathscr{M}} \setminus (\mathscr{P} \cup \widetilde{\mathscr{A}})$.

Lemma A.14. *Let $f = [f_0, \ldots, f_n]$ be a symmetric signature of arity n. Then $f \in \widehat{\mathscr{M}} \setminus (\mathscr{P} \cup \widetilde{\mathscr{A}})$ iff*

- *$n = 2$ and $f = \lambda[1, a, 1]$, where $a^4 \notin \{0, 1\}$ and $\lambda \ne 0$; or*
- *$n \ge 3$ and $f = [s, t]^{\otimes n} \pm [t, s]^{\otimes n}$, where $st \ne 0$ and $s^4 \ne t^4$; or*
- *$n \ge 3$ and $f_k = \lambda(\pm 1)^k (n - 2k)$, where $\lambda \ne 0$.*

Similarly, $f \in \widehat{\mathscr{M}}^\dagger \setminus (\mathscr{P} \cup \widetilde{\mathscr{A}})$ iff

- *$n = 2$ and $f = \lambda[1, b, -1]$, where $b^4 \notin \{0, 1\}$ and $\lambda \ne 0$; or*
- *$n \ge 3$ and $f = [s, ti]^{\otimes n} \pm [t, si]^{\otimes n}$ where $st \ne 0$ and $s^4 \ne t^4$; or*
- *$n \ge 3$ and $f_k = \lambda(\pm i)^k (n - 2k)$, where $\lambda \ne 0$.*

Proof. We prove the lemma for $\widehat{\mathscr{M}}$. The proof for $\widehat{\mathscr{M}}^\dagger$ follows from a holographic transformation by $\left[\begin{smallmatrix} 1 & 0 \\ 0 & i \end{smallmatrix}\right]$.

By Lemma A.8, a binary symmetric signature $f \in \widehat{\mathscr{M}}$ has the form $[a, b, a]$ or $[a, 0, -a]$. Since $[a, 0, -a] \in \mathscr{A}$ as a multiple of $[1, 0, -1]$, we exclude it. For $[a, b, a]$, if $ab = 0$, then $f \in \mathscr{P}$. Also if $a^4 = b^4$, then $[a, b, a] \in \mathscr{A}$, being a multiple of $[1, \pm 1]^{\otimes 2}$ or $[1, \pm i, 1]$. This gives the form $f = \lambda[1, b, 1]$ with $b^4 \notin \{0, 1\}$ and $\lambda \ne 0$. Conversely, any f of this form belongs to $\widehat{\mathscr{M}} \setminus (\mathscr{P} \cup \widetilde{\mathscr{A}})$.

For arity $n \ge 3$, $f \in \widehat{\mathscr{M}}$ iff f takes the form $[s, t]^{\otimes n} \pm [t, s]^{\otimes n}$ or $f_k = \lambda(\pm 1)^k (n - 2k)$. For the latter case $f \in \widehat{\mathscr{M}} \setminus (\mathscr{P} \cup \widetilde{\mathscr{A}})$ follows from its type $\langle 1, \pm 2, 1\rangle$.

For $f = [s,t]^{\otimes n} \pm [t,s]^{\otimes n}$, if $st = 0$, then $f \in \mathscr{P}$. If $s^2 = t^2$, then f is degenerate, thus $f \in \mathscr{P}$. If $s^2 = -t^2$, then $f \in \mathscr{A}$. Conversely, if $st \neq 0$ and $s^4 \neq t^4$, then f is non-degenerate and f_k has type $\langle 1, \frac{s}{t} + \frac{t}{s}, 1\rangle$. Note that $\frac{s}{t} + \frac{t}{s} \neq 0$ by $s^4 \neq t^4$. Thus $f \in \widehat{\mathscr{M}} \setminus (\mathscr{P} \cup \widehat{\mathscr{A}})$ by Lemma A.11. □

By the second recurrence relation of the signatures in $\widetilde{\mathscr{M}} \setminus (\mathscr{P} \cup \widetilde{\mathscr{A}})$, we have the following lemma that will be used in the proof of Theorem C.11.

Corollary A.15. *If $f \in \widetilde{\mathscr{M}} \setminus (\mathscr{P} \cup \widetilde{\mathscr{A}})$, then f does not satisfy parity constraints.*

Proof. For $f \in \widetilde{\mathscr{M}} \setminus (\mathscr{P} \cup \widetilde{\mathscr{A}})$, if f has arity 2, then $f = \lambda[1, a, 1]$ for some $\lambda \neq 0$, $a^4 \neq 0, 1$ by Lemma A.14. Thus it does not satisfy parity constraints.

For f with arity $n \geq 3$, by Lemma A.11, there exists a constants $c \neq 0$ such that $f \in \langle 1, c, 1\rangle$. Note that there exists $f_k \neq 0$, where $1 \leq k \leq n-1$ by $f \notin \mathscr{P}$. If f satisfies parity constraints, then $f_{k-1} = f_{k+1} = 0$. Moreover, by $f_{k-1} - cf_k + f_{k+1} = 0$, we have $c = 0$. This is a contradiction.

The proof for $f \in \widehat{\mathscr{M}^\dagger} \setminus (\mathscr{P} \cup \widetilde{\mathscr{A}})$ follows from a holographic transformation by $\left[\begin{smallmatrix} 1 & 0 \\ 0 & i \end{smallmatrix}\right]$. □

The following lemma gives a characterization of nonzero signatures in $\widehat{\mathscr{M}}$. A Gen-Eq is a signature of the form $f = [a, 0, \ldots, 0, b]$, called a generalized equality (with $a = 0$ or $b = 0$ allowed.)

Lemma A.16. *A* Gen-Eq *signature f is in $\widehat{\mathscr{M}}$ iff $f = \lambda[1, 0, \ldots, 0, \pm 1]$, for some λ.*

Suppose f is a symmetric signature that is not a Gen-Eq*. Then $f \in \widehat{\mathscr{M}}$ iff f satisfies a second order recurrence $f_k - cf_{k+1} + f_{k+2} = 0$ (for $0 \leq k \leq \mathrm{arity}(f) - 2$) and the following conditions hold.*

If f has arity $2n$, then

- *$f_{n-k} = f_{n+k}$ (for $0 \leq k \leq n$), $f_n \neq 0$, $c = \frac{2f_{n-1}}{f_n}$; or*
- *$f_{n-k} = -f_{n+k}$ (for $0 \leq k \leq n$), $f_{n-1} \neq 0$, $c = \frac{f_{n-2}}{f_{n-1}}$.*

If f has arity $2n+1$, then

- *$f_{n-k} = f_{n+1+k}$ (for $0 \leq k \leq n$), $f_n \neq 0$, $c = \frac{f_{n-1}}{f_n} + 1$; or*
- *$f_{n-k} = -f_{n+1+k}$ (for $0 \leq k \leq n$), $f_n \neq 0$, $c = \frac{f_{n-1}}{f_n} - 1$.*

Proof. Symmetric signatures in $\widehat{\mathscr{M}}$ have the following forms, $f = [s,t]^{\otimes m} \pm [t,s]^{\otimes m}$, or $f_k = \lambda(\pm 1)^k(m-2k)$ $(0 \leq k \leq m)$. A Gen-Eq $f \in \widehat{\mathscr{M}}$ iff it takes the first form with $st = 0$. Suppose f is not a Gen-Eq, then we have $st \neq 0$ in the first form. In particular f is not identically zero. In both forms, f satisfies a second order recurrence $f_k - cf_{k+1} + f_{k+2} = 0$ $(0 \leq k \leq m-2)$, for some c. For example in the first form with a tensor sum, the product of the eigenvalues $s/t \cdot t/s = 1$.

For even arity $m = 2n$, and $f = [s,t]^{\otimes 2n} + [t,s]^{\otimes 2n}$, we have the symmetry $f_{n+k} = f_{n-k}$. Thus $f_{n-1} = f_{n+1}$ and $cf_n = 2f_{n-1}$. If $f_n = 0$, then f is identically zero, a contradiction. Therefore, we have $c = \frac{2f_{n-1}}{f_n}$.

For $f = [s,t]^{\otimes 2n} - [t,s]^{\otimes 2n}$, or $f_k = \lambda(\pm 1)^k(2n-2k)$, we have $f_{n+k} = -f_{n-k}$. Thus we have $f_n = 0$ and $cf_{n-1} = f_{n-2}$. If $f_{n-1} = 0$, then f is identically zero, a contradiction. Therefore, we have $c = \frac{f_{n-2}}{f_{n-1}}$.

Conversely, the second order recurrence $f_k - cf_{k+1} + f_{k+2} = 0$ gives the expression $f = c_1[s,t]^{\otimes 2n} + c_2[t,s]^{\otimes 2n}$, or in the double root case when $c = \pm 2$, we have the form $f_k = \lambda(\pm 1)^k(2n - \mu k)$. If $f_{n+k} = -f_{n-k}$, then $f_n = 0$, the double root case must be $f_k = \lambda(\pm 1)^k(2n-2k)$, and the tensor sum takes the form $f = [s,t]^{\otimes 2n} - [t,s]^{\otimes 2n}$. If $f_{n+k} = f_{n-k}$, then we only have the form $f = [s,t]^{\otimes 2n} + [t,s]^{\otimes 2n}$.

For odd arity, the proof is similar. We omit it here. □

(a) (b)

Figure 24: Two gadgets used to obtain $[1, 0, -1]^{\otimes 2}$. The circle vertices are assigned f and the square vertices are assign $=_4$.

Corollary A.17. *If $f \in \widehat{\mathscr{M}^\dagger}$ has even arity $2n$, then for all $0 \le k \le 2n$,*

$$f_k = f_{2n-k} \qquad \text{or} \qquad f_k = -f_{2n-k}$$

and the signs strictly alternate.

Proof. By definition, $\widehat{\mathscr{M}^\dagger} = \left[\begin{smallmatrix} 1 & 0 \\ 0 & i \end{smallmatrix}\right]^{\otimes 2n} \widehat{\mathscr{M}}$. By Lemma A.16, for some $\epsilon = \pm 1$, we have $i^{n-k} f_{n-k} = \epsilon i^{n+k} f_{n+k}$ for all k. The Corollary follows. □

In the proof of Pl-#CSP^2 dichotomy, we often use the following Corollary. It gives a characterization of a signature of arity 4 in $\widehat{\mathscr{M}}$. It follows directly from Lemma A.16 and the definition of $\widehat{\mathscr{M}^\dagger}$.

Corollary A.18. *An arity 4 signature $f \in \widehat{\mathscr{M}}$ has the following forms:*

- $[u, v, w, v, u]$ *and* $(u + w)w = 2v^2$*; or*
- $[u, v, 0, -v, -u]$.

An arity 4 signature $f \in \widehat{\mathscr{M}^\dagger}$ has the following forms:

- $[u, v, w, -v, u]$ *and* $(u - w)w = 2v^2$,
- $[u, v, 0, v, -u]$.

The following lemma can be proved by domain pairing. We can use it to derive #P-hardness of Pl-#CSP^2 problems by applying the known dichotomy of Pl-#CSP.

Lemma A.19. *Suppose $f = [f_0, f_1, \ldots, f_{2n}]$ is a symmetric signature of arity $2n$. Let $g = [f_0, f_2, \ldots, f_{2n}]$ be a symmetric signature of arity n consisting of all even indexed entries of f. Then*

$$\text{Pl-\#CSP}(g) \le \text{Pl-\#CSP}^2(f).$$

Proof. For any instance of Pl-#CSP(g), we replace each edge e by two edges that connect the same incident nodes of e. For each variable node that is connected to k edges, we replace its label $=_k$ by $=_{2k}$. We replace each occurrence of g by f as a constraint. Then the new instance is a problem in Pl-#CSP$^2(f)$ and has the same value as the given instance of Pl-#CSP(g), because $g_k = f_{2k}$. Note that the values f_{2k+1} with an odd index contribute nothing to the partition function in this instance. □

The case when $f = [1, i]^{\otimes 4} + a[1, -i]^{\otimes 4}$ poses some special difficulty, mainly because $\partial(f)$ is identically 0. The following lemma shows that in this case, with $a \neq 0$, we can construct $[1, 0, -1]^{\otimes 2}$ in the LHS in a Pl-Holant problem with f on the RHS. Its utility is that after a holographic transformation by $\left[\begin{smallmatrix} 1 & 0 \\ 0 & i \end{smallmatrix}\right]$ or by $\left[\begin{smallmatrix} 1 & 1 \\ i & -i \end{smallmatrix}\right] = \left[\begin{smallmatrix} 1 & 0 \\ 0 & i \end{smallmatrix}\right] \left[\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}\right]$ we have $[1, 0, 1]^{\otimes 2}$ on the LHS.

Lemma A.20. *Let $\mathcal{F}$ be a set of signatures containing $f = [1, i]^{\otimes 4} + a[1, -i]^{\otimes 4}$. Then*

$$\text{Pl-Holant}([1, 0, -1]^{\otimes 2} \cup \mathcal{EQ}_2 \mid \mathcal{F}) \equiv \text{Pl-}\#\text{CSP}^2(\mathcal{F}).$$

Proof. Suppose $a \neq -1$ and consider the gadget in Figure 24a. We assign f to the circle vertex and $=_4$ to the square vertices. This gives $(1 + a)[1, 0, -1]^{\otimes 2}$ on the left as desired.

Otherwise $a = -1$. Consider the gadget in Figure 24b. We assign f to the circle vertices and $=_4$ to the square vertices. This gives $-8[1, 0, -1]^{\otimes 2}$ on the left as desired. □

Coming up next are a couple of complexity dichotomy theorems that were previously shown. They are also quoted in Section 2 of Part I. Here we restate them for easier reference. The first is a dichotomy theorem about counting complex weighted graph homomorphisms over degree prescribed graphs. It includes Pl-$\#\text{CSP}^2(f)$, where f is a symmetric *binary* signature, as a special case. This is also quoted as Theorem 2.24 in Part I.

Theorem A.21 (Theorem 3 in [8])**.** *Let $S \subseteq \mathbb{Z}^+$ contain $k \geq 3$, let $\mathcal{G} = \{=_k \mid k \in S\}$, and let $d = \gcd(S)$. Further suppose that $f_0, f_1, f_2 \in \mathbb{C}$. Then* Pl-Holant $([f_0, f_1, f_2] \mid \mathcal{G})$ *is* #P*-hard unless one of the following conditions holds:*

1. $f_0 f_2 = f_1^2$;
2. $f_0 = f_2 = 0$;
3. $f_1 = 0$;
4. $f_0 f_2 = -f_1^2$ *and* $f_0^d = -f_2^d \neq 0$;
5. $f_0^d = f_2^d \neq 0$.

In any exceptional case, the problem is computable in polynomial time.

Theorem A.21 is the original statement as in [8]. It is explicit and easy to apply. Conceptually, it can be restated as Theorem A.21′, which supports the putative form of the Pl-$\#\text{CSP}^d$ dichotomy.

Theorem A.21′ (Theorem 3 in [8])**.** *Let $S \subseteq \mathbb{Z}^+$ contain $k \geq 3$, let $\mathcal{G} = \{=_k \mid k \in S\}$, and let $d = \gcd(S)$. Further suppose that f is a non-degenerate, symmetric, complex-valued binary signature in Boolean variables. Then* Pl-Holant $(f \mid \mathcal{G})$ *is* #P*-hard unless f satisfies one of the following conditions, in which case, the problem is computable in polynomial time:*

1. *there exists $T \in \mathcal{T}_{4d}$ such that $T^{\otimes 2} f \in \mathscr{A}$;*
2. *$f \in \mathscr{P}$;*
3. *there exists $T \in \mathcal{T}_{2d}$ such that $T^{\otimes 2} f \in \widehat{\mathscr{M}}$.*

The following theorem is the dichotomy theorem of Pl-$\#\text{CSP}(\mathcal{F})$, where $\mathcal{F}$ is a set of symmetric signatures. This is also quoted as Theorem 2.25 in Part I.

Theorem A.22 (Theorem 19 in [20])**.** *Let $\mathcal{F}$ be any set of symmetric, complex-valued signatures in Boolean variables. Then* Pl-$\#\text{CSP}(\mathcal{F})$ *is* #P*-hard unless $\mathcal{F} \subseteq \mathscr{A}$, $\mathcal{F} \subseteq \mathscr{P}$, or $\mathcal{F} \subseteq \widehat{\mathscr{M}}$, in which case the problem is computable in polynomial time.*

We repeat the definition of redundant matrices in Section 2.7.

Definition A.23 (Definition 6.1 in [6])**.** *A 4-by-4 matrix is* redundant *if its middle two rows and middle two columns are the same.*

An example of a redundant matrix is the signature matrix of a symmetric arity 4 signature.

Definition A.24 (Definition 6.2 in [6]). *The* signature matrix *of a symmetric arity 4 signature* $f = [f_0, f_1, f_2, f_3, f_4]$ *is*

$$M_f = \begin{bmatrix} f_0 & f_1 & f_1 & f_2 \\ f_1 & f_2 & f_2 & f_3 \\ f_1 & f_2 & f_2 & f_3 \\ f_2 & f_3 & f_3 & f_4 \end{bmatrix}.$$

This definition extends to an asymmetric signature g *as*

$$M_g = \begin{bmatrix} g^{0000} & g^{0010} & g^{0001} & g^{0011} \\ g^{0100} & g^{0110} & g^{0101} & g^{0111} \\ g^{1000} & g^{1010} & g^{1001} & g^{1011} \\ g^{1100} & g^{1110} & g^{1101} & g^{1111} \end{bmatrix}.$$

When we present g *as an* $\mathcal{F}$*-gate, we order the four external edges* $ABCD$ *counterclockwise. In* M_g*, the row index bits are ordered* AB *and the column index bits are ordered* DC*, in reverse order. This is for convenience so that the signature matrix of the linking of two arity 4* $\mathcal{F}$*-gates is the matrix product of the signature matrices of the two* $\mathcal{F}$*-gates.*

If M_g *is redundant, we also define the* compressed signature matrix *of* g *as*

$$\widetilde{M_g} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} M_g \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

The definition of *compressed signature matrix* is a slight change from [20] where $\widetilde{M_g}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ is called by that name. It does not affect the following lemma. We repeat the following lemma from [20], which is very convenient to apply.

Lemma A.25 (Corollary 3.8 in [20]). *Let* f *be an arity* 4 *signature with complex weights. If* M_f *is redundant and* $\widetilde{M_f}$ *is nonsingular, then* Pl-Holant(f) *is* #P*-hard.*

B Reduction from Pl-#CSP to Pl-#CSP^2

Definition B.1. *For* $k \geq 1$*,* $\ell \geq 0$ *and any* ω*, we define* $E_k^\ell(\omega) = [1, 0, \ldots, 0, \omega^\ell]$ *to be a signature of arity* k*, and define* $E(\omega) = \{E_k^\ell(\omega) \mid k \equiv \ell \pmod 2\}$*. We also write* E_k^ℓ *for* $E_k^\ell(\omega)$ *when* ω *is clear from the context.*

The following lemma shows that if we have a unary $[1, \omega] \in \mathcal{F}$ with $\omega \neq 0$, then either $\mathcal{F}$ is contained in one single tractable set or Pl-#CSP$^2(\mathcal{F})$ is #P-hard. We will use this lemma for the case that $\mathcal{F}$ contains at least one nonzero signature of odd arity. The proof of this lemma also demonstrates in a simple setting the idea that will be used in the proof of Lemma B.3.

Lemma B.2. *Let* $\omega \neq 0$ *and let* $\mathcal{F}$ *be a set of symmetric signatures containing* $[1, \omega] \in \mathcal{F}$*. If* $\mathcal{F} \not\subseteq \mathscr{P}$*,* $\mathcal{F} \not\subseteq \mathscr{A}$*,* $\mathcal{F} \not\subseteq \mathscr{A}^\dagger$*,* $\mathcal{F} \not\subseteq \widehat{\mathscr{M}}$*, and* $\mathcal{F} \not\subseteq \widehat{\mathscr{M}^\dagger}$*, then* Pl-#CSP$^2(\mathcal{F})$ *is* #P*-hard.*

Proof. Firstly, we have $E_k^k(\omega) = \partial^k_{[1,\omega]}(=_{2k})$ of arity k on the LHS in Pl-#CSP$^2(\mathcal{F})$, for all $k \geq 1$. By a holographic transformation using T^{-1}, where $T = \left[\begin{smallmatrix} 1 & 0 \\ 0 & \omega \end{smallmatrix}\right]$, we have $(E_k^k(\omega))T^{-1} = (=_k)$ on the LHS, and

$$\text{Pl-\#CSP}(T\mathcal{F}) \leq_T \text{Pl-Holant}(\mathcal{EQ} \cup \mathcal{EQ}_2 T^{-1} \mid T\mathcal{F}) \leq_T \text{Pl-\#CSP}^2(\mathcal{F}),$$

where $\mathcal{EQ}$ on LHS of the Holant instance comes from $E_k^k(\omega)$ in the second step of the reduction. If $T\mathcal{F} \nsubseteq \mathscr{P}$, $T\mathcal{F} \nsubseteq \mathscr{A}$ and $T\mathcal{F} \nsubseteq \widehat{\mathscr{M}}$, then Pl-#CSP$(T\mathcal{F})$ is #P-hard by Theorem A.22. Thus Pl-#CSP$^2(\mathcal{F})$ is #P-hard.

Otherwise, $T\mathcal{F} \subseteq \mathscr{P}$, $T\mathcal{F} \subseteq \mathscr{A}$ or $T\mathcal{F} \subseteq \widehat{\mathscr{M}}$. If $T\mathcal{F} \subseteq \mathscr{P}$, then $\mathcal{F} \subseteq \mathscr{P}$. In the following, assume that $T\mathcal{F} \nsubseteq \mathscr{P}$, then $T\mathcal{F} \subseteq \mathscr{A}$ or $T\mathcal{F} \subseteq \widehat{\mathscr{M}}$.

Note that $[1, \omega^2] \in T\mathcal{F}$. If $\omega^8 \neq 1$, then $[1, \omega^2] \notin \mathscr{A} \cup \widehat{\mathscr{M}}$. This is a contradiction.

If $\omega^4 = -1$, then $[1, \omega^2] \notin \widehat{\mathscr{M}}$. Thus $T\mathcal{F} \subseteq \mathscr{A}$. It follows that $\mathcal{F} \subseteq \mathscr{A}^\dagger$.

For $\omega^4 = 1$, if $T\mathcal{F} \subseteq \mathscr{A}$, then $\mathcal{F} \subseteq \mathscr{A}$. If $T\mathcal{F} \subseteq \widehat{\mathscr{M}}$, then either $\mathcal{F} \subseteq \widehat{\mathscr{M}}$ if $\omega^2 = 1$, or $\mathcal{F} \subseteq \widehat{\mathscr{M}}^\dagger$ if $\omega^2 = -1$. □

Lemma B.2 allows us to transfer the complexity question of Pl-#CSP^2 to that of Pl-#CSP, to which we can apply the known dichotomy (Theorem A.22). However it requires a unary signature. We observe that if all signatures in $\mathcal{F}$ have even arities, then there is no way to construct a unary in Pl-#CSP$^2(\mathcal{F})$. In this case, we use the next lemma, which is similar to Lemma B.2. It shows that if we have $[1, \omega]^{\otimes 2}$ with $\omega \neq 0$ in $\mathcal{F}$, then we can still transfer the question of Pl-#CSP^2 to that of Pl-#CSP. It is proved using a global simulation of Pl-#CSP by Pl-#CSP^2.

Lemma B.3. *Let $\mathcal{F}$ be a set of signatures of even arities. Suppose $[1, \omega]^{\otimes 2} \in \mathcal{F}$ for some $\omega \neq 0$. If $\mathcal{F} \nsubseteq \mathscr{P}$, $\mathcal{F} \nsubseteq \mathscr{A}$, $\mathcal{F} \nsubseteq \mathscr{A}^\dagger$, $\mathcal{F} \nsubseteq \widehat{\mathscr{M}}$ and $\mathcal{F} \nsubseteq \widehat{\mathscr{M}}^\dagger$, then* Pl-#CSP$^2(\mathcal{F})$ *is* #P*-hard.*

Proof. We first prove that Pl-Holant$(E(\omega) \mid \mathcal{F}) \leq_T$ Pl-#CSP$^2(\mathcal{F})$.

For $k \geq 1$ and $\ell \geq 0$, we have all of $E_{2k}^{2\ell} = E_{2k}^{2\ell}(\omega) = \partial^\ell_{[1,\omega]^{\otimes 2}}(=_{2k+2\ell})$ on LHS in Pl-#CSP$^2(\mathcal{F})$. Given any instance Ω of Pl-Holant$(E(\omega) \mid \mathcal{F})$, since all signatures in $\mathcal{F}$ have even arities, the number of E_k^ℓ of odd arity must be even. In each connected component of Ω, we can connect all E_k^ℓ of odd arity in pairs, by some copies of $[1, \omega]^{\otimes 2}$ in a planar way. Note that when one input of E_k^ℓ is connected to a unary $[1, \omega]$, it becomes $E_{k-1}^{\ell+1}$. Hence a pair E_{2v-1}^{2u+1} and $E_{2v'-1}^{2u'+1}$ can be functionally replaced by a pair E_{2v}^{2u} and $E_{2v'}^{2u'}$ that are connected by $[1, \omega]^{\otimes 2}$.

Formally, we may assume the plane graph Ω is connected, since the Holant value on Ω is the product over its connected components, and the number of $E_k^\ell \in E(\omega)$ of odd arity is even in each connected component of Ω. We will connect pairs of E_k^ℓ of odd arity by copies of $[1, \omega]^{\otimes 2}$ within each connected component.

Let T be a spanning tree of the dual graph of Ω, and pick any node as the root of T. For definiteness we pick the node of T that corresponds to the external face of Ω as root. If on a leaf node of T, i.e., a face of Ω, there are an even number of E_k^ℓ of odd arity, we can connect them in pairs within the face by copies of $[1, \omega]^{\otimes 2}$, maintaining planarity. If there are an odd number of them, we can pick any one, and still connect the others in pairs within the face by copies of $[1, \omega]^{\otimes 2}$, maintaining planarity. On the edge connecting the leaf to its parent in the tree T, the corresponding edge in Ω has an E_s^t in one of the two incident nodes of Ω. If s is odd, we pick this E_s^t. If s is even, we pick the first E_k^ℓ of odd arity in clockwise order in the face of Ω, which is the leaf node in T, and connect it to that E_s^t by one copy of $[1, \omega]^{\otimes 2}$. This effectively transforms that E_s^t to E_{s-1}^{t+1} of odd arity. We then delete the leaf node from T.

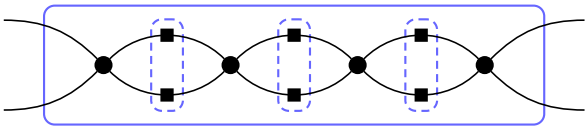


Figure 25: Gadget used to obtain $=_4$. The circle vertices are assigned $\hat{f}$ and the dashed subgadgets are assigned $[1,0,1]^{\otimes 2}$ aligned horizontally so that it is equivalent to assigning $[1,0,1]$ to the square vertices.

The proof is completed by induction. Note that at the root of T, there must be an even number of E_k^ℓ of odd arity, including those which have been transformed by its children in T. Thus we can simulate the Pl-Holant$(E(\omega) \mid \mathcal{F})$ problem Ω by Pl-#CSP$^2(\mathcal{F})$.

Note that $E_k^k \in E(\omega)$, for all $k \geq 1$. Thus we have

$$\text{Pl-Holant}(E_1^1, E_2^2, \ldots, E_k^k, \ldots \mid \mathcal{F}) \leq_T \text{Pl-Holant}(E(\omega) \mid \mathcal{F}) \leq_T \text{Pl-\#CSP}^2(\mathcal{F}).$$

Then by a holographic transformation using $T = \left[\begin{smallmatrix} 1 & 0 \\ 0 & \omega^{-1} \end{smallmatrix}\right]$, we have

$$\text{Pl-\#CSP}(T^{-1}\mathcal{F}) \equiv \text{Pl-Holant}(E_1^1, E_2^2, \ldots, E_k^k, \ldots \mid \mathcal{F}) \leq_T \text{Pl-\#CSP}^2(\mathcal{F}).$$

The rest of the proof is the same as the proof of Lemma B.2. We omit it here. □

The next lemma shows that when we obtain $[1,0,1]^{\otimes 2}$, we can reduce a Pl-#CSP^2 problem to a Pl-#CSP^4 problem, when all signatures in $\mathcal{F}$ have arity divisible by 4.

Lemma B.4. Pl-#CSP$^2(\mathcal{F}) \leq_T$ Pl-#CSP$^4(\mathcal{F}, [1,0,1]^{\otimes 2})$, *if all signatures in* $\mathcal{F}$ *have arity* $\equiv 0 \pmod 4$.

Proof. Let Ω be an instance of Pl-#CSP$^2(\mathcal{F})$. Since all signatures in $\mathcal{F}$ have arity $\equiv 0 \pmod 4$, the number of EQUALITIES of arity $\equiv 2 \pmod 4$ must be even. We can connect in pairs all EQUALITIES of arity $\equiv 2 \pmod 4$ by some copies of $[1,0,1]^{\otimes 2}$ maintaining planarity similarly as in the proof of Lemma B.3. When two inputs of $=_{m+2}$ are connected to $[1,0,1]$ it becomes $\partial(=_{m+2}) = (=_m)$. Hence a pair $=_{4k-2}$ and $=_{4\ell-2}$ can be functionally replaced by a pair $=_{4k}$ and $=_{4\ell}$ that are connected by $[1,0,1]^{\otimes 2}$. The rest of the proof is the same as in Lemma B.3 and we omit it here. □

The next corollary is used in the proof of the No-Mixing theorems. We present it here since the proof uses a global simulation that is similar to Lemma B.4.

Corollary B.5. *Suppose* $f = [1,i]^{\otimes 4} + i^r[1,-i]^{\otimes 4}$ $(0 \leq r \leq 3)$ *and* $g = [g_0, \ldots, g_{2n}]$ *with* $g_k = (\pm i)^k(2n-2k)$. *Furthermore, let* $\hat{g} = (Z^{-1})^{\otimes 2n} g$, *where* $Z = \left[\begin{smallmatrix} 1 & 1 \\ i & -i \end{smallmatrix}\right]$. *Then*

$$\text{Pl-\#CSP}^2(\hat{g}) \leq_T \text{Pl-\#CSP}^2(f, g).$$

Proof. Clearly $\hat{g} = [0,1,0,\ldots,0]$ or $\hat{g} = [0,\ldots,0,1,0]$, the perfect matching signature or its reversal. By applying Lemma A.20 to $f = [1,i]^{\otimes 4} + i^r[1,-i]^{\otimes 4}$, we get $[1,0,-1]^{\otimes 2}$ on the left:

$$\text{Pl-Holant}([1,0,-1]^{\otimes 2} \cup \mathcal{EQ}_2 \mid f, g) \leq_T \text{Pl-\#CSP}^2(f, g).$$

Under a holographic transformation by Z, we have

$$\text{Pl-Holant}([1,0,1]^{\otimes 2} \mid \hat{f}, \hat{g}) \leq_T \text{Pl-Holant}([1,0,-1]^{\otimes 2} \cup \mathcal{EQ}_2 \mid f, g),$$

where $\hat{f} = (Z^{-1})^{\otimes 4} f = [1, 0, 0, 0, i^r]$. Note that $[1, 0, -1]Z^{\otimes 2} = 2[1, 0, 1]$, as $Z^{\mathrm{T}} \left[\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right] Z = 2 \left[\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right]$. Consider the gadget in Figure 25. We assign $\hat{f}$ to the circle vertices and $[1, 0, 1]^{\otimes 2}$ the dashed subgadgets rotated appropriately so that it is equivalent to assigning $[1, 0, 1]$ to the square vertices. The signature of this gadget is $=_4$, for any $0 \le r \le 3$. Thus

$$\text{Pl-Holant}([1, 0, 1]^{\otimes 2} \mid =_4, \hat{g}) \le_T \text{Pl-Holant}([1, 0, 1]^{\otimes 2} \mid \hat{f}, \hat{g}).$$

In Pl-Holant$([1, 0, 1]^{\otimes 2} \mid =_4, \hat{g})$, by $[1, 0, 1]^{\otimes 2}$ and $=_4$, we can get all of $=_{4k}$ for $k \ge 1$ on RHS and then move them to LHS by $[1, 0, 1]^{\otimes 2}$. Moreover, we have $[1, 0, 1]^{\otimes 2}$ on RHS by connecting two copies of $=_4$ by $[1, 0, 1]^{\otimes 2}$. Thus

$$\text{Pl-Holant}(\mathcal{EQ}_4 \mid [1, 0, 1]^{\otimes 2}, \hat{g}) \le_T \text{Pl-Holant}([1, 0, 1]^{\otimes 2} \mid =_4, \hat{g}).$$

Now we simulate Pl-#CSP$^2(\hat{g})$ by Pl-Holant$(\mathcal{EQ}_4 \mid [1, 0, 1]^{\otimes 2}, \hat{g})$. If $\hat{g}$ has arity $2n \equiv 0 \pmod 4$, then we are done by Lemma B.4.

If $\hat{g}$ has arity $2n \equiv 2 \pmod 4$, then in an instance Ω of Pl-#CSP$^2(\hat{g})$, the number of occurrences of EQUALITIES of arity $\equiv 2 \pmod 4$ has the same parity as the number of occurrences of $\hat{g}$, which could be odd. However, we observe that all entries of signatures in Pl-#CSP$^2(\hat{g})$ are nonnegative integers. Thus the value of Ω is a nonnegative integer. Let $\Omega \uplus \Omega$ be the disjoint union of two copies of Ω as a plane graph with a common external face, then the value of $\Omega \uplus \Omega$ is the square of the value of Ω. Thus computing the values on $\Omega \uplus \Omega$ and Ω are equivalent. In $\Omega \uplus \Omega$, the number of EQUALITIES of arity $\equiv 2 \pmod 4$ is even. Now we can use the same global simulation as in Lemma B.4, except that in the last step we may use one extra copy of $[1, 0, 1]^{\otimes 2}$ to connect two EQUALITIES of arity $\equiv 2 \pmod 4$ at the two root nodes of the two spanning trees of the dual graphs of Ω, if the number of occurrences of EQUALITIES of arity $\equiv 2 \pmod 4$ in Ω is odd. Thus we have

$$\text{Pl-\#CSP}^2(\hat{g}) \le \text{Pl-Holant}(\mathcal{EQ}_4 \mid [1, 0, 1]^{\otimes 2}, \hat{g}). \qquad \square$$

C Dichotomy Theorem when $\mathcal{F}$ Contains an Odd Arity Signature

In this section, we give a dichotomy theorem for Pl-#CSP$^2(\mathcal{F})$, where $\mathcal{F}$ includes at least one nonzero signature f that has odd arity.

The next result is similar to Lemma 6.2 in [20].

Lemma C.1. *Let $x, y \in \mathbb{C}$ and $f = [x, 0, y, 0]$. If $y \ne 0$ and $x^4 \ne y^4$, then* Pl-#CSP$^2(f)$ *is #P-hard.*

Proof. We reduce from Pl-#CSP$([x^2, y^2, y^2])$ to Pl-#CSP$^2(f)$. Since Pl-#CSP$([x^2, y^2, y^2])$ is #P-hard when $y \ne 0$ and $x^4 \ne y^4$ by Theorem A.21, this shows that Pl-#CSP$^2(f)$ is also #P-hard.

An instance of Pl-#CSP$([x^2, y^2, y^2])$ is a signature grid Ω with underlying graph $G = (U, V, E)$, where G is bipartite and planar, and every vertex in U has degree 2. We replace every vertex in V of degree k (which is assigned $=_k \in \mathcal{EQ}$) with a vertex of degree $2k$, and bundle two adjacent variables to form k bundles of 2 edges each. The k bundles correspond to the k incident edges of the original vertex with degree k. We assign $=_{2k}$ to the new vertices of degree $2k$.

If the inputs to these equality signatures are restricted to $\{(0, 0), (1, 1)\}$ on each bundle, then these equality signatures take value 1 on $((0, 0), \ldots, (0, 0))$ and $((1, 1), \ldots, (1, 1))$ and take value 0 elsewhere. Thus, if we restrict the domain to $\{(0, 0), (1, 1)\}$, it is the equality signature $=_k$.

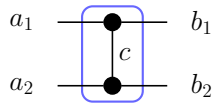


Figure 26: Gadget designed for the paired domain. Both vertices are assigned $[x, 0, y, 0]$.

To simulate $[x^2, y^2, y^2]$, we connect two copies of $f = [x, 0, y, 0]$ by a single edge as shown in Figure 26 to form a gadget with signature

$$h(a_1, a_2, b_1, b_2) = \sum_{c=0,1} f(a_1, b_1, c) f(a_2, b_2, c).$$

We replace every (degree 2) vertex in U (which is assigned $[x^2, y^2, y^2]$) by a degree 4 vertex assigned h, where the variables of h are bundled as (a_1, a_2) and (b_1, b_2).

The vertices in this new graph G' are connected as in the original graph G, except that every original edge is replaced by two edges that connect to the same side of the gadget in Figure 26. Notice that h is only connected by (a_1, a_2) and (b_1, b_2) to some bundle of two incident edges of an equality signature. Since this equality signature enforces that the value on each bundle is either $(0, 0)$ or $(1, 1)$, we only need to consider the restriction of h to the domain $\{(0, 0), (1, 1)\}$. On this domain, $h = [x^2, y^2, y^2]$ is a *symmetric* signature of arity 2. Therefore, the signature grid Ω' with underlying graph G' has the same Holant value as the original signature grid Ω. □

The following lemma is a dichotomy for Pl-#CSP$^2(f)$ where f is a symmetric ternary signature.

Lemma C.2. *Let f be a symmetric signature of arity 3, then* Pl-#CSP$^2(f)$ *is #P-hard unless $f \in \mathscr{P} \cup \widetilde{\mathscr{A}} \cup \widetilde{\mathscr{M}}$.*

Proof. Let $f = [f_0, f_1, f_2, f_3]$. If f satisfies parity constraints, then $f = [f_0, 0, f_2, 0]$ or $f = [0, f_1, 0, f_3]$.

For $f = [f_0, 0, f_2, 0]$, if $f_2 = 0$, then $f \in \mathscr{P}$. If $f_0^2 = f_2^2$, then $f \in \mathscr{A}$. If $f_0^2 = -f_2^2$, then $f \in \mathscr{A}^\dagger$. Otherwise, we have $f_2 \neq 0$ and $f_0^4 \neq f_2^4$. Thus Pl-#CSP$^2(f)$ is #P-hard by Lemma C.1. For $f = [0, f_1, 0, f_3]$, the proof follows from a holographic transformation using $\left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right]$.

In the following, assume that f does not satisfy parity constraints. Firstly, we have $\partial(f) = [f_0 + f_2, f_1 + f_3]$.

- For $(f_0 + f_2)(f_1 + f_3) \neq 0$, we are done by Lemma B.2.
- For $f_0 + f_2 = f_1 + f_3 = 0$, $f = [f_0, f_1, -f_0, -f_1]$. Since f does not satisfy parity constraints, we have $f_0 f_1 \neq 0$. If $f_0^2 = f_1^2$, then $f \in \mathscr{A}$.
  Otherwise, we have $\partial_f(=_4) = [f_0, -f_1]$ on LHS and $\partial_{[f_0, -f_1]}(f) = [f_0^2 - f_1^2, 2f_0 f_1, f_1^2 - f_0^2]$ on RHS. Moreover, we have $\partial_{[f_0^2 - f_1^2, 2f_0 f_1, f_1^2 - f_0^2]}(=_4) = (f_0^2 - f_1^2)[1, 0, -1]$ on LHS, where $f_0^2 - f_1^2 \neq 0$. So we have $\partial_{[1,0,-1]}(f) = 2[f_0, f_1]$ on RHS. Then we are done by Lemma B.2 and $f_0 f_1 \neq 0$.
- For $f_0 + f_2 \neq 0$, $f_1 + f_3 = 0$, we have $f_1 = -f_3 \neq 0$ since f does not satisfy parity constraints. Note that we have $\partial(f) = (f_0 + f_2)[1, 0]$ in RHS, so we have $\partial^2_{[1,0]}(f) = [f_0, f_1]$ in RHS. If $f_0 \neq 0$, then we are done by Lemma B.2. If $f_0 = 0$, then $f_2 \neq 0$ since $f_0 + f_2 \neq 0$. Note that we have $f_1[0, 1]$ and $f_2[1, 0]$ now. Thus we have $\partial_{[1,0]}[\partial_{[0,1]}(f)] = [f_1, f_2]$. Then we are done by Lemma B.2.
- For $f_0 + f_2 = 0, f_1 + f_3 \neq 0$, the proof follows from a holographic transformation using $\left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right]$.

□

The next lemma shows that if we have an odd arity signature in $\widetilde{\mathscr{M}} \setminus (\mathscr{P} \cup \widetilde{\mathscr{A}})$, then we can prove Theorem A.2 directly. The key point is that we can use such a signature to get a unary $[1, \omega]$ with $\omega \neq 0$.

Lemma C.3. *Let $\mathcal{F}$ be a symmetric signature set and $f \in \mathcal{F}$ has odd arity.*

- *If $f \in \widehat{\mathscr{M}} \setminus (\mathscr{P} \cup \widetilde{\mathscr{A}})$, then either $\mathcal{F} \subseteq \widehat{\mathscr{M}}$ or* Pl-#CSP$^2(\mathcal{F})$ *is #P-hard.*
- *If $f \in \widehat{\mathscr{M}^\dagger} \setminus (\mathscr{P} \cup \widetilde{\mathscr{A}})$, then either $\mathcal{F} \subseteq \widehat{\mathscr{M}^\dagger}$ or* Pl-#CSP$^2(\mathcal{F})$ *is #P-hard.*

Proof. We will use our calculus with the derivative operator ∂. Firstly, we prove the lemma for $f \in \widehat{\mathscr{M}} \setminus (\mathscr{P} \cup \widetilde{\mathscr{A}})$. We already have $\mathcal{F} \nsubseteq \mathscr{P}$, $\mathcal{F} \nsubseteq \mathscr{A}$, $\mathcal{F} \nsubseteq \mathscr{A}^\dagger$ by the presence of f, and $\mathcal{F} \nsubseteq \widehat{\mathscr{M}^\dagger}$ by Corollary A.13. If we can construct a unary $[a, b]$ with $ab \neq 0$, then we can finish the proof by Lemma B.2.

As $f \notin \mathscr{P}$ and has odd arity, its arity $n \geq 3$. By Lemma A.14, the signature $f \in \widehat{\mathscr{M}} \setminus (\mathscr{P} \cup \widetilde{\mathscr{A}})$ can take one of the following two forms (see the Calculus after Definition A.4):

- For $f = [s, t]^{\otimes n} \pm [t, s]^{\otimes n}$, where $n \geq 3$ is odd, we have $st \neq 0$ and $s^4 \neq t^4$. Thus we have $\partial^{\frac{n-1}{2}}(f) = (s^2+t^2)^{\frac{n-1}{2}}([s, t] \pm [t, s]) = (s^2+t^2)^{\frac{n-1}{2}}(s \pm t)[1, \pm 1]$, a nonzero multiple of $[1, \pm 1]$. So we are done by Lemma B.2.
- For $f_k = \lambda(\pm 1)^k(n - 2k)$, we have $\partial^{\frac{n-1}{2}}(f) = 2^{\frac{n-1}{2}}\lambda[1, \mp 1]$ and we are done by Lemma B.2.

Similarly, for $f \in \widehat{\mathscr{M}^\dagger} \setminus (\mathscr{P} \cup \widetilde{\mathscr{A}})$, we just need to construct a unary $[a, b]$ with $ab \neq 0$.

- For $f = [s, ti]^{\otimes n} \pm [t, si]^{\otimes n}$, we have $\partial^{\frac{n-1}{2}}(f) = (s^2 - t^2)^{\frac{n-1}{2}}[s, ti] \pm (t^2 - s^2)^{\frac{n-1}{2}}[t, si] = (s^2 - t^2)^{\frac{n-1}{2}}(s \pm t)[1, \pm i]$. By Lemma A.14, we have $st \neq 0$ and $s^4 \neq t^4$, and so this is a nonzero multiple of $[1, \pm i]$. So we are done by Lemma B.2.
- For $f_k = \lambda(\pm i)^k(n - 2k)$, if $n \equiv 1 \pmod 4$, we have $\partial_{=_4}^{\frac{n-1}{4}}(f) = 2^{\frac{n-1}{4}}\lambda[1, \mp i]$ and we are done by Lemma B.2. If $n \equiv 3 \pmod 4$, we have $\partial[\partial_{=_4}^{\frac{n-3}{4}}(f)] = 2^{\frac{n+5}{4}}\lambda[1, \pm i]$ and we are done by Lemma B.2. □

We remark that the use of $\partial_{=_4}$ instead of just ∂ in this proof is necessary, because $\partial^2(f) = 0$ when $f_k = \lambda(\pm i)^k(n - 2k)$, and $n \geq 5$. One may also suppose that the case for $\widehat{\mathscr{M}^\dagger} \setminus (\mathscr{P} \cup \widetilde{\mathscr{A}})$ can be reduced to the case for $\widehat{\mathscr{M}} \setminus (\mathscr{P} \cup \widetilde{\mathscr{A}})$ by the transformation $T = \left[\begin{smallmatrix} 1 & 0 \\ 0 & i \end{smallmatrix}\right]$. While T transforms $\widehat{\mathscr{M}^\dagger}$ to $\widehat{\mathscr{M}}$, and keeps $\mathscr{P}$ and $\widetilde{\mathscr{A}}$ invariant, this transformation does not keep $\mathcal{EQ}_2$ invariant. In fact $[1, 0, 1]T^{\otimes 2} = [1, 0, -1] \notin \mathcal{EQ}_2$. Therefore we need to handle the proof for $\widehat{\mathscr{M}^\dagger} \setminus (\mathscr{P} \cup \widetilde{\mathscr{A}})$ separately.

By definitions of $\mathscr{P}$ and $\widetilde{\mathscr{A}}$, we have the following simple lemma.

Lemma C.4. *If $f \in \mathscr{P} \cup \widetilde{\mathscr{A}}$, then f satisfies parity constraints iff f belongs to the following set, up to a scalar factor*

$$\left\{[1, 0]^{\otimes n}, [0, 1]^{\otimes n}, [1, 0]^{\otimes 2n} + t[0, 1]^{\otimes 2n}, [1, \rho]^{\otimes n} \pm [1, -\rho]^{\otimes n}, [1, \alpha]^{\otimes n} \pm [1, -\alpha]^{\otimes n} \;\middle|\; t \neq 0, n \geq 1\right\}.$$

The next lemma shows that if we have a nonzero odd arity signature $f \in \mathscr{P} \cup \widetilde{\mathscr{A}}$ that does not satisfy parity constraints, then we can obtain a unary $[a, b]$ with $ab \neq 0$. Note that if we have a unary $[a, b]$ with $ab \neq 0$, then we can apply Lemma B.2.

Lemma C.5. *If $f \in \mathscr{P} \cup \widetilde{\mathscr{A}}$ has odd arity and does not satisfy parity constraints, then we can construct a unary $[a, b]$ with $ab \neq 0$ in* Pl-#CSP$^2(f)$.

Proof. Let f have arity $2n+1$, $n \geq 0$. Not satisfying parity constraints implies that f is not identically 0. Up to a nonzero factor, f has the following forms.

If $f \in \mathscr{P}$, then $f = [a,b]^{\otimes 2n+1}$ with $ab \neq 0$ or $f = [1,0,\ldots,0,x]$ with $x \neq 0$.

- If $f = [1,0,\ldots,0,x], x \neq 0$, then $\partial^n(f) = [1,x]$.
- If $f = [a,b]^{\otimes 2n+1}$, $a^2+b^2 \neq 0$, then $\partial^n(f) = (a^2+b^2)^n[a,b]$.
- For $f = [1,\pm i]^{\otimes 2n+1}$, if n is even, then $\partial^{\frac{n}{2}}_{=_4}(f) = 2^{\frac{n}{2}}[1,\pm i]$. If n is odd, then we have $\partial_f(=_{2n+2}) = [1,\mp i]$ on LHS and we have $\partial^{2n}_{[1,\mp i]}(f) = 2^{2n}[1,\pm i]$ on RHS.

For $f \in \widetilde{\mathscr{A}} \setminus \mathscr{P}$, we have $f = [1,\rho]^{\otimes 2n+1} \pm i[1,-\rho]^{\otimes 2n+1}$ or $f = [1,\alpha]^{\otimes 2n+1} \pm i[1,-\alpha]^{\otimes 2n+1}$.

- If $f = [1,\alpha]^{\otimes 2n+1} \pm i[1,-\alpha]^{\otimes 2n+1}$, then $\partial^n(f) = (1+\alpha^2)^n[1 \pm i, (1 \mp i)\alpha]$.
- If $f = [1,\rho]^{\otimes 2n+1} \pm i[1,-\rho]^{\otimes 2n+1}$ with $\rho^2 = 1$, then $\partial^n(f) = (1+\rho^2)^n[1 \pm i, (1 \mp i)\rho]$.
- For $f = [1,\rho]^{\otimes 2n+1} \pm i[1,-\rho]^{\otimes 2n+1}$ with $\rho^2 = -1$, and if n is even, then we have $\partial^{\frac{n}{2}}_{=_4}(f) = 2^{\frac{n}{2}}[1 \pm i, (1 \mp i)\rho]$. If n is odd, then $2n+1 \equiv 3 \pmod 4$, and $(\pm\rho)^{2n+1} = \pm\rho^3 = \mp\rho$, by $\rho^2 = -1$. Then we have $\partial_f(=_{2n+2}) = [1,\rho^{2n+1}] \pm i[1,(-\rho)^{2n+1}] = [1,-\rho] \pm i[1,\rho] = (1 \pm i)[1, \pm i\rho]$ on LHS. Note that $(\frac{1 \mp i}{1 \pm i})^{2n} = (\mp i)^{2n} = (-1)^n = -1$ since n is odd.
  Then we have $\partial^{2n}_{[1,\pm i\rho]}(f) = (1 \mp i)^{2n}[1,\rho] \pm i(1 \pm i)^{2n}[1,-\rho] = (1 \pm i)^{2n}\{(\frac{1 \mp i}{1 \pm i})^{2n}[1,\rho] \pm i[1,-\rho]\} = (1 \pm i)^{2n}\{-[1,\rho] \pm i[1,-\rho]\} = (1 \pm i)^{2n}[-1 \pm i, -\rho(1 \pm i)]$. □

If a signature f satisfies parity constraints, then there is no way to construct $[a,b]$ with $ab \neq 0$ from f. In fact in Pl-#CSP^2 using f, the signature of any $\{f\}$-gate will also satisfy the parity constraints, and in particular for unary signature, it can only be a multiple of $[1,0]$ or $[0,1]$. The next lemma shows that if we have a nonzero odd arity signature $f \in \mathscr{P} \cup \widetilde{\mathscr{A}}$ that satisfies parity constraints, then we can obtain $[1,0]$ or $[0,1]$. We also remark that in Pl-#CSP^2 using signatures of even arity one can only produce signatures of even arity, and thus no unary signatures.

Lemma C.6. *If a nonzero $f \in \mathscr{P} \cup \widetilde{\mathscr{A}}$ has odd arity and satisfies parity constraints, then we can construct a unary $[1,0]$ or $[0,1]$ in* Pl-#CSP$^2(f)$.

Proof. By Lemma C.4, an nonzero f of odd arity belongs to the following set, up to a nonzero factor,

$$\left\{[1,0]^{\otimes 2n+1}, [0,1]^{\otimes 2n+1}, [1,\rho]^{\otimes 2n+1} \pm [1,-\rho]^{\otimes 2n+1}], [1,\alpha]^{\otimes 2n+1} \pm [1,-\alpha]^{\otimes 2n+1}] \mid n \geq 0\right\}.$$

For $f = [1,0]^{\otimes 2n+1}$ or $f = [0,1]^{\otimes 2n+1}$ we have $\partial^n(f) = [1,0]$ or $[0,1]$ respectively.

For $f = [1,\alpha]^{\otimes 2n+1} \pm [1,-\alpha]^{\otimes 2n+1}$, $\partial^n(f) = (1+\alpha^2)^n[1 \pm 1, (1 \mp 1)\alpha]$, a nonzero multiple of $[1,0]$ or $[0,1]$.

For $f = [1,\rho]^{\otimes 2n+1} \pm [1,-\rho]^{\otimes 2n+1}$ with $\rho^2 = 1$, $\partial^n(f) = (1+\rho^2)^n[1 \pm 1, (1 \mp 1)\rho]$, a nonzero multiple of $[1,0]$ or $[0,1]$.

For $f = [1,\rho]^{\otimes 2n+1} \pm [1,-\rho]^{\otimes 2n+1}$, with $\rho^2 = -1$, if $2n+1 \equiv 1 \pmod 4$, then $\partial^{\frac{n}{2}}_{=_4}(f) = 2^{\frac{n}{2}}[1 \pm 1, (1 \mp 1)\rho]$, a nonzero multiple of $[1,0]$ or $[0,1]$. If $2n+1 \equiv 3 \pmod 4$, then $(\pm\rho)^{2n+1} = \mp\rho$. If we write $f = [f_0, f_1, \ldots, f_{2n+1}]$, then exactly one of f_0 and f_{2n+1} is nonzero. We have the unary $u = \partial_f(=_{2n+2}) = [f_0, f_{2n+1}]$ in LHS, a nonzero multiple of $[1,0]$ or $[0,1]$. Then we get $\partial^{2n}_u(f)$ in RHS, also a nonzero multiple of $[1,0]$ or $[0,1]$. □

The next lemma shows that if we already have $[1,0]$ or $[0,1]$ and also a signature f of *any* arity that does not satisfy the parity constraints, then we can construct a unary $[a,b]$ with $ab \neq 0$.

Lemma C.7. *If f does not satisfy the parity constraints, then we can construct a unary $[a, b]$ with $ab \neq 0$ in* Pl-#CSP$^2([1, 0], f)$ *or* Pl-#CSP$^2([0, 1], f)$.

Proof. We prove the lemma for Pl-#CSP$^2([1, 0], f)$. The proof for the other case follows from a holographic transformation by $\left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right]$.

Let $f = [f_0, f_1, \ldots, f_n]$. Since f does not satisfy the parity constraints, there exist $0 \leq i < j \leq n$ such that $[f_i, f_{i+1}, \ldots, f_{j-1}, f_j] = [f_i, 0, \ldots, 0, f_j]$, where $f_i f_j \neq 0$ and $j - i$ is odd. We can get both $f' = \partial^{n-i}_{[1,0]} = [f_0, f_1, \ldots, f_i]$ and $f'' = \partial^{n-j}_{[1,0]} = [f_0, f_1, \ldots, f_j]$ on RHS. Either i or j is odd. And so we have either $=_{i+1}$ or $=_{j+1}$, and we can get either $\partial_{f'}(=_{i+1}) = [f_0, f_i]$ or $\partial_{f''}(=_{j+1}) = [f_0, f_j]$ on LHS. Without loss of generality, assume that we have $[f_0, f_i]$ on LHS.

If $f_0 = 0$, then we have $[0, 1]$ on LHS and $f''' = \partial^{\frac{j-i-1}{2}}(\partial^i_{[0,1]}(f'')) = \partial^{\frac{j-i-1}{2}}([f_i, 0, \ldots, 0, f_j]) = [f_i, f_j]$ on RHS, and we are done.

If $f_0 \neq 0$, let $m = \min_{1 \leq k \leq n} \{k \mid f_k \neq 0\}$. (As $j > 0$ and $f_j \neq 0$, this m is well-defined.) Then $f^{(4)} = \partial^{n-m}_{[1,0]}(f) = [f_0, 0, \ldots, 0, f_m]$. Moreover, we have $\partial^{m-1}_{[f_0,f_i]}(f^{(4)}) = [f_0^m, f_i^{m-1} f_m]$. □

The next lemma assumes the presence of a non-degenerate binary Gen-Eq. The conclusion is about a transformed signature but still in the Pl-#CSP^2 setting.

Lemma C.8. *For any $x \neq 0$ and any signature f of arity $2n$, let $\hat{f} = \left[\begin{smallmatrix} 1 & 0 \\ 0 & x^{-\frac{1}{2}} \end{smallmatrix}\right]^{\otimes 2n} f$. Then* Pl-#CSP$^2(\hat{f}) \leq_T$ Pl-#CSP$^2(f, [1, 0, x])$.

Proof. After a holographic transformation by $\left[\begin{smallmatrix} 1 & 0 \\ 0 & x^{\frac{1}{2}} \end{smallmatrix}\right]$, we have

$$\text{Pl-\#CSP}^2([1, 0, x], f) \equiv_T \text{Pl-Holant}([1, 0, x], [1, 0, 0, 0, x^2], \cdots \mid [1, 0, 1], \hat{f}).$$

If x is a root of unity, then there exists a $t \geq 1$ such that $x^t = 1$. Thus we have $=_{2kt}$ for all $k \geq 1$ on LHS. Moreover, we have $=_{2k}$ by $\partial^{k(t-1)}(=_{2kt})$ on LHS for all $k \geq 1$. Thus we are done.

If x is not a root of unity, then we have $\partial^{d-2}(E^d_{2d}(x)) = [1, 0, 0, 0, x^d]$ of arity 4 on LHS for all $d \geq 2$, where $E^d_{2d}(x) = [1, 0, \ldots, 0, x^d]$ has arity $2d$. Thus we can get $[1, 0, 0, 0, 1]$ on LHS by interpolation. Then we can get all of $=_{2k}$ on LHS since we have $[1, 0, 1]$ on RHS. □

Lemma C.9. *Suppose either $f = [1, \rho]^{\otimes 3} \pm [1, -\rho]^{\otimes 3}$ or $f = [1, \alpha]^{\otimes 3} \pm [1, -\alpha]^{\otimes 3}$, and let $h = [1, 0, x]$. If $x^4 \notin \{0, 1\}$, then* Pl-#CSP$^2(f, h)$ *is* #P*-hard.*

Proof. We prove the lemma for $f = [1, \rho]^{\otimes 3} \pm [1, -\rho]^{\otimes 3}$. The proof for $f = [1, \alpha]^{\otimes 3} \pm [1, -\alpha]^{\otimes 3}$ is similar and we omit it here.

Let $\hat{f} = [1, x^{-\frac{1}{2}}\rho]^{\otimes 3} \pm [1, -x^{-\frac{1}{2}}\rho]^{\otimes 3}$, then Pl-#CSP$^2(\hat{f}) \leq$Pl-#CSP$^2(f, h)$ by Lemma C.8. $\hat{f}$ satisfies a second order recurrence with eigenvalues $\pm x^{-\frac{1}{2}}\rho$ with sum 0 and product $-\rho^2/x$. Hence $\hat{f}$ has type $\langle -\rho^2/x, 0, 1\rangle$. Moreover, the second recurrence relation is unique up to a scalar since $\hat{f}$ is non-degenerate and has arity 3. By $(x^{-1}\rho^2)^4 \neq 1$, we have $\hat{f} \notin \mathscr{P} \cup \widetilde{\mathscr{A}} \cup \widetilde{\mathscr{M}}$ by Lemma A.11. So Pl-#CSP$^2(\hat{f})$ is #P-hard by Lemma C.2. Thus Pl-#CSP$^2(f, h)$ is #P-hard. □

Lemma C.10. *Let $f = [1, \rho]^{\otimes 3} \pm [1, -\rho]^{\otimes 3}$ and $g = [1, \alpha]^{\otimes 3} \pm [1, -\alpha]^{\otimes 3}$. Then* Pl-#CSP$^2(f, g)$ *is* #P*-hard.*

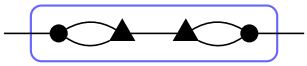


Figure 27: Gadget used to obtain a signature of the form $[a, 0, b]$ with $|a| \neq |b|$. The circle vertices are assigned f and the triangle vertices are assigned g.

Proof. Consider the gadget in Figure 27. We assign f to the circle vertices and g to the triangle vertices. Let h be the signature of this gadget.

- If $f = [1, \rho]^{\otimes 3} + [1, -\rho]^{\otimes 3}$ and $g = [1, \alpha]^{\otimes 3} + [1, -\alpha]^{\otimes 3}$, then $h = 32[\rho^2\alpha^2, 0, -2]$.
- If $f = [1, \rho]^{\otimes 3} - [1, -\rho]^{\otimes 3}$ and $g = [1, \alpha]^{\otimes 3} + [1, -\alpha]^{\otimes 3}$, then $h = 32\rho^2[-2, 0, \rho^2\alpha^2]$.
- If $f = [1, \rho]^{\otimes 3} + [1, -\rho]^{\otimes 3}$ and $g = [1, \alpha]^{\otimes 3} - [1, -\alpha]^{\otimes 3}$, then $h = 32\alpha^2[\rho^2\alpha^2, 0, 2]$.
- If $f = [1, \rho]^{\otimes 3} - [1, -\rho]^{\otimes 3}$ and $g = [1, \alpha]^{\otimes 3} - [1, -\alpha]^{\otimes 3}$, then $h = 32\rho^2\alpha^2[2, 0, \rho^2\alpha^2]$.

Note that both f and g satisfy parity constraints, and thus h also satisfies that. Hence, e.g., in the first case, $f = 2[1, 0, \rho^2, 0]$ and $g = 2[1, 0, \alpha^2, 0]$, we only need to calculate h_0 and h_2, since $h_1 = 0$ by parity. In fact the left half of Figure 27, connecting f to g, also satisfies parity and has the signature $4[1 + \rho^2\alpha^2, 0, 2\rho^2\alpha^2]$, and thus $h = 16[(1 + \rho^2\alpha^2)^2, 0, 4(\rho^2\alpha^2)^2] = 32[\rho^2\alpha^2, 0, -2]$.

Since $|\alpha\rho| = 1 \neq 2$, we are done by Lemma C.9. □

Now we can prove a conditional No-Mixing theorem for Pl-#CSP^2 when a set of signatures $\mathcal{F}$ is assumed to consist of only tractable signatures *and* has at least one nonzero signature of odd arity.

Theorem C.11. *Let $\mathcal{F} \subseteq \bigcup_{k=1}^{5} S_k$ be a set of symmetric signatures that includes at least one nonzero signature of odd arity. If $\mathcal{F} \nsubseteq S_k$ for all $1 \leq k \leq 5$, then* Pl-#CSP$^2(\mathcal{F})$ *is* #P*-hard.*

Proof. If $\mathcal{F}$ contains a signature of odd arity in $\widetilde{\mathscr{M}} \setminus (\mathscr{P} \cup \widetilde{\mathscr{A}})$, then we are done by Lemma C.3. Thus we can assume that $\mathcal{F}$ contains at least one nonzero signature of odd arity $f \in \mathscr{P} \cup \widetilde{\mathscr{A}}$.

By Lemma C.5, if f does not satisfy parity constraints, then we have a unary $[a, b]$ with $ab \neq 0$ and we are done by Lemma B.2. Otherwise, we have $[1, 0]$ or $[0, 1]$ by Lemma C.6. If there exists a signature in $\mathcal{F}$ that does not satisfy parity constraints, then we can obtain a unary $[a, b]$ with $ab \neq 0$ by Lemma C.7. Thus we are done by Lemma B.2.

Now we can assume that $\mathcal{F}$ includes a nonzero odd arity signature $f \in \mathscr{P} \cup \widetilde{\mathscr{A}}$ and all signatures in $\mathcal{F}$ satisfy parity constraints. Thus $\mathcal{F} \bigcap \left(\widetilde{\mathscr{M}} \setminus (\mathscr{P} \cup \widetilde{\mathscr{A}})\right) = \emptyset$ by Corollary A.15. So we have $\mathcal{F} \subseteq \mathscr{P} \cup \widetilde{\mathscr{A}}$, i.e., $\mathcal{F} \subseteq \bigcup_{k=3}^{5} S_k$. Then by Lemma C.4, we have, up to scalar multiples,

$$\mathcal{F} \subseteq \left\{ \begin{array}{ccc} [1,0]^{\otimes n}, & [0,1]^{\otimes n}, & [1,0]^{\otimes 2n} + t[0,1]^{\otimes 2n}, \\ [1,\rho]^{\otimes n} \pm [1,-\rho]^{\otimes n}, & & [1,\alpha]^{\otimes n} \pm [1,-\alpha]^{\otimes n} \end{array} \;\middle|\; t \neq 0 \text{ and } n \geq 1 \right\}.$$

Note that the following signatures are all in $\bigcap_{k=3}^{5} S_k$ (see Figure 35):

$$\begin{array}{llll} [1,0]^{\otimes n} & \text{and} & [0,1]^{\otimes n}, & \\ [1,0]^{\otimes 2n} + t[0,1]^{\otimes 2n} & \text{with} & t^4 = 1, & \\ [1,\rho]^{\otimes m} \pm [1,-\rho]^{\otimes m} & \text{and} & [1,\alpha]^{\otimes \ell} \pm [1,-\alpha]^{\otimes \ell} & \text{with} \quad 1 \leq m, \ell \leq 2. \end{array}$$

Let

$$\mathcal{F}' = \mathcal{F} \bigcap \left\{ \begin{array}{c} [1,0]^{\otimes 2n} + t[0,1]^{\otimes 2n}, \\ [1,\rho]^{\otimes m} \pm [1,-\rho]^{\otimes m}, \quad [1,\alpha]^{\otimes \ell} \pm [1,-\alpha]^{\otimes \ell} \end{array} \;\middle|\; t^4 \notin \{0,1\} \text{ and } m, \ell \geq 3 \right\}.$$

Then $\mathcal{F}' \nsubseteq S_k$ for $3 \le k \le 5$. Indeed if $\mathcal{F}' \subseteq S_k$ for some $3 \le k \le 5$, then $\mathcal{F} \subseteq S_k$. Let

$$S = \mathcal{F}' \bigcap \left\{ [1,\rho]^{\otimes m} \pm [1,-\rho]^{\otimes m} \mid m \ge 3 \right\} \qquad \text{and} \qquad T = \mathcal{F}' \bigcap \left\{ [1,\alpha]^{\otimes \ell} \pm [1,-\alpha]^{\otimes \ell} \mid \ell \ge 3 \right\}.$$

If $S \neq \emptyset$ and $T \neq \emptyset$, then there exist $g, h \in \mathcal{F}'$ such that $g = [1,\alpha]^{\otimes m} \pm [1,-\alpha]^{\otimes m}$ and $h = [1,\rho]^{\otimes \ell} \pm [1,-\rho]^{\otimes \ell}$, where $m, \ell \ge 3$. By Lemma C.6, we can get $[1,0]$ or $[0,1]$ from f. If we have $[1,0]$, then we have $g' = \partial_{[1,0]}^{m-3}(g) = [1,\alpha]^{\otimes 3} \pm [1,-\alpha]^{\otimes 3}$ and $h' = \partial_{[1,0]}^{\ell-3}(h) = [1,\rho]^{\otimes 3} \pm [1,-\rho]^{\otimes 3}$, and are done by Lemma C.10. If we have $[0,1]$, then the proof follows from a transformation by $\left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right]$.

If exactly one of S and T is not empty, then there exists some $[1,0]^{\otimes 2n} + t[0,1]^{\otimes 2n}$ with $t^4 \notin \{0,1\}$ in $\mathcal{F}'$, since otherwise $\mathcal{F}'$ would be contained in either $\mathscr{A}$ or $\mathscr{A}^\dagger$. This contradicts $\mathcal{F}' \nsubseteq S_k$ for $3 \le k \le 5$. By taking ∂^{n-1}, we have $[1,0,t]$. Moreover, we have $g = [1,\alpha]^{\otimes m} \pm [1,-\alpha]^{\otimes m}$ or $h = [1,\rho]^{\otimes \ell} \pm [1,-\rho]^{\otimes \ell}$ in $\mathcal{F}'$, where $m, \ell \ge 3$. By a similar proof with the previous case, first getting $[0,1]$ or $[1,0]$ by Lemma C.6, we can have $g' = [1,\alpha]^{\otimes 3} \pm [1,-\alpha]^{\otimes 3}$ or $h' = [1,\rho]^{\otimes 3} \pm [1,-\rho]^{\otimes 3}$ in $\mathcal{F}'$. Thus Pl-#CSP$^2(\mathcal{F}')$ is #P-hard by Lemma C.9. So Pl-#CSP$^2(\mathcal{F})$ is #P-hard.

If $S = \emptyset$ and $T = \emptyset$, then $\mathcal{F}' \subseteq \left\{ [1,0]^{\otimes 2n} + t[0,1]^{\otimes 2n} \mid t^4 \notin \{0,1\} \right\} \subseteq \mathscr{P}$. This contradicts that $\mathcal{F}' \nsubseteq S_k$ for $3 \le k \le 5$. □

Now we can prove the dichotomy for Pl-#CSP^2 with a single symmetric signature of odd arity.

Theorem C.12. *If f is a symmetric signature of odd arity, then either* Pl-#CSP$^2(f)$ *is* #P*-hard or $f \in \mathscr{P} \cup \widetilde{\mathscr{A}} \cup \widetilde{\mathscr{M}}$.*

Proof. Let f have arity $2n+1$. If $2n+1 = 1$, then $f \in \mathscr{P}$. If $2n+1 = 3$, then we are done by Lemma C.2. In the following, assume that $2n+1 \ge 5$. Let $f' = \partial(f)$. If $f' \notin \mathscr{P} \cup \widetilde{\mathscr{A}} \cup \widetilde{\mathscr{M}}$, then Pl-#CSP$^2(f')$ is #P-hard by induction. Thus Pl-#CSP$^2(f)$ is #P-hard as well. Otherwise, $f' \in \mathscr{P} \cup \widetilde{\mathscr{A}} \cup \widetilde{\mathscr{M}}$.

If $f' \in \widetilde{\mathscr{M}} \setminus (\mathscr{P} \cup \widetilde{\mathscr{A}})$, then we are done by Lemma C.3. So we can assume that $f' \in \mathscr{P} \cup \widetilde{\mathscr{A}}$. Note that f' has odd arity, so if f' does not satisfy parity constraints, then we have $[a,b]$ with $ab \neq 0$ by Lemma C.5 and we are done by Lemma B.2. Otherwise, either f' is identically zero or, as f' has odd arity and satisfies parity, by Lemma C.4

$$f' \in \left\{ [1,0]^{\otimes 2n-1}, [0,1]^{\otimes 2n-1}, [1,\rho]^{\otimes 2n-1} \pm [1,-\rho]^{\otimes 2n-1}, [1,\alpha]^{\otimes 2n-1} \pm [1,-\alpha]^{\otimes 2n-1} \right\}.$$

If $f' \equiv 0$, then $f = x[1,i]^{\otimes 2n+1} + y[1,-i]^{2n+1}$ by Lemma A.5. If $x = 0$ or $y = 0$ or $[xy \neq 0 \wedge x^4 = y^4]$, then $f \in \mathscr{A}$. Otherwise, $xy \neq 0 \wedge x^4 \neq y^4$.

- For $2n+1 \equiv 1 \pmod 4$, we have $\partial_{=_4}^{\frac{n}{2}}(f) = 2^{\frac{n}{2}}\{x[1,i] + y[1,-i]\} = 2^{\frac{n}{2}}[x+y, (x-y)i]$. Note that $x+y \neq 0, x-y \neq 0$ by $x^4 \neq y^4$. Then we are done by Lemma B.2.
- For $2n+1 \equiv 3 \pmod 4$, we have $f'' = \partial_{=_4}^{\frac{n-1}{2}}(f) = 2^{\frac{n-1}{2}}\{x[1,i]^{\otimes 3} + y[1,-i]^{\otimes 3}\}$. Note that $xy \neq 0$ and f is non-degenerate. And by its second order recurrence, $f \in \langle 1,0,1 \rangle$. it follows from Lemma A.11 that $f'' \notin \mathscr{P} \cup \widetilde{\mathscr{A}} \cup \widetilde{\mathscr{M}}$ since $x^4 \neq y^4$. Thus Pl-#CSP$^2(f'')$ is #P-hard by Lemma C.2. So Pl-#CSP$^2(f)$ is #P-hard.

If $f' \in \{[1,0]^{\otimes 2n-1}, [0,1]^{\otimes 2n-1}, [1,\rho]^{\otimes 2n-1} \pm [1,-\rho]^{\otimes 2n-1}, [1,\alpha]^{\otimes 2n-1} \pm [1,-\alpha]^{\otimes 2n-1}\}$, then we have $[1,0]$ or $[0,1]$ by Lemma C.6. So if f does not satisfy parity constraints, then we have $[a,b]$ with $ab \neq 0$ by Lemma C.7 and we are done by Lemma B.2. So we can assume that f satisfies parity constraints in the following.

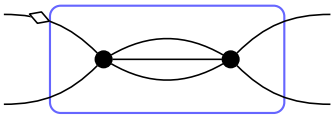


Figure 28: Gadget used to obtain a signature whose signature matrix is redundant.

- For $f' = [1,0]^{\otimes 2n-1}$, $f = x[1,i]^{\otimes 2n+1} + y[1,-i]^{2n+1} + [1,0]^{\otimes 2n+1}$ by Lemma A.5. If $x = y = 0$, then $f \in \mathscr{P}$. Otherwise, $(x,y) \neq (0,0)$.
  Let $a = x + y$, $b = (x-y)i$, then $(a,b) \neq (0,0)$. Note that $f = [1+a, b, -a, -b, \ldots, \pm a, \pm b]$. Since $1+a$ and $-a$ cannot be both 0, by the parity constraints, we have $b = 0$. And thus $a \neq 0$. Moreover we have $\partial^{n-1}([1,0]^{2n-1}) = [1,0]$ and $f''' = \partial^{2n-3}_{[1,0]}(f) = [1+a, 0, -a, 0, a]$. We note that $2n-3 \geq 1$ and so $\partial^{2n-3}_{[1,0]}$ is defined. Note that f''' is a redundant signature and its compressed signature matrix $\left[\begin{smallmatrix} 1+a & 0 & -a \\ 0 & -a & 0 \\ -a & 0 & a \end{smallmatrix}\right]$ is nonsingular, so Pl-#CSP$^2(f''')$ is #P-hard by Lemma A.25. Thus Pl-#CSP$^2(f)$ is #P-hard.
- For $f' = [0,1]^{\otimes 2n-1}$, the proof follows from the previous case by a transformation using $\left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right]$.
- For $f' = [1,\alpha]^{\otimes 2n-1} \pm [1,-\alpha]^{\otimes 2n-1}$, $f = x[1,i]^{\otimes 2n+1} + y[1,-i]^{2n+1} + \frac{1}{1+\alpha^2}\{[1,\alpha]^{\otimes 2n+1} \pm [1,-\alpha]^{\otimes 2n+1}\}$ by Lemma A.5. If $x = y = 0$, then $f \in \mathscr{A}^\dagger$. Otherwise, $(x,y) \neq (0,0)$. Firstly, we construct $[1,0,\alpha^2]$ by f. Note that we have $f^{(4)} = \partial^{n-1}(f) = (1+\alpha^2)^{n-2}\{[1,\alpha]^{\otimes 3} \pm [1,-\alpha]^{\otimes 3}\}$.
  If $f^{(4)} = (1+\alpha^2)^{n-2}\{[1,\alpha]^{\otimes 3} + [1,-\alpha]^{\otimes 3}\}$ with a $+$ sign, we have $\partial(f^{(4)}) = 2(1+\alpha^2)^{n-1}[1,0]$ and $\partial_{[1,0]}(f^{(4)}) = 2(1+\alpha^2)^{n-2}[1,0,\alpha^2]$.
  If $f^{(4)} = (1+\alpha^2)^{n-2}\{[1,\alpha]^{\otimes 3} - [1,-\alpha]^{\otimes 3}\}$ with a $-$ sign, we have $\partial(f^{(4)}) = 2\alpha(1+\alpha^2)^{n-1}[0,1]$ and $\partial_{[0,1]}(f^{(4)}) = 2\alpha(1+\alpha^2)^{n-2}[1,0,\alpha^2]$.
  In either case, we have $[1,0,\alpha^2]$. Then we have $f^{(5)} = \partial^{n-1}_{[1,0,\alpha^2]}(f) = (1-\alpha^2)^{n-1}\{x[1,i]^{\otimes 3} + y[1,-i]^3\}$. If $x = 0$ or $y = 0$ or $[xy \neq 0 \wedge x^4 = y^4]$, then $f^{(5)} \in \mathscr{A} \setminus \mathscr{A}^\dagger$. By the eigenvalues, $f' \in \langle 1, 0 \pm i\rangle$, hence $f' \in \mathscr{A}^\dagger \setminus (\mathscr{P} \cup \mathscr{A} \cup \widetilde{\mathscr{M}})$ in this case. So Pl-#CSP$^2(f^{(5)}, f')$ is #P-hard by Theorem C.11. Otherwise, $xy \neq 0$ and $x^4 \neq y^4$. Then $f^{(5)} \notin \mathscr{P} \cup \widetilde{\mathscr{A}} \cup \widetilde{\mathscr{M}}$. Thus Pl-#CSP$^2(f^{(5)})$ is #P-hard by Lemma C.2. So Pl-#CSP$^2(f)$ is #P-hard.

The final case is $f' = [1,\rho]^{\otimes 2n-1} \pm [1,-\rho]^{\otimes 2n-1}$.

- For $f' = [1,1]^{\otimes 2n-1} + [1,-1]^{\otimes 2n-1}$,

$$f = x[1,i]^{\otimes 2n+1} + y[1,-i]^{2n+1} + \frac{1}{2}\left\{[1,1]^{\otimes 2n+1} + [1,-1]^{\otimes 2n+1}\right\}.$$

  If $x = y = 0$, then $f \in \mathscr{A}$. In the following, assume that $(x,y) \neq (0,0)$. Let $a = x + y$, $b = (x-y)i$, then $(a,b) \neq (0,0)$. Moreover, $f = [a, b, -a, -b, \ldots, \pm b] + [1,0,1,0,\ldots,0] = [a+1, b, -a+1, -b, a+1, \ldots, \pm b]$. Since $a+1$ and $-a+1$ cannot be both 0, and f satisfies parity, we have $b = 0$. Thus $f = [a+1, 0, -a+1, 0, a+1, \ldots, \pm a+1, 0]$. As $b = 0$ we have $a \neq 0$. Note that we have $\partial^n(f) = 2^n[1,0]$. Thus we have $f^{(6)} = \partial^{2n-3}_{[1,0]}(f) = [a+1, 0, -a+1, 0, a+1]$. The compressed signature matrix of $f^{(6)}$ is $\left[\begin{smallmatrix} a+1 & 0 & -a+1 \\ 0 & -a+1 & 0 \\ -a+1 & 0 & a+1 \end{smallmatrix}\right]$ with determinant $4a(1-a)$. If $a \neq 1$, then by $a \neq 0$, this determinant is nonzero. Thus the compressed signature matrix of $f^{(6)}$ is nonsingular and Pl-#CSP$^2(f^{(6)})$ is #P-hard by Lemma A.25. So Pl-#CSP$^2(f)$ is #P-hard.

If $a = 1$, then we have $f^{(7)} = \partial^{2n-4}_{[1,0]}(f) = 2[1,0,0,0,1,0]$ of arity 5 (note that $2n-4 \geq 0$). Consider the gadget in Figure 28. We assign $[1,0,0,0,1,0]$ to both vertices. The signature of this gadget is redundant, and its compressed signature matrix is $\left[\begin{smallmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 3 \end{smallmatrix}\right]$. Since this matrix is nonsingular, we are done by Lemma A.25.

- For $f' = [1,1]^{\otimes 2n-1} - [1,-1]^{\otimes 2n-1}$,

$$f = x[1,i]^{\otimes 2n+1} + y[1,-i]^{2n+1} + \frac{1}{2}\left\{[1,1]^{\otimes 2n+1} - [1,-1]^{\otimes 2n+1}\right\}.$$

After the holographic transformation by $[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}]$, we have Pl-#CSP$^2(f, f') \equiv$ Pl-#CSP$^2(\widehat{f}, \widehat{f'})$, where $\widehat{f'} = [1,1]^{\otimes 2n-1} + [1,-1]^{\otimes 2n-1}$, and $\widehat{f} = xi^{2n+1}[1,-i]^{\otimes 2n+1} + y(-i)^{2n+1}[1,i]^{2n+1} + \frac{1}{2}\{[1,1]^{\otimes 2n+1} + [1,-1]^{\otimes 2n+1}\}$. Thus we are done by the previous case.

- For $f' = [1,i]^{\otimes 2n-1} + [1,-i]^{\otimes 2n-1}$, f has arity $2n+1$ and using Proposition A.7 (the Explicit List for $\int(f')$), $\int([1,\pm i]^{\otimes 2n-1})$ is a sum of $\lambda[1,\pm i]^{\otimes 2n+1}$ with a signature having the k-th term of the form $-\frac{1}{2}k(\pm i)^k$. Thus, we can write $f_k = (x - \frac{1}{2}k)i^k + (y - \frac{1}{2}k)(-i)^k$ by Lemma A.5. We have $\partial_{f'}(=_{2n}) = \partial_{\partial(f)}(=_{2n}) = 2[1,0]$ on LHS.
Let $a = x+y$, $b = (x-y)i$, then $f = [a, b, -a+2, -b, a-4, \ldots, \pm b]$. Since a and $-a+2$ cannot be both 0, and f satisfies parity, we have $b = 0$. Then we have $f^{(8)} = \partial^{2n-3}_{[1,0]}(f) = [a, 0, -a+2, 0, a-4]$. If $a \neq 2$, then the compressed signature matrix of $f^{(8)}$ is $\left[\begin{smallmatrix} a & 0 & -a+2 \\ 0 & -a+2 & 0 \\ -a+2 & 0 & a-4 \end{smallmatrix}\right]$, and is nonsingular and we are done by Lemma A.25.
For $a = 2$, we have $\partial^{2n-4}_{[1,0]}(f) = 2[1,0,0,0,-1,0]$. Consider the gadget in Figure 28. We assign $[1,0,0,0,-1,0]$ to both vertices. The signature of this gadget is redundant, and its compressed signature matrix is $\left[\begin{smallmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 3 \end{smallmatrix}\right]$. Since this matrix is nonsingular, we are done by Lemma A.25.
- For $f' = [1,i]^{\otimes 2n-1} - [1,-i]^{\otimes 2n-1}$, the proof follows from the previous case by a holographic transformation using $[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}]$. □

By Theorem C.11 and Theorem C.12, we have the following dichotomy theorem.

Theorem C.13. *For any set of symmetric signatures $\mathcal{F}$ which contains at least one nonzero signature with odd arity, if $\mathcal{F} \subseteq \mathscr{P}$, or $\mathscr{A}$, or $\mathscr{A}^\dagger$, or $\widehat{\mathscr{M}}$, or $\widehat{\mathscr{M}}^\dagger$, then* Pl-#CSP$^2(\mathcal{F})$ *is tractable. Otherwise,* Pl-#CSP$^2(\mathcal{F})$ *is* #P*-hard.*

D The Arity 4 Dichotomy

The goal of this section is a dichotomy theorem for Pl-#CSP$^2(f)$ when f is a symmetric signature of arity 4. Frequently our first test uses the determinantal criterion of a redundant signature of arity 4 based on Lemma A.25.

Lemma D.1. *Let f be an arity 4 signature. If the signature matrix of f is redundant, and its compressed form is nonsingular, then* Pl-#CSP$^2(f)$ *is #P-hard.*

Proof. Since Pl-Holant$(f) \leq_T$ Pl-#CSP$^2(f)$, we are done by Lemma A.25. □

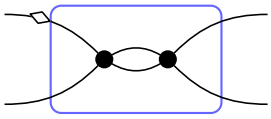


Figure 29: Gadget used in Lemma D.2. Both vertices are assigned f.

Next we introduce a trick which we call the "Three Stooges". For $f = [a, b, c, d, e]$, define

$$\begin{aligned} f^{\times} &= [a, c, e] \\ f^{\times\times} &= [a^2 + c^2 + 2b^2, ac + ce + 2bd, c^2 + e^2 + 2d^2], \qquad \text{and} \\ f^{\overline{\times}} &= [a^2 + c^2 + 2b^2, b^2 + d^2 + 2c^2, c^2 + e^2 + 2d^2]. \end{aligned}$$

The following lemma is proved by the technique of domain pairing.

Lemma D.2. *If $f = [a, b, c, d, e]$, then* Pl-#CSP$(f^{\times}, f^{\overline{\times}}, f^{\times\times}) \le_T$ Pl-#CSP$^2(f)$.

Proof. Let f' be the signature of the gadget in Figure 29 and f'' be the signature of the gadget in Figure 29 rotated 90°. Then f' has a signature matrix on the left, and f'' has a signature matrix on the right:

$$\begin{bmatrix} a^2+c^2+2b^2 & ab+cd+2bc & ab+cd+2bc & ac+ce+2bd \\ ab+cd+2bc & b^2+d^2+2c^2 & b^2+d^2+2c^2 & bc+de+2cd \\ ab+cd+2bc & b^2+d^2+2c^2 & b^2+d^2+2c^2 & bc+de+2cd \\ ac+ce+2bd & bc+de+2cd & bc+de+2cd & c^2+e^2+2d^2 \end{bmatrix}; \quad \begin{bmatrix} a^2+c^2+2b^2 & ab+cd+2bc & ab+cd+2bc & b^2+d^2+2c^2 \\ ab+cd+2bc & ac+ce+2bd & b^2+d^2+2c^2 & bc+de+2cd \\ ab+cd+2bc & b^2+d^2+2c^2 & ac+ce+2bd & bc+de+2cd \\ b^2+d^2+2c^2 & bc+de+2cd & bc+de+2cd & c^2+e^2+2d^2 \end{bmatrix}.$$

We highlight the relevant entries in the display below (in fact, readers should only focus on the entries highlighted; see Figure 2 in Part I for an illustration of the rotation operation):

$$\begin{bmatrix} a^2+c^2+2b^2 & * & * & ac+ce+2bd \\ * & b^2+d^2+2c^2 & * & * \\ * & * & b^2+d^2+2c^2 & * \\ ac+ce+2bd & * & * & c^2+e^2+2d^2 \end{bmatrix}; \quad \begin{bmatrix} a^2+c^2+2b^2 & * & * & b^2+d^2+2c^2 \\ * & ac+ce+2bd & * & * \\ * & * & ac+ce+2bd & * \\ b^2+d^2+2c^2 & * & * & c^2+e^2+2d^{2\text{`}} \end{bmatrix}.$$

For any instance of Pl-#CSP$(f^{\times}, f^{\overline{\times}}, f^{\times\times})$, we replace each edge e by two edges that connect the same incident nodes of e. For each variable node that is connected to k edges, we replace its label $=_k$ by $=_{2k}$. We replace each occurrence of $f^{\times}, f^{\overline{\times}}, f^{\times\times}$ by f, f', f'' as a constraint respectively. Then the new instance is a problem in Pl-#CSP$^2(f, f', f'')$ and has the same value as the given instance of Pl-#CSP$(f^{\times}, f^{\overline{\times}}, f^{\times\times})$. By Pl-#CSP$^2(f, f', f'') \equiv$ Pl-#CSP$^2(f)$, we complete the proof. □

We demonstrate a simple use of the "Three Stooges" in the following lemma.

Lemma D.3. *If $a^4 \notin \{0, 1\}$, then* Pl-#CSP$^2([1, 0, a, 0, a^2])$ *is* #P*-hard.*

Proof. For $f = [1, 0, a, 0, a^2]$, we have $f^{\times} = [1, a, a^2]$ and $f^{\overline{\times}} = [1 + a^2, 2a^2, a^2(1 + a^2)]$. By Lemma A.8, $f^{\times} \notin \mathscr{A} \cup \widehat{\mathscr{M}}$ since $a^4 \notin \{0, 1\}$. By the same reason and Lemma A.8, the only possibility for $f^{\overline{\times}} \in \mathscr{P}$ is being degenerate. Thus $a^2(1 + a^2)^2 = 4a^4$. This implies that $a = 0$ or $a = \pm 1$; a contradiction. This implies that $f^{\times}$ and $f^{\overline{\times}}$ cannot both be in $\mathscr{P}$, $\mathscr{A}$, or $\widehat{\mathscr{M}}$. Thus Pl-#CSP$(f^{\times}, f^{\overline{\times}})$ is #P-hard by Theorem A.22. Then by Lemma D.2, Pl-#CSP$^2(f)$ is #P-hard. □

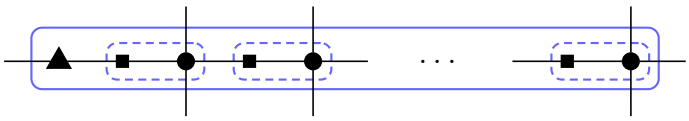


Figure 30: Gadget Γ_k, which has $k-1$ copies of the dashed box. Circle vertices are assigned $\hat{f}$, square vertices are assigned $=_2$, and the triangle vertex is assigned $[1,0,a]$.

Lemma D.4. *Let* $f = [1,1]^{\otimes 4} + a[1,-1]^{\otimes 4}$*, where* $a^4 \neq 0, 1$*. Then* Pl-#CSP$^2(f)$ *is #P-hard.*

Proof. Under a holographic transformation by $H = \left[\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}\right]$, we have

$$\text{Pl-\#CSP}^2(f) \equiv \text{Pl-Holant}(\mathcal{EQ}_2 \mid f) \tag{D.12}$$

$$\equiv \text{Pl-Holant}([1,0,1],[1,0,1,0,1],\dots \mid \hat{f}), \tag{D.13}$$

where $\hat{f} = (H^{-1})^{\otimes 4} f = [1,0,0,0,a]$. By Lemma D.3, Pl-#CSP$^2([1,0,a,0,a^2])$ is #P-hard, and we have

$$\text{Pl-\#CSP}^2([1,0,a,0,a^2]) \equiv \text{Pl-Holant}(\mathcal{EQ}_2 \mid [1,0,a,0,a^2]) \tag{D.14}$$

$$\equiv \text{Pl-Holant}([1,0,a],[1,0,0,0,a^2],\cdots \mid [1,0,1,0,1]) \tag{D.15}$$

$$\leq \text{Pl-Holant}([1,0,a],[1,0,0,0,a^2],\cdots \mid [1,0,1],[1,0,1,0,1],\cdots), \tag{D.16}$$

where the second equivalence $\equiv$ is by a holographic transformation with $\left[\begin{smallmatrix} 1 & 0 \\ 0 & \sqrt{a} \end{smallmatrix}\right]$.

The problem in (D.13) can simulate the problem in (D.16). With $[1,0,1]$ on the left and $\hat{f}$ on the right in (D.13), we can get $\partial(\hat{f}) = [1,0,a]$ on the right. Now consider the gadget in Figure 30. We assign $\hat{f}$ to the circle vertices, $=_2$ to the square vertices, and $[1,0,a]$ to the triangle vertex. If there are $k-1$ occurrences of the dashed subgadget, then the signature of this gadget is $[1,0,\dots,0,a^k]$ of arity $2k$. Thus

$$\begin{aligned} &\text{Pl-Holant}([1,0,1],[1,0,1,0,1],\cdots \mid [1,0,a],[1,0,0,0,a^2],\cdots) \\ \leq\ &\text{Pl-Holant}([1,0,1],[1,0,1,0,1],\cdots \mid \hat{f}). \end{aligned}$$

Then combining three reductions, we have Pl-#CSP$^2([1,0,a,0,a^2]) \leq$ Pl-#CSP$^2(f)$, where $a^4 \neq 0, 1$. Thus Pl-#CSP$^2(f)$ is #P-hard by Lemma D.3. ☐

Now we are ready to prove the following theorem.

Theorem D.5. *Let* f *be a signature of arity 4, then* Pl-#CSP$^2(f)$ *is* #P*-hard or* $f \in \mathscr{P} \cup \widetilde{\mathscr{A}} \cup \widetilde{\mathscr{M}}$*.*

Proof. The first step is to apply Lemma D.2 to $f^\times$. For $f = [f_0, f_1, f_2, f_3, f_4]$ we have $f^\times = [f_0, f_2, f_4]$. If Pl-#CSP$(f^\times)$ is #P-hard, then Pl-#CSP$^2(f)$ is #P-hard by Lemma D.2. In the following, assume that Pl-#CSP$(f^\times)$ is not #P-hard, and hence tractable by the dichotomy Theorem A.21, and $[f_0, f_2, f_4]$ takes the following form

$$[0,0,0],\ [1,0,0],\ [0,0,1],\ [1,r,r^2],\ [0,1,0],\ [1,0,a],\ [1,1,-1],\ [1,-1,-1],\ \text{ or }\ [1,b,1]$$

up to a scalar, where $r \neq 0$, $a \neq 0$, and $b^2 \notin \{0,1\}$.

Case 1: $[f_0, f_2, f_4] = [0, 0, 0]$

In this case, $f = [0, x, 0, y, 0]$ and $f^{\overline{\times}} = [2x^2, x^2 + y^2, 2y^2]$.

- If $x^2 = y^2$, then $f = [0, x, 0, \pm x, 0] \in \mathscr{A}$.
- If $x^2 = -y^2$, then $f = [0, 1, 0, \pm i, 0] \in \mathscr{A}^\dagger$ since $\begin{bmatrix} 1 & 0 \\ 0 & \sqrt{i} \end{bmatrix}^{\otimes 4} f \in \mathscr{A}$.
- If $x^4 \neq y^4$, then Pl-#CSP$(f^{\overline{\times}})$ is #P-hard by Theorem A.21, so Pl-#CSP$^2(f)$ is #P-hard by Lemma D.2.

Case 2: $[f_0, f_2, f_4] = [1, 0, 0]$ or $[0, 0, 1]$

We prove the case for $[f_0, f_2, f_4] = [1, 0, 0]$, i.e., $f = [1, x, 0, y, 0]$. The other case is similar. Note that we have $\partial(f) = [1, x + y, 0]$. If $x + y \neq 0$, then Pl-#CSP$^2([1, x + y, 0])$ is #P-hard by Theorem A.21. Thus Pl-#CSP$^2(f)$ is #P-hard.
If $x = -y \neq 0$, then Pl-#CSP$^2(f)$ is #P-hard by Lemma D.1.
If $x = -y = 0$, then $f = [1, 0]^{\otimes 4} \in \mathscr{P}$.

Case 3: $[f_0, f_2, f_4] = [1, r, r^2]$ with $r \neq 0$

In this case, $f = [1, x, r, y, r^2]$. If $rx \neq y$, then Pl-#CSP$^2(f)$ is #P-hard by Lemma D.1. Otherwise, $f = [1, x, r, xr, r^2]$. Then we have $\partial(f) = (1 + r)[1, x, r]$. If $r \neq -1$, then we have $[1, x, r]$. In the following we will separate out the cases according to value of r.

For $r^4 \neq 1$ in $f = [1, x, r, xr, r^2]$.

- If $x = 0$, then $f = [1, 0, r, 0, r^2]$, and Pl-#CSP$^2(f)$ is #P-hard by Lemma D.3.
- If $x^2 = r$, then $f = [1, x]^{\otimes 4} \in \mathscr{P}$.
- If $x^2 \neq r$ and $x \neq 0$, then Pl-#CSP$^2([1, x, r])$ is #P-hard by Theorem A.21. Thus Pl-#CSP$^2(f)$ is #P-hard.

For $r = 1$, then $f = [1, x, 1, x, 1]$.

- If $x^4 = 0$ or 1, then $f \in \mathscr{A}$.
- If $x^4 \neq 0, 1$, then let $a = \frac{1-x}{1+x}$ and we have $a^4 \neq 0, 1$ by Lemma A.3. Note that $f = \frac{1}{1+a}\left\{[1, 1]^{\otimes 4} + a[1, -1]^{\otimes 4}\right\}$. By Lemma D.4, Pl-#CSP$^2(f)$ is #P-hard.

For $r = -1$, then $f = [1, x, -1, -x, 1]$.

- If $x^4 = 0$ or 1, then $f \in \mathscr{A}$.
- If $x^4 \neq 0, 1$, then let $a = \frac{1+xi}{1-xi}$ and we have $a^4 \neq 0, 1$ by Lemma A.3. Note that $f = \frac{1}{a+1}\left\{[1, i]^{\otimes 4} + a[1, -i]^{\otimes 4}\right\}$. Thus we have $[1, 0, -1]^{\otimes 2}$ on the left by Lemma A.20. Under the holographic transformation by $\left[\begin{smallmatrix} 1 & 0 \\ 0 & i \end{smallmatrix}\right]$, this $[1, 0, -1]^{\otimes 2}$ is transformed to $[1, 0, 1]^{\otimes 2}$, and we have

$$\text{Pl-\#CSP}^2(f) \equiv_T \text{Pl-Holant}(\mathcal{EQ}_4 \cup \{[1, 0, 1]^{\otimes 2}, [1, 0, -1], [1, 0, 0, 0, 0, 0, -1], \cdots\} \mid f'), \tag{D.17}$$

where $f' = \frac{1}{1+a}\left\{[1, 1]^{\otimes 4} + a[1, -1]^{\otimes 4}\right\}$. Now having $[1, 0, 1]^{\otimes 2}$ on the left, we can form a pair of self loops in a planar way for a pair of adjacent f' and get $(\partial(f'))^{\otimes 2} = \left(\frac{2}{1+a}[1 + a, 1 - a, 1 + a]\right)^{\otimes 2}$ on the right side. Since we have $[1, 0, 1]^{\otimes 2}$ on the left side, we can obtain $[1, 0, 1]^{\otimes 2}$ on the right side by interpolation using $[1 + a, 1 - a, 1 + a]^{\otimes 2}$.

Note that the matrix $\left[\begin{smallmatrix}1+a & 1-a\\ 1-a & 1+a\end{smallmatrix}\right]$ can be diagonalized by $\left[\begin{smallmatrix}1 & 1\\ 1 & -1\end{smallmatrix}\right]$. This implies that

$$\text{Pl-\#CSP}^4(f', [1,0,1]^{\otimes 2}) \le_T \text{Pl-Holant}(\mathcal{EQ}_4 \cup \{[1,0,1]^{\otimes 2}, [1,0,-1], [1,0,0,0,0,0,-1], \cdots\} \mid f').$$

Then by (D.17) and Lemma B.4, we have

$$\text{Pl-\#CSP}^2(f') \le \text{Pl-\#CSP}^4(f', [1,0,1]^{\otimes 2}) \le \text{Pl-\#CSP}^2(f). \tag{D.18}$$

By Lemma D.4, Pl-#CSP(f') is #P-hard. Thus Pl-#CSP(f) is #P-hard.

For $r^2 = -1$, then $r = \pm i$ in $f = [1, x, r, xr, -1]$.

- If $x = 0$, then $f = [1, 0, r, 0, -1] \in \mathscr{A}^\dagger$ since $\left[\begin{smallmatrix}1 & 0\\ 0 & \sqrt{i}\end{smallmatrix}\right]^{\otimes 4} f = [1, 0, \pm 1, 0, 1] \in \mathscr{A}$.
- If $x^2 = r$, then $f = [1, x]^{\otimes 4} \in \mathscr{P}$.
- If $x^2 = -r$, then $f = [1, x, -x^2, -x^3, -1] \in \mathscr{A}^\dagger$ since $\left[\begin{smallmatrix}1 & 0\\ 0 & x^{-1}\end{smallmatrix}\right]^{\otimes 4} f \in \mathscr{A}$, with $x^4 = -1$.
- If $x^4 \ne 0, -1$, thus $x^2 \ne \pm r$. Then Pl-#CSP$^2([1, x, r])$ is #P-hard by Theorem A.21. Thus Pl-#CSP$^2(f)$ is #P-hard.

Case 4: $[f_0, f_2, f_4] = [0, 1, 0]$

In this case, $f = [0, x, 1, y, 0]$. We first apply Lemma D.1 and calculate the determinant of the compressed matrix for f, which is $2xy - 1$. If $xy \ne \frac{1}{2}$, then Pl-#CSP$^2(f)$ is #P-hard by Lemma D.1.

If $xy = \frac{1}{2}$ and $x - y = 0$, then $f = [0, \frac{1}{\sqrt{2}}, 1, \frac{1}{\sqrt{2}}, 0]$ or $f = [0, -\frac{1}{\sqrt{2}}, 1, -\frac{1}{\sqrt{2}}, 0]$. Both are in $\widehat{\mathscr{M}}$, by Lemma A.18.

If $xy = \frac{1}{2}$ and $x + y = 0$, then $f = [0, \frac{i}{\sqrt{2}}, 1, -\frac{i}{\sqrt{2}}, 0]$ or $f = [0, -\frac{i}{\sqrt{2}}, 1, \frac{i}{\sqrt{2}}, 0]$. Both are in $\widehat{\mathscr{M}}^\dagger$, by Lemma A.18. In fact from the previous line with $[0, \pm\frac{1}{\sqrt{2}}, 1, \pm\frac{1}{\sqrt{2}}, 0]$, we can see directly $\left[\begin{smallmatrix}1 & 0\\ 0 & i\end{smallmatrix}\right]^{\otimes 4} f \in \widehat{\mathscr{M}}$.

In the following we have $xy = \frac{1}{2}$ and $x^2 \ne y^2$. Then $f^{\aleph} = [1 + 2x^2, 2 + x^2 + y^2, 1 + 2y^2]$ and $f^{\asymp} = [1 + 2x^2, 1, 1 + 2y^2]$. We will prove that Pl-#CSP$(f^{\aleph}, f^{\asymp})$ is #P-hard by showing that $f^{\aleph}, f^{\asymp}$ cannot be both in the same $\mathscr{P}, \mathscr{A}$, or $\widehat{\mathscr{M}}$.

- By $x^2 \ne y^2$ and Lemma A.8, we have $f^{\asymp} \notin \widehat{\mathscr{M}}$.
- Suppose $\{f^{\aleph}, f^{\asymp}\} \subset \mathscr{P}$. $f^{\asymp}$ is not of the form $[1, 0, a]$, and also not of the form $[0, 1, 0]$ since $1 + 2x^2 \ne 1 + 2y^2$. Thus $f^{\asymp}$ is degenerate, i.e., $(1 + 2x^2)(1 + 2y^2) = 1$. Note that $f^{\aleph}$ is not of the form $[0, 1, 0]$ since $1 + 2x^2 \ne 1 + 2y^2$. If $f^{\aleph}$ is of the form $[1, 0, a]$, then $x^2 + y^2 = -2$. Then together with $xy = \frac{1}{2}$ we obtain $(1 + 2x^2)(1 + 2y^2) = -2 \ne 1$. This contradicts that $f^{\asymp}$ is degenerate. Thus $f^{\aleph}$ and $f^{\asymp}$ are both degenerate. Then we have

$$\begin{aligned}(1 + 2x^2)(1 + 2y^2) &= (x^2 + y^2 + 2)^2,\\ (1 + 2x^2)(1 + 2y^2) &= 1.\end{aligned} \tag{D.19}$$

Together we have $(x^2 + y^2 + 2)^2 = 1$, i.e., $x^2 + y^2 = -3$ or $x^2 + y^2 = -1$. However both possibilities contradict (D.19) and $xy = \frac{1}{2}$. Thus $f^{\aleph}$ and $f^{\asymp}$ cannot both belong to $\mathscr{P}$.

- Suppose $\{f^{\divideontimes}, f^{\asymp}\} \subset \mathscr{A}$. By $f^{\asymp} \in \mathscr{A}$ and the middle term is nonzero, by Corollary A.9 we have $1 + 2x^2 = \pm(1 + 2y^2)$. Since $x^2 \neq y^2$, we have $1 + 2x^2 = -1 - 2y^2$. This leads to $(x + y)^2 = 0$ by using $xy = \frac{1}{2}$. This contradicts $x^2 \neq y^2$.

We have proved that $f^{\divideontimes}, f^{\asymp}$ cannot be both in $\mathscr{P}$, or $\mathscr{A}$, or $\widehat{\mathscr{M}}$. Thus Pl-#CSP$(f^{\divideontimes}, f^{\asymp})$ is #P-hard by Theorem A.22. So Pl-#CSP$^2(f)$ is #P-hard by Lemma D.2.

Case 5: $[f_0, f_2, f_4] = [1, 0, a]$ with $a \neq 0$

In this case, $f = [1, x, 0, y, a]$. We first apply Lemma D.1 and calculate the determinant of the compressed matrix for f, which is $-(ax^2 + y^2)$. If $ax^2 + y^2 \neq 0$, then Pl-#CSP$^2(f)$ is #P-hard by Lemma D.1. In the following we assume $ax^2 + y^2 = 0$.

If $x = y = 0$, then $f \in \mathscr{P}$.

If $x = y \neq 0$, then $a = -1$. So $f = [1, x, 0, x, -1] \in \widehat{\mathscr{M}^{\dagger}}$, by Corollary A.18.

If $x = -y \neq 0$, then $a = -1$. So $f = [1, x, 0, -x, -1] \in \widehat{\mathscr{M}}$, by Corollary A.18.

Now we assume $ax^2 + y^2 = 0$ and $x^2 \neq y^2$. Then $a \neq -1$ and $xy \neq 0$ by $a \neq 0$. In this case, the "Three Stooges" are

$$f^{\times} = [1, 0, a], \qquad f^{\divideontimes} = [1 + 2x^2, x^2 + y^2, a^2 + 2y^2], \qquad \text{and} \qquad f^{\asymp} = [1 + 2x^2, 2xy, a^2 + 2y^2].$$

By $ax^2 + y^2 = 0$, we have

$$f^{\divideontimes} = [1 + 2x^2, (1 - a)x^2, a^2 - 2ax^2] \qquad \text{and} \qquad f^{\asymp} = [1 + 2x^2, 2xy, a^2 - 2ax^2].$$

We will prove that Pl-#CSP$(f^{\times}, f^{\divideontimes}, f^{\asymp})$ is #P-hard by showing that $f^{\times}, f^{\divideontimes}$ and $f^{\asymp}$ cannot be all in the same $\mathscr{P}, \mathscr{A}$, or $\widehat{\mathscr{M}}$.

- Assume that $\{f^{\times}, f^{\asymp}\} \subseteq \widehat{\mathscr{M}}$. Note that $a \neq -1$. If $f^{\times} \in \widehat{\mathscr{M}}$, we have $a = 1$ by Lemma A.8. Then by $f^{\asymp} \in \widehat{\mathscr{M}}$ and Lemma A.8, we have $1 + 2x^2 = 1 - 2x^2$ or $2xy = 0$. This is a contradiction.
- Assume that $\{f^{\divideontimes}, f^{\asymp}\} \subseteq \mathscr{P}$. If $1 + 2x^2$ and $a^2 - 2ax^2$ are both zero, then $a = 0$ or -1. This is a contradiction. Thus $f^{\divideontimes}, f^{\asymp}$ are not of the form $[0, 1, 0]$. By $xy \neq 0$, $f^{\asymp}$ is not of the form $[1, 0, c]$ with $c \neq 0$. Thus $f^{\asymp}$ is degenerate by Lemma A.8, i.e.,

$$(1 + 2x^2)(a^2 - 2ax^2) = 4x^2y^2 = -4ax^4, \tag{D.20}$$

  where the last equality is by $ax^2 + y^2 = 0$.

  If $a = 1$, we have $1 - 4x^4 = -4x^4$ by (D.20). This is a contradiction.

  If $a \neq 1$, then $f^{\divideontimes}$ is not of the form $[1, 0, c]$ with $c \neq 0$. Thus $f^{\divideontimes}$ is degenerate by $f^{\divideontimes} \in \mathscr{P}$, i.e.,

$$(1 + 2x^2)(a^2 - 2ax^2) = (1 - a)^2x^4.$$

  Then by (D.20), we have $-4ax^4 = (1 - a)^2x^4$. This implies that $-4a = (1 - a)^2$ by $x \neq 0$. Then $(1 + a)^2 = 0$, contradicting $a \neq \pm 1$.
- Suppose $\{f^{\times}, f^{\asymp}\} \subset \mathscr{A}$. By $f^{\times} \in \mathscr{A}$, and $a \neq 0$, we get $a^4 = 1$ from Lemma A.8. It follows that $a = 1$ or $a^2 = -1$, as we have $a \neq -1$.

  For $a = 1$, the equation $ax^2 + y^2 = 0$ gives us $y^2 = -x^2$. Then from Corollary A.9 we have

$$(1 + 2x^2)^2 = (1 - 2x^2)^2$$

by $f^{\times\times} \in \mathscr{A}$ and $2xy \neq 0$. Thus $x = 0$. This is a contradiction.
For $a^2 = -1$, by $f^{\times\times} \in \mathscr{A}$ and $2xy \neq 0$, we have $(1 + 2x^2)^2 = (-1 - 2ax^2)^2$ by Corollary A.9. $1 + 2x^2 = 1 + 2ax^2$ leads to a contradiction $a = 1$, hence $1 + 2x^2 = -(1 + 2ax^2)$. Then $x^2 = -\frac{1}{a+1}$ and $f^{\times\times} = [\frac{a-1}{a+1}, 2xy, \frac{a-1}{a+1}]$. Note that $a + 1 \neq 0$. We observe that the norm of x^2 is $\frac{1}{\sqrt{2}}$ and the norm of x is equal to the norm of y by $ax^2 = -y^2$ and $a^2 = -1$. Thus the norm of $2xy$ is $\sqrt{2}$. Moreover, the norm of $\frac{a-1}{a+1}$ is 1, as $a = \pm i$. Thus the norm of $2xy$ is not equal to the norm of $\frac{a-1}{a+1}$, and are nonzero. So $f^{\times\times} \notin \mathscr{A}$ by Corollary A.9.

This implies that $f^{\times}, f^{\overline{\times}}$ and $f^{\times\times}$ cannot be all in $\mathscr{P}$, or all in $\mathscr{A}$, or all in $\widehat{\mathscr{M}}$. Thus the problem Pl-#CSP$(f^{\times}, f^{\overline{\times}}, f^{\times\times})$ is #P-hard by Theorem A.22. So Pl-#CSP$^2(f)$ is #P-hard.

Case 6: $[f_0, f_2, f_4] = [1, \pm 1, -1]$

In this case, $f = [1, x, 1, y, -1]$ or $[1, x, -1, y, -1]$. We consider the first case; the second case is similar.
We have $\partial(f) = [2, x + y, 0]$. If $x + y \neq 0$, then Pl-#CSP$^2([2, x + y, 0])$ is #P-hard by Theorem A.21. Thus Pl-#CSP$^2(f)$ is #P-hard. Now we assume $x + y = 0$. Next we apply Lemma D.1 and calculate the determinant of the compressed matrix for f, which is a nonzero constant multiple of $x^2 + 1$. If $x^2 + 1 \neq 0$, then Pl-#CSP$^2(f)$ is #P-hard by Lemma D.1.
If $x + y = 0$ and $x^2 + 1 = 0$, then $f = [1, \pm i, 1, \mp i, -1]$. We have

$$\partial(f) = 2[1, 0, 0], \qquad \partial_{[1,0,0]}(f) = [1, \pm i, 1], \qquad \text{and} \qquad \partial_{[1,\pm i,1]}(f) = [0, \pm 2i, 2].$$

Then Pl-#CSP$^2([0, \pm 2i, 2])$ is #P-hard by Theorem A.21. Thus Pl-#CSP$^2(f)$ is #P-hard.

Case 7: $[f_0, f_2, f_4] = [1, b, 1]$ with $b^2 \neq 0, 1$

In this last case of Theorem D.5, $f = [1, x, b, y, 1]$ and the determinant of the compress signature matrix is

$$D = b + 2bxy - b^3 - x^2 - y^2. \tag{D.21}$$

If $D \neq 0$, then Pl-#CSP$^2(f)$ is #P-hard by Lemma D.1. In the following we assume that $D = 0$.
If $x = y = 0$, then $b = 0$ or $b^2 = 1$ by $D = b(1 - b^2) = 0$. This is a contradiction.
If $x = y \neq 0$, then $D = (1 - b)[b(1 + b) - 2x^2] = 0$. By $b \neq 1$, we have $b(1 + b) = 2x^2$. This implies that $f \in \widehat{\mathscr{M}}$ by Corollary A.18.
Similarly, if $x = -y \neq 0$, then $D = (1 + b)[b(1 - b) - 2x^2] = 0$. By $b \neq -1$, we have $b(1 - b) = 2x^2$. This implies that $f \in \widehat{\mathscr{M}^{\dagger}}$ by Corollary A.18.
In the following, assume that $x^2 \neq y^2$ in addition to $D = 0$. In this case, the "Three Stooges" are

$$\begin{aligned} f^{\times} &= [1, b, 1], \\ f^{\overline{\times}} &= [1 + b^2 + 2x^2, 2b^2 + x^2 + y^2, 1 + b^2 + 2y^2], \qquad \text{and} \\ f^{\times\times} &= [1 + b^2 + 2x^2, 2b + 2xy, 1 + b^2 + 2y^2]. \end{aligned}$$

We will prove that Pl-#CSP$(f^{\times}, f^{\overline{\times}}, f^{\times\times})$ is #P-hard by showing that $f^{\times}, f^{\overline{\times}}, f^{\times\times}$ cannot all be in the same $\mathscr{P}$, or $\mathscr{A}$, or $\widehat{\mathscr{M}}$.

By $b^2 \neq 0, 1$, we have $f^{\times} \notin \mathscr{P}$ by Lemma A.8.

- Suppose $b^2 \neq -1$. Then in addition to $b^2 \neq 0, 1$, we have $b^4 \neq 0, 1$. Then $f^{\times} \notin \mathscr{A}$ by Lemma A.8. Moreover, if $f^{\divideontimes} \in \widehat{\mathscr{M}}$, then by Lemma A.8 and the fact that $x^2 \neq y^2$, we must have

$$1 + b^2 + 2x^2 = -(1 + b^2 + 2y^2) \qquad \text{and} \qquad 2b^2 + x^2 + y^2 = 0. \tag{D.22}$$

  From (D.22), we get $b^2 = 1$. This is a contradiction. This implies that $f^{\times}, f^{\divideontimes}$ cannot be all in $\mathscr{P}$, or all in $\mathscr{A}$, or all in $\widehat{\mathscr{M}}$ when $b^2 \neq -1$.

- Now suppose $b^2 = -1$. Then $f^{\divideontimes} = [2x^2, x^2 + y^2 - 2, 2y^2]$ and $f^{\times\times} = 2[x^2, b + xy, y^2]$. If $f^{\divideontimes} \in \widehat{\mathscr{M}}$, then by $x^2 \neq y^2$ and Lemma A.8, we have

$$x^2 = -y^2 \qquad \text{and} \qquad x^2 + y^2 - 2 = 0$$

  This is a contradiction.

  Finally suppose $\{f^{\divideontimes}, f^{\times\times}\} \subset \mathscr{A}$.

  – If $x^2 + y^2 = 0$, then $xy = -1$ by $b^2 = -1$ and

$$D = b + 2bxy - b^3 - x^2 - y^2 = 0.$$

    Then $f^{\divideontimes} = 2[x^2, -1, y^2]$, $f^{\times\times} = 2[x^2, b - 1, y^2]$ both have all nonzero entries. If they are both in $\mathscr{A}$, the norm of their entries must be all the same $|b-1| = |x^2| = |-1| = 1$, by Corollary A.9. However $b - 1$ does not have norm 1 since $b^2 = -1$.

  – If $x^2 + y^2 \neq 0$, then, since we also have $x^2 \neq y^2$, the first and the last entries of both $f^{\divideontimes}$ and $f^{\times\times}$ are neither equal nor negative of each other. It follows from membership in $\mathscr{A}$ that $x^2 + y^2 - 2 = 0$ and $b + xy = 0$ by Corollary A.9. Then by $D = b + 2bxy - b^3 - x^2 - y^2 = 0$ and $b^2 = -1$, we get a contradiction.

  We have proved that $f^{\times}, f^{\divideontimes}, f^{\times\times}$ cannot be all in $\mathscr{P}$, or all in $\mathscr{A}$, or all in $\widehat{\mathscr{M}}$ when $b^2 = -1$.

From above, $f^{\times}, f^{\divideontimes}, f^{\times\times}$ cannot be all in $\mathscr{P}$, or all in $\mathscr{A}$, or all in $\widehat{\mathscr{M}}$ when $x^2 \neq y^2$ and $D = 0$. Thus Pl-#CSP$(f^{\times}, f^{\divideontimes}, f^{\times\times})$ is #P-hard by Theorem A.22. So Pl-#CSP$^2(f)$ is #P-hard. This completes the proof of Case 7.

This completes the proof of Theorem D.5. □

E An Application of Cyclotomic Field

E.1 Dichotomy Theorem with a Signature in $\widehat{\mathscr{M}} \setminus (\mathscr{P} \cup \widetilde{\mathscr{A}})$

The next three lemmas are crucial. The purpose of these lemmas is to give a similar result as Lemma C.3 when the signature set $\mathcal{F}$ contains some $f \in \widehat{\mathscr{M}} \setminus (\mathscr{P} \cup \widetilde{\mathscr{A}})$, and all signatures in $\mathcal{F}$ have even arity. The proof uses an argument involving the degree of extension of a *cyclotomic field*.

We first prove that if we have an even arity signature in $\widehat{\mathscr{M}} \setminus (\mathscr{P} \cup \widetilde{\mathscr{A}})$, then we can construct a binary $[1, a, 1]$ with $a^4 \notin \{0, 1\}$.

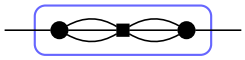


Figure 31: Gadget used in the proof of Lemma E.1.

Lemma E.1. *Let $\mathcal{F}$ be a set of symmetric signatures containing some $f \in \widehat{\mathscr{M}} \setminus (\mathscr{P} \cup \widetilde{\mathscr{A}})$, which has even arity. Then*

$$\text{Pl-}\#\text{CSP}^2([1,a,1],\mathcal{F}) \leq_T \text{Pl-}\#\text{CSP}^2(\mathcal{F})$$

for some a satisfying $a^4 \notin \{0,1\}$.

Proof. If f has arity 2, then we are done by Lemma A.14. Thus, we assume that f has arity $2n \geq 4$. By Lemma A.14, we have either $f = [s,t]^{\otimes 2n} \pm [t,s]^{\otimes 2n}$ with $s^4 \neq t^4$ and $st \neq 0$ or $f_k = (\pm 1)^k(2n-2k)$ up to a scalar.

For $f = [s,t]^{\otimes 2n} + [t,s]^{\otimes 2n}$, we have $\partial^{n-1}(f) = (s^2+t^2)^{n-1}\{[s,t]^{\otimes 2} + [t,s]^{\otimes 2}\} = (s^2+t^2)^n[1,a,1]$, where $a = \frac{2st}{s^2+t^2}$. Note that $s^2+t^2 \neq 0$ and $a \neq 0, \pm 1$. If $a \neq \pm i$, then we are done. Suppose $a = \pm i$. Then $g = \partial^{n-2}(f) = (s^2+t^2)^{n-2}\{[s,t]^{\otimes 4} + [t,s]^{\otimes 4}\}$. A simple calculation shows that $g = -2s^2t^2(s^2+t^2)^{n-2}[3, \pm i, -1, \pm i, 3]$. Consider the gadget in Figure 31. We assign $[3, \pm i, -1, \pm i, 3]$ to the circle vertices and $=_6$ to the square vertex. Its signature is $[8, \pm 6i, 8]$, so we are done.

For $f = [s,t]^{\otimes 2n} - [t,s]^{\otimes 2n}$, we have $\partial^{n-1}(f) = (s^2+t^2)^{n-1}\{[s,t]^{\otimes 2} - [t,s]^{\otimes 2}\} = \lambda[1,0,-1]$, where $\lambda = (s^2+t^2)^{n-1}(s^2-t^2) \neq 0$. For $2n \geq 6$, we have $\partial_{[1,0,-1]}(f) = (s^2-t^2)\{[s,t]^{\otimes 2n-2} + [t,s]^{\otimes 2n-2}\}$ and we are done by the proof of the previous case, as $2n-2 \geq 4$. For $2n = 4$, we have $\partial_{[1,0,-1]}(f) = (s^2-t^2)\{[s,t]^{\otimes 2} + [t,s]^{\otimes 2}\} = (s^4-t^4)[1,a,1]$, where $a = \frac{2st}{s^2+t^2} \neq 0, \pm 1$. If $a \neq \pm i$, then we are done. Suppose $a = \pm i$, then a simple calculation shows that f is a nonzero multiple of $[2i, \mp 1, 0, \pm 1, -2i]$. (One can verify that $\frac{s^3t - st^3}{s^4-t^4} = \frac{st}{s^2+t^2} = \frac{a}{2} = \pm\frac{i}{2}$.) Consider the gadget in Figure 31. We assign $[2i, \mp 1, 0, \pm 1, -2i]$ to the circle vertices and $=_6$ to the square vertex.. The signature of this gadget is $[-3, \mp 4i, -3]$, so we are done.

For $f_k = (\pm 1)^k(2n-2k)$, we have $\partial^{n-2}(f) = 2^{n-1}[2, \pm 1, 0, \mp 1, -2]$. Consider the gadget in Figure 31. We assign $[2, \pm 1, 0, \mp 1, -2]$ to the circle vertices and $=_6$ to the square vertex. The signature of this gadget is $[5, \pm 4, 5]$, so we are done. ☐

The next lemma shows that if we have $[1,a,1]$ with $a^4 \neq 0, 1$, then we can obtain $[1,1]^{\otimes 2}$ by interpolation.

Lemma E.2. *For any signature set $\mathcal{F}$ and any $a^4 \notin \{0,1\}$,*

$$\text{Pl-}\#\text{CSP}^2(\{[1,1]^{\otimes 2}\} \cup \mathcal{F}) \leq_T \text{Pl-}\#\text{CSP}^2(\{[1,a,1]\} \cup \mathcal{F}).$$

Proof. The eigenvalues of $\left[\begin{smallmatrix} 1 & b \\ b & 1 \end{smallmatrix}\right]$ are $1+b$ and $1-b$ respectively. If we have a signature $[1,b,1]$, for some $b \neq 1$, such that ratio $\frac{1+b}{1-b}$ of eigenvalues is not a root of unity, then we can interpolate any binary signature $[1,x,1]$ for $x \in \mathbb{C}$. In particular, we could interpolate the desired $[1,1]^{\otimes 2}$.

Indeed, let Ω be an instance of Pl-$\#$CSP$^2(\{[1,x,1]\} \cup \mathcal{F})$ in which $[1,x,1]$ occurs n times. Write $\left[\begin{smallmatrix} 1 & x \\ x & 1 \end{smallmatrix}\right]$ as $H \left[\begin{smallmatrix} 1+x & 0 \\ 0 & 1-x \end{smallmatrix}\right] H$, where $H = \frac{1}{\sqrt{2}}\left[\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}\right]$. We can stratify the partition function value on Ω as $\mathrm{Z}(\Omega) = \sum_{\ell=1}^{n} c_\ell (1+x)^\ell (1-x)^{n-\ell}$, where c_ℓ is the sum, over all assignments that assign 00 to ℓ copies of $\left[\begin{smallmatrix} 1+x & 0 \\ 0 & 1-x \end{smallmatrix}\right]$ and 11 to the remaining $n-\ell$ copies, of the product of evaluations of all other signatures from $\mathcal{F}$ and those copies of H. If we construct a sequence Ω_k of instances of Pl-$\#$CSP$^2(\{[1,b,1]\} \cup \mathcal{F})$, where we replace each occurrence of $[1,x,1]$ by a chain of k linked copies

of $[1, b, 1]$, then since $\left[\begin{smallmatrix} 1 & b \\ b & 1 \end{smallmatrix}\right]^k = H \left[\begin{smallmatrix} (1+b)^k & 0 \\ 0 & (1-b)^k \end{smallmatrix}\right] H$, we have $\mathrm{Z}(\Omega_k) = (1-b)^{kn} \sum_{\ell=1}^{n} c_\ell (\frac{1+b}{1-b})^{k\ell}$, for $0 \le k \le n$. This is a Vandermonde system of full rank, and we can solve for all c_ℓ and find the value $\mathrm{Z}(\Omega)$.

The simple gadget with two copies of $=_{2k}$ connected by $2k-1$ parallel copies of $[1, a, 1]$ has signature $[1, a^{2k-1}, 1]$. Our key claim is that there exists a $k \ge 1$, depending only on a, such that $\frac{1+a^{2k-1}}{1-a^{2k-1}}$ is *not* a root of unity. Then we are done by the interpolation given above.

For a contradiction, assume that $\frac{1+a^{2k-1}}{1-a^{2k-1}}$ is a root of unity for all $k \ge 1$. For $k = 1$, $\frac{1+a}{1-a}$ is some root of unity $e^{2\pi i j/m}$, where $\gcd(j, m) = 1$. Then $a \in \Phi_m = \mathbb{Q}(e^{2\pi i/m})$, the m-th cyclotomic field. Therefore $a^{2k-1} \in \Phi_m$ as well for all $k \ge 1$. Furthermore, $|\frac{1+a}{1-a}| = 1$, so a is purely imaginary, i.e. $a = ih$ for some real $h \notin \{0, \pm 1\}$ since $a^4 \notin \{0, 1\}$. First we consider the case $0 < |h| < 1$. Then $a^{2k-1} = \pm i h^{2k-1}$ and $\lim_{k\to\infty} h^{2k-1} = 0$.

For all $k \ge 1$, $\frac{1+a^{2k-1}}{1-a^{2k-1}}$ is some root of unity $e^{2\pi i J/M}$ (in which J and M depend on k), where $0 < |J| < M/2$ with $\gcd(J, M) = 1$. Then $e^{2\pi i/M} \in \Phi_m$ as well, so $\Phi_M \subseteq \Phi_m$. Note that $|\tan(\pi J/M)| = |h|^{2k-1}$. Hence $|h|^{2k-1} \ge \tan(\pi/M) \ge \pi/M$. Thus $M \ge \pi/|h|^{2k-1}$.

However, the M-th cyclotomic field Φ_M has degree of extension $[\mathbb{Q}(e^{2\pi i/M}) : \mathbb{Q}] = \varphi(M)$, where φ is the Euler totient function. We have a crude estimate $(\varphi(M))^2 \ge M/2$, which is obvious by considering each prime dividing M. Then it follows that $\lim_{M\to\infty} \varphi(M) = \infty$, which contradicts $\varphi(M) \le \varphi(m) < \infty$.

The remaining case $|h| > 1$ can be handled similarly. In fact, if $|h^{2k-1}|$ is large, then the angle $\tan^{-1}(|h^{2k-1}|)$ is close (but unequal) to $\pi/2$. Then the angle of $\left(\frac{1+a^{2k-1}}{1-a^{2k-1}}\right)^2$ is close (but unequal) to 0 mod 2π. □

Combining Lemma B.3, Lemma E.1 and Lemma E.2, we have proved the following.

Lemma E.3. *Let $\mathcal{F}$ be a set of even-arity signatures containing f. If $f \in \widehat{\mathscr{M}} \setminus (\mathscr{P} \cup \widetilde{\mathscr{A}})$, then* Pl-#CSP$^2(\mathcal{F})$ *is* #P*-hard unless $\mathcal{F} \subseteq \widehat{\mathscr{M}}$.*

E.2 Dichotomy Theorem with a Signature in $\widehat{\mathscr{M}^\dagger} \setminus (\mathscr{P} \cup \widetilde{\mathscr{A}})$

We would like to prove a corresponding statement to Lemma E.3 after replacing the condition $f \in \widehat{\mathscr{M}} \setminus (\mathscr{P} \cup \widetilde{\mathscr{A}})$ by $f \in \widehat{\mathscr{M}^\dagger} \setminus (\mathscr{P} \cup \widetilde{\mathscr{A}})$. This corresponding statement is indeed true and is implied by Theorem A.2, the final dichotomy theorem for Pl-#CSP^2. However, at this point leading up to the proof of Theorem A.2, we are not able to prove it. Instead, we prove a weaker version, Lemma E.7, in which f is assisted by a binary signature other than a multiple of $[1, 0, 1]$.

Remark 5. Here we explain some of the difficulties in the proof caused by structural complications of the signatures involved.

When we prove the No-Mixing statements for $\widehat{\mathscr{M}}$ the crucial step is the ability to construct $[1, \omega]^{\otimes 2}$ with $\omega \ne 0$ in the Pl-#CSP^2 setting (cf. Lemma E.1 and Lemma E.2). This is the key, and the only known method, for us to leverage the existing dichotomy for Pl-#CSP (cf. Lemma B.3). Then in a similar spirit, to prove the No-Mixing statements for $\widehat{\mathscr{M}^\dagger}$, we would like to be able to construct $[1, \omega]^{\otimes 2}$ as well.

A signature $f = [f_0, \ldots, f_n]$ is called an *odd* signature if $f_{2k} = 0$ for all $k \ge 0$, and an *even* signature if $f_{2k+1} = 0$ for all $k \ge 0$.

In any $\mathcal{F}$-gate H, if every signature in $\mathcal{F}$ satisfies parity constraints, then the signature of H also satisfies parity constraints. In fact the parity of the signature of H is the same as the parity of the

number of occurrences of odd signatures of $\mathcal{F}$ in H. To see this, suppose σ is a $\{0,1\}$-assignment to all the edges of H, including internal and external edges, that has a nonzero evaluation on H. By parity constraints, each odd (resp. even) signature appearing in H has an odd (resp. even) number of incident edges assigned 1. Adding up all these numbers mod 2, noting that each internal edge of H assigned 1 contributes 2 to the sum while each external edge of H assigned 1 contributes 1, we get $N \equiv 2X + Y \equiv Y \pmod 2$, where N is the number of occurrences of odd signatures of $\mathcal{F}$ in H, and X (resp. Y) is the number of internal (resp. external) edges assigned to 1 by σ. Hence H has the same parity as N.

For any signature of the form $f = [s, ti]^{\otimes m} \pm [t, si]^{\otimes m}$, or $f_k = (\pm i)^k (m - 2k)$, for any arity m, $(Z^{-1})^{\otimes m} f$ satisfies the parity constraints, where $Z = \frac{1}{2}\left[\begin{smallmatrix} 1 & 1 \\ i & -i \end{smallmatrix}\right]$. In fact for f of the first type, $(Z^{-1})^{\otimes m} f = [u, v]^{\otimes m} \pm [u, -v]^{\otimes m}$ for $u = s + t$ and $v = s - t$, and for f of the second type, $(Z^{-1})^{\otimes m} f = 2^m [0, 1, 0, \ldots, 0]$ or $2^m [0, \ldots, 0, 1, 0]$. Note that

$$\left[\begin{smallmatrix} 1 & 1 \\ i & -1 \end{smallmatrix}\right]^{\otimes m} [0, 1, 0, \ldots, 0] = \mathrm{Sym}_n^{n-1}(\left[\begin{smallmatrix} 1 \\ i \end{smallmatrix}\right]; \left[\begin{smallmatrix} 1 \\ -i \end{smallmatrix}\right])$$

has its k-th term $i^k(m - 2k)$. Similarly, $\left[\begin{smallmatrix} 1 & 1 \\ i & -i \end{smallmatrix}\right]^{\otimes m} [0, \ldots, 0, 1, 0]$ has its k-th term $(-i)^k(m - 2k)$.

Under the holographic transformation Z, we have

$$\text{Pl-\#CSP}^2(f) \equiv_T \text{Pl-Holant}([0, 1, 0], [1, 0, 1, 0, 1], \ldots \mid \hat{f}), \tag{E.23}$$

where $\hat{f} = (Z^{-1})^{\otimes m} f$, and $\frac{1}{2}[0, 1, 0] = (=_2) Z^{\otimes 2}$, $\frac{1}{2^3}[1, 0, 1, 0, 1] = (=_4) Z^{\otimes 4}$, etc. Notice that for the signatures $(=_{2n}) Z^{\otimes 2n}$, if the arity $2n \equiv 2 \pmod 4$ then the signature is odd, and if $2n \equiv 0 \pmod 4$ then the signature is even.

Every signature of the form $[s, ti]^{\otimes m} + [t, si]^{\otimes m}$ is even, every signature of the form $[s, ti]^{\otimes m} - [t, si]^{\otimes m}$ is odd, and for even arity $2n$ the signatures $[0, 1, 0, \ldots, 0]$ and $[0, \ldots, 0, 1, 0]$ are both odd.

Thus, if we focus on signatures $f = [s, ti]^{\otimes 2n} + [t, si]^{\otimes 2n}$ with arity $2n \equiv 0 \pmod 4$, or $f = [s, ti]^{\otimes 2n} - [t, si]^{\otimes 2n}$ with arity $2n \equiv 2 \pmod 4$, or $f_k = (\pm i)^k (2n - 2k)$ with arity $2n \equiv 2 \pmod 4$, then the following property holds for *all* the signatures in the bipartite Pl-Holant problem in (E.23):

> All signatures of arity $2n \equiv 2 \pmod 4$ satisfy odd parity and all signatures of arity $2n \equiv 0 \pmod 4$ satisfy even parity.

It follows that, for such f, any gadget constructed from (E.23) has the same parity as the number of occurrences of signatures of arity $2n \equiv 2 \pmod 4$.

Furthermore, in a bipartite gadget construction in Pl-Holant$([0, 1, 0], [1, 0, 1, 0, 1], \ldots \mid \hat{f})$, if the resulting signature of the gadget is *binary*, the number of occurrences of signatures of arity $2n \equiv 2 \pmod 4$ in this gadget *must* be odd. Indeed let N_0 (resp. N_2) denote the number of occurrences of signatures of arity $2n \equiv 0 \pmod 4$ (resp. $2n \equiv 2 \pmod 4$) in this bipartite gadget, and we add up the arities of all signatures modulo 4, we get $0N_0 + 2N_2 \equiv 2N_I + 2 \pmod 4$, where N_I is the number of internal edges in the bipartite gadget, and the additive term 2 is because the gadget is a binary gadget. Thus $N_2 \equiv N_I + 1 \pmod 2$. On the other hand, since the gadget is bipartite, N_I is the sum of all arities of signatures from RHS, and minus 2 if the external 2 edges come from the RHS. As all signatures in this gadget have even arity, $N_I \equiv 0 \pmod 2$. Hence $N_2 \equiv 1 \pmod 2$.

This implies that any binary signature constructed in Pl-Holant$([0, 1, 0], [1, 0, 1, 0, 1], \ldots \mid \hat{f})$ must have odd parity, i.e., they are all of the form $\lambda[0, 1, 0]$. Thus, before the Z-transformation, one can only construct binary signatures of the form $\frac{\lambda}{2}[1, 0, 1] = \lambda Z^{\otimes 2}[0, 1, 0]$ in Pl-\#CSP(f) by gadget construction. This can be verified as $\left[\begin{smallmatrix} 1 & 1 \\ i & -i \end{smallmatrix}\right] \left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right] \left[\begin{smallmatrix} 1 & i \\ 1 & -i \end{smallmatrix}\right] = 2 \left[\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right]$.

In particular one *cannot* construct $[1,\omega]^{\otimes 2}$ in Pl-#CSP$^2(f)$ by gadget construction. This explains the extra mile we have to travel in this proof.

As indicated, therefore, we prove a weaker version of Lemma E.3 in this subsection, namely Lemma E.7, in which f is assisted by a binary signature other than a multiple of $[1,0,1]$.

We begin with the following lemma.

Lemma E.4. *Let $\mathcal{F}$ be any set of symmetric signatures of even arities, and suppose $\mathcal{F}$ contains signatures f and g, where $f \in \widehat{\mathscr{M}^\dagger} \setminus (\mathscr{P} \cup \widetilde{\mathscr{A}})$, and $g = [g_0, g_1, \ldots, g_{2n}]$ and there exists a positive integer s such that $g_0^s = -g_{2n}^s \neq 0$. Then either $\mathcal{F} \subseteq \widehat{\mathscr{M}^\dagger}$ or Pl-#CSP$^2(\mathcal{F})$ is #P-hard.*

Proof. Let $E_{2k}(-1) = [1, 0, \ldots, 0, -1]$ have arity $2k$ and $\mathcal{E}(-1) = \{E_{2k}(-1) \mid k \geq 1\}$. Firstly, by our calculus we have $\partial_g^s(=_{2ns+2k}) = g_0^s E_{2k}(-1)$ on LHS for $k \geq 1$. Thus we have

$$\text{Pl-Holant}(\mathcal{E}(-1) \cup \mathcal{EQ}_2 \mid \mathcal{F}) \leq_T \text{Pl-\#CSP}^2(\mathcal{F}).$$

Under a holographic transformation by $T^{-1} = \left[\begin{smallmatrix} 1 & 0 \\ 0 & i \end{smallmatrix}\right]$, the set $\mathcal{E}(-1) \cup \mathcal{EQ}_2$ is set-wise invariant. Indeed, for all $k \geq 1$, signatures of arity $4k$ in $\mathcal{E}(-1) \cup \mathcal{EQ}_2$ are pointwise fixed, and signatures of arity $4k-2$ in $\mathcal{E}(-1)$ and in $\mathcal{EQ}_2$ are interchanged. Thus,

$$\text{Pl-\#CSP}^2(T\mathcal{F}) \leq_T \text{Pl-Holant}(\mathcal{E}(-1) \cup \mathcal{EQ}_2 \mid T\mathcal{F}) \equiv_T \text{Pl-Holant}(\mathcal{E}(-1) \cup \mathcal{EQ}_2 \mid \mathcal{F}).$$

Note that $T^{\otimes 2n} f \in T\mathcal{F}$ is in $\widehat{\mathscr{M}} \setminus (\mathscr{P} \cup \widetilde{\mathscr{A}})$. Thus either $T\mathcal{F} \subseteq \widehat{\mathscr{M}}$ or Pl-#CSP$^2(T\mathcal{F})$ is #P-hard by Lemma E.3. Note that $T\mathcal{F} \subseteq \widehat{\mathscr{M}}$ iff $\mathcal{F} \subseteq \widehat{\mathscr{M}^\dagger}$. Thus either $\mathcal{F} \subseteq \widehat{\mathscr{M}^\dagger}$ or Pl-#CSP$^2(\mathcal{F})$ is #P-hard. □

The next two lemmas show that if we have a signature in $\widehat{\mathscr{M}^\dagger} \setminus (\mathscr{P} \cup \widetilde{\mathscr{A}})$ and a binary signature that is not a multiple of $[1,0,1]$, then we have the same statement for $\widehat{\mathscr{M}^\dagger}$, as Lemma E.3 is for $\widehat{\mathscr{M}}$. This will be stated as Lemma E.7. Note that if $f \in \widehat{\mathscr{M}^\dagger} \setminus (\mathscr{P} \cup \widetilde{\mathscr{A}})$ is a binary signature, then f takes the form $[1, b, -1]$ by Lemma A.14, and this case is covered by Lemma E.4, where f also plays the role of g. Thus we assume $f \in \widehat{\mathscr{M}^\dagger} \setminus (\mathscr{P} \cup \widetilde{\mathscr{A}})$ has arity ≥ 4. By Lemma A.14, such a signature f has two forms. Lemma E.5 and E.6 handle these two cases respectively.

Lemma E.5. *Let $\mathcal{F}$ be any set of symmetric signatures of even arities, and suppose $\mathcal{F}$ contains signatures f and h, where $f = [s, ti]^{\otimes 2n} \pm [t, si]^{\otimes 2n}$ with $2n \geq 4$, $s^4 \neq t^4$ and $st \neq 0$, and h is any nonzero binary signature other than $\lambda[1,0,1]$. Then either $\mathcal{F} \subseteq \widehat{\mathscr{M}^\dagger}$ or Pl-#CSP$^2(\mathcal{F})$ is #P-hard.*

Proof. Firstly, by our calculus, ignoring the nonzero factor $(s^2 - t^2)^{n-2}$ in $\partial^{n-2}(f)$, we have $g = [s, ti]^{\otimes 4} \pm (-1)^{n-2}[t, si]^{\otimes 4}$. If $g = [s, ti]^{\otimes 4} - [t, si]^{\otimes 4}$, then we have $\partial(g) = (s^2 - t^2)\{[s, ti]^{\otimes 2} + [t, si]^{\otimes 2}\} = (s^2 - t^2)[s^2 + t^2, 2sti, -(s^2 + t^2)]$ and we are done by Lemma E.4.

Suppose $g = [s, ti]^{\otimes 4} + [t, si]^{\otimes 4}$, and we also have $h \neq \lambda[1,0,1]$. If $h \notin \mathscr{P} \cup \widetilde{\mathscr{A}} \cup \widetilde{\mathscr{M}}$, then Pl-#CSP$^2(\mathcal{F})$ is #P-hard by Theorem A.21′. Otherwise, by Lemma A.8, the possibilities for h, after normalizing, are

$$[a, b]^{\otimes 2}, \quad [1, 0, x], \quad [0, 1, 0], \quad [1, \rho, -\rho^2], \quad [1, \alpha, -\alpha^2], \quad [1, u, 1], \quad \text{and} \quad [1, v, -1],$$

where $x \notin \{0, 1\}$, $\rho^4 = 1$, $\alpha^4 = -1$, $u^4 \notin \{0, 1\}$, and $v^4 \notin \{0, 1\}$.

- If $h = [a, b]^{\otimes 2}$ with $ab \neq 0$, then we are done by Lemma B.3.

- If $h \in \{[1,0,-1],[1,0,\pm i],[1,\pm 1,-1],[1,\alpha,-\alpha^2],[1,v,-1]\}$, then we are done by Lemma E.4.
- If $h = [1,u,1]$ with $u^4 \neq 0,1$, then $h \in \widehat{\mathscr{M}} \setminus (\mathscr{P} \cup \widetilde{\mathscr{A}})$ by Lemma A.14. Thus we are done by Lemma E.3.

The remaining cases are $h = [1,0]^{\otimes 2}$, $[0,1]^{\otimes 2}$, $[1,0,x]$, $[0,1,0]$ or $[1,\pm i,1]$, where $x^4 \neq 0,1$.

- If $h = [1,0,x]$ with $x^4 \neq 0,1$, then by taking 4 copies of h and connecting one input of h to each edge of g, we have $\hat{g} = [\begin{smallmatrix}1 & 0\\ 0 & x\end{smallmatrix}]^{\otimes 2n} g = [s,xti]^{\otimes 4} + [t,xsi]^{\otimes 4}$. The signature $\hat{g}$ is non-degenerate, has arity 4, and satisfies a second recurrence relation. The eigenvalues of the recurrence relation are $\frac{xti}{s}$ and $\frac{xsi}{t}$. By the trace and product, $\hat{g}$ has type $\langle -x^2, \frac{xti}{s} + \frac{xsi}{t}, 1\rangle$. Thus $\hat{g} \notin \mathscr{P} \cup \widetilde{\mathscr{A}} \cup \widetilde{\mathscr{M}}$ by Lemma A.11, since $(-x^2)^2 \neq 0,1$ and $\frac{t}{s} + \frac{s}{t} \neq 0$. So Pl-#CSP$^2(\hat{g})$ is #P-hard by Theorem D.5. Thus Pl-#CSP$^2(\mathcal{F})$ is #P-hard.
- If $h = [0,1,0]$, then $\partial_h(g) = 2sti\{[s,ti]^{\otimes 2} + [t,si]^{\otimes 2}\} = 2sti[s^2+t^2, 2sti, -(s^2+t^2)]$. Then we are done by Lemma E.4.
- If $h = [1,\pm i,1]$, by connecting two copies of $[1,\pm i,1]$ we have $\pm 2i[0,1,0]$, as $[\begin{smallmatrix}1 & \pm i\\ \pm i & 1\end{smallmatrix}]^2 = [\begin{smallmatrix}0 & \pm 2i\\ \pm 2i & 0\end{smallmatrix}]$. Then we are done by the previous case.
- If $h = [1,0]^{\otimes 2}$, then we have $g' = \partial_h(g) = s^2[s,ti]^{\otimes 2} + t^2[t,si]^{\otimes 2} = [s^4+t^4, (s^2+t^2)sti, -2s^2t^2]$. We claim that $g' \notin \mathscr{P} \cup \widetilde{\mathscr{A}} \cup \widetilde{\mathscr{M}}$.
  - If $g' \in \mathscr{P}$, then g' is degenerate by $(s^2+t^2)sti \neq 0$ and $-2s^2t^2 \neq 0$. So $-2s^2t^2(s^4+t^4) = -(s^2+t^2)^2 s^2t^2$. Thus $st = 0$ or $(s^2-t^2)^2 = 0$. This is a contradiction.
  - If $g' \in \mathscr{A} \setminus \mathscr{P}$, then $g' = [1,\rho,-\rho^2]$ up to a scalar by Corollary A.9, where $\rho^4 = 1$. By $\rho^2 = \pm 1$, we have $s^4 + t^4 = \pm 2s^2t^2$. This contradicts that $s^4 \neq t^4$.
  - If $g' \in \mathscr{A}^\dagger \setminus \mathscr{P}$, then $g' = [1,\alpha,-\alpha^2]$ up to a scalar by Corollary A.9, where $\alpha^4 = -1$. Thus $2s^2t^2(s^4+t^4) = -(s^2+t^2)^2 s^2t^2$. Then, by $st \neq 0$, we have $3(s^4+t^4) = -2s^2t^2 \neq 0$, and so $|s^4+t^4| \neq |-2s^2t^2|$. This implies that the norms of two nonzero entries of g' are not equal. This contradicts the form $g' = \lambda[1,\alpha,-\alpha^2]$.
  - Since $s^4 \neq t^4$ we have $s^4 + t^4 \neq \pm 2s^2t^2$. Hence $g' \notin \widetilde{\mathscr{M}}$ by Corollary A.9.

  Then by Theorem A.21′, Pl-#CSP$^2(g')$ is #P-hard. Thus Pl-#CSP$^2(\mathcal{F})$ is #P-hard.
- If $h = [0,1]^{\otimes 2}$, then we apply the transformation $[\begin{smallmatrix}0 & 1\\ 1 & 0\end{smallmatrix}]$ and are done by the previous case. □

Lemma E.6. *Let $\mathcal{F}$ be any set of symmetric signatures of even arities, and suppose $\mathcal{F}$ contains signatures f and h, where f has arity $2n \geq 4$ and $f_k = (\pm i)^k(2n-2k)$, and h is any nonzero binary signature other than $\lambda[1,0,1]$. Then either $\mathcal{F} \subseteq \widehat{\mathscr{M}}^\dagger$ or* Pl-#CSP$^2(\mathcal{F})$ *is #P-hard.*

Proof. If $2n \equiv 0 \pmod 4$, then $f_0 = -f_{2n} = 2n$. Thus we are done by Lemma E.4.

Suppose $2n \equiv 2 \pmod 4$. Thus $n \geq 3$ and we have $g = \partial_{=4}^{\frac{n-3}{2}}(f)$ of arity 6. Ignoring the nonzero factor $2^{\frac{n-3}{2}}$, we have $g_k = (\pm i)^k(6-2k)$. Removing another factor 2, we have

$$g = [3, \pm 2i, -1, 0, -1, \mp 2i, 3].$$

We also have a nonzero binary signature $h \neq \lambda[1,0,1]$. If $h \notin \mathscr{P} \cup \widetilde{\mathscr{A}} \cup \widetilde{\mathscr{M}}$, then Pl-#CSP$^2(\mathcal{F})$ is #P-hard by Theorem A.21′. Otherwise (similar to the proof of Lemma E.5), by Lemma A.8, the possibilities for h, after normalizing, are

$$[a,b]^{\otimes 2}, \quad [1,0,x], \quad [0,1,0], \quad [1,\rho,-\rho^2], \quad [1,\alpha,-\alpha^2], \quad [1,u,1], \quad \text{and} \quad [1,v,-1],$$

where $x \notin \{0,1\}$, $\rho^4 = 1$, $\alpha^4 = -1$, $u^4 \notin \{0,1\}$, and $v^4 \notin \{0,1\}$. If $h = [1,0,-1]$, $[1,0,\pm i]$, $[1,\pm 1,-1]$, $[1,\alpha,-\alpha^2]$, $[1,v,-1]$, $[1,u,1]$, or $[a,b]^{\otimes 2}$ with $ab \neq 0$, then we are done with the same proof as in Lemma E.5.

The remaining cases are $h = [1,0]^{\otimes 2}$, $[0,1]^{\otimes 2}$, $[1,0,x]$, $[0,1,0]$, or $[1,\pm i,1]$, where $x^4 \notin \{0,1\}$.

- For $h = [1,0,x]$ with $x^4 \notin \{0,1\}$, we have $g' = \partial_h(g) = [3-x, \pm 2i, -1-x, \mp 2xi, -1+3x]$. The signature g' is non-degenerate because $(\pm 2i)(\mp 2xi) \neq (-1-x)^2$ by $x \neq 1$. Moreover, g' satisfies the second recurrence relation with type $\langle 1, \mp 2i, -1\rangle$. Thus $g' \notin \mathscr{P} \cup \widehat{\mathscr{A}} \cup \widehat{\mathscr{M}}$ by Lemma A.11. Moreover, by $x \neq \pm 1$, we have $3 - x \neq \pm(-1+3x)$, so $g' \notin \widehat{\mathscr{M}^\dagger}$ by Corollary A.17. So Pl-#CSP$^2(g')$ is #P-hard by Theorem D.5. Thus Pl-#CSP$^2(\mathcal{F})$ is #P-hard.
- If $h = [0,1,0]$, then $\partial_h(g) = [\pm 2i, -1, 0, -1, \mp 2i]$. Then we are done by Lemma E.4.
- If $h = [1, \pm i, 1]$, by connecting two copies of $[1, \pm i, 1]$ we have $\pm 2i[0,1,0]$. Then we are done by the proof of the previous case.
- If $h = [1,0]^{\otimes 2}$, then we have $g'' = \partial_h^2(g) = [3, \pm 2i, -1]$. By Corollary A.9, we have $g'' \notin \mathscr{P} \cup \widetilde{\mathscr{A}} \cup \widetilde{\mathscr{M}}$. Then by Theorem A.21′, Pl-#CSP$^2(g'')$ is #P-hard. Thus Pl-#CSP$^2(\mathcal{F})$ is #P-hard.
- If $h = [0,1]^{\otimes 2}$, we apply the transformation $\left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right]$ and it follows from the previous case. □

Lemma E.7. *Let $\mathcal{F}$ be any set of symmetric signatures of even arities, and suppose $\mathcal{F}$ contains signatures f and h, where $f \in \widehat{\mathscr{M}^\dagger} \setminus (\mathscr{P} \cup \widetilde{\mathscr{A}})$, and h is any nonzero binary signature other than $\lambda[1,0,1]$. Then either $\mathcal{F} \subseteq \widehat{\mathscr{M}^\dagger}$ or Pl-#CSP$^2(\mathcal{F})$ is #P-hard.*

Proof. If f has arity 2, then $f = [1, b, -1]$ by Lemma A.14. Then we are done by Lemma E.4.

If f has arity $2n \geq 4$, then by Lemma A.14, we have $f = [s, ti]^{\otimes 2n} \pm [t, si]^{\otimes 2n}$ with $st \neq 0$, $s^4 \neq t^4$, or $f_k = (\pm i)^k(2n - 2k)$ up to a scalar. These two cases are proved in Lemma E.5, and E.6 respectively. □

Remark 6. Lemma E.3 and Lemma E.7 will substantially simplify the succeeding proof for No-Mixing Lemmas concerning $\widehat{\mathscr{M}}$ and $\widehat{\mathscr{M}^\dagger}$. Thus it is natural that we wish to do the same for $\mathscr{A}$, and that means we would like to construct $[1,\omega]^{\otimes 2}$ with $\omega \neq 0$ in Pl-#CSP$^2(f)$ for $f \in \mathscr{A}$. Unfortunately, for most cases of $f \in \mathscr{A}$ this is impossible.

First, for a signature $f \in \mathscr{A}$, if f satisfies parity constraints, then all signatures constructed in Pl-#CSP$^2(f)$ satisfy parity constraints, since all $\mathcal{EQ}_2$ also satisfy parity constraints. So it is impossible to construct $[1,\omega]^{\otimes 2}$ with $\omega \neq 0$ in Pl-#CSP$^2(f)$.

If a signature $f \in \mathscr{A}$ is degenerate and does not satisfy parity constraints, then $f = [1, \pm 1]^{\otimes 2n}$ or $f = [1, \pm i]^{\otimes 2n}$ up to a scalar. For $f = [1, \pm 1]^{\otimes 2n}$, we have $\partial^{n-1}(f) = 2^{n-1}[1, \pm 1]^{\otimes 2}$. For $f = [1, \pm i]^{\otimes 2n}$ and $2n \equiv 2 \pmod 4$, we have $\partial_{=4}^{\frac{n-1}{2}}(f) = 2^{\frac{n-1}{2}}[1, \pm i]^{\otimes 2}$. Thus in these two particular cases we *can* get $[1,\omega]^{\otimes 2}$ with $\omega \neq 0$. We will show that these are the only cases that this is possible.

Let $f = [1, \pm i]^{\otimes 2n}$ and $2n \equiv 0 \pmod 4$. After a holographic transformation by $Z = \left[\begin{smallmatrix} 1 & 1 \\ i & -i \end{smallmatrix}\right]$, we have

$$\text{Pl-\#CSP}^2(f) \equiv_T \text{Pl-Holant}([0,1,0], [1,0,1,0,1], \ldots \mid \hat{f}),$$

where $\hat{f} = (Z^{-1})^{\otimes 2n} f$, i.e., $\hat{f} = [1,0]^{\otimes 2n}$ or $\hat{f} = [0,1]^{\otimes 2n}$. In Pl-Holant$([0,1,0], [1,0,1,0,1], \ldots \mid \hat{f})$, all signatures of arity $\equiv 0 \pmod 4$ have even parity and all signatures of arity $\equiv 2 \pmod 4$ have odd parity. By the same proof in Remark 6, all nonzero binary signatures that can be constructed in Pl-Holant$([0,1,0], [1,0,1,0,1], \ldots \mid \hat{f})$ are multiples of $[0,1,0]$. In terms of signatures that can be constructed before the Z-transformation, this is equivalent to say that all nonzero binary signatures that can be constructed in Pl-#CSP$^2(f)$ must be multiples of $[1,0,1]$. In particular, one *cannot* construct $[1,\omega]^{\otimes 2}$ with $\omega \neq 0$ in Pl-#CSP$^2(f)$.

If $f \in \mathscr{A}$ is non-degenerate and does not satisfy parity constraints, then $f = [1,i]^{\otimes 2n} \pm i[1,-i]^{\otimes 2n}$ or $f = [1,1]^{\otimes 2n} \pm i[1,-1]^{\otimes 2n}$. If we can construct $[1,\omega]^{\otimes 2}$ with $\omega \neq 0$ in Pl-#CSP$^2(f)$, then $[1,\omega]^{\otimes 2}$ must be in $\mathscr{A}$. Thus $[1,\omega]^{\otimes 2}$=$[1,\pm 1]^{\otimes 2}$ or $[1,\pm i]^{\otimes 2}$.

For $f = [1,i]^{\otimes 2n} \pm i[1,-i]^{\otimes 2n}$, $f = [1,\pm 1,-1,\mp 1,\ldots,(-1)^n]$ up to the scalar $1 \pm i$. In any construction in Pl-#CSP$^2(f)$, if we ignore a global scalar factor which is a power of $1 \pm i$, all entries of the constructed signature are real numbers. Thus the ratio of any two nonzero entries is a real number. But this is not the case with $[1,\pm i]^{\otimes 2}$. This implies that we *cannot* construct $[1,\pm i]^{\otimes 2}$ in Pl-#CSP$^2(f)$ by gadget construction.

Moreover, we claim that it is impossible to get $[1,\pm 1]^{\otimes 2}$ in Pl-#CSP$^2(f)$ by gadget construction. After a holographic transformation by $Z = \left[\begin{smallmatrix} 1 & 1 \\ i & -i \end{smallmatrix}\right]$, we have

$$\text{Pl-\#CSP}^2(f) \equiv_T \text{Pl-Holant}([0,1,0],[1,0,1,0,1],\ldots \mid \hat{f}),$$

where $\hat{f} = (Z^{-1})^{\otimes 2n} f = [1,0,\ldots,0,\pm i]$. All signatures in Pl-Holant$([0,1,0],[1,0,1,0,1],\ldots \mid \hat{f})$ satisfy parity constraints. Thus we cannot construct $(Z^{-1})^{\otimes 2}[1,\pm 1]^{\otimes 2} = \mp\frac{i}{2}[1,\pm i]^{\otimes 2}$, which does not satisfy parity constraints, by gadget construction. Thus we *cannot* get $[1,\pm 1]^{\otimes 2}$ in Pl-#CSP$^2(f)$ by gadget construction.

For $f = [1,1]^{\otimes 2n} \pm i[1,-1]^{\otimes 2n}$, after a holographic transformation by $\left[\begin{smallmatrix} 1 & 0 \\ 0 & i \end{smallmatrix}\right]$, we can use the same argument as the previous case for $[1,\pm i]^{\otimes 2}$ to prove that we cannot get $[1,\pm 1]^{\otimes 2}$ in Pl-#CSP$^2(f)$ by gadget construction. Moreover, it is also impossible to get $[1,\pm i]^{\otimes 2}$ in Pl-#CSP$^2(f)$ by gadget construction. After a holographic transformation by $H = \left[\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}\right]$, we have

$$\text{Pl-\#CSP}^2(f) \equiv_T \text{Pl-Holant}([1,0,1],[1,0,1,0,1],\ldots \mid \hat{f}),$$

where $\hat{f} = (H^{-1})^{\otimes 2n} f = [1,0,\ldots,0,\pm i]$. All signatures in Pl-Holant$([1,0,1],[1,0,1,0,1],\ldots \mid \hat{f})$ satisfy parity constraints. Thus we cannot construct $(H^{-1})^{\otimes 2}[1,\pm i]^{\otimes 2} = \pm\frac{i}{2}[1,\mp i]^{\otimes 2}$ by gadget construction. This implies that we cannot get $[1,\pm i]^{\otimes 2}$ in Pl-#CSP$^2(f)$ by gadget construction.

F No-Mixing of a Pair of Signatures of Even Arity

The general theme of this section and the next is that, for planar Pl-#CSP^2 problems, various tractable signatures of different types cannot mix. In these two sections, all signatures are of even arity. In this section we prove a No-Mixing theorem for a pair of signatures. This will be extended to a set of signatures in the next section.

The general form of the No-Mixing theorem to be proved in this section is as follows: Let f and g be two symmetric signatures of even arity. Suppose for some $1 \leq j < i \leq 5$, $f \in S_i \setminus S_j$ and $g \in S_j \setminus S_i$, and for all $1 \leq k \leq 5$, $\{f,g\} \not\subseteq S_k$. Then Pl-#CSP$^2(f,g)$ is #P-hard. We will call such a statement No-Mixing-(i,j).

It is easy to see that, with possibly switching the names f and g, the condition stated above is equivalent to the following assumption:

$$\{f,g\} \subset \bigcup_{k=1}^{5} S_k \text{ but for any } 1 \leq k \leq 5, \text{ we have } \{f,g\} \not\subseteq S_k.$$

However under this assumption, we make the following observation that any index i for which $f \in S_i$ can be chosen as the distinguishing index:

If $f \in S_i$ for some i, then there exists some $j \neq i$ such that $g \in S_j \setminus S_i$ and $f \in S_i \setminus S_j$.

In particular, neither f nor g can be identically 0.

We will prove the No-Mixing theorem-(i, j) in a reverse lexicographic order of (i, j): We order the statements as $(5, 4), (5, 3), (5, 2), (5, 1), (4, 3), (4, 2), (4, 1), (3, 2), (3, 1), (2, 1)$. After having proved all No-Mixing theorem-(i', j') preceding (i, j) in this order, we assume there are two signatures f and g such that $f \in S_i \setminus S_j$ and $g \in S_j \setminus S_i$. Now we *may* make the following additional assumption:

$$f, g \notin \bigcup_{i<k\leq 5} S_k \qquad \text{and} \qquad g \notin \bigcup_{j<k\leq i} S_k.$$

Indeed, if f or g belongs to S_k for some $k > i$, then let k be the maximum index such that S_k contains either f or g. Then by the observation above, there exists some $j \neq k$ such that one signature belongs to $S_j \setminus S_k$, and the other one belongs to $S_k \setminus S_j$. By the maximality of k, we have $k > j$. Since $k > i$ and No-Mixing theorem-(k, j) has already been proved, we have Pl-#CSP$^2(f, g)$ is #P-hard. Moreover, if $g \in \bigcup_{j<\ell\leq i} S_\ell$, then $g \in S_\ell$ for some $j < \ell < i$, as $g \notin S_i$. Then $f \in S_i \setminus S_\ell$ since $\{f, g\} \not\subseteq S_\ell$, and also $g \in S_\ell \setminus S_i$. Hence Pl-#CSP$^2(f, g)$ is #P-hard by No-Mixing-(i, k) already proved.

We now proceed with this plan. We first prove a preliminary result, which allows us to construct signatures of arbitrarily high even arities from a given binary signature.

Lemma F.1. *For any binary signature $[a, b, c]$, any integer $k \geq 1$, and any signature set $\mathcal{F}$,*

$$\text{Pl-\#CSP}^2([a, b]^{\otimes 2k} + [b, c]^{\otimes 2k}, \mathcal{F}) \leq_T \text{Pl-\#CSP}^2([a, b, c], \mathcal{F}).$$

Proof. We take $2k$ copies of $[a, b, c]$ and connect one input of each $[a, b, c]$ to an edge of $=_{2k}$. The resulting signature is $[a, b]^{\otimes 2k} + [b, c]^{\otimes 2k}$, since $\left[\begin{smallmatrix} a & b \\ b & c \end{smallmatrix}\right]^{\otimes n} \left(\left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right]^{\otimes n} + \left[\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right]^{\otimes n} \right) = \left[\begin{smallmatrix} a \\ b \end{smallmatrix}\right]^{\otimes n} + \left[\begin{smallmatrix} b \\ c \end{smallmatrix}\right]^{\otimes n}$. □

In the next lemma, we will prove that for any symmetric signature $f \in \mathscr{A} \setminus \mathscr{P}$ of even arity, we can construct an arity 4 signature $g \in \mathscr{A} \setminus \mathscr{P}$ in Pl-#CSP$^2(\{f\} \cup \mathcal{F})$. Thus we can assume that we have an arity 4 signature $g \in \mathscr{A} \setminus \mathscr{P}$ in the proof of the No-Mixing lemma of $\mathscr{P}$ versus $\mathscr{A}$, namely No-Mixing-$(5, 4)$. We can prove a similar result for $\mathscr{A}^\dagger \setminus \mathscr{P}$. This is for the proof of No-Mixing-$(5, 3)$.

Lemma F.2. *For any symmetric signature $f \in \mathscr{A} \setminus \mathscr{P}$ (respectively, $f \in \mathscr{A}^\dagger \setminus \mathscr{P}$) of even arity $2n \geq 2$, there exists a symmetric signature $g \in \mathscr{A} \setminus \mathscr{P}$ (respectively, $g \in \mathscr{A}^\dagger \setminus \mathscr{P}$) of arity 4, such that for any set $\mathcal{F}$,*

$$\text{Pl-\#CSP}^2(\{g\} \cup \mathcal{F}) \leq_T \text{Pl-\#CSP}^2(\{f\} \cup \mathcal{F}).$$

Proof. If f has arity $2n = 4$, then there is nothing to prove. Suppose $2n \neq 4$. For $f \in \mathscr{A}^\dagger \setminus \mathscr{P}$, if $2n = 2$, then $f = [1, \alpha, -\alpha^2]$ by Corollary A.9. By Lemma F.1, we have $g = [1, \alpha]^{\otimes 4} - [1, -\alpha]^{\otimes 4}$, since $\alpha^4 = -1$. Clearly $g \in \mathscr{A}^\dagger$ and is non-degenerate. Note that g satisfies a second recurrence relation of type $\langle -\alpha^2, 0, 1\rangle$, since the eigenvalues of the recurrence are $\pm\alpha$ with trace 0 and product $-\alpha^2$. Thus $g \notin \mathscr{P}$ by Lemma A.11. For $2n \geq 6$, we have $f = [1, \alpha]^{\otimes 2n} + i^r[1, -\alpha]^{\otimes 2n}$ by definitions (see Figure 35). Then by our calculus, we have $\partial^{n-2}(f) = (1 + \alpha^2)^{n-2}\{[1, \alpha]^{\otimes 4} + i^r[1, -\alpha]^{\otimes 4}\}$. Clearly it is in $\mathscr{A}^\dagger$ and is non-degenerate. It also has type $\langle -\alpha^2, 0, 1\rangle$ and therefore it is not in $\mathscr{P}$.

For $f \in \mathscr{A} \setminus \mathscr{P}$, if $2n = 2$, then $f = [1, \rho, -\rho^2]$ by Corollary A.9. By Lemma F.1, we have $g = [1, \rho]^{\otimes 4} + [1, -\rho]^{\otimes 4}$, since $\rho^4 = 1$. Clearly $g \in \mathscr{A}$ and is non-degenerate. Note that g has type $\langle -\rho^2, 0, 1\rangle$, since the eigenvalues of its second recurrence relation are $\pm\rho$ with trace 0 and product $-\rho^2$. Thus $g \notin \mathscr{P}$ by Lemma A.11.

For $2n \geq 6$, we have $f = [1, \rho]^{\otimes 2n} + i^r [1, -\rho]^{\otimes 2n}$ by definitions (see Figure 35). If $2n \equiv 0 \pmod 4$, then n is even, and we have $\partial_{=_4^{\frac{n-2}{2}}}(f) = 2^{\frac{n-2}{2}}\{[1, \rho]^{\otimes 4} + i^r [1, -\rho]^{\otimes 4}\}$ that is in $\mathscr{A}$, and not in $\mathscr{P}$ by its type $\langle -\rho^2, 0, 1\rangle$. For $2n \equiv 2 \pmod 4$, we have $h = \partial_{=_4^{\frac{n-1}{2}}}(f) = 2^{\frac{n-1}{2}}\{[1, \rho]^{\otimes 2} + i^r [1, -\rho]^{\otimes 2}\}$.

- If $r = 2$, then we have $h = 2^{\frac{n-1}{2}}[0, 2\rho, 0]$. Thus we have $[0, 1, 0]$ up to a nonzero scalar and $\partial^{n-2}_{[0,1,0]}(f) = (2\rho)^{n-2}\{[1, \rho]^{\otimes 4} + i^r (-1)^{n-2} [1, -\rho]^{\otimes 4}\}$ that is in $\mathscr{A}$, and not in $\mathscr{P}$ by its type $\langle -\rho^2, 0, 1\rangle$.
- If $r \neq 2$, then $h = 2^{\frac{n-1}{2}}(1 + i^r)[1, \frac{1-i^r}{1+i^r}\rho, \rho^2]$. Then we have $\partial_{[1, \frac{1-i^r}{1+i^r}\rho, \rho^2]}(=_4) = [1, 0, \rho^2]$ on LHS and $\partial^{n-2}_{[1,0,\rho^2]}(f) = 2^{n-2}\{[1, \rho]^{\otimes 4} + i^r [1, -\rho]^{\otimes 4}\}$ by $\rho^4 = 1$, that is in $\mathscr{A} \setminus \mathscr{P}$ by the same reason. □

We note that the complication for the case $f \in \mathscr{A} \setminus \mathscr{P}$ is unavoidable since if $\rho = \pm i$, then $\partial(f) = 0$, therefore we need to use $\partial_{=_4}(f)$.

F.1 Mixing with $\mathscr{P}$

In this subsection, we prove No-Mixing-$(5, j)$, for $1 \leq j \leq 4$, namely the No-Mixing of one signature in $\mathscr{P}$ and another signature in a different tractable set. Thus we assume there is some $f \in S_5 = \mathscr{P}$, and some $g \in S_k$ for some $1 \leq k \leq 4$, and for no $1 \leq k \leq 5$, $\{f, g\} \subset S_k$. Under this assumption we show that Pl-#CSP$^2(f, g)$ is #P-hard. As explained earlier, for $j < k < 5$, when we prove No-Mixing-$(5, j)$, we can make logical use of No-Mixing-$(5, k)$.

Lemma F.3. *Let $\{f, g\} \subseteq \bigcup_{k=1}^{5} S_k$ and $\{f, g\} \nsubseteq S_j$ for every $1 \leq j \leq 5$. Assume that $f \in \mathscr{P}$, then* Pl-#CSP$^2(f, g)$ *is* #P*-hard.*

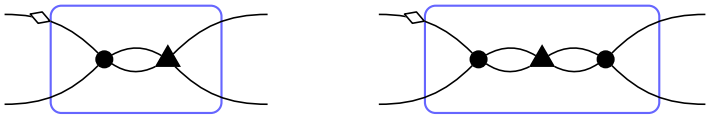
Proof. As explained earlier, since $f \in \mathscr{P}$, there exists some $1 \leq k \leq 4$, such that $g \in S_k \setminus \mathscr{P}$ and $f \in \mathscr{P} \setminus S_k$. Since $[0, 1, 0] \in \bigcap_{k=1}^{5} S_k$, we know that f is not a multiple of $[0, 1, 0]$. Then by $f \in \mathscr{P}$ (see Figure 35), we have $f = [a, b]^{\otimes 2n}$ with a and b not both 0 (because f is not identically 0), or $f = [1, 0, \ldots, 0, x]$ with $x \neq 0$.

We first consider the case $f = [a, b]^{\otimes 2n}$, with $(a, b) \neq (0, 0)$. It has three subcases.

- If $ab \neq 0$ (i.e., a and b both nonzero) and $a^2 + b^2 \neq 0$, then we have $\partial^{n-1}(f) = (a^2 + b^2)^{n-1}[a, b]^{\otimes 2}$. We are done by Lemma B.3.
- If $ab \neq 0$ and $a^2 + b^2 = 0$, then $f = [1, \pm i]^{\otimes 2n}$ up to a nonzero scalar. Note that $f \in \mathscr{P} \cap \mathscr{A} \cap \widehat{\mathscr{M}^\dagger}$. Hence $g \in \mathscr{A}^\dagger \setminus (\mathscr{P} \cup \mathscr{A} \cup \widehat{\mathscr{M}^\dagger})$ or $g \in \widehat{\mathscr{M}} \setminus (\mathscr{P} \cup \mathscr{A} \cup \widehat{\mathscr{M}^\dagger})$.
  If $g \in \widehat{\mathscr{M}} \setminus (\mathscr{P} \cup \widetilde{\mathscr{A}} \cup \widehat{\mathscr{M}^\dagger})$, then a fortiori, $g \in \widehat{\mathscr{M}} \setminus (\mathscr{P} \cup \widetilde{\mathscr{A}})$. Therefore we are done by Lemma E.3.
  The other case is $g \in \mathscr{A}^\dagger \setminus (\mathscr{P} \cup \mathscr{A} \cup \widehat{\mathscr{M}^\dagger})$, then a fortiori, $g \in \mathscr{A}^\dagger \setminus \mathscr{P}$, and by Lemma F.2, we have an arity 4 signature $g' \in \mathscr{A}^\dagger \setminus \mathscr{P}$. By definition (see Figure 35), $g' = [1, \alpha]^{\otimes 4} + i^r [1, -\alpha]^{\otimes 4}$. For $r = 2$, we have $\partial(g') = 2\alpha(1 + \alpha^2)[0, 1, 0]$ and $\partial^{n-1}_{[0,1,0]}(f) = (\pm 2i)^{n-1}[1, \pm i]^{\otimes 2}$. Then we are done by Lemma B.3. For $r \neq 2$, we have on LHS

$$\partial_{g'}(=_6) = \partial_{[1,\alpha]^{\otimes 4}}(=_6) + i^r \partial_{[1,-\alpha]^{\otimes 4}}(=_6) = [1, 0, \alpha^4] + i^r [1, 0, (-\alpha)^4] = (1 + i^r)[1, 0, -1]$$

  and $\partial^{n-1}_{[1,0,-1]}(f) = 2^{n-1}[1, \pm i]^{\otimes 2}$. Then again we are done by Lemma B.3.



(a) (b)

Figure 32: Two gadgets used in the proof of Lemma F.4.

- For $f = [1,0]^{\otimes 2n}$ or $[0,1]^{\otimes 2n}$, we have $\partial^{n-1}(f) = [1,0]^{\otimes 2}$ or $[0,1]^{\otimes 2}$. Note that $f \in \mathscr{P} \cap \mathscr{A} \cap \mathscr{A}^\dagger$. Thus $g \in \widetilde{\mathscr{M}} \setminus (\mathscr{P} \cup \widetilde{\mathscr{A}})$. If $g \in \widehat{\mathscr{M}} \setminus (\mathscr{P} \cup \widetilde{\mathscr{A}})$, then we are done by Lemma E.3. If $g \in \widehat{\mathscr{M}^\dagger} \setminus (\mathscr{P} \cup \widetilde{\mathscr{A}})$, then we are done by Lemma E.7, where the binary signature is supplied by $\partial^{n-1}(f) = [1,0]^{\otimes 2}$ or $[0,1]^{\otimes 2}$.

The remaining case is $f = [1,0,\ldots,0,x]$ with $x \neq 0$. We have $\partial^{n-1}(f) = [1,0,x]$.

Suppose $g \in \widetilde{\mathscr{A}}$. As $f \in \mathscr{P}$, we have $g \notin \mathscr{P}$. Then we have an arity 4 signature $g' \in \widetilde{\mathscr{A}} \setminus \mathscr{P}$ by Lemma F.2. Moreover, by definition (see Figure 35), we have $g' = [1,\gamma]^{\otimes 4} + i^r[1,-\gamma]^{\otimes 4}$ where $\gamma^8 = 1$. Depending on whether $g \in \mathscr{A}$ or $\mathscr{A}^\dagger$, we have either $f \in \mathscr{P} \setminus \mathscr{A}$, or $f \in \mathscr{P} \setminus \mathscr{A}^\dagger$. Then we claim that $x^4 \neq 1$. Note that f has even arity $2n$. If $x^4 = 1$, then $f = [1,0,\ldots,0,x] \in \mathscr{A}$ as well as $\left[\begin{smallmatrix}1 & 0\\ 0 & \alpha\end{smallmatrix}\right]^{\otimes 2n} f = [1,0,\ldots,0,xi^n] \in \mathscr{A}$ thus $f \in \mathscr{A}^\dagger$. This is a contradiction. Thus we have $x^4 \neq 0,1$. Let $\widehat{g'} = [1, x^{-\frac{1}{2}}\gamma]^{\otimes 4} + i^r[1,-x^{-\frac{1}{2}}\gamma]^{\otimes 4}$. Then by Lemma C.8, Pl-#CSP$^2(\widehat{g'}) \leq$Pl-#CSP$^2(f,g)$. Note that $\widehat{g'}$ has type $\langle -x^{-1}\gamma^2, 0, 1\rangle$ by calculating the trace and product of the eigenvalues of the second recurrence relation. Note that $(-x^{-1}\gamma^2)^4 = x^{-4} \neq 0,1$. Thus $\widehat{g'} \notin \mathscr{P} \cup \widetilde{\mathscr{A}} \cup \widetilde{\mathscr{M}}$ by Lemma A.11. This implies that Pl-#CSP$^2(\widehat{g'})$ is #P-hard by Theorem D.5. So Pl-#CSP$^2(f,g)$ is #P-hard.

Now we may assume that $g \notin \widetilde{\mathscr{A}}$. Thus $g \in \widetilde{\mathscr{M}} \setminus (\mathscr{P} \cup \widetilde{\mathscr{A}})$. If $g \in \widehat{\mathscr{M}} \setminus (\mathscr{P} \cup \widetilde{\mathscr{A}})$, then we are done by Lemma E.3. If $g \in \widehat{\mathscr{M}^\dagger} \setminus (\mathscr{P} \cup \widetilde{\mathscr{A}})$, then $f \in \mathscr{P} \setminus \widehat{\mathscr{M}^\dagger}$. In this case we claim that $x \neq 1$. Suppose for a contradiction that $x = 1$, then we show that $f \in \widehat{\mathscr{M}^\dagger}$. Notice that $f = [1,0,\ldots,0,1] = (=_{2n})$ and $\widehat{\mathscr{M}^\dagger} = Z\mathscr{M}$, where $Z = \left[\begin{smallmatrix}1 & 1\\ i & -i\end{smallmatrix}\right]$. Crucially recall that f has even arity. Then, up to a nonzero scalar, $(Z^{-1})^{\otimes 2n} f = [1,0,1,\ldots,0,1] \in \mathscr{M}$ of arity $2n$ (if n is even) or $(Z^{-1})^{\otimes 2n} f = [0,1,0,\ldots,1,0] \in \mathscr{M}$ of arity $2n$ (if n is odd). Hence $x \neq 1$. Then we are done by Lemma E.7, with $g \in \widehat{\mathscr{M}^\dagger} \setminus (\mathscr{P} \cup \widetilde{\mathscr{A}})$, and the help of $\partial^{n-1}(f) = [1,0,x]$. □

F.2 Mixing with $\mathscr{A}$

In this subsection, we prove the No-Mixing lemma of $\mathscr{A}$ with other tractable sets. Because we have already proved Lemma F.3, the No-Mixing lemma for $S_5 = \mathscr{P}$, we only need to consider No-Mixing-$(4,j)$ of $S_4 = \mathscr{A}$ with S_j for $1 \leq j \leq 3$.

There is a particular case involving $\mathscr{A}$ and $\mathscr{A}^\dagger$ that requires some special care. This is when two signatures $f \in \mathscr{A}$ and $g \in \mathscr{A}^\dagger$ both satisfy the parity constraint. We deal with this case first. Furthermore, by Lemma F.2, for two signatures $f \in \mathscr{A} \setminus \mathscr{P}$ and $g \in \mathscr{A}^\dagger \setminus \mathscr{P}$ we may assume the signatures f and g have arity 4. Hence the next lemma considers signatures f and g of arity 4.

Lemma F.4. *Let* $f = [1,\rho]^{\otimes 4} \pm [1,-\rho]^{\otimes 4} \in \mathscr{A}$ *and* $g = [1,\alpha]^{\otimes 4} \pm [1,-\alpha]^{\otimes 4} \in \mathscr{A}^\dagger$. *Then* Pl-#CSP$^2(f,g)$ *is* #P*-hard.*

Proof. There are four cases depending on the combination of the two $\pm$ signs. Suppose $f = [1,\rho]^{\otimes 4} + [1,-\rho]^{\otimes 4}$ and $g = [1,\alpha]^{\otimes 4} + [1,-\alpha]^{\otimes 4}$. Consider the gadget in Figure 32a. We assign g to the circle vertex and f to the triangle vertex. Since both $f = 2[1,0,\rho^2,0,1]$ and $g = 2[1,0,\alpha^2,0,-1]$

have even parity, the signature of this gadget also has even parity. It is also clearly a redundant signature by design. Hence there are only five signature entries we need to compute. E.g., the entry of Hamming weight 0 is $g_0 f_0 + g_2 f_2 = 4(1 + \alpha^2\rho^2)$. Up to a factor of 4, the signature of this gadget has signature matrix

$$\begin{bmatrix} \alpha^2\rho^2+1 & 0 & 0 & \alpha^2+\rho^2 \\ 0 & 2\alpha^2\rho^2 & 2\alpha^2\rho^2 & 0 \\ 0 & 2\alpha^2\rho^2 & 2\alpha^2\rho^2 & 0 \\ \alpha^2-\rho^2 & 0 & 0 & \alpha^2\rho^2-1 \end{bmatrix}, \text{ which becomes } \begin{bmatrix} \alpha^2\rho^2+1 & 0 & 0 & 2\alpha^2\rho^2 \\ 0 & \alpha^2-\rho^2 & 2\alpha^2\rho^2 & 0 \\ 0 & 2\alpha^2\rho^2 & \alpha^2+\rho^2 & 0 \\ 2\alpha^2\rho^2 & 0 & 0 & \alpha^2\rho^2-1 \end{bmatrix}$$

after a 90° counterclockwise rotation of the gadget. (See Figure 2 in Part I for an illustration of the rotation operation.) Taking the four corner entries, we define the binary signature $h = [\alpha^2\rho^2+1, 2\alpha^2\rho^2, \alpha^2\rho^2-1]$. By domain pairing, Pl-#CSP$(h) \le_T$ Pl-#CSP$^2(f, g)$. (Domain pairing is the following reduction: In an instance of Pl-#CSP(h) replace every occurrence of h by a copy of the 90° counterclockwise rotated gadget, and replace both edges of h by two parallel edges each, and replace every $(=_k)$ in the Pl-#CSP(h) instance by $(=_{2k})$ in Pl-#CSP$^2(f, g)$. Note that the rotation is necessary to create a *symmetric* binary signature h in the paired domain.)

Note that $\alpha^2 = \pm i$ and $\rho^2 = \pm 1$, so $\alpha^2\rho^2 \pm 1$ has norm $\sqrt{2}$, while $2\alpha^2\rho^2$ has norm 2. Also $\alpha^2\rho^2 + 1 \neq \alpha^2\rho^2 - 1$. Hence $h \notin \mathscr{P} \cup \mathscr{A}$ by Corollary A.9 and also $h \notin \widehat{\mathscr{M}}$ by Lemma A.8. Thus Pl-#CSP(h) is #P-hard by Theorem A.22. So Pl-#CSP$^2(f, g)$ is #P-hard.

Suppose $f = [1, \rho]^{\otimes 4} - [1, -\rho]^{\otimes 4}$ and $g = [1, \alpha]^{\otimes 4} - [1, -\alpha]^{\otimes 4}$. Consider the same construction. Up to a nonzero factor of $4\alpha\rho$, the signature of this gadget has the signature matrix

$$\begin{bmatrix} 2 & 0 & 0 & 2\rho^2 \\ 0 & 1+\alpha^2\rho^2 & 1+\alpha^2\rho^2 & 0 \\ 0 & 1+\alpha^2\rho^2 & 1+\alpha^2\rho^2 & 0 \\ 2\alpha^2 & 0 & 0 & 2\alpha^2\rho^2 \end{bmatrix}, \text{ which becomes } \begin{bmatrix} 2 & 0 & 0 & 1+\alpha^2\rho^2 \\ 0 & 2\alpha^2 & 1+\alpha^2\rho^2 & 0 \\ 0 & 1+\alpha^2\rho^2 & 2\rho^2 & 0 \\ 1+\alpha^2\rho^2 & 0 & 0 & 2\alpha^2\rho^2 \end{bmatrix}$$

after a 90° counterclockwise rotation of the gadget. Let $h = [2, 1+\alpha^2\rho^2, 2\alpha^2\rho^2]$. By domain pairing, we have Pl-#CSP$(h) \le_T$ Pl-#CSP$^2(f, g)$. Note that $1+\alpha^2\rho^2 = 1 \pm i$ has norm $\sqrt{2}$ while $2\alpha^2\rho^2 \neq 2$ but has norm 2. Hence $h \notin \mathscr{P} \cup \mathscr{A} \cup \widehat{\mathscr{M}}$ by Corollary A.9 and Lemma A.8. Thus we are done by Theorem A.22.

Suppose $f = [1, \rho]^{\otimes 4} - [1, -\rho]^{\otimes 4}$ and $g = [1, \alpha]^{\otimes 4} + [1, -\alpha]^{\otimes 4}$. Consider the gadget in Figure 32b. We assign f to the circle vertices and g to the triangle vertex. Up to a nonzero factor of $16\alpha^2\rho^2$, the signature of this gadget has the signature matrix

$$\begin{bmatrix} 2 & 0 & 0 & 2\rho^2 \\ 0 & \rho^2 & \rho^2 & 0 \\ 0 & \rho^2 & \rho^2 & 0 \\ 2\rho^2 & 0 & 0 & 2 \end{bmatrix}, \quad \text{which becomes} \quad \begin{bmatrix} 2 & 0 & 0 & \rho^2 \\ 0 & 2\rho^2 & \rho^2 & 0 \\ 0 & \rho^2 & 2\rho^2 & 0 \\ \rho^2 & 0 & 0 & 2 \end{bmatrix}$$

after a 90° rotation of the gadget. Let $h = [2, \rho^2, 2]$. We also have $g^\times = 2[1, \alpha^2]^{\otimes 2}$ by domain pairing with g (see Lemma A.19). Then Pl-#CSP$(g^\times, h) \le_T$ Pl-#CSP$^2(f, g)$. Note that $|\rho^2| = 1 \neq 2$, so by Lemma A.8 and Corollary A.9, $h \in \widehat{\mathscr{M}} \setminus (\mathscr{P} \cup \mathscr{A})$. Also by Lemma A.8 and $(\alpha^2)^2 = -1 \neq 1$ we have $g^\times \notin \widehat{\mathscr{M}}$. Thus we are done by Theorem A.22. Note that in this case, the rotation is necessary to create a *non-degenerate* binary signature h in the paired domain.

Finally, suppose $f = [1,\rho]^{\otimes 4} + [1,-\rho]^{\otimes 4}$ and $g = [1,\alpha]^{\otimes 4} - [1,-\alpha]^{\otimes 4}$. Consider the gadget in Figure 32b. We assign g to the circle vertices and f to the triangle vertex. Up to a nonzero factor of $16\alpha^2\rho^2$, the signature of this gadget has the signature matrix

$$\begin{bmatrix} 2 & 0 & 0 & 2\alpha^2 \\ 0 & \alpha^2 & \alpha^2 & 0 \\ 0 & \alpha^2 & \alpha^2 & 0 \\ 2\alpha^2 & 0 & 0 & -2 \end{bmatrix}, \quad \text{which becomes} \quad \begin{bmatrix} 2 & 0 & 0 & \alpha^2 \\ 0 & \alpha^2 & 2\alpha^2 & 0 \\ 0 & 2\alpha^2 & \alpha^2 & 0 \\ \alpha^2 & 0 & 0 & -2 \end{bmatrix}$$

after a 90° rotation of the gadget. Let $h = [2, \alpha^2, -2]$, then Pl-#CSP$(h) \le_T$ Pl-#CSP$^2(f,g)$ by domain pairing. Since $|\alpha^2| = 1 \neq 2$, we have $h \notin \mathscr{P} \cup \mathscr{A}$ by Corollary A.9 and also $h \notin \cup\widehat{\mathscr{M}}$ by Lemma A.8. Thus we are done by Theorem A.22. Note that in this case, the rotation is also necessary to create a *non-degenerate* binary signature h in the paired domain. □

Remark 7. The use of a more complicated construction in the third case is necessary. Notice that $g = [1,\alpha]^{\otimes 4} + [1,-\alpha]^{\otimes 4} = 2[1,0,\alpha^2,0,-1]$ has an even parity, while $f = [1,\rho]^{\otimes 4} - [1,-\rho]^{\otimes 4} = 2\rho[0,1,0,\rho^2,0]$ has an odd parity. Then in any construction of a signature using f and g, if the number of occurrences N_f of f is odd (resp. even), then the resulting signature also has an odd (resp. even) parity. To see this, let H be an arbitrary $\{f,g\}$-gate with N_f occurrences of f. Suppose σ is a $\{0,1\}$-assignment to all the edges of H, including internal and external edges, that has a nonzero evaluation on H. Then each copy of f has an odd number of incident edges assigned to 1. Summing these numbers (mod 2) over all copies of f we get a number $\equiv N_f \pmod 2$, since each of these numbers is $\equiv 1 \pmod 2$. Similarly each copy of g has an even number of incident edges assigned to 1. Summing these numbers (mod 2) over all copies of g we get a number $\equiv 0 \pmod 2$. On the other hand, if we add these two sums together we get $2X + Y$ where X is the number of internal edges and Y is the number of external edges assigned to 1 by σ. This is because each internal edge assigned to 1 appears exactly twice in the sum. Hence this number is $\equiv Y \pmod 2$. We conclude that $N_f \equiv Y \pmod 2$, the Hamming weight of σ on the external edges.

If N_f is odd, from any constructed signature of arity 4, by rotation and domain pairing we can only get the identically zero binary signature. Thus we must use f an even number of times. Using g alone will not get out of $\mathscr{A}^\dagger$, which is a tractable set. Thus we must use f at least twice. Also using g alone will not get out of $\mathscr{A}$, another tractable set. Therefore we must use g at least once. Therefore the construction we give is the simplest possible.

The same consideration applies for the construction in the fourth case.

The next Lemma deals with the situation when we have a binary signature in $\mathscr{A} \setminus \mathscr{P}$ and an arity 4 signature in $\mathscr{A}^\dagger \setminus \mathscr{P}$.

Lemma F.5. *Let* $f = [1,\rho,-\rho^2]$ *and* $g = [1,\alpha]^{\otimes 4} + i^r[1,-\alpha]^{\otimes 4}$. *Then* Pl-#CSP$^2(f,g)$ *is* #P*-hard.*

Proof. By our calculus, we have $\partial_{[1,\rho,-\rho^2]}(g) = \lambda[1,\alpha]^{\otimes 2} + i^r\mu[1,-\alpha]^{\otimes 2}$, where $\lambda = 1 - \rho^2\alpha^2 + 2\rho\alpha$ and $\mu = 1 - \rho^2\alpha^2 - 2\rho\alpha$. Note that $1 - \rho^2\alpha^2 = 1 \pm i$ has norm $\sqrt{2}$ and $|2\rho\alpha| = 2$, we have $\lambda \neq 0$. Let $x = i^r\mu/\lambda$, then $\partial_{[1,\rho,-\rho^2]}(g) = \lambda(1+x)[1, \frac{1-x}{1+x}\alpha, \alpha^2]$. By norm, $(1-\rho^2\alpha^2)^4 \neq (2\rho\alpha)^4$ and $(1-\rho^2\alpha^2)(2\rho\alpha) \neq 0$, we have $x^4 \neq 0, 1$ by Lemma A.3. By Lemma A.3 again, we have $(\frac{1-x}{1+x})^4 \neq 0, 1$. Thus $[1, \frac{1-x}{1+x}\alpha, \alpha^2] \notin \mathscr{P} \cup \widetilde{\mathscr{A}} \cup \widetilde{\mathscr{M}}$ by Corollary A.9. This implies that Pl-#CSP$^2([1, \frac{1-x}{1+x}\alpha, \alpha^2])$ is #P-hard by Theorem A.21′. Thus Pl-#CSP$^2(f,g)$ is #P-hard. □

The next lemma is the No-Mixing lemma of $\mathscr{A}$ with the other tractable sets, namely the statements No-Mixing-$(4, j)$ for $1 \le j \le 3$. Having already proved Lemma F.3, we can assume that both f and g are not in $S_5 = \mathscr{P}$.

Lemma F.6. *Let* $\{f, g\} \subseteq \left(\bigcup_{k=1}^{4} S_k\right) \setminus S_5$ *and* $\{f, g\} \nsubseteq S_j$ *for every* $1 \le j \le 4$. *Assume that* $f \in \mathscr{A}$, *then* Pl-#CSP$^2(f, g)$ *is* #P*-hard.*

Proof. By $f \in \mathscr{A}$, we have $g \notin \mathscr{A}$. Thus, $g \in (\mathscr{A}^\dagger \cup \widehat{\mathscr{M}} \cup \widehat{\mathscr{M}^\dagger}) \setminus (\mathscr{P} \cup \mathscr{A})$.

1. Suppose $g \in \mathscr{A}^\dagger \setminus (\mathscr{P} \cup \mathscr{A})$. Then a fortiori, $g \in \mathscr{A}^\dagger \setminus \mathscr{P}$. As $f \in \mathscr{A} \setminus \mathscr{P}$, by Lemma F.2, we have some $f' \in \mathscr{A} \setminus \mathscr{P}$ and $g' \in \mathscr{A}^\dagger \setminus \mathscr{P}$, both of arity 4. Without loss of generality, we will assume the given f and g are of arity 4. By definition (see Figure 35), we can assume that

$$f = [1, \rho]^{\otimes 4} + i^r[1, -\rho]^{\otimes 4} \quad \text{and} \quad g = [1, \alpha]^{\otimes 4} + i^s[1, -\alpha]^{\otimes 4} \quad \text{where } r, s = 0, 1, 2, 3.$$

   - If both $r, s \equiv 0 \pmod 2$, then $f = [1, \rho]^{\otimes 4} \pm [1, -\rho]^{\otimes 4}$ and $g = [1, \alpha]^{\otimes 4} \pm [1, -\alpha]^{\otimes 4}$. This is the case where both f and g satisfy the parity constraint, and it is proved in Lemma F.4.
   - If $r \equiv 1 \pmod 2$ then $f = [1, \rho]^{\otimes 4} \pm i[1, -\rho]^{\otimes 4}$. For $\rho^2 = 1$, by our calculus we have

$$\partial(f) = 2\{[1, \rho]^{\otimes 2} \pm i[1, -\rho]^{\otimes 2}\} = 2(1 \pm i)[1, \mp i\rho, \rho^2] = 2(1 \pm i)[1, \rho', -\rho'^2],$$

     where $\rho' = \mp i\rho$, and $\rho'^4 = 1$. Thus Pl$-$#CSP$^2([1, \rho', -\rho'^2], g)$ is #P-hard by Lemma F.5. So Pl$-$#CSP$^2(f, g)$ is #P-hard.
     For $\rho^2 = -1$, we cannot use $[1, 0, 1]$ to reduce the arity of f, because $\partial(f) = 0$ in this case. Instead we construct a suitable binary signature from g. If $s \neq 2$, then we have $g_0 = 1 + i^s \neq 0$ and $g_4 = \alpha^4 + i^s(-\alpha)^4 = -(1 + i^s) = -g_0$, and therefore $\partial_g(=_6) = g_0[1, 0, -1]$ on the LHS. Then we have $\partial_{[1,0,-1]}(f) = 2\{[1, \rho]^{\otimes 2} \pm i[1, -\rho]^{\otimes 2}\} = 2(1 \pm i)[1, \mp i\rho, \rho^2] = 2(1 \pm i)[1, \rho', -\rho'^2]$, where $\rho' = \mp i\rho$ and $\rho'^4 = 1$. Then we are done by Lemma F.5. If $s = 2$, then $\partial(g) = (1 + \alpha^2)\{[1, \alpha]^{\otimes 2} - [1, -\alpha]^{\otimes 2}\}$, a nonzero multiple of $[0, 1, 0]$. Thus we have $\partial_{[0,1,0]}(f) = 2\rho\{[1, \rho]^{\otimes 2} \mp i[1, -\rho]^{\otimes 2}\} = 2\rho(1 \mp i)[1, \pm i\rho, \rho^2] = 2\rho(1 \mp i)[1, \rho', -\rho'^2]$, where $\rho' = \pm i\rho$ and $\rho'^4 = 1$. Then we are done by Lemma F.5 again.
   - If $r \equiv 0 \pmod 2$ and $s \equiv 1 \pmod 2$, i.e., $f = [1, \rho]^{\otimes 4} \pm [1, -\rho]^{\otimes 4}$ and $g = [1, \alpha]^{\otimes 4} \pm i[1, -\alpha]^{\otimes 4}$, then we will construct a binary signature $h = [1, b, \pm 1]$. Note that $h \in \widetilde{\mathscr{M}}$ by Lemma A.8. Furthermore, we will ensure that $b^4 \neq 0, 1$, thus $h \notin \mathscr{P} \cup \widetilde{\mathscr{A}}$ by Corollary A.9. Then we are done by Lemma E.3 and Lemma E.7.
     We have $\partial(g) = (1 + \alpha^2)\{[1, \alpha]^{\otimes 2} \pm i[1, -\alpha]^{\otimes 2}\} = (1 + \alpha^2)(1 \pm i)[1, \mp i\alpha, \alpha^2]$, a nonzero multiple of $[1, \alpha', -\alpha'^2]$, where $\alpha' = \mp i\alpha$ and $\alpha'^4 = -1$. Moreover, we have $h = \partial_{[1,\alpha',-\alpha'^2]}(f) = \lambda[1, \rho]^{\otimes 2} \pm \mu[1, -\rho]^{\otimes 2}$, where $\lambda = 1 - \rho^2\alpha'^2 + 2\rho\alpha'$ and $\mu = 1 - \rho^2\alpha'^2 - 2\rho\alpha'$. Then $h = \lambda(1 \pm x)[1, a\rho, \rho^2]$, where $x = \mu/\lambda$ and $a = \frac{1 \mp x}{1 \pm x}$. Note that $1 - \rho^2\alpha'^2 = 1 \pm i$ has norm $\sqrt{2}$ and $|2\rho\alpha'| = 2$, thus $\lambda \neq 0$ and $(1 - \rho^2\alpha'^2)^4 \neq (2\rho\alpha')^4$ by norm, therefore $x^4 \neq 0, 1$ by Lemma A.3. Then by Lemma A.3 again, $a^4 \neq 0, 1$, and so $(a\rho)^4 \neq 0, 1$ as well. As $\lambda \neq 0$, $1 \pm x \neq 0$, $\rho^2 = \pm 1$, we have a nonzero multiple of $[1, a\rho, \pm 1]$, our desired binary signature, and we are done by Lemma E.3 and Lemma E.7.

   In the following we may assume $g \notin \mathscr{A}^\dagger$.
2. Suppose $g \in \widehat{\mathscr{M}} \setminus (\mathscr{P} \cup \mathscr{A})$, then $g \in \widehat{\mathscr{M}} \setminus (\mathscr{P} \cup \widetilde{\mathscr{A}})$. We also have $f \notin \widehat{\mathscr{M}}$, lest $\{f, g\} \subseteq \widehat{\mathscr{M}}$, and we are done by Lemma E.3.

3. Suppose $g \in \widehat{\mathscr{M}^\dagger} \setminus (\mathscr{P} \cup \mathscr{A})$, then $g \in \widehat{\mathscr{M}^\dagger} \setminus (\mathscr{P} \cup \widetilde{\mathscr{A}})$. Now $f \in \mathscr{A} \setminus (\mathscr{P} \cup \widehat{\mathscr{M}^\dagger})$. Note that $[1,1]^{\otimes 2n} \pm [1,-1]^{\otimes 2n} \in \widehat{\mathscr{M}^\dagger}$. This can be verified as follows: Let $Z = \left[\begin{smallmatrix} 1 & 1 \\ i & -i \end{smallmatrix}\right]$, then $\widehat{\mathscr{M}^\dagger} = Z\mathscr{M}$, and $Z^{-1} = \frac{1}{2}\left[\begin{smallmatrix} 1 & -i \\ 1 & i \end{smallmatrix}\right]$. We first verify that $[1,0]^{\otimes 2n} \pm [0,1]^{\otimes 2n} \in \widehat{\mathscr{M}^\dagger}$, by

$$\begin{aligned}\left[\begin{smallmatrix} 1 & -i \\ 1 & i \end{smallmatrix}\right]^{\otimes 2n} \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}^{\otimes 2n} \pm \begin{bmatrix} 0 \\ 1 \end{bmatrix}^{\otimes 2n} \right\} &= \begin{bmatrix} 1 \\ 1 \end{bmatrix}^{\otimes 2n} \pm (-i)^{2n} \begin{bmatrix} 1 \\ -1 \end{bmatrix}^{\otimes 2n} \\ &= \begin{bmatrix} 1 \\ 1 \end{bmatrix}^{\otimes 2n} \pm (-1)^{n} \begin{bmatrix} 1 \\ -1 \end{bmatrix}^{\otimes 2n} \in \mathscr{M}.\end{aligned}$$

Then notice that

$$\begin{bmatrix} 1 \\ 1 \end{bmatrix}^{\otimes 2n} \pm \begin{bmatrix} 1 \\ -1 \end{bmatrix}^{\otimes 2n} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}^{\otimes 2n} \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}^{\otimes 2n} \pm \begin{bmatrix} 0 \\ 1 \end{bmatrix}^{\otimes 2n} \right\} \in \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \widehat{\mathscr{M}^\dagger}.$$

However

$$\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ i & -i \end{bmatrix} = \begin{bmatrix} 1+i & 1-i \\ 1-i & 1+i \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ i & -i \end{bmatrix} \begin{bmatrix} 0 & 1-i \\ 1+i & 0 \end{bmatrix}$$

and $\left[\begin{smallmatrix} 0 & 1-i \\ 1+i & 0 \end{smallmatrix}\right] \mathscr{M} = \mathscr{M}$, therefore $\left[\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}\right] \widehat{\mathscr{M}^\dagger} = \left[\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}\right] Z\mathscr{M} = Z \left[\begin{smallmatrix} 0 & 1-i \\ 1+i & 0 \end{smallmatrix}\right] \mathscr{M} = Z\mathscr{M} = \widehat{\mathscr{M}^\dagger}$. (Also see Figure 36).
Since $[1,1]^{\otimes 2n} \pm [1,-1]^{\otimes 2n} \in \widehat{\mathscr{M}^\dagger}$ and $f \notin \widehat{\mathscr{M}^\dagger}$, f cannot take the form $[1,1]^{\otimes 2n} \pm [1,-1]^{\otimes 2n}$. Then by definition (see Figure 35) f takes the form

$$[1, \rho, -\rho^2], \quad \text{or} \quad [1,1]^{\otimes 2n} \pm i[1,-1]^{\otimes 2n}, \quad \text{or} \quad [1,i]^{\otimes 2n} + i^r[1,-i]^{\otimes 2n}, \quad \text{where} \quad 2n \geq 4.$$

The following three cases are immediately done by Lemma E.7:

- $f = [1, \rho, -\rho^2]$.
- $f = [1,1]^{\otimes 2n} \pm i[1,-1]^{\otimes 2n}$ with $2n \geq 4$, then we have $\partial^{n-1}(f) = 2^{n-1}[1 \pm i, 1 \mp i, 1 \pm i]$ which is not $\lambda[1,0,1]$.
- If $f = [1,i]^{\otimes 2n} + i^r[1,-i]^{\otimes 2n}$ with $2n \equiv 2 \pmod 4$, then we have $\partial^{\frac{n-1}{2}}_{=4}(f) = 2^{\frac{n-1}{2}}[1 + i^r, (1-i^r)i, -(1+i^r)]$ which is not $\lambda[1,0,1]$, no matter what value r takes.

The remaining case is that $f = [1,i]^{\otimes 2n} + i^r[1,-i]^{\otimes 2n}$ with $2n \equiv 0 \pmod 4$. In this case, we have

$$\partial^{\frac{n-2}{2}}_{=4}(f) = 2^{\frac{n-2}{2}}\{[1,i]^{\otimes 4} + i^r[1,-i]^{\otimes 4}\}.$$

We will denote by $f' = [1,i]^{\otimes 4} + i^r[1,-i]^{\otimes 4}$. If g has arity 2, then up to a nonzero scalar, $g = [1,b,-1]$ with $b^4 \neq 0, 1$ by Lemma A.14, and we are done by Lemma E.7. In the following, assume that g has arity $2m \geq 4$. By Lemma A.14, either $g = [s,ti]^{\otimes 2m} \pm [t,si]^{\otimes 2m}$ with $s^4 \neq t^4$ and $st \neq 0$, or g has arity $2m$ and $g_k = (\pm i)^k(2m-2k)$.

- If g has arity $2m \geq 4$ and $g_k = (\pm i)^k(2m-2k)$ up to a nonzero scalar, then let $\hat{g} = (Z^{-1})^{\otimes 2m} g$, where $Z = \left[\begin{smallmatrix} 1 & 1 \\ i & -i \end{smallmatrix}\right]$. Then $\hat{g} = [0,1,0,\ldots,0]$ or $\hat{g} = [0,\ldots,0,1,0]$ of arity $2m$. By Corollary B.5, we have

  $$\text{Pl-}\#\text{CSP}^2(\hat{g}) \leq \text{Pl-}\#\text{CSP}^2(f', g).$$

  Let $\hat{g}' = \partial^{m-2}(\hat{g}) = [0,1,0,0,0]$ or $[0,0,0,1,0]$. Clearly $\hat{g}'$ is non-degenerate. It also has a second order recurrence of type $\langle 0,0,1\rangle$ or $\langle 1,0,0\rangle$. By Lemma A.11, $\hat{g}' \notin \mathscr{P} \cup \widetilde{\mathscr{A}} \cup \widetilde{\mathscr{M}}$. Then Pl-$\#\text{CSP}^2(\hat{g}')$ is #P-hard by Theorem D.5 and we are done.

- If $g = [s, ti]^{\otimes 2m} \pm [t, si]^{\otimes 2m}$, we have

$$g' = \partial^{m-2}(g) = (s^2 - t^2)^{m-2} \left\{ [s, ti]^{\otimes 4} \pm (-1)^{m-2} [t, si]^{\otimes 4} \right\}$$

and from f' we get $[1, 0, -1]^{\otimes 2}$ on LHS by Lemma A.20, thus

$$\text{Pl-Holant}([1, 0, -1]^{\otimes 2} \cup \mathcal{EQ}_2 \mid f', g') \leq \text{Pl-\#CSP}^2(f, g).$$

After a holographic transformation using $T = \left[\begin{smallmatrix} 1 & 0 \\ 0 & -i \end{smallmatrix}\right]$, we have

$$\begin{aligned} &\text{Pl-Holant}([1, 0, 1]^{\otimes 2}, [1, 0, -1], [1, 0, 0, 0, 1], \ldots \mid \hat{f}', \hat{g}') \\ &\qquad \equiv \text{Pl-Holant}([1, 0, -1]^{\otimes 2} \cup \mathcal{EQ}_2 \mid f', g'), \end{aligned}$$

where $\hat{f}' = (T^{-1})^{\otimes 4} f' = [1, 1]^{\otimes 4} + i^r [1, -1]^{\otimes 4}$ and $\hat{g}' = (T^{-1})^{\otimes 4} g'$. Note that $\hat{f}'$ satisfies a second order recurrence of type $\langle -1, 0, 1 \rangle$. Thus $\hat{f}' \notin \widehat{\mathscr{M}}$ by Lemma A.11. Also note that $\mathscr{P}$ and $\widetilde{\mathscr{A}}$ are invariant under T, and since $g' \in \widehat{\mathscr{M}^\dagger} \setminus (\mathscr{P} \cup \widetilde{\mathscr{A}})$, we have $\hat{g}' \in \widehat{\mathscr{M}} \setminus (\mathscr{P} \cup \widetilde{\mathscr{A}})$. In the following, we will construct $[1, 0, 1]^{\otimes 2}$ on RHS for

$$\text{Pl-Holant}([1, 0, 1]^{\otimes 2}, [1, 0, -1], [1, 0, 0, 0, 1], \ldots \mid \hat{f}', \hat{g}').$$

Since we have $[1, 0, 1]^{\otimes 2}$ on LHS, we can get $[\partial(\hat{f}')]^{\otimes 2} = 4[1 + i^r, 1 - i^r, 1 + i^r]^{\otimes 2}$ on RHS.

- If $r = 0$, then we directly have $[1, 0, 1]^{\otimes 2}$ on RHS.
- If $r = 2$, then we have $[0, 1, 0]^{\otimes 2}$ on RHS. Thus we can move $[1, 0, 1]^{\otimes 2}$ on LHS to RHS.
- If $r = 1$ or 3, then we have $[1, \pm i, 1]^{\otimes 2}$ on RHS. By connecting two copies of $[1, \pm i, 1]^{\otimes 2}$ by $[1, 0, 1]^{\otimes 2}$ of LHS, we have a nonzero multiple of $[0, 1, 0]^{\otimes 2}$ on RHS. Then we can move $[1, 0, 1]^{\otimes 2}$ on LHS to RHS.

From the above, we have

$$\begin{aligned} &\text{Pl-Holant}([1, 0, 1]^{\otimes 2}, [1, 0, -1], [1, 0, 0, 0, 1], \ldots \mid \hat{f}', \hat{g}', [1, 0, 1]^{\otimes 2}) \\ \leq\ &\text{Pl-Holant}([1, 0, 1]^{\otimes 2}, [1, 0, -1], [1, 0, 0, 0, 1], \ldots \mid \hat{f}', \hat{g}'). \end{aligned}$$

Note that we have all of $=_{4k}$ on the LHS. Thus by Lemma B.4,

$$\text{Pl-\#CSP}^2(\hat{f}', \hat{g}') \leq \text{Pl-Holant}([1, 0, 1]^{\otimes 2}, [1, 0, -1], [1, 0, 0, 0, 1], \ldots \mid \hat{f}', \hat{g}', [1, 0, 1]^{\otimes 2}).$$

Recall that $\hat{g}' \in \widehat{\mathscr{M}} \setminus (\mathscr{P} \cup \widetilde{\mathscr{A}})$ and $\hat{f}' \notin \widehat{\mathscr{M}}$. Thus we are done by Lemma E.3. □

F.3 Mixing with $\mathscr{A}^\dagger$

In this subsection, we prove the No-Mixing lemma for $\mathscr{A}^\dagger$ with other tractable sets, namely the statements No-Mixing-$(3, j)$, for $1 \leq j \leq 2$. Because we have already proved Lemma F.3 and Lemma F.6, the No-Mixing lemmas for $S_5 = \mathscr{P}$ and $S_4 = \mathscr{A}$ respectively, we only need to consider the mixing of $S_3 = \mathscr{A}^\dagger$ with S_j for $1 \leq j \leq 2$. Thus we may assume $f \in \mathscr{A}^\dagger$ and $g \in \widetilde{\mathscr{M}} \setminus \mathscr{A}^\dagger$. Moreover, we can assume that $f, g \notin \mathscr{P} \cup \mathscr{A}$.

Lemma F.7. *Let* $\{f, g\} \subseteq \left(\bigcup_{k=1}^{3} S_k\right) \setminus (S_4 \cup S_5)$ *and* $\{f, g\} \nsubseteq S_j$ *for* $1 \leq j \leq 3$. *Assume that* $f \in \mathscr{A}^\dagger$, *then* Pl-\#CSP$^2(f, g)$ *is* \#P*-hard.*

Proof. Firstly, we have $f \in \mathscr{A}^\dagger \setminus \mathscr{P}$, thus $f \in \{[1,\alpha,-\alpha^2],[1,\alpha]^{\otimes 2n} + i^r[1,-\alpha]^{\otimes 2n} \mid 2n \geq 4\}$ (see Figure 35). Clearly $[1,\alpha,-\alpha^2]$ is not $\lambda[1,0,1]$. If $f = [1,\alpha]^{\otimes 2n} + i^r[1,-\alpha]^{\otimes 2n}$, then we have $\partial^{n-1}(f) = (1+\alpha^2)^{n-1}\{[1,\alpha]^{\otimes 2} + i^r[1,-\alpha]^{\otimes 2}\} = (1+\alpha^2)^{n-1}[1+i^r,(1-i^r)\alpha,(1+i^r)\alpha^2]$ which is not $\lambda[1,0,1]$. Hence we can always obtain a nonzero binary signature that is not $\lambda[1,0,1]$ from f.

Note that $g \in \widetilde{\mathscr{M}} \setminus (\mathscr{P} \cup \widetilde{\mathscr{A}})$. If $g \in \widehat{\mathscr{M}} \setminus (\mathscr{P} \cup \widetilde{\mathscr{A}})$, we are done by Lemma E.3. For $g \in \widehat{\mathscr{M}}^\dagger \setminus (\mathscr{P} \cup \widetilde{\mathscr{A}})$, since we have a nonzero binary signature that is not $\lambda[1,0,1]$, we are done by Lemma E.7. □

F.4 Mixing with $\widehat{\mathscr{M}}$

In this subsection, we prove the No-Mixing lemma for $\widehat{\mathscr{M}}$ with other tractable sets. Because we have already proved Lemma F.3, Lemma F.6, and Lemma F.7, the No-Mixing lemmas for $S_5 = \mathscr{P}$, $S_4 = \mathscr{A}$, and $S_3 = \mathscr{A}^\dagger$ respectively, we only need to consider the No-Mixing of $S_2 = \widehat{\mathscr{M}}$ with $S_1 = \widehat{\mathscr{M}}^\dagger$.

Lemma F.8. *Let* $\{f,g\} \subseteq \left(\bigcup_{k=1}^{2} S_k\right) \setminus (S_3 \cup S_4 \cup S_5)$ *and* $\{f,g\} \nsubseteq S_j$ *for* $1 \leq j \leq 2$. *Then* Pl-#CSP$^2(f,g)$ *is* #P*-hard.*

Proof. Either f or $g \in \widehat{\mathscr{M}}$, otherwise $\{f,g\} \subseteq \widehat{\mathscr{M}}^\dagger$. As they do not belong to $S_3 \cup S_4 \cup S_5 = \mathscr{P} \cup \widetilde{\mathscr{A}}$, we have a signature in $\widehat{\mathscr{M}} \setminus (\mathscr{P} \cup \widetilde{\mathscr{A}})$. Thus we are done by Lemma E.3. □

By Lemma F.3, Lemma F.6, Lemma F.7 and Lemma F.8, we have the following No-Mixing theorem for two signatures with even arities.

Theorem F.9. *Let* f *and* g *be two symmetric signatures of even arity. If* $\{f,g\} \subseteq \bigcup_{k=1}^{5} S_k$ *and* $\{f,g\} \nsubseteq S_j$ *for* $1 \leq j \leq 5$, *then* Pl-#CSP$^2(f,g)$ *is* #P*-hard.*

G No-Mixing of Even Arity Signature Set

In this section, we extend Theorem F.9, the No-Mixing theorem for a pair of two sigatures of even arity, to Theorem G.4, the No-Mixing theorem for a set of signatures of even arity. For convenience, we explicitly list some signature sets that are be used in the proof of Theorem G.4.

Lemma G.1. *For nonzero even arity signatures, ignoring a nonzero factor, we have*

1. $\mathscr{A}^\dagger \cap (\mathscr{P} \cup \mathscr{A})$ *is the set*

$$\{[1,\alpha]^{\otimes 2n},[1,0]^{\otimes 2n},[0,1]^{\otimes 2n},[0,1,0],[1,0,\ldots,0,i^r] \mid n \geq 1,\ 0 \leq r \leq 3\}.$$

2. $\widehat{\mathscr{M}} \cap (\mathscr{P} \cup \widetilde{\mathscr{A}})$ *is the set*

$$\{[1,\pm 1]^{\otimes 2m},[0,1,0],[1,\pm i,1],[1,0,\ldots,0,\pm 1],[1,i]^{\otimes 2n} \pm [1,-i]^{\otimes 2n} \mid m \geq 1,\ n \geq 2\}.$$

3. $\widehat{\mathscr{M}}^\dagger \cap (\mathscr{P} \cup \widetilde{\mathscr{A}})$ *is the set*

$$\{[1,\pm i]^{\otimes 2m},[0,1,0],[1,\pm 1,-1],[1,0,\ldots,0,\pm 1],[1,1]^{\otimes 2n} \pm [1,-1]^{\otimes 2n} \mid m \geq 1,\ n \geq 2\}.$$

4. $\bigcap_{3 \leq k \leq 5} S_k$ *is the set* $\{[1,0]^{\otimes 2n},[0,1]^{\otimes 2n},[0,1,0],[1,0,\ldots,0,i^r] \mid n \geq 1,\ 0 \leq r \leq 3\}$.

5. $\bigcap_{1\le k\le 5} S_k = \bigcap_{2\le k\le 5} S_k$ *is the set* $\{[0,1,0],[1,0,\ldots,0,\pm 1]\}$.

Proof. For all five cases, it is easy to show that the listed signatures in the displayed set are indeed members of the respective stated intersection, bear in mind that the signatures all have even arity. E.g., the signature $f=[1,0,\ldots,0,i^r]$ is clearly in $\mathscr{P}$ (as well as $\mathscr{A}$), and it has even arity $2n$, and thus under the transformation $T=\left[\begin{smallmatrix}1&0\\0&\alpha\end{smallmatrix}\right]$, $(T^{-1})^{2n}f=[1,0,\ldots,0,i^s]\in\mathscr{A}$, for some $0\le s\le 3$. Thus $f\in\mathscr{A}^\dagger$.

In the following, we prove that if f has even arity and is in the stated intersection then it is among the listed types.

1. (a.) Firstly, suppose that $f\in\mathscr{A}^\dagger\cap(\mathscr{P}\cup\mathscr{A})$ is degenerate, i.e., $f=[a,b]^{2n}$. If $f=[1,0]^{2n}$ or $[0,1]^{2n}$ up to a nonzero scalar, then f is among the listed. Suppose $ab\neq 0$. Then up to a nonzero scalar, $f=[1,\omega]^{2n}$, for some $\omega\neq 0$. By $f\in\mathscr{A}^\dagger$, we have $\left[\begin{smallmatrix}1&0\\0&\alpha\end{smallmatrix}\right]^{2n}f=[1,\alpha\omega]^{2n}\in\mathscr{A}$. Thus $(\alpha\omega)^4=1$, i.e., $\omega^4=-1$. So f is among the listed types.

   (b.) If $f\in\mathscr{A}^\dagger\cap(\mathscr{P}\cup\mathscr{A})$ is a non-degenerate binary signature, by $f\in\mathscr{A}^\dagger$ and Lemma A.8, we have $f=[1,\alpha,-\alpha^2]$, or $[0,1,0]$, or $[1,0,\rho]$ up to a scalar, where $\alpha^4=-1,\rho^4=1$. Note that $[1,\alpha,-\alpha^2]\notin\mathscr{P}\cup\mathscr{A}$ by Corollary A.9. Thus $f=[0,1,0]$ or $[1,0,\rho]$; these are among the listed types.

   (c.) If $f\in\mathscr{A}^\dagger\cap(\mathscr{P}\cup\mathscr{A})$ is non-degenerate and and has arity $2n\ge 4$, by $f\in\mathscr{A}^\dagger$ and Lemma A.11, f has type $\langle 0,1,0\rangle$ or $\langle 1,0,\pm i\rangle$ and the second order recurrence relation is unique up to a scalar. If f has type $\langle 1,0,\pm i\rangle$, then $f\notin\mathscr{P}\cup\mathscr{A}$ by Lemma A.11. This contradicts that $f\in\mathscr{A}^\dagger\cap(\mathscr{P}\cup\mathscr{A})$. If f has type $\langle 0,1,0\rangle$, then $f=[1,0,\ldots,0,x]$ with $x\neq 0$ up to a nonzero scalar, because f is non-degenerate. Moreover, if $x^4\neq 1$, bear in mind that f has even arity, then $f\notin\mathscr{A}^\dagger$ and this contradicts that $f\in\mathscr{A}^\dagger\cap(\mathscr{P}\cup\mathscr{A})$. Hence $x^4=1$ and $f=[1,0,\ldots,0,i^r]$, for some $0\le r\le 3$; this is among the listed types.
   Summarizing, we proved that if $f\in\mathscr{A}^\dagger\cap(\mathscr{P}\cup\mathscr{A})$ then f is among the listed types.

2. (a.) Suppose $f\in\widehat{\mathscr{M}}\cap(\mathscr{P}\cup\widehat{\mathscr{A}})$ is a nonzero degenerate signature, i.e., $f=[a,b]^{\otimes 2n}$. By $f\in\widehat{\mathscr{M}}$ we have $\left[\begin{smallmatrix}1&1\\1&-1\end{smallmatrix}\right]^{\otimes 2n}f=[a+b,a-b]^{\otimes 2n}\in\mathscr{M}$, which must satisty the parity constraints. Thus $a=\pm b$ and $f=[1,\pm 1]^{\otimes 2n}$ up to a nonzero scalar.

   (b.) If $f\in\widehat{\mathscr{M}}\cap(\mathscr{P}\cup\widehat{\mathscr{A}})$ is a non-degenerate binary signature, by $f\in\widehat{\mathscr{M}}$ and Lemma A.8, we have $f=[0,1,0]$, or $[1,b,1]$, or $[1,0,-1]$ up to a nonzero scalar. If $f=[1,b,1]$ and $b^4\neq 0,1$, then $f\notin\mathscr{P}\cup\widetilde{\mathscr{A}}$, by Corollary A.9. This contradicts that $f\in\widehat{\mathscr{M}}\cap(\mathscr{P}\cup\widetilde{\mathscr{A}})$. Thus, $f=[0,1,0]$, $[1,0,1]$, $[1,i^r,1]$ or $[1,0,-1]$, where $0\le r\le 3$. Note that if $r=0$ or 2, then $[1,i^r,1]=[1,\pm 1,1]=[1,\pm 1]^{\otimes 2}$. Thus all these binary signatures are in the listed types.

   (c.) If $f\in\widehat{\mathscr{M}}\cap(\mathscr{P}\cup\widetilde{\mathscr{A}})$ is non-degenerate and has arity $2n\ge 4$, by $f\in\widehat{\mathscr{M}}$ and Lemma A.11, f has type $\langle 0,1,0\rangle$ or $\langle 1,c,1\rangle$, and the second order recurrence relation is unique up to a scalar. If f has type $\langle 1,c,1\rangle$ with $c\neq 0$, then $f\notin\mathscr{P}\cup\widetilde{\mathscr{A}}$ by Lemma A.11 and this contradicts that $f\in\widehat{\mathscr{M}}\cap(\mathscr{P}\cup\widetilde{\mathscr{A}})$. If f has type $\langle 1,0,1\rangle$, then there exist constants x and y such that $f=x[1,i]^{\otimes 2n}+y[1,-i]^{\otimes 2n}$. By non-degeneracy, we get $xy\neq 0$, and by its type $\langle 1,0,1\rangle$, $f\notin\mathscr{P}$ by Lemma A.11. Thus $f\in\widetilde{\mathscr{A}}$. In fact by Lemma A.11 and its type $\langle 1,0,1\rangle$, $f\notin\mathscr{A}^\dagger\setminus\mathscr{P}$, thus it follows that $f\in\mathscr{A}\setminus\mathscr{P}$. Then there are two possiblities: Either $f=\left[\begin{smallmatrix}1&1\\1&-1\end{smallmatrix}\right]^{\otimes 2n}\left\{[1,0]^{\otimes 2n}+i^r[0,1]^{\otimes 2n}\right\}$, or $f=\left[\begin{smallmatrix}1&1\\i&-i\end{smallmatrix}\right]^{\otimes 2n}\left\{[1,0]^{\otimes 2n}+i^r[0,1]^{\otimes 2n}\right\}$, up to a nonzero scalar, where $0\le r\le 3$. By $\left[\begin{smallmatrix}1&1\\1&-1\end{smallmatrix}\right]^{-1}Z=\frac{1}{2}Z\left[\begin{smallmatrix}0&1-i\\1+i&0\end{smallmatrix}\right]$ the first possiblity quickly reaches a contradiction. Thus $f=[1,i]^{\otimes 2n}+i^r[1,-i]^{\otimes 2n}$ up to a nonzero scalar, for some $0\le r\le 3$. If $f=[1,i]^{\otimes 2n}\pm i[1,-i]^{\otimes 2n}$, then $\left[\begin{smallmatrix}1&1\\1&-1\end{smallmatrix}\right]^{\otimes 2n}f$ is a nonzero multiple of the

form $[1,i]^{\otimes 2n} \pm i[1,-i]^{\otimes 2n}$, which does not satisfy parity, and hence not in $\mathscr{M}$. So f is not in $\widehat{\mathscr{M}}$. Hence $f = [1,i]^{\otimes 2n} \pm [1,-i]^{\otimes 2n}$, which is among the listed types.
If f has type $\langle 0,1,0\rangle$, then $f = [1,0,\ldots,0,x]$ with $x \neq 0$, up to a nonzero scalar. By $f \in \widehat{\mathscr{M}}$ and Lemma A.16, we have $x^2 = 1$. Thus $f = [1,0,\ldots,0,\pm 1]$, which is among the listed types.
Summarizing, we proved that if $f \in \widehat{\mathscr{M}} \cap (\mathscr{P} \cup \widetilde{\mathscr{A}})$, then f is among the listed types.

3. Note that $\mathscr{P} \cup \widetilde{\mathscr{A}}$ is unchanged under the transformation by $\left[\begin{smallmatrix}1 & 0\\ 0 & i\end{smallmatrix}\right]$. Thus
$$\widehat{\mathscr{M}^{\dagger}} \cap (\mathscr{P} \cup \widetilde{\mathscr{A}}) = \left[\begin{smallmatrix}1 & 0\\ 0 & i\end{smallmatrix}\right] \left\{ \widehat{\mathscr{M}} \cap (\mathscr{P} \cup \widetilde{\mathscr{A}}) \right\}.$$
Then the proof of this case follows from the previous case by a transformation using $\left[\begin{smallmatrix}1 & 0\\ 0 & i\end{smallmatrix}\right]$.
4. If $f \in \bigcap_{k=3}^{5} S_k$, then a fortiori, $f \in \mathscr{A}^{\dagger} \cap (\mathscr{P} \cup \mathscr{A})$. This implies that
$$f \in \{[1,\alpha]^{\otimes 2n}, [1,0]^{\otimes 2n}, [0,1]^{\otimes 2n}, [0,1,0], [1,0,\ldots,0,i^r] \mid n \geq 1,\ 0 \leq r \leq 3\}.$$
Note that $[1,\alpha]^{\otimes 2m} \notin \mathscr{A}$. Thus $f = [1,0]^{\otimes 2n}$, or $[0,1]^{\otimes 2n}$, or $[0,1,0]$, or $[1,0,\ldots,0,i^r]$. All of these four types are among the listed.
5. We already have
$$\{[0,1,0], [1,0,\ldots,0,\pm 1]\} \subseteq \bigcap_{1 \leq k \leq 5} S_k \subseteq \bigcap_{2 \leq k \leq 5} S_k.$$
If $f \in \bigcap_{k=2}^{5} S_k$, then $f \in \bigcap_{k=3}^{5} S_k$. This implies that
$$f \in \{[1,0]^{\otimes 2n}, [0,1]^{\otimes 2n}, [0,1,0], [1,0,\ldots,0,i^r] \mid n \geq 1,\ 0 \leq r \leq 3\}.$$
Moreover, if $f = [1,0]^{\otimes 2n}$, or $[0,1]^{\otimes 2n}$ or $[1,0,\ldots,0,\pm i]$, then $f \notin \widehat{\mathscr{M}}$, because $\left[\begin{smallmatrix}1 & 1\\ 1 & -1\end{smallmatrix}\right]^{\otimes 2n} f$ does not satisfy parity constraints. Hence $f = [0,1,0]$, or $[1,0,\ldots,0,\pm 1]$, and both types are among the listed.

□

We state the following simple lemma which allows us to replace a signature set $\mathcal{F}$ in the proof of the No-Mixing Theorem by a smaller set $\mathcal{F}'$ that subtracts from $\mathcal{F}$ those signatures that belong to all common tractable signature sets.

Lemma G.2. *Let $\mathcal{F}$ be a set of symmetric signatures such that for all $1 \leq k \leq 5$, $\mathcal{F} \not\subseteq S_k$. Let $\mathcal{F}' = \mathcal{F} \setminus (\bigcap_{k=1}^{5} S_k)$. Then for all $1 \leq k \leq 5$, $\mathcal{F}' \not\subseteq S_k$ and $\text{Pl-\#CSP}^2(\mathcal{F}') \leq \text{Pl-\#CSP}^2(\mathcal{F})$.*

Proof. Suppose for some $1 \leq k \leq 5$, $\mathcal{F}' \subseteq S_k$, then clearly $\mathcal{F} \subseteq S_k$. The reduction is trivial since $\mathcal{F}' \subseteq \mathcal{F}$. □

Suppose $\mathcal{F}$ is as given in Lemma G.2, and $\mathcal{F} \cap (\bigcup_{k=1}^{5} S_k) \neq \emptyset$. Let $j = \min\{k \mid \mathcal{F} \cap S_k \neq \emptyset, 1 \leq k \leq 5\}$. Then j is well defined. The same proof shows that $\mathcal{F}' = \mathcal{F} \setminus (\bigcap_{k=j}^{5} S_k)$ also has the property that $\mathcal{F}' \not\subseteq S_k$, for $j \leq k \leq 5$, and $\mathcal{F}' \cap S_k = \emptyset$ for $1 \leq k < j$.

Corollary G.3. *Let $\mathcal{F}$ be a set of symmetric signatures such that for all $1 \leq k \leq 5$, $\mathcal{F} \not\subseteq S_k$. Furthermore suppose $\mathcal{F} \cap (\bigcup_{k=1}^{5} S_k) \neq \emptyset$ and let $j = \min\{k \mid \mathcal{F} \cap S_k \neq \emptyset, 1 \leq k \leq 5\}$. Let $\mathcal{F}' = \mathcal{F} \setminus (\bigcap_{k=j}^{5} S_k)$. Then for all $1 \leq k \leq 5$, $\mathcal{F}' \not\subseteq S_k$ and $\text{Pl-\#CSP}^2(\mathcal{F}') \leq \text{Pl-\#CSP}^2(\mathcal{F})$.*

Recall that $S_1 = \widehat{\mathscr{M}}$, $S_2 = \widehat{\mathscr{M}^{\dagger}}$, $S_3 = \mathscr{A}^{\dagger}$, $S_4 = \mathscr{A}$ and $S_5 = \mathscr{P}$.

Theorem G.4. *Let $\mathcal{F} \subseteq \bigcup_{k=1}^{5} S_k$ be a set of symmetric signatures of even arities. If $\mathcal{F} \subseteq S_k$ for some $1 \le k \le 5$, then* Pl-#CSP$^2(\mathcal{F})$ *is tractable. Otherwise,* Pl-#CSP$^2(\mathcal{F})$ *is* #P*-hard.*

Proof. If $\mathcal{F} \subseteq S_k$ for some $1 \le k \le 5$, then tractability follows from the definition of $\mathscr{P}$-transformability, $\mathscr{A}$-transformability and $\mathscr{M}$-transformability.

Now suppose $\mathcal{F} \nsubseteq S_k$ for all $1 \le k \le 5$. We first replace $\mathcal{F}$ by $\mathcal{F}' = \mathcal{F} \setminus (\bigcap_{k=1}^{5} S_k)$. This also excludes the identically 0 signature. By Lemma G.2, we still have $\mathcal{F}' \nsubseteq S_k$ for $1 \le k \le 5$, and we only need to prove Pl-#CSP$^2(\mathcal{F}')$ is #P-hard.

We will treat the tractable sets in the order $S_1, S_2, \ldots, S_5$, starting with $S_1 = \widehat{\mathscr{M}}$.

1. Suppose that $\mathcal{F}' \cap S_1 \neq \emptyset$.
   Let $\mathcal{G}_1 = \mathcal{F}' \cap S_1$, and $\mathcal{H}_1 = \mathcal{F}' \setminus S_1$. Then $\mathcal{G}_1 \neq \emptyset$, and since $\mathcal{F}' \nsubseteq S_1$ we also have $\mathcal{H}_1 \neq \emptyset$. If there exists $g \in \mathcal{G}_1$ such that $g \in \widehat{\mathscr{M}} \setminus (\mathscr{P} \cup \widetilde{\mathscr{A}})$, then we are done by Lemma E.3. Otherwise, $\mathcal{G}_1 \subseteq \widehat{\mathscr{M}} \cap (\mathscr{P} \cup \widetilde{\mathscr{A}})$. Then by the forms given in Lemma G.1, ignoring nonzero scalars, $\mathcal{G}_1 \subseteq \{[1, \pm 1]^{\otimes 2m}, [1, \pm i, 1], [1, i]^{\otimes 2n} \pm [1, -i]^{\otimes 2n} \mid m \ge 1, n \ge 2\}$. Note that we have excluded $\bigcap_{k=1}^{5} S_k$ in $\mathcal{F}'$, hence also in $\mathcal{G}_1$. By Lemma G.1, $[1, 0, \ldots, 0, \pm 1], [0, 1, 0] \notin \mathcal{F}'$.
   If $[1, \pm 1]^{\otimes 2m} \in \mathcal{G}_1$ for some $m \ge 1$, then we can construct $\partial^{m-1}([1, \pm 1]^{\otimes 2m}) = 2^{m-1}[1, \pm 1]^{\otimes 2}$, and we are done by Lemma B.3.
   Otherwise, by the forms in

   $$\mathcal{G}_1 \subseteq \{[1, \pm i, 1], [1, i]^{\otimes 2n} \pm [1, -i]^{\otimes 2n} \mid n \ge 2\}, \tag{G.24}$$

   we have $\mathcal{G}_1 \subseteq \mathscr{A}$. If $\mathcal{H}_1 \subseteq \mathscr{A}$, then we would have $\mathcal{F}' \subseteq \mathscr{A}$, a contradiction. Thus $\mathcal{H}_1 \nsubseteq \mathscr{A}$. Thus there exists $h \in \mathcal{H}_1 \setminus \mathscr{A}$. By definition of $\mathcal{H}_1$, $h \notin \widehat{\mathscr{M}}$. Also $\mathcal{H}_1 \subseteq \bigcup_{k=1}^{5} S_k$, thus $h \in (\mathscr{P} \cup \mathscr{A}^{\dagger} \cup \widehat{\mathscr{M}}^{\dagger}) \setminus (\mathscr{A} \cup \widehat{\mathscr{M}})$. By the forms of signatures in the nonempty set $\mathcal{G}_1$ in (G.24) we have $\mathcal{G}_1 \cap (\mathscr{P} \cup \mathscr{A}^{\dagger} \cup \widehat{\mathscr{M}}^{\dagger}) = \emptyset$. To check this: for the binary $[1, \pm i, 1]$, we apply Lemma A.8; for $[1, i]^{\otimes 2n} \pm [1, -i]^{\otimes 2n}$ we use its second order recurrence of type $\langle 1, 0, 1 \rangle$ and then we apply Lemma A.11. Thus Pl-#CSP$^2(\mathcal{F}')$ is #P-hard by Theorem F.9.
2. We have $\mathcal{F}' \cap S_1 = \emptyset$. We replace $\mathcal{F}'$ by $\mathcal{F}'' = \mathcal{F}' \setminus (\bigcap_{k=2}^{5} S_k)$. By Corollary G.3, we still have $\mathcal{F}'' \nsubseteq S_k$ for $2 \le k \le 5$, $\mathcal{F}'' \cap S_1 = \emptyset$, and we only need to prove Pl-#CSP$^2(\mathcal{F}'')$ is #P-hard.
   Now suppose that $\mathcal{F}'' \cap S_2 \neq \emptyset$.
   By Lemma G.1, $[1, 0, \ldots, 0, \pm 1], [0, 1, 0] \notin \mathcal{F}''$.
   Let $\mathcal{G}_2 = \mathcal{F}'' \cap S_2$ and $\mathcal{H}_2 = \mathcal{F}'' \setminus S_2$. Both $\mathcal{G}_2, \mathcal{H}_2 \neq \emptyset$ and by definition $\mathcal{H}_2 \cap \widetilde{\mathscr{M}} = \emptyset$. Thus there exists $h \in \mathcal{H}_2 \setminus \widetilde{\mathscr{M}}$. If there exists $g \in \mathcal{G}_2$ such that $g \in \widehat{\mathscr{M}}^{\dagger} \setminus (\mathscr{P} \cup \widetilde{\mathscr{A}})$, then Pl-#CSP$^2(f, g)$ is #P-hard by Theorem F.9.
   Otherwise, $\mathcal{G}_2 \subseteq \widehat{\mathscr{M}}^{\dagger} \cap (\mathscr{P} \cup \widetilde{\mathscr{A}})$. Then $\mathcal{G}_2 \subseteq \{[1, \pm i]^{2m}, [1, \pm 1, -1], [1, 1]^{\otimes 2n} \pm [1, -1]^{\otimes 2n} \mid m \ge 1, n \ge 2\}$ by Lemma G.1. By its form $\mathcal{G}_2 \subseteq \mathscr{A}$. If $\mathcal{H}_2 \subseteq \mathscr{A}$, then we would have $\mathcal{F}'' \subseteq \mathscr{A}$, a contradiction. Thus $\mathcal{H}_2 \nsubseteq \mathscr{A}$. Hence there exists $h' \in \mathcal{H}_2 \setminus \mathscr{A}$. By definition of $\mathcal{H}_2$, $h' \notin \widetilde{\mathscr{M}}$. As $\mathcal{F}'' \subseteq \bigcup_{k=2}^{5} S_k$, $h' \in (\mathscr{P} \cup \mathscr{A}^{\dagger}) \setminus (\mathscr{A} \cup \widetilde{\mathscr{M}})$. If $\mathcal{G}_2$ includes either $[1, \pm 1, -1]$ or $[1, 1]^{\otimes 2n} \pm [1, -1]^{\otimes 2n}$ for some $n \ge 2$, both are not in $\mathscr{P} \cup \mathscr{A}^{\dagger}$. To see this, we apply Corollary A.9 to the binary $[1, \pm 1, -1]$. For $[1, 1]^{\otimes 2n} \pm [1, -1]^{\otimes 2n}$ with $n \ge 2$, we note its recurrence type $\langle -1, 0, 1 \rangle$ and then apply Lemma A.11. Then Pl-#CSP$^2(\mathcal{F}'')$ is #P-hard by Theorem F.9.
   We are left with the case where the nonempty set $\mathcal{G}_2 \subseteq \{[1, \pm i]^{2m} \mid m \ge 1\}$. By its form $\mathcal{G}_2 \subseteq \mathscr{P} \cap \mathscr{A} \cap \widehat{\mathscr{M}}^{\dagger}$ and $\mathcal{G}_2 \cap \mathscr{A}^{\dagger} = \emptyset$. If there exists $h'' \in \mathcal{H}_2 \setminus (\mathscr{A} \cup \mathscr{P})$, then by definition of $\mathcal{H}_2$ this $h'' \notin \widetilde{\mathscr{M}}$ as well, and we conclude that Pl-#CSP$^2(\mathcal{F}'')$ is #P-hard by Theorem F.9.

So we may assume $\mathcal{H}_2 \subseteq \mathscr{A} \cup \mathscr{P}$. If $\mathcal{H}_2 \subseteq \mathscr{A}$, then we would have $\mathcal{F}'' \subseteq \mathscr{A}$, a contradiction. Thus there exists $h''' \in (\mathcal{H}_2 \cap \mathscr{P}) \setminus \mathscr{A}$. Considering the forms of signatures in $\mathscr{P} \setminus \mathscr{A}$, it takes the form $h''' = [a, b]^{\otimes 2n}$ with $a^4 \neq b^4$, $ab \neq 0$, or $h''' = [1, 0, \ldots, 0, x]$ of arity $2n$, with $x^4 \neq 0, 1$, for some $n \geq 1$. Taking $h^{(4)} = \partial^{n-1}(h''')$, we get a nonzero multiple of either $[a, b]^{\otimes 2}$ or $[1, 0, x]$. Then taking $\partial^{m-1}_{h^{(4)}}([1, \pm i]^{\otimes 2m})$, for some $m \geq 1$, where $[1, \pm i]^{\otimes 2m} \in \mathcal{G}_2$ which is nonempty, we get a nonzero multiple of $[1, \pm i]^{\otimes 2}$, and we are done by Lemma B.3.

3. Now we have $\mathcal{F}'' \cap S_2 = \emptyset$.
We replace $\mathcal{F}''$ by $\mathcal{F}''' = \mathcal{F}'' \setminus (\bigcap_{k=3}^{5} S_k)$. By Lemma G.2, we still have $\mathcal{F}''' \nsubseteq S_k$ for $3 \leq k \leq 5$, $\mathcal{F}''' \cap (S_1 \cup S_2) = \emptyset$, and we only need to prove Pl-#CSP$^2(\mathcal{F}''')$ is #P-hard.
Suppose that $\mathcal{F}''' \cap S_3 \neq \emptyset$.
By Lemma G.1, the following signatures $[1, 0, \ldots, 0, i^r]$ of arity $2n$, $[0, 1, 0], [1, 0]^{\otimes 2n}, [0, 1]^{\otimes 2n}$ are all out of $\mathcal{F}'''$, for any $0 \leq r \leq 3$ and any $n \geq 1$.
Let $\mathcal{G}_3 = \mathcal{F}''' \cap S_3$, $\mathcal{H}_3 = \mathcal{F}''' \setminus S_3$. Both $\mathcal{G}_3, \mathcal{H}_3 \neq \emptyset$. Thus there exists $h \in \mathcal{H}_3$ such that $h \in (\mathscr{P} \cup \mathscr{A}) \setminus (\mathscr{A}^\dagger \cup \widetilde{\mathscr{M}})$. If there exists $g \in \mathcal{G}_3$ such that $g \in \mathscr{A}^\dagger \setminus \mathscr{P}$, then by Corollary A.12, $g \notin \mathscr{A}$. Thus Pl-#CSP$^2(g, h)$ is #P-hard by Theorem F.9.
Otherwise, we have $\mathcal{G}_3 \subseteq \mathscr{A}^\dagger \cap \mathscr{P}$. Thus we have $\mathcal{G}_3 \subseteq \{[1, \alpha]^{\otimes 2m} \mid m \geq 1\}$. Note that by Lemma G.1, we have excluded $[1, 0, \ldots, 0, i^r]$ of arity $2n$, $[0, 1, 0], [1, 0]^{\otimes 2n}, [0, 1]^{\otimes 2n}$ which are all in $\bigcap_{k=3}^{5} S_k$. (See Figure 35.)
Then we have $\partial^{m-1}([1, \alpha]^{\otimes 2m}) = (1 + \alpha^2)[1, \alpha]^{\otimes 2}$ and we are done by Lemma B.3.

4. Finally we have $\mathcal{F}''' \cap S_3 = \emptyset$.
We have $\mathcal{F}''' \nsubseteq S_k$ for $4 \leq k \leq 5$, $\mathcal{F}''' \cap (S_1 \cup S_2 \cup S_3) = \emptyset$, and thus $\mathcal{F}''' \subseteq S_4 \cup S_5$. Then we are done directly by Theorem F.9. □

H Dichotomy Theorem for an Even-Arity Signature

In this section, we prove the dichotomy theorem for Pl-#CSP$^2(f)$, where f has a general even arity $2n$. If $2n = 2$ or 4, then it has been proved in Theorem A.21′ and Theorem D.5 respectively. Thus we will assume $2n \geq 6$.

The next simple lemma is to determine if a symmetric signature satisfies a second order recurrence relation. In the following proof, we often argue that a signature f does not belong to $\mathscr{P} \cup \widetilde{\mathscr{A}} \cup \widetilde{\mathscr{M}}$ by Lemma A.11, and by showing that f does not satisfy any second order recurrence relation.

Lemma H.1. *For a symmetric signature $f = [f_0, f_1, \ldots, f_n]$, let $M_f = \begin{bmatrix} f_0 & f_1 & f_2 \\ f_1 & f_2 & f_3 \\ \vdots & \vdots & \vdots \\ f_{n-2} & f_{n-1} & f_n \end{bmatrix}$, then f satisfies a second order recurrence relation iff* $\operatorname{rank}(M_f) \leq 2$.

Proof. The signature f satisfies a second order recurrence relation $af_k + bf_{k+1} + cf_{k+2} = 0$ for $0 \leq k \leq n-2$ iff the linear system $M_f X = 0$ has a nonzero solution $(a, b, c)^T$ iff $\operatorname{rank}(M_f) \leq 2$. □

We often use the following argument to prove hardness: Firstly, we prove $f \notin \mathscr{P} \cup \widetilde{\mathscr{A}} \cup \widetilde{\mathscr{M}}$ using Lemma H.1. Moreover, if we can get $[1, \omega]^{\otimes 2}$ in Pl-#CSP$^2(f)$ for some $\omega \neq 0$, then Pl-#CSP$^2(f, [1, \omega]^{\otimes 2})$ is #P-hard by Lemma B.3. Or if we can get a signature $g \in \widehat{\mathscr{M}} \setminus (\mathscr{P} \cup \widetilde{\mathscr{A}})$ in Pl-#CSP$^2(f)$, then Pl-#CSP$^2(f, g)$ is #P-hard by Lemma E.3.

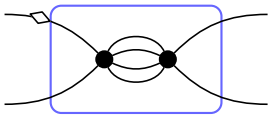


Figure 33: Gadget used to obtain a signature whose signature matrix is redundant. Both vertices are assigned f.

The next three lemmas are some special cases of Theorem H.5 which is the main result of this section. We prove these lemmas separately to facilitate the presentation of the proof of Theorem H.5.

Lemma H.2. *Suppose* $ab \neq 0$ *and* $f = [1, a, 0, -a, 0, a, b]$, *then* Pl-#CSP$^2(f)$ *is* #P*-hard.*

Proof. Note that $M_f = \begin{bmatrix} f_0 & f_1 & f_2 \\ f_1 & f_2 & f_3 \\ f_2 & f_3 & f_4 \\ f_3 & f_4 & f_5 \\ f_4 & f_5 & f_6 \end{bmatrix} = \begin{bmatrix} 1 & a & 0 \\ a & 0 & -a \\ 0 & -a & 0 \\ -a & 0 & a \\ 0 & a & b \end{bmatrix}$ has rank 3. Thus f does not satisfy any second order recurrence relation by Lemma H.1. So $f \notin \mathscr{P} \cup \widetilde{\mathscr{A}} \cup \widetilde{\mathscr{M}}$ by Lemma A.11.

Moreover, we have $\partial_{=_4}(f) = [1, 2a, b]$. If $[1, 2a, b]$ is degenerate, then $[1, 2a, b] = [1, 2a]^{\otimes 2}$. We are done since Pl-#CSP$^2(f, [1, 2a]^{\otimes 2})$ is #P-hard by Lemma B.3. Otherwise,

- For $b^4 \neq 1$, we have $[1, 2a, b] \notin \mathscr{P} \cup \widetilde{\mathscr{A}} \cup \widetilde{\mathscr{M}}$ by Corollary A.9 and Lemma A.8. Thus Pl-#CSP$^2([1, 2a, b])$ is #P-hard by Theorem A.21′ and we are done.
- For $b^2 = -1$, we have $\partial^2_{[1,2a,b]}(=_6) = [1, 0, -1]$ on the left and we have $f' = \partial_{[1,0,-1]}(f) = [1, 2a, 0, -2a, -b]$. Note that f' is redundant and the determinant of its compressed signature matrix is $4(b-1)a^2 \neq 0$. Thus Pl-#CSP$^2(f')$ is #P-hard by Lemma A.25 and we are done.
- For $b^2 = 1$, if $(2a)^4 \neq 1$, then we have $[1, 2a, b] \in \widetilde{\mathscr{M}} \setminus (\mathscr{P} \cup \widetilde{\mathscr{A}})$ by Lemma A.14. Thus Pl-#CSP$^2(f, [1, 2a, b])$ is #P-hard by Lemma E.3 and Lemma E.7 and we are done.
  Otherwise, we have $(2a)^4 = 1$. This implies that $(2a)^2 = \pm b$. Since $[1, 2a, b]$ is non-degenerate, we have $(2a)^2 \neq b$, thus $(2a)^2 = -b$. Moreover, we have $f'' = \partial_{[1,2a,b]}(f) = [1 + (2a)^2, (1-b)a, -(2a)^2, -(1-b)a, b^2 + (2a)^2]$. Note that $f'' = [0, 0, 1, 0, 0]$ for $b = 1$ and $f'' = [2, \pm 1, -1, \mp 1, 2]$ for $b = -1$. Both of $[0, 0, 1, 0, 0]$ and $[2, \pm 1, -1, \mp 1, 2]$ are redundant and their compressed signature matrices are nonsingular. Thus Pl-#CSP$^2(f'')$ is #P-hard by Lemma A.25 and we are done. □

The next lemma shows that if $\partial(f) = [1, 0]^{\otimes 2n-2} + t[0, 1]^{\otimes 2n-2}$ with $t \neq 0$, then either $f = [1, 0]^{\otimes 2n} + t[0, 1]^{\otimes 2n}$ or Pl-#CSP(f) is #P-hard. We will use this lemma in Theorem H.5 for the cases where $\partial(f)$ is a non-degenerate generalized equality Gen-Eq.

For $f = [a, b]^{\otimes 2n} = [f_0, f_1, \ldots, f_{2n}]$ we have $f_k = a^{n-k}b^k$. Then it is easy to see that $\bar{f} = [a^2, b^2]^{\otimes n} = [f_0, f_2, \ldots, f_{2n}]$, consisting of even indexed entries of f. This observation also extends to a sum of tensor powers by linearity. We will use this simple fact in the next lemma.

Lemma H.3. *Suppose that* $(x, y) \neq (0, 0)$ *and* $f = x[1, i]^{\otimes 2n} + y[1, -i]^{\otimes 2n} + [1, 0]^{\otimes 2n} + t[0, 1]^{\otimes 2n}$, *where* $2n \geq 6$ *and* $t \neq 0$, *then* Pl-#CSP(f) *is* #P*-hard.*

Proof. Let $a = x + y$, $b = (x - y)i$, then $(a, b) \neq (0, 0)$. Note that

$$f = [a, b, -a, -b, \ldots, \pm b, \mp a] + [1, 0, \ldots, 0, t] = [a + 1, b, -a, -b, a, \ldots, \pm b, \mp a + t].$$

Since M_f has a rank 3 submatrix $\begin{bmatrix} f_0 & f_1 & f_2 \\ f_1 & f_2 & f_3 \\ f_2 & f_3 & f_4 \\ f_{2n-2} & f_{2n-1} & f_{2n} \end{bmatrix} = \begin{bmatrix} a+1 & b & -a \\ b & -a & -b \\ -a & -b & a \\ \pm a & \pm b & \mp a+t \end{bmatrix}$, M_f has rank 3. By Lemma H.1, f does not satisfy any second order recurrence relation. So $f \notin \mathscr{P} \cup \widetilde{\mathscr{A}} \cup \widetilde{\mathscr{M}}$ by Lemma A.11.

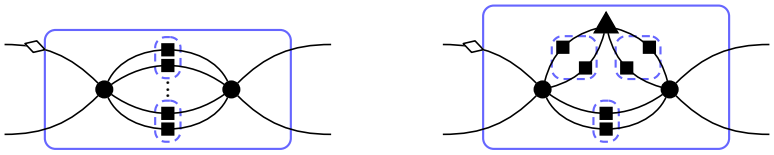
1. For $a \neq 0$, let $\bar{f} = [f_0, f_2, \ldots, f_{2n}]$, then $\bar{f} = a[1,-1]^{\otimes n} + [1,0]^{\otimes n} + t[0,1]^{\otimes n}$ and Pl-#CSP$(\bar{f}) \leq$ Pl-#CSP$^2(f)$ by Lemma A.19. Note that $\bar{f} = [a+1, -a, a, \ldots, \pm a, \mp a + t]$ has arity $n \geq 3$.
   - For $2n \geq 8$ or $[2n = 6$ and $t \neq -1]$, we claim that $\bar{f} \notin \mathscr{P} \cup \mathscr{A} \cup \widehat{\mathscr{M}}$.
     For $2n \geq 8$, Since $M_{\bar{f}}$ has a rank 3 submatrix $\begin{bmatrix} \bar{f}_0 & \bar{f}_1 & \bar{f}_2 \\ \bar{f}_1 & \bar{f}_2 & \bar{f}_3 \\ \bar{f}_{n-2} & \bar{f}_{n-1} & \bar{f}_n \end{bmatrix} = \begin{bmatrix} a+1 & -a & a \\ -a & a & -a \\ \mp a & \pm a & \mp a+t \end{bmatrix}$, $M_{\bar{f}}$ has rank 3. Thus $\bar{f}$ does not satisfy any second order recurrence relation by Lemma H.1. So $\bar{f} \notin \mathscr{P} \cup \mathscr{A} \cup \widehat{\mathscr{M}}$ by Lemma A.11.
     For $2n = 6$ and $t \neq -1$, $M_{\bar{f}}$ is a 2×3 matrix and has rank less than 3. So it always satisfies a second order recurrence relation. But we still show that $\bar{f} \notin \mathscr{P} \cup \mathscr{A} \cup \widehat{\mathscr{M}}$. Note that $\bar{f} = [a+1, -a, a, -a+t]$ when $n = 3$.
     - $\bar{f}$ is non-degenerate by $(\bar{f}_1)^2 \neq \bar{f}_0\bar{f}_2$ and $\bar{f}$ is not Gen-Eq since $\bar{f}_1 \neq 0$, so $\bar{f} \notin \mathscr{P}$.
     - If $\bar{f} \in \mathscr{A} \setminus \mathscr{P}$, then $\bar{f}$ has type $\langle 1, 0, \pm 1\rangle$ by Lemma A.11. By $\bar{f}_0 - \bar{f}_2 \neq 0$, $\bar{f}$ does not have type $\langle 1, 0, -1\rangle$. If $\bar{f}$ has type $\langle 1, 0, 1\rangle$, then $\bar{f}_0 + \bar{f}_2 = 0$, $\bar{f}_1 + \bar{f}_3 = 0$. This implies $t = -1$. It is a contradiction. Thus $\bar{f} \notin \mathscr{A} \setminus \mathscr{P}$.
     - By $\bar{f}_1 = -\bar{f}_2 \neq 0$ and Lemma A.16, if $\bar{f} \in \widehat{\mathscr{M}}$, then $\bar{f}_0 = -\bar{f}_3$. This contradicts that $t \neq -1$. Thus $\bar{f} \notin \widehat{\mathscr{M}}$.

     To summarize, $\bar{f} \notin \mathscr{P} \cup \mathscr{A} \cup \widehat{\mathscr{M}}$ for $2n \geq 8$, or $[2n = 6$ and $t \neq -1]$. Thus Pl-#CSP$(\bar{f})$ is #P-hard by Theorem A.22. So Pl-#CSP$^2(f)$ is #P-hard.
   - For $2n = 6$ and $t = -1$, we have $f = [a+1, b, -a, -b, a, b, -a-1]$. Firstly, we have $\partial^2(f) = [1, 0, -1]$ and $f' = \partial_{[1,0,-1]}(f) = [1+2a, 2b, -2a, -2b, 1+2a]$. The compressed signature matrix of f' is $\begin{bmatrix} 1+2a & 2b & -2a \\ 2b & -2a & -2b \\ -2a & -2b & 1+2a \end{bmatrix}$ and its determinant is $-2(4a^2 + 4b^2 + a)$. If $4a^2 + 4b^2 + a \neq 0$, then it is nonsingular, and we are done by Lemma A.25.
     Otherwise we have $4a^2 + 4b^2 + a = 0$. Consider the gadget in Figure 33. We assign f to both vertices. The signature of this gadget is redundant, and its compressed signature matrix is

$$\begin{bmatrix} 1+2a+8a^2+8b^2 & b & -2a-8a^2-8b^2 \\ b & 8a^2+8b^2 & -b \\ -2a-8a^2-8b^2 & -b & 1+2a+8a^2+8b^2 \end{bmatrix} = \begin{bmatrix} 1 & b & 0 \\ b & -2a & -b \\ 0 & -b & 1 \end{bmatrix}.$$

     If $a + b^2 \neq 0$, then this matrix is nonsingular, so we are done by Lemma A.25.
     Otherwise we have $4a^2 + 4b^2 + a = 0$ and $a + b^2 = 0$. Also we have $a \neq 0$. By solving these two equations, $a = \frac{3}{4}$ and $b = \pm\frac{\sqrt{3}}{2}i$. Moreover, we have $\partial_{=_4}(f) = [1+2a, 2b, -1-2a] = [\frac{5}{2}, \pm\sqrt{3}i, -\frac{5}{2}]$. By Lemma A.14, $\partial_{=_4}(f) \in \widehat{\mathscr{M}^\dagger} \setminus (\mathscr{P} \cup \mathscr{A})$. Recall that $f \notin \widehat{\mathscr{M}^\dagger}$. Thus Pl-#CSP$^2(f, [1+2a, 2b, -1-2a])$ is #P-hard by Lemma E.7 and we are done.
2. For $a = 0$, then $b \neq 0$ by $(a, b) \neq (0, 0)$.
   - if $2n \equiv 0 \pmod 4$ and $t \neq -1$, then

$$f'' = \partial_{=_4}^{\frac{n-2}{2}}(f) = 2^{\frac{n-2}{2}} x[1, i]^{\otimes 4} + 2^{\frac{n-2}{2}} y[1, -i]^{\otimes 4} + [1, 0]^{\otimes 4} + t[0, 1]^{\otimes 4},$$



(a) (b)

Figure 34: Two gadgets used to obtain a signature whose signature matrix is redundant. The dashed subgadgets are assigned $[1,0,1]^{\otimes 2}$ rotated so that it is equivalent to assigning $[1,0,1]$ to the square vertices.

i.e., $f'' = [1, 2^{\frac{n-2}{2}}b, 0, -2^{\frac{n-2}{2}}b, t]$. Note that f'' is redundant and the determinant of its compressed signature matrix is $-2^{n-2}b^2(t+1)$. By $t \neq -1$ and $b \neq 0$, the compressed signature matrix is nonsingular. So Pl-#CSP$^2(f'')$ is #P-hard by Lemma A.25. Thus Pl-#CSP$^2(f)$ is #P-hard.

- if $2n \equiv 0 \pmod 4$ and $t = -1$, we have $\partial^{n-1}(f) = [1, 0, -1]$ and

$$f''' = \partial^{n-3}_{[1,0,-1]}(f) = 2^{n-3}x[1,i]^{\otimes 6} + 2^{n-3}y[1,-i]^{\otimes 6} + [1,0]^{\otimes 6} + (-1)^{n-2}[0,1]^{\otimes 6},$$

i.e., $f''' = [1, 2^{n-3}b, 0, -2^{n-3}b, 0, 2^{n-3}b, (-1)^{n-2}]$. By Lemma H.2, Pl-#CSP$^2(f''')$ is #P-hard and we are done.

- if $2n \equiv 2 \pmod 4$, we have

$$f^{(4)} = \partial^{\frac{n-3}{2}}_{=_4}(f) = 2^{\frac{n-3}{2}}x[1,i]^{\otimes 6} + 2^{\frac{n-3}{2}}y[1,-i]^{\otimes 6} + [1,0]^{\otimes 6} + t[0,1]^{\otimes 6}.$$

Note that $f^{(4)} = [1, 2^{\frac{n-3}{2}}b, 0, -2^{\frac{n-3}{2}}b, 0, 2^{\frac{n-3}{2}}b, t]$. By Lemma H.2, Pl-#CSP$^2(f^{(4)})$ is #P-hard and we are done. □

We will use the next lemma in the proof of Theorem H.5 for the case that $\partial(f) = [1,i]^{\otimes 2n-2} + i^r[1,-i]^{\otimes 2n-2}$. In this case, we will transform Pl-#CSP^2 to Pl-#CSP^4 by holographic transformation and gadget construction. This is why we have to deal with Pl-#CSP^4 problems in the next lemma.

Lemma H.4. *Suppose $f = [0,1,0,\ldots,0,a,0]$ has arity $2n \geq 6$. If $a^4 = 1$, then the problem* Pl-#CSP$^4(f, [1,0,1]^{\otimes 2}, [1,0,1,0,1])$ *is* #P*-hard.*

Proof. In Pl-#CSP$^4(f, [1,0,1]^{\otimes 2}, [1,0,1,0,1])$, we do not have $=_2$ on the left, so we cannot connect the two edges on the right freely. But we do have $[1,0,1]^{\otimes 2}$ on the right and $=_4$ on the left, so we can do a loop to a pair of $=_4$ on the left respectively and we get $[1,0,1]^{\otimes 2}$ on the left.

Suppose $a^2 = 1$. Consider the gadget in Figure 34a. We assign f to the circle vertices and $[1,0,1]^{\otimes 2}$ to the dashed subgadgets rotated so that it is equivalent to assigning $[1,0,1]$ to the square vertices, where there are $2n-2$ parallel edges connecting the 2 copies of f with $2n-2$ square vertices. The signature of this gadget is redundant, and its compressed signature matrix is $\begin{bmatrix} 2n-2 & 0 & 0 \\ 0 & 1+a^2 & 0 \\ 0 & 0 & (2n-2)a^2 \end{bmatrix}$, which is nonsingular, by $a^2 = 1$. Thus we have

$$\text{Pl-\#CSP}^2(f') \leq_T \text{Pl-\#CSP}^4(f', [1,0,1]^{\otimes 2}) \leq_T \text{Pl-\#CSP}^4(f, [1,0,1]^{\otimes 2}, [1,0,1,0,1]),$$

where the first $\leq_T$ is by Lemma B.4. Then we are done by Lemma A.25.

For $a^2 = -1$, the gadget in Figure 34a cannot work since the compressed signature matrix of its resulting signature is $\begin{bmatrix} 2n-2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -2n+2 \end{bmatrix}$ which is singular.

We consider two cases.

- Suppose $2n \equiv 0 \pmod 4$. Then by Lemma B.4, we have

$$\text{Pl-\#CSP}^2(f, [1,0,1,0,1]) \le_T \text{Pl-\#CSP}^4(f, [1,0,1,0,1], [1,0,1]^{\otimes 2}). \tag{H.25}$$

  In Pl-#CSP$^2(f, [1,0,1,0,1])$, we have $f' = \partial^{n-2}(f) = [0,1,0,\pm i,0]$. Note that $f' \in \mathscr{A}^\dagger$ by considering $\left[\begin{smallmatrix} 1 & 0 \\ 0 & \alpha \end{smallmatrix}\right]^{\otimes 4} f'$, and also $f' \notin \mathscr{P} \cup \mathscr{A} \cup \mathscr{M}$ by considering its type $\langle 1, 0, \pm i\rangle$, and by Lemma A.11. Furthermore we have $[1,0,1,0,1] \in \mathscr{A}$, and also $[1,0,1,0,1] \notin \mathscr{A}^\dagger$ by its type $\langle 1, 0, -1\rangle$, and by Lemma A.11. Thus Pl-#CSP$^2(f', [1,0,1,0,1])$ is #P-hard by Theorem G.4 and we are done.
- For $2n \equiv 2 \pmod 4$, we cannot use Lemma B.4 to get the reduction in (H.25) since Lemma B.4 requires that all signatures on the right have arity $\equiv 0 \pmod 4$. But we have $f' = \partial_{=_4}^{\frac{n-3}{2}}(f) = [0,1,0,0,0,\pm i,0]$ as well as $\partial_{=_4}(f') = (1 \pm i)[0,1,0]$. We may use $[1,0,1]^{\otimes 2}$ of the LHS to transport this $[0,1,0]$ from the RHS to the LHS as follows: Let $f(x_1, y_1, x_2, y_2)$ be the function $[1,0,1]^{\otimes 2}$ which is 1 iff $x_1 = y_1$ and $x_2 = y_2$, and 0 otherwise. Then we connect x_1 and x_2 with the two edges of $[0,1,0]$ from the RHS. This creates $[0,1,0]$ on the LHS, with which we can take derivative of f' from the RHS. Then we have $\partial_{[0,1,0]}(f') = [1,0,0,0,\pm i]$. Consider the gadget in Figure 34b. We assign f' to the circle vertices, $[1,0,0,0,\pm i]$ to the triangle vertex, and $[1,0,1]^{\otimes 2}$ to the dashed subgadgets rotated so that it is equivalent to assigning $[1,0,1]$ to the square vertices. The signature f'' of this gadget is redundant, and its compressed signature matrix is $\begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 \mp i & 0 \\ 0 & 0 & \mp 2i \end{bmatrix}$, which is nonsingular. Thus Pl-#CSP$^2(f'')$ is #P-hard by Lemma A.25. Moreover, we have

$$\text{Pl-\#CSP}^4(f'', [1,0,1]^{\otimes 2}) \le_T \text{Pl-\#CSP}^4(f, [1,0,1,0,1], [1,0,1]^{\otimes 2})$$

  and

$$\text{Pl-\#CSP}^2(f'') \le_T \text{Pl-\#CSP}^4(f'', [1,0,1]^{\otimes 2})$$

  by Lemma B.4 and we are done. Now Lemma B.4 can work since f'' has arity 4. □

Now we are ready to prove the main theorem of this section, the dichotomy of Pl-#CSP$^2(f)$, where f has a general even arity $2n$. We will prove the theorem by induction on the arity $2n$. The base cases $2n = 2$ and $2n = 4$ are already done in Theorem A.21′ and Theorem D.5, respectively. We always have $f' = \partial(f)$ in Pl-#CSP$^2(f)$ which has arity $2n-2$. If $f' \notin \mathscr{P} \cup \widetilde{\mathscr{A}} \cup \widetilde{\mathscr{M}}$, then Pl-#CSP$^2(f')$ is #P-hard by induction and Pl-#CSP$^2(f)$ is #P-hard. Otherwise, for $f' \in \mathscr{P} \cup \widetilde{\mathscr{A}} \cup \widetilde{\mathscr{M}}$, we can explicitly express f by the integral operator $\int(f')$. We will prove the theorem in the following order:
(1) $f' \in \mathscr{P}$, (2) $f' \in \mathscr{A}^\dagger \setminus \mathscr{P}$, (3) $f' \in \mathscr{A} \setminus \mathscr{P}$, (4) $f' \in \widehat{\mathscr{M}} \setminus (\mathscr{P} \cup \widetilde{\mathscr{A}})$, and (5) $f' \in \widehat{\mathscr{M}^\dagger} \setminus (\mathscr{P} \cup \widetilde{\mathscr{A}})$.

Note that by Corollary A.13, Case (4) is equivalent to $f' \in \widehat{\mathscr{M}} \setminus (\mathscr{P} \cup \widetilde{\mathscr{A}} \cup \widehat{\mathscr{M}^\dagger})$, and Case (5) is equivalent to $f' \in \widehat{\mathscr{M}^\dagger} \setminus (\mathscr{P} \cup \widetilde{\mathscr{A}} \cup \widehat{\mathscr{M}})$.

In the proof, to use Theorem D.5, we often construct arity 4 signatures by our Calculus with binary signatures or $=_4$.

Theorem H.5. *Let f be a symmetric signature of even arity $2n$. If $f \in \mathscr{P} \cup \widetilde{\mathscr{A}} \cup \widetilde{\mathscr{M}}$, then* Pl-#CSP$^2(f)$ *is tractable. Otherwise,* Pl-#CSP$^2(f)$ *is* #P*-hard.*

Proof. If $f \in \mathscr{P} \cup \widetilde{\mathscr{A}} \cup \widetilde{\mathscr{M}}$, then tractability follows from the definition of $\mathscr{P}$-transformability, $\mathscr{A}$-transformability, and $\mathscr{M}$-transformability. Now suppose $f \notin \mathscr{P} \cup \widetilde{\mathscr{A}} \cup \widetilde{\mathscr{M}}$. If $2n \in \{2, 4\}$, then we are done by Theorem A.21′ and Theorem D.5 respectively.

For $2n \geq 6$, we will prove the theorem by induction on arity $2n$. If $f' = \partial(f) \notin \mathscr{P} \cup \widetilde{\mathscr{A}} \cup \widetilde{\mathscr{M}}$, then Pl-#CSP$^2(f')$ is #P-hard by induction. Thus Pl-#CSP$^2(f)$ is #P-hard. Otherwise, $f' \in \mathscr{P} \cup \widetilde{\mathscr{A}} \cup \widetilde{\mathscr{M}}$.

1. For $f' \in \mathscr{P}$, we have $f' \equiv 0$ or $f' = [a, b]^{\otimes 2n-2}$ (where $(a, b) \neq (0, 0)$) or $f' = [1, 0]^{\otimes 2n-2} + t[0, 1]^{\otimes 2n-2}$ with $t \neq 0$ by definition. Note that $2n - 2 \geq 4$.
   (a) $f' \equiv 0$. Then $f = x[1, i]^{\otimes 2n} + y[1, -i]^{\otimes 2n}$ by Proposition A.7 (the Explicit List for $\int(f')$). If $x = 0$ or $y = 0$, then $f \in \mathscr{P}$. If $xy \neq 0$ and $x^4 = y^4$, then $f \in \mathscr{A}$. In the following, assume that $xy \neq 0$ and $x^4 \neq y^4$.
   - For $2n \equiv 0 \pmod 4$, we have $f'' = \partial_{=_4}^{\frac{n-2}{2}}(f) = 2^{\frac{n-2}{2}} x[1, i]^{\otimes 4} + 2^{\frac{n-2}{2}} y[1, -i]^{\otimes 4}$. By $xy \neq 0$, f'' is non-degenerate, and has the unique recurrence type $\langle 1, 0, 1\rangle$. Therefore $f'' \notin \mathscr{P} \cup \mathscr{A}^\dagger \cup \widetilde{\mathscr{M}}$ by Lemma A.11. By $x^4 \neq y^4$ it is also not in $\mathscr{A}$. Thus $f'' \notin \mathscr{P} \cup \widetilde{\mathscr{A}} \cup \widetilde{\mathscr{M}}$. Therefore Pl-#CSP$^2(f'')$ is #P-hard by Theorem D.5. So Pl-#CSP$^2(f)$ is #P-hard.
   - For $2n \equiv 2 \pmod 4$, we cannot reduce the arity of f to 4 by $=_4$ directly as in the previous case. We will construct a binary signature that is not $\lambda[1, 0, 1]$ to reduce the arity of f. Firstly, we have $f''' = \partial_{=_4}^{\frac{n-1}{2}}(f) = 2^{\frac{n-1}{2}} x[1, i]^{\otimes 2} + 2^{\frac{n-1}{2}} y[1, -i]^{\otimes 2} = 2^{\frac{n-1}{2}}[a, b, -a]$, where $a = x + y, b = (x - y)i$. We remark that $[a, b, -a]$ can reduce the arity of f, but it involves a case analysis of a and b. Instead we use $[a, b, -a]$ to construct a simpler binary signature.
     Note that $a \neq 0$ by $x^4 \neq y^4$. Then we have $\partial_{[a,b,-a]}(=_4) = a[1, 0, -1]$ on the left. Thus we have $f^{(4)} = \partial_{[1,0,-1]}^{n-2}(f) = 2^{n-2} x[1, i]^{\otimes 4} + 2^{n-2} y[1, -i]^{\otimes 4}$. With the same reason as in the previous case, $f^{(4)} \notin \mathscr{P} \cup \widetilde{\mathscr{A}} \cup \widetilde{\mathscr{M}}$ by its type, and by $xy \neq 0, x^4 \neq y^4$. Thus Pl-#CSP$^2(f^{(4)})$ is #P-hard by Theorem D.5. So Pl-#CSP$^2(f)$ is #P-hard.
   (b) $f' = [a, b]^{\otimes 2n-2}$ with $ab \neq 0$. If $a^2 + b^2 \neq 0$, we have $\partial^{n-2}(f) = (a^2 + b^2)^{n-2}[a, b]^{\otimes 2}$ and we are done by Lemma B.3.
   Suppose $a^2 + b^2 = 0$, i.e., $f' = [1, \pm i]^{\otimes 2n-2}$ up to a scalar.
   - For $2n \equiv 0 \pmod 4$, we have $\partial_{=_4}^{\frac{n-2}{2}}(f') = 2^{\frac{n-2}{2}}[1, \pm i]^{\otimes 2}$ and are done by Lemma B.3.
   - For $2n \equiv 2 \pmod 4$, we cannot get $[1, \pm i]^{\otimes 2}$ in Pl-#CSP$^2(f')$ by Remark 6 (note that the arity of f' is $2n - 2 \equiv 0 \pmod 4$). To get $[1, \pm i]^{\otimes 2}$, we need the help of f. By Proposition A.7 (the Explicit List for $\int(f')$), $f = x[1, i]^{\otimes 2n} + y[1, -i]^{\otimes 2n} + g$, where g has arity $2n$ and $g_k = \frac{1}{4}(\pm i)^k(2n - 2k)$. If $x = y = 0$, then $f \in \widehat{\mathscr{M}^\dagger}$. Otherwise, let $a = x + y$, $b = (x - y)i$, then $(a, b) \neq (0, 0)$. We have $\partial_{=_4}^{\frac{n-1}{2}}(f) = 2^{\frac{n-1}{2}} x[1, i]^{\otimes 2} + 2^{\frac{n-1}{2}} y[1, -i]^{\otimes 2} + 2^{\frac{n-3}{2}}[1, 0, 1]$, i.e., $\partial_{=_4}^{\frac{n-1}{2}}(f) = 2^{\frac{n-3}{2}}[2a+1, 2b, -2a+1]$. If $a \neq 0$, then we have $\partial_{[2a+1,2b,-2a+1]}(=_4) = [2a + 1, 0, -2a + 1]$ on the left and $\partial_{[2a+1,0,-2a+1]}^{n-2}(f') = (4a)^{n-2}[1, \pm i]^{\otimes 2}$. Then we are done by Lemma B.3.
     If $a = 0$, then $b \neq 0$ and we have $[1, 2b, 1]$ and $\partial_{[1,2b,1]}^{n-2}(f') = (\pm 4bi)^{n-2}[1, \pm i]^{\otimes 2}$. Then we are done by Lemma B.3 again.
   (c) $f' = [1, 0]^{\otimes 2n-2}$. Then $f = x[1, i]^{\otimes 2n} + y[1, -i]^{\otimes 2n} + [1, 0]^{\otimes 2n}$ by Proposition A.7 (the Explicit List for $\int(f')$). If $x = y = 0$, then $f \in \mathscr{P}$. In the following, assume that

$(x,y) \neq (0,0)$. Let $a = x + y$, $b = (x - y)i$, then $(a,b) \neq (0,0)$.
We have $\partial^{n-1}(f) = [1,0]^{\otimes 2}$ and $f'' = \partial^{n-2}_{[1,0]^{\otimes 2}}(f) = x[1,i]^{\otimes 4} + y[1,-i]^{\otimes 4} + [1,0]^{\otimes 4}$, i.e., $f'' = [1+a, b, -a, -b, a]$. Note that f'' is redundant. If $a^2 + b^2 \neq 0$, then the compressed signature matrix of f'' is nonsingular and we are done by Lemma A.25.
Otherwise, we have $a = \pm ib$. We claim that $f'' \notin \mathscr{P} \cup \widetilde{\mathscr{A}} \cup \widetilde{\mathscr{M}}$. Note that $ab \neq 0$ by $(a,b) \neq (0,0)$ and $a = \pm ib$. If f'' is degenerate, then by $(f''_1)^2 = f''_0 f''_2$, we have $-a - a^2 = b^2$. This implies that $a = 0$. It is a contradiction. Moreover, note that $f'' = [1+a, \mp ia, -a, \pm ia, a]$ and has type $\langle 0, 1, \pm i\rangle$. Since f'' is non-degenerate and has arity ≥ 3, the second order recurrence relation $\langle 0, 1, \pm i\rangle$ is unique up to a scalar. Thus $f'' \notin \mathscr{P} \cup \widetilde{\mathscr{A}} \cup \widetilde{\mathscr{M}}$ by Lemma A.11. So Pl-#CSP$^2(f'')$ is #P-hard by Theorem D.5 and we are done.

(d) $f' = [0,1]^{\otimes 2n-2}$. The proof follows from the previous case by a holographic transformation using $\left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right]$.

(e) $f' = [1,0]^{\otimes 2n-2} + t[0,1]^{\otimes 2n-2}$ with $t \neq 0$. Then $f = x[1,i]^{\otimes 2n} + y[1,-i]^{\otimes 2n} + [1,0]^{\otimes 2n} + t[0,1]^{\otimes 2n}$ by Proposition A.7 (the Explicit List for $\int(f')$). If $x = y = 0$, then $f \in \mathscr{P}$. Otherwise, we have $(x,y) \neq (0,0)$ and we are done by Lemma H.3.

2. For $f' \in \mathscr{A}^\dagger \setminus \mathscr{P}$, we have $f' = [1,\alpha]^{\otimes 2n-2} + i^r[1,-\alpha]^{\otimes 2n-2}$ by definition (See Figure 35). Then $f = x[1,i]^{\otimes 2n} + y[1,-i]^{\otimes 2n} + \frac{1}{1+\alpha^2}\{[1,\alpha]^{\otimes 2n} + i^r[1,-\alpha]^{\otimes 2n}\}$ by Proposition A.7 (the Explicit List for $\int(f')$). If $x = y = 0$, then $f \in \mathscr{A}^\dagger$. In the following, assume that $(x,y) \neq (0,0)$.
Note that f' has type $\langle 1, 0, \pm i\rangle$ up to a scalar. And this second order recurrence relation is unique up to a scalar. Thus $f' \in \mathscr{A}^\dagger \setminus (\mathscr{P} \cup \mathscr{A} \cup \widetilde{\mathscr{M}})$ by Lemma A.11. In the following, we complete the proof by constructing a signature of even arity in $(\mathscr{P} \cup \mathscr{A} \cup \widetilde{\mathscr{M}}) \setminus \mathscr{A}^\dagger$ and apply Theorem G.4, or constructing an arity 4 signature that is not in $\mathscr{P} \cup \widetilde{\mathscr{A}} \cup \widetilde{\mathscr{M}}$ and apply Theorem D.5.
Firstly, we have $f'' = \partial^{n-3}(f') = (1+\alpha^2)^{n-3}\{[1,\alpha]^{\otimes 4} + i^r[1,-\alpha]^{\otimes 4}\}$. We will discard the nonzero factor that are powers of $1+\alpha^2$. If $r \neq 2$, we have $\partial(f'') = (1+i^r)[1, \frac{1-i^r}{1+i^r}\alpha, \alpha^2]$ and we have $\partial_{[1,\frac{1-i^r}{1+i^r}\alpha,\alpha^2]}(=_4) = [1,0,\alpha^2]$ on the left. For $r = 2$, $\partial(f'')$ is a nonzero multiple of $[0,1,0]$ and we have $\partial_{[0,1,0]}(f'') = 2\alpha[1,0,\alpha^2]$ on the right. Either way, we can take the derivative (for $[1,0,\alpha^2]$ in RHS we connect it via $(=_2)$ of LHS to f)

$$f''' = \partial^{n-2}_{[1,0,\alpha^2]}(f) = (1-\alpha^2)^{n-2}\{x[1,i]^{\otimes 4} + y[1,-i]^{\otimes 4}\}.$$

Note that $\partial_{[1,0,\alpha^2]}([1,\pm\alpha]^{2n})$ is the identically zero signature, since $\alpha^4 = -1$.
If $xy = 0$, or $[xy \neq 0$ and $x^4 = y^4]$, then $f''' \in \mathscr{A} \setminus \mathscr{A}^\dagger$. So Pl-#CSP$^2(f', f''')$ is #P-hard by Theorem G.4. Thus Pl-#CSP$^2(f)$ is #P-hard.
Otherwise, $xy \neq 0$ and $x^4 \neq y^4$, so $f''' \notin \mathscr{P} \cup \widetilde{\mathscr{A}} \cup \widetilde{\mathscr{M}}$ (by the same reason as before: first by its type $\langle 1, 0, 1\rangle$ it could only possibly be in $\mathscr{A}$ among the five classes by Lemma A.11; but $x^4 \neq y^4$ rules that out too). Thus Pl-#CSP$^2(f''')$ is #P-hard by Theorem D.5. So Pl-#CSP$^2(f)$ is #P-hard.

3. For $f' \in \mathscr{A} \setminus \mathscr{P}$, we have $f' = [1,\rho]^{\otimes 2n-2} + i^r[1,-\rho]^{\otimes 2n-2}$ by definition (See Figure 35).
   - If $f' = [1,1]^{\otimes 2n-2} + i^r[1,-1]^{\otimes 2n-2}$, then $f = x[1,i]^{\otimes 2n} + y[1,-i]^{\otimes 2n} + \frac{1}{2}\{[1,1]^{\otimes 2n} + i^r[1,-1]^{\otimes 2n}\}$ by Proposition A.7 (the Explicit List for $\int(f')$). If $x = y = 0$, then $f \in \mathscr{A}$. In the following, assume that $(x,y) \neq (0,0)$.
     By a holographic transformation using $H = \left[\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}\right]$, we have

$$\text{Pl-Holant}([1,0,1],[1,0,1,0,1],\ldots \mid \hat{f}', \hat{f}) \equiv \text{Pl-\#CSP}^2(f', f),$$

where $\hat{f}' = (H^{-1})^{\otimes 2n-2} f' = [1,0]^{2n-2} + i^r[0,1]^{2n-2} = [1,0,\ldots,0,i^r]$, $\hat{f} = (H^{-1})^{\otimes 2n} f = x'[1,-i]^{\otimes 2n} + y'[1,i]^{\otimes 2n} + \frac{1}{2}\{[1,0]^{\otimes 2n} + i^r[0,1]^{\otimes 2n}\}$, where $x' = \frac{(1+i)^{2n}}{2^{2n}}x$, $y' = \frac{(1-i)^{2n}}{2^{2n}}y$. Note that $(x', y') \neq (0,0)$.
Since we have $[1,0,1]$ on the left and $[1,0,\ldots,0,i^r]$ of arity $2n-2 \geq 4$ on the right in

$$\text{Pl-Holant}([1,0,1],[1,0,1,0,1],\ldots \mid \hat{f}', \hat{f}),$$

we can construct $=_{2k}$ on the right for $k \geq 1$ in the following way: Firstly, connect four copies of $[1,0,\ldots,0,i^r]$ by three copies of $[1,0,1]$ in a planar fashion, to form an equality $[1,0,\ldots,0,1]$ of arity $4(2n-2)-6 = 8n-14$. Then use $4n-9$ copies of $[1,0,1]$ to form loops on $(=_{8n-14})$, and we get $(=_4)$. From this, and $(=_2) = [1,0,1]$ on the left, we can get all $(=_{2k})$ on the right for $k \geq 1$. Then by $=_2$ on the left, we can construct all of $=_{2k}$ on the left. Thus

$$\text{Pl-\#CSP}^2(\hat{f}', \hat{f}) \leq \text{Pl-Holant}([1,0,1],[1,0,1,0,1],\ldots \mid \hat{f}', \hat{f}).$$

By Lemma H.3 Pl-$\#\text{CSP}^2(\hat{f})$ is #P-hard. Thus Pl-$\#\text{CSP}^2(f)$ is #P-hard.

- If $f' = [1,i]^{\otimes 2n-2} + i^r[1,-i]^{\otimes 2n-2}$, then $f = x[1,i]^{\otimes 2n} + y[1,-i]^{\otimes 2n} + \tilde{f}$, where $\tilde{f}$ has arity $2n$ and $\tilde{f}_k = \frac{1}{4}\left\{i^k(2n-2k) + i^r(-i)^k(2n-2k)\right\}$ by Proposition A.7 (the Explicit List for $\int(f')$). Under the holographic transformation by $Z = \left[\begin{smallmatrix} 1 & 1 \\ i & -i \end{smallmatrix}\right]$, the expressions are more revealing: $f = Z^{\otimes 2n}[x,1,0,\ldots,0,i^r,y]$, and $f' = \partial(f) = Z^{\otimes(2n-2)}[1,0,\ldots,0,i^r]$. However, if we apply the holographic transformation Z to Pl-$\#\text{CSP}^2(f,f')$, we have

$$\text{Pl-Holant}([0,1,0],[1,0,1,0,1],\ldots \mid \widehat{f}, \widehat{f'}) \equiv \text{Pl-Holant}(\mathcal{EQ}_2 \mid f, f'),$$

where $\widehat{f} = (Z^{-1})^{\otimes 2n} f = [x,1,0,\ldots,0,i^r,y]$, and $\widehat{f'} = (Z^{-1})^{\otimes 4} f = [1,0,\ldots,0,i^r]$. Note that now we do not have $=_2$ on the left in Pl-Holant$([0,1,0],[1,0,1,0,1],\ldots \mid \widehat{f}, \widehat{f'})$. This is inconvenient to construct gadget. So, in the following steps we first try to construct $[1,0,-1]^{\otimes 2}$ on the LHS of Pl-$\#\text{CSP}^2(f)$ to get Pl-Holant$([1,0,-1]^{\otimes 2} \cup \mathcal{EQ}_2 \mid f)$. This will be done with the help of Lemma A.20. Then after the holographic transformation by Z, we have $[1,0,-1]^{\otimes 2} Z^{\otimes 2} = 4[1,0,1]^{\otimes 2}$ on the left.
To apply Lemma A.20, we construct $[1,i]^{\otimes 4} + i^s[1,-i]^{\otimes 4}$ in Pl-$\#\text{CSP}^2(f,f')$ for some $0 \leq s \leq 3$ as follows.
  - If $2n \equiv 2 \pmod 4$, then we have $\partial_{=_4}^{\frac{n-3}{2}}(f') = 2^{\frac{n-3}{2}}\{[1,i]^{\otimes 4} + i^r[1,-i]^{\otimes 4}\}$.
  - If $2n \equiv 0 \pmod 4$, then we have $\partial_{=_4}^{\frac{n-2}{2}}(f') = 2^{\frac{n-2}{2}}\{[1,i]^{\otimes 2} + i^r[1,-i]^{\otimes 2}\} = 2^{\frac{n-2}{2}}[1 + i^r, (1-i^r)i, -(1+i^r)]$. This is a nonzero multiple of $[1,\pm 1,-1]$ if $r \neq 2$, and a nonzero multiple of $[0,1,0]$ if $r = 2$.
    If $r \neq 2$, then we have $\partial_{[1,\pm 1,-1]}(=_4) = [1,0,-1]$ on the left and

$$\partial_{[1,0,-1]}^{n-2}(f') = 2^{n-2}\{[1,i]^{\otimes 4} + i^r[1,-i]^{\otimes 4}\}.$$

    If $r = 2$, we have $\partial_{[0,1,0]}^{n-2}(f') = (2i)^{n-2}\{[1,i]^{\otimes 4} + i^r(-1)^{n-2}[1,-i]^{\otimes 4}\}$.

Thus we have $f'' = [1,i]^{\otimes 4} + i^s[1,-i]^{\otimes 4}$, for some $0 \leq s \leq 3$, in Pl-$\#\text{CSP}^2(f,f')$. Then by Lemma A.20, we have $[1,0,-1]^{\otimes 2}$ on the left, i.e., we have

$$\text{Pl-Holant}(\mathcal{EQ}_2, [1,0,-1]^{\otimes 2} \mid f, f'') \equiv \text{Pl-\#CSP}^2(f).$$

By a holographic transformation using $Z = \left[\begin{smallmatrix} 1 & 1 \\ i & -i \end{smallmatrix}\right]$, we have

$$\text{Pl-Holant}([1,0,1]^{\otimes 2}, [0,1,0], [1,0,1,0,1], \dots | \widehat{f}, \widehat{f''}) \equiv \text{Pl-Holant}(\mathcal{EQ}_2, [1,0,-1]^{\otimes 2} | f, f''),$$

where $\widehat{f} = (Z^{-1})^{\otimes 2n} f = [x, 1, 0, \dots, 0, i^r, y]$, and $\widehat{f''} = (Z^{-1})^{\otimes 4} f = [1,0,0,0,i^s]$.
In $\text{Pl-Holant}([1,0,1]^{\otimes 2}, [0,1,0], [1,0,1,0,1], \dots | \widehat{f}, \widehat{f''})$, by $[1,0,1]^{\otimes 2}$ on the left and $\widehat{f''}$ on the right, we get $=_4$ on the right as follows: Use 4 copies of $\widehat{f''}$, connected together by 3 copies of $[1,0,1]^{\otimes 2}$ in a planar way. Each copy of $[1,0,1]^{\otimes 2}$ connects two edges of one copy of $\widehat{f''}$ to another copy of $\widehat{f''}$ in such a way that the effect is equivalent to connecting them by two copies of $(=_2) = [1,0,1]$. This way we get an arity $16 - 12 = 4$ signature $(=_4) = [1,0,0,0,(i^s)^4]$. Moreover, we have $=_{4k}$ for $k \geq 1$ on the right by $[1,0,1]^{\otimes 2}$ on the left and $=_4$ on the right in a similar way. Then we can move $\widehat{f}$ to LHS by $[1,0,1]^{\otimes 2}$ because $\widehat{f}$ has even arity. Thus we have

$$\text{Pl-Holant}([1,0,1]^{\otimes 2}, [1,0,1,0,1], \widehat{f} | \mathcal{EQ}_4) \leq \text{Pl-Holant}([1,0,1]^{\otimes 2}, [1,0,1,0,1], \dots | \widehat{f}, \widehat{f''}).$$

Note that

$$\text{Pl-\#CSP}^4(\widehat{f}, [1,0,1,0,1], [1,0,1]^{\otimes 2}) \equiv \text{Pl-Holant}([1,0,1]^{\otimes 2}, [1,0,1,0,1], \widehat{f} \mid \mathcal{EQ}_4).$$

We will prove that $\text{Pl-\#CSP}^4(\widehat{f}, [1,0,1,0,1], [1,0,1]^{\otimes 2})$ is #P-hard to complete the proof of this case.
Note that $\begin{bmatrix} \widehat{f}_0 & \widehat{f}_1 & \widehat{f}_2 \\ \widehat{f}_1 & \widehat{f}_2 & \widehat{f}_3 \\ \widehat{f}_{2n-3} & \widehat{f}_{2n-2} & \widehat{f}_{2n-1} \end{bmatrix} = \left[\begin{smallmatrix} x & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & i^r \end{smallmatrix}\right]$ has rank 3. Thus $\widehat{f}$ does not satisfy any second order recurrence relation by Lemma H.1. So $\widehat{f} \notin \widehat{\mathscr{P}} \cup \widehat{\mathscr{A}} \cup \widehat{\mathscr{M}}$ by Lemma A.11.
If $(x, y) = (0, 0)$, we are done by Lemma H.4. In the following, assume that $(x, y) \neq (0, 0)$.

– If $2n \equiv 0 \pmod 4$, then

$$\text{Pl-\#CSP}^2(\widehat{f}) \leq \text{Pl-\#CSP}^4(\widehat{f}, [1,0,1]^{\otimes 2}) \tag{H.26}$$

by Lemma B.4.
For $\text{Pl-\#CSP}^2(\widehat{f})$, we have $\widehat{f'''} = \partial^{n-2}(\widehat{f}) = [x, 1, 0, i^r, y]$. Note that $\widehat{f'''}$ is redundant. If $(-1)^r x + y \neq 0$, then the compressed signature matrix of $\widehat{f'''}$ is nonsingular and we are done by Lemma A.25.
Otherwise, we have $x = \pm y$, and thus both $x, y \neq 0$. It is easy to see that $\widehat{f'''}$ does not satisfy the second order recurrence relations $\langle 0, 1, 0\rangle$, $\langle 1, 0, \pm 1\rangle$, $\langle 1, 0, \pm i\rangle$. Thus $\widehat{f'''} \notin \widehat{\mathscr{P}} \cup \widehat{\mathscr{A}}$ by Lemma A.11.
We consider three possibilities for $\widehat{f'''}$.
• If $\widehat{f'''} \in \widehat{\mathscr{M}} \setminus (\widehat{\mathscr{P}} \cup \widehat{\mathscr{A}})$, then $\text{Pl-\#CSP}^2(\widehat{f}, \widehat{f'''})$ is #P-hard by Lemma E.3, where we have $\widehat{f} \notin \widehat{\mathscr{M}}$ because we have noted earlier that $\widehat{f} \notin \widehat{\mathscr{P}} \cup \widehat{\mathscr{A}} \cup \widehat{\mathscr{M}}$. Thus $\text{Pl-\#CSP}^4(\widehat{f}, [1,0,1]^{\otimes 2})$ is #P-hard by (H.26) and we are done.
• If $\widehat{f'''} \in \widehat{\mathscr{M}^\dagger} \setminus (\widehat{\mathscr{P}} \cup \widehat{\mathscr{A}})$, then $\widehat{f'''} = [x, 1, 0, 1, -x]$ by Corollary A.18 (the other form $[u, v, w, v, u]$ with $(u + w)w = 2v^2$ in Corollary A.18 is impossible because $w = 0$ here and $(u + w)w = 2v^2$ would force $v = 0$.) Then we are done by Lemma E.4, because $\widehat{f'''}$ plays the role of g in Lemma E.4, and $\widehat{f} \notin \widehat{\mathscr{M}^\dagger}$ by $\widehat{f} \notin \widehat{\mathscr{P}} \cup \widehat{\mathscr{A}} \cup \widehat{\mathscr{M}}$.
• If $\widehat{f'''} \notin \widehat{\mathscr{P}} \cup \widehat{\mathscr{A}} \cup \widehat{\mathscr{M}}$, then $\text{Pl-\#CSP}^2(\widehat{f'''})$ is #P-hard by Theorem D.5 and we are done.

– For $2n \equiv 2 \pmod 4$, we cannot use the reduction in (H.26) since Lemma B.4 requires that all signatures on the right have arity $\equiv 0 \pmod 4$. We get around this difficulty by constructing some arity 4 signatures in Pl-#CSP$^4(\widehat{f})$, and then use Lemma B.4 for these arity 4 signatures.

Firstly, we have $\widehat{g} = \partial_{=_4}^{\frac{n-3}{2}}(\widehat{f}) = [x, 1, 0, 0, 0, i^r, y]$. We also have $\partial_{=_4}(\widehat{g}) = [x, 1+i^r, y]$. They are both on the right. Then we have $\partial_{[x,1+i^r,y]}(=_4) = [x, 0, y]$ on the left. We also connect $[x, 0, y]$ and $[x, 1 + i^r, y]$ and then $[x, 0, y]$ in a chain, to get another binary signature $h = [x^3, (1 + i^r)xy, y^3]$ on the left. This can be verified by

$$\begin{bmatrix} x & 0 \\ 0 & y \end{bmatrix} \begin{bmatrix} x & 1+i^r \\ 1+i^r & y \end{bmatrix} \begin{bmatrix} x & 0 \\ 0 & y \end{bmatrix} = \begin{bmatrix} x^3 & (1+i^r)xy \\ (1+i^r)xy & y^3 \end{bmatrix}.$$

From these we produce two arity 4 signatures on the right:

$$\begin{aligned} \widehat{g'} &= \partial_{[x,0,y]}(\widehat{g}) = [x^2, x, 0, i^r y, y^2] \\ \widehat{g''} &= \partial_h(\widehat{g}) \quad = [x^4 + 2(1 + i^r)xy, x^3, 0, i^r y^3, y^4 + 2i^r(1 + i^r)xy]. \end{aligned}$$

Thus

$$\text{Pl-\#CSP}^4(\widehat{g'}, \widehat{g''}, [1, 0, 1]^{\otimes 2}, [1, 0, 1, 0, 1]) \le \text{Pl-\#CSP}^4(\widehat{f}, [1, 0, 1]^{\otimes 2}, [1, 0, 1, 0, 1]).$$

Moreover, note that all signatures in $\{\widehat{g'}, \widehat{g''}, [1, 0, 1, 0, 1])\}$ have arity 4. Then by Lemma B.4, we have

$$\text{Pl-\#CSP}^2(\widehat{g'}, \widehat{g''}, [1, 0, 1, 0, 1]) \le \text{Pl-\#CSP}^4(\widehat{g''}, \widehat{g''}, [1, 0, 1]^{\otimes 2}, [1, 0, 1, 0, 1]).$$

It is easy to see that $\widehat{g'}$ is non-degenerate and does not satisfy the second order recurrence relations $\langle 0, 1, 0\rangle$, $\langle 1, 0, \pm 1\rangle$, $\langle 1, 0, \pm i\rangle$, because $(x, y) \neq (0, 0)$. Thus $\widehat{g'} \notin \mathscr{P} \cup \widetilde{\mathscr{A}}$ by Lemma A.11. If $\widehat{g'} \notin \widehat{\mathscr{M}}$, then Pl-#CSP$^2(\widehat{g'})$ is #P-hard by Theorem D.5 and we are done.

Otherwise, $\widehat{g'} \in \widehat{\mathscr{M}} \setminus (\mathscr{P} \cup \widetilde{\mathscr{A}})$ or $\widehat{g'} \in \widehat{\mathscr{M}^\dagger} \setminus (\mathscr{P} \cup \widetilde{\mathscr{A}})$.

Note that $[1, 0, 1, 0, 1]$ has type $\langle 1, 0, -1\rangle$ and the second order recurrence relation is unique up to a scalar. Thus $[1, 0, 1, 0, 1] \notin \widehat{\mathscr{M}}$ by Lemma A.11. If $\widehat{g'} \in \widehat{\mathscr{M}} \setminus (\mathscr{P} \cup \widetilde{\mathscr{A}})$, then Pl-#CSP$^2(\widehat{g'}, \widehat{g''}, [1, 0, 1, 0, 1])$ is #P-hard by Lemma E.3 and we are done.

Therefore we may assume $\widehat{g'} \in \widehat{\mathscr{M}^\dagger} \setminus (\mathscr{P} \cup \widetilde{\mathscr{A}} \cup \widehat{\mathscr{M}})$.

By Corollary A.18, for $\widehat{g'} \in \widehat{\mathscr{M}^\dagger} \setminus (\mathscr{P} \cup \widetilde{\mathscr{A}} \cup \widehat{\mathscr{M}})$, it cannot be of the form $[u, v, w, -v, u]$ with $(u-w)w = 2v^2$; for if it were so, then by $w = 0$ in this case, we would have $v = 0$, and this would imply that $x = i^r y = 0$ in $\widehat{g'}$. It contradicts that $(x, y) \neq (0, 0)$. So $\widehat{g'}$ must be of the form $[u, v, 0, v, -u]$, i.e., $x^2 = -y^2$, $x = i^r y$. Thus we have $x = \epsilon i y$ and $i^r = \epsilon i$, for some $\epsilon = \pm 1$, and $x \neq 0$. Hence both $x, y \neq 0$ and $1 + i^r \neq 0$. It follows that $x^3 = -\epsilon i y^3 \neq \epsilon i y^3 = i^r y^3$.

Moreover, if $\widehat{g''} \in \widehat{\mathscr{M}^\dagger}$, it cannot take the form $[u, v, w, -v, u]$ with $(u - w)w = 2v^2$ in Corollary A.18 because if so then $w = 0$ would force $v = 0$ and that would force both $x = y = 0$. Then $\widehat{g''}$ must be of the form $[u, v, 0, v, -u]$. But this would force $x^3 = i^r y^3$, a contradiction. Thus $\widehat{g''} \notin \widehat{\mathscr{M}^\dagger}$.

If $\widehat{g''} \notin \mathscr{P} \cup \widetilde{\mathscr{A}} \cup \widehat{\mathscr{M}}$, then Pl-#CSP$^2(\widehat{g''})$ is #P-hard by Theorem D.5 and we are done. Otherwise, $\widehat{g''} \in (\mathscr{P} \cup \widetilde{\mathscr{A}} \cup \widehat{\mathscr{M}}) \setminus \widehat{\mathscr{M}^\dagger}$, Pl-#CSP$^2(\widehat{g'}, \widehat{g''}, [1, 0, 1, 0, 1])$ is #P-hard by Lemma E.4 and we are done.

4. For $f' \in \widehat{\mathscr{M}} \setminus (\mathscr{P} \cup \widetilde{\mathscr{A}})$, we are done by Lemma E.3.
5. For $f' \in \widehat{\mathscr{M}^\dagger} \setminus (\mathscr{P} \cup \widetilde{\mathscr{A}})$, or equivalently, $f' \in \widehat{\mathscr{M}^\dagger} \setminus (\mathscr{P} \cup \widetilde{\mathscr{A}} \cup \widehat{\mathscr{M}})$, $f' = [s, ti]^{\otimes 2n-2} \pm [t, si]^{\otimes 2n-2}$, $st \neq 0$, $s^4 \neq t^4$, or f' has arity $2n-2$ and $f'_k = (\pm i)^k (2n-2-2k)$ by Lemma A.14. Note that we are done if we have a nonzero binary signature that is not $\lambda[1,0,1]$ by Lemma E.7. Moreover, if we have an arity 4 signature h that is not in $\widehat{\mathscr{M}^\dagger}$ then we are done by the following argument: if $h \in (\mathscr{P} \cup \widetilde{\mathscr{A}} \cup \widehat{\mathscr{M}}) \setminus \widehat{\mathscr{M}^\dagger}$, then Pl-#CSP$^2(h, f')$ is #P-hard by Theorem G.4 since $f' \in \widehat{\mathscr{M}^\dagger} \setminus (\mathscr{P} \cup \widetilde{\mathscr{A}} \cup \widehat{\mathscr{M}})$; if $h \notin \mathscr{P} \cup \widetilde{\mathscr{A}} \cup \widehat{\mathscr{M}}$, then Pl-#CSP$^2(h)$ is #P-hard by Theorem D.5.
   - For $f' = [s, ti]^{\otimes 2n-2} + [t, si]^{\otimes 2n-2}$ with $2n \equiv 0 \pmod 4$ or $f' = [s, ti]^{\otimes 2n-2} - [t, si]^{\otimes 2n-2}$ with $2n \equiv 2 \pmod 4$, we have $\partial^{n-1}(f) = (s^2+t^2)(s^2-t^2)^{n-1}[1, \frac{2sti}{s^2+t^2}, -1] \neq \lambda[1,0,1]$.
   - For $f' = [s, ti]^{\otimes 2n-2} + [t, si]^{\otimes 2n-2}$ with $2n \equiv 2 \pmod 4$, $f = x[1,i]^{\otimes 2n} + y[1,-i]^{\otimes 2n} + \frac{1}{s^2-t^2}\{[s,ti]^{\otimes 2n} - [t,si]^{\otimes 2n}\}$ by Proposition A.7 (the Explicit List for $\int(f')$). If $x = y = 0$, then $f \in \widehat{\mathscr{M}^\dagger}$. Otherwise, we have

$$\begin{aligned} f''' = \partial_{=_4}^{\frac{n-1}{2}}(f) &= 2^{\frac{n-1}{2}} x[1,i]^{\otimes 2} + 2^{\frac{n-1}{2}} y[1,-i]^{\otimes 2} + \frac{(s^4+t^4)^{\frac{n-1}{2}}}{s^2-t^2}\left\{[s,ti]^{\otimes 2} - [t,si]^{\otimes 2}\right\} \\ &= 2^{\frac{n-1}{2}} x[1,i]^{\otimes 2} + 2^{\frac{n-1}{2}} y[1,-i]^{\otimes 2} + (s^4+t^4)^{\frac{n-1}{2}}[1,0,1] \end{aligned}$$

     Let $a = 2^{\frac{n-1}{2}}(x+y)$, $b = 2^{\frac{n-1}{2}}(x-y)i$ and $c = (s^4+t^4)^{\frac{n-1}{2}}$, then $f''' = [c+a, b, c-a]$. Note that $(a,b) \neq (0,0)$. If $b \neq 0$, it is obvious that $f''' \neq \lambda[1,0,1]$. If $b = 0$, then $a \neq 0$. Then $f''' \neq \lambda[1,0,1]$ by $c + a \neq c - a$.
   - For the case that f' has arity $2n-2$ and $f'_k = (\pm i)^k(2n-2-2k)$ with $2n \equiv 2 \pmod 4$, we have $f'' = \partial_{=_4}^{\frac{n-3}{2}}(f')$ which has arity 4 and $f''_k = 2^{\frac{n-3}{2}}(\pm i)^k(4-2k)$. Moreover, we have $\partial(f'') = 2^{\frac{n+1}{2}}[1, \pm i, -1] \neq \lambda[1,0,1]$. We remark that it is necessary to use $=_4$ that many times, since f with two loops by $=_2$ is already identically zero.
   - For the case that f' has arity $2n-2$ and $f'_k = (\pm i)^k(2n-2-2k)$ with $2n \equiv 0 \pmod 4$, we may consider only the case where the sign $\pm$ is $+$. Indeed under $Z = \left[\begin{smallmatrix} 1 & 1 \\ i & -i \end{smallmatrix}\right]$, for the $+$ sign $f' = Z^{\otimes(2n-2)}[0,1,0,\ldots,0]$ and for the $-$ sign $f' = Z^{\otimes(2n-2)}[0,\ldots,0,1,0]$, a reversal under the Z-transformation. If we take a holographic transformation by $T = \left[\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right]$, we have $TZ = \left[\begin{smallmatrix} 1 & 1 \\ -i & i \end{smallmatrix}\right] = Z\left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right]$, and so $(TZ)^{\otimes(2n-2)}[0,\ldots,0,1,0] = Z^{\otimes(2n-2)}[0,1,0,\ldots,0]$. Meanwhile, $\mathcal{EQ}_2$ is invariant under T.
     Thus we consider f' of arity $2n-2$ where $f'_k = i^k(2n-2-2k)$ with $2n \equiv 0 \pmod 4$. Let $\hat{f}' = (Z^{-1})^{\otimes(2n-2)} f' = [0,1,0,\ldots,0]$ and let $\hat{f} = (Z^{-1})^{\otimes(2n-2)} f$. Then we have $(Z^{-1})^{\otimes(2n-2)}(\partial(f)) = \partial_{[0,1,0]}(\hat{f})$ up to a scalar. This implies $\partial_{[0,1,0]}(\hat{f}) = [0,1,0,\ldots,0]$. Thus there exist constants x and y such that $\hat{f} = [x,0,1,0,\ldots,0,y]$. By the holographic transformation using Z, we have

$$\text{Pl-\#CSP}^2(f) \equiv \text{Pl-Holant}([0,1,0],[1,0,1,0,1],\ldots \mid \hat{f}).$$

     We remark that, in Pl-Holant$([0,1,0],[1,0,1,0,1],\ldots \mid \hat{f})$, all signatures have even arities. And all signatures of arity $2m \equiv 2 \pmod 4$ satisfy odd parity and all signatures of arity $2m \equiv 0 \mod 4$ satisfy even parity. Then by the statement of Remark 5, any binary signature constructed in Pl-Holant$([0,1,0],[1,0,1,0,1],\ldots \mid \hat{f})$ can only be of the form $\lambda[0,1,0]$. This implies that the binary signature constructed in Pl-#CSP$^2(f)$ can only be of the form $\lambda[1,0,1]$ before the Z-transformation. This forces us to construct signatures of arity at least 4 to prove hardness.

In Pl-Holant$([0,1,0],[1,0,1,0,1],\dots|\hat{f})$, note that by $2n \equiv 0 \mod 4$ we have $2n \geq 8$, and $\hat{g} = \partial^{\frac{n-2}{2}}_{[1,0,1,0,1]}(\hat{f}) = [x + \frac{n-2}{2} \cdot 6, 0, 1, 0, y]$. It has arity 4. If $(x + \frac{n-2}{2} \cdot 6)y \neq 1$, then $\hat{g} \notin \mathscr{M}$, because symmetric matchgate signatures must form geometric series in alternate terms. Thus we have $Z^{\otimes 4}(\hat{g}) \notin \widehat{\mathscr{M}^\dagger}$ in Pl-#CSP$^2(f, f')$ and we are done.
If $(x + \frac{n-2}{2} \cdot 6)y = 1$, then $y \neq 0$. Firstly, we have an arity 8 signature

$$\widehat{g'} = \partial^{\frac{n-4}{2}}_{[1,0,1,0,1]}(\hat{f}) = [x + \frac{n-4}{2} \cdot 6, 0, 1, 0, 0, 0, 0, 0, y]$$

(note that $n \geq 4$ when $2n \equiv 0 \mod 4$), and we have $\partial^2_{[0,1,0]}(\widehat{g'}) = [1, 0]^{\otimes 4}$ on the right. So we have $[0, 1]^{\otimes 4}$ on the left. Moreover, we have $\partial_{[0,1]^{\otimes 4}}(\widehat{g'}) = y[0, 1]^{\otimes 4}$ on the right. So we have $[1, 0]^{\otimes 4}$ on the left. Then we have $\widehat{g''} = \partial^{\frac{n-2}{2}}_{[1,0]^{\otimes 4}}(\hat{f}) = [x, 0, 1, 0, 0]$ on the right. Note that $\widehat{g''} \notin \mathscr{M}$. Thus we have $Z^{\otimes 4}(\widehat{g''}) \notin \widehat{\mathscr{M}^\dagger}$ in Pl-#CSP$^2(f, f')$ and we are done.

- For the last case of Case 5, $f' = [s, ti]^{\otimes 2n-2} - [t, si]^{\otimes 2n-2}$ with $2n \equiv 0 \pmod 4$, we let $u = \frac{s-t}{s+t}$, then $u^4 \neq 0, 1$ by Lemma A.3. Let $Z = \left[\begin{smallmatrix} 1 & 1 \\ i & -i \end{smallmatrix}\right]$, then

$$\begin{aligned}
\widehat{f'} &= (Z^{-1})^{\otimes 2n-2}(f') \\
&= \frac{1}{2^{2n-2}} \left\{[s + t, s - t]^{\otimes 2n-2} - [s + t, t - s]^{\otimes 2n-2}\right\} \\
&= \frac{(s + t)^{2n-2}}{2^{2n-2}} \left\{[1, u]^{\otimes 2n-2} - [1, -u]^{\otimes 2n-2}\right\} \\
&= \lambda[0, u^2, 0, u^4, \dots, u^{2n-2}, 0],
\end{aligned}$$

where $\lambda = \frac{(s+t)^{2n-2}}{2^{2n-3}u} \neq 0$. Let $(Z^{-1})^{\otimes 2n} f = \widehat{f}$, then $(Z^{-1})^{\otimes 2n-2}(\partial(f)) = \partial_{[0,1,0]}(\widehat{f})$ up to a scalar. This implies that $\partial_{[0,1,0]}(\widehat{f}) = \lambda[0, u^2, 0, u^4, \dots, u^{2n-2}, 0]$. Thus there exist constants x and y such that $\widehat{f} = (Z^{-1})^{\otimes 2n} f = \lambda[1 + x, 0, u^2, 0, u^4, \dots, u^{2n-2}, 0, u^{2n} + y]$, where we append the terms 1 and u^{2n} for future convenience. (This can be accommodated by naming different x and y.) If $x = y = 0$, then $\widehat{f} \in \mathscr{M}$ and $f \in \widehat{\mathscr{M}^\dagger}$. In the following, assume that $(x, y) \neq (0, 0)$. By the holographic transformation using Z, we have

$$\text{Pl-\#CSP}^2(f) \equiv \text{Pl-Holant}([0,1,0],[1,0,1,0,1],\dots \mid \widehat{f}).$$

By the same argument as the previous case, it is impossible to construct a "good" binary signature in this case. So we have to construct signatures of arity at least 4 to prove hardness.
We will repeatedly use the following computation in the remainder of this proof: Let $\bar{g} = \partial_{[1,0,v,0,v^2]}(g)$ for some v, then $\text{arity}(\bar{g}) = \text{arity}(g) - 4$ and $\bar{g}_k = g_k + 6vg_{k+2} + v^2 g_{k+4}$.
We will complete the proof by constructing some arity 4 signatures $\widehat{h}$ in the setting after the Z-transformation Pl-Holant$([0,1,0],[1,0,1,0,1],\dots \mid \widehat{f})$ that cannot all belong to $\mathscr{M}$. We note that if $\widehat{h} \notin \mathscr{M}$ then $h = Z^{\otimes 4}\widehat{h} \notin \widehat{\mathscr{M}^\dagger}$. This will imply Pl-#CSP$^2(h, f')$ is #P-hard as noted earlier, thus complete the proof of this Case 5.
In Pl-Holant$([0,1,0],[1,0,1,0,1],\dots \mid \widehat{f})$, we have $\partial^{n-2}_{[0,1,0]}(\widehat{f})$ which is a nonzero multiple of $[1, 0, u^2, 0, u^4]$. Then we have $[u^4, 0, u^2, 0, 1] = u^4[1, 0, u^{-2}, 0, u^{-4}]$ on the left. Ignoring $\lambda \neq 0$, we write

$$\widehat{f} = [1, 0, u^2, 0, u^4, 0, \dots, 0, u^{2n}] + [x, 0, 0, 0, 0, \dots, 0, y]$$

which has arity $2n \geq 8$, and we have

$$\begin{aligned}\widehat{f^{(4)}} &= \partial_{[1,0,u^{-2},0,u^{-4}]}^{\frac{n-4}{2}}(\widehat{f}) \\ &= 8^{\frac{n-4}{2}}[1,0,u^2,0,u^4,0,u^6,0,u^8] + [x,0,0,0,0,0,0,0,yu^{-2(n-4)}] \\ &= [x+8^{\frac{n-4}{2}},0,8^{\frac{n-4}{2}}u^2,0,8^{\frac{n-4}{2}}u^4,0,8^{\frac{n-4}{2}}u^6,0,8^{\frac{n-4}{2}}u^8+yu^{-2(n-4)}].\end{aligned}$$

Let $x' = \frac{x}{8^{\frac{n-4}{2}}}$, $y' = \frac{yu^{-2(n-4)}}{8^{\frac{n-4}{2}}}$, then $\widehat{f^{(4)}} = [x'+1,0,u^2,0,u^4,0,u^6,0,u^8+y']$ up to the scalar $8^{\frac{n-4}{2}}$. Further, we have $\widehat{f^{(5)}} = \partial_{[1,0,u^{-2},0,u^{-4}]}(\widehat{f^{(4)}}) = [x'+8,0,8u^2,0,8u^4+y'u^{-4}]$. If $x'=0$ or $y'=0$, then $\widehat{f^{(5)}} \notin \mathscr{M}$ by $(x'+8)(8u^4+y'u^{-4}) \neq (8u^2)^2$ and we are done. So we can assume that $x'y' \neq 0$ in the following.

In the following, if we have the signature $[1,0,v,0,v^2]$ with $v \neq 0$ on the left, then we have $\partial_{[1,0,v,0,v^2]}(\widehat{f^{(4)}}) = [x'+c,0,cu^2,0,y'v^2+cu^4]$, where $c = 1+6u^2v+u^4v^2$. If $c=0$, then we have $[x',0,0,0,y'v^2] \notin \mathscr{M}$ and we are done. So in the following, we always suppose that $c = 1+6u^2v+u^4v^2 \neq 0$. Moreover, if $(x'+c)(y'v^2+cu^4) \neq (cu^2)^2$, then $[x'+c,0,cu^2,0,y'v^2+cu^4] \notin \mathscr{M}$ and we are done. So we assume that $(x'+c)(y'v^2+cu^4) = (cu^2)^2$. This implies that $x'+c \neq 0$ and $x'y'v^2 + (x'u^4+y'v^2)c = 0$. To summerize, in the following if we have $[1,0,v,0,v^2]$ with $v \neq 0$ on the left, then we have

$$\begin{aligned} &c = 1+6u^2v+u^4v^2 \neq 0, \\ &x'+c \neq 0, \\ &x'y'v^2 + (x'u^4+y'v^2)c = 0. \end{aligned} \tag{H.27}$$

Firstly, by $\widehat{f^{(5)}} = \partial_{[1,0,u^{-2},0,u^{-4}]}(\widehat{f^{(4)}}) = [x'+8,0,8u^2,0,8u^4+y'u^{-4}]$ and (H.27), we have

$$\begin{aligned} &x'+8 \neq 0, \\ &x'y'u^{-4} + 8(x'u^4+y'u^{-4}) = 0. \end{aligned} \tag{H.28}$$

Note that we have $[1,0,1,0,1]$ on the left, so we have $\widehat{f^{(6)}} = \partial_{[1,0,1,0,1]}(\widehat{f^{(4)}}) = [x'+c_1,0,c_1u^2,0,y'+c_1u^4]$, where $c_1 = 1+6u^2+u^4$. Then by (H.27), we have $c_1 \neq 0$ and

$$\begin{aligned} &x'+c_1 \neq 0, \\ &x'y' + (x'u^4+y')c_1 = 0. \end{aligned} \tag{H.29}$$

By (H.28), (H.29), and $x'y' \neq 0$, we have

$$\begin{aligned} 1+(\frac{u^8}{y'}+\frac{1}{x'})8 = 0, \\ 1+(\frac{u^4}{y'}+\frac{1}{x'})c_1 = 0. \end{aligned}$$

Then we have

$$\begin{aligned} \frac{1}{x'} = \frac{c_1-8u^4}{8c_1(u^4-1)} &= -\frac{1+7u^2}{8(1+u^2)(1+6u^2+u^4)}, \\ \frac{1}{y'} = \frac{8-c_1}{8c_1(u^8-u^4)} &= -\frac{7+u^2}{8u^4(1+u^2)(1+6u^2+u^4)}. \end{aligned}$$

Since $x' \neq 0$, we have $1 + 7u^2 \neq 0$.
For $\widehat{f^{(5)}}$, $\widehat{f^{(6)}}$, let $v_2 = \frac{x'+8}{8u^2}$ and $v_3 = \frac{x'+c_1}{c_1u^2}$, then $v_2 \neq 0$, $v_3 \neq 0$ by $x' + 8 \neq 0$ and $x' + c_1 \neq 0$, and $\widehat{f^{(5)}} = [1, 0, v_2^{-1}, 0, v_2^{-2}]$, $\widehat{f^{(6)}} = [1, 0, v_3^{-1}, 0, v_3^{-2}]$ up to the scalars $x' + 8$, $x' + c_1$ respectively. So we have $[1, 0, v_2, 0, v_2^2]$, $[1, 0, v_3, 0, v_3^2]$ on the left. Moreover, let $c_2 = 1 + 6u^2v_2 + u^4v_2^2$, $c_3 = 1 + 6u^2v_3 + u^4v_3^2$, then we have by (H.27)

$$\begin{aligned} x'y'v_2^2 + (x'u^4 + y'v_2^2)c_2 = 0, \\ x'y'v_3^2 + (x'u^4 + y'v_3^2)c_3 = 0. \end{aligned} \tag{H.30}$$

In (H.30), we have

$$\begin{aligned} c_1 &= 1 + 6u^2 + u^4, \\ \frac{1}{x'} &= \frac{c_1 - 8u^4}{8c_1(u^4 - 1)} = -\frac{7u^2 + 1}{8(u^2 + 1)(u^4 + 6u^2 + 1)}, \\ \frac{1}{y'} &= \frac{8 - c_1}{8c_1(u^8 - u^4)} = -\frac{u^2 + 7}{8u^4(u^2 + 1)(u^4 + 6u^2 + 1)}, \\ v_2 &= \frac{x' + 8}{8u^2} = \frac{-7u^2 - u^4}{7u^2 + 1}, \\ c_2 &= 1 + 6u^2v_2 + u^4v_2^2 = \frac{u^{12} + 14u^{10} + 7u^8 - 300u^6 + 7u^4 + 14u^2 + 1}{(7u^2 + 1)^2}, \\ v_3 &= \frac{x' + c_1}{c_1u^2} = -\frac{7 + u^2}{u^2(1 + 7u^2)}, \\ c_3 &= 1 + 6u^2v_3 + u^4v_3^2 = \frac{8u^4 - 272u^2 + 8}{(1 + 7u^2)^2}. \end{aligned}$$

Note that all of them are functions of u. Thus (H.30) gives two equations of u as following:

$$\begin{aligned} \frac{8u^4c_1^2(1 + u^2)^2 \cdot p_1(u)}{(1 + 7u^2)^4} = 0, \\ \frac{3072u^2(1 + u^2)^2c_1 \cdot p_2(u)}{(1 + 7u^2)^4} = 0, \end{aligned} \tag{H.31}$$

where $p_1(u) = u^{12} + 14u^{10} - 49u^8 - 700u^6 - 49u^4 + 14u^2 + 1$, $p_2(u) = 7u^4 + 2u^2 + 7$. Note that $q_1(u)p_1(u) + q_2(u)p_2(u) = 244224$, where $q_1(u) = -188 - 315u^2$, $q_2(u) = 34916 - 9555u^2 - 32872u^4 - 2058u^6 + 644u^8 + 45u^{10}$, thus $\gcd(p_1(u), p_2(u)) = 1$. Then by $u^4 \neq 0, 1$, $c_1 \neq 0$, the two equations in (H.30) have no common solution in u. This is a contradiction and we finish the proof. □

We hereby finish the proof of Theorem H.5, and hence we complete the proof of the main theorem of Part II—Theorem A.2 is a straightforward combination of Theorem C.13, Theorem H.5 and Theorem G.4.

References

[1] Rodney J. Baxter. *Exactly solved models in statistical mechanics*. Academic press London, 1982.

[2] Jin-Yi Cai, Xi Chen, Richard J. Lipton, and Pinyan Lu. On tractable exponential sums. In *FAW*, pages 148–159. Springer Berlin Heidelberg, 2010.

[3] Jin-Yi Cai and Vinay Choudhary. Some results on matchgates and holographic algorithms. *Int. J. Software and Informatics*, 1(1):3–36, 2007.

[4] Jin-Yi Cai, Vinay Choudhary, and Pinyan Lu. On the theory of matchgate computations. *Theory Comput. Syst.*, 45(1):108–132, 2009.

[5] Jin-Yi Cai and Aaron Gorenstein. Matchgates revisited. *Theory Comput.*, 10(7):167–197, 2014.

[6] Jin-Yi Cai, Heng Guo, and Tyson Williams. A complete dichotomy rises from the capture of vanishing signatures (extended abstract). In *STOC*, pages 635–644. ACM, 2013. *CoRR*, abs/1204.6445.

[7] Jin-Yi Cai, Heng Guo, and Tyson Williams. Holographic algorithms beyond matchgates. In *ICALP*, pages 271–282. Springer Berlin Heidelberg, 2014. *CoRR*, abs/1307.7430.

[8] Jin-Yi Cai and Michael Kowalczyk. Spin systems on k-regular graphs with complex edge functions. *Theoretical Computer Science*, 2012. DOI:10.1016/j.tcs.2012.01.021.

[9] Jin-Yi Cai, Michael Kowalczyk, and Tyson Williams. Gadgets and anti-gadgets leading to a complexity dichotomy. In *ITCS*, pages 452–467. ACM, 2012.

[10] Jin-Yi Cai and Pinyan Lu. On symmetric signatures in holographic algorithms. *Theory Comput. Syst.*, 46(3):398–415, 2010.

[11] Jin-Yi Cai and Pinyan Lu. Holographic algorithms: From art to science. *J. Comput. Syst. Sci.*, 77(1):41–61, 2011.

[12] Jin-Yi Cai, Pinyan Lu, and Mingji Xia. Holographic algorithms with matchgates capture precisely tractable planar #CSP. In *FOCS*, pages 427–436. IEEE Computer Society, 2010.

[13] Jin-Yi Cai, Pinyan Lu, and Mingji Xia. Computational complexity of Holant problems. *SIAM J. Comput.*, 40(4):1101–1132, 2011.

[14] Jin-Yi Cai, Pinyan Lu, and Mingji Xia. Dichotomy for Holant* problems of Boolean domain. In *SODA*, pages 1714–1728, 2011.

[15] Jin-Yi Cai, Pinyan Lu, and Mingji Xia. The complexity of complex weighted Boolean #CSP. *J. Comput. System Sci.*, 80(1):217–236, 2014.

[16] C. T. J. Dodson and T. Poston. *Tensor Geometry*. Graduate Texts in Mathematics. Springer-Verlag, second edition, 1991.

[17] Jan Draisma, Dion C. Gijswijt, László Lovász, Guus Regts, and Alexander Schrijver. Characterizing partition functions of the vertex model. *J. Algebra*, 350:197–206, 2012.

[18] Michael Freedman, László Lovász, and Alexander Schrijver. Reflection positivity, rank connectivity, and homomorphism of graphs. *J. Amer. Math. Soc.*, 20(1):37–51, 2007.

[19] Heng Guo, Pinyan Lu, and Leslie G. Valiant. The complexity of symmetric Boolean parity Holant problems. *SIAM J. Comput.*, 42(1):324–356, 2013.

[20] Heng Guo and Tyson Williams. The complexity of planar Boolean #CSP with complex weights. In *ICALP*, pages 516–527. Springer Berlin Heidelberg, 2013. *CoRR*, abs/1212.2284.

[21] Sangxia Huang and Pinyan Lu. A dichotomy for real weighted Holant problems. In *CCC*, pages 96–106. IEEE Computer Society, 2012. Full version available at http://www.csc.kth.se/~sangxia/papers/2012-ccc.pdf.

[22] Ernst Ising. Beitrag zür theorie des ferromagnetismus. *Zeitschrift für Physik*, 31(1):253–258, 1925.

[23] P. W. Kasteleyn. The statistics of dimers on a lattice. *Physica*, 27:1209–1225, 1961.

[24] P. W. Kasteleyn. Graph theory and crystal physics. In F. Harary, editor, *Graph Theory and Theoretical Physics*, pages 43–110. Academic Press, London, 1967.

[25] Michael Kowalczyk. *Dichotomy theorems for Holant problems.* PhD thesis, University of Wisconsin—Madison, 2010. http://cs.nmu.edu/~mkowalcz/research/main.pdf.

[26] J. M. Landsberg, Jason Morton, and Serguei Norine. Holographic algorithms without matchgates. *Linear Algebra Appl.*, 438(2):782–795, 2013.

[27] T. D. Lee and C. N. Yang. Statistical theory of equations of state and phase transitions. II. Lattice gas and Ising model. *Phys. Rev.*, 87(3):410–419, 1952.

[28] Elliott H. Lieb. Residual entropy of square ice. *Phys. Rev.*, 162(1):162–172, 1967.

[29] Elliott H. Lieb and Alan D. Sokal. A general Lee-Yang theorem for one-component and multicomponent ferromagnets. *Comm. Math. Phys.*, 80(2):153–179, 1981.

[30] Jason Morton. Pfaffian circuits. *CoRR*, abs/1101.0129, 2011.

[31] Jason Morton and Susan Margulies. Polynomial-time solvable #CSP problems via algebraic models and Pfaffian circuits. *CoRR*, abs/1311.4066, 2013. To appear in Journal of Symbolic Computation.

[32] Lars Onsager. Crystal statistics. I. A two-dimensional model with an order-disorder transition. *Phys. Rev.*, 65(3-4):117–149, 1944.

[33] Alexander Schrijver. Characterizing partition functions of the spin model by rank growth. *Indag. Math. (N.S.)*, 24(4):1018–1023, 2013.

[34] H. N. V. Temperley and M. E. Fisher. Dimer problem in statistical mechanics—an exact result. *Philosophical Magazine*, 6:1061–1063, 1961.

[35] Leslie G. Valiant. Expressiveness of matchgates. *Theor. Comput. Sci.*, 289(1):457–471, 2002.

[36] Leslie G. Valiant. Quantum circuits that can be simulated classically in polynomial time. *SIAM J. Comput.*, 31(4):1229–1254, 2002.

[37] Leslie G. Valiant. Accidental algorthims. In *FOCS*, pages 509–517. IEEE Computer Society, 2006.

[38] Leslie G. Valiant. Holographic algorithms. *SIAM J. Comput.*, 37(5):1565–1594, 2008.

[39] Leslie G. Valiant. Some observations on holographic algorithms. In *LATIN*, pages 577–590. Springer Berlin Heidelberg, 2010.

[40] Dirk Vertigan. The computational complexity of Tutte invariants for planar graphs. *SIAM Journal on Computing*, 35(3):690–712, 2005.

[41] Dirk Llewellyn Vertigan. *On the computational complexity of Tutte, Jones, Homfly and Kauffman invariants*. PhD thesis, University of Oxford, 1991.

[42] Dominic Welsh. *Complexity: Knots, Colourings and Countings*. London Mathematical Society Lecture Note Series. Cambridge University Press, 1993.

[43] C. N. Yang. The spontaneous magnetization of a two-dimensional Ising model. *Phys. Rev.*, 85(5):808–816, 1952.

[44] C. N. Yang and T. D. Lee. Statistical theory of equations of state and phase transitions. I. Theory of condensation. *Phys. Rev.*, 87(3):404–409, 1952.

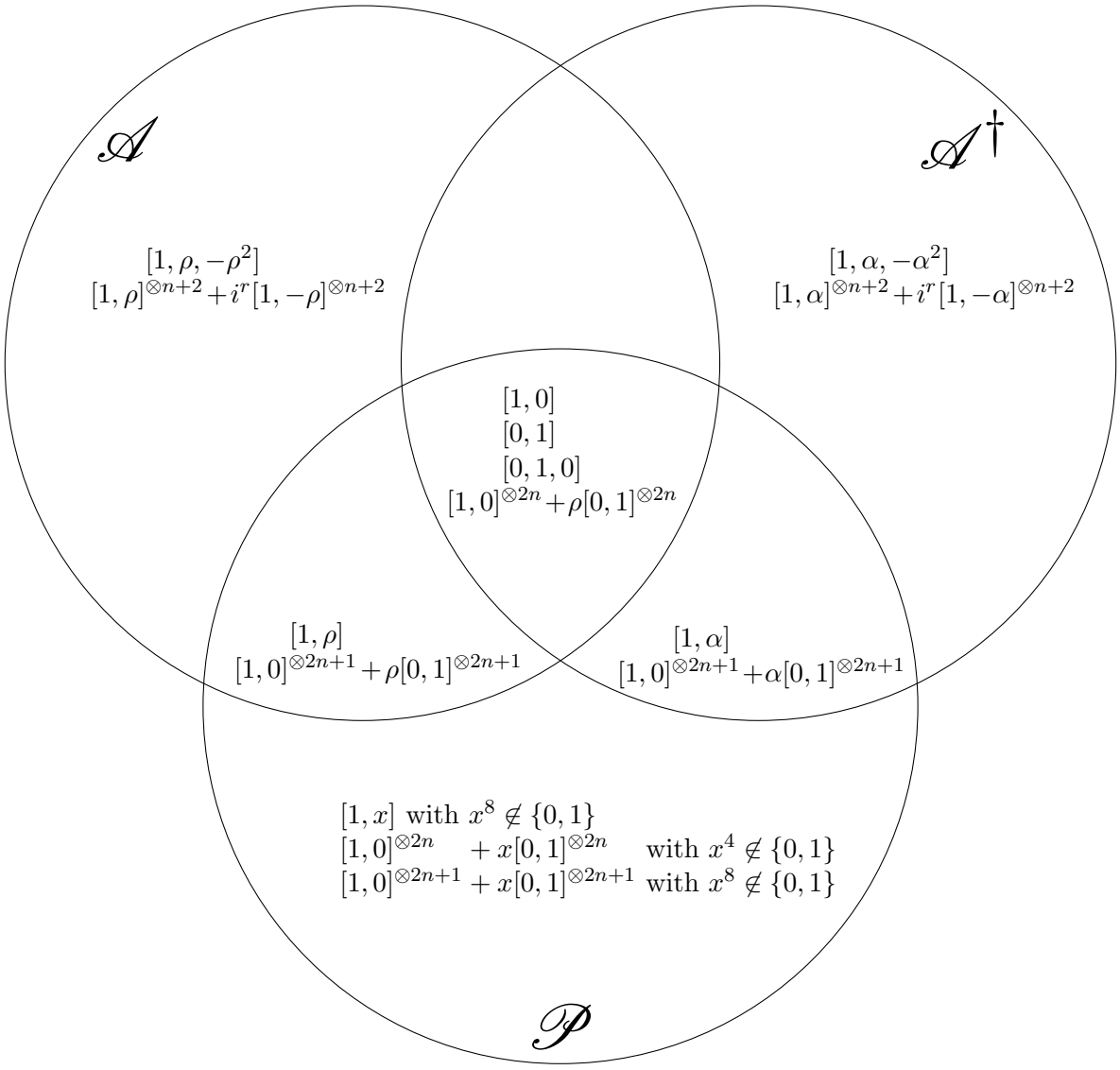


Figure 35: A Venn diagram of the #CSP^2 tractable sets $\mathscr{A}$, $\mathscr{A}^\dagger$, and $\mathscr{P}$. Note that $\rho^4 = 1$, $\alpha^4 = -1$, and $n \geq 1$. Excluded are tensor products of unary signatures.

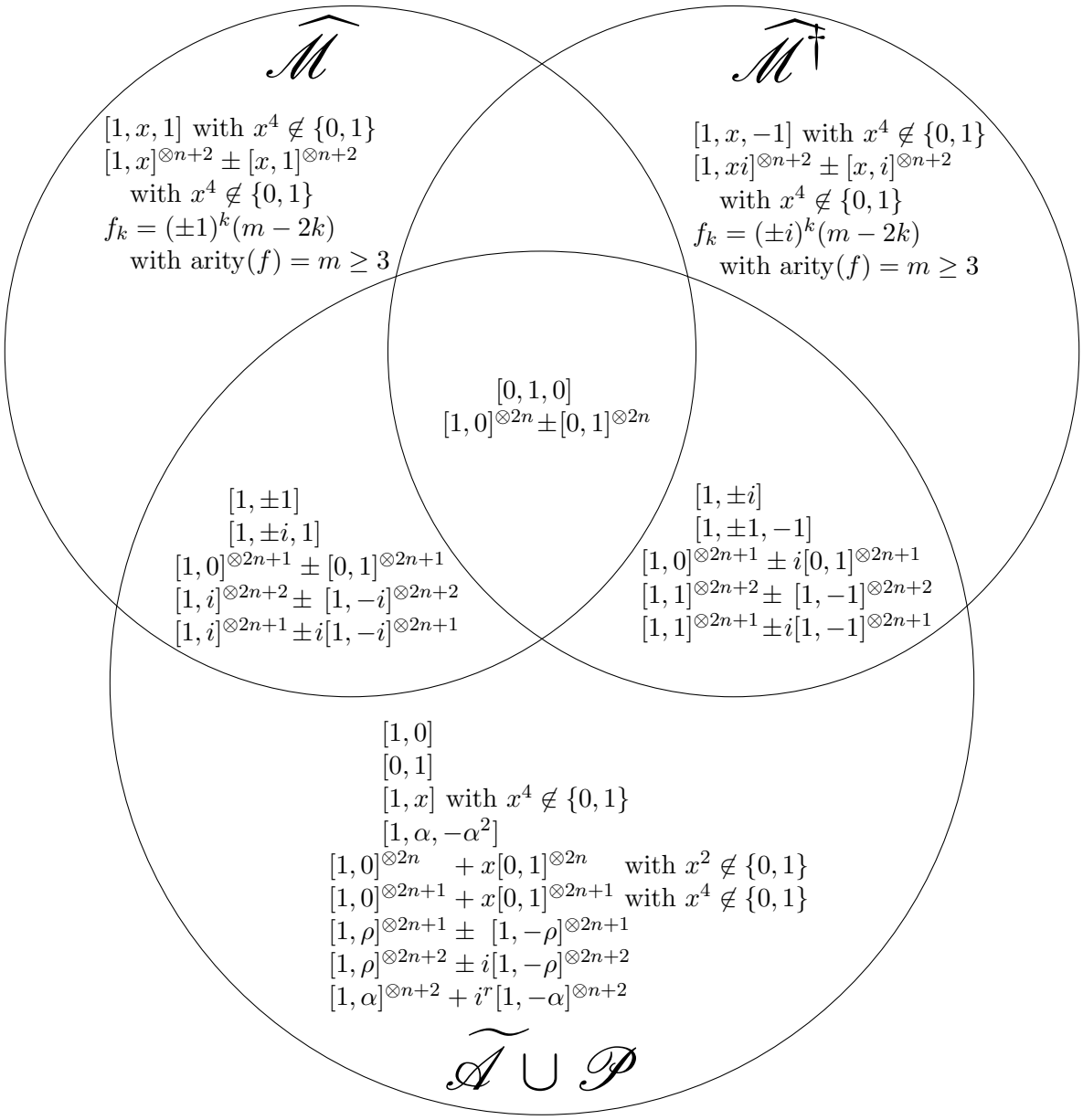


Figure 36: A Venn diagram of the Pl-#CSP^2 tractable sets $\widehat{\mathscr{M}}$ and $\widehat{\mathscr{M}^{\dagger}}$ along with the set $\widetilde{\mathscr{A} \cup \mathscr{P}}$ of all tractable #CSP^2 signatures. Note that $\rho^4 = 1$, $\alpha^4 = -1$, and $n \geq 1$. Excluded are tensor products of unary signatures.

Paper 4

2014 IEEE Annual Symposium on Foundations of Computer Science

The Complexity of Counting Edge Colorings and a Dichotomy for Some Higher Domain Holant Problems (Extended Abstract)

Jin-Yi Cai, Heng Guo, Tyson Williams

Computer Science Department
University of Wisconsin–Madison
Madison, WI USA
{jyc,hguo,tdw}@cs.wisc.edu

***Abstract*—We show that an effective version of Siegel's Theorem on finiteness of integer solutions for a specific algebraic curve and an application of elementary Galois theory are key ingredients in a complexity classification of some Holant problems. These Holant problems, denoted by $\mathrm{Holant}(f)$, are defined by a symmetric ternary function f that is invariant under any permutation of the $\kappa \geq 3$ domain elements. We prove that $\mathrm{Holant}(f)$ exhibits a complexity dichotomy. The hardness, and thus the dichotomy, holds even when restricted to planar graphs. A special case of this result is that counting edge κ-colorings is #P-hard over planar 3-regular multigraphs for all $\kappa \geq 3$. In fact, we prove that counting edge κ-colorings is #P-hard over planar r-regular multigraphs for all $\kappa \geq r \geq 3$. The problem is polynomial-time computable in all other parameter settings. The proof of the dichotomy theorem for $\mathrm{Holant}(f)$ depends on the fact that a specific polynomial $p(x, y)$ has an explicitly listed finite set of integer solutions, and the determination of the Galois groups of some specific polynomials. In the process, we also encounter the Tutte polynomial, medial graphs, Eulerian partitions, Puiseux series, and a certain lattice condition on the (logarithm of) the roots of polynomials.**

***Keywords*-counting problems; dichotomy theorem; Holant problems; edge coloring;**

I. Introduction

What do Siegel's Theorem and Galois theory have to do with complexity theory? In this paper, we show that an effective version of Siegel's Theorem on finiteness of integer solutions for a specific algebraic curve and an application of elementary Galois theory are key ingredients in a chain of steps that lead to a complexity classification of some counting problems. More specifically, we consider a certain class of counting problems that are expressible as Holant problems with an arbitrary domain of size κ over 3-regular multigraphs (i.e. self-loops and parallel edges are allowed), and prove a dichotomy theorem for this class of problems. The hardness, and thus the dichotomy, holds even when restricted to planar multigraphs. Among other things, the proof of the dichotomy theorem depends on the following: (A) the specific polynomial

$$p(x, y) = x^5 - 2x^3y - x^2y^2 - x^3 + xy^2 + y^3 - 2x^2 - xy$$

has only the integer solutions

$$(x, y) = (-1, 1), (0, 0), (1, -1), (1, 2), (3, 3),$$

and (B) the determination of the Galois groups of some specific polynomials. In the process, we also encounter the Tutte polynomial, medial graphs, Eulerian partitions, Puiseux series, and a certain lattice condition on the (logarithm of) the roots of polynomials such as $p(x, y)$.

A special case of this dichotomy theorem is the problem of counting edge colorings over planar 3-regular multigraphs using κ colors. In this case, the corresponding constraint function is the $\text{ALL-DISTINCT}_{3,\kappa}$ function, which takes value 1 when all three inputs from $[\kappa]$ are distinct and 0 otherwise. We further prove that the problem using κ colors over r-regular multigraphs is #P-hard for all $\kappa \geq r \geq 3$, even when restricted to planar multigraphs. The problem is polynomial-time computable in all other parameter settings. This solves a long-standing open problem.

We give a brief description of the framework of Holant problems [20], [18], [15], [17]. The problem $\mathrm{Holant}(\mathcal{F})$, defined by a set of functions $\mathcal{F}$, takes as input a *signature grid* $\Omega = (G, \pi)$, where $G = (V, E)$ is a multigraph, π assigns each $v \in V$ a function $f_v \in \mathcal{F}$, and f_v maps $[\kappa]^{\deg(v)}$ to $\mathbb{C}$ for some integer $\kappa \geq 2$. An edge κ-labeling $\sigma : E \to [\kappa]$ gives an evaluation $\prod_{v \in V} f_v(\sigma \mid_{E(v)})$, where $E(v)$ denotes the incident edges of v and $\sigma \mid_{E(v)}$ denotes the restriction of σ to $E(v)$. The counting problem on the instance Ω is to compute

$$\mathrm{Holant}(\Omega, \mathcal{F}) = \sum_{\sigma : E \to [\kappa]} \prod_{v \in V} f_v \left(\sigma \mid_{E(v)}\right).$$

0272-5428/14 $31.00 © 2014 IEEE
DOI 10.1109/FOCS.2014.70

Counting edge κ-colorings over r-regular multigraphs amounts to setting $f_v = \text{ALL-DISTINCT}_{r,\kappa}$ for all v. We also use $\text{Pl-Holant}(\mathcal{F})$ to denote the restriction of $\text{Holant}(\mathcal{F})$ to planar multigraphs.

Holant problems appear in many areas under a variety of different names. They are equivalent to counting Constraint Satisfaction Problems (#CSP) [5], [7] with the restriction that all variables are read twice,[1] to the contraction of a tensor network [21], [31], and to the partition function of graphical models in Forney normal form [32], [35] from artificial intelligence, coding theory, and signal processing. Special cases of Holant problems include simulating quantum circuits [42], [36], counting graph homomorphisms [1], [23], [3], [28], [9], and evaluating the partition function of the edge-coloring model [1, Section 3.6].

An edge κ-coloring of a graph G is an edge κ-labeling of G such that any two incident edges have different colors. A fundamental problem in graph theory is to determine how many colors are required to edge color G. The obvious lower bound is $\Delta(G)$, the maximum degree of the graph. By Vizing's Theorem [44], an edge coloring using just $\Delta(G) + 1$ colors always exists for simple graphs (i.e. graphs without self-loops or parallel edges). Whether $\Delta(G)$ colors suffice depends on the graph G.

Consider the edge coloring problem over 3-regular graphs. It follows from the parity condition (Lemma IV.4) that any graph containing a bridge does not have an edge 3-coloring. For bridgeless planar simple graphs, Tait [41] showed that the existence of an edge 3-coloring is equivalent to the Four-Color Theorem. Thus, the answer for the decision problem over planar 3-regular simple graphs is that there is an edge 3-coloring iff the graph is bridgeless.

Without the planarity restriction, determining if a 3-regular (simple) graph has an edge 3-coloring is NP-complete [30]. This hardness extends to finding an edge κ-coloring over κ-regular (simple) graphs for all $\kappa \geq 3$ [33]. However, these reductions are not parsimonious, and, in fact, it is claimed that no parsimonious reduction exists unless $\text{P} = \text{NP}$ [46, p. 118]. The counting complexity of this problem has remained open.

We prove that counting edge colorings over planar regular multigraphs is $\#\text{P}$-hard.[2]

[1]Without this restriction, #CSPs are a special case of Holant problems.

[2]Vizing's Theorem is for simple graphs. In Holant problems as well as counting complexity such as graph homomorphism or #CSP, one typically considers multigraphs (i.e. self-loops and parallel edges are allowed). However, our hardness result for counting edge 3-colorings over planar 3-regular multigraphs also holds for simple graphs. See Theorem 4.9 in [11].

Theorem I.1. *#κ-EDGECOLORING is $\#\text{P}$-hard over planar r-regular multigraphs for all $\kappa \geq r \geq 3$.*

See Theorem IV.8 for the proof when $\kappa = r$. Theorem 4.20 in [11] considers $\kappa > r$.

The techniques we develop to prove Theorem I.1 naturally extend to a class of Holant problems with domain size $\kappa \geq 3$ over planar 3-regular multigraphs. Functions such as $\text{ALL-DISTINCT}_{3,\kappa}$ are symmetric, which means that they are invariant under any permutation of its three inputs. But $\text{ALL-DISTINCT}_{3,\kappa}$ has another invariance—it is invariant under any permutation of the κ domain elements. We call the second property *domain invariance*.

A ternary function that is both symmetric and domain invariant is specified by three values, which we denote by $\langle a, b, c\rangle$. The output is a when all inputs are the same, c when all inputs are distinct, and b when two inputs are the same but the third input is different.

We prove a dichotomy theorem for such functions with complex weights.

Theorem I.2. *Suppose $\kappa \geq 3$ is the domain size and $a, b, c \in \mathbb{C}$. Then either $\text{Holant}(\langle a, b, c\rangle)$ is computable in polynomial time or $\text{Pl-Holant}(\langle a, b, c\rangle)$ is $\#\text{P}$-hard. Furthermore, given a, b, c, there is a polynomial-time algorithm that decides which is the case.*

See Theorem 10.1 in [11] for an explicit listing of the tractable cases. Note that counting edge κ-colorings over 3-regular multigraphs is the special case when $\langle a, b, c\rangle = \langle 0, 0, 1\rangle$.

There is only one previous dichotomy theorem for higher domain Holant problems [19]. The important difference is that the present work is for general domain size $\kappa \geq 3$ while the previous result is for domain size $\kappa = 3$. When restricted to domain size 3, the result in [19] assumes that all unary functions are available, while this dichotomy does not assume that; however it does assume domain invariance. Dichotomy theorems for an arbitrary domain size are generally difficult to prove. The Feder-Vardi Conjecture for decision Constraint Satisfaction Problems (CSP) is still open [27]. It was a major achievement to prove this conjecture for domain size 3 [4]. The #CSP dichotomy was proved after a long series of work [6], [5], [3], [22], [2], [15], [8], [12], [24], [29], [13], [7].

Our proof of Theorem I.2 has many components, and a number of new ideas are introduced in this proof. We discuss some of these ideas and give an outline of our proof in Section II.

II. Proof Outline and Techniques

As usual, the difficult part of a dichotomy theorem is to carve out *exactly* the tractable problems in the class, and prove all the rest #P-hard. A dichotomy theorem for Holant problems has the additional difficulty that some tractable problems are only shown to be tractable under a holographic transformation, which can make the appearance of the problem rather unexpected. For example, we show [11] that $\mathrm{Holant}(\langle -3-4i, 1, -1+2i\rangle)$ on domain size 4 is tractable. Despite its appearance, this problem is intimately connected with a tractable graph homomorphism problem defined by the Hadamard matrix $\begin{bmatrix} 1 & -1 & -1 & -1 \\ -1 & 1 & -1 & -1 \\ -1 & -1 & 1 & -1 \\ -1 & -1 & -1 & 1 \end{bmatrix}$. In order to understand all problems in a Holant problem class, we must deal with such problems. Dichotomy theorems for graph homomorphisms and for #CSP do not have to deal with as varied a class of such problems, since they implicitly assume all EQUALITY functions are available and must be preserved. This restricts the possible transformations.

After isolating a set of tractable problems, our #P-hardness results in both Theorem I.1 and Theorem I.2 are obtained by reducing from evaluations of the Tutte polynomial over planar graphs. A dichotomy is known for such problems (Theorem IV.1).

The chromatic polynomial, a specialization of the Tutte polynomial, is concerned with vertex colorings. On domain size κ, one starting point of our hardness proofs is the chromatic polynomial, which contains the problem of counting vertex colorings using at most κ colors. By the planar dichotomy for the Tutte polynomial, this problem is #P-hard for all $\kappa \geq 3$.

Another starting point for our hardness reductions is the evaluation of the Tutte polynomial at an integer diagonal point (x, x), which is #P-hard for all $x \geq 3$ by the same planar Tutte dichotomy. These are new starting places for reductions involving Holant problems. These problems were known to have a so-called state-sum expression (Lemma IV.3), which is a sum over weighted Eulerian partitions. This sum is not over the original planar graph but over its directed medial graph, which is always a planar 4-regular graph (Figure 1). We show that this state-sum expression is naturally expressed as a Holant problem with a particular quaternary constraint function (Lemma IV.6).

To reduce from these two problems, we execute the following strategy. First, we attempt to construct the unary constraint function $\langle 1\rangle$, which takes value 1 on all κ inputs. Second, we attempt to interpolate all succinct binary signatures assuming that we have $\langle 1\rangle$. (See Section III for the definition of a succinct signature.) Lastly, we attempt to construct a ternary signature with a special property assuming that all these binary signatures are available. At each step, there are some problems specified by certain signatures $\langle a, b, c\rangle$ for which our attempts fail. In such cases, we directly obtain a dichotomy without the help of additional signatures.

Below we highlight some of our proof techniques.

Interpolation within an orthogonal subspace: We develop the ability to interpolate when faced with some nontrivial null spaces inherently present in interpolation constructions. In any construction involving an initial signature and a recurrence matrix, it is possible that the initial signature is orthogonal to some row eigenvectors of the recurrence matrix. Previous interpolation results always attempt to find a construction that avoids this. In the present work, this avoidance seems impossible. We prove an interpolation result that can succeed in this situation to the greatest extent possible. We prove that one can interpolate any signature provided that it is orthogonal to the same set of row eigenvectors, and the relevant eigenvalues satisfy a lattice condition.

Satisfy lattice condition via Galois theory: A key requirement for this interpolation to succeed is the lattice condition (Definition V.1), which involves the roots of the characteristic polynomial of the recurrence matrix. We use Galois theory to prove that our constructions satisfy this condition. If a polynomial has a large Galois group, such as S_n or A_n, and its roots do not all have the same complex norm, then we show that its roots satisfy the lattice condition.

Effective Siegel's Theorem via Puiseux series: We need to determine the Galois groups for an infinite family of polynomials, one for each domain size. If these polynomials are irreducible, then we can show they all have the full symmetric group as their Galois group, and hence fulfill the lattice condition. We suspect that these polynomials are all irreducible but are unable to prove it.

A necessary condition for irreducibility is the absence of any linear factor. This infinite family of polynomials, as a single bivariate polynomial in (x, κ), defines an algebraic curve, which has genus 3. By a well-known theorem of Siegel [39], there are only a finite number of integer values of κ for which the corresponding polynomial has a linear factor. However this theorem and others like it are not *effective* in general. There are some effective versions of Siegel's Theorem that can be applied to the algebraic curve, but the best general effective bound is over $10^{20,000}$ [45] and hence cannot be checked in practice. Instead, we use Puiseux series in

Section V to show that this algebraic curve has exactly five explicitly listed integer solutions.

Eigenvalue Shifted Triples: For a pair of eigenvalues, the lattice condition is equivalent to the statement that the ratio of these eigenvalues is not a root of unity. A sufficient condition is that the eigenvalues have distinct complex norms. We prove three results, each of which is a different way to satisfy this sufficient condition. Chief among them is the technique we call an *Eigenvalue Shifted Triple* (EST). In an EST, we have three recurrence matrices, each of which differs from the other two by a nonzero additive multiple of the identity matrix. Provided these two multiples are linearly independent over $\mathbb{R}$, we show at least one of these matrices has eigenvalues with distinct complex norms. (However determining which one succeeds is a difficult task; but we need not know that).

E Pluribus Unum: When the ratio of a pair of eigenvalues is a root of unity, it is a challenge to effectively use this failure condition. Direct application of this cyclotomic condition is often of limited use. We introduce an approach that uses this cyclotomic condition effectively. A direct recursive construction involving these two eigenvalues only creates a finite number of different signatures. We reuse all of these signatures in a multitude of new interpolation constructions, one of which we hope will succeed. If the eigenvalues in all of these constructions also satisfy a cyclotomic condition, then we obtain a more useful condition than any of the previous cyclotomic conditions. This idea generalizes the anti-gadget technique [14], which only reuses the "last" of these signatures.

Local holographic transformation: One reason to obtain all succinct binary signatures is for use in the gadget construction known as a local holographic transformation. This construction mimics the effect of a holographic transformation applied on a single signature. In particular, using this construction, we attempt to obtain a succinct ternary signature of the form $\langle a, b, b\rangle$, where $a \neq b$. This signature turns out to have some magical properties in the Bobby Fischer gadget, which we discuss next.

Bobby Fischer gadget: Typically, any combinatorial construction for higher domain Holant problems produces very intimidating looking expressions that are nearly impossible to analyze. In our case, it seems necessary to consider a construction that has to satisfy multiple requirements involving at least nine polynomials. However, we are able to combine the signature $\langle a, b, b\rangle$, where $a \neq b$, with a succinct binary signature of our choice in a special construction that we call the *Bobby Fischer gadget*. This gadget is able to satisfy seven conditions using just one degree of freedom. This ability to satisfy a multitude of constraints simultaneously in one magic stroke reminds us of some unfathomably brilliant moves by Bobby Fischer, the chess genius extraordinaire.

III. Preliminaries

In this paper, we investigate some complex-weighted Holant problems on domain size $\kappa \geq 3$. A constraint function, or *signature*, of arity n, maps from $[\kappa]^n \to \mathbb{C}$. For consideration of models of computation, functions take complex algebraic numbers.

Graphs (called multigraphs in Section I) may have self-loops and parallel edges. A graph without self-loops or parallel edges is a *simple* graph. A *signature grid* $\Omega = (G, \pi)$ of $\mathrm{Holant}(\mathcal{F})$ consists of a graph $G = (V, E)$, where π assigns each vertex $v \in V$ and its incident edges with some $f_v \in \mathcal{F}$ and its input variables. We say Ω is a *planar signature grid* if G is planar, where the variables of f_v are ordered counterclockwise. The Holant problem on instance Ω is to evaluate

$$\mathrm{Holant}(\Omega; \mathcal{F}) = \sum_{\sigma} \prod_{v \in V} f_v(\sigma \mid_{E(v)}),$$

a sum over all edge labelings $\sigma : E \to [\kappa]$, where $E(v)$ denotes the incident edges of v and $\sigma \mid_{E(v)}$ denotes the restriction of σ to $E(v)$.

A function f_v can be represented by listing its values in lexicographical order as in a truth table, which is a vector in $\mathbb{C}^{\kappa^{\deg(v)}}$, or as a tensor in $(\mathbb{C}^{\kappa})^{\otimes \deg(v)}$. In this paper, we consider symmetric signatures. An example of a symmetric signature is the Equality signature $=_r$ of arity r. A Holant problem is parametrized by a set of signatures.

Definition III.1. *Given a set of signatures $\mathcal{F}$, we define the counting problem* $\mathrm{Holant}(\mathcal{F})$ *as:*
Input: A signature grid $\Omega = (G, \pi)$*;*
Output: $\mathrm{Holant}(\Omega; \mathcal{F})$.

The problem $\text{Pl-Holant}(\mathcal{F})$ is defined similarly using a planar signature grid. Replacing a signature $f \in \mathcal{F}$ by a constant multiple cf, where $c \neq 0$, does not change the complexity of $\mathrm{Holant}(\mathcal{F})$. It introduces a global nonzero factor to $\mathrm{Holant}(\Omega; \mathcal{F})$. We follow the usual conventions about polynomial time Turing reduction $\leq_T$.

We say a signature f is *realizable* or *constructible* from a signature set $\mathcal{F}$ if there is a gadget with some dangling edges such that each vertex is assigned a signature from $\mathcal{F}$, and the resulting graph, when viewed

as a black-box signature with inputs on the dangling edges, is exactly f. If f is realizable from a set $\mathcal{F}$, then we can freely add f into $\mathcal{F}$ while preserving the complexity.

Formally, such a notion is defined by an $\mathcal{F}$-gate [15], [16]. An $\mathcal{F}$-gate is similar to a signature grid (G, π) for $\text{Holant}(\mathcal{F})$ except that $G = (V, E, D)$ is a graph with some dangling edges D. The dangling edges define external variables for the $\mathcal{F}$-gate. (See Figure 3 for an example.) We denote the regular edges in E by $1, 2, \ldots, m$ and the dangling edges in D by $m+1, \ldots, m+n$. Then we can define a function Γ for this $\mathcal{F}$-gate as

$$\Gamma(y_1, y_2, \ldots, y_n) = \sum_{x_1, x_2, \ldots, x_m \in [\kappa]} H(x_1, x_2, \ldots, x_m, y_1, y_2, \ldots, y_n),$$

where $(y_1, \ldots, y_n) \in [\kappa]^n$ denotes a labeling on the dangling edges and $H(x_1, \ldots, x_m, y_1, \ldots, y_n)$ denotes the value of the signature grid on a labeling of all edges in G, which is the product of evaluations at all vertices. We also call this function Γ the signature of the $\mathcal{F}$-gate. An $\mathcal{F}$-gate is planar if the underlying graph G is a planar graph, and the dangling edges, ordered counterclockwise corresponding to the order of the input variables, are in the outer face in a planar embedding. A planar $\mathcal{F}$-gate can be used in a planar signature grid as if it is just a single vertex with the particular signature.

Using the idea of planar $\mathcal{F}$-gates, we can reduce one planar Holant problem to another. Suppose g is the signature of some planar $\mathcal{F}$-gate. Then we obtain $\text{Pl-Holant}(\mathcal{F} \cup \{g\}) \leq_T \text{Pl-Holant}(\mathcal{F})$, by replacing every appearance of g by the $\mathcal{F}$-gate. Since the signature of the $\mathcal{F}$-gate is g, the Holant values for these two signature grids are identical.

Our main results are about symmetric signatures (i.e. signatures that are invariant under any permutation of inputs). However, we also need some asymmetric signatures in our proofs. When a gadget has an asymmetric signature, we place a diamond on the edge corresponding to the first input. The remaining inputs are ordered counterclockwise around the vertex. (See Figure 3 for an example.)

An arity r signature on domain size κ is fully specified by κ^r values. However, some special cases can be defined using far fewer values. Consider the signature $\text{ALL-DISTINCT}_{r,\kappa}$ of arity r on domain size κ that outputs 1 when all inputs are distinct and 0 otherwise. We also denote this signature by $\text{AD}_{r,\kappa}$. In addition to being symmetric, it is also invariant under any permutation of the κ domain elements. We call the second property *domain invariance*. The signature of an $\mathcal{F}$-gate in which all signatures in $\mathcal{F}$ are domain invariant is itself domain invariant.

Definition III.2 (Succinct signature)**.** *Let $\tau = (P_1, P_2, \ldots, P_\ell)$ be a partition of $[\kappa]^r$ listed in some order. We say that f is a* succinct signature *of type τ if f is constant on each P_i. A set $\mathcal{F}$ of signatures is of type τ if every $f \in \mathcal{F}$ has type τ. We denote a succinct signature f of type τ by $\langle f(P_1), \ldots, f(P_\ell)\rangle$, where $f(P) = f(x)$ for any $x \in P$.*

Furthermore, we may omit 0 entries. If f is a succinct signature of type τ, we also say f is a succinct signature *of type τ' with length ℓ', where τ' lists ℓ' parts of the partition τ and we write f as $\langle f_1, f_2, \ldots, f_{\ell'}\rangle$, provided all nonzero values $f(P_i)$ are listed. When using this notation, we will make it clear which zero entries have been omitted.*

For example, a symmetric signature in the Boolean domain (i.e. $\kappa = 2$) has been denoted in previous work [10] by $[f_0, f_1, \ldots, f_r]$, where f_w is the output on inputs of Hamming weight w. This corresponds to the succinct signature type $(P_0, P_1, \ldots, P_r)$, where P_w is the set of inputs of Hamming weight w.

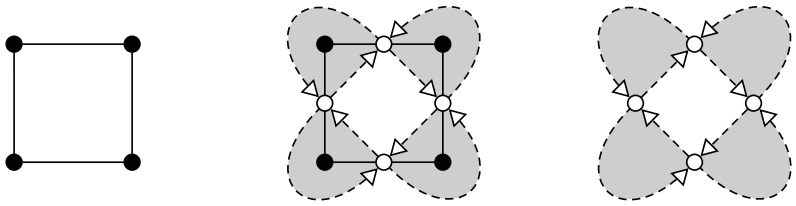
We prove a dichotomy theorem for $\text{Pl-Holant}(f)$ when f is a succinct ternary signature of type τ_3 on domain size $\kappa \geq 3$. For $\kappa \geq 3$, the succinct signature of type $\tau_3 = (P_1, P_2, P_3)$ is a partition of $[\kappa]^3$ with $P_i = \{(x, y, z) \in [\kappa]^3 : |\{x, y, z\}| = i\}$ for $1 \leq i \leq 3$. The notation $\{x, y, z\}$ denotes a multiset and $|\{x, y, z\}|$ denotes the number of distinct elements in it. Succinct signatures of type τ_3 are exactly the symmetric and domain invariant ternary signatures. In particular, the succinct ternary signature for $\text{AD}_{3,\kappa}$ is $\langle 0, 0, 1\rangle$.

IV. Complexity of Counting Edge Colorings

Here we prove that counting edge κ-colorings over planar r-regular graphs is #P-hard provided $\kappa = r \geq 3$. For the proof when $\kappa > r \geq 3$, see Theorem 4.20 in [11]. We reduce from evaluating the Tutte polynomial of a planar graph at the positive integer points on the diagonal $x = y$. For $x \geq 3$, evaluating the Tutte polynomial of a planar graph at (x, x) is #P-hard.

Theorem IV.1 (Theorem 5.1 in [43])**.** *For $x, y \in \mathbb{C}$, evaluating the Tutte polynomial at (x, y) is #P-hard over planar graphs unless $(x-1)(y-1) \in \{1, 2\}$ or $(x, y) \in \{(1, 1), (-1, -1), (\omega, \omega^2), (\omega^2, \omega)\}$, where $\omega = e^{2\pi i/3}$. In each exceptional case, the computation can be done in polynomial time.*

To state the connection with the diagonal of the Tutte polynomial, we need to consider Eulerian subgraphs in directed medial graphs. We say a graph is an Eulerian



(a) (b) (c)

Figure 1: A plane graph ((a)), its directed medial graph ((c)), and both superimposed ((b)).

(di)graph if every vertex has even degree (resp. in-degree equal to out-degree), but connectedness is not required. Now recall the definition of a medial graph and its directed variant.

Definition IV.2 (cf. Section 4 in [25])**.** *For a connected plane graph G (i.e. a planar embedding of a connected planar graph), its* medial graph G_m *has a vertex on each edge of G and two vertices in G_m are joined by an edge for each face of G in which their corresponding edges occur consecutively.*

The directed medial graph $\vec{G}_m$ *of G colors the faces of G_m black or white depending on whether they contain or do not contain, respectively, a vertex of G. Then the edges of the medial graph are directed so that the black face is on the left.*

See Figure 1 for an example. Notice that the directed medial graph is always a planar 4-regular graph. Now we can give the connection with the diagonal of the Tutte polynomial. A monochromatic vertex is a vertex with all its incident edges having the same color.

Lemma IV.3 (Equation (17) in [25])**.** *Suppose G is a connected plane graph and $\vec{G}_m$ is its directed medial graph. For $\kappa \in \mathbb{N}$, let $\mathcal{C}(\vec{G}_m)$ be the set of all edge κ-labelings of $\vec{G}_m$ so that each (possibly empty) set of monochromatic edges forms an Eulerian digraph. Then*

$$\kappa\,\mathrm{T}(G;\kappa+1,\kappa+1) = \sum_{c\,\in\,\mathcal{C}(\vec{G}_m)} 2^{m(c)}, \qquad (1)$$

where $m(c)$ is the number of monochromatic vertices in the coloring c.

The Eulerian partitions in $\mathcal{C}(\vec{G}_m)$ have the property that the subgraphs induced by each partition do not intersect (or crossover) each other due to the orientation of the edges in the medial graph. We call the counting problem defined by the sum on the right-hand side of (1) as counting weighted Eulerian partitions over planar 4-regular graphs. This problem also has an expression as a Holant problem using a succinct signature. To define this succinct signature, it helps to know the following basic result about edge colorings.

When the number of available colors coincides with the regularity parameter of the graph, the cuts in any coloring satisfy a well-known parity condition. The parity condition we state here follows from a more general parity argument (see (1.2) and the Parity Argument on page 95 in [40]).

Lemma IV.4 (Parity Condition)**.** *Let G be a κ-regular graph and consider a cut C in G. For any edge κ-coloring of G, $c_1 \equiv c_2 \equiv \cdots \equiv c_\kappa \pmod 2$, where c_i is the number of edges in C colored i.*

Consider all quaternary $\{\mathrm{AD}_{\kappa,\kappa}\}$-gates on domain size $\kappa \geq 3$. These gadgets have a succinct signature of type $\tau_{\text{color}} = (P_{\substack{1\,1\\1\,1}}, P_{\substack{1\,2\\1\,2}}, P_{\substack{1\,2\\2\,1}}, P_{\substack{1\,1\\2\,2}}, P_{\substack{1\,4\\2\,3}}, P_0)$, where

$$\begin{aligned}
P_{\substack{1\,1\\1\,1}} &= \{(w,x,y,z) \in [\kappa]^4 \mid w = x = y = z\},\\
P_{\substack{1\,2\\1\,2}} &= \{(w,x,y,z) \in [\kappa]^4 \mid w = x \neq y = z\},\\
P_{\substack{1\,2\\2\,1}} &= \{(w,x,y,z) \in [\kappa]^4 \mid w = y \neq x = z\},\\
P_{\substack{1\,1\\2\,2}} &= \{(w,x,y,z) \in [\kappa]^4 \mid w = z \neq x = y\},\\
P_{\substack{1\,4\\2\,3}} &= \{(w,x,y,z) \in [\kappa]^4 \mid w,x,y,z \text{ are distinct}\},\\
P_0 &= [\kappa]^4 - P_{\substack{1\,1\\1\,1}} - P_{\substack{1\,2\\1\,2}} - P_{\substack{1\,2\\2\,1}} - P_{\substack{1\,1\\2\,2}} - P_{\substack{1\,4\\2\,3}}.
\end{aligned}$$

Any quaternary signature of an $\{\mathrm{AD}_{\kappa,\kappa}\}$-gate is constant on the first five entries of τ_{color} since $\mathrm{AD}_{\kappa,\kappa}$ is domain invariant. Using Lemma IV.4, we can show that the entry corresponding to P_0 is 0.

Lemma IV.5. *Suppose $\kappa \geq 3$ is the domain size and let F be a quaternary $\{\mathrm{AD}_{\kappa,\kappa}\}$-gate with succinct signature f of type τ_{color}. Then $f(P_0) = 0$.*

By Lemma IV.5, we denote a quaternary signature f of an $\{\mathrm{AD}_{\kappa,\kappa}\}$-gate by the succinct signature $\langle f(P_{\substack{1\,1\\1\,1}}), f(P_{\substack{1\,2\\1\,2}}), f(P_{\substack{1\,2\\2\,1}}), f(P_{\substack{1\,1\\2\,2}}), f(P_{\substack{1\,4\\2\,3}})\rangle$ of type τ_{color}, which has the entry for P_0 omitted. When $\kappa = 3$, $P_{\substack{1\,4\\2\,3}}$ is empty and we define its entry in the succinct signature to be 0.

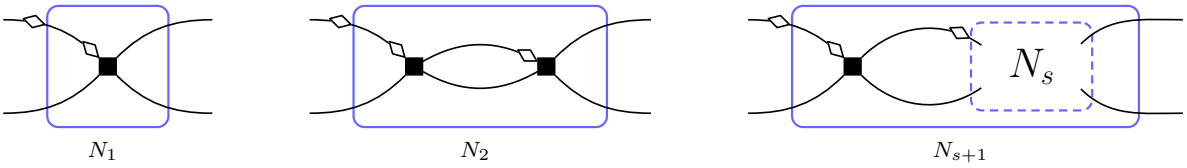


Figure 2: Recursive construction to interpolate $\langle 2,1,0,1,0\rangle$. Vertices assigned the signature of the gadget in Figure 3.

Lemma IV.6. *Let G be a connected plane graph and let G_m be the medial graph of G. Then*

$$\kappa\,\mathrm{T}(G;\kappa+1,\kappa+1) = \text{Pl-Holant}(G_m;\langle 2,1,0,1,0\rangle),$$

where the Holant problem has domain size κ and $\langle 2,1,0,1,0\rangle$ is a succinct signature of type τ_{color}.

Proof: Let $f = \langle 2,1,0,1,0\rangle$. By Lemma IV.3, we only need to prove that

$$\sum_{c\in\mathcal{C}(\vec{G}_m)} 2^{m(c)} = \text{Pl-Holant}(G_m;f), \qquad (2)$$

where the notation is from Lemma IV.3.

Each $c \in \mathcal{C}(\vec{G}_m)$ is also an edge κ-labeling of G_m. At each vertex $v \in V(\vec{G}_m)$, the four incident edges are assigned at most two distinct colors by c. If all four edges are assigned the same color, then the constraint f on v contributes a factor of 2 to the total weight. This is given by the value in the first entry of f. Otherwise, there are two different colors, say x and y. Because the orientation at v in $\vec{G}_m$ is cyclically "in, out, in, out", the coloring around v can only be of the form $xxyy$ or $xyyx$. These correspond to the second and fourth entries of f. Therefore, every term in the summation on the left-hand side of (2) appears (with the same nonzero weight) in the summation $\text{Pl-Holant}(G_m;f)$.

It is also easy to see that every nonzero term in $\text{Pl-Holant}(G_m;f)$ appears in the sum on the left-hand side of (2) with the same weight of 2 to the power of the number of monochromatic vertices. In particular, any coloring with a vertex that is cyclically colored $xyxy$ for different colors x and y does not contribute because $f(P^{1\,2}_{2\,1}) = 0$. ∎

By Theorem IV.1 and Lemma IV.6, the problem $\text{Pl-Holant}(\langle 2,1,0,1,0\rangle)$ is #P-hard.

Corollary IV.7. *Suppose $\kappa \geq 3$ is the domain size. Let $\langle 2,1,0,1,0\rangle$ be a succinct quaternary signature of type τ_{color}. Then $\text{Pl-Holant}(\langle 2,1,0,1,0\rangle)$ is #P-hard.*

With this connection established, we can now show that counting edge colorings is #P-hard over planar regular graphs when the number of colors and the regularity parameter coincide.

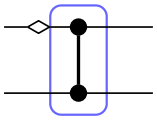


Figure 3: Quaternary gadget f used in the interpolation construction below. All vertices are assigned $\mathrm{AD}_{\kappa,\kappa}$. The bold edge represents $\kappa-2$ parallel edges.

Theorem IV.8. *#κ-EdgeColoring is #P-hard over planar κ-regular graphs for all $\kappa \geq 3$.*

Proof: Let κ be the domain size of all Holant problems in this proof and let $\langle 2,1,0,1,0\rangle$ be a succinct quaternary signature of type τ_{color}. We reduce from $\text{Pl-Holant}(\langle 2,1,0,1,0\rangle)$ to $\text{Pl-Holant}(\mathrm{AD}_{\kappa,\kappa})$, which denotes the problem of counting edge κ-colorings in planar κ-regular graphs as a Holant problem. Then by Corollary IV.7, we conclude that $\text{Pl-Holant}(\mathrm{AD}_{\kappa,\kappa})$ is #P-hard.

Consider the gadget in Figure 3, where the bold edge represents $\kappa-2$ parallel edges. We assign $\mathrm{AD}_{\kappa,\kappa}$ to both vertices. Up to a nonzero factor of $(\kappa-2)!$, this gadget has the succinct quaternary signature $f = \langle 0,1,1,0,0\rangle$ of type τ_{color}. Consider the recursive construction in Figure 2. All vertices are assigned the signature f. Let f_s be the succinct quaternary signature of type τ_{color} for the sth gadget of the recursive construction. Then $f_1 = f$ and $f_s = M^s f_0$, where

$$M = \begin{bmatrix} 0 & \kappa-1 & 0 & 0 & 0 \\ 1 & \kappa-2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \qquad \text{and} \qquad f_0 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}.$$

The signature f_0 is simply the succinct quaternary signature of type τ_{color} for two parallel edges. We can express M via the Jordan decomposition $M = P\Lambda P^{-1}$, where

$$P = \begin{bmatrix} 1 & 1-\kappa & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

and $\Lambda = \mathrm{diag}(\kappa-1,-1,1,-1,1)$. Then for $t = 2s$, we

have

$$f_t = P \begin{bmatrix} \kappa-1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}^t P^{-1} \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}$$
$$= P \begin{bmatrix} x & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} P^{-1} \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} y+1 \\ y \\ 0 \\ 1 \\ 0 \end{bmatrix},$$

where $x = (\kappa-1)^t$ and $y = \frac{x-1}{\kappa}$.

Consider an instance Ω of Pl-Holant$(\langle 2,1,0,1,0\rangle)$ on domain size κ. Suppose $\langle 2,1,0,1,0\rangle$ appears n times in Ω. We construct from Ω a sequence of instances Ω_t of Pl-Holant(f) indexed by t, where $t = 2s$ with $s \geq 0$. We obtain Ω_t from Ω by replacing each occurrence of $\langle 2,1,0,1,0\rangle$ with the gadget f_t.

As a polynomial in $x = (\kappa-1)^t$, Pl-Holant$(\Omega_t; f)$ is independent of t and has degree at most n with integer coefficients. Using our oracle for Pl-Holant(f), we can evaluate this polynomial at $n+1$ distinct points $x = (\kappa-1)^{2s}$ for $0 \leq s \leq n$. Then via polynomial interpolation, we can recover the coefficients of this polynomial efficiently. Evaluating this polynomial at $x = \kappa+1$ (so that $y = 1$) gives the value of Pl-Holant$(\Omega; \langle 2,1,0,1,0\rangle)$, as desired. ∎

V. Dose of an effective Siegel's Theorem

We jump into the middle of our proof for Theorem I.2. Consider the polynomial $p(x,y) \in \mathbb{Z}[x,y]$ defined by

$$p(x,y) = x^5 - (2y+1)x^3 - (y^2+2)x^2 + y(y-1)x + y^3.$$

We consider $y = \kappa+1$ as an integer parameter $y \geq 4$, and treat $p(x,y)$ as an infinite family of quintic polynomials in x with integer coefficients. We want to show that the roots of all these quintic polynomials satisfy the lattice condition. (For $\kappa \in \{3,4\}$, we need alternative proofs.)

Definition V.1. *Fix some $\ell \in \mathbb{N}$. We say that $\lambda_1, \lambda_2, \ldots, \lambda_\ell \in \mathbb{C} - \{0\}$ satisfy the* lattice condition *if for all $x \in \mathbb{Z}^\ell - \{\mathbf{0}\}$ with $\sum_{i=1}^{\ell} x_i = 0$, we have*

$$\prod_{i=1}^{\ell} \lambda_i^{x_i} \neq 1.$$

We suspect that for any integer $y \geq 4$, $p(x,y)$ is in fact irreducible over $\mathbb{Q}$ as a polynomial in x. We can show that this is a sufficient condition for the roots of $p(x,y)$ to satisfy the lattice condition for any integer $y \geq 4$. When considering y as an indeterminate, the bivariate polynomial $p(x,y)$ is irreducible over $\mathbb{Q}$ and the algebraic curve defined by it has genus 3, so by Theorem 1.2 in [37], $p(x,y)$ is reducible over $\mathbb{Q}$ for at most a finite number of $y \in \mathbb{Z}$.

We know of just five values of $y \in \mathbb{Z}$ for which $p(x,y)$ is reducible as a polynomial in x:

$$p(x,y) = \begin{cases} (x-1)(x^4+x^3+2x^2-x+1) & y = -1 \\ x^2(x^3-x-2) & y = 0 \\ (x+1)(x^4-x^3-2x^2-x+1) & y = 1 \\ (x-1)(x^2-x-4)(x^2+2x+2) & y = 2 \\ (x-3)(x^4+3x^3+2x^2-5x-9) & y = 3. \end{cases}$$

These five factorizations also give five integer solutions to $p(x,y) = 0$. It is a well-known theorem of Siegel [39] that an algebraic curve of genus at least 1 has only a finite number of integral points. For this curve of genus 3, Faltings' Theorem [26] says that there can be only a finite number of rational points. However these theorems are not *effective* in general. There are some effective versions of Siegel's Theorem that can be applied to our polynomial, but the best effective bound that we can find is over $10^{20,000}$ [45] and hence cannot be checked in practice.

However, it is shown in the next lemma that in fact these five are the only integer solutions. In particular, for any integer $y \geq 4$, $p(x,y)$ does not have a linear factor in $\mathbb{Z}[x]$. The following proof is based on a key auxiliary function $g_2(x,y) = \frac{y^2}{x} + y - x^2 + 1$ due to Aaron Levin [34]. We thank Aaron and also thank Bjorn Poonen [38] who suggested a similar proof.

Lemma V.2. *The only integer solutions to $p(x,y) = 0$ are $(1,-1)$, $(0,0)$, $(-1,1)$, $(1,2)$, and $(3,3)$.*

Proof sketch: Clearly these five points are solutions to $p(x,y) = 0$. Let $(a,b) \in \mathbb{Z}^2$ be a solution to $p(x,y) = 0$ with $a \neq 0$. We claim $a \mid b^2$. By definition of $p(x,y)$, clearly $a \mid b^3$. If p is a prime that divides a, then let $\mathrm{ord}_p(a) = e$ and $\mathrm{ord}_p(b) = f$ be the exact orders with which p divides a and b respectively. Then $f \geq 1$ since $3f \geq e$ and our claim is that $2f \geq e$. Suppose for a contradiction that $2f < e$. From $p(a,b) = 0$, we have $a^2(a^3 - 2ab - a - b^2 - 2) = -b^3 - ab(b-1)$. The order with respect to p of the left-hand side is $\mathrm{ord}_p\left(a^2(a^3-2ab-a-b^2-2)\right) \geq \mathrm{ord}_p\left(a^2\right) = 2e$. Since p is relatively prime to $b-1$, $\mathrm{ord}_p\left(ab(b-1)\right) = e + f > 3f$, and therefore the order of the right-hand side with respect to p is $\mathrm{ord}_p\left(-b^3 - ab(b-1)\right) = \mathrm{ord}_p(b^3) = 3f$. However, $2e > 3f$, a contradiction. This proves the claim.

Now consider the functions $g_1(x,y) = y - x^2$ and $g_2(x,y) = \frac{y^2}{x} + y - x^2 + 1$. Whenever $(a,b) \in \mathbb{Z}^2$ is a solution to $p(x,y) = 0$ with $a \neq 0$, $g_1(a,b)$ and $g_2(a,b)$ are integers. We compute the Puiseux series expansions

$y_1(x)$ for $x \in \mathbb{R}$, $y_2(x)$ for $x > 0$, and $y_3(x)$ for $x > 0$, where

$$y_1(x) = x^2 + 2x^{-1} + 2x^{-2} - 6x^{-4} - 18x^{-5} + O(x^{-6}),$$
$$y_2(x) = x^{3/2} - \frac{1}{2}x + \frac{1}{8}x^{1/2} - \frac{65}{128}x^{-1/2} - x^{-1} - \frac{1471}{1024}x^{-3/2} - x^{-2} + O(x^{-5/2}), \quad \text{and}$$
$$y_3(x) = -x^{3/2} - \frac{1}{2}x - \frac{1}{8}x^{1/2} + \frac{65}{128}x^{-1/2} - x^{-1} + \frac{1471}{1024}x^{-3/2} - x^{-2} + O(x^{-5/2}),$$

to obtain asymptotic approximations to the roots of these polynomials with y expanded as a Puiseux series of x. These series converge to the actual roots of $p(x, y)$ for large x. The basic idea of the proof—called Runge's method—is that, for example, when we substitute $y_2(x)$ in $g_2(x, y)$, we get $g_2(x, y_2(x)) = O(x^{-1/2})$, where the multiplier in the O-notation is bounded both above and below by a nonzero constant in absolute value. Thus for large x, this cannot be an integer. However, for integer solutions (x, y) of $p(x, y)$, this must be an integer. We prove that $|x| > 16$ suffices to show this for each asymptotic approximation. For $|x| \leq 16$, one can directly check that there are no other integer solutions. ■

ACKNOWLEDGMENTS

We thank Joanna Ellis-Monaghan for bringing [25] to our attention. We are thankful to Mingji Xia who discussed with us an early version of this work. We are very grateful to Bjorn Poonen and especially Aaron Levin for sharing their expertise on Runge's method, and in particular for the auxiliary function $g_2(x, y)$ in the proof of Lemma V.2. We benefited from discussions with William Whistler on a draft of this work, whom we thank. We also thank the anonymous referees for their helpful comments. All authors were supported by NSF CCF-1217549. The second author was also supported by a Simons Award for Graduate Students in Theoretical Computer Science from the Simons Foundation. The third author was also supported by a Cisco Systems Distinguished Graduate Fellowship.

REFERENCES

[1] Christian Borgs, Jennifer Chayes, László Lovász, Vera T. Sós, and Katalin Vesztergombi. Counting graph homomorphisms. In Martin Klazar, Jan Kratochvíl, Martin Loebl, Jiří Matoušek, Pavel Valtr, and Robin Thomas, editors, *Topics in Discrete Mathematics*, volume 26 of *Algorithms and Combinatorics*, pages 315–371. Springer Berlin Heidelberg, 2006.

[2] Andrei Bulatov, Martin Dyer, Leslie Ann Goldberg, Markus Jalsenius, and David Richerby. The complexity of weighted Boolean #CSP with mixed signs. *Theor. Comput. Sci.*, 410(38-40):3949–3961, 2009.

[3] Andrei Bulatov and Martin Grohe. The complexity of partition functions. *Theor. Comput. Sci.*, 348(2):148–186, 2005.

[4] Andrei A. Bulatov. A dichotomy theorem for constraint satisfaction problems on a 3-element set. *J. ACM*, 53(1):66–120, 2006.

[5] Andrei A. Bulatov. The complexity of the counting constraint satisfaction problem. *J. ACM*, 60(5):34:1–34:41, 2013.

[6] Andrei A. Bulatov and Víctor Dalmau. Towards a dichotomy theorem for the counting constraint satisfaction problem. *Inform. and Comput.*, 205(5):651–678, 2007.

[7] Jin-Yi Cai and Xi Chen. Complexity of counting CSP with complex weights. In *STOC*, pages 909–920. ACM, 2012.

[8] Jin-Yi Cai, Xi Chen, and Pinyan Lu. Non-negatively weighted #CSP: An effective complexity dichotomy. In *IEEE Conference on Computational Complexity*, pages 45–54. IEEE Computer Society, 2011.

[9] Jin-Yi Cai, Xi Chen, and Pinyan Lu. Graph homomorphisms with complex values: A dichotomy theorem. *SIAM J. Comput.*, 42(3):924–1029, 2013.

[10] Jin-Yi Cai, Heng Guo, and Tyson Williams. A complete dichotomy rises from the capture of vanishing signatures (extended abstract). In *STOC*, pages 635–644. ACM, 2013.

[11] Jin-Yi Cai, Heng Guo, and Tyson Williams. The complexity of counting edge colorings and a dichotomy for some higher domain Holant problems. *CoRR*, abs/1404.4020, 2014.

[12] Jin-Yi Cai, Sangxia Huang, and Pinyan Lu. From Holant to #CSP and back: Dichotomy for Holant^c problems. *Algorithmica*, 64(3):511–533, 2012.

[13] Jin-Yi Cai and Michael Kowalczyk. Spin systems on k-regular graphs with complex edge functions. *Theor. Comput. Sci.*, 461:2–16, 2012.

[14] Jin-Yi Cai, Michael Kowalczyk, and Tyson Williams. Gadgets and anti-gadgets leading to a complexity dichotomy. In *ITCS*, pages 452–467. ACM, 2012.

[15] Jin-Yi Cai, Pinyan Lu, and Mingji Xia. Holant problems and counting CSP. In *STOC*, pages 715–724. ACM, 2009.

[16] Jin-Yi Cai, Pinyan Lu, and Mingji Xia. Holographic algorithms with matchgates capture precisely tractable planar #CSP. In *FOCS*, pages 427–436. IEEE Computer Society, 2010.

[17] Jin-Yi Cai, Pinyan Lu, and Mingji Xia. Computational complexity of Holant problems. *SIAM J. Comput.*, 40(4):1101–1132, 2011.

[18] Jin-Yi Cai, Pinyan Lu, and Mingji Xia. Holographic reduction, interpolation and hardness. *Computational Complexity*, 21(4):573–604, 2012.

[19] Jin-Yi Cai, Pinyan Lu, and Mingji Xia. Dichotomy for Holant* problems with domain size 3. In *SODA*, pages 1278–1295. SIAM, 2013.

[20] Jin-Yi Cai, Pinyan Lu, and Mingji Xia. Holographic algorithms by Fibonacci gates. *Linear Algebra Appl.*, 438(2):690–707, 2013.

[21] C. T. J. Dodson and T. Poston. *Tensor Geometry*, volume 130 of *Graduate Texts in Mathematics*. Springer-Verlag, second edition, 1991.

[22] Martin Dyer, Leslie Ann Goldberg, and Mark Jerrum. The complexity of weighted Boolean #CSP. *SIAM J. Comput.*, 38(5):1970–1986, 2009.

[23] Martin Dyer and Catherine Greenhill. The complexity of counting graph homomorphisms. *Random Struct. Algorithms*, 17(3-4):260–289, 2000.

[24] Martin Dyer and David Richerby. On the complexity of #CSP. In *STOC*, pages 725–734. ACM, 2010.

[25] Joanna A. Ellis-Monaghan. Identities for circuit partition polynomials, with applications to the Tutte polynomial. *Adv. Appl. Math.*, 32(1-2):188–197, 2004.

[26] Gerd Faltings. Endlichkeitssätze für abelsche Varietäten über Zahlkörpern. *Invent. Math.*, 73(3):349–366, 1983.

[27] Tomás Feder and Moshe Y. Vardi. The computational structure of monotone monadic SNP and constraint satisfaction: A study through Datalog and group theory. *SIAM J. Comput.*, 28(1):57–104, 1999.

[28] Leslie Ann Goldberg, Martin Grohe, Mark Jerrum, and Marc Thurley. A complexity dichotomy for partition functions with mixed signs. *SIAM J. Comput.*, 39(7):3336–3402, 2010.

[29] Heng Guo, Sangxia Huang, Pinyan Lu, and Mingji Xia. The complexity of weighted Boolean #CSP modulo k. In *STACS*, pages 249–260. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2011.

[30] Ian Holyer. The NP-completeness of edge-coloring. *SIAM J. Comput.*, 10(4):718–720, 1981.

[31] A. W. Joshi. *Matrices And Tensors In Physics*. New Age International, revised third edition, 1995.

[32] G. David Forney Jr. Codes on graphs: normal realizations. *Information Theory, IEEE Transactions on*, 47(2):520–548, 2001.

[33] Daniel Leven and Zvi Galil. NP completeness of finding the chromatic index of regular graphs. *Journal of Algorithms*, 4(1):35–44, 1983.

[34] Aaron Levin. private communication, 2013.

[35] Hans-Andrea Loeliger. An introduction to factor graphs. *Signal Processing Magazine, IEEE*, 21(1):28–41, 2004.

[36] Igor L. Markov and Yaoyun Shi. Simulating quantum computation by contracting tensor networks. *SIAM J. Comput.*, 38(3):963–981, 2008.

[37] Peter Müller. Hilbert's irreducibility theorem for prime degree and general polynomials. *Israel J. Math.*, 109(1):319–337, 1999.

[38] Bjorn Poonen. private communication, 2013.

[39] Carl Ludwig Siegel. Über einige anwendungen diophantischer approximationen. *Abh. Pruess. Akad. Wiss. Phys. Math. Kl.*, pages 41–69, 1929.

[40] Michael Stiebitz, Diego Scheide, Bjarne Toft, and Lene M. Favrholdt. *Graph Edge Coloring: Vizing's Theorem and Goldberg's Conjecture*. Wiley, 2012.

[41] Peter Tait. Remarks on the colourings of maps. *Proc. R. Soc. Edinburgh*, 10:729, 1880.

[42] Leslie G. Valiant. Quantum circuits that can be simulated classically in polynomial time. *SIAM J. Comput.*, 31(4):1229–1254, 2002.

[43] Dirk Vertigan. The computational complexity of Tutte invariants for planar graphs. *SIAM J. Comput.*, 35(3):690–712, 2005.

[44] Vadim G. Vizing. Critical graphs with given chromatic class. *Metody Diskret. Analiz.*, 5:9–17, 1965.

[45] P. G. Walsh. A quantitative version of Runge's theorem on Diophantine equations. *Acta Arith.*, 62(2):157–172, 1992.

[46] Dominic Welsh. *Complexity: Knots, Colourings and Countings*. London Mathematical Society Lecture Note Series. Cambridge University Press, 1993.

Full version of Paper 4 attached

Cai et al. *Res Math Sci(2016)3:18*
DOI 10.1186/s40687-016-0067-8

Research in the Mathematical Sciences

RESEARCH Open Access

The complexity of counting edge colorings and a dichotomy for some higher domain Holant problems

Jin-Yi Cai[1], Heng Guo[2] and Tyson Williams[1*]

*Correspondence:
tdw@cs.wisc.edu
[1]University of Wisconsin-Madison, Madison, WI, USA
Full list of author information is available at the end of the article

Abstract

We show that an effective version of Siegel's theorem on finiteness of integer solutions for a specific algebraic curve and an application of elementary Galois theory are key ingredients in a complexity classification of some Holant problems. These Holant problems, denoted by Holant(f), are defined by a symmetric ternary function f that is invariant under any permutation of the $\kappa \geq 3$ domain elements. We prove that Holant(f) exhibits a complexity dichotomy. The hardness, and thus the dichotomy, holds even when restricted to planar multigraphs. A special case of this result is that counting edge κ-colorings is #P-hard over planar 3-regular multigraphs for all $\kappa \geq 3$. In fact, we prove that counting edge κ-colorings is #P-hard over planar r-regular multigraphs for all $\kappa \geq r \geq 3$. The problem is polynomial time computable in all other parameter settings. The proof of the dichotomy theorem for Holant(f) depends on the fact that a specific polynomial $p(x, y)$ has an explicitly listed finite set of integer solutions and the determination of the Galois groups of some specific polynomials. In the process, we also encounter the Tutte polynomial, medial graphs, Eulerian partitions, Puiseux series, and a certain lattice condition on the (logarithm of) the roots of polynomials.

1 Introduction

What do Siegel's theorem and Galois theory have to do with complexity theory? In this paper, we show that an effective version of Siegel's theorem on finiteness of integer solutions for a specific algebraic curve and an application of elementary Galois theory are key ingredients in a chain of steps that lead to a complexity classification of some counting problems. More specifically, we consider a certain class of counting problems that are expressible as Holant problems with an arbitrary domain of size κ over 3-regular multigraphs (i.e., self-loops and parallel edges are allowed) and prove a dichotomy theorem for this class of problems. The hardness, and thus the dichotomy, holds even when restricted to planar multigraphs. Among other things, the proof of the dichotomy theorem depends on the following: (A) the specific polynomial $p(x, y) = x^5 - 2x^3y - x^2y^2 - x^3 + xy^2 + y^3 - 2x^2 - xy$ has only the integer solutions $(x, y) = (-1, 1), (0, 0), (1, -1), (1, 2), (3, 3)$, and (B) the determination of the Galois groups of some specific polynomials. In the process, we also encounter the Tutte polynomial,

Springer Open

medial graphs, Eulerian partitions, Puiseux series, and a certain lattice condition on the (logarithm of) the roots of polynomials such as $p(x, y)$.

A special case of this dichotomy theorem is the problem of counting edge colorings over planar 3-regular multigraphs using κ colors. In this case, the corresponding constraint function is the ALL-DISTINCT$_{3,\kappa}$ function, which takes value 1 when all three inputs from $[\kappa]$ are distinct and 0 otherwise. We further prove that the problem using κ colors over r-regular multigraphs is #P-hard for all $\kappa \geq r \geq 3$, even when restricted to planar multigraphs. The problem is polynomial time computable in all other parameter settings. This solves a long-standing open problem.

We give a brief description of the framework of Holant problems [18,20,21,23]. The problem Holant($\mathcal{F}$), defined by a set of functions $\mathcal{F}$, takes as input a *signature grid* $\Omega = (G, \pi)$, where $G = (V, E)$ is a multigraph, π assigns each $v \in V$ a function $f_v \in \mathcal{F}$, and f_v maps $[\kappa]^{\deg(v)}$ to $\mathbb{C}$ for some integer $\kappa \geq 2$. An edge κ-labeling $\sigma : E \to [\kappa]$ gives an evaluation $\prod_{v \in V} f_v(\sigma \mid_{E(v)})$, where $E(v)$ denotes the incident edges of v and $\sigma \mid_{E(v)}$ denotes the restriction of σ to $E(v)$. The counting problem on the instance Ω is to compute

$$\text{Holant}(\Omega, \mathcal{F}) = \sum_{\sigma : E \to [\kappa]} \prod_{v \in V} f_v \left(\sigma \mid_{E(v)}\right).$$

Counting edge κ-colorings over r-regular multigraphs amounts to setting $f_v = \text{ALL-DISTINCT}_{r,\kappa}$ for all v. We also use Pl-Holant($\mathcal{F}$) to denote the restriction of Holant($\mathcal{F}$) to planar multigraphs.

Holant problems appear in many areas under a variety of different names. They are equivalent to counting constraint satisfaction problems (#CSPs) [7,9] with the restriction that all variables are read twice,[1] to the contraction of a tensor network [25,41], and to the partition function of graphical models in Forney normal form [42,47] from artificial intelligence, coding theory, and signal processing. Special cases of Holant problems include simulating quantum circuits [48,56], counting graph homomorphisms [2,5,12,27,34], and evaluating the partition function of the edge-coloring model [2, Section 3.6].

An edge κ-coloring of a graph G is an edge κ-labeling of G such that any two incident edges have different colors. A fundamental problem in graph theory is to determine how many colors are required to edge color G. The obvious lower bound is $\Delta(G)$, the maximum degree of the graph. By Vizing's theorem [60], an edge coloring using just $\Delta(G) + 1$ colors always exists for simple graphs (i.e., graphs without self-loops or parallel edges). Whether $\Delta(G)$ colors suffice depends on the graph G.

Consider the edge-coloring problem over 3-regular graphs. It follows from the parity condition (Lemma 4.4) that any graph containing a bridge does not have an edge 3-coloring. For bridgeless planar simple graphs, Tait [55] showed that the existence of an edge 3-coloring is equivalent to the four-color theorem. Thus, the answer for the decision problem over planar 3-regular simple graphs is that there is an edge 3-coloring iff the graph is bridgeless.

Without the planarity restriction, determining whether a 3-regular (simple) graph has an edge 3-coloring is NP-complete [39]. This hardness extends to finding an edge κ-coloring over κ-regular (simple) graphs for all $\kappa \geq 3$ [45]. However, these reductions are not parsimonious, and, in fact, it is claimed that no parsimonious reduction exists

[1] Without this restriction, #CSPs are a special case of Holant problems.

unless P = NP [62, p. 118]. The counting complexity of this problem has remained open.

We prove that counting edge colorings over planar regular multigraphs is #P-hard.[2]

Theorem 1.1 *#κ-*EDGECOLORING *is* #P*-hard over planar r-regular multigraphs if* $\kappa \geq r \geq 3$.

This theorem is proved in Theorem 4.8 for $\kappa = r$ and Theorem 4.20 for $\kappa > r$.

The techniques we develop to prove Theorem 1.1 naturally extend to a class of Holant problems with domain size $\kappa \geq 3$ over planar 3-regular multigraphs. Functions such as ALL-DISTINCT$_{3,\kappa}$ are symmetric, which means that they are invariant under any permutation of its three inputs. But ALL-DISTINCT$_{3,\kappa}$ has another invariance—it is invariant under any permutation of the κ domain elements. We call the second property *domain invariance*.

A ternary function that is both symmetric and domain invariant is specified by three values, which we denote by $\langle a, b, c\rangle$. The output is a when all inputs are the same, the output is c when all inputs are distinct, and the output is b when two inputs are the same but the third input is different.

We prove a dichotomy theorem for such functions with complex weights.

Theorem 1.2 *Suppose* $\kappa \geq 3$ *is the domain size and* $a, b, c \in \mathbb{C}$. *Then either* Holant$(\langle a, b, c\rangle)$ *is computable in polynomial time or* Pl-Holant$(\langle a, b, c\rangle)$ *is* #P*-hard. Furthermore, given* a, b, c, *there is a polynomial-time algorithm that decides which is the case.*

See Theorem 10.1 for an explicit listing of the tractable cases. Note that counting edge κ-colorings over 3-regular multigraphs is the special case when $\langle a, b, c\rangle = \langle 0, 0, 1\rangle$.

There is only one previous dichotomy theorem for higher domain Holant problems [22] (see Theorem 5.1). The important difference is that the present work is for general domain size $\kappa \geq 3$, while the previous result is for domain size $\kappa = 3$. When restricted to domain size 3, the result in [22] assumes that all unary functions are available, while this dichotomy does not assume that; however, it does assume domain invariance. Dichotomy theorems for an arbitrary domain size are generally difficult to prove. The Feder-Vardi conjecture for decision constraint satisfaction problems (CSPs) is still open [32]. It was a major achievement to prove this conjecture for domain size 3 [6]. The #CSP dichotomy was proved after a long series of papers [4,5,7–9,11,15,16,24,26,28,35].

Our proof of Theorem 1.2 has many components, and a number of new ideas are introduced in this proof. We discuss some of these ideas and give an outline of our proof in Sect. 2. In Sect. 3, we review basic terminology and define the notation of a *succinct signature*. Section 4 contains our proof of Theorem 1.1 about edge coloring. In Sect. 5, we discuss the tractable cases of Theorem 1.2. In Sect. 6, we extend our main proof technique of polynomial interpolation. Then in Sects. 7, 8, and 9, we develop our hardness proof and tie everything together in Sect. 10.

[2] Vizing's theorem is for simple graphs. In Holant problems as well as counting complexity such as graph homomorphism or #CSP, one typically considers multigraphs (i.e., self-loops and parallel edges are allowed). However, our hardness result for counting edge 3-colorings over planar 3-regular multigraphs also holds for simple graphs (Theorem 4.9).

2 Proof outline and techniques

As usual, the difficult part of a dichotomy theorem is to carve out *exactly* the tractable problems in the class and prove all the rest #P-hard. A dichotomy theorem for Holant problems has the additional difficulty that some tractable problems are only shown to be tractable under a holographic transformation, which can make the appearance of the problem rather unexpected. For example, we show in Sect. 5 that the problem Holant($\langle -3-4i, 1, -1+2i\rangle$) on domain size 4 is tractable. Despite its appearance, this problem is intimately connected with a tractable graph homomorphism problem defined by the Hadamard matrix $\begin{bmatrix} 1 & -1 & -1 & -1 \\ -1 & 1 & -1 & -1 \\ -1 & -1 & 1 & -1 \\ -1 & -1 & -1 & 1 \end{bmatrix}$. In order to understand all problems in a Holant problem class, we must deal with such problems. Dichotomy theorems for graph homomorphisms and for #CSP do not have to deal with as varied a class of such problems, since they implicitly assume all EQUALITY functions are available and must be preserved. This restricts the possible transformations.

After isolating a set of tractable problems, our #P-hardness results in both Theorem 1.1 and Theorem 1.2 are obtained by reducing from evaluations of the Tutte polynomial over planar graphs. A dichotomy is known for such problems (Theorem 4.1).

The chromatic polynomial, a specialization of the Tutte polynomial (Proposition 4.10), is concerned with vertex colorings. On domain size κ, one starting point of our hardness proofs is the chromatic polynomial, which contains the problem of counting vertex colorings using at most κ colors. By the planar dichotomy for the Tutte polynomial, this problem is #P-hard for all $\kappa \geq 3$.

Another starting point for our hardness reductions is the evaluation of the Tutte polynomial at an integer diagonal point (x, x), which is #P-hard for all $x \geq 3$ by the same planar Tutte dichotomy. These are new starting places for reductions involving Holant problems. These problems were known to have a so-called state-sum expression (Lemma 4.3), which is a sum over weighted Eulerian partitions. This sum is not over the original planar graph but over its directed medial graph, which is always a planar 4-regular graph (Fig. 4). We show that this state-sum expression is naturally expressed as a Holant problem with a particular quaternary constraint function (Lemma 4.6).

To reduce from these two problems, we execute the following strategy. First, we attempt to construct the unary constraint function $\langle 1\rangle$, which takes value 1 on all κ inputs (Lemma 8.1). Second, we attempt to interpolate all succinct binary signatures, assuming that we have $\langle 1\rangle$ (Sect. 9). (See Sect. 3 for the definition of a succinct signature.) Lastly, we attempt to construct a ternary signature with a special property, assuming that all these binary signatures are available (Lemma 7.1). At each step, there are some problems specified by certain signatures $\langle a, b, c\rangle$ for which our attempts fail. In such cases, we directly obtain a dichotomy without the help of additional signatures. See Fig. 1 for a flowchart of hardness reductions.

Below we highlight some of our proof techniques.

Interpolation within an orthogonal subspace We develop the ability to interpolate when faced with some nontrivial null spaces inherently present in interpolation constructions. In any construction involving an initial signature and a recurrence matrix, it is possible that the initial signature is orthogonal to some row eigenvectors of the recurrence matrix. Previous interpolation results always attempt to find a construction that

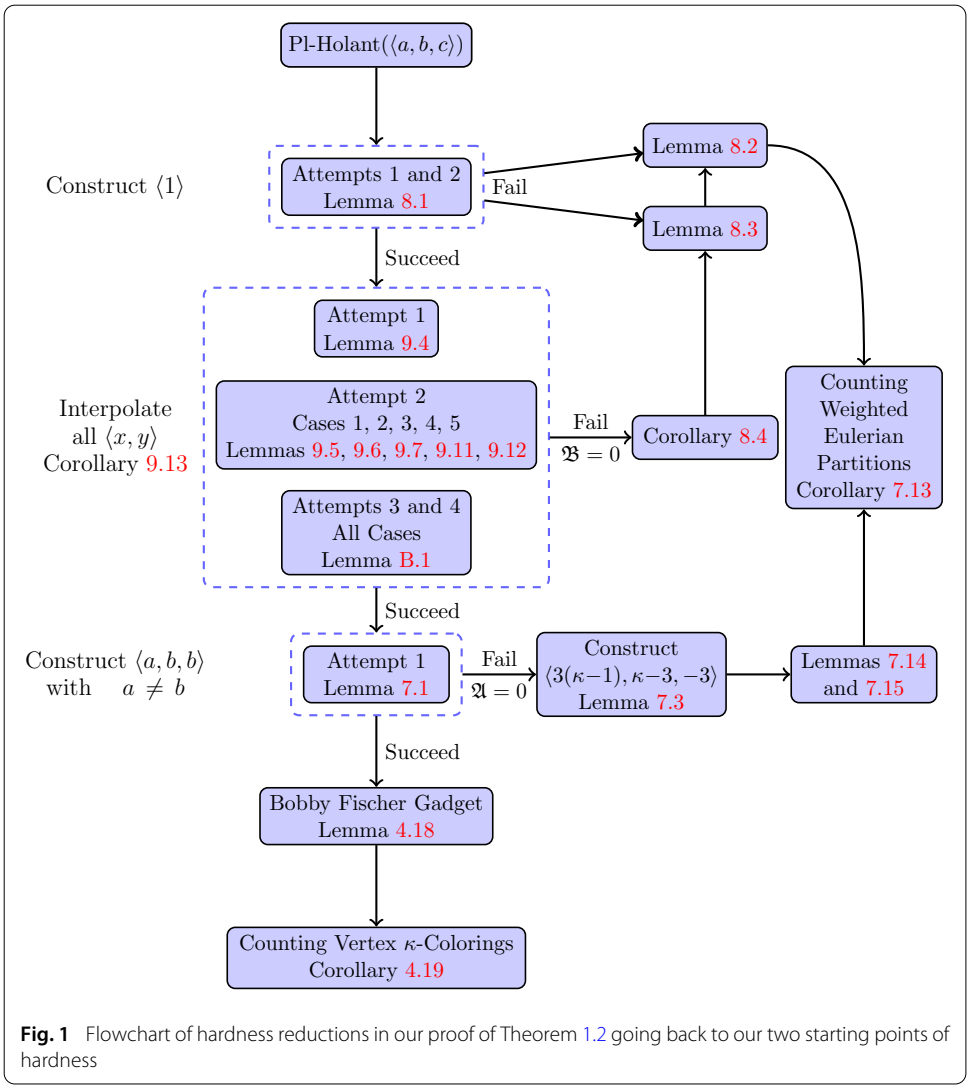


Fig. 1 Flowchart of hardness reductions in our proof of Theorem 1.2 going back to our two starting points of hardness

avoids this. In the present work, this avoidance seems impossible. In Sect. 6, we prove an interpolation result that can succeed in this situation to the greatest extent possible. We prove that one can interpolate any signature, provided that it is orthogonal to the same set of row eigenvectors, and the relevant eigenvalues satisfy a lattice condition (Lemma 6.6).

Satisfy lattice condition via Galois theory A key requirement for this interpolation to succeed is the lattice condition (Definition 6.3), which involves the roots of the characteristic polynomial of the recurrence matrix. We use Galois theory to prove that our constructions satisfy this condition. If a polynomial has a large Galois group, such as S_n or A_n, and its roots do not all have the same complex norm, then we show that its roots satisfy the lattice condition (Lemma 6.5).

Effective Siegel's theorem via Puiseux series We need to determine the Galois groups for an infinite family of polynomials, one for each domain size. If these polynomials are irreducible, then we can show they all have the full symmetric group as their Galois group and hence fulfill the lattice condition. We suspect that these polynomials are all irreducible but are unable to prove it.

A necessary condition for irreducibility is the absence of any linear factor. This infinite family of polynomials, as a single bivariate polynomial in (x, κ), defines an algebraic curve, which has genus 3. By a well-known theorem of Siegel [52], there are only a finite number of integer values of κ for which the corresponding polynomial has a linear factor. However, this theorem and others like it are not *effective* in general. There are some effective versions of Siegel's theorem that can be applied to the algebraic curve, but the best general effective bound is over $10^{20,000}$ [61] and hence cannot be checked in practice. Instead, we use Puiseux series to show that this algebraic curve has exactly five explicitly listed integer solutions (Lemma 7.6).

Eigenvalue shifted triples For a pair of eigenvalues, the lattice condition is equivalent to the statement that the ratio of these eigenvalues is not a root of unity. A sufficient condition is that the eigenvalues have distinct complex norms. We prove three results, each of which is a different way to satisfy this sufficient condition. Chief among them is the technique we call an *Eigenvalue Shifted Triple* (EST). These generalize the technique of Eigenvalue Shifted Pairs from [43]. In an EST, we have three recurrence matrices, each of which differs from the other two by a nonzero additive multiple of the identity matrix. Provided these two multiples are linearly independent over $\mathbb{R}$, we show at least one of these matrices has eigenvalues with distinct complex norms (Lemma 9.10). (However, determining which one succeeds is a difficult task, but we need not know that).

E Pluribus Unum When the ratio of a pair of eigenvalues is a root of unity, it is a challenge to effectively use this failure condition. Direct application of this cyclotomic condition is often of limited use. We introduce an approach that uses this cyclotomic condition effectively. A direct recursive construction involving these two eigenvalues only creates a finite number of different signatures. We reuse all of these signatures in a multitude of new interpolation constructions (Lemma 9.3), one of which we hope will succeed. If the eigenvalues in all of these constructions also satisfy a cyclotomic condition, then we obtain a more useful condition than any of the previous cyclotomic conditions. This idea generalizes the anti-gadget technique [17], which only reuses the "last" of these signatures.

Local holographic transformation One reason to obtain all succinct binary signatures is for use in the gadget construction known as a local holographic transformation (Fig. 11). This construction mimics the effect of a holographic transformation applied on a single signature. In particular, using this construction, we attempt to obtain a succinct ternary signature of the form $\langle a, b, b\rangle$, where $a \neq b$ (Lemma 7.1). This signature turns out to have some magical properties in the Bobby Fischer gadget, which we discuss next.

Bobby Fischer gadget Typically, any combinatorial construction for higher domain Holant problems produces very intimidating looking expressions that are nearly impossible to analyze. In our case, it seems necessary to consider a construction that has to satisfy multiple requirements involving at least nine polynomials. However, we are able to combine the signature $\langle a, b, b\rangle$, where $a \neq b$, with a succinct binary signature of our choice in a special construction that we call the *Bobby Fischer gadget* (Fig. 9). This gadget is able to satisfy seven conditions using just one degree of freedom (Lemma 4.18). This ability to satisfy a multitude of constraints simultaneously in one magic stroke reminds us of some unfathomably brilliant moves by Bobby Fischer, the chess genius extraordinaire.

3 Preliminaries

3.1 Problems and definitions

The framework of Holant problems is defined for functions mapping any $[\kappa]^n \to R$ for a finite κ and some commutative semiring R. In this paper, we investigate some complex-weighted Holant problems on domain size $\kappa \geq 3$. A constraint function, or *signature*, of arity n maps from $[\kappa]^n \to \mathbb{C}$. For consideration of models of computation, functions take complex algebraic numbers.

Graphs (called multigraphs in Sect. 1) may have self-loops and parallel edges. A graph without self-loops or parallel edges is a *simple* graph. A *signature grid* $\Omega = (G, \pi)$ of Holant($\mathcal{F}$) consists of a graph $G = (V, E)$, where π assigns to each vertex $v \in V$ and its incident edges some $f_v \in \mathcal{F}$ and its input variables. We say Ω is a *planar signature grid* if G is planar, where the variables of f_v are ordered counterclockwise. The Holant problem on instance Ω is to evaluate $\text{Holant}(\Omega; \mathcal{F}) = \sum_\sigma \prod_{v \in V} f_v(\sigma \mid_{E(v)})$, a sum over all edge assignments $\sigma : E \to [\kappa]$, where $E(v)$ denotes the incident edges of v and $\sigma \mid_{E(v)}$ denotes the restriction of σ to $E(v)$.

A function f_v can be represented by listing its values in lexicographical order as in a truth table, which is a vector in $\mathbb{C}^{\kappa^{\deg(v)}}$, or as a tensor in $(\mathbb{C}^\kappa)^{\otimes \deg(v)}$. In this paper, we consider symmetric signatures. An example of which is the EQUALITY signature $=_r$ of arity r. Sometimes we represent f as a matrix M_f that we call its *signature matrix*, which has row index $(x_1, \ldots, x_t)$ and column index $(x_k, \ldots, x_{t+1})$ (in reverse order) for some t that will be clear from context.

A Holant problem is parametrized by a set of signatures.

Definition 3.1 Given a set of signatures $\mathcal{F}$, we define the counting problem Holant($\mathcal{F}$) as:

Input: A signature grid $\Omega = (G, \pi)$;
Output: Holant($\Omega; \mathcal{F}$).

The problem Pl-Holant($\mathcal{F}$) is defined similarly using a planar signature grid.

A signature f of arity n is *degenerate* if there exist unary signatures $u_j \in \mathbb{C}^\kappa$ $(1 \leq j \leq n)$ such that $f = u_1 \otimes \cdots \otimes u_n$. A symmetric degenerate signature has the form $u^{\otimes n}$. For such signatures, it is equivalent to replace it by n copies of the corresponding unary signature. Replacing a signature $f \in \mathcal{F}$ by a constant multiple cf, where $c \neq 0$, does not change the complexity of Holant($\mathcal{F}$). It introduces a global nonzero factor to Holant($\Omega; \mathcal{F}$).

We allow $\mathcal{F}$ to be an infinite set. For Holant($\mathcal{F}$) to be tractable, the problem must be computable in polynomial time even when the description of the signatures in the input Ω is included in the input size. In contrast, we say Holant($\mathcal{F}$) is #P-hard if there exists a finite subset of $\mathcal{F}$ for which the problem is #P-hard. The same definitions apply for Pl-Holant($\mathcal{F}$) when Ω is a planar signature grid. We say a signature set $\mathcal{F}$ is tractable (resp. #P-hard) if the corresponding counting problem Holant($\mathcal{F}$) is tractable (resp. #P-hard). We say $\mathcal{F}$ is tractable (resp. #P-hard) for planar problems if Pl-Holant($\mathcal{F}$) tractable (resp. #P-hard). Similarly for a signature f, we say f is tractable (resp. #P-hard) if $\{f\}$ is.

We follow the usual conventions about polynomial-time Turing reduction $\leq_T$ and polynomial-time Turing equivalence $\equiv_T$. We use I_n and J_n to denote the n-by-n identity matrix and n-by-n matrix of all 1's, respectively.

3.2 Holographic reduction

To introduce the idea of holographic reductions, it is convenient to consider bipartite graphs. For a general graph, we can always transform it into a bipartite graph while preserving the Holant value, as follows. For each edge in the graph, we replace it by a path of length two. (This operation is called the *2-stretch* of the graph and yields the edge-vertex incidence graph.) Each new vertex is assigned the binary EQUALITY signature $=_2$.

We use Holant($\mathcal{F} \mid \mathcal{G}$) to denote the Holant problem on bipartite graphs $H = (U, V, E)$, where each vertex in U or V is assigned a signature in $\mathcal{F}$ or $\mathcal{G}$, respectively. Signatures in $\mathcal{F}$ are considered as row vectors (or covariant tensors); signatures in $\mathcal{G}$ are considered as column vectors (or contravariant tensors) [25]. Similarly, Pl-Holant($\mathcal{F} \mid \mathcal{G}$) denotes the Holant problem over signature grids with a planar bipartite graph.

For a κ-by-κ matrix T and a signature set $\mathcal{F}$, define $T\mathcal{F} = \{g \mid \exists f \in \mathcal{F}$ of arity n, $g = T^{\otimes n} f\}$, similarly for $\mathcal{F}T$. Whenever we write $T^{\otimes n} f$ or $T\mathcal{F}$, we view the signatures as column vectors, similarly for $fT^{\otimes n}$ or $\mathcal{F}T$ as row vectors.

Let T be an invertible κ-by-κ matrix. The holographic transformation defined by T is the following operation: given a signature grid $\Omega = (H, \pi)$ of Holant($\mathcal{F} \mid \mathcal{G}$), for the same bipartite graph H, we get a new grid $\Omega' = (H, \pi')$ of Holant($\mathcal{F}T \mid T^{-1}\mathcal{G}$) by replacing each signature in $\mathcal{F}$ or $\mathcal{G}$ with the corresponding signature in $\mathcal{F}T$ or $T^{-1}\mathcal{G}$. Valiant's Holant Theorem [57] (see also [13]) is easily generalized to domain size $\kappa \geq 3$.

Theorem 3.2 *Suppose $\kappa \geq 3$ is the domain size. If $T \in \mathbb{C}^{\kappa \times \kappa}$ is an invertible matrix, then* Holant($\Omega; \mathcal{F} \mid \mathcal{G}$) = Holant($\Omega'; \mathcal{F}T \mid T^{-1}\mathcal{G}$).

Therefore, an invertible holographic transformation does not change the complexity of the Holant problem in the bipartite setting. Furthermore, there is a special kind of holographic transformation, the orthogonal transformation, that preserves the binary equality and thus can be used freely in the standard setting. For $\kappa = 2$, this first appeared in [18] as Theorem 2.2.

Theorem 3.3 *Suppose $\kappa \geq 3$ is the domain size. If $T \in \mathbb{C}^{\kappa \times \kappa}$ is an orthogonal matrix (i.e., $TT^T = I_\kappa$), then* Holant($\Omega; \mathcal{F}$) = Holant($\Omega'; T\mathcal{F}$).

Since the complexity of a signature is unchanged by a nonzero constant multiple, we also call a transformation T such that $TT^{\mathrm{T}} = \lambda I$ for some $\lambda \neq 0$ an orthogonal transformation. Such transformations do not change the complexity of a problem.

3.3 Realization

One basic notion used throughout the paper is realization. We say a signature f is *realizable* or *constructible* from a signature set $\mathcal{F}$ if there is a gadget with some dangling edges such that each vertex is assigned a signature from $\mathcal{F}$, and the resulting graph, when viewed as a black-box signature with inputs on the dangling edges, is exactly f. If f is realizable from a set $\mathcal{F}$, then we can freely add f into $\mathcal{F}$ while preserving the complexity.

Formally, such a notion is defined by an $\mathcal{F}$-gate [18,19]. An $\mathcal{F}$-gate is similar to a signature grid (G, π) for Holant($\mathcal{F}$) except that $G = (V, E, D)$ is a graph with some dangling edges D. The dangling edges define external variables for the $\mathcal{F}$-gate. (See Fig. 2 for an example.) We denote the regular edges in E by $1, 2, \ldots, m$ and the dangling edges in D by $m+1, \ldots, m+n$. Then we can define a function Γ for this $\mathcal{F}$-gate as

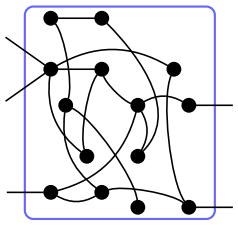


Fig. 2 An $\mathcal{F}$-gate with 5 dangling edges

$$\Gamma(y_1, \ldots, y_n) = \sum_{x_1, \ldots, x_m \in [\kappa]} H(x_1, \ldots, x_m, y_1, \ldots, y_n),$$

where $(y_1, \ldots, y_n) \in [\kappa]^n$ is an assignment on the dangling edges and $H(x_1, \ldots, x_m, y_1, \ldots, y_n)$ is the value of the signature grid on an assignment of all edges in G, which is the product of evaluations at all internal vertices. We also call this function Γ the signature of the $\mathcal{F}$-gate.

An $\mathcal{F}$-gate is planar if the underlying graph G is a planar graph, and the dangling edges, ordered counterclockwise corresponding to the order of the input variables, are in the outer face in a planar embedding. A planar $\mathcal{F}$-gate can be used in a planar signature grid as if it is just a single vertex with the particular signature.

Using the idea of planar $\mathcal{F}$-gates, we can reduce one planar Holant problem to another. Suppose g is the signature of some planar $\mathcal{F}$-gate. Then Pl-Holant$(\mathcal{F} \cup \{g\}) \leq_T$ Pl-Holant$(\mathcal{F})$. The reduction is simple. Given an instance of Pl-Holant$(\mathcal{F} \cup \{g\})$, by replacing every appearance of g by the $\mathcal{F}$-gate, we get an instance of Pl-Holant$(\mathcal{F})$. Since the signature of the $\mathcal{F}$-gate is g, the Holant values for these two signature grids are identical.

Although our main results are about symmetric signatures (i.e., signatures that are invariant under any permutation of inputs), some of our proofs utilize asymmetric signatures. When a gadget has an asymmetric signature, we place a diamond on the edge corresponding to the first input. The remaining inputs are ordered counterclockwise around the vertex. (See Fig. 5 for an example.)

We note that even for a very simple signature set $\mathcal{F}$, the signatures for all $\mathcal{F}$-gates can be quite complicated and expressive.

3.4 Succinct signatures

An arity r signature on domain size κ is fully specified by κ^r values. However, some special cases can be defined using far fewer values. Consider the signature ALL-DISTINCT$_{r,\kappa}$ of arity r on domain size κ that outputs 1 when all inputs are distinct and 0 otherwise. We also denote this signature by $\text{AD}_{r,\kappa}$. In addition to being symmetric, it is also invariant under any permutation of the κ domain elements. We call the second property *domain invariance*. The signature of an $\mathcal{F}$-gate in which all signatures in $\mathcal{F}$ are domain invariant is itself domain invariant.

Definition 3.4 *(Succinct signature)* Let $\tau = (P_1, P_2, \ldots, P_\ell)$ be a partition of $[\kappa]^r$ listed in some order. We say that f is a *succinct signature* of type τ if f is constant on each P_i. A set $\mathcal{F}$ of signatures is of type τ if every $f \in \mathcal{F}$ has type τ. We denote a succinct signature f of type τ by $\langle f(P_1), \ldots, f(P_\ell)\rangle$, where $f(P) = f(x)$ for any $x \in P$.

Furthermore, we may omit 0 entries. If f is a succinct signature of type τ, we also say f is a *succinct signature* of type τ' with length ℓ', where τ' lists ℓ' parts of the partition τ, and we write f as $\langle f_1, f_2, \ldots, f_{\ell'}\rangle$, provided that all nonzero values $f(P_i)$ are listed. When using this notation, we will make it clear which zero entries have been omitted.

For example, a symmetric signature in the Boolean domain (i.e., $\kappa = 2$) has been denoted in previous work [14] by $[f_0, f_1, \ldots, f_r]$, where f_w is the output on inputs of Hamming weight w. This corresponds to the succinct signature type $(P_0, P_1, \ldots, P_r)$, where P_w is the set of inputs of Hamming weight w. A similar succinct signature notation was used for symmetric signatures on domain size 3 [22, p. 1282].

We prove a dichotomy theorem for Pl-Holant(f) when f is a succinct ternary signature of type τ_3 on domain size $\kappa \geq 3$. For $\kappa \geq 3$, the succinct signature of type $\tau_3 = (P_1, P_2, P_3)$ is a partition of $[\kappa]^3$ with $P_i = \{(x, y, z) \in [\kappa]^3 : |\{x, y, z\}| = i\}$ for $1 \leq i \leq 3$. The notation $\{x, y, z\}$ denotes a multiset, and $|\{x, y, z\}|$ denotes the number of distinct elements in it. Succinct signatures of type τ_3 are exactly the symmetric and domain invariant ternary signatures. In particular, the succinct ternary signature for $\mathrm{AD}_{3,\kappa}$ is $\langle 0, 0, 1\rangle$.

We use several other succinct signature types as well. For domain invariant unary signatures, there are only two signatures up to a nonzero scalar. Using the trivial partition that contains all inputs, we denote these two succinct unary signatures as $\langle 0\rangle$ and $\langle 1\rangle$ and say that they have succinct type τ_1. We also need a succinct signature type for domain invariant binary signatures. Such signatures are necessarily symmetric. We call their succinct signature type $\tau_2 = (P_1, P_2)$, where $P_i = \{(x, y) \in [\kappa]^2 : |\{x, y\}| = i\}$ for $1 \leq i \leq 2$.

We note that the number of succinct signature types for arity r signatures on domain size κ that are both symmetric and domain invariant is the number of partitions of r into at most κ parts. This is related to the partition function from number theory, which is not to be confused with the partition function with its origins in statistical mechanics and has been intensively studied in complexity theory of counting problems.

While there are some other succinct signature types that we define later as needed, there is one more important type that we define here. Any quaternary signature f that is domain invariant has a succinct signature of length at most 15. When a signature has both vertical and horizontal symmetry, there is a shorter succinct signature that has only length 9. We say a signature f has vertical symmetry if $f(w, x, y, z) = f(x, w, z, y)$ and horizontal symmetry if $f(w, x, y, z) = f(z, y, x, w)$. For example, the signature of the gadget in Fig. 9 has both vertical and horizontal symmetry. Accordingly, let $\tau_4 = (P_{\begin{smallmatrix}1&1\\1&1\end{smallmatrix}}, P_{\begin{smallmatrix}1&2\\1&1\end{smallmatrix}}, P_{\begin{smallmatrix}1&2\\1&2\end{smallmatrix}}, P_{\begin{smallmatrix}1&3\\1&2\end{smallmatrix}}, P_{\begin{smallmatrix}1&2\\2&1\end{smallmatrix}}, P_{\begin{smallmatrix}1&3\\2&1\end{smallmatrix}}, P_{\begin{smallmatrix}1&1\\2&2\end{smallmatrix}}, P_{\begin{smallmatrix}1&1\\2&3\end{smallmatrix}}, P_{\begin{smallmatrix}1&4\\2&3\end{smallmatrix}})$ be a type of succinct quaternary signature with partitions

$$
\begin{aligned}
P_{\begin{smallmatrix}1&1\\1&1\end{smallmatrix}} &= \{(w, x, y, z) \in [\kappa]^4 \mid w = x = y = z\},\\
P_{\begin{smallmatrix}1&2\\1&1\end{smallmatrix}} &= \left\{(w, x, y, z) \in [\kappa]^4 \middle| \begin{array}{l}(w = x = y \neq z) \vee (w = x = z \neq y)\\ \vee (w = y = z \neq x) \vee (x = y = z \neq w)\end{array}\right\},\\
P_{\begin{smallmatrix}1&2\\1&2\end{smallmatrix}} &= \{(w, x, y, z) \in [\kappa]^4 \mid w = x \neq y = z\},\\
P_{\begin{smallmatrix}1&3\\1&2\end{smallmatrix}} &= \{(w, x, y, z) \in [\kappa]^4 \mid (w = x \neq y \neq z \neq x) \vee (y = z \neq w \neq x \neq z)\},\\
P_{\begin{smallmatrix}1&2\\2&1\end{smallmatrix}} &= \{(w, x, y, z) \in [\kappa]^4 \mid w = y \neq x = z\},\\
P_{\begin{smallmatrix}1&3\\2&1\end{smallmatrix}} &= \{(w, x, y, z) \in [\kappa]^4 \mid (w = y \neq x \neq z \neq y) \vee (x = z \neq w \neq y \neq z)\},\\
P_{\begin{smallmatrix}1&1\\2&2\end{smallmatrix}} &= \{(w, x, y, z) \in [\kappa]^4 \mid w = z \neq x = y\},
\end{aligned}
$$

$P_{\substack{1\,1\\2\,3}} = \{(w,x,y,z) \in [\kappa]^4 \mid (w = z \neq x \neq y \neq z) \vee (x = y \neq w \neq z \neq y)\}$, and
$P_{\substack{1\,4\\2\,3}} = \{(w,x,y,z) \in [\kappa]^4 \mid w,x,y,z \text{ are all distinct}\}$.

4 Counting edge κ-colorings over planar r-regular graphs

In this section, we show that counting edge κ-colorings over planar r-regular graphs is #P-hard provided $\kappa \geq r \geq 3$. When this condition fails to hold, the problem is trivially tractable. There are two cases depending on whether $\kappa = r$ or not.

4.1 The Case $\kappa = r$

When $\kappa = r$, we reduce from evaluating the Tutte polynomial of a planar graph at the positive integer points on the diagonal $x = y$. For $x \geq 3$, evaluating the Tutte polynomial of a planar graph at (x,x) is #P-hard.

Theorem 4.1 (Theorem 5.1 in [59]) *For $x,y \in \mathbb{C}$, evaluating the Tutte polynomial at (x,y) is #P-hard over planar graphs unless $(x-1)(y-1) \in \{1,2\}$ or $(x,y) \in \{(1,1),(-1,-1),(\omega,\omega^2),(\omega^2,\omega)\}$, where $\omega = e^{2\pi i/3}$. In each exceptional case, the computation can be done in polynomial time.*

To state the connection with the diagonal of the Tutte polynomial, we need to consider Eulerian subgraphs in directed medial graphs. We say a graph is an Eulerian (di)graph if every vertex has even degree (resp. in-degree equal to out-degree), but connectedness is not required. Now recall the definition of a medial graph and its directed variant.

Definition 4.2 (*cf. Section 4 in* [30]) For a connected plane graph G (i.e., a planar embedding of a connected planar graph), its *medial graph* G_m has a vertex on each edge of G and two vertices in G_m are joined by an edge for each face of G in which their corresponding edges occur consecutively.

The *directed medial graph* $\vec{G}_m$ of G colors the faces of G_m black or white depending on whether they contain or do not contain, respectively, a vertex of G. Then the edges of the medial graph are directed so that the black face is on the left.

Figures 3 and 4 give examples of a medial graph and a directed medial graph, respectively. Notice that the (directed) medial graph is always a planar 4-regular graph.

Building on previous work [1,29,49,58], Ellis-Monaghan gave the following connection with the diagonal of the Tutte polynomial. A monochromatic vertex is a vertex with all its incident edges having the same color.

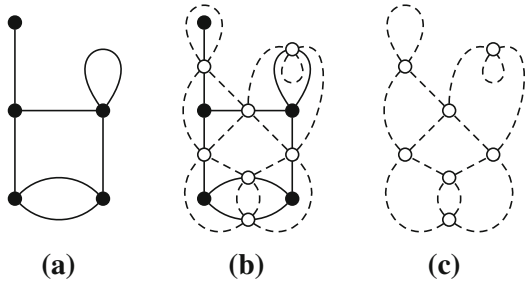


Fig. 3 A plane graph (**a**), its medial graph (**c**), and the two graphs superimposed (**b**)

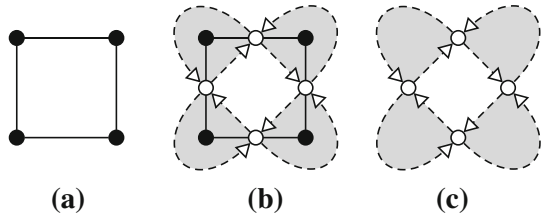


Fig. 4 A plane graph (**a**), its directed medial graph (**c**), and both superimposed (**b**)

Lemma 4.3 (Equation (17) in [30]) *Suppose G is a connected plane graph and $\vec{G}_m$ is its directed medial graph. For $\kappa \in \mathbb{N}$, let $\mathcal{C}(\vec{G}_m)$ be the set of all edge κ-labelings of $\vec{G}_m$ so that each (possibly empty) set of monochromatic edges forms an Eulerian digraph. Then*

$$\kappa \, \mathrm{T}(G; \kappa+1, \kappa+1) = \sum_{c \in \mathcal{C}(\vec{G}_m)} 2^{m(c)}, \tag{1}$$

where $m(c)$ is the number of monochromatic vertices in the coloring c.

The Eulerian partitions in $\mathcal{C}(\vec{G}_m)$ have the property that the subgraphs induced by each partition do not intersect (or crossover) each other due to the orientation of the edges in the medial graph. We call the counting problem defined by the sum on the right-hand side of (1) counting weighted Eulerian partitions over planar 4-regular graphs. This problem also has an expression as a Holant problem using a succinct signature. To define this succinct signature, it helps to know the following basic result about edge colorings.

When the number of available colors coincides with the regularity parameter of the graph, the cuts in any coloring satisfy a well-known parity condition. This parity condition follows from a more general parity argument (see (1.2) and the parity argument on page 95 in [54]). We state this simpler parity condition and provide a short proof for completeness.

Lemma 4.4 (Parity Condition) *Let G be a κ-regular graph and consider a cut C in G. For any edge κ-coloring of G,*

$$c_1 \equiv c_2 \equiv \cdots \equiv c_\kappa \pmod 2,$$

where c_i is the number of edges in C colored i for $1 \le i \le \kappa$.

Proof Consider two distinct colors i and j. Remove from G all edges not colored i or j. The resulting graph is a disjoint union of cycles consisting of alternating colors i and j. Each cycle in this graph must cross the cut C an even number of times. Therefore, $c_i \equiv c_j \pmod 2$. □

Consider all quaternary $\{\mathrm{AD}_{\kappa,\kappa}\}$-gates on domain size $\kappa \ge 3$. These gadgets have a succinct signature of type $\tau_{\text{color}} = (P_{\substack{1\,1\\1\,1}}, P_{\substack{1\,2\\1\,2}}, P_{\substack{1\,2\\2\,1}}, P_{\substack{1\,1\\2\,2}}, P_{\substack{1\,4\\2\,3}}, P_0)$, where

$$P_{\substack{1\,1\\1\,1}} = \{(w,x,y,z) \in [\kappa]^4 \mid w = x = y = z\},$$
$$P_{\substack{1\,2\\1\,2}} = \{(w,x,y,z) \in [\kappa]^4 \mid w = x \neq y = z\},$$
$$P_{\substack{1\,2\\2\,1}} = \{(w,x,y,z) \in [\kappa]^4 \mid w = y \neq x = z\},$$
$$P_{\substack{1\,1\\2\,2}} = \{(w,x,y,z) \in [\kappa]^4 \mid w = z \neq x = y\},$$
$$P_{\substack{1\,4\\2\,3}} = \{(w,x,y,z) \in [\kappa]^4 \mid w,x,y,z \text{ are distinct}\}, \text{ and}$$

$$P_0 = [\kappa]^4 - P_{\substack{1\,1\\1\,1}} - P_{\substack{1\,2\\1\,2}} - P_{\substack{1\,2\\2\,1}} - P_{\substack{1\,1\\2\,2}} - P_{\substack{1\,4\\2\,3}}.$$

Any quaternary signature of an $\{\mathrm{AD}_{\kappa,\kappa}\}$-gate is constant on the first five parts of τ_{color} since $\mathrm{AD}_{\kappa,\kappa}$ is domain invariant. Using Lemma 4.4, we can show that the entry corresponding to P_0 is 0.

Lemma 4.5 *Suppose $\kappa \geq 3$ is the domain size and let F be a quaternary $\{\mathrm{AD}_{\kappa,\kappa}\}$-gate with succinct signature f of type τ_{color}. Then $f(P_0) = 0$.*

Proof Let $\sigma_0 \in P_0$ be an edge κ-labeling of the external edges of F. Assume for a contradiction that σ_0 can be extended to an edge κ-coloring of F. We form a graph G from two copies of F (namely, F_1 and F_2) by connecting their corresponding external edges. Then G has a coloring σ that extends σ_0. Consider the cut $C = (F_1, F_2)$ in G whose cut set contains exactly those edges labeled by σ_0. By Lemma 4.4, the counts of the colors assigned by σ_0 must satisfy the parity condition. However, this is a contradiction since no edge κ-labeling in P_0 satisfies the parity condition. □

By Lemma 4.5, we denote a quaternary signature f of an $\{\mathrm{AD}_{\kappa,\kappa}\}$-gate by the succinct signature $\langle f(P_{\substack{1\,1\\1\,1}}), f(P_{\substack{1\,2\\1\,2}}), f(P_{\substack{1\,2\\2\,1}}), f(P_{\substack{1\,1\\2\,2}}), f(P_{\substack{1\,4\\2\,3}})\rangle$ of type τ_{color}, which has the entry for P_0 omitted.[3] When $\kappa = 3$, $P_{\substack{1\,4\\2\,3}}$ is empty and we define its entry in the succinct signature to be 0.

Lemma 4.6 *Let G be a connected plane graph and let G_m be the medial graph of G. Then*

$$\kappa\ \mathrm{T}(G; \kappa+1, \kappa+1) = \text{Pl-Holant}(G_m; \langle 2, 1, 0, 1, 0\rangle),$$

where the Holant problem has domain size κ and $\langle 2, 1, 0, 1, 0\rangle$ is a succinct signature of type τ_{color}.

Proof Let $f = \langle 2, 1, 0, 1, 0\rangle$. By Lemma 4.3, we only need to prove that

$$\sum_{c \in \mathcal{C}(\vec{G}_m)} 2^{m(c)} = \text{Pl-Holant}(G_m; f), \tag{2}$$

where the notation is from Lemma 4.3.

Each $c \in \mathcal{C}(\vec{G}_m)$ is also an edge κ-labeling of G_m. At each vertex $v \in V(\vec{G}_m)$, the four incident edges are assigned at most two distinct colors by c. If all four edges are assigned the same color, then the constraint f on v contributes a factor of 2 to the total weight. This is given by the value in the first entry of f. Otherwise, there are two different colors, say x and y. Because the orientation at v in $\vec{G}_m$ is cyclically "in, out, in, out," the coloring around v can only be of the form $xxyy$ or $xyyx$. These correspond to the second and fourth entries of f. Therefore, every term in the summation on the left-hand side of (2) appears (with the same nonzero weight) in the summation Pl-Holant$(G_m; f)$.

It is also easy to see that every nonzero term in Pl-Holant$(G_m; f)$ appears in the sum on the left-hand side of (2) with the same weight of 2 to the power of the number of

[3] If $\kappa > 4$, then Lemma 4.4 further implies that these signatures are also 0 on $P_{\substack{1\,4\\2\,3}}$. However, when $\kappa = 4$, this value might be nonzero. The $\mathrm{AD}_{4,4}$ signature is an example of this. Instead of using this observation that depends on κ in our proof, we only construct gadgets such that their signatures are 0 on $P_{\substack{1\,4\\2\,3}}$ for any value of κ.

monochromatic vertices. In particular, any coloring with a vertex that is cyclically colored *xyxy* for different colors x and y does not contribute because $f(P_{\begin{smallmatrix}1&2\\2&1\end{smallmatrix}}) = 0$. □

Remark This result shows that this planar Holant problem is at least as hard as computing the Tutte polynomial at the point $(\kappa + 1, \kappa + 1)$ over planar graphs, which implies #P-hardness. Of course they are equally difficult in the sense that both are #P-complete. In fact, they are more directly related since every 4-regular plane graph is the medial graph of some plane graph.

By Theorem 4.1 and Lemma 4.6, the problem Pl-Holant($\langle 2, 1, 0, 1, 0\rangle$) is #P-hard. We state this as a corollary.

Corollary 4.7 *Suppose* $\kappa \geq 3$ *is the domain size. Let* $\langle 2, 1, 0, 1, 0\rangle$ *be a succinct quaternary signature of type* τ_{color}. *Then* Pl-Holant($\langle 2, 1, 0, 1, 0\rangle$) *is* #P*-hard.*

With this connection established, we can now show that counting edge colorings is #P-hard over planar regular graphs when the number of colors and the regularity parameter coincide.

Theorem 4.8 #κ-EdgeColoring *is* #P*-hard over planar* κ*-regular graphs for all* $\kappa \geq 3$.

Proof Let κ be the domain size of all Holant problems in this proof and let $\langle 2, 1, 0, 1, 0\rangle$ be a succinct quaternary signature of type τ_{color}. We reduce from Pl-Holant($\langle 2, 1, 0, 1, 0\rangle$) to Pl-Holant($\text{AD}_{\kappa,\kappa}$), which denotes the problem of counting edge κ-colorings over planar κ-regular graphs as a Holant problem. Then by Corollary 4.7, we conclude that Pl-Holant($\text{AD}_{\kappa,\kappa}$) is #P-hard.

Consider the gadget in Fig. 5, where the bold edge represents $\kappa - 2$ parallel edges. We assign $\text{AD}_{\kappa,\kappa}$ to both vertices. Up to a nonzero factor of $(\kappa - 2)!$, this gadget has the succinct quaternary signature $f = \langle 0, 1, 1, 0, 0\rangle$ of type τ_{color}. Now consider the recursive construction in Fig. 6. All vertices are assigned the signature f. Let f_s be the succinct quaternary signature of type τ_{color} for the sth gadget of the recursive construction. Then $f_1 = f$ and $f_s = M^s f_0$, where

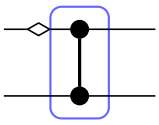


Fig. 5 Quaternary gadget used in the interpolation construction below. All vertices are assigned $\text{AD}_{\kappa,\kappa}$. The *bold edge* represents $\kappa - 2$ parallel edges

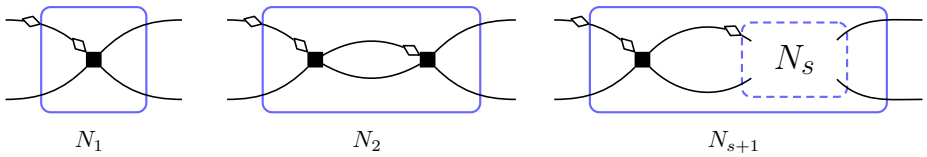


Fig. 6 Recursive construction to interpolate $\langle 2, 1, 0, 1, 0\rangle$. The vertices are assigned the signature of the gadget in Fig. 5

$$M = \begin{bmatrix} 0 & \kappa-1 & 0 & 0 & 0 \\ 1 & \kappa-2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad f_0 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}.$$

The signature f_0 is simply the succinct quaternary signature of type τ_{color} for two parallel edges. We can express M via the Jordan decomposition $M = P\Lambda P^{-1}$, where

$$P = \begin{bmatrix} 1 & 1-\kappa & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

and $\Lambda = \text{diag}(\kappa-1, -1, 1, -1, 1)$. Then for $t = 2s$, we have

$$f_t = P \begin{bmatrix} \kappa-1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}^t P^{-1} \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} = P \begin{bmatrix} x & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} P^{-1} \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} y+1 \\ y \\ 0 \\ 1 \\ 0 \end{bmatrix},$$

where $x = (\kappa-1)^t$ and $y = \frac{x-1}{\kappa}$.

Consider an instance Ω of Pl-Holant($\langle 2, 1, 0, 1, 0\rangle$) on domain size κ. Suppose $\langle 2, 1, 0, 1, 0\rangle$ appears n times in Ω. We construct from Ω a sequence of instances Ω_t of Pl-Holant(f) indexed by t, where $t = 2s$ with $s \geq 0$. We obtain Ω_t from Ω by replacing each occurrence of $\langle 2, 1, 0, 1, 0\rangle$ with the gadget f_t.

As a polynomial in $x = (\kappa-1)^t$, Pl-Holant($\Omega_t; f$) is independent of t and has degree at most n with integer coefficients. Using our oracle for Pl-Holant(f), we can evaluate this polynomial at $n+1$ distinct points $x = (\kappa-1)^{2s}$ for $0 \leq s \leq n$. Then via polynomial interpolation, we can recover the coefficients of this polynomial efficiently. Evaluating this polynomial at $x = \kappa + 1$ (so that $y = 1$) gives the value of Pl-Holant(Ω; $\langle 2, 1, 0, 1, 0\rangle$), as desired. □

Remark For $\kappa = 3$, the interpolation step is actually unnecessary since the succinct signature of f_2 happens to be $\langle 2, 1, 0, 1, 0\rangle$.

When $\kappa = 3$, it is easy to extend Theorem 4.8 by further restricting to simple graphs, i.e., graphs without self-loops or parallel edges.

Theorem 4.9 #3-EdgeColoring *is* #P*-hard over simple planar* 3*-regular graphs.*

Proof By Theorem 4.8, it suffices to efficiently compute the number of edge 3-colorings of a planar 3-regular graph G that might have self-loops and parallel edges. Furthermore, we can assume that G is connected since the number of edge colorings is multiplicative over connected components. If G contains a self-loop, then there are no edge colorings in G, so assume G contains no self-loops. If G also contains no parallel edges, then G is simple and we are done.

Thus, assume that G contains n vertices and parallel edges between some distinct vertices u and v. If u and v are connected by three edges, then this constitutes the whole

graph, which has six edge 3-colorings. Otherwise, u and v are connected by two edges. Then there exist vertices u' and v' such that u and u' are connected by a single edge, v and v' are connected by a single edge, and $u' \neq v'$. In any edge 3-coloring of G, it is easy to see that the edges (u, u') and (v, v') must be assigned the same color. By removing u, v, and their incident edges while adding an edge between u' and v', we have a planar 3-regular graph G' on $n-2$ vertices with half as many edge colorings as G. Then by induction, we can efficiently compute the number of edge 3-colorings in G'. □

In "Appendix 3", we give an alternative proof of Theorem 4.8, which uses the interpolation techniques we develop in Sect. 6. The purpose of "Appendix 3" is to show how a recursive construction in an interpolation proof can be used to form a hypothesis about possible invariance properties. One example of an invariance property is that any planar $\{\mathrm{AD}_{\kappa,\kappa}\}$-gate with a succinct quaternary signature $\langle a, b, c, d, e\rangle$ of type τ_{color} must satisfy $a + c = b + d$ (Lemma 13.1).

4.2 The case $\kappa > r$

Now we consider $\kappa > r \geq 3$. This time, we reduce from the problem of counting vertex κ-colorings over planar graphs. This problem is also #P-hard by the same dichotomy for the Tutte polynomial (Theorem 4.1) since the chromatic polynomial is a specialization the Tutte polynomial.

Proposition 4.10 (Proposition 6.3.1 in [3]) *Let $G = (V, E)$ be a graph. Then $\chi(G; \lambda)$, the chromatic polynomial of G, is expressed as a specialization of the Tutte polynomial via the relation*

$$\chi(G; \lambda) = (-1)^{|V|-k(G)} \lambda^{k(G)}\, \mathrm{T}(G; 1-\lambda, 0),$$

where $k(G)$ is the number of connected components of the graph G.

The first step in the proof is to interpolate every possible binary signature that is domain invariant, which are necessarily symmetric. These signatures have the succinct signature type τ_2.

Lemma 4.11 *Suppose $\kappa \geq 3$ is the domain size and let $x, y \in \mathbb{C}$. If we assign the succinct binary signature $\langle x, y\rangle$ of type τ_2 to every vertex of the recursive construction in Fig. 7, then the corresponding recurrence matrix is $\begin{bmatrix} x & (\kappa-1)y \\ y & x+(\kappa-2)y \end{bmatrix}$ with eigenvalues $x + (\kappa - 1)y$ and $x - y$.*

Proof Let f_ℓ be the signature of the ℓth gadget in this construction. To obtain $f_{\ell+1}$ from f_ℓ, we view f_ℓ as a column vector and multiply it by the recurrence matrix $M = \begin{bmatrix} x & (\kappa-1)y \\ y & x+(\kappa-2)y \end{bmatrix}$.

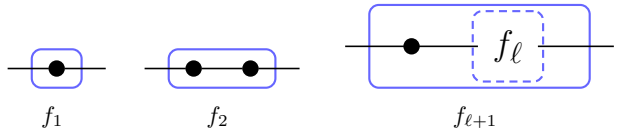


Fig. 7 Recursive construction to interpolate any succinct binary signature of type τ_2. All vertices are assigned the same succinct binary signature of type τ_2

In general, we have $f_\ell = M^\ell f_0$, where f_0 is the initial signature, which is just a single edge and has the succinct binary signature $\langle 1, 0\rangle$ of type τ_2. The (column) eigenvectors of M are $\left[\begin{smallmatrix}1\\1\end{smallmatrix}\right]$ and $\left[\begin{smallmatrix}1-\kappa\\1\end{smallmatrix}\right]$ with eigenvalues $x + (\kappa - 1)y$ and $x - y$, respectively, as claimed. □

The success of interpolation depends on these eigenvalues and the relationship between the recurrence matrix and the initial signature of the construction. To show that the interpolation succeeds, we use a result from [36], the full version of [37]. This result is about interpolating unary signatures on a Boolean domain for planar Holant problems, but the same proof applies equally well for higher domains using a binary recursive construction (like that in Fig. 7) and a succinct signature type with length 2.

Lemma 4.12 (Lemma 4.4 in [36]) *Suppose $\mathcal{F}$ is a set of signatures and τ is a succinct signature type with length* 2. *If there exists an infinite sequence of planar $\mathcal{F}$-gates defined by an initial succinct signature $s \in \mathbb{C}^{2\times 1}$ of type τ and recurrence matrix $M \in \mathbb{C}^{2\times 2}$ satisfying the following conditions,*

1. $\det(M) \neq 0$;
2. $\det([s\ Ms]) \neq 0$;
3. *M has infinite order modulo a scalar;*

then

$$\text{Pl-Holant}(\mathcal{F} \cup \{\langle x, y\rangle\}) \leq_T \text{Pl-Holant}(\mathcal{F}),$$

for any $x, y \in \mathbb{C}$, where $\langle x, y\rangle$ is a succinct binary signature of type τ.

Consider the recursive construction in Fig. 7. To every vertex, we assign the succinct binary signature $\langle x, y\rangle$. Since the initial signature is $s = \langle 1, 0\rangle$, the determinant of the matrix $[s\ Ms]$ is simply y. In order to interpolate all binary succinct signatures of type τ_2, we need to satisfy the second condition of Lemma 4.12, which is $y \neq 0$. However, when $y = 0$, the recurrence matrix is a scalar multiple of the identity matrix, which implies that the eigenvalues are the same. For two-dimensional interpolation using a matrix with a full basis of eigenvectors, as is the case here, the third condition of Lemma 4.12 is equivalent to the ratio of the eigenvalues not being a root of unity. In particular, the eigenvalues cannot be the same. Therefore, when using the recursive construction in Fig. 7, it suffices to satisfy the first and third conditions of Lemma 4.12. We state this as a corollary.

Corollary 4.13 *Suppose $\mathcal{F}$ is a set of signatures. Let $s = \langle 1, 0\rangle$ of type τ_2 be the initial succinct binary signature and let $M \in \mathbb{C}^{2\times 2}$ be the recurrence matrix for some infinite sequence of planar $\mathcal{F}$-gates defined by the recursive construction in Fig. 7. If M satisfies the following conditions,*

1. $\det(M) \neq 0$;
2. *M has infinite order modulo a scalar;*

then

$$\text{Pl-Holant}(\mathcal{F} \cup \{\langle x, y\rangle\}) \leq_T \text{Pl-Holant}(\mathcal{F}),$$

for any $x, y \in \mathbb{C}$, where $\langle x, y\rangle$ is a succinct binary signature of type τ_2.

Lemma 4.14 *Suppose κ is the domain size with $\kappa > r$ for any integer $r \geq 3$, and $x, y \in \mathbb{C}$. Let $\mathcal{F}$ be a signature set containing $\mathrm{AD}_{r,\kappa}$. Then*

$$\text{Pl-Holant}(\mathcal{F} \cup \{\langle x, y\rangle\}) \leq_T \text{Pl-Holant}(\mathcal{F}),$$

where $\langle x, y\rangle$ is a succinct binary signature of type τ_2.

Proof Let $(n)_k = n(n-1)\cdots(n-k+1)$ be the kth falling power of n. Consider the gadget in Fig. 8. We assign $\mathrm{AD}_{r,\kappa}$ to both vertices. The succinct binary signature of type τ_2 for this gadget is $f = \langle(\kappa-1)_{r-1}, (\kappa-2)_{r-1}\rangle$. Up to a nonzero factor of $(\kappa-2)_{r-2}$, we have the signature $f' = \frac{1}{(\kappa-2)_{r-2}} f = \langle \kappa - 1, \kappa - r\rangle$.

Consider the recursive construction in Fig. 7. We assign f' to all vertices. By Lemma 4.11, the eigenvalues of the corresponding recurrence matrix are $(r-1) > 0$ and $(\kappa-1)(\kappa - r + 1) > 0$. Thus, M is nonsingular. Furthermore, the eigenvalues are not equal since $\kappa \notin \{0, r\}$. Therefore, we are done by Corollary 4.13. □

Next we show that $\text{Pl-Holant}(\mathrm{AD}_{r,\kappa})$ is at least as hard as $\text{Pl-Holant}(\mathrm{AD}_{3,\kappa})$. To overcome a difficulty when r is even, we apply the following result, which uses the notion of a planar pairing.

Definition 4.15 *(Definition 11 in* [37]*)* A *planar pairing* in a graph $G = (V, E)$ is a set of edges $P \subset V \times V$ such that P is a perfect matching in the graph $(V, V \times V)$, and the graph $(V, E \cup P)$ is planar.

Lemma 4.16 (Lemma 12 in [37]) *For any planar 3-regular graph G, there exists a planar pairing that can be computed in polynomial time.*

Lemma 4.17 *Suppose κ is the domain size with $\kappa > r$ for any integer $r \geq 3$. Then*

$$\text{Pl-Holant}(\mathrm{AD}_{3,\kappa}) \leq_T \text{Pl-Holant}(\mathrm{AD}_{r,\kappa}).$$

Proof By Lemma 4.14, we can assume that $\langle 1, 1\rangle$ is available. Take $\mathrm{AD}_{r,\kappa}$ and first form $t = \left\lceil \frac{r-4}{2} \right\rceil$ self-loops. Then add a new vertex on each self-loop and assign $\langle 1, 1\rangle$ to each of these new vertices. Up to a nonzero factor of $(\kappa-3)_{2t}$, the resulting signature is $\mathrm{AD}_{3,\kappa}$ if r is odd and $\mathrm{AD}_{4,\kappa}$ if r is even. To reduce from $r = 3$ to $r = 4$, we use a planar pairing, which can be efficiently computed by Lemma 4.16. We add a new vertex on each edge in a planar pairing and assign $\langle 1, 1\rangle$ to each of these new vertices. Then up to a nonzero factor of $\kappa - 3$, the signature at each vertex of the initial graph is effectively $\mathrm{AD}_{3,\kappa}$. □

The succinct binary signature $\langle 1 - \kappa, 1\rangle$ of type τ_2 has a special property. Let u be any constant unary signature, which has a succinct signature of type τ_1. If $\langle 1 - \kappa, 1\rangle$ is connected to u, then the resulting unary signature is identically 0.

This observation is the key to what follows. We use it in the next lemma to achieve what would appear to be an impossible task. The requirements, if duly specified, would result in multiple conditions to be satisfied by nine separate polynomials pertaining to

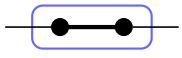


Fig. 8 Binary gadget used in the interpolation construction of Fig. 7. Both vertices are assigned $\mathrm{AD}_{r,\kappa}$, and the *bold edge* represents $r - 1$ parallel edges

some construction in place of the gadget in Fig. 9. And yet we are able to use just one degree of freedom to cause seven of the polynomials to vanish, satisfying most of these conditions. In addition, the other two polynomials are not forgotten. They are nonzero, and their ratio is not a root of unity, which allows interpolation to succeed.

This ability to satisfy a multitude of constraints simultaneously in one magic stroke reminds us of some unfathomably brilliant moves by Bobby Fischer, the chess genius extraordinaire, and so we name this gadget (Fig. 9) the *Bobby Fischer gadget*.

This gadget is the new idea that allows us to prove Theorem 4.20.

Lemma 4.18 *Suppose* $\kappa \geq 3$ *is the domain size and* $a, b \in \mathbb{C}$. *Let* $\mathcal{F}$ *be a signature set containing the succinct ternary signature* $\langle a, b, b\rangle$ *of type* τ_3 *and the succinct binary signature* $\langle 1-\kappa, 1\rangle$ *of type* τ_2. *If* $a \neq b$, *then*

$$\text{Pl-Holant}(\mathcal{F} \cup \{=_4\}) \leq_T \text{Pl-Holant}(\mathcal{F}).$$

Proof Consider the gadget in Fig. 9. We assign $\langle a, b, b\rangle$ to the circle vertices and $\langle 1-\kappa, 1\rangle$ to the square vertex. This gadget has a succinct quaternary signature of type τ_4, which has length 9. However, all but two of the entries in this succinct signature must be 0.

To see this, consider an assignment that assigns different values to the two edges incident to the circle vertex on top. Since the assignment to these two edges differs, the signature $\langle a, b, b\rangle$ contributes a factor of b regardless of the value of its third edge, which is connected to the square vertex assigned $\langle 1-\kappa, 1\rangle$. From the perspective of $\langle 1-\kappa, 1\rangle$, this behavior is equivalent to connecting it to the succinct unary signature $b\langle 1\rangle$ of type τ_1. Thus, the sum over the possible assignments to this third edge is 0. The same argument shows that the two edges incident to the circle vertex on the bottom do not contribute anything to the Holant sum when assigned different values.

Thus, any nonzero contribution to the Holant sum comes from assignments where the top two dangling edges are assigned the same value and the bottom two dangling edges are assigned the same value. There are only two entries that satisfy this requirement in the succinct quaternary signature of type τ_4 for this gadget, which are the entries for $P_{\substack{1\,1\\1\,1}}$ and $P_{\substack{1\,1\\2\,2}}$. To compute those two entries, we use the following trick. Since the two external edges of each circle vertex must be assigned the same value, we think of them as just a single edge. Then the effective succinct binary signature of type τ_2 for the circle vertices is just $\langle a, b\rangle$. By connecting the first $\langle a, b\rangle$ with $\langle 1-\kappa, 1\rangle$, the result is $(a-b)\langle 1-\kappa, 1\rangle$ like it is an eigenvector. Connecting the other copy of $\langle a, b\rangle$ to $(a-b)\langle 1-\kappa, 1\rangle$ gives $(a-b)^2\langle 1-\kappa, 1\rangle$. This computation is expressed via the matrix multiplication $[bJ_\kappa + (a-b)I_\kappa][J_\kappa - \kappa I_\kappa][bJ_\kappa + (a-b)I_\kappa] = (a-b)[J_\kappa - \kappa I_\kappa][bJ_\kappa + (a-b)I_\kappa] = (a-b)^2[J_\kappa - \kappa I_\kappa]$. Thus up to a nonzero factor of $(a-b)^2$, the corresponding succinct quaternary signature of type τ_4 for this gadget is $f = \langle 1-\kappa, 0, 0, 0, 0, 0, 1, 0, 0\rangle$.

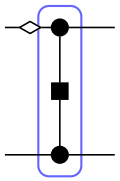


Fig. 9 The Bobby Fischer gadget, which achieves many objectives using only a single degree of freedom

Consider the recursive construction in Fig. 6. We assign f to all vertices. Let f_s be the signature of the sth gadget in this construction. The seven entries that are 0 in the succinct signature of type τ_4 for f are also 0 in the succinct signature of type τ_4 for f_s. Thus, we can express f_s via a succinct signature of type τ_4' with length 2, defined as follows. The first two parts in τ_4' are $P_{\begin{smallmatrix}1&1\\1&1\end{smallmatrix}}$ and $P_{\begin{smallmatrix}1&1\\2&2\end{smallmatrix}}$ from the succinct signature type τ_4. The last part contains all the remaining assignments. Then the succinct signature for f_s of type τ_4' is $M^s f_0$, where $M = \left[\begin{smallmatrix}1-\kappa & 0\\ 0 & 1\end{smallmatrix}\right]$ and $f_0 = \langle 1, 1\rangle$, which is just the succinct signature of type τ_4' for two parallel edges.

Clearly the conditions in Lemma 4.12 hold, so we can interpolate any succinct signature of type τ_4'. In particular, we can interpolate our target signature $=_4$, which is $\langle 1, 0\rangle$ when expressed as a succinct signature of type τ_4'. □

Remark The nine polynomials mentioned before Lemma 4.18 correspond to the nine entries of some quaternary gadget with a succinct signature of type τ_4. In light of Lemma 4.14, this gadget might involve many succinct binary signatures $\langle x, y\rangle$ of type τ_2 for various choices of $x, y \in \mathbb{C}$. Each distinct binary signature provides an additional degree of freedom to these polynomials. Our construction in Fig. 9 only requires one binary signature $\langle x, y\rangle$, and we use our one degree of freedom to set $\frac{x}{y} = 1 - \kappa$.

With the aid of the succinct unary signature $\langle 1\rangle$ of type τ_1 and the succinct binary signature $\langle 0, 1\rangle$ of type τ_2, the assumptions in the previous lemma are sufficient to prove #P-hardness.

Corollary 4.19 *Suppose $\kappa \geq 3$ is the domain size and $a, b \in \mathbb{C}$. Let $\mathcal{F}$ be a signature set containing the succinct ternary signature $\langle a, b, b\rangle$ of type τ_3, the succinct unary signature $\langle 1\rangle$ of type τ_1, and the succinct binary signatures $\langle 1-\kappa, 1\rangle$ and $\langle 0, 1\rangle$ of type τ_2. If $a \neq b$, then* Pl-Holant($\mathcal{F}$) *is* #P*-hard.*

Proof By Lemma 4.18, we have $=_4$. Connecting $\langle 1\rangle$ to $=_4$ gives $=_3$. With $=_3$, we can construct the equality signatures of every arity. Along with the binary disequality signature $\neq_2$, which is the succinct binary signature $\langle 0, 1\rangle$ of type τ_2, we can express the problem of counting the number of vertex κ-colorings over planar graphs. By Proposition 4.10, this is, up to a nonzero factor, the problem of evaluating the Tutte polynomial at $(1-\kappa, 0)$, which is #P-hard by Theorem 4.1. □

Now we can show that counting edge colorings is #P-hard over planar regular graphs when there are more colors than the regularity parameter.

Theorem 4.20 *#κ-*EdgeColoring *is* #P*-hard over planar r-regular graphs for all $\kappa > r \geq 3$.*

Proof By Lemma 4.17, it suffices to consider $r = 3$. By Lemma 4.14, we can assume that any succinct binary signature of type τ_2 is available.

Consider the gadget in Fig. 10. We assign $\mathrm{AD}_{3,\kappa}$ to the circle vertex and $\langle 3-\kappa, 1\rangle$ to the square vertices. By Lemma 11.6, the succinct ternary signature of type τ_3 for this gadget is $f = 2(\kappa-2)\langle -(\kappa-3)(\kappa-1), 1, 1\rangle$.

Now take two edges of $\mathrm{AD}_{3,\kappa}$ and connect them to the two edges of $\langle 1, 1\rangle$. Up to a nonzero factor of $(\kappa-1)(\kappa-2)$, this gadget has the succinct unary signature $\langle 1\rangle$ of type τ_1. Then we are done by Corollary 4.19. □

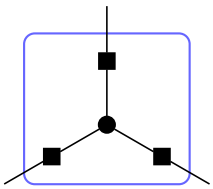


Fig. 10 Local holographic transformation gadget construction for a ternary signature

5 Tractable problems

In the rest of the paper, we adapt and extend our previous proof techniques to obtain a dichotomy for Pl-Holant($\langle a, b, c\rangle$), where $\langle a, b, c\rangle$ is a succinct ternary signature of type τ_3 on domain size $\kappa \geq 3$, for any $a, b, c \in \mathbb{C}$. In this section, we show how to compute a few of these problems in polynomial time.

5.1 Previous dichotomy theorem

There is only one previous dichotomy theorem for higher domain Holant problems. It is a dichotomy for a single symmetric ternary signature on domain size $\kappa = 3$ in the framework of Holant* problems, which means that all unary signatures are assumed to be freely available.

In Theorem 5.1, the notation $f \frown g$ denotes the signature that results from connecting one edge incident to a vertex assigned the signature f to one edge incident to a vertex assigned the signature g. When f and g are both unary signatures, which are represented by vectors, then the resulting 0-ary signature is just a scalar.

Theorem 5.1 (Theorem 3.1 in [22]) *Let f be a symmetric ternary signature on domain size* 3. *Then* Holant*(f) *is* #P*-hard unless f is of one of the following forms, in which case, the problem is computable in polynomial time.*

1. *There exists $\alpha, \beta, \gamma \in \mathbb{C}^3$ that are mutually orthogonal (i.e., $\alpha \frown \beta = \alpha \frown \gamma = \beta \frown \gamma = 0$) and*
$$f = \alpha^{\otimes 3} + \beta^{\otimes 3} + \gamma^{\otimes 3};$$
2. *There exists $\alpha, \beta_1, \beta_2 \in \mathbb{C}^3$ such that $\alpha \frown \beta_1 = \alpha \frown \beta_2 = \beta_1 \frown \beta_1 = \beta_2 \frown \beta_2 = 0$ and*
$$f = \alpha^{\otimes 3} + \beta_1^{\otimes 3} + \beta_2^{\otimes 3};$$
3. *There exists $\beta, \gamma \in \mathbb{C}^3$ and $f_\beta \in (\mathbb{C}^3)^{\otimes 3}$ such that $\beta \neq \mathbf{0}$, $\beta \frown \beta = 0$, $f_\beta \frown \beta = \mathbf{0}$, and*
$$f = f_\beta + \beta^{\otimes 2} \otimes \gamma + \beta \otimes \gamma \otimes \beta + \gamma \otimes \beta^{\otimes 2}.$$

Some domain invariant signatures are tractable by Theorem 5.1.

Corollary 5.2 *Suppose the domain size is* 3 *and $a, b, \lambda \in \mathbb{C}$. Let f be a succinct ternary signature of type τ_3. Then* Holant(f) *is computable in polynomial time when f has one of the following forms:*

1. $f = \lambda\langle 1, 0, 0\rangle = \lambda\left[(1, 0, 0)^{\otimes 3} + (0, 1, 0)^{\otimes 3} + (0, 0, 1)^{\otimes 3}\right]$;
2. $f = 3\lambda\langle -5, -2, 4\rangle = \lambda\left[(1, -2, -2)^{\otimes 3} + (-2, 1, -2)^{\otimes 3} + (-2, -2, 1)^{\otimes 3}\right]$;

3. $f = \langle a, b, a\rangle = \frac{a+2b}{3}(1,1,1)^{\otimes 3} + \frac{a-b}{3}\left[(1,\omega,\omega^2)^{\otimes 3} + (1,\omega^2,\omega)^{\otimes 3}\right]$,

where ω is a primitive third root of unity.

In Corollary 5.2, form 1 is the ternary equality signature $=_3$, which is trivially tractable for any domain size. Then form 2 is just form 1 after a holographic transformation by the matrix $T = \begin{bmatrix} 1 & -2 & -2 \\ -2 & 1 & -2 \\ -2 & -2 & 1 \end{bmatrix}$, which is orthogonal after scaling by $\frac{1}{3}$. This is an example of two problems that must have the same complexity by Theorem 3.3.

The tractability of these two problems for higher domain sizes is stated in the following corollary.

Corollary 5.3 *Suppose $\kappa \geq 3$ is the domain size and $\lambda \in \mathbb{C}$. Let f be a succinct ternary signature of type τ_3. Then* Holant(f) *is computable in polynomial time if f has one of the following forms:*

1. $f = \lambda\langle 1,0,0\rangle$;
2. $f = \lambda T^{\otimes 3}\langle 1,0,0\rangle = \lambda\kappa\langle \kappa^2 - 6\kappa + 4, -2(\kappa-2), 4\rangle$,

where $T = \kappa I_\kappa - 2J_\kappa$.

Note that $T = \kappa I_\kappa - 2J_\kappa$ is an orthogonal matrix after scaling by $\frac{1}{\kappa}$.

5.2 Affine signatures

Let ω be a primitive third root of unity. Consider the ternary signature $f(x,y,z)$ with succinct ternary signature $\langle 1,0,c\rangle$ of type τ_3 on domain size 3, where $c^3 = 1$. Its support is an affine subspace of $\mathbb{Z}_3$ defined by $x+y+z=0$. Furthermore, consider the quadratic polynomial $q_c(x,y,z) = \lambda_c(xy + xz + yz)$, where $\lambda_1 = 0$, $\lambda_\omega = 2$, and $\lambda_{\omega^2} = 1$. Then $\omega^{q_c(x,y,z)}$ agrees with f when $x+y+z=0$. This function f is an example of a ternary domain affine signature.

Definition 5.4 A k-ary function $f(x_1,\ldots,x_k)$ is *affine* on domain size 3 if it has the form

$$\lambda \cdot \chi_{Ax=0} \cdot e^{\frac{2\pi i}{3}q(x)},$$

where $\lambda \in \mathbb{C}$, $x = (x_1, x_2, \ldots, x_k, 1)^{\mathrm{T}}$, A is a matrix over $\mathbb{Z}_3$, $q(x) \in \mathbb{Z}_3$ is a quadratic polynomial, and χ is a 0-1 indicator function such that $\chi_{Ax=0}$ is 1 iff $Ax = 0$. We use $\mathscr{A}$ to denote the set of all affine functions.

The ternary domain affine signatures are tractable just as those in the Boolean domain are [10].

Lemma 5.5 *Suppose the domain size is* 3. *Then* Holant$(\mathscr{A})$ *is computable in polynomial time.*

Proof Given an instance of Holant$(\mathscr{A})$, the output can be expressed as the summation of a single function $F = \chi_{Ax=0} \cdot e^{\frac{2\pi i}{3}q(x_1,x_2,\ldots,x_k)}$ since $\mathscr{A}$ is closed under multiplication. In polynomial time, we can solve the linear system $Ax = 0$ over $\mathbb{Z}_3$ and decide whether it is feasible. If the linear system is infeasible, then the function is the identically 0 function, so the output is just 0.

Otherwise, the linear system is feasible (including possibly vacuous). Without loss of generality, we can assume that $y_1, y_2, \ldots, y_s$ are independent variables over $\mathbb{Z}_3$ while all oth-

ers are dependent variables, where $0 \le s \le k$. Each dependent variable can be expressed by an affine linear form of $y_1, y_2, \ldots, y_s$. We can substitute for all of the dependent variables in $q(x_1, x_2, \ldots, x_k)$, which gives a new quadratic polynomial $q'(y_1, y_2, \ldots, y_s)$. Thus, we have

$$\sum_{x_1,\ldots,x_k \in \mathbb{Z}_3} \chi_{Ax=0} \cdot e^{\frac{2\pi i}{3} q(x_1,x_2,\ldots,x_k)} = \sum_{y_1,\ldots,y_s \in \mathbb{Z}_3} e^{\frac{2\pi i}{3} q'(y_1,y_2,\ldots,y_s)}. \tag{3}$$

Then the right-hand side of (3) is computable in polynomial time by Theorem 1 in [24]. □

After multiplying the function $\langle 1, 0, c\rangle$ by a scalar, we obtain the succinct ternary signature $\langle a, 0, c\rangle$ of type τ_3 such that $a^3 = c^3$. Since undergoing an orthogonal transformation does not change the complexity of the problem by Theorem 3.3, we obtain the following corollary of the previous result.

Corollary 5.6 *Suppose the domain size is* 3 *and* $a, c \in \mathbb{C}$. *Let* $T \in \mathbf{O}_3(\mathbb{C})$ *and let* $\langle a, 0, c\rangle$ *be a succinct ternary signature of type* τ_3. *If* $a^3 = c^3$, *then* $\text{Holant}(T^{\otimes 3}\langle a, 0, c\rangle)$ *is computable in polynomial time.*

For domain size 3, the only orthogonal matrix T such that $T^{\otimes 3}\langle a, b, c\rangle$ is still a succinct ternary signature of type τ_3 is $\pm\frac{1}{3}\begin{bmatrix} 1 & -2 & -2 \\ -2 & 1 & -2 \\ -2 & -2 & 1 \end{bmatrix}$. However, the tractability in Corollary 5.6 holds for any orthogonal matrix T.

We introduce another affine signature. It can be considered as a signature of arity 4 on the Boolean domain. When placed in a planar signature grid, its input variables are listed in a cyclic order x_1, x_2, y_2, y_1 counterclockwise. We then consider it as a binary signature on domain size 4, where the two variables (x_1, x_2) and (y_1, y_2) range over the four values in $\{0, 1\}^2$. Notice the reversal of the order y_2, y_1. This is to allow a planar connection between these signatures. We list its values as the matrix $H_4 = \begin{bmatrix} 1 & -1 & -1 & -1 \\ -1 & 1 & -1 & -1 \\ -1 & -1 & 1 & -1 \\ -1 & -1 & -1 & 1 \end{bmatrix}$, which is an Hadamard matrix, where the row index is (x_1, x_2) and the column index is (y_1, y_2), both ordered lexicographically. A closed-form expression showing that this is an affine signature on the Boolean domain is $f(x_1, x_2, y_2, y_1) = (-1)^{q(x_1,x_2,y_1,y_2)}$, where q is the quadratic polynomial

$$q(x_1, x_2, y_1, y_2) = x_1 + x_2 + x_1x_2 + y_1 + y_2 + y_1y_2 + x_1y_2 + x_2y_1. \tag{4}$$

As a binary signature on domain size 4, f has the succinct signature $\langle 1, -1\rangle$ of type τ_2. The fact that f is an affine signature on the Boolean domain shows that the Holant problem defined by f on domain size 4 is tractable. This follows from Theorem 1.4 in [24], or the more general graph homomorphism dichotomy theorems [12,34].

We are interested in this problem because its tractability implies the tractability of a set of problems defined by a succinct ternary signature of type τ_3.

Lemma 5.7 *Suppose the domain size is* 4 *and* $\lambda, \mu \in \mathbb{C}$. *Let* $\langle \mu^2, 1, \mu\rangle$ *be a succinct ternary signature of type* τ_3. *If* $\mu = -1 + \varepsilon 2i$ *with* $\varepsilon = \pm 1$, *then* $\text{Holant}(\lambda\langle \mu^2, 1, \mu\rangle)$ *is computable in polynomial time.*

Proof Let $T = \frac{1}{2}\begin{bmatrix} x & y & y & y \\ y & x & y & y \\ y & y & x & y \\ y & y & y & x \end{bmatrix}$, where $x = -3 - \varepsilon i$ and $y = 1 - \varepsilon i$. Then up to a factor of λ^n on graphs with n vertices, the output of $\text{Holant}(\lambda\langle \mu^2, 1, \mu\rangle)$ is the same as the output for

$$\begin{aligned}\text{Holant}(\langle \mu^2, 1, \mu\rangle) &= \text{Holant}(\langle -3-\varepsilon 4i, 1, -1+\varepsilon 2i\rangle)\\ &\equiv_T \text{Holant}(=_2 \mid T^{\otimes 3}(=_3))\\ &= \text{Holant}((=_2)T^{\otimes 2} \mid =_3)\\ &= \text{Holant}(2\langle 1, -1\rangle \mid =_3)\\ &\leq_T \text{Holant}(\langle 1, -1\rangle \mid \{=_k \mid k \in \mathbb{Z}^+\}).\end{aligned}$$

Since $\text{Holant}(\langle 1, -1\rangle \mid \{=_k \mid k \in \mathbb{Z}^+\})$ is the Holant expression for the graph homomorphism problem defined by the Hadamard matrix $\begin{bmatrix} 1 & -1 & -1 & -1 \\ -1 & 1 & -1 & -1 \\ -1 & -1 & 1 & -1 \\ -1 & -1 & -1 & 1 \end{bmatrix}$, we can finish the proof by applying the dichotomy theorems for symmetric matrices in [12,34]. For example, this problem is tractable by Theorem 1.2 in [34] (see also [24]), where the quadratic representation is $q(x_1, x_2, y_1, y_2)$ from (4). □

We restate this lemma as a simple corollary for later convenience.

Corollary 5.8 *Suppose the domain size is* 4 *and* $a, b, c \in \mathbb{C}$. *Let* $\langle a, b, c\rangle$ *be a succinct ternary signature of type* τ_3. *If* $a+5b+2c = 0$ *and* $5b^2+2bc+c^2 = 0$, *then* $\text{Holant}(\langle a, b, c\rangle)$ *is computable in polynomial time.*

Proof Since $a = -5b - 2c$ and $b = \frac{1}{5}(-1 \pm 2i)c$, after scaling by $\mu = -1 \mp 2i$, we have $\mu\langle a, b, c\rangle = c\langle \mu^2, 1, \mu\rangle$ and are done by Lemma 5.7. □

6 An interpolation result

The goal of this section is to generalize an interpolation result from [21], which we rephrase using our notion of a succinct signature (cf. Lemma 4.12).

Theorem 6.1 (Theorem 3.5 in [21]) *Suppose* $\mathcal{F}$ *is a set of signatures and* τ *is a succinct signature type with length* 3. *If there exists an infinite sequence of planar* $\mathcal{F}$*-gates defined by an initial succinct signature* $s \in \mathbb{C}^{3\times 1}$ *of type* τ *and a recurrence matrix* $M \in \mathbb{C}^{3\times 3}$ *with eigenvalues* α, β, *and* γ *satisfying the following conditions:*

1. $\det(M) \neq 0$;
2. *s is not orthogonal to any row eigenvector of M;*
3. *for all* $(i, j, k) \in \mathbb{Z}^3 - \{(0, 0, 0)\}$ *with* $i + j + k = 0$, *we have* $\alpha^i\beta^j\gamma^k \neq 1$;

then

$$\text{Pl-Holant}(\mathcal{F} \cup \{f\}) \leq_T \text{Pl-Holant}(\mathcal{F}),$$

for any succinct ternary signature f of type τ.

Our generalization of this result is designed to relax the second condition so that s can be orthogonal to some row eigenvectors of M. Suppose r is a row eigenvector of M, with eigenvalue λ, that is orthogonal to s (i.e., the dot product $r \cdot s$ is 0). Consider $M^k s$, the kth signature in the infinite sequence defined by M and s. This signature is also orthogonal to r since $r \cdot M^k s = \lambda^k r \cdot s = 0$. We do not know of any way of interpolating a signature using this infinite sequence that is not also orthogonal to r. On the other hand, we would like to interpolate those signatures that do satisfy this orthogonality condition. Our interpolation result gives a sufficient condition to achieve this.

We assume our n-by-n matrix M is diagonalizable, i.e., it has n linearly independent (row and column) eigenvectors. We do not assume that M necessarily has n distinct eigenvalues (although this would be a sufficient condition for it to be diagonalizable). The relaxation of the second condition is that, for some positive integer ℓ, the initial signature s is *not* orthogonal to exactly ℓ of these linearly independent row eigenvectors of M. To satisfy this condition, we use a two-step approach. First, we explicitly exhibit $n-\ell$ linearly independent row eigenvectors of M that are orthogonal to s. Then we use the following lemma to show that the remaining row eigenvectors of M are not orthogonal to s. The justification for this approach is that the eigenvectors orthogonal to s are often simple to express while the eigenvectors not orthogonal to s tend to be more complicated.

Lemma 6.2 *For $n \in \mathbb{Z}^+$, let $s \in \mathbb{C}^{n\times 1}$ be a vector and let $M \in \mathbb{C}^{n\times n}$ be a diagonalizable matrix. If* $\text{rank}([s\ Ms\ \ldots\ M^{n-1}s]) \geq \ell$, *then for any set of n linearly independent row eigenvectors, s is not orthogonal to at least ℓ of them.*

Proof Since M is diagonalizable, it has n linearly independent eigenvectors. Suppose for a contradiction that there exists a set of n linearly independent row eigenvectors of M such that s is orthogonal to $t > n-\ell$ of them. Let $N \in \mathbb{C}^{t\times n}$ be the matrix whose t rows are the row eigenvectors of M that are orthogonal to s. Then $N[s\ Ms\ \ldots\ M^{n-1}s]$ is the zero matrix. From this, it follows that $\text{rank}([s\ Ms\ \ldots\ M^{n-1}s]) < \ell$, a contradiction. □

The third condition of Theorem 6.1 is also known as the lattice condition.

Definition 6.3 Fix some $\ell \in \mathbb{N}$. We say that $\lambda_1, \lambda_2, \ldots, \lambda_\ell \in \mathbb{C} - \{0\}$ satisfy the *lattice condition* if for all $x \in \mathbb{Z}^\ell - \{\mathbf{0}\}$ with $\sum_{i=1}^{\ell} x_i = 0$, we have $\prod_{i=1}^{\ell} \lambda_i^{x_i} \neq 1$.

When $\ell \geq 3$, we use Galois theory to show that the lattice condition is satisfied. The idea is that the lattice condition must hold if the Galois group of the polynomial, whose roots are the λ_i's, is large enough. In [21], for the special case $n = \ell = 3$, it was shown that the roots of most cubic polynomials satisfy the lattice condition using this technique.

Lemma 6.4 (Lemma 5.2 in [21]) *Let $f(x) \in \mathbb{Q}[x]$ be an irreducible cubic polynomial. Then the roots of $f(x)$ satisfy the lattice condition iff $f(x)$ is not of the form $ax^3 + b$ for some $a, b \in \mathbb{Q}$.*

In the following lemma, we show that if the Galois group for a polynomial of degree n is one of the two largest possible groups, S_n or A_n, then its roots satisfy the lattice condition provided these roots do not all have the same complex norm.

Lemma 6.5 *Let f be a polynomial of degree $n \geq 2$ with rational coefficients. If the Galois group of f over $\mathbb{Q}$ is S_n or A_n and the roots of f do not all have the same complex norm, then the roots of f satisfy the lattice condition.*

Proof We consider A_n since the same argument applies to $S_n \supset A_n$. For $1 \leq i \leq n$, let a_i be the roots of f such that $|a_1| \leq \cdots \leq |a_n|$. By assumption, as least one of these inequalities is strict. Suppose for a contradiction that these roots fail to satisfy the lattice condition. This means there exists $x \in \mathbb{Z}^n - \{\mathbf{0}\}$ satisfying $\sum_{i=1}^{n} x_i = 0$ such that

$$a_1^{x_1} \cdots a_n^{x_n} = 1. \tag{5}$$

Since x is not all 0, it must contain some positive entries and some negative entries. We can rewrite (5) as $b_1^{y_1} \cdots b_s^{y_s} = c_1^{z_1} \cdots c_t^{z_t}$, where $s, t \geq 1$, $b_1, \ldots, b_s, c_1, \ldots, c_t$ are $s + t$ distinct members from $\{a_1, \ldots, a_n\}$, $y_i > 0$ for $1 \leq i \leq s$, $z_i > 0$ for $1 \leq i \leq t$, and $y_1 + \cdots + y_s = z_1 + \cdots + z_t$. We omit factors in (5) with exponent 0.

If $n = 2$, then $s = t = 1$ and $|b_1| = |c_1|$. This is a contradiction to the assumption that roots of f do not all have the same complex norm. Otherwise, assume $n \geq 3$. If $s = t = 1$, then $|b_1| = |c_1|$ again. We apply 3-cycles from A_n to conclude that all roots of f have the same complex norm, a contradiction. Otherwise, $s + t > 2$. Without loss of generality, suppose $s \geq t$, which implies $s \geq 2$. Pick $j \in \{0, \ldots, n - s - t\}$ such that $|a_{j+1}| \leq \cdots \leq |a_{j+s+t}|$ contains a strict inequality. We permute the roots so that $b_i = a_{j+i}$ for $1 \leq i \leq s$ and $c_i = a_{j+s+i}$ for $1 \leq i \leq t$ (or possibly swapping b_1 and b_2 if necessary to ensure the permutation is in A_n). Then taking the complex norm of both sides gives a contradiction. □

Remark This result can simplify the interpolation arguments in [21]. Since each of their cubic polynomials is irreducible, the corresponding Galois groups are transitive subgroups of S_3, namely S_3 or A_3 (and in fact by inspection, they are all S_3). Then Lemma 4.5 from [44] (the full version of [43]) shows that the eigenvalues of these polynomials do not all have the same complex norm. Thus, the roots of all polynomials exhibited in [21] satisfy the lattice condition by Lemma 6.5.

In the current paper, we apply Lemma 6.5 to an infinite family of quintic polynomials that we encounter in Sect. 7. If the polynomials are irreducible, then we are able to apply this lemma. Unfortunately, we are unable to show that all these polynomials are irreducible and thus also have to consider the possible ways in which they could factor. Nevertheless, we are still able to show that all these polynomials satisfy the lattice condition.

To conclude, we state and prove our new interpolation result.

Lemma 6.6 *Suppose $\mathcal{F}$ is a set of signatures and τ is a succinct signature type with length $n \in \mathbb{Z}^+$. If there exists an infinite sequence of planar $\mathcal{F}$-gates defined by an initial succinct signature $s \in \mathbb{C}^{n \times 1}$ of type τ and a recurrence matrix $M \in \mathbb{C}^{n \times n}$ satisfying the following conditions,*

1. *M is diagonalizable with n linearly independent eigenvectors;*
2. *s is not orthogonal to exactly ℓ of these linearly independent row eigenvectors of M with eigenvalues $\lambda_1, \ldots, \lambda_\ell$;*
3. *$\lambda_1, \ldots, \lambda_\ell$ satisfy the lattice condition;*

then

$$\text{Pl-Holant}(\mathcal{F} \cup \{f\}) \leq_T \text{Pl-Holant}(\mathcal{F})$$

for any succinct signature f of type τ that is orthogonal to the $n - \ell$ of these linearly independent eigenvectors of M to which s is also orthogonal.

Proof Let $\lambda_1, \ldots, \lambda_n$ be the n eigenvalues of M, with possible repetition. Since M is diagonalizable, we can write M as $T \Lambda T^{-1}$, where Λ is the diagonal matrix $\begin{bmatrix} B_1 & \mathbf{0} \\ \mathbf{0} & B_2 \end{bmatrix}$ with $B_1 = \text{diag}(\lambda_1, \ldots, \lambda_\ell)$ and $B_2 = \text{diag}(\lambda_{\ell+1}, \ldots, \lambda_n)$. Notice that the columns of T are the

column eigenvectors of M and the rows of T^{-1} are the row eigenvectors of M. Let $\mathbf{t}_i$ be the ith column T and let $T^{-1}s = [\alpha_1 \ \dots \ \alpha_n]^{\mathrm{T}}$. Then $\alpha_i \neq 0$ for $1 \leq i \leq \ell$ and $\alpha_i = 0$ for $\ell < i \leq n$, since s is not orthogonal to exactly the first ℓ row eigenvectors of M.

Now we can write

$$M^k s = T \begin{bmatrix} B_1^k & \mathbf{0} \\ \mathbf{0} & B_2^k \end{bmatrix} T^{-1} s = T \begin{bmatrix} B_1^k & \mathbf{0} \\ \mathbf{0} & B_2^k \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_\ell \\ 0 \\ \vdots \\ 0 \end{bmatrix} = T \operatorname{diag}(\alpha_1 \lambda_1^k, \dots, \alpha_\ell \lambda_\ell^k, 0, \dots, 0)$$

$$= T \operatorname{diag}(\alpha_1, \dots, \alpha_\ell, 0, \dots, 0) \begin{bmatrix} \lambda_1^k \\ \vdots \\ \lambda_\ell^k \\ 0 \\ \vdots \\ 0 \end{bmatrix} = [\alpha_1 \mathbf{t}_1, \dots, \alpha_\ell \mathbf{t}_\ell, \mathbf{0}, \dots, \mathbf{0}] \begin{bmatrix} \lambda_1^k \\ \vdots \\ \lambda_\ell^k \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

For $1 \leq i \leq \ell$, let $\mathbf{t}'_i = \alpha_i \mathbf{t}_i$. Both the columns of T and the rows of T^{-1} are linearly independent. From $T^{-1}T = I_m$, we see that $\mathbf{t}_i$ for $1 \leq i \leq \ell$ is orthogonal to the last $n - \ell$ rows of T^{-1}. Thus $\operatorname{span}\{\mathbf{t}_1, \dots, \mathbf{t}_\ell\} = \operatorname{span}\{\mathbf{t}'_1, \dots, \mathbf{t}'_\ell\}$ is precisely the space of vectors orthogonal to the last $n - \ell$ rows of T^{-1}.

Consider an instance Ω of Pl-Holant($\mathcal{F} \cup \{f\}$). Let V_f be the subset of vertices assigned f with $n_f = |V_f|$. Since f is orthogonal to any row eigenvector of M to which s is also orthogonal, we have $f \in \operatorname{span}\{\mathbf{t}'_1, \dots, \mathbf{t}'_\ell\}$. Let $f = \beta_1 \mathbf{t}'_1 + \cdots + \beta_\ell \mathbf{t}'_\ell$. Then Pl-Holant($\Omega; \mathcal{F} \cup \{f\}$) is a homogeneous polynomial in the β_i's of total degree n_f. For $y = (y_1, \dots, y_\ell) \in \mathbb{N}^\ell$, let c_y be the coefficient of $\beta_1^{y_1} \cdots \beta_\ell^{y_\ell}$ in Pl-Holant($\Omega; \mathcal{F} \cup \{f\}$) so that

$$\text{Pl-Holant}(\Omega; \mathcal{F} \cup \{f\}) = \sum_{y_1 + \cdots + y_\ell = n_f} c_y \beta_1^{y_1} \cdots \beta_\ell^{y_\ell}.$$

We construct from Ω a sequence of instances Ω_k of Pl-Holant($\mathcal{F}$) indexed by $k \in \mathbb{N}$. We obtain Ω_k from Ω by replacing each occurrence of f with $M^k s$, for $k \geq 0$. Then

$$\text{Pl-Holant}(\Omega_k; \mathcal{F}) = \sum_{y_1 + \cdots + y_\ell = n_f} c_y \left(\lambda_1^{y_1} \cdots \lambda_\ell^{y_\ell}\right)^k.$$

Note that, crucially, the same c_y coefficients appear. We treat this as a linear system with the c_y's as the unknowns. The coefficient matrix is a Vandermonde matrix of order $\binom{n_f + \ell - 1}{\ell - 1}$, which is polynomial in n_f and thus polynomial in the size of Ω. It is nonsingular if every $\lambda_1^{y_1} \cdots \lambda_\ell^{y_\ell}$ is distinct, which is indeed the case since $\lambda_1, \dots, \lambda_\ell$ satisfy the lattice condition.

Therefore, we can solve for the c_y's in polynomial time and compute Pl-Holant($\Omega; \mathcal{F} \cup \{f\}$). □

Remark When restricted to $n = \ell = 3$, this proof is simpler than the one given in [21] for Theorem 6.1 due to our implicit use of a local holographic transformation (i.e., the writing of f as a linear combination of $\mathbf{t}'_1, \dots, \mathbf{t}'_\ell$ and expressing Pl-Holant($\Omega; \mathcal{F} \cup \{f\}$) in terms of this).

7 Puiseux series, Siegel's theorem, and Galois theory

We prove our main dichotomy theorem in three stages. This section covers the last stage, which assumes that all succinct binary signatures of type τ_2 are available. Among

the ways we utilize this assumption is to build the gadget known as a local holographic transformation (see Fig. 11), which is the focus of Sect. 7.1. Then in Sect. 7.2, our efforts are largely spent, proving that a certain interpolation succeeds. To that end, we employ Galois theory aided by an effective version of Siegel's theorem for a specific algebraic curve, which is made possible by analyzing Puiseux series expansions.

We define the following expressions which appear throughout the rest of the paper:

$$\mathfrak{A} = a - 3b + 2c; \tag{6}$$

$$\mathfrak{B} = \mathfrak{A} + \kappa(b - c) = a + (\kappa - 3)b - (\kappa - 2)c; \quad \text{and} \tag{7}$$

$$\mathfrak{C} = \mathfrak{B} + \kappa[2b + (\kappa - 2)c] = a + 3(\kappa - 1)b + (\kappa - 1)(\kappa - 2)c. \tag{8}$$

7.1 Constructing a special ternary signature

We construct one of two special ternary signatures. Either we construct a signature of the form $\langle a, b, b\rangle$ with $a \neq b$ and can finish the proof with Corollary 4.19 or we construct $\langle 3(\kappa - 1), \kappa - 3, -3\rangle$. With this latter signature, we can interpolate the weighted Eulerian partition signature.

A key step in our dichotomy theorem occurred back in Sect. 4.2 through Lemma 4.18 with the Bobby Fischer gadget. To apply this lemma, we need to construct a gadget with a succinct ternary signature of type τ_3 such that the last two entries are equal and different from the first. This is the goal of the next lemma, which assumes $\mathfrak{B} \neq 0$. We will determine the complexity of the case $\mathfrak{B} = 0$ in Corollary 8.4 without using the results from this section.

Lemma 7.1 *Suppose $\kappa \geq 3$ is the domain size and $a, b, c \in \mathbb{C}$. Let $\mathcal{F}$ be a signature set containing the succinct ternary signature $\langle a, b, c\rangle$ of type τ_3 and the succinct binary signature $\langle x, y\rangle$ of type τ_2 for all $x, y \in \mathbb{C}$. If $\mathfrak{A}\mathfrak{B} \neq 0$, then there exist $a', b' \in \mathbb{C}$ satisfying $a' \neq b'$ such that*

$$\text{Pl-Holant}(\mathcal{F} \cup \{\langle a', b', b'\rangle\}) \leq_T \text{Pl-Holant}(\mathcal{F}),$$

where $\langle a', b', b'\rangle$ is a succinct ternary signature of type τ_3.

Proof Consider the gadget in Fig. 11. We assign $\langle a, b, c\rangle$ to the circle vertex and $\langle x, y\rangle$ to the square vertices for some $x, y \in \mathbb{C}$ of our choice, to be determined shortly. By Lemma 11.6, the succinct ternary signature of type τ_3 for the resulting gadget is $\langle a', b', c'\rangle$, where

$$a' - b' = (x - y)^2[2\mathfrak{D} + \mathfrak{A}(x - y)] \quad \text{and} \quad b' - c' = (x - y)^2\mathfrak{D}$$

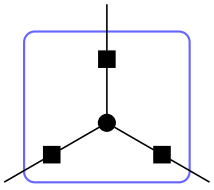


Fig. 11 Local holographic transformation gadget construction for a ternary signature

with $\mathfrak{D} = (b-c)(x-y) + \mathfrak{B}y$. We pick $x = \mathfrak{B} + y$ and $y = -(b-c)$ so that $\mathfrak{D} = 0$ and thus $b' - c' = 0$. Then the first difference simplifies to $a' - b' = \mathfrak{A}\mathfrak{B}^3 \neq 0$. This signature has the desired properties, so we are done. □

The previous proof fails when $\mathfrak{A} = 0$ because such signatures are invariant set-wise under this type of local holographic transformation. With the exception of a single point, we can use this same gadget construction to reduce between any two of these points.

Lemma 7.2 *Suppose $\kappa \geq 3$ is the domain size and $b, c, s, t \in \mathbb{C}$. Let $\mathcal{F}$ be a signature set containing the succinct ternary signature $\langle 3b-2c, b, c\rangle$ of type τ_3 and the succinct binary signature $\langle x, y\rangle$ of type τ_2 for all $x, y \in \mathbb{C}$. If $b \neq c$, $3b + (\kappa-3)c \neq 0$, and $3s + (\kappa-3)t \neq 0$, then*

$$\text{Pl-Holant}(\mathcal{F} \cup \{\langle 3s-2t, s, t\rangle\}) \leq_T \text{Pl-Holant}(\mathcal{F}),$$

where $\langle 3s-2t, s, t\rangle$ is a succinct ternary signature of type τ_3.

Proof Consider the gadget in Fig. 11. We assign $\langle 3b-2c, b, c\rangle$ to the circle vertex and $\langle x, y\rangle$ to the square vertices for some $x, y \in \mathbb{C}$ of our choice, to be determined shortly. By Lemma 11.6, the signature of this gadget is $f = [x + (\kappa-1)y]\langle 3\hat{b} - 2\hat{c}, \hat{b}, \hat{c}\rangle$, where

$$\hat{b} = bx^2 + 2[2b + (\kappa-3)c]xy + [(3\kappa-5)b + (\kappa^2 - 5\kappa + 6)c]y^2 \qquad \text{and}$$
$$\hat{c} = cx^2 + 2[3b + (\kappa-4)c]xy + [(3\kappa-6)b + (\kappa^2 - 5\kappa + 7)c]y^2.$$

We note that the difference $\hat{b} - \hat{c}$ nicely factors as

$$\hat{b} - \hat{c} = (b-c)(x-y)^2.$$

We pick $x = y + \sqrt{s-t}$ so that $\hat{b} - \hat{c} = (b-c)(s-t)$ is the desired difference $s-t$ up to a nonzero factor of $b-c$. Then we want to set $\hat{c}$ to be $(b-c)t$. With $x = y + \sqrt{s-t}$, we can simplify $(b-c)t - \hat{c}$ to

$$(b-c)t - \hat{c} = -\kappa[3b + (\kappa-3)c]y^2 - 2\sqrt{s-t}[3b + (\kappa-3)c]y + bt - cs. \tag{9}$$

Since $\kappa[3b+(\kappa-3)c] \neq 0$, (9) is a nontrivial quadratic polynomial in y, so we can set y such that this expression vanishes. Then the signature is $f = (b-c)[x + (\kappa-1)y]\langle 3s-2t, s, t\rangle$. It remains to check that $x + (\kappa-1)y \neq 0$.

If $x + (\kappa-1)y = 0$, then $y = -\frac{\sqrt{s-t}}{\kappa}$. However, plugging this into (9) gives $\frac{(b-c)[3s+(\kappa-3)t]}{\kappa} \neq 0$, so $x + (\kappa-1)y$ is indeed nonzero. □

If $\mathfrak{A} = 0$ and $3b + (\kappa-3)c = 0$, then $-3\langle a, b, c\rangle$ simplifies to $c\langle 3(\kappa-1), \kappa-3, -3\rangle$, which is a failure condition of the previous lemma. The reason is that this signature is pointwise invariant under such local holographic transformations. However, a different ternary construction can reach this point.

Lemma 7.3 *Suppose $\kappa \geq 3$ is the domain size and $b, c \in \mathbb{C}$. Let $\mathcal{F}$ be a signature set containing the succinct ternary signature $\langle 3b-2c, b, c\rangle$ of type τ_3 and the succinct binary signature $\langle x, y\rangle$ of type τ_2 for every $x, y \in \mathbb{C}$. If $b \neq c$, then*

$$\text{Pl-Holant}(\mathcal{F} \cup \{\langle 3(\kappa-1), \kappa-3, -3\rangle\}) \leq_T \text{Pl-Holant}(\mathcal{F}),$$

where $\langle 3(\kappa-1), \kappa-3, -3\rangle$ is a succinct ternary signature of type τ_3.

Proof If $3b + (\kappa - 3)c = 0$, then up to a nonzero factor of $\frac{-c}{3}$, $\langle 3b - 2c, b, c\rangle$ is already the desired signature. Otherwise, $3b + (\kappa - 3)c \neq 0$. By Lemma 7.2, we have $\langle 3s - 2t, s, t\rangle$ for any $s, t \in \mathbb{C}$ satisfying $3s + (\kappa - 3)t \neq 0$.

Consider the gadget in Fig. 12. We assign $\langle 3s - 2t, s, t\rangle$ to vertices for some $s, t \in \mathbb{C}$ satisfying $3s + (\kappa - 3)t \neq 0$ of our choice, to be determined shortly. By Lemma 11.4, the signature of this gadget is $\langle 3s' - 2t', s', t'\rangle$, where

$$s' = (5\kappa + 14)s^3 + (\kappa^2 + 9\kappa - 42)s^2t + (7\kappa^2 - 33\kappa + 42)st^2 + (\kappa - 2)(\kappa^2 - 6\kappa + 7)t^3,$$

and

$$t' = (\kappa + 14)s^3 + 21(\kappa - 2)s^2t + 3(3\kappa^2 - 15\kappa + 14)st^2 + (\kappa^3 - 9\kappa^2 + 23\kappa - 14)t^3.$$

It suffices to pick s and t satisfying $3s + (\kappa - 3)t \neq 0$ such that $s' = \kappa - 3$ and $t' = -3$ up to a common nonzero factor.

We note that the difference $s' - t'$ factors as

$$s' - t' = \kappa(s - t)^2[4s + (\kappa - 4)t].$$

We pick $s = \frac{-(\kappa-4)t+1}{4}$ so that $s' - t' = \kappa(s - t)^2$ is the desired difference κ up to a factor of $(s - t)^2$. Then we want to set t' to be $-3(s - t)^2$. With $s = \frac{-(\kappa-4)t+1}{4}$, we can simplify $-3(s - t)^2 - t'$ to

$$-3(s - t)^2 - t' = \frac{1}{64}\left[\kappa^3(\kappa - 2)t^3 - 3\kappa^2(\kappa + 2)t^2 + 3\kappa(\kappa - 10)t - (\kappa + 26)\right]. \quad (10)$$

Since $\kappa \geq 3$, (10) is a nontrivial cubic polynomial in t, so we can set t such that this expression vanishes. Then $\langle 3s' - 2t', s', t'\rangle = (s - t)^2\langle 3(\kappa - 1), \kappa - 3, -3\rangle$. It remains to check that $s \neq t$ and $3s + (\kappa - 3)t \neq 0$.

If $s = t$, then $t = \frac{1}{\kappa}$. Plugging this into (10) gives -1, so $s \neq t$. If $3s + (\kappa - 3)t = 0$, then $t = -\frac{3}{\kappa}$. Plugging this into (10) gives $1 - \kappa \neq 0$, so $3s + (\kappa - 3)t \neq 0$. □

7.2 Dose of an effective Siegel's theorem and Galois theory

It suffices to show that $\langle 3(\kappa-1), \kappa-3, -3\rangle$ is #P-hard for all $\kappa \geq 3$. The general strategy is to use interpolation. However, proving that this interpolation succeeds presents a significant challenge.

Consider the polynomial $p(x, y) \in \mathbb{Z}[x, y]$ defined by

$$\begin{aligned} p(x, y) &= x^5 - 2x^3y - x^2y^2 - x^3 + xy^2 + y^3 - 2x^2 - xy \\ &= x^5 - (2y + 1)x^3 - (y^2 + 2)x^2 + y(y - 1)x + y^3. \end{aligned}$$

We consider y as an integer parameter $y \geq 4$ and treat $p(x, y)$ as an infinite family of quintic polynomials in x with integer coefficients. We want to show that the roots of all

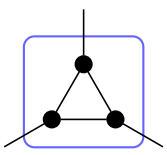


Fig. 12 Triangle gadget used to construct $\langle 3(\kappa - 1), \kappa - 3, -3\rangle$

these quintic polynomials satisfy the lattice condition. First, we determine the number of real and nonreal roots.

Lemma 7.4 *For any integer $y \geq 1$, the polynomial $p(x, y)$ in x has three distinct real roots and two nonreal complex conjugate roots.*

Proof Up to a factor of $-4y^2$, the discriminant of $p(x, y)$ (with respect to x) is

$$27y^{11} - 4y^{10} + 726y^9 - 493y^8 + 2712y^7 - 400y^6 - 2503y^5 + 475y^4 + 956y^3 - 904y^2 + 460y + 104.$$

By replacing y with $z + 1$, we get

$$27z^{11} + 293z^{10} + 2171z^9 + 10316z^8 + 33334z^7 + 77398z^6 + 127383z^5 + 141916z^4 + 102097z^3 + 44373z^2 + 10336z + 1156,$$

which is positive for any $z \geq 0$. Thus, the discriminant is negative.

Therefore, $p(x, y)$ has distinct roots in x for all $y \geq 1$. Furthermore, with a negative discriminant, $p(x, y)$ has $2s$ nonreal complex conjugate roots for some odd integer s. Since $p(x, y)$ is a quintic polynomial (in x), the only possibility is $s = 1$. □

We suspect that for any integer $y \geq 4$, $p(x, y)$ is in fact irreducible over $\mathbb{Q}$ as a polynomial in x. When considering y as an indeterminate, the bivariate polynomial $p(x, y)$ is irreducible over $\mathbb{Q}$ and the algebraic curve it defines has genus 3, so by Theorem 1.2 in [50], $p(x, y)$ is reducible over $\mathbb{Q}$ for at most a finite number of $y \in \mathbb{Z}$. For any integer $y \geq 4$, if $p(x, y)$ is irreducible over $\mathbb{Q}$ as a polynomial in x, then its Galois group is S_5 and its roots satisfy the lattice condition.

Lemma 7.5 *For any integer $y \geq 4$, if $p(x, y)$ is irreducible in $\mathbb{Q}[x]$, then the roots of $p(x, y)$ satisfy the lattice condition.*

Proof By Lemma 7.4, $p(x, y)$ has three distinct real roots and two nonreal complex conjugate roots. With three distinct real roots, we know that not all the roots have the same complex norm. It is well known that an irreducible polynomial of prime degree n with exactly two nonreal roots has S_n as a Galois group over $\mathbb{Q}$ (for example, Theorem 10.15 in [53]). Then we are done by Lemma 6.5. □

We know of just five values of $y \in \mathbb{Z}$ for which $p(x, y)$ is reducible as a polynomial in x:

$$p(x, y) = \begin{cases} (x-1)(x^4 + x^3 + 2x^2 - x + 1) & y = -1, \\ x^2(x^3 - x - 2) & y = 0, \\ (x+1)(x^4 - x^3 - 2x^2 - x + 1) & y = 1, \\ (x-1)(x^2 - x - 4)(x^2 + 2x + 2) & y = 2, \\ (x-3)(x^4 + 3x^3 + 2x^2 - 5x - 9) & y = 3. \end{cases}$$

These five factorizations also give five integer solutions to $p(x, y) = 0$. It is a well-known theorem of Siegel [52] that an algebraic curve of genus at least 1 has only a finite number of integral points. For this curve of genus 3, Faltings' theorem [31] says that there can be only a finite number of rational points. However, these theorems are not *effective* in general. There are some effective versions of Siegel's theorem that can be applied to our

polynomial, but the best effective bound that we can find is over $10^{20,000}$ [61] and hence cannot be checked in practice.

However, it is shown in the next lemma that in fact these five are the only integer solutions. In particular, for any integer $y \geq 4$, $p(x, y)$ does not have a linear factor in $\mathbb{Z}[x]$, and hence by Gauss's Lemma, also no linear factor in $\mathbb{Q}[x]$. The following proof is essentially due to Aaron Levin [46]. We thank Aaron for suggesting the key auxiliary function $g_2(x, y) = \frac{y^2}{x} + y - x^2 + 1$, as well as for his permission to include the proof here. We also thank Bjorn Poonen [51] who suggested a similar proof. After the proof, we will explain certain complications in the proof.

Lemma 7.6 *The only integer solutions to* $p(x, y) = 0$ *are* $(1, -1)$, $(0, 0)$, $(-1, 1)$, $(1, 2)$, *and* $(3, 3)$.

Proof Clearly these five points are solutions to $p(x, y) = 0$. For $a \in \mathbb{Z}$ with $-3 < a < 17$, one can directly check that $p(a, y) = 0$ has no other integer solutions in y.

Let $(a, b) \in \mathbb{Z}^2$ be a solution to $p(x, y) = 0$ with $a \neq 0$. We claim $a \mid b^2$. By definition of $p(x, y)$, clearly $a \mid b^3$. If p is a prime that divides a, then let $\mathrm{ord}_p(a) = e$ and $\mathrm{ord}_p(b) = f$ be the exact orders with which p divides a and b, respectively. Then $f \geq 1$ since $3f \geq e$ and our claim is that $2f \geq e$. Suppose for a contradiction that $2f < e$. From $p(a, b) = 0$, we have

$$a^2(a^3 - 2ab - a - b^2 - 2) = -b^3 - ab(b - 1).$$

The order with respect to p of the left-hand side is

$$\mathrm{ord}_p\left(a^2(a^3 - 2ab - a - b^2 - 2)\right) \geq \mathrm{ord}_p\left(a^2\right) = 2e.$$

Since p is relatively prime to $b - 1$, $\mathrm{ord}_p\left(ab(b - 1)\right) = e + f > 3f$, and therefore, the order of the right-hand side with respect to p is

$$\mathrm{ord}_p\left(-b^3 - ab(b - 1)\right) = \mathrm{ord}_p(b^3) = 3f.$$

However, $2e > 3f$, a contradiction. This proves the claim.

Now consider the functions $g_1(x, y) = y - x^2$ and $g_2(x, y) = \frac{y^2}{x} + y - x^2 + 1$. Whenever $(a, b) \in \mathbb{Z}^2$ is a solution to $p(x, y) = 0$ with $a \neq 0$, $g_1(a, b)$ and $g_2(a, b)$ are integers. However, we show that if $a \leq -3$ or $a \geq 17$, then either $g_1(a, b)$ or $g_2(a, b)$ is not an integer.

Let $c_2 = -(x - 1)x$, $c_1 = -x(2x^2 + 1)$, and $c_0 = x^2(x^3 - x - 2)$ so that $p(x, y) = y^3 + c_2y^2 + c_1y + c_0$. Then the discriminant of $p(x, y)$ with respect to y is

$$\begin{aligned}\mathrm{disc}_y(p(x, y)) &= c_2^2c_1^2 - 4c_1^3 - 4c_2^3c_0 - 27c_0^2 + 18c_2c_1c_0 \\ &= (x - 1)x^3(4x^7 + 5x^6 + x^5 + 45x^4 + 151x^3 + 163x^2 + 67x - 4). \quad (11)\end{aligned}$$

Suppose $x \leq -3$. Replacing x with $-z - 1$ in (11) gives

$$-(z + 1)^3(z + 2)(4z^7 + 23z^6 + 55z^5 + 25z^4 + 21z^3 + 39z^2 + 17z + 14).$$

This is clearly negative (for $z \geq 0$), so (11) is negative. Thus $p(x, y)$ only has one real root as a polynomial in y. Let $y_1(x)$ be that root and consider $y_1^-(x) = x^2 + 2x^{-1}$ and $y_1^+(x) = x^2 + 2x^{-1} + 2x^{-2}$. We have $p(x, y_1^-(x)) = -2x^2 + 6 + 4x^{-1} + 8x^{-3} < 0$. Also $p(x, y_1^+(x)) = 6 + 18x^{-1} + 16x^{-2} + 12x^{-3} + 24x^{-4} + 24x^{-5} + 8x^{-6} > 0$.

Hence, $y_1^-(x) < y_1(x) < y_1^+(x)$, and all three are positive since $y_1^-(x)$ is positive. Then for $x \le -3$,

$$-1 < 2x^{-1} = g_1(x, y_1^-(x)) < g_1(x, y_1(x)) < g_1(x, y_1^+(x)) = 2x^{-1} + 2x^{-2} < 0,$$

so $g_1(x, y_1(x))$ is not an integer. Therefore, $y_1(x)$, the only real root for any integer $x \le -3$, is not an integer.

Now suppose $x \ge 17$. Then (11) is positive, and there are three distinct real roots. Similar to the previous argument, we have $p(x, y_1^-(x)) < 0$ and $p(x, y_1^+(x)) > 0$. Hence, there is some root $y_1(x)$ in the open interval $(y_1^-(x), y_1^+(x))$. All three terms $y_1^-(x) < y_1(x) < y_1^+(x)$ are positive because $y_1^-(x) > 0$. Then

$$0 < 2x^{-1} = g_1(x, y_1^-(x)) < g_1(x, y_1(x)) < g_1(x, y_1^+(x)) = 2x^{-1} + 2x^{-2} < 1,$$

so $g_1(x, y_1(x))$ is not an integer.

There are two more real roots. Consider

$$y_2^-(x) = x^{3/2} - \frac{1}{2}x + \frac{1}{8}x^{1/2} - \frac{65}{128}x^{-1/2} - 2x^{-1} \qquad \text{and}$$
$$y_2^+(x) = x^{3/2} - \frac{1}{2}x + \frac{1}{8}x^{1/2} - \frac{65}{128}x^{-1/2}.$$

Replacing x with $(z+2)^2$ in

$$\begin{aligned} p(x, y_2^-(x)) = 2x^{5/2} &- \frac{2495}{512}x^2 + \frac{1087}{512}x^{3/2} - \frac{19569}{16384}x - \frac{8579}{16384}x^{1/2} + \frac{126847}{32768} \\ &+ \frac{1452419}{131072}x^{-1/2} - \frac{317}{256}x^{-1} + \frac{2871103}{2097152}x^{-3/2} - \frac{12675}{8192}x^{-2} \\ &- \frac{195}{32}x^{-5/2} - 8x^{-3} \end{aligned}$$

gives

$$\frac{1}{2097152(z+2)^6} \times \begin{pmatrix} 4194304z^{11} + 82055168z^{10} + 722808832z^9 + 3774605184z^8 \\ + 12935149184z^7 + 30375187136z^6 + 49489164080z^5 + 55372934880z^4 \\ + 41238374079z^3 + 19431701370z^2 + 5465401844z + 812262392 \end{pmatrix},$$

which is clearly positive ($z \ge 0$). Thus, $p(x, y_2^-(x)) > 0$. Also

$$\begin{aligned} p(x, y_2^+(x)) = -2x^{5/2} &- \frac{447}{512}x^2 - \frac{193}{512}x^{3/2} - \frac{3185}{16384}x + \frac{20605}{16384}x^{1/2} - \frac{4225}{32768} \\ &+ \frac{12675}{131072}x^{-1/2} - \frac{274625}{2097152}x^{-3/2} < 0. \end{aligned}$$

Hence, there is some root $y_2(x)$ in the open interval $(y_2^-(x), y_2^+(x))$. All three terms $y_2^-(x) < y_2(x) < y_2^+(x)$ are positive because $y_2^-(x) > 0$. Hence, for $x \ge 17$,

$$\begin{aligned} -1 < &-4x^{-1/2} - \frac{65}{512}x^{-1} - \frac{1}{2}x^{-3/2} + \frac{4225}{16384}x^{-2} + \frac{65}{32}x^{-5/2} + 4x^{-3} \\ &= g_2(x, y_2^-(x)) < g_2(x, y_2(x)) < g_2(x, y_2^+(x)) = -\frac{65}{512}x^{-1} + \frac{4225}{16384}x^{-2} < 0, \end{aligned}$$

so $g_2(x, y_2(x))$ is not an integer.

Finally, consider

$$y_3^-(x) = -x^{3/2} - \frac{1}{2}x - \frac{1}{8}x^{1/2} + \frac{65}{128}x^{-1/2} - x^{-1} \quad \text{and}$$
$$y_3^+(x) = -x^{3/2} - \frac{1}{2}x - \frac{1}{8}x^{1/2} + \frac{65}{128}x^{-1/2} - \frac{1}{2}x^{-1}.$$

We have

$$\begin{aligned} p(x, y_3^-(x)) = &-\frac{1471}{512}x^2 - \frac{447}{512}x^{3/2} - \frac{11377}{16384}x - \frac{6013}{16384}x^{1/2} + \frac{94079}{32768} - \frac{339331}{131072}x^{-1/2} \\ &- \frac{61}{512}x^{-1} - \frac{511807}{2097152}x^{-3/2} - \frac{12675}{16384}x^{-2} + \frac{195}{128}x^{-5/2} - x^{-3} \\ &< 0. \end{aligned}$$

Replacing x with $(z+3)^2$ in

$$\begin{aligned} p(x, y_3^+(x)) = x^{5/2} &- \frac{959}{512}x^2 - \frac{127}{512}x^{3/2} - \frac{7281}{16384}x - \frac{13309}{16384}x^{1/2} + \frac{53119}{32768} - \frac{77699}{131072}x^{-1/2} \\ &+ \frac{67}{1024}x^{-1} + \frac{78017}{2097152}x^{-3/2} - \frac{12675}{32768}x^{-2} + \frac{195}{512}x^{-5/2} - \frac{1}{8}x^{-3} \end{aligned}$$

gives

$$\begin{aligned} &\frac{1}{2097152(z+3)^6} \\ &\times \begin{pmatrix} 2097152z^{11} + 65277952z^{10} + 919728128z^9 + 7736969088z^8 \\ + 43137332608z^7 + 167175471424z^6 + 458797435600z^5 + 889807335920z^4 \\ + 1191781601633z^3 + 1045691960361z^2 + 537771428331z + 121660965323 \end{pmatrix}, \end{aligned}$$

which is clearly positive ($z \geq 0$). Thus, $p(x, y_3^+(x)) > 0$. Hence, there is some root $y_3(x)$ in the open interval $(y_3^-(x), y_3^+(x))$. All three terms $y_3^-(x) < y_3(x) < y_3^+(x)$ are negative because $y_3^+(x) < 0$. Furthermore, the partial derivative $\frac{\partial g_2(x,y)}{\partial y} = 2x^{-1}y + 1$ and $\frac{\partial^2 g_2(x,y)}{\partial y^2} = 2x^{-1} > 0$. Thus, $\frac{\partial g_2(x,y)}{\partial y} \leq \frac{\partial g_2(x,y)}{\partial y}\,|_{y=y_3^+(x)} = -2x^{1/2} - \frac{1}{4}x^{-1/2} + \frac{65}{64}x^{-3/2} - x^{-2} < 0$, for all $y \in (-\infty, y_3^+(x)]$. Thus, $g_2(x, y)$ is decreasing monotonically in y over the interval $(-\infty, y_3^+(x)]$. Then

$$\begin{aligned} 0 &< x^{-1/2} - \frac{65}{512}x^{-1} + \frac{1}{8}x^{-3/2} + \frac{4225}{16384}x^{-2} - \frac{65}{128}x^{-5/2} + \frac{1}{4}x^{-3} \\ &= g_2(x, y_3^+(x)) < g_2(x, y_3(x)) < g_2(x, y_3^-(x)) \\ &= 2x^{-1/2} - \frac{65}{512}x^{-1} + \frac{1}{4}x^{-3/2} + \frac{4225}{16384}x^{-2} - \frac{65}{64}x^{-5/2} + x^{-3} < 1, \end{aligned}$$

so $g_2(x, y_3(x))$ is not an integer. To complete the proof, notice that the intervals $(y_1^-(x), y_1^+(x))$, $(y_2^-(x), y_2^+(x))$, and $(y_3^-(x), y_3^+(x))$ are disjoint. Therefore, we have shown that none of the three roots is an integer for any integer $x \geq 17$. □

Remark One can obtain the Puiseux series expansions for $p(x, y)$, which are

$$\begin{aligned} y_1(x) &= x^2 + 2x^{-1} + 2x^{-2} - 6x^{-4} - 18x^{-5} + O(x^{-6}) && \text{for } x \in \mathbb{R}, \\ y_2(x) &= x^{3/2} - \tfrac{1}{2}x + \tfrac{1}{8}x^{1/2} - \tfrac{65}{128}x^{-1/2} - x^{-1} - \tfrac{1471}{1024}x^{-3/2} - x^{-2} + O(x^{-5/2}) && \text{for } x > 0, \text{ and} \\ y_3(x) &= -x^{3/2} - \tfrac{1}{2}x - \tfrac{1}{8}x^{1/2} + \tfrac{65}{128}x^{-1/2} - x^{-1} + \tfrac{1471}{1024}x^{-3/2} - x^{-2} + O(x^{-5/2}) && \text{for } x > 0. \end{aligned}$$

These series converge to the actual roots of $p(x, y)$ for large x. The basic idea of the proof—called Runge's method—is that, for example, when we substitute $y_2(x)$ in $g_2(x, y)$, we get $g_2(x, y_2(x)) = O(x^{-1/2})$, where the multiplier in the O-notation is bounded both above and below by a nonzero constant in absolute value. Thus, for large x, this cannot be an integer. However, for integer solutions (x, y) of $p(x, y)$, this must be an integer.

We note that the expressions for the $y_i^+(x)$ and $y_i^-(x)$ are the truncated or rounded Puiseux series expansions. The reason we discuss $y_i^+(x)$ and $y_i^-(x)$ is because we want to prove an absolute bound, instead of the asymptotic bound implied by the O-notation.

By Lemma 7.6, if $p(x, y)$ is reducible over $\mathbb{Q}$ as a polynomial in x for any integer $y \geq 4$, then the only way it can factor is as a product of an irreducible quadratic and an irreducible cubic. The next lemma handles this possibility.

Lemma 7.7 *For any integer $y_0 \geq 4$, if $p(x, y_0)$ is reducible over $\mathbb{Q}$, then the roots of $p(x, y_0)$ satisfy the lattice condition.*

Proof Let $q(x) = p(x, y_0)$ for a fixed integer $y_0 \geq 4$. Suppose that $q(x) = f(x)g(x)$, where $f(x), g(x) \in \mathbb{Q}[x]$ are monic polynomials of degree at least 1. By Lemma 7.6, the degree of each factor must be at least 2. Then without loss of generality, let $f(x)$ and $g(x)$ be quadratic and cubic polynomials, respectively, both of which are irreducible over $\mathbb{Q}$. By Gauss' Lemma, we can further assume $f(x), g(x) \in \mathbb{Z}[x]$.

Let $\mathbb{Q}_f$ and $\mathbb{Q}_g$ denote the splitting fields over $\mathbb{Q}$ of f and g, respectively. Suppose α, β are the roots of $f(x)$ and γ, δ, ϵ are the roots of $g(x)$. Of course none of these roots are 0. Suppose there exist $i, j, k, m, n \in \mathbb{Z}$ such that

$$\alpha^i \beta^j = \gamma^k \delta^m \epsilon^n \quad \text{and} \quad i + j = k + m + n. \tag{12}$$

We want to show that $i = j = k = m = n = 0$.

We first show that if $i = j$ and $k = m = n$, then $i = j = k = m = n = 0$. By (12), we have $(\alpha\beta)^i = (\gamma\delta\epsilon)^k$ and $2i = 3k$. Suppose $i \neq 0$, then also $k \neq 0$. We can write $i = 3t$ and $k = 2t$ for some nonzero $t \in \mathbb{Z}$. Let $A = \alpha\beta$ and $B = \gamma\delta\epsilon$. Then, both A and B are integers and $AB = y_0^3$. From $A^{3t} = B^{2t}$, we have $A^3 = \pm B^2$. Then $y_0^6 = A^2B^2 = \pm A^5$, and since $y_0 > 3$, there is a nonzero integer $s > 1$ such that $y_0 = s^5$. This implies $A = \pm s^6$ and $B = \pm s^9$ (with the same $\pm$ sign). Then $f(x) = x^2 + c_1 x \pm s^6$, $g(x) = x^3 + c_2' x^2 + c_1' x \pm s^9$, and $q(x) = x^5 - (2s^5 + 1)x^3 - (s^{10} + 2)x^2 + s^5(s^5 - 1)x + s^{15}$. We consider the coefficient of x in $q(x) = f(x)g(x)$. This is $s^{10} - s^5 = \pm c_1' s^6 \pm c_1 s^9$. Since $s > 1$, there is a prime p such that $p^u \mid s$ and $p^{u+1} \nmid s$, for some $u \geq 1$. But then p^{6u} divides $s^5 = s^{10} \pm c_1' s^6 \pm c_1 s^9$. This is a contradiction. Hence, $i = j$ and $k = m = n$ imply $i = j = k = m = n = 0$.

Now we claim that $\omega = \alpha/\beta$ is not a root of unity. For a contradiction, suppose that ω is a primitive dth root of unity. Since $\omega \in \mathbb{Q}_f$, which is a degree 2 extension over $\mathbb{Q}$, we have $\phi(d) \mid 2$, where $\phi(\cdot)$ is Euler's totient function. Hence, $d \in \{1, 2, 3, 4, 6\}$. The quadratic polynomial $f(x)$ has the form $x^2 - (1 + \omega)\beta x + \omega\beta^2 \in \mathbb{Z}[x]$. Hence, $r = \frac{(1+\omega)}{\omega\beta} \in \mathbb{Q}$. We prove the claim separately according to whether $r = 0$ or not.

If $r = 0$, then $\omega = -1$ and $d = 2$. In this case, $f(x)$ has the form $x^2 + a$ for some $a \in \mathbb{Z}$. It is easy to check that $q(x)$ has no such polynomial factor in $\mathbb{Z}[x]$ unless $y_0 = 0$. In fact, suppose $x^2 + a \mid q(x)$ in $\mathbb{Z}[x]$. Then $q(x) = (x^2 + a)(x^3 + bx + c)$ since the coefficient of x^4 in $q(x)$ is 0. Also $a + b = -(2y_0 + 1)$, $c = -(y_0^2 + 2)$, $ab = y_0(y_0 - 1)$ and $ac = y_0^3$. It follows that

a and b are the two roots of the quadratic polynomial $X^2 + (2y_0 + 1)X + y_0^2 - y_0 \in \mathbb{Z}[X]$. Since $a, b \in \mathbb{Z}$, the discriminant $8y_0 + 1$ must be a perfect square, and in fact an odd perfect square $(2z - 1)^2$ for some $z \in \mathbb{Z}$. Thus, $y_0 = z(z - 1)/2$. By the quadratic formula, $a = -y_0 + z - 1$ or $-y_0 - z$. On the other hand, $a = ac/c = -y_0^3/(y_0^2 + 2)$. In both cases, this leads to a polynomial in z in $\mathbb{Z}[z]$ that has no integer solutions other than $z = 0$, which gives $y_0 = 0$.

Now suppose $r \neq 0$. Plugging r back in $f(x)$, we have $f(x) = x^2 - (2 + \omega + \omega^{-1})r^{-1}x + (2 + \omega + \omega^{-1})r^{-2}$. The quantity $2 + \omega + \omega^{-1} = 4, 1, 2, 3$ when $d = 1, 3, 4, 6$, respectively. Since $(2 + \omega + \omega^{-1})r^{-2} \in \mathbb{Z}$, the rational number r^{-1} must be an integer when $d = 3, 4, 6$ and half an integer when $d = 1$. In all cases, it is easy to check that a polynomial $f(x)$ of the specified form does not divide $q(x)$ unless $y = 0$ or $y = -1$. Thus, we have proved the claim that $\omega = \alpha/\beta$ is not a root of unity.

Next consider the case that $f(x)$ is irreducible over $\mathbb{Q}_g$. Let E be the splitting field of f over $\mathbb{Q}_g$. Then, $[E : \mathbb{Q}_g] = 2$. Therefore, there exists an automorphism $\tau \in \mathrm{Gal}(E/\mathbb{Q}_g)$ that swaps α and β but fixes $\mathbb{Q}_g$ and thus fixes γ, δ, ϵ pointwise. By applying τ to (12), we have $\alpha^j\beta^i = \gamma^k\delta^m\epsilon^n$. Dividing by (12) gives $(\alpha/\beta)^{j-i} = 1$. Since α/β is not a root of unity, we get $i = j$. Hence, we have $(\alpha\beta)^i = \gamma^k\delta^m\epsilon^n$. The order of $\mathrm{Gal}(\mathbb{Q}_g/\mathbb{Q})$ is $[\mathbb{Q}_g : \mathbb{Q}]$, which is divisible by 3. Thus, $\mathrm{Gal}(\mathbb{Q}_g/\mathbb{Q}) \subseteq S_3$ contains an element of order 3, which must act as a 3-cycle on γ, δ, ϵ. Since $\alpha\beta \in \mathbb{Q}$, applying this cyclic permutation gives $(\alpha\beta)^i = \gamma^m\delta^n\epsilon^k$. Therefore, $\gamma^{k-m}\delta^{m-n}\epsilon^{n-k} = 1$. Notice that $(k - m) + (m - n) + (n - k) = 0$.

It can be directly checked that $q(x)$ is not divisible by any $x^3 + c \in \mathbb{Z}[x]$,

and therefore by Lemma 6.4, the roots γ, δ, ϵ of the cubic polynomial $g(x)$ satisfy the lattice condition. Therefore, $k = m = n$. Again, we have shown that $i = j$ and $k = m = n$ imply $i = j = k = m = n = 0$.

The last case is when $f(x)$ splits in $\mathbb{Q}_g[x]$. Then $\mathbb{Q}_f$ is a subfield of $\mathbb{Q}_g$, and $2 = [\mathbb{Q}_f : \mathbb{Q}]|[\mathbb{Q}_g : \mathbb{Q}]$. Therefore, $[\mathbb{Q}_g : \mathbb{Q}] = 6$ and $\mathrm{Gal}(\mathbb{Q}_g/\mathbb{Q}) = S_3$. Since $\mathbb{Q}_f$ is normal over $\mathbb{Q}$, being a splitting field of a separable polynomial in characteristic 0, by the fundamental theorem of Galois theory, the corresponding subgroup for $\mathbb{Q}_f$ is $\mathrm{Gal}(\mathbb{Q}_g/\mathbb{Q}_f)$, which is a normal subgroup of S_3 with index 2. Such a subgroup of S_3 is unique, namely A_3. In particular, the transposition τ' that swaps γ and δ but fixes ϵ is an element in $\mathrm{Gal}(\mathbb{Q}_g/\mathbb{Q}) = S_3$ but not in $\mathrm{Gal}(\mathbb{Q}_g/\mathbb{Q}_f) = A_3$. This transposition must fix α and β setwise but not pointwise. Hence, it must swap α and β.

By applying τ' to (12), we have $\alpha^j\beta^i = \gamma^m\delta^k\epsilon^n$. Then dividing these two equations gives $(\alpha/\beta)^{i-j} = (\delta/\gamma)^{m-k}$. Similarly, by considering the transposition that switches γ and ϵ and fixes δ, we get $(\alpha/\beta)^{i-j} = (\gamma/\epsilon)^{k-n}$. By combining these two equations, we have $\gamma^{n-m}\delta^{m-k}\epsilon^{k-n} = 1$. Note that $(n - m) + (m - k) + (k - n) = 0$.

As we noted above, the roots of the irreducible $g(x)$ satisfy the lattice condition, so we conclude that $k = n = m$. From $(\alpha/\beta)^{i-j} = (\delta/\gamma)^{m-k} = 1$, we get $i = j$ since α/β is not a root of unity. We conclude that $i = j = k = m = n = 0$, so the roots of $q(x)$ satisfy the lattice condition. □

Even though $p(x, 3) = (x - 3)(x^4 + 3x^3 + 2x^2 - 5x - 9)$ is reducible, its roots still satisfy the lattice condition. To show this, we utilize a few results, Theorem 7.8, Lemma 7.9, and Lemma 7.10.

The first is a well-known theorem of Dedekind.

Theorem 7.8 (Theorem 4.37 [40]) *Suppose $f(x) \in \mathbb{Z}[x]$ is a monic polynomial of degree n. For a prime p, let $f_p(x)$ be the corresponding polynomial in $\mathbb{Z}_p[x]$. If $f_p(x)$ has distinct roots and factors over $\mathbb{Z}_p[x]$ as a product of irreducible factors with degrees $d_1, d_2, \ldots, d_r$, then the Galois group of f over $\mathbb{Q}$ contains an element with cycle type $(d_1, d_2, \ldots, d_r)$.*

With the second result, we can show that $x^4 + 3x^3 + 2x^2 - 5x - 9$ has Galois group S_4 over $\mathbb{Q}$.

Lemma 7.9 (Lemma on page 98 in [33]) *For $n \geq 2$, let G be a subgroup of S_n. If G is transitive, contains a transposition and contains a p-cycle for some prime $p > n/2$, then $G = S_n$.*

In the contrapositive, the third result shows that the roots of $x^4 + 3x^3 + 2x^2 - 5x - 9$ do not all have the same complex norm.

Lemma 7.10 (Lemma D.2 in [17]) *If all roots of $x^4 + a_3x^3 + a_2x^2 + a_1x + a_0 \in \mathbb{C}[x]$ have the same complex norm, then $a_2|a_1|^2 = |a_3|^2\overline{a_2}a_0$.*

Theorem 7.11 *The roots of $p(x, 3) = (x - 3)(x^4 + 3x^3 + 2x^2 - 5x - 9)$ satisfy the lattice condition.*

Proof Let $f(x) = x^4 + 3x^3 + 2x^2 - 5x - 9$ and let G_f be the Galois group of f over $\mathbb{Q}$. We claim that $G_f = S_4$. As a polynomial over $\mathbb{Z}_5$, $f(x) \equiv x^4 + 3x^3 + 2x^2 + 1$ is irreducible, so $f(x)$ is also irreducible over $\mathbb{Z}$. By Gauss' Lemma, this implies irreducibility over $\mathbb{Q}$. Over $\mathbb{Z}_{13}$, $f(x)$ factors into the product of irreducibles $(x^2 + 7)(x + 6)(x + 10)$ and clearly has distinct roots, so by Theorem 7.8, G_f contains a transposition. Over $\mathbb{Z}_3$, $f(x)$ factors into the product of irreducibles $x(x^3 + 2x + 1)$ and has distinct roots because its discriminant is $1 \not\equiv 0 \pmod 3$, so by Theorem 7.8, G_f contains a 3-cycle. Then by Lemma 7.9, $G_f = S_4$.

Let $\alpha, \beta, \gamma, \delta$ be the roots of $f(x)$. Suppose there exist $i, j, k, \ell, n \in \mathbb{Z}$ satisfying $n = i + j + k + \ell$ such that $3^n = \alpha^i\beta^j\gamma^k\delta^\ell$. Now $G_f = S_4$ contains the 4-cycle (1 2 3 4) that cyclically permutes the roots of $f(x)$ but fixes $\mathbb{Q}$. We apply it zero, one, two, and three times to get

$$\begin{aligned} 3^n &= \alpha^i\beta^j\gamma^k\delta^\ell, \\ &= \beta^i\gamma^j\delta^k\alpha^\ell, \\ &= \gamma^i\delta^j\alpha^k\beta^\ell, \text{ and} \\ &= \delta^i\alpha^j\beta^k\gamma^\ell. \end{aligned}$$

Then $3^{4n} = (\alpha\beta\gamma\delta)^{i+j+k+\ell} = (-9)^{i+j+k+\ell}$. Since $n = i + j + k + \ell$, this can only hold when $n = 0$.

Thus, it suffices to show that the roots of $f(x)$ satisfy the lattice condition. By the contrapositive of Lemma 7.10, the roots of $f(x)$ do not all have the same complex norm. Then we are done by Lemma 6.5. □

From Lemma 7.5, Lemma 7.7, and Theorem 7.11, we obtain the following Theorem.

Theorem 7.12 *For any integer $y_0 \geq 3$, the roots of $p(x, y_0)$ satisfy the lattice condition.*

We use Theorem 7.12 to prove Lemma 7.14. We note that the succinct signature type τ_4 is a refinement of τ_{color}, so any succinct signature of type τ_{color} can also be expressed as a succinct signature of type τ_4. In particular, the succinct signature $\langle 2, 1, 0, 1, 0\rangle$ of type τ_{color} is written $\langle 2, 0, 1, 0, 0, 0, 1, 0, 0\rangle$ of type τ_4. Then the following is a restatement of Corollary 4.7.

Corollary 7.13 *Suppose $\kappa \geq 3$ is the domain size. Let $\langle 2, 0, 1, 0, 0, 0, 1, 0, 0\rangle$ be a succinct quaternary signature of type τ_4. Then* Pl-Holant($\langle 2, 0, 1, 0, 0, 0, 1, 0, 0\rangle$) *is* #P*-hard.*

Lemma 7.14 *Suppose $\kappa \geq 4$ is the domain size. Then* Pl-Holant($\langle 3(\kappa - 1), \kappa - 3, -3\rangle$) *is* #P*-hard.*

Proof Let $\langle 2, 0, 1, 0, 0, 0, 1, 0, 0\rangle$ be a succinct quaternary signature of type τ_4. We reduce from Pl-Holant($\langle 2, 0, 1, 0, 0, 0, 1, 0, 0\rangle$), which is #P-hard by Corollary 7.13.

Consider the gadget in Fig. 13. We assign $\langle 3(\kappa - 1), \kappa - 3, -3\rangle$ to the vertices. By Lemma 11.3, the signature of this gadget is $f = \langle f_{\substack{1\,1\\1\,1}}, f_{\substack{1\,2\\1\,1}}, f_{\substack{1\,2\\1\,2}}, f_{\substack{1\,3\\1\,2}}, f_{\substack{1\,2\\2\,1}}, f_{\substack{1\,3\\2\,1}}, f_{\substack{1\,1\\2\,2}}, f_{\substack{1\,1\\2\,3}}, f_{\substack{1\,4\\2\,3}}\rangle$ up to a nonzero factor of κ, where

$$
\begin{aligned}
f_{\substack{1\,1\\1\,1}} &= (\kappa - 1)(\kappa + 3),\\
f_{\substack{1\,2\\1\,1}} &= \kappa - 3,\\
f_{\substack{1\,2\\1\,2}} &= 2\kappa - 3,\\
f_{\substack{1\,3\\1\,2}} &= \kappa - 3,\\
f_{\substack{1\,2\\2\,1}} &= 2\kappa - 3,\\
f_{\substack{1\,3\\2\,1}} &= \kappa - 3,\\
f_{\substack{1\,1\\2\,2}} &= (\kappa - 3)(\kappa + 1),\\
f_{\substack{1\,1\\2\,3}} &= \kappa - 3, \text{ and}\\
f_{\substack{1\,4\\2\,3}} &= -3.
\end{aligned}
$$

Now consider the recursive construction in Fig. 14. We assign f to every vertex. Up to a nonzero factor of κ^s, let g_s be the succinct signature of type τ_4 for the sth gadget in this construction. Then $g_0 = \langle 1, 0, 0, 0, 0, 0, 1, 0, 0\rangle$ and $g_s = M^s g_0$, where M is the matrix in Table 1.

The row vectors

$$
\begin{aligned}
&(0, 0, 0, 0, -1, 0, 0, 0, 1),\\
&(0, -1, 0, 1, -1, 0, 0, 1, 0),\\
&(-1, 0, 1, 0, -1, 0, 1, 0, 0), \text{ and}\\
&(0, 0, 0, 0, -1, 1, 0, 0, 0)
\end{aligned}
$$

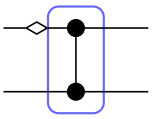


Fig. 13 Quaternary gadget used in the interpolation construction below. All vertices are assigned $\langle 3(\kappa - 1), \kappa - 3, -3\rangle$

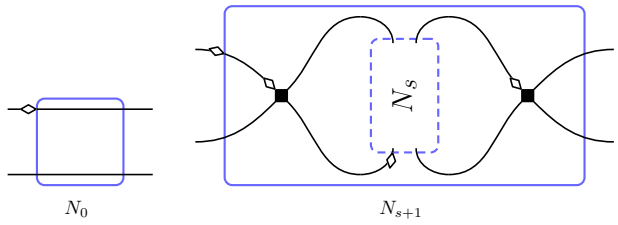


Fig. 14 Recursive construction to interpolate the weighted Eulerian partition signature. The vertices are assigned the signature of the gadget in Fig. 13

are linearly independent row eigenvectors of M, all with eigenvalue κ^3, that are orthogonal to the initial signature g_0. Note that our target signature $\langle 2, 0, 1, 0, 0, 0, 1, 0, 0\rangle$ is also orthogonal to these four row eigenvectors.

Up to a factor of $(x-\kappa^3)^4$, the characteristic polynomial of M is

$$h(x,\kappa) = x^5 - \kappa^6(2\kappa-1)x^3 - \kappa^9(\kappa^2-2\kappa+3)x^2 + (\kappa-2)(\kappa-1)\kappa^{12}x + (\kappa-1)^3\kappa^{15}.$$

Since $h(\kappa^3,\kappa) = (\kappa-3)\kappa^{17}$ and $\kappa \geq 4$, we know that κ^3 is not a root of $h(x,\kappa)$ as a polynomial in x. Thus, none of the remaining eigenvalues are κ^3. The roots of $h(x,\kappa)$ satisfy the lattice condition iff the roots of

$$\begin{aligned}\tilde{h}(x,\kappa) &= \frac{1}{\kappa^{15}}h(\kappa^3 x,\kappa)\\ &= x^5 - (2\kappa-1)x^3 - (\kappa^2-2\kappa+3)x^2 + (\kappa-2)(\kappa-1)x + (\kappa-1)^3\end{aligned}$$

satisfy the lattice condition. In $\tilde{h}(x,\kappa)$, we replace κ by $y+1$ to get $p(x,y) = x^5 - (2y+1)x^3 - (y^2+2)x^2 + (y-1)yx + y^3$. By Theorem 7.12, the roots $p(x,y_0)$ satisfy the lattice condition for any integer $y_0 \geq 3$. Thus, the roots of $\tilde{h}(x,\kappa)$ satisfy the lattice for any $\kappa \geq 4$. In particular, this means that the five eigenvalues of M different from κ^3 are distinct, so M is diagonalizable.

The 5-by-5 matrix in the upper-left corner of $[g_0\ Mg_0\ \ldots\ M^8g_0]$ is

$$\begin{bmatrix} 1 & 9(\kappa-1)^2\kappa & (\kappa-1)\kappa^4(\kappa^3-3\kappa^2+11\kappa+3) & (\kappa-1)\kappa^7(\kappa^3+12\kappa^2-11\kappa+6) & (\kappa-1)\kappa^{10}(\kappa^4+4\kappa^3-4\kappa^2+44\kappa-33)\\ 0 & 3(\kappa-3)(\kappa-1)\kappa & -(\kappa-3)\kappa^4(\kappa^2-2\kappa-1) & (\kappa-3)\kappa^7(3\kappa^2-3\kappa+2) & (\kappa-3)\kappa^{10}(\kappa^3-4\kappa^2+16\kappa-11)\\ 0 & 9(\kappa-1)^2\kappa & \kappa^4(\kappa^4-4\kappa^3+6\kappa^2+4\kappa-3) & \kappa^7(15\kappa^3-28\kappa^2+11\kappa-6) & \kappa^{10}(\kappa^5+3\kappa^4-22\kappa^3+72\kappa^2-83\kappa+33)\\ 0 & 3(\kappa-3)(\kappa-1)\kappa & -(\kappa-3)(\kappa-1)\kappa^4(\kappa+1) & 2(\kappa-3)\kappa^7(2\kappa^2-\kappa+1) & (\kappa-3)(\kappa-1)\kappa^{10}(\kappa^2-6\kappa+11)\\ 0 & (\kappa-3)^2\kappa & (\kappa-3)\kappa^4(\kappa+1) & (\kappa-3)\kappa^7(\kappa^2-\kappa+2) & (\kappa-3)\kappa^{10}(\kappa^3-2\kappa^2+10\kappa-11)\end{bmatrix}.$$

Its determinant is $(\kappa-3)^3(\kappa-1)^2\kappa^{26}(\kappa^4+\kappa^3+17\kappa^2+3\kappa+2)$, which is nonzero since $\kappa \geq 4$. Thus, $[g_0\ Mg_0\ \ldots\ M^8g_0]$ has rank at least 5, so by Lemma 6.2, g_0 is not orthogonal to the five remaining row eigenvectors of M.

Therefore, by Lemma 6.6, we can interpolate $\langle 2, 0, 1, 0, 0, 0, 1, 0, 0\rangle$, which completes the proof. □

When $\kappa = 3$, $\langle 3(\kappa-1), \kappa-3, -3\rangle$ simplifies to $-3\langle -2, 0, 1\rangle$. We have a much simpler proof that this signature is #P-hard.

Lemma 7.15 *Suppose the domain size is* 3. *Then* Pl-Holant($\langle -2, 0, 1\rangle$) *is* #P*-hard.*

Proof Let $g = \langle 2, 0, 1, 0, 0, 0, 1, 0\rangle$ be a succinct quaternary signature of type τ_4. We reduce from Pl-Holant(g), which is #P-hard by Corollary 7.13.

Consider the gadget in Fig. 15. The vertices are assigned $\langle -2, 0, 1\rangle$. Up to a factor of 9, the signature of this gadget is g, as desired. □

Table 1 Recurrence matrix for the recursive construction in the proof of Lemma 7.14

$(\kappa-1)\left(\kappa^2+9\kappa-9\right)$	$12(\kappa-3)(\kappa-1)^2$	$(\kappa-3)^2(\kappa-1)$	$2(\kappa-3)^2(\kappa-2)(\kappa-1)$	$(\kappa-3)^2(\kappa-1)$	$2(\kappa-3)^2(\kappa-2)(\kappa-1)$	$(\kappa-1)(2\kappa-3)(4\kappa-3)$	$6(\kappa-3)(\kappa-2)(\kappa-1)^2$	$(\kappa-3)^3(\kappa-2)(\kappa-1)$
$3(\kappa-3)(\kappa-1)$	$3\kappa^3-28\kappa^2+60\kappa-36$	$-(\kappa-3)(2\kappa-3)$	$-2(\kappa-3)(\kappa-2)(2\kappa-3)$	$-(\kappa-3)(2\kappa-3)$	$-2(\kappa-3)(\kappa-2)(2\kappa-3)$	$3(\kappa-3)(\kappa-1)^2$	$(\kappa-2)\left(\kappa^3-14\kappa^2+30\kappa-18\right)$	$-(\kappa-3)^2(\kappa-2)(2\kappa-3)$
$(2\kappa-3)(4\kappa-3)$	$12(\kappa-3)(\kappa-1)^2$	$(\kappa-3)^2(\kappa-1)$	$2(\kappa-3)^2(\kappa-2)(\kappa-1)$	$(\kappa-3)^2(\kappa-1)$	$2(\kappa-3)^2(\kappa-2)(\kappa-1)$	$9\kappa^3-26\kappa^2+27\kappa-9$	$6(\kappa-3)(\kappa-2)(\kappa-1)^2$	$(\kappa-3)^3(\kappa-2)(\kappa-1)$
$3(\kappa-3)(\kappa-1)$	$2\left(\kappa^3-14\kappa^2+30\kappa-18\right)$	$-(\kappa-3)(2\kappa-3)$	$-2(\kappa-3)(\kappa-2)(2\kappa-3)$	$-(\kappa-3)(2\kappa-3)$	$-2(\kappa-3)(\kappa-2)(2\kappa-3)$	$3(\kappa-3)(\kappa-1)^2$	$(\kappa-3)\left(\kappa^3-12\kappa^2+22\kappa-12\right)$	$-(\kappa-3)^2(\kappa-2)(2\kappa-3)$
$(\kappa-3)^2$	$-4(\kappa-3)(2\kappa-3)$	$3(\kappa-3)$	$6(\kappa-3)(\kappa-2)$	$\kappa^3+3\kappa-9$	$6(\kappa-3)(\kappa-2)$	$(\kappa-3)^2(\kappa-1)$	$-2(\kappa-3)(\kappa-2)(2\kappa-3)$	$3(\kappa-3)^2(\kappa-2)$
$(\kappa-3)^2$	$-4(\kappa-3)(2\kappa-3)$	$3(\kappa-3)$	$6(\kappa-3)(\kappa-2)$	$3(\kappa-3)$	$\kappa^3+6\kappa^2-30\kappa+36$	$(\kappa-3)^2(\kappa-1)$	$-2(\kappa-3)(\kappa-2)(2\kappa-3)$	$3(\kappa-3)^2(\kappa-2)$
$(\kappa-3)^2$	$-4(\kappa-3)(2\kappa-3)$	$\kappa^3+3\kappa-9$	$6(\kappa-3)(\kappa-2)$	$3(\kappa-3)$	$6(\kappa-3)(\kappa-2)$	$(\kappa-3)^2(\kappa-1)$	$-2(\kappa-3)(\kappa-2)(2\kappa-3)$	$3(\kappa-3)^2(\kappa-2)$
$(\kappa-3)^2$	$-4(\kappa-3)(2\kappa-3)$	$3(\kappa-3)$	$\kappa^3+6\kappa^2-30\kappa+36$	$3(\kappa-3)$	$6(\kappa-3)(\kappa-2)$	$(\kappa-3)^2(\kappa-1)$	$-2(\kappa-3)(\kappa-2)(2\kappa-3)$	$3(\kappa-3)^2(\kappa-2)$
$(\kappa-3)^2$	$-4(\kappa-3)(2\kappa-3)$	$3(\kappa-3)$	$6(\kappa-3)(\kappa-2)$	$3(\kappa-3)$	$6(\kappa-3)(\kappa-2)$	$(\kappa-3)^2(\kappa-1)$	$-2(\kappa-3)(\kappa-2)(2\kappa-3)$	$(2\kappa-3)\left(2\kappa^2-9\kappa+18\right)$

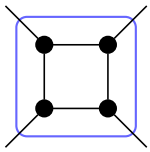


Fig. 15 Square gadget used to construct the weighted Eulerian partition signature

We summarize this section with the following result. With all succinct binary signatures of type τ_2 available as well as the succinct unary signature $\langle 1\rangle$ of type τ_1, any succinct ternary signature $\langle a, b, c\rangle$ of type τ_3 satisfying $\mathfrak{B} \neq 0$ is #P-hard.

Lemma 7.16 *Suppose $\kappa \geq 3$ is the domain size and $a, b, c \in \mathbb{C}$. Let $\mathcal{F}$ be a signature set containing the succinct ternary signature $\langle a, b, c\rangle$ of type τ_3, the succinct unary signature $\langle 1\rangle$ of type τ_1, and the succinct binary signature $\langle x, y\rangle$ of type τ_2 for all $x, y \in \mathbb{C}$. If $\mathfrak{B} \neq 0$, then* Pl-Holant($\mathcal{F}$) *is* #P*-hard.*

Proof Suppose $\mathfrak{A} \neq 0$. By Lemma 7.1, we have a succinct ternary signature $\langle a', b', b'\rangle$ of type τ_3 with $a' \neq b'$. Then we are done by Corollary 4.19.

Otherwise, $\mathfrak{A} = 0$. Since $\mathfrak{B} \neq 0$, we have $b \neq c$. By Lemma 7.3, we have $\langle 3(\kappa - 1), \kappa - 3, -3\rangle$. If $\kappa \geq 4$, then we are done by Lemma 7.14. Otherwise, $\kappa = 3$ and we are done by Lemma 7.15. □

8 Constructing a nonzero unary signature

The primary goal of this section is to construct the succinct unary signature $\langle 1\rangle$ of type τ_1. However, this is not always possible. For example, the succinct ternary signature $\langle 0, 0, 1\rangle = \mathrm{AD}_{3,3}$ of type τ_3 (on domain size 3) cannot construct $\langle 1\rangle$. This follows from the parity condition (Lemma 4.4). In such cases, we show that the problem is either computable in polynomial time or #P-hard without the help of additional signatures.

Lemma 8.1 handles two easy cases for which it is possible to construct $\langle 1\rangle$.

Lemma 8.1 *Suppose $\kappa \geq 3$ is the domain size and $a, b, c \in \mathbb{C}$. Let $\mathcal{F}$ be a signature set containing the succinct ternary signature $\langle a, b, c\rangle$ of type τ_3. If $a + (\kappa - 1)b \neq 0$ or $[2b + (\kappa - 2)c][b^2 - 4bc - (\kappa - 3)c^2] \neq 0$, then*

$$\text{Pl-Holant}(\mathcal{F} \cup \{\langle 1\rangle\}) \leq_T \text{Pl-Holant}(\mathcal{F}),$$

where $\langle 1\rangle$ is a succinct unary signature of type τ_1.

Proof Suppose $a + (\kappa - 1)b \neq 0$. Consider the gadget in Fig. 16a. We assign $\langle a, b, c\rangle$ to its vertex. By Lemma 11.1, this gadget has the succinct unary signature $\langle u\rangle$ of type τ_1, where $u = a + (\kappa - 1)b$. Since $u \neq 0$, this signature is equivalent to $\langle 1\rangle$.

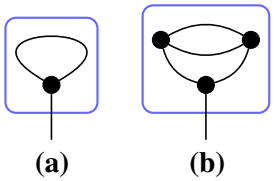


Fig. 16 Two simple unary gadgets **a** is a simple self-loop and **b** contains parallel edges

Otherwise, $a + (\kappa - 1)b = 0$, and $[2b + (\kappa - 2)c][b^2 - 4bc - (\kappa - 3)c^2] \neq 0$. Consider the gadget in Fig. 16b. We assign $\langle a, b, c\rangle$ to all three vertices. By Lemma 11.1, this gadget has the succinct unary signature $\langle u'\rangle$ of type τ_1, where $u' = -(\kappa - 1)(\kappa - 2)[2b + (\kappa - 2)c][b^2 - 4bc - (\kappa - 3)c^2]$. Since $u' \neq 0$, this signature is equivalent to $\langle 1\rangle$. □

One of the failure conditions of Lemma 8.1 is when both $a+(\kappa-1)b = 0$ and $b^2-4bc-(\kappa-3)c^2 = 0$ hold. In this case, $\langle a, b, c\rangle = c\langle -(\kappa-1)(2\pm\sqrt{\kappa+1}), 2\pm\sqrt{\kappa+1}, 1\rangle$. If $c = 0$, then $a = b = c = 0$ and the signature is trivial. Otherwise, $c \neq 0$. Then up to a nonzero factor of c, this signature further simplifies to $\mathrm{AD}_{3,3}$ by taking the minus sign when $\kappa = 3$. Just like $\mathrm{AD}_{3,3}$, we show (in Lemma 8.2) that all of these signatures are #P-hard.

Similar to the proof of Theorem 4.8, we prove the hardness in Lemma 8.2 by reducing from counting weighted Eulerian partitions.

Lemma 8.2 *Suppose $\kappa \geq 3$ is the domain size and $a, b, c \in \mathbb{C}$. Let $\langle a, b, c\rangle$ be a succinct ternary signature of type τ_3. If $a + (\kappa - 1)b = 0$ and $b^2 - 4bc - (\kappa - 3)c^2 = 0$, then*

$$\langle a, b, c\rangle = c\langle -(\kappa - 1)(2 + \varepsilon\sqrt{\kappa + 1}), 2 + \varepsilon\sqrt{\kappa + 1}, 1\rangle,$$

where $\varepsilon = \pm 1$, and Pl-Holant$(\langle a, b, c\rangle)$ *is* #P*-hard unless $c = 0$, in which case, the problem is computable in polynomial time.*

Proof If $c = 0$, then $a = b = c = 0$ so the output is always 0. Otherwise, $c \neq 0$. Up to a nonzero factor of c, $\langle a, b, c\rangle$ can be written as $\langle -(\kappa - 1)(2 + \varepsilon\sqrt{\kappa + 1}), 2 + \varepsilon\sqrt{\kappa + 1}, 1\rangle$ under the given assumptions, where $\varepsilon = \pm 1$.

Suppose $\kappa = 3$. If $\varepsilon = -1$, then we have $\langle 0, 0, 1\rangle = \mathrm{AD}_{3,3}$ and we are done by Theorem 4.8. Otherwise, $\varepsilon = 1$ and we have $\langle 8, -4, -1\rangle$. Let $T = \frac{1}{3}\begin{bmatrix} 1 & -2 & -2 \\ -2 & 1 & -2 \\ -2 & -2 & 1 \end{bmatrix}$, which is an orthogonal matrix. It follows from Theorem 3.3 and Lemma 11.6 that

$$\text{Pl-Holant}(\langle 8, -4, -1\rangle) \equiv_T \text{Pl-Holant}(T^{\otimes 3}\langle 8, -4, -1\rangle) \equiv_T \text{Pl-Holant}(\langle 0, 0, 1\rangle),$$

so again we are done by Theorem 4.8.

Now we suppose $\kappa \geq 4$. Let $g = \langle 2, 0, 1, 0, 0, 0, 1, 0, 0\rangle$ be a succinct quaternary signature of type τ_4. We reduce from Pl-Holant(g) to Pl-Holant$(\langle a, b, c\rangle)$. Then by Corollary 7.13, Pl-Holant$(\langle a, b, c\rangle)$ is #P-hard. We write this signature as $\langle -(\kappa - 1)\gamma, \gamma, 1\rangle$, where $\gamma = 2 + \varepsilon\sqrt{\kappa + 1}$. Consider the gadget in Fig. 17. We assign $\langle -(\kappa - 1)\gamma, \gamma, 1\rangle$ to both vertices. By Lemma 11.3, up to a nonzero factor of $\gamma - 1$, this gadget has the succinct quaternary signature f of type τ_4, where

$$f = \big\langle (\kappa - 1)(\gamma - 3)\gamma^2,\ -(\kappa - 2)\gamma,\ 3\gamma - 1,\ 2\gamma,\ 3\gamma - 1,\ 2\gamma,\\ -(\gamma - 3)\gamma^2,\ 2\gamma,\ \gamma + 1\big\rangle.$$

Now consider the recursive construction in Fig. 6. We assign f to all vertices. Let f_s be the succinct signature of type τ_4 for the sth gadget in this recursive construction.

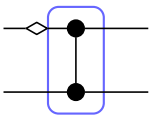


Fig. 17 Quaternary gadget used in the interpolation construction below. All vertices are assigned $\langle -(\kappa - 1)\gamma, \gamma, 1\rangle$

The initial signature, which is just two parallel edges, has the succinct signature $f_0 = \langle 1, 0, 0, 0, 0, 0, 1, 0, 0\rangle$ of type τ_4. We can express f_s as $M^s f_0$, where M is the matrix in Table 2.

Consider an instance Ω of Pl-Holant(g). Suppose g appears n times in Ω. We construct from Ω a sequence of instances Ω_s of Pl-Holant(f) indexed by $s \geq 0$. We obtain Ω_s from Ω by replacing each occurrence of g with the gadget f_s.

We can express M as $(\gamma-1)^3 P^{-1}\Lambda P$, where P is the matrix in Table 3,

$$\Lambda = \mathrm{diag}(-1, -1, -1, -1, \kappa-2, \kappa-2, \kappa-1, \kappa-1, \lambda),$$

and $\lambda = \frac{(\kappa-2)(\kappa+2\gamma-4)}{(\gamma-1)^2}$. The rows of P are linearly independent since

$$\det(P) = (\kappa-1)(\kappa-2)^2(\gamma-1)^6(\gamma-3)^3\gamma \neq 0.$$

For $1 \leq i \leq 9$, let r_i be the ith row of P. Notice that the initial signature f_0 and the target signature g are orthogonal to the same set of row eigenvectors of M, namely $\{r_1, r_2, r_3, r_5, r_7, r_9\}$. Up to a common factor of $(\gamma-1)^3$, the eigenvalues for M corresponding to r_4, r_6, and r_8 (the three row eigenvectors of M *not* orthogonal to f_0) are -1, $\kappa-2$, and $\kappa-1$, respectively. Since $\kappa \geq 4$, $\kappa-2$ and $\kappa-1$ are relatively prime and greater than 1, so these three eigenvalues satisfy the lattice condition. Thus by Lemma 6.6, we can interpolate g as desired. □

Remark Although the matrices in Table 2 and Table 3 seem large, they are probably the smallest possible to succeed in this recursive quaternary construction. In fact, for quaternary signatures one would normally expect these matrices to be even larger since there are typically fifteen different entries in a domain invariant signature of arity 4.

The other failure condition of Lemma 8.1 is when both $a+(\kappa-1)b = 0$ and $2b+(\kappa-2)c = 0$ hold. In this case, $\langle a, b, c\rangle = c\langle(\kappa-1)(\kappa-2), -(\kappa-2), 2\rangle$. If this signature is connected to $\langle 1\rangle$, then the first entry of the resulting succinct binary signature of type τ_2 is $(\kappa-1)(\kappa-2)\cdot 1-(\kappa-2)\cdot(\kappa-1) = 0$ while the second entry is $-(\kappa-2)\cdot 2+2\cdot(\kappa-2) = 0$. That is, the resulting binary signature is identically 0. This suggests we apply a holographic transformation such that the support of the resulting signature is only on $\kappa-1$ of the domain elements.

If $c = 0$, then $a = b = c = 0$ and the signature is trivial. Otherwise, $c \neq 0$. If $\kappa = 3$, then up to a nonzero factor of c, this signature further simplifies to $\langle 2, -1, 2\rangle$, which is tractable by case 3 of Corollary 5.2. Otherwise, $\kappa \geq 4$, and we show the problem is #P-hard.

Lemma 8.3 *Suppose $\kappa \geq 4$ is the domain size. Let $f = \langle(\kappa-1)(\kappa-2), -(\kappa-2), 2\rangle$ be a succinct ternary signature of type τ_3. Then* Pl-Holant(f) *is* #P*-hard.*

Proof Consider the matrix $T = \left[\begin{smallmatrix} 1 & \mathbf{1} \\ \mathbf{1} & T' \end{smallmatrix}\right] \in \mathbb{C}^{\kappa\times\kappa}$, where $T' = yJ_{\kappa-1} + (x-y)I_{\kappa-1}$ with $x = -\frac{\kappa+\sqrt{\kappa}-1}{\sqrt{\kappa}+1}$ and $y = \frac{1}{\sqrt{\kappa}+1}$. After scaling by $\frac{1}{\sqrt{\kappa}}$, we claim that T is an orthogonal matrix.

Let r_i be the ith row of $\frac{1}{\sqrt{\kappa}}T$. First we compute the diagonal entries of $\frac{1}{\kappa}TT^{\mathrm{T}}$. Clearly $r_1 r_1^{\mathrm{T}} = 1$. For $2 \leq i \leq \kappa$, we have

$$r_i r_i^{\mathrm{T}} = \frac{1}{\kappa}\left[1 + x^2 + (\kappa-2)y^2\right] = \frac{1}{\kappa}\left[1 + \frac{(\kappa+\sqrt{\kappa}-1)^2}{(\sqrt{\kappa}+1)^2} + \frac{\kappa-2}{(\sqrt{k}+1)^2}\right] = 1.$$

Table 2 The recurrence matrix M, up to a factor of $(\gamma+1)$, for the recursive construction in the proof of Lemma 8.2

$(\kappa-1)(\gamma-3)\gamma^2$	$-2(\kappa-2)(\kappa-1)\gamma$	$(\kappa-1)(3\gamma-1)$	$2(\kappa-2)(\kappa-1)\gamma$	0	0	0	0	0
0	$\kappa^2(\gamma+1)-4\kappa\gamma+2(\gamma+1)$	0	$(\kappa-2)(3\gamma-1)$	$-(\kappa-2)\gamma$	$-(\kappa-4)(\kappa-2)\gamma$	$-(\kappa-2)\gamma$	$-(\kappa-4)(\kappa-2)\gamma$	$2(\kappa-2)(\gamma-4)\gamma^2$
$3\gamma-1$	$2(\kappa-2)\gamma$	$\kappa^2(\gamma+1)+\kappa(3\gamma-5)-7\gamma+5$	$-2(\kappa-2)\gamma$	0	0	0	0	0
0	$2(3\gamma-1)$	0	$(\kappa-2)\gamma(\kappa+\gamma+1)$	2γ	$2(\kappa-4)\gamma$	2γ	$2(\kappa-4)\gamma$	$-4(\gamma-4)\gamma^2$
0	$-2(\kappa-2)\gamma$	0	$2(\kappa-2)\gamma$	$-(\gamma-3)\gamma^2$	$4(\kappa-2)\gamma$	$3\gamma-1$	$4(\kappa-2)\gamma$	$(\kappa-2)(\gamma-4)\gamma(\gamma+1)$
0	$-(\kappa-4)\gamma$	0	$(\kappa-4)\gamma$	2γ	$2(\kappa-4)\gamma$	2γ	$\kappa(3\gamma+1)-4(\gamma+1)$	$(\gamma-4)\gamma(\gamma\kappa+\kappa-4)$
0	$-2(\kappa-2)\gamma$	0	$2(\kappa-2)\gamma$	$3\gamma-1$	$4(\kappa-2)\gamma$	$-(\gamma-3)\gamma^2$	$4(\kappa-2)\gamma$	$(\kappa-2)(\gamma-4)\gamma(\gamma+1)$
0	$-(\kappa-4)\gamma$	0	$(\kappa-4)\gamma$	2γ	$\kappa(3\gamma+1)-4(\gamma+1)$	2γ	$2(\kappa-4)\gamma$	$(\gamma-4)\gamma(\gamma\kappa+\kappa-4)$
0	4γ	0	-4γ	$\gamma+1$	$2(\gamma\kappa+\kappa-4)$	$\gamma+1$	$2(\gamma\kappa+\kappa-4)$	$\kappa^2(\gamma+1)-2(\gamma+5)-2(5\gamma-11)$

Table 3 The matrix P whose rows are the row eigenvectors of the matrix in Table 2

0	0	0	0	1	-2	0	0	1
0	0	0	0	0	-1	0	1	0
0	-1	0	1	$-(\gamma-3)\gamma$	$(\gamma-3)\gamma$	0	0	0
0	0	0	0	-1	0	1	0	0
0	$(\kappa-2)\gamma$	0	$-(\kappa-2)\gamma$	0	$(\kappa-2)(\gamma-1)$	0	$(\kappa-2)(\gamma-1)$	$(\kappa-2)(\gamma-4)(\gamma-1)\gamma$
0	$-(\kappa-2)\gamma$	0	$(\kappa-2)\gamma$	$\gamma-1$	$(\kappa-2)(\gamma-1)$	$\gamma-1$	$(\kappa-2)(\gamma-1)$	0
0	2	0	$\kappa-2$	0	0	0	0	0
1	0	$\kappa-1$	0	0	0	0	0	0
$(\gamma-3)\gamma$	$\kappa^2+\kappa(2\gamma-7)-2(\gamma-5)$	$-(\gamma-3)\gamma$	$-\kappa^2-\kappa(2\gamma-7)+2(\gamma-5)$	$-(\gamma-3)\gamma$	$-(\kappa-4)(\gamma-3)\gamma$	$-(\gamma-3)\gamma$	$-(\kappa-4)(\gamma-3)\gamma$	$2(\gamma-4)(\gamma-3)\gamma^2$

Now we compute the off-diagonal entries. For $2 \le i \le \kappa$, we have

$$r_1 r_i^{\mathrm{T}} = \frac{1}{\kappa}\left[1 + x + (\kappa - 2)y\right] = \frac{1}{\kappa}\left[1 - \frac{\kappa + \sqrt{\kappa} - 1}{\sqrt{\kappa} + 1} + \frac{\kappa - 2}{\sqrt{\kappa} + 1}\right] = 0.$$

For $2 \le i < j \le \kappa$, we have

$$r_i r_j^{\mathrm{T}} = \frac{1}{\kappa}\left[1 + 2xy + (\kappa - 3)y^2\right] = \frac{1}{\kappa}\left[1 - \frac{2(\kappa + \sqrt{\kappa} - 1)}{(\sqrt{\kappa} + 1)^2} + \frac{\kappa - 3}{(\sqrt{k} + 1)^2}\right] = 0.$$

This proves the claim.

We apply a holographic transformation by T to the signature f to obtain $\widehat{f} = T^{\otimes 3} f$, which does not change the complexity of the problem by Theorem 3.3. Since the first row of T is a row of all 1's, the output of $\widehat{f}$ on any input containing the first domain element is 0. When restricted to the remaining $\kappa - 1$ domain elements, $\widehat{f}$ is domain invariant and symmetric, so it can be expressed as a succinct ternary signature of type τ_3.

Up to a nonzero factor of $\frac{\kappa^3}{(\sqrt{\kappa}+1)^2}$, it can be verified that $\widehat{f} = \langle -(\kappa-2)(2+\sqrt{\kappa}), 2+\sqrt{\kappa}, 1\rangle$. One way to do this is as follows. We write $f = \langle a, b, 2\rangle$ and $T = \left[\begin{smallmatrix} 1 & \mathbf{1} \\ \mathbf{1} & T' \end{smallmatrix}\right] \in \mathbb{C}^{\kappa \times \kappa}$, where $T' = yJ_{\kappa-1} + (x - y)I_{\kappa-1}$. The entries of $\widehat{f}$ are polynomials in κ with coefficients from $\mathbb{Z}[a, b, x, y]$. The degree of these polynomials is at most 3 since the arity of f is 3. After computing the entries of $\widehat{f}$ for domain sizes $3 \le \kappa \le 6$ as elements in $\mathbb{Z}[a, b, x, y]$, we interpolate the entries of $\widehat{f}$ as elements in $(\mathbb{Z}[a, b, x, y])[\kappa]$. Then replacing a, b, x, y with their actual values gives the claimed expression for the signature.

Since $\kappa \ge 4$, $\widehat{f}$ is #P-hard by Lemma 8.2, which completes the proof. □

At this point, we have achieved the broader goal of this section. For any $a, b, c \in \mathbb{C}$ and domain size $\kappa \ge 3$, either Pl-Holant($\langle a, b, c\rangle$) is computable in polynomial time, or Pl-Holant($\langle a, b, c\rangle$) is #P-hard, or we can use $\langle a, b, c\rangle$ to construct $\langle 1\rangle$ (i.e., the reduction Pl-Holant($\{\langle a, b, c, \rangle, \langle 1\rangle\}$ $\le_T$ Pl-Holant($\langle a, b, c\rangle$) holds). However, Lemma 8.3 is easily generalized, and this generalization turns out to be necessary to obtain our dichotomy.

Recall that connecting $f = \langle(\kappa - 1)(\kappa - 2), -(\kappa - 2), 2\rangle$ to $\langle 1\rangle$ results in an identically 0 signature. This suggests that we consider the more general signature $\widetilde{f} = \alpha\langle 1\rangle^{\otimes 3} + \beta f$ for any $\alpha \in \mathbb{C}$ and any nonzero $\beta \in \mathbb{C}$ since this does not change the complexity (as we argue in Corollary 8.4). For any $a, b, c \in \mathbb{C}$ satisfying $\mathfrak{B} = 0$ (cf. (7)), if $\alpha = \frac{2b+(\kappa-2)c}{\kappa}$ and $\beta = \frac{-b+c}{\kappa}$, then $\widetilde{f} = \langle a, b, c\rangle$. We note that the condition $\mathfrak{B} = 0$ can also be written as $(\kappa - 2)(b - c) = b - a$. We now prove a dichotomy for the signature $\widetilde{f}$.

Corollary 8.4 *Suppose $\kappa \ge 3$ is the domain size and $a, b, c \in \mathbb{C}$. Let $\langle a, b, c\rangle$ be a succinct ternary signature of type τ_3. If $\mathfrak{B} = 0$, then* Pl-Holant($\langle a, b, c\rangle$) *is* #P*-hard unless $b = c$ or $\kappa = 3$, in which case, the problem is computable in polynomial time.*

Proof If $b = c$, then by $\mathfrak{B} = 0$ we have $a = b = c$, which means the signature is degenerate and the problem is trivially tractable. If $\kappa = 3$, then $a = c$ and the problem is tractable by case 3 of Corollary 5.2. Otherwise $b \ne c$ and $\kappa \ge 4$.

Since $\mathfrak{B} = 0$, it can be verified that $\langle a, b, c\rangle = \frac{2b+(\kappa-2)c}{\kappa}\langle 1\rangle^{\otimes 3} + \frac{-b+c}{\kappa} f$, where $f = \langle(\kappa - 1)(\kappa - 2), -(\kappa - 2), 2\rangle$. We show that Pl-Holant($\langle a, b, c\rangle$) is #P-hard iff Pl-Holant(f) is. Since Pl-Holant(f) is #P-hard by Lemma 8.3, this proves the result.

Let $G = (V, E)$ be a connected planar 3-regular graph with $n = |V|$ and $m = |E|$. We can view Pl-Holant(G; $\langle a, b, c\rangle$) as a sum of 2^n Holant computations using the signatures $\alpha\langle 1\rangle^{\otimes 3}$ and βf. Each of these Holant computations considers a different assignment of

either $\alpha\langle 1\rangle^{\otimes 3}$ or βf to each vertex. Since connecting f to $\langle 1\rangle$ gives an identically 0 signature, if any connected signature grid contains both $\alpha\langle 1\rangle^{\otimes 3}$ and βf, then that particular Holant computation is 0. This is because a vertex of degree three assigned $\langle 1\rangle^{\otimes 3}$ is equivalent to three vertices of degree one connected to the same three neighboring vertices and each assigned $\langle 1\rangle$. There are only two possible assignments that could be nonzero. If all vertices are assigned $\alpha\langle 1\rangle^{\otimes 3}$, then the Holant is $\alpha^n\kappa^m$. Otherwise, all vertices are assigned βf and the Holant is β^n Pl-Holant$(G;f)$. Thus, Pl-Holant$(G;\alpha\langle 1\rangle^{\otimes 3}+\beta f) = \alpha^n\kappa^m+\beta^n$ Pl-Holant$(G;f)$. Since $\beta\neq 0$, one can solve for either Holant value given the other. □

9 Interpolating all binary signatures of type τ_2

In this section, we show how to interpolate all binary succinct signatures of type τ_2 in most settings. We use two general techniques to achieve this goal. In the first subsection, we use a generalization of the anti-gadget technique that creates a multitude of gadgets. They are so numerous that one is most likely to succeed. In the second subsection, we introduce a new technique called *Eigenvalue Shifted Triples* (ESTs). These generalize the technique of Eigenvalue Shifted Pairs from [43], and we use EST to interpolate binary succinct signatures in cases where the anti-gadget technique cannot handle. There are a few isolated problems for which neither technique works. However, these problems are easily handled separately in Lemma 12.1 in "Appendix 2".

From Sect. 8, every problem fits into one of three cases: either (1) the problem is tractable, (2) the problem is #P-hard, or (3) we can construct the succinct unary signature $\langle 1\rangle$ of type τ_1. Thus, many results in this section assume that $\langle 1\rangle$ is available.

9.1 E pluribus unum

we use Lemma 4.12 to prove our interpolation results. The main technical difficulty is to satisfy the third condition of Lemma 4.12, which is to prove that some recurrence matrix (that defines a sequence of gadgets) has infinite order up to a scalar. When the matrix has a finite order up to a scalar, we can utilize this failure condition to our advantage by constructing an anti-gadget [17], which is the "last" gadget with a distinct signature (up to a scalar) in the infinite sequence of gadgets. To make sure that we construct a multitude of nontrivial gadgets without cancelation, we put the anti-gadget inside another gadget (contrast the gadget in Fig. 18 with the gadget in Fig. 19b). From among this plethora of gadgets, at least one must succeed under the right conditions.

Although this idea works quite well in that some gadget among those constructed does succeed, we still must prove that one such gadget succeeds in every setting. We aim to exhibit a recurrence matrix whose ratio of eigenvalues is not a root of unity. We consider three related recurrence matrices at once. The next two lemmas consider two similar situations involving the eigenvalues of three such matrices. When applied, these lemmas show that some recurrence matrix must have eigenvalues with distinct complex norms, even though exactly which one among them succeeds may depend on the parameters in a complicated way.

Lemma 9.1 *Let $d_0, d_1, d_2, \Psi \in \mathbb{C}$. If d_0, d_1, and d_2 have the same argument but are distinct, then for all $\rho \in \mathbb{R}$, there exists $i \in \{0,1,2\}$ such that $|\Psi + d_i| \neq \rho$.*

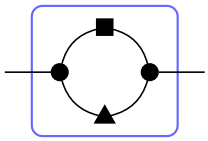


Fig. 18 Binary gadget that generalizes the anti-gadget technique. The *circle* vertices are assigned $\langle a, b, c\rangle$, while the *square* and *triangle* vertices are each assigned the signature of some gadget

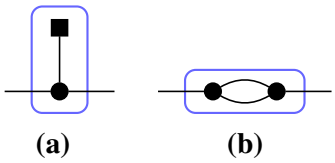


Fig. 19 Binary gadgets used to interpolate any succinct binary signature of type τ_2. In **a** *circle* vertices are assigned $\langle a, b, c\rangle$, and the *square* vertex is assigned $\langle 1\rangle$. In **b** Both circle vertices are assigned $\langle a, b, c\rangle$

Proof Assume to the contrary that there exists $\rho \in \mathbb{R}$ such that $|\Psi + d_i| = \rho$ for every $i \in \{0, 1, 2\}$. In the complex plane, consider the circle centered at the origin of radius ρ. Each $\Psi + d_i$ is a distinct point on this circle as well as a distinct point on a common line through Ψ. However, the line intersects the circle in at most two points, a contradiction. □

Lemma 9.2 *Let $d_0, d_1, d_2, \Psi \in \mathbb{C}$. If d_0, d_1, and d_2 have the same complex norm but are distinct and $\Psi \neq 0$, then for all $\rho \in \mathbb{R}$, there exists $i \in \{0, 1, 2\}$ such that $|\Psi + d_i| \neq \rho$.*

Proof Let $\ell = |d_0|$. Assume to the contrary that there exists $\rho \in \mathbb{R}$ such that $|\Psi + d_i| = \rho$ for every $i \in \{0, 1, 2\}$. In the complex plane, consider the circle centered at the origin of radius ρ and the circle centered at Ψ of radius ℓ. Since $\Psi \neq 0$, these circles are distinct. Each $\Psi + d_i$ is a distinct point on both circles. However, these circles intersect in at most two points, a contradiction. □

Now we use Lemma 9.1 and Lemma 9.2 as well as our generalization of the anti-gadget technique to establish a crucial lemma.

Lemma 9.3 *Suppose $\kappa \geq 3$ is the domain size and $a, b, c, \omega \in \mathbb{C}$. Let $\mathcal{F}$ be a set of signatures containing the succinct binary signature $\langle \omega+\kappa-1, \omega-1\rangle$ of type τ_2 and the succinct ternary signature $\langle a, b, c\rangle$ of type τ_3. If the following three conditions are satisfied:*

1. *$\omega \notin \{0, \pm 1\}$,*
2. *$\mathfrak{B} \neq 0$, and*
3. *at least one of the following holds:*
   - *(i) $\mathfrak{C} = 0$ or*
   - *(ii) $\mathfrak{C}^2 = \omega^{2\ell}\mathfrak{B}^2$ for some $\ell \in \{0, 1\}$ but either $\mathfrak{C}^2 \neq \mathfrak{A}^2$ or $\kappa \neq 3$,*

then

$$\text{Pl-Holant}(\mathcal{F} \cup \{\langle x, y\rangle\}) \leq_T \text{Pl-Holant}(\mathcal{F})$$

for any $x, y \in \mathbb{C}$, where $\langle x, y\rangle$ is a succinct binary signature of type τ_2.

We use this lemma to establish that various 2-by-2 recurrence matrices have infinite order modulo scalars. When applied, ω will be the ratio of two eigenvalues, one of which is a multiple of $\mathfrak{B}$ or $\mathfrak{B}^2$ by a nonzero function of κ.

Proof of Lemma 9.3 Let $\Phi = \frac{\mathfrak{C}^2}{\mathfrak{B}^2}$ and $\Psi = \frac{(\kappa-2)\mathfrak{A}^2}{\mathfrak{B}^2}$. Consider the recursive construction in Fig. 7. After scaling by a nonzero factor of κ, we assign $f = \frac{1}{\kappa}\langle \omega + \kappa - 1, \omega - 1\rangle$ to every vertex. Let f_s be the succinct binary signature of type τ_2 for the sth gadget in this construction. We can express f_s as $M^s \left[\begin{smallmatrix}1\\0\end{smallmatrix}\right]$, where $M = \frac{1}{\kappa}\left[\begin{smallmatrix}\omega+\kappa-1 & (\kappa-1)(\omega-1)\\ \omega-1 & (\kappa-1)\omega+1\end{smallmatrix}\right] = \left[\begin{smallmatrix}1 & 1-\kappa\\ 1 & 1\end{smallmatrix}\right]\left[\begin{smallmatrix}\omega & 0\\ 0 & 1\end{smallmatrix}\right]\left[\begin{smallmatrix}1 & 1-\kappa\\ 1 & 1\end{smallmatrix}\right]^{-1}$ by Lemma 4.11. Then $f_s = \frac{1}{\kappa}\langle \omega^s + \kappa - 1, \omega^s - 1\rangle$. The eigenvalues of M are 1 and ω, so the determinant of M is $\omega \neq 0$. If ω is not a root of unity, then we are done by Corollary 4.13.

Otherwise, suppose ω is a primitive root of unity of order n. Since $\omega \neq \pm 1$ by assumption, $n \geq 3$. Now consider the gadget in Fig. 18. We assign $\langle a, b, c\rangle$ to the circle vertices, $f_r = \frac{1}{\kappa}\langle \omega^r + \kappa - 1, \omega^r - 1\rangle$ to the square vertex, and $f_s = \frac{1}{\kappa}\langle \omega^s + \kappa - 1, \omega^s - 1\rangle$ to the triangle vertex, where $r, s \geq 0$ are parameters of our choice. By Lemma 11.5, up to a nonzero factor of $\frac{\mathfrak{B}^2}{\kappa}$, this gadget has the succinct binary signature

$$f(r,s) = \tfrac{1}{\kappa}\langle \Phi\omega^{r+s} + (\kappa-1)(\omega^r + \omega^s + \Psi + 1),\ \Phi\omega^{r+s} - (\omega^r + \omega^s + \Psi + 1) + \kappa\rangle$$

of type τ_2. Consider using this gadget in the recursive construction of Fig. 7. Let $f_t(r, s)$ be the succinct binary signature of type τ_2 for the tth gadget in this recursive construction. Then $f_1(r, s) = f(r, s)$ and $f_t(r, s) = (M(r, s))^t \left[\begin{smallmatrix}1\\0\end{smallmatrix}\right]$, where the eigenvalues of $M(r, s)$ are $\Phi\omega^{r+s} + \kappa - 1$ and $\omega^r + \omega^s + \Psi$ by Lemma 4.11. Thus, the determinant of $M(r, s)$ is $(\Phi\omega^{r+s} + \kappa - 1)(\omega^r + \omega^s + \Psi)$. Since Φ is either 0 or a power of ω by condition 3, the first factor is nonzero for any choice of r and s. However, for some r and s, it might be that $g(r, s) = \omega^r + \omega^s + \Psi = 0$.

Suppose $\Psi = 0$. We consider the two possible cases of Φ in order to finish the proof under this assumption.

1. Suppose $\Phi = 0$. Consider the gadget $M(0, 1)$. The determinant of $M(0, 1)$ is nonzero since $g(0, 1) \neq 0$ and the ratio of its eigenvalues is not a root of unity because they have distinct complex norms. Thus, we are done by Corollary 4.13.
2. Suppose $\Phi = \omega^{2\ell}$ for some $\ell \in \{0, 1\}$. Consider the gadget $M(n - \ell, n - \ell)$. The determinant of $M(n - \ell, n - \ell)$ is nonzero since $g(n - \ell, n - \ell) \neq 0$ and the ratio of its eigenvalues is not a root of unity because they have distinct complex norms. Thus, we are done by Corollary 4.13.

Otherwise, $\Psi \neq 0$. We claim that $g(r, s) = 0$ can hold for at most one choice of $r, s \in \mathbb{Z}_n$ (modulo the swapping of r and s). To see this, consider r_1, s_1, r_2, s_2 such that $g(r_1, s_1) = 0 = g(r_2, s_2)$. Then $\omega^{r_1} + \omega^{s_1} = -\Psi = \omega^{r_2} + \omega^{s_2}$. By taking complex norms and applying the law of cosines, we have $\cos\theta_1 = \cos\theta_2$, where $\theta_j = \arg(\omega^{s_j - r_j})$ is the angle from ω^{r_j} to ω^{s_j} for $j \in \{1, 2\}$. Thus, $\theta_1 = \pm\theta_2$. Since $\Psi \neq 0$, we have $\theta_1 \neq \pm\pi$. If $\theta_1 = \theta_2$, then $\omega^{r_1}(1 + e^{i\theta_1}) = \omega^{r_2}(1 + e^{i\theta_1})$. Since $\theta_1 \neq \pm\pi$, the factor $1 + e^{i\theta_1}$ is nonzero. After dividing by this factor, we conclude that $r_1 = r_2$ and thus $s_1 = s_2$. Otherwise, $\theta_1 = -\theta_2$. Then $\omega^{r_1}(1 + e^{i\theta_1}) = \omega^{s_2}(1 + e^{i\theta_1})$, and we conclude that $r_1 = s_2$ and $s_1 = r_2$. This proves the claim.

Suppose $n \geq 4$ and let $S_0 = \{(0, 0), (1, n-1), (2, n-2)\}$ and $S_1 = \{(1, 1), (2, 0), (3, n-1)\}$. Then $g(r, s) = 0$ holds for at most one $(r, s) \in S_0 \cup S_1$. In particular, $g(r, s)$ is either nonzero

for all $(r, s) \in S_0$ or nonzero for all $(r, s) \in S_1$. Pick $j \in \{0, 1\}$ such that $g(r, s)$ is nonzero for all $(r, s) \in S_j$. By Lemma 9.1 with $d_i = (\omega^i + \omega^{-i})\omega^j$ and $\rho = |\Phi\omega^{2j} + \kappa - 1|$, there exists some $(r, s) \in S_j$ such that the eigenvalues of $M(r, s)$ have distinct complex norms, so we are done by Corollary 4.13.

Otherwise, $n = 3$. We consider the two possible cases of Φ in order to finish the proof.

1. Suppose $\Phi = 0$. Let $S_j = \{(0, j), (1, j + 1), (2, j + 2)\}$. Then $g(r, s) = 0$ holds for at most one $(r, s) \in S_0 \cup S_1$. In particular, $g(r, s)$ is either nonzero for all $(r, s) \in S_0$ or nonzero for all $(r, s) \in S_1$. Pick $j \in \{0, 1\}$ such that $g(r, s)$ is nonzero for all $(r, s) \in S_j$. By Lemma 9.2 with $d_i = (1 + \omega^j)\omega^i$ and $\rho = \kappa - 1$, there exists some $(r, s) \in S_j$ such that the eigenvalues of $M(r, s)$ have distinct complex norms, so we are done by Corollary 4.13.
2. Suppose $\Phi = \omega^{2\ell}$ for some $\ell \in \{0, 1\}$ but either $\mathfrak{C}^2 \neq \mathfrak{A}^2$ or $\kappa \neq 3$. Note that this is equivalent to $\Phi \neq \Psi$ or $\kappa \neq 3$. Consider the set $S = \{(0, 0), (0, 1), (0, 2), (1, 1), (1, 2), (2, 2)\}$. If there exists some $(r, s) \in S$ such that $g(r, s) \neq 0$ and the eigenvalues of $M(r, s)$ have distinct complex norms, then we are done by Corollary 4.13.

Otherwise, for every $(r, s) \in S$, either $g(r, s) = 0$ or the eigenvalues of $M(r, s)$ have the same complex norm. If the latter condition were to always hold, then we would have

$$\begin{aligned} |2 + \Psi| &= \left|\omega^{2\ell} + \kappa - 1\right| = |-1 + \Psi|, \\ \left|2\omega^2 + \Psi\right| &= \left|\omega^{2\ell+1} + \kappa - 1\right| = \left|-\omega^2 + \Psi\right|, \text{ and} \\ |2\omega + \Psi| &= \left|\omega^{2\ell+2} + \kappa - 1\right| = |-\omega + \Psi|, \end{aligned}$$

where each equality corresponds to one of the six $M(r, s)$ having eigenvalues of equal complex norm for $(r, s) \in S$. Of the six equalities, at most one may not hold since $g(r, s) = 0$ for at most one $(r, s) \in S$. Since $n = 3$, two of the three terms of the form $|\omega^{2\ell+m} + \kappa - 1|$ must be equal, so we can write the stronger condition

$$\begin{array}{ccccc} \left|2\omega^2 + \Psi\omega^\ell\right| & = & |\omega + \kappa - 1| & = & \left|-\omega^2 + \Psi\omega^\ell\right| \\ & & \| & & \\ \left|2\omega + \Psi\omega^\ell\right| & = & \left|\omega^2 + \kappa - 1\right| & = & \left|-\omega + \Psi\omega^\ell\right|. \end{array} \tag{13}$$

As it is, one of the horizontal equalities in (13) may not hold. However, even without one of these equalities, we can still reach a contradiction.

We show that $\Psi\omega^\ell \in \mathbb{R}$ even if one of the equalities in (13) does not hold. In fact, either the left or the right half of the equalities in (13) hold. In the first case, $|2\omega^2 + \Psi\omega^\ell| = |2\omega + \Psi\omega^\ell|$ holds and we get $\Psi\omega^\ell \in \mathbb{R}$. Similarly in the second case, $|-\omega^2 + \Psi\omega^\ell| = |-\omega + \Psi\omega^\ell|$ holds and we get $\Psi\omega^\ell \in \mathbb{R}$ as well. Next, we use real and imaginary parts to calculate the complex norms even if one of the equalities in (13) does not hold. Either the top half of the equalities hold and thus $|2\omega^2 + \Psi\omega^\ell| = |-\omega^2 + \Psi\omega^\ell|$, or the bottom half of the equalities hold and thus $|2\omega + \Psi\omega^\ell| = |-\omega + \Psi\omega^\ell|$. In any case, it readily follows that $\Psi\omega^\ell = 1$. This implies $\Psi = \omega^{2\ell}$, so we can rewrite (13) as

$$\begin{array}{ccc} \sqrt{3} = |\omega + \kappa - 1| & = & \sqrt{3} \\ & \| & \\ \sqrt{3} = \left|\omega^2 + \kappa - 1\right| & = & \sqrt{3}, \end{array}$$

where at most one equation may not hold. This forces $\kappa = 3$. However, $\Phi = \omega^{2\ell} = \Psi$ and $\kappa = 3$ is a contradiction. □

The previous lemma is strong enough to handle the typical case.

Lemma 9.4 *Suppose $\kappa \geq 3$ is the domain size and $a, b, c \in \mathbb{C}$. Let $\mathcal{F}$ be a signature set containing the succinct ternary signature $\langle a, b, c\rangle$ of type τ_3 and the succinct unary signature $\langle 1\rangle$ of type τ_1. If*

1. $\mathfrak{B} \neq 0$,
2. $\mathfrak{C} \neq 0$,
3. $\mathfrak{C}^2 \neq \mathfrak{B}^2$, *and*
4. *either* $\mathfrak{C}^2 \neq \mathfrak{A}^2$ *or* $\kappa \neq 3$,

then

$$\text{Pl-Holant}(\mathcal{F} \cup \{\langle x, y\rangle\}) \leq_T \text{Pl-Holant}(\mathcal{F})$$

for any $x, y \in \mathbb{C}$, where $\langle x, y\rangle$ is a succinct binary signature of type τ_2.

Proof Let $\omega = \frac{\mathfrak{C}}{\mathfrak{B}}$, which is well defined. Consider the gadget in Fig. 19a. We assign $\langle a, b, c\rangle$ to the circle vertex and $\langle 1\rangle$ to the square vertex. Up to a nonzero factor of $\frac{\mathfrak{B}}{\kappa}$, this gadget has the succinct binary signature

$$\frac{\kappa}{\mathfrak{B}}\langle a + (\kappa - 1)b, 2b + (\kappa - 2)c\rangle = \langle \omega + \kappa - 1, \omega - 1\rangle$$

of type τ_2. Then we are done by Lemma 9.3 with $\ell = 1$ in case (ii) of condition 3. □

If $\mathfrak{B} = 0$, then we already know the complexity by Corollary 8.4. The other failure conditions from the previous lemma are:

$$\mathfrak{C} - \mathfrak{B} = \kappa[2b + (\kappa - 2)c] = 0; \tag{14}$$

$$\mathfrak{C} + \mathfrak{B} = 2a + 2(2\kappa - 3)b + (\kappa - 2)^2 c = 0; \tag{15}$$

$$\mathfrak{C} = 0; \tag{16}$$

$$\kappa = 3 \text{ and } \mathfrak{C} - \mathfrak{A} = 0, \quad \text{or equivalently} \quad \kappa = 3 \quad \text{and} \quad b = 0; \tag{17}$$

$$\kappa = 3 \text{ and } \mathfrak{C} + \mathfrak{A} = 0, \quad \text{or equivalently} \quad \kappa = 3 \quad \text{and} \quad 2a + 3b + 4c = 0. \tag{18}$$

Notice that these five failure conditions are *linear* in a, b, c.

By starting the proof with a different gadget, Lemma 9.3 can handle the first three failure conditions. The last two failure conditions require a new idea, Eigenvalue Shifted Triples, which we introduce in Sect. 9.2. In fact, these two cases are equivalent under an orthogonal holographic transformation.

The next lemma considers the failure condition in (14). Note that $\mathfrak{C} = \mathfrak{B}$ iff the signature can be written as $\langle 2a, -(\kappa-2)c, 2c\rangle$ up to a factor of 2. The first excluded case in Lemma 9.5 is handled by Corollary 8.4, and the last two excluded cases are tractable by Corollary 5.3.

Lemma 9.5 *Suppose $\kappa \geq 3$ is the domain size and $a, c \in \mathbb{C}$. Let $\mathcal{F}$ be a signature set containing the succinct ternary signature $\langle 2a, -(\kappa - 2)c, 2c\rangle$ of type τ_3 and the succinct unary signature $\langle 1\rangle$ of type τ_1. If*

1. $2a \neq (\kappa - 1)(\kappa - 2)c$,

2. $4a \neq (\kappa^2 - 6\kappa + 4)c$, *and*
3. $c \neq 0$,

then

$$\text{Pl-Holant}(\mathcal{F} \cup \{\langle x, y\rangle\}) \leq_T \text{Pl-Holant}(\mathcal{F})$$

for any $x, y \in \mathbb{C}$, *where* $\langle x, y\rangle$ *is a succinct binary signature of type* τ_2.

Proof Note that when $2b = -(\kappa-2)c$, we have $\mathfrak{B} = \mathfrak{C} = 2a-(\kappa-1)(\kappa-2)c$ by (14), which is nonzero by condition 1 of the lemma. Let $\omega_0 = 4a^2+(\kappa-2)[4ac+(2\kappa^2+\kappa-2)c^2]$ and assume $\omega_0 \neq 0$. Then let $\omega = \frac{\mathfrak{B}^2}{\omega_0} \neq 0$. By conditions 2 and 3, it follows that $\omega \neq 1$. Also we note that when $2b = -(\kappa-2)c$, we have $2\mathfrak{A} = 2a+(3\kappa-2)c$ and $2\mathfrak{C} = 2a-(\kappa-1)(\kappa-2)c$. By the same conditions, 2 and 3, we have $\mathfrak{C}^2 \neq \mathfrak{A}^2$. We further assume that $\omega \neq -1$, which is equivalent to $8a^2 - 4(\kappa-2)^2ac + (\kappa-2)(\kappa^3-2\kappa^2+6\kappa-4)c^2 \neq 0$.

Consider the gadget in Fig. 19b. We assign $\langle 2a, -(\kappa-2)c, 2c\rangle$ to the vertices. Up to a nonzero factor of $\frac{\omega_0}{\kappa}$, this gadget has the succinct binary signature

$$\begin{aligned}&\frac{\kappa}{\omega_0}\langle 4a^2+(\kappa-1)(\kappa-2)(3\kappa-2)c^2, \quad -(\kappa-2)[4ac-(\kappa^2-6\kappa+4)c^2]\rangle\\ &\quad = \langle \omega+\kappa-1, \quad \omega-1\rangle\end{aligned}$$

of type τ_2. Then we are done by Lemma 9.3 with $\ell = 0$ in case (ii) of condition 3.

Now we deal with the following exceptional cases.

1. If $\omega_0 = 0$, then $2a = -\left[\kappa-2\pm i\kappa\sqrt{2(\kappa-2)}\right]c$. Up to a nonzero factor of $-c$, we have $-\frac{1}{c}\langle 2a, -(\kappa-2)c, 2c\rangle = \langle \kappa-2\pm i\kappa\sqrt{2(\kappa-2)}, \kappa-2, -2\rangle$ and are done by case 1 of Lemma 12.1.
2. If $8a^2 - 4(\kappa-2)^2ac + (\kappa-2)(\kappa^3-2\kappa^2+6\kappa-4)c^2 = 0$, then $4a = \left[(\kappa-2)^2 \pm i\kappa\sqrt{\kappa^2-4}\right]c$. Up to a nonzero factor of $\frac{c}{2}$, we have

   $$\frac{2}{c}\langle 2a, -(\kappa-2)c, 2c\rangle = \langle (\kappa-2)^2 \pm i\kappa\sqrt{\kappa^2-4}, -2(\kappa-2), 4\rangle$$

   and are done by case 2 of Lemma 12.1. □

The next lemma considers the failure condition in (15). Note that $\mathfrak{C} = -\mathfrak{B}$ iff the signature can be written as $\langle -2(2\kappa-3)b-(\kappa-2)^2c, 2b, 2c\rangle$ up to a factor of 2. The first excluded case in Lemma 9.6 is handled by Corollary 8.4, and the last excluded case is tractable by Corollary 5.8.

Lemma 9.6 *Suppose* $\kappa \geq 3$ *is the domain size and* $a, b \in \mathbb{C}$. *Let* $\mathcal{F}$ *be a signature set containing the succinct ternary signature* $\langle -2(2\kappa-3)b-(\kappa-2)^2c, 2b, 2c\rangle$ *of type* τ_3 *and the succinct unary signature* $\langle 1\rangle$ *of type* τ_1. *If*

1. $2b \neq -(\kappa-2)c$ *and*
2. $\kappa \neq 4$ *or* $5b^2 + 2bc + c^2 \neq 0$,

then

$$\text{Pl-Holant}(\mathcal{F} \cup \{\langle x, y\rangle\}) \leq_T \text{Pl-Holant}(\mathcal{F})$$

for any $x, y \in \mathbb{C}$, *where* $\langle x, y\rangle$ *is a succinct binary signature of type* τ_2.

Proof Note that when $2a = -2(2\kappa-3)b - (\kappa-2)^2 c$, we have $\mathfrak{B} = -\mathfrak{C}$ by (15) and $2\mathfrak{B} = -\kappa[2b + (\kappa-2)c]$, which is nonzero by condition 1 of the lemma. Let $\omega_0 = 8(2\kappa-3)b^2 + (\kappa-2)\left[8(\kappa-3)bc + (\kappa^2-6\kappa+12)c^2\right]$ and assume $\omega_0 \neq 0$. Then let $\omega = \frac{\kappa[2b+(\kappa-2)c]^2}{\omega_0}$. By condition 1, $\omega \neq 0$. It can be shown that $\kappa[2b + (\kappa-2)c]^2 = \omega_0$ is equivalent to $(b-c)[3b + (\kappa-3)c] = 0$. Thus, assume $b \neq c$ and $3b \neq -(\kappa-3)c$. Then $\omega \neq 1$. Also we note that when $2a = -2(2\kappa-3)b - (\kappa-2)^2 c$, we have $2\mathfrak{A} = -\kappa[4b + (\kappa-4)c]$ and $2\mathfrak{C} = \kappa[2b + (\kappa-2)c]$. By the same assumptions, $b \neq c$ and $3b \neq -(\kappa-3)c$, we have $\mathfrak{C}^2 \neq \mathfrak{A}^2$. Further assume that $\omega \neq -1$, which is equivalent to $2(5\kappa-6)b^2 + (\kappa-2)[6(\kappa-2)bc + (\kappa^2-4\kappa+6)c^2] \neq 0$.

Consider the gadget in Fig. 19b. We assign $\langle -2(2\kappa-3)b - (\kappa-2)^2 c, 2b, 2c\rangle$ to the vertices. Up to a nonzero factor of $\frac{\omega_0}{4}$, this gadget has the succinct binary signature $\frac{1}{\omega_0}\langle x, y\rangle = \langle \omega + \kappa - 1, \omega - 1\rangle$ of type τ_2, where

$$x = 4(4\kappa^2 - 9\kappa + 6)b^2 + (\kappa-2)\left[4(\kappa-2)(2\kappa-3)bc + (\kappa^3 - 6\kappa^2 + 16\kappa - 12)c^2\right]$$

and

$$y = -4(\kappa-2)\left[3b^3 + (\kappa-6)bc - (\kappa-3)c^2\right].$$

Then we are done by Lemma 9.3 with $\ell = 0$ in case (ii) of condition 3.

Now we deal with the following exceptional cases.

1. If $\omega_0 = 0$, then we have $-4(2\kappa-3)b = \left[2(\kappa-3)(\kappa-2) \pm i\kappa\sqrt{2(\kappa-2)}\right]c$ but $\kappa \neq 4$ by condition 2 since otherwise $\omega_0 = 8(5b^2 + 2bc + c^2) \neq 0$. Up to a nonzero factor of $\frac{c}{2(2\kappa-3)}$,

$$\begin{aligned}&\frac{2(2\kappa-3)}{c}\langle -2(2\kappa-3)b - (\kappa-2)^2 c, 2b, 2c\rangle \\ &= \Big\langle -(2\kappa-3)\left[2(\kappa-2) \mp i\kappa\sqrt{2(\kappa-2)}\right], \\ &\quad -2\,(\kappa-3)(\kappa-2) \mp i\kappa\sqrt{2(\kappa-2)}, 4(2\kappa-3)\Big\rangle\end{aligned}$$

   and are done by case 3 of Lemma 12.1.
2. If $b = c$, then up to a nonzero factor of c, we have $\frac{1}{c}\langle -2(2\kappa-3)b - (\kappa-2)^2 c, 2b, 2c\rangle = \langle -\kappa^2 + 2, 2, 2\rangle$ and are done by case 4 Lemma 12.1.
3. If $3b = -(\kappa-3)c$, then up to a nonzero factor of $\frac{c}{3}$, we have $\frac{3}{c}\langle -2(2\kappa-3)b - (\kappa-2)^2 c, 2b, 2c\rangle = \langle \kappa^2 - 6\kappa + 6, -2(\kappa-3), 6\rangle$ and are done by case 5 of Lemma 12.1.
4. If $2(5\kappa-6)b^2 + (\kappa-2)[6(\kappa-2)bc + (\kappa^2-4\kappa+6)c^2] = 0$, then $-2(5\kappa-6)b = \left[3(\kappa-2)^2 \pm i\kappa\sqrt{\kappa^2-4}\right]c$. Up to a nonzero factor of $\frac{c}{5\kappa-6}$,

$$\begin{aligned}&\frac{5\kappa-6}{c}\langle -2(2\kappa-3)b - (\kappa-2)^2 c, 2b, 2c\rangle \\ &= \Big\langle (\kappa-3)(\kappa-2)^2 \pm i\kappa(2\kappa-3)\sqrt{\kappa^2-4}, \\ &\quad -3\,(\kappa-2)^2 \mp i\kappa\sqrt{\kappa^2-4}, 2(5\kappa-6)\Big\rangle\end{aligned}$$

   and are done by case 6 of Lemma 12.1. □

The next lemma considers the failure condition in (16). Note that $\mathfrak{C} = 0$ iff the signature can be written as $\langle -3(\kappa-1)b - (\kappa-1)(\kappa-2)c, b, c\rangle$. The excluded case in Lemma 9.7 is handled by Corollary 8.4.

Lemma 9.7 *Suppose $\kappa \geq 3$ is the domain size and $b, c \in \mathbb{C}$. Let $\mathcal{F}$ be a signature set containing the succinct ternary signature $\langle -3(\kappa-1)b-(\kappa-1)(\kappa-2)c, b, c\rangle$ of type τ_3 and the succinct unary signature $\langle 1\rangle$ of type τ_1. If $2b \neq -(\kappa-2)c$, then*

$$\text{Pl-Holant}(\mathcal{F} \cup \{\langle x, y\rangle\}) \leq_T \text{Pl-Holant}(\mathcal{F})$$

for any $x, y \in \mathbb{C}$, where $\langle x, y\rangle$ is a succinct binary signature of type τ_2.

Proof Note that when $a = -3(\kappa-1)b-(\kappa-2)(\kappa-1)c$, we have $\mathfrak{C} = 0$ and $2\mathfrak{B} = -\kappa[2b+(\kappa-2)c]$, which is nonzero by assumption. Let $\omega_0 = (9\kappa-10)b^2+(\kappa-2)[2(3\kappa-5)bc+(\kappa^2-4\kappa+5)c^2]$ and assume $\omega_0 \neq 0$. Then let $\omega = \frac{(\kappa-1)[2b+(\kappa-2)c]^2}{\omega_0}$. By assumption, $\omega \neq 0$. Assume $\omega \neq 1$, which is equivalent to $-(5\kappa-6)b^2-(\kappa-3)(\kappa-2)(2b-c)c \neq 0$. Further assume $\omega \neq -1$, which is equivalent to $(13\kappa-14)b^2+(\kappa-2)[2(5\kappa-7)bc+(2\kappa^2-7\kappa+7)c^2] \neq 0$.

Consider the gadget in Fig. 19b. We assign $\langle -3(\kappa-1)b-(\kappa-1)(\kappa-2)c, b, c\rangle$ to the vertices. Up to a nonzero factor of ω_0, this gadget has the succinct binary signature $\frac{1}{\omega_0}\langle x, y\rangle = \langle \omega+\kappa-1, \omega-1\rangle$ of type τ_2, where

$$x = (\kappa-1)\left\{3(3\kappa-2)b^2+(\kappa-2)\left[6bc+(\kappa^2-3\kappa+3)c^2\right]\right\} \quad \text{and}$$
$$y = -(5\kappa-6)b^2-(\kappa-3)(\kappa-2)(2b-c)c.$$

Then we are done by Lemma 9.3 via case (i) of condition 3.

Now we deal with the following exceptional cases.

1. If $\omega_0 = 0$, then $-(9\kappa-10)b = [(\kappa-2)(3\kappa-5) \pm i\kappa\sqrt{2(\kappa-2)}]c$. Up to a nonzero factor of $\frac{c}{9\kappa-10}$, we have

$$\begin{aligned}&\frac{9\kappa-10}{c}\langle -3(\kappa-1)b-(\kappa-1)(\kappa-2)c, b, c\rangle\\ &= \langle -(\kappa-1)\left[5(\kappa-2) \mp 3i\kappa\sqrt{2(\kappa-2)}\right],\\ &\quad -(\kappa-2)(3\kappa-5) \mp i\kappa\sqrt{2(\kappa-2)}, \quad 9\kappa-10\rangle\end{aligned}$$

   and we are done by case 7 of Lemma 12.1.
2. If $-(5\kappa-6)b^2-(\kappa-3)(\kappa-2)(2b-c)c = 0$, then $-(5\kappa-6)b = \left[(\kappa-3)(\kappa-2) \pm \kappa\sqrt{\kappa^2-5\kappa+6}\right]c$. Up to a nonzero factor of $-\frac{c}{5\kappa-6}$, we have

$$\begin{aligned}&-\frac{5\kappa-6}{c}\langle -3(\kappa-1)b-(\kappa-1)(\kappa-2)c, b, c\rangle\\ &= \langle (\kappa-1)\left[(\kappa-2)(2\kappa+3) \mp 3\kappa\sqrt{\kappa^2-5\kappa+6}\right],\\ &(\kappa-3)(\kappa-2) \pm \kappa\sqrt{\kappa^2-5\kappa+6}, \quad -5\kappa+6\rangle\end{aligned}$$

   and are done by case 8 Lemma 12.1.
3. If $(13\kappa-14)b^2+(\kappa-2)[2(5\kappa-7)bc+(2\kappa^2-7\kappa+7)c^2] = 0$, then $-(13\kappa-14)b = \left[(\kappa-2)(5\kappa-7) \pm i\kappa\sqrt{\kappa^2-\kappa-2}\right]c$. Up to a nonzero factor of $\frac{c}{13\kappa-14}$, we have

$$\begin{aligned}&\frac{13\kappa-14}{c}\langle -3(\kappa-1)b-(\kappa-1)(\kappa-2)c, b, c\rangle\\ &= \langle (\kappa-1)\left[(\kappa-2)(2\kappa-7) \pm 3i\kappa\sqrt{\kappa^2-\kappa-2}\right],\\ &\quad -(\kappa-2)(5\kappa-7) \mp i\kappa\sqrt{\kappa^2-\kappa-2}, \quad 13\kappa-14\rangle\end{aligned}$$

   and are done by case 9 of Lemma 12.1. □

9.2 Eigenvalue shifted triples

To handle failure conditions (17) and (18) from Lemma 9.4, we need another technique. We introduce an Eigenvalue Shifted Triple, which extends the concept of an Eigenvalue Shifted Pair.

Definition 9.8 (*Definition 4.6 in* [43]) A pair of nonsingular matrices $M, M' \in \mathbb{C}^{2\times 2}$ is called an *Eigenvalue Shifted Pair* if $M' = M + \delta I$ for some nonzero $\delta \in \mathbb{C}$, and M has distinct eigenvalues.

Eigenvalue Shifted Pairs were used in [43] to show that interpolation succeeds in most cases since these matrices correspond to some recursive gadget constructions and at least one of them usually has eigenvalues with distinct complex norms. In [43], it is shown that the interpolation succeeds unless the variables in question take real values. Then other techniques were developed to handle the real case. We use Eigenvalue Shifted Pairs in a stronger way. We exhibit three matrices such that any two form an Eigenvalue Shifted Pair. Provided that these shifts are linearly independent over $\mathbb{R}$, this is enough to show that interpolation succeeds for both real and complex settings of the variables. We call this an Eigenvalue Shifted Triple.

Definition 9.9 A trio of nonsingular matrices $M_0, M_1, M_2 \in \mathbb{C}^{2\times 2}$ is called an *Eigenvalue Shifted Triple* (EST) if M_0 has distinct eigenvalues and there exist nonzero $\delta_1, \delta_2 \in \mathbb{C}$ satisfying $\frac{\delta_1}{\delta_2} \notin \mathbb{R}$ such that $M_1 = M_0 + \delta_1 I$, and $M_2 = M_0 + \delta_2 I$.

We note that if M_0, M_1, and M_2 form an Eigenvalue Shifted Triple, then any permutation of the matrices is also an Eigenvalue Shifted Triple.

The proof of the next lemma is similar to the proof of Lemma 4.7 in [44], the full version of [43].

Lemma 9.10 *Suppose $\alpha, \beta, \delta_1, \delta_2 \in \mathbb{C}$. If $\alpha \neq \beta$, $\delta_1, \delta_2 \neq 0$, and $\frac{\delta_1}{\delta_2} \notin \mathbb{R}$, then $|\alpha| \neq |\beta|$ or $|\alpha + \delta_1| \neq |\beta + \delta_1|$ or $|\alpha + \delta_2| \neq |\beta + \delta_2|$.*

Proof Assume for a contradiction that $|\alpha| = |\beta|$, $|\alpha+\delta_1| = |\beta+\delta_1|$, and $|\alpha+\delta_2| = |\beta+\delta_2|$. After a rotation in the complex plane, we can assume that $\alpha = \overline{\beta}$. Note that all of our assumptions are unchanged by this rotation. For $i \in \{1, 2\}$, we have

$$\begin{aligned}(\alpha + \delta_i)\overline{(\alpha + \delta_i)} &= |\alpha + \delta_i|^2 \\ &= |\beta + \delta_i|^2 \\ &= (\beta + \delta_i)\overline{(\beta + \delta_i)} = (\overline{\alpha} + \delta_i)(\alpha + \overline{\delta_i}).\end{aligned}$$

This implies $(\overline{\alpha} - \alpha)(\overline{\delta_i} - \delta_i) = 0$. Since $\alpha \neq \beta = \overline{\alpha}$, we have $\delta_i \in \mathbb{R}$. Then $\frac{\delta_1}{\delta_2} \in \mathbb{R}$, a contradiction. □

The next lemma considers the failure condition in (17), which is $\kappa = 3$ and $b = 0$, so the signature has the form $\langle a, 0, c\rangle$. If $a = 0$, then the problem is already #P-hard by Theorem 4.8. If $c = 0$, then the problem is tractable by case 1 of Corollary 5.2. If $a^3 = c^3$, then the problem is tractable by Corollary 5.6.

Lemma 9.11 *Suppose the domain size is 3 and $a, c \in \mathbb{C}$. Let $\mathcal{F}$ be a signature set containing the succinct ternary signature $\langle a, 0, c\rangle$ of type τ_3 and the succinct unary signature $\langle 1\rangle$ of type τ_1. If $ac \neq 0$ and $a^3 \neq c^3$, then*

$$\text{Pl-Holant}(\mathcal{F} \cup \{\langle x, y\rangle\}) \leq_T \text{Pl-Holant}(\mathcal{F})$$

for any $x, y \in \mathbb{C}$, where $\langle x, y\rangle$ is a succinct binary signature of type τ_2.

Proof Assume $2a + c \neq 0$ and let $\omega = \frac{a^2+2c^2}{c(2a+c)}$. Assume $a^2 + 2c^2$ so that $\omega \neq 0$. Further assume $a^2 + 2ac + 3c^2 \neq 0$ so that $\omega^2 \neq 1$ as well as $a^2 + ac + 7c^2 \neq 0$ so that $\omega^3 \neq 1$. Note that these conclusions also require $a \neq c$ and $a^3 \neq c^3$, respectively.

Consider using the recursive construction in Fig. 20. The circle vertices are assigned $\langle a, 0, c\rangle$, and the square vertex is assigned $\langle 1\rangle$. Let $z = \frac{c}{a}$, which is well defined by assumption. The succinct signature of type τ_2 for the initial gadget N_0 in this construction is $\langle a, c\rangle$. Up to a nonzero factor of a, this signature is $f_0 = \frac{1}{a}\langle a, c\rangle = \langle 1, z\rangle$. Then up to a nonzero factor of $c(2a + c)$, the succinct signature of type τ_2 for the sth gadget in this construction is $f_s = \langle \omega^k, z\rangle = M^s f_0$, where

$$M = \frac{1}{c(2a+c)} \begin{bmatrix} a^2 + 2c^2 & 0 \\ 0 & c(2a+c) \end{bmatrix} = \begin{bmatrix} \omega & 0 \\ 0 & 1 \end{bmatrix}.$$

Clearly M is nonsingular. The determinant of $[f_0\, M f_0] = \left[\begin{smallmatrix} a & a\omega \\ c & c \end{smallmatrix}\right]$ is $z(1 - \omega) \neq 0$. If ω is not a root of unity, then we are done by Lemma 4.12.

Otherwise, suppose ω is a primitive root of unity of order n. By assumption, $n \geq 4$. Now consider the recursive construction in Fig. 7. We assign f_s to every vertex, where $s \geq 0$ is a parameter of our choice. Let $g_t(s)$ be the signature of the tth gadget in this recursive construction when using f_s. Then $g_1(s) = f_s$ and $g_t(s) = (N(s))^t \left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right]$, where $N(s) = \left[\begin{smallmatrix} \omega^s & 2z \\ z & \omega^s + z \end{smallmatrix}\right]$.

By Lemma 4.11, the eigenvalues of $N(s)$ are $\omega^s + 2z$ and $\omega^s - z$, which means the determinant of $N(s)$ is $(\omega^s + 2z)(\omega^s - z)$. Each eigenvalue can vanish for at most one value of $s \in \mathbb{Z}_n$ since both eigenvalues are linear polynomials in ω^s that are not identically 0. Furthermore, at least one of the eigenvalues never vanishes for all $s \in \mathbb{Z}_n$ since otherwise $1 = |z| = \frac{1}{2}$.

Thus, at most one matrix among $N(0), N(1), N(2)$, and $N(3)$ can be singular. Pick distinct $j, k, \ell \in \{0, 1, 2, 3\}$ such that $N(j)$, $N(k)$, and $N(\ell)$ are nonsingular. To finish the proof, we show that $N(j)$, $N(k)$, and $N(\ell)$ form an Eigenvalue Shifted Triple. Then by Lemma 9.10, at least one of the matrices has eigenvalues with distinct complex norms, so we are done by Corollary 4.13.

The eigenvalue shift from $N(j)$ to $N(k)$ is $\delta_{j,k} = \omega^j(\omega^{k-j} - 1)$, which is nonzero since j and k are distinct in $\mathbb{Z}_n$. Assume for a contradiction that $\frac{\delta_{j,k}}{\delta_{j,\ell}} \in \mathbb{R}$, which is equivalent to $\arg(\delta_{j,k}) = \arg(\pm\delta_{j,\ell})$. Then we have

$$\arg\left(\omega^{k-j} - 1\right) = \arg\left(\pm(\omega^{\ell-j} - 1)\right). \tag{19}$$

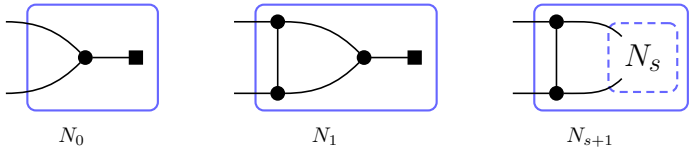


Fig. 20 Alternative recursive construction to interpolate a binary signature (cf. Fig. 7). The *circle* vertices are assigned $\langle a, b, c\rangle$, and the *square* vertex is assigned $\langle 1\rangle$

In the complex plane, any nonzero $x-1 \in \mathbb{C}$ with $|x|=1$ lies on the circle of radius 1 centered at $(-1,0)$. Such x satisfy $\frac{\pi}{2} < \arg(x-1) < \frac{3\pi}{2}$. Thus, the argument of $x-1$ is unique, even up to a sign, contradicting (19). Therefore, M_j, M_k, and M_ℓ form an Eigenvalue Shifted Triple as claimed.

Now we deal with the following exceptional cases.

1. If $2a+c=0$, then up to a nonzero factor of a, we have $\frac{1}{a}\langle a,0,c\rangle = \langle 1,0,-2\rangle$ and are done by case 10 of Lemma 12.1.
2. If $a^2+2c^2=0$, then $a=\pm i\sqrt{2}c$. Up to a nonzero factor of c, we have $\frac{1}{c}\langle a,0,c\rangle = \langle \pm i\sqrt{2},0,1\rangle$ and are done by case 11 of Lemma 12.1.
3. If $a^2+2ac+3c^2=0$, then $a=c(-1\pm i\sqrt{2})$. Up to a nonzero factor of c, we have $\frac{1}{c}\langle a,0,c\rangle = \langle -1\pm i\sqrt{2},0,1\rangle$ and are done by case 12 of Lemma 12.1.
4. If $a^2+ac+7c^2=0$, then $2a=c(-1\pm 3i\sqrt{3})$. Up to a nonzero factor of $\frac{c}{2}$, we have $\frac{2}{c}\langle a,0,c\rangle = \langle -1\pm 3i\sqrt{3},0,2\rangle$ and are done by case 13 of Lemma 12.1. □

The next lemma considers the failure condition in (18). Since this failure condition is just a holographic transformation of the failure condition in (17), the excluded cases in this lemma are handled exactly as those preceding Lemma 9.11.

Lemma 9.12 *Suppose the domain size is* 3 *and* $b,c \in \mathbb{C}$. *Let* $\mathcal{F}$ *be a signature set containing the succinct ternary signature* $\langle -3b-4c,2b,2c\rangle$ *of type* τ_3 *and the succinct unary signature* $\langle 1\rangle$ *of type* τ_1. *Assume* $T^{\otimes 3}\langle -3b-4c,2b,2c\rangle = \langle \hat{a},\hat{b},\hat{c}\rangle$, *where* $T=\begin{bmatrix}1&-2&-2\\-2&1&-2\\-2&-2&1\end{bmatrix}$. *If* $\hat{a}\hat{c}\neq 0$ *and* $\hat{a}^3\neq\hat{c}^3$, *then*

$$\text{Pl-Holant}(\mathcal{F}\cup\{\langle x,y\rangle\}) \leq_T \text{Pl-Holant}(\mathcal{F})$$

for any $x,y\in\mathbb{C}$, *where* $\langle x,y\rangle$ *is a succinct binary signature of type* τ_2.

Proof By Lemma 11.6 with $x=1$ and $y=-2$, we have $\hat{b}=0$. Thus, after a holographic transformation by T, we are in the case covered by Lemma 9.11. Since T is orthogonal after scaling by $\frac{1}{3}$, the complexity of these problems is unchanged by Theorem 3.3. □

We summarize this section with the following lemma.

Corollary 9.13 *Suppose the domain size is* $\kappa\geq 3$ *and* $a,b,c\in\mathbb{C}$. *Let* $\mathcal{F}$ *be a signature set containing the succinct ternary signature* $\langle a,b,c\rangle$ *of type* τ_3 *and the succinct unary signature* $\langle 1\rangle$ *of type* τ_1. *Then*

$$\text{Pl-Holant}(\mathcal{F}\cup\{\langle x,y\rangle\}) \leq_T \text{Pl-Holant}(\mathcal{F})$$

for any $x,y\in\mathbb{C}$, *where* $\langle x,y\rangle$ *is a succinct binary signature of type* τ_2, *unless*

- $\mathfrak{B}=0$ *or*
- *there exist* $\lambda\in\mathbb{C}$ *and* $T\in\{I_\kappa,\kappa I_\kappa-2J_\kappa\}$ *such that*

$$\langle a,b,c\rangle = \begin{cases} T^{\otimes 3}\lambda\langle 1,0,0\rangle, \text{ or}\\ T^{\otimes 3}\lambda\langle 0,0,1\rangle \text{ and } \kappa=3, \text{ or}\\ T^{\otimes 3}\lambda\langle 1,0,\omega\rangle \text{ and } \kappa=3 \text{ where } \omega^3=1, \text{ or}\\ T^{\otimes 3}\lambda\langle \mu^2,1,\mu\rangle \text{ and } \kappa=4 \text{ where } \mu=-1\pm 2i. \end{cases}$$

Proof If failure condition (14), (15), (16), (17), or (18) holds, then we are done by Lemma 9.5, Lemma 9.6, Lemma 9.7, Lemma 9.11, or Lemma 9.12, respectively, with the various excluded cases listed. If none of (14), (15), (16), (17), and (18) hold, then we are done by Lemma 9.4. □

10 The main dichotomy

Now we can prove our main dichotomy theorem.

Theorem 10.1 *Suppose $\kappa \geq 3$ is the domain size and $a, b, c \in \mathbb{C}$. Let $\langle a, b, c\rangle$ be a succinct ternary signature of type τ_3. Then* Pl-Holant($\langle a, b, c\rangle$) *is* #P*-hard unless at least one of the following holds:*

1. $a = b = c$;
2. $a = c$ *and* $\kappa = 3$;

there exists $\lambda \in \mathbb{C}$ and $T \in \{I_\kappa, \kappa I_\kappa - 2J_\kappa\}$ such that

3. $\langle a, b, c\rangle = T^{\otimes 3}\lambda\langle 1, 0, 0\rangle$;
4. $\langle a, b, c\rangle = T^{\otimes 3}\lambda\langle 1, 0, \omega\rangle$ *and* $\kappa = 3$ *where* $\omega^3 = 1$;
5. $\langle a, b, c\rangle = T^{\otimes 3}\lambda\langle \mu^2, 1, \mu\rangle$ *and* $\kappa = 4$ *where* $\mu = -1 \pm 2i$;

in which case, the computation can be done in polynomial time.

Proof The signature in case 1 is degenerate, which is trivially tractable. Case 2 is tractable by case 3 of Corollary 5.2. Case 3 is tractable by Corollary 5.3. Case 4 is tractable by Corollary 5.6. Case 5 is tractable by Lemma 5.7.

Otherwise, $\langle a, b, c\rangle$ is none of these tractable cases. If $\mathfrak{B} = 0$, then we are done by Corollary 8.4, so assume that $\mathfrak{B} \neq 0$. If $a + (\kappa - 1)b = 0$ and $b^2 - 4bc - (\kappa - 3)c^2 = 0$, then we are done by Lemma 8.2, so assume that $a + (\kappa - 1)b \neq 0$ or $b^2 - 4bc - (\kappa - 3)c^2 \neq 0$.

If $a + (\kappa - 1)b \neq 0$, then we have the succinct unary signature $\langle 1\rangle$ of type τ_1 by Lemma 8.1. Otherwise, $a + (\kappa - 1)b = 0$ and $b^2 - 4bc - (\kappa - 3)c^2 \neq 0$. Since $\mathfrak{B} \neq 0$, we have $2b + (\kappa - 2)c \neq 0$. Then again we have $\langle 1\rangle$ by Lemma 8.1. Thus, in either case, we have $\langle 1\rangle$.

By Corollary 9.13, we have all binary succinct signatures $\langle x, y\rangle$ for any $x, y \in \mathbb{C}$. Then we are done by Lemma 7.16. □

Author details
[1]University of Wisconsin-Madison, Madison, WI, USA , [2]Queen Mary, University of London, London, UK.

Acknowledgements
We thank Joanna Ellis-Monaghan for bringing [30] to our attention. We are thankful to Mingji Xia who discussed with us an early version of this work. We are very grateful to Bjorn Poonen and especially Aaron Levin for sharing their expertise on Runge's method, in particular for the auxiliary function $g_2(x, y)$ in the proof of Lemma 7.6. We benefited from discussions with William Whistler on a draft of this work, whom we thank. We also thank the anonymous referees for their helpful comments. All authors were supported by NSF CCF-1217549. The second author was also supported by a Simons Award for Graduate Students in Theoretical Computer Science from the Simons Foundation. The third author was also supported by a Cisco Systems Distinguished Graduate Fellowship.

11 Appendix 1: Computing gadget signatures

In this paper, some of the more difficult claims to verify are those when we say that a particular $\mathcal{F}$-gate (or gadget) has a particular signature. This is an essential difficultly that cannot be avoided. We are proving that Pl-Holant($\mathcal{F}$) is #P-hard for various $\mathcal{F}$ (and computing the signature of an $\mathcal{F}$-gate is a generalization of this problem). Thus, one

should not expect to be able to compute these signatures significantly faster in general than what the naive algorithm can do.

This has always been an issue for any dichotomy theorem about counting problems, but with larger domain sizes, we seem to be reaching the limit of what can be computed by hand for the signatures of gadget constructions that are presented in our proofs. To counter this, the standard techniques are to utilize the smallest gadgets (that suffice) or an infinite family of related gadgets with a (small) description of finite size, which we certainly employ. Additionally, we point out some tricks, where they exist, to save as much work as possible.

Beyond all this, we also face another problem. We would like to express the signature of a gadget as a function of the domain size. To compute the signature of a gadget for every domain size is no longer a finite computation. However, each entry of the gadget's signature is a polynomial in the domain size of degree at most the number of internal edges in the gadget. To obtain these polynomials, one can interpolate them by computing the signature for small domain sizes. It is easy to write a program to do this.

When computing by hand, there is another possibility that works quite well. One partitions the internal edge assignments into a limited number of parts such that the assignments in each part contribute the same quantity to the Holant sum. This is best explained with some examples.

Lemma 11.1 *Suppose* $\kappa \geq 3$ *is the domain size and* $a, b, c, x, y \in \mathbb{C}$. *Let* $\langle a, b, c\rangle$ *be a succinct ternary signature of type* τ_3 *and let* $\langle x, y\rangle$ *be a succinct signature of type* τ_2. *If we assign* $\langle a, b, c\rangle$ *to the circle vertex and* $\langle x, y\rangle$ *to the square vertex of the gadget in Fig. 21c, then the succinct unary signature of type* τ_1 *of the resulting gadget is* $\langle x[a + (\kappa - 1)b] + y(\kappa - 1)[2b + (\kappa - 2)c]\rangle$.

If the square vertex is replaced by Fig. 21d, then the resulting signature is $\langle a + (\kappa - 1)b\rangle$. *If the square vertex is replaced by Fig. 21e, and* $a + (\kappa - 1)b = 0$, *then the resulting signature is*

$$\langle -(\kappa - 1)(\kappa - 2)[2b + (\kappa - 2)c][b^2 - 4bc - (\kappa - 3)c^2]\rangle. \tag{20}$$

Proof Since $\langle a, b, c\rangle$ and $\langle x, y\rangle$ are domain invariant, the signatures of these gadgets are also domain invariant. Any domain invariant unary signature has a succinct signature of type τ_1.

Let $g \in [\kappa]$ be a possible edge assignment, which we call a color. Suppose the external edge is assigned g and consider all internal edge assignments that assign the same colors to both edges. For such assignments, $\langle x, y\rangle$ contributes a factor of x. Now if this color assigned to both internal edges is also g, then $\langle a, b, c\rangle$ contributes a factor of a. Thus, the

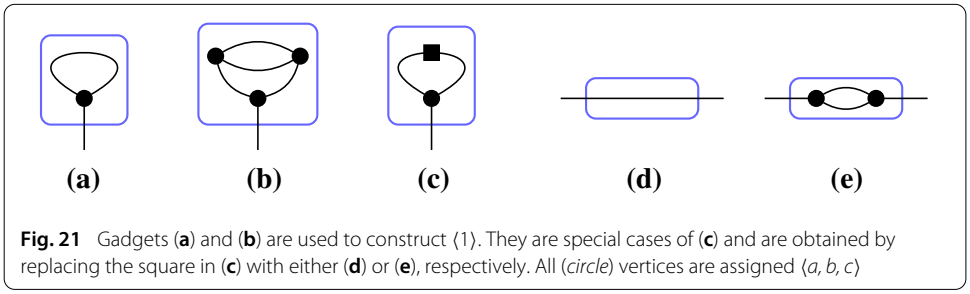


Fig. 21 Gadgets (**a**) and (**b**) are used to construct ⟨1⟩. They are special cases of (**c**) and are obtained by replacing the square in (**c**) with either (**d**) or (**e**), respectively. All (*circle*) vertices are assigned $\langle a, b, c\rangle$

Holant sum includes one factor of ax. If the two internal edges are assigned any color different from g, then $\langle a, b, c\rangle$ contributes a factor of b. Since there are $\kappa - 1$ such colors, this adds $(\kappa - 1)bx$ to the Holant sum.

Now consider all internal assignments that assign different colors to the edges. For such assignments, $\langle x, y\rangle$ contributes a factor of y. First, suppose that one of the internal edges is assigned g. There are two ways this could happen and $\langle a, b, c\rangle$ contributes a factor of b. Since there are $\kappa - 1$ choices for the remaining edge assignment, this adds $2(\kappa - 1)by$ to the Holant sum. Lastly, suppose that the two internal edges are not assigned g. Then $\langle a, b, c\rangle$ contributes a factor of c. Since there are $(\kappa - 1)(\kappa - 2)$ such assignments, this adds $(\kappa - 1)(\kappa - 2)cy$ to the Holant sum. Thus, the resulting signature is $\langle x[a + (\kappa - 1)b] + y(\kappa - 1)[2b + (\kappa - 2)c]\rangle$ as claimed.

Replacing the square by Fig. 21d is equivalent to setting $x = 1$ and $y = 0$, which gives $\langle a + (\kappa - 1)b\rangle$. Replacing the square by Fig. 21e is equivalent to setting x and y to the values given in Lemma 11.2. The resulting signature is indeed (20). □

Lemma 11.2 *Suppose $\kappa \geq 3$ is the domain size and $a, b, c \in \mathbb{C}$. Let $\langle a, b, c\rangle$ be a succinct ternary signature of type τ_3. If we assign $\langle a, b, c\rangle$ to both vertices of the gadget in Fig. 22, then the succinct binary signature of type τ_2 of the resulting gadget is $\langle x, y\rangle$, where*

$$x = a^2 + 3(\kappa - 1)b^2 + (\kappa - 1)(\kappa - 2)c^2 \qquad \text{and}$$
$$y = 2ab + \kappa b^2 + 4(\kappa - 2)bc + (\kappa - 2)(\kappa - 3)c^2.$$

Proof Since $\langle a, b, c\rangle$ is domain invariant, the signature of this gadget is also domain invariant. Any domain invariant binary signature has a succinct signature of type τ_2.

Let $g, r \in [\kappa]$ be distinct edge assignments. We have two entries to compute. To compute x, suppose that both external edges are assigned g. We begin with the case where both internal edges have the same assignment. If this assignment is g, then a^2 is contributed to the sum. If this assignment is not g, then b^2 is contributed to the sum for a total contribution of $(\kappa - 1)b^2$. Now consider the case that the two internal edges have a different assignment. If one of these assignments is g, then b^2 is contributed to the sum for a total contribution of $2(\kappa - 1)b^2$. If neither assignment is g, then c^2 is contributed to the sum for a total contribution of $(\kappa - 1)(\kappa - 2)c^2$. These total contributions sum to the value for x given in Lemma 11.2.

To compute y, suppose one external edge is assigned g and the other is assigned r. We begin with the case where both internal edges have the same assignment. If this assignment is g or r, then ab is contributed to the sum for a total contribution of $2ab$. If this assignment is not g or r, then b^2 is contributed to the sum for a total contribution of $(\kappa - 2)b^2$. Now consider the case that the two internal edges have a different assignment. If both are assigned g or r, then b^2 is contributed to the sum for a total contribution of $2b^2$. If exactly one is assigned g or r, then bc is contributed to the sum for a total contribution of $4(\kappa - 2)bc$. If neither is assigned g or r, then c^3 is contributed to the sum for a total

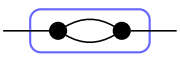


Fig. 22 A simple binary gadget

contribution of $(\kappa-2)(\kappa-3)c^3$. These total contributions sum to the value for y given in Lemma 11.2. □

When checking these proofs, a concern is that some assignments might not have been counted. One sanity check to address this concern is to set $a=b=c=1$ and inspect the resulting expression. If computed correctly, the result will be κ^m, where m is the number of internal edges, which is the number of internal edge assignments.

Lemma 11.3 *Suppose $\kappa \geq 3$ is the domain size and $a, b, c \in \mathbb{C}$. Let $\langle a, b, c\rangle$ be a succinct ternary signature of type τ_3. If we assign $\langle a, b, c\rangle$ to both vertices of the gadget in Fig. 23, then the succinct quaternary signature of type τ_4 of the resulting gadget is*

$$f = \left\langle f_{\begin{smallmatrix}1&1\\1&1\end{smallmatrix}}, f_{\begin{smallmatrix}1&2\\1&1\end{smallmatrix}}, f_{\begin{smallmatrix}1&2\\1&2\end{smallmatrix}}, f_{\begin{smallmatrix}1&3\\1&2\end{smallmatrix}}, f_{\begin{smallmatrix}1&2\\2&1\end{smallmatrix}}, f_{\begin{smallmatrix}1&3\\2&1\end{smallmatrix}}, f_{\begin{smallmatrix}1&1\\2&2\end{smallmatrix}}, f_{\begin{smallmatrix}1&1\\2&3\end{smallmatrix}}, f_{\begin{smallmatrix}1&4\\2&3\end{smallmatrix}} \right\rangle,$$

where

$$\begin{aligned}
f_{\begin{smallmatrix}1&1\\1&1\end{smallmatrix}} &= a^2 + (\kappa-1)b^2,\\
f_{\begin{smallmatrix}1&2\\1&1\end{smallmatrix}} &= b[a + b + (\kappa-2)c],\\
f_{\begin{smallmatrix}1&2\\1&2\end{smallmatrix}} &= 2b^2 + (\kappa-2)c^2,\\
f_{\begin{smallmatrix}1&3\\1&2\end{smallmatrix}} &= b^2 + 2bc + (\kappa-3)c^2,\\
f_{\begin{smallmatrix}1&2\\2&1\end{smallmatrix}} &= f_{\begin{smallmatrix}1&2\\1&2\end{smallmatrix}},\\
f_{\begin{smallmatrix}1&3\\2&1\end{smallmatrix}} &= f_{\begin{smallmatrix}1&3\\1&2\end{smallmatrix}},\\
f_{\begin{smallmatrix}1&1\\2&2\end{smallmatrix}} &= b[2a + (\kappa-2)b],\\
f_{\begin{smallmatrix}1&1\\2&3\end{smallmatrix}} &= ac + 2b^2 + (\kappa-3)bc, \text{ and}\\
f_{\begin{smallmatrix}1&4\\2&3\end{smallmatrix}} &= c[4b + (\kappa-4)c].
\end{aligned}$$

Proof Since $\langle a, b, c\rangle$ is domain invariant, the signature of this gadget is also domain invariant. The vertical and horizontal symmetry of this gadget implies that its signature has a succinct signature of type τ_4.

Let $w, x, y, z \in [\kappa]$ be distinct edge assignments. We have nine entries to compute. Recall that the edge with the diamond is considered the first input and the rest are ordered counterclockwise.

1. To compute $f_{\begin{smallmatrix}1&1\\1&1\end{smallmatrix}}$, suppose the external assignment is (w, w, w, w). If the internal edge is also assigned w, then a^2 is contributed to the sum. If the internal edge is not assigned w, then b^2 is contributed to the sum for a total contribution of $(\kappa-1)b^2$.
2. To compute $f_{\begin{smallmatrix}1&2\\1&1\end{smallmatrix}}$, suppose the external assignment is (w, w, w, x). If the internal edge is assigned w, then ab is contributed to the sum. If the internal edge is assigned x, then b^2 is contributed to the sum. If the internal edge is not assigned w or x, then bc is contributed to the sum for a total contribution of $(\kappa-2)bc$.

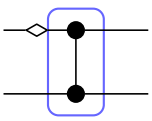


Fig. 23 A simple quaternary gadget

3. To compute $f_{\begin{smallmatrix}1&2\\1&2\end{smallmatrix}}$, suppose the external assignment is (w, w, x, x). If the internal edge is assigned w, then b^2 is contributed to the sum. If the internal edge is assigned x, then b^2 is contributed to the sum. If the internal edge is not assigned w or x, then c^2 is contributed to the sum for a total contribution of $(\kappa - 2)c^2$.
4. To compute $f_{\begin{smallmatrix}1&3\\1&2\end{smallmatrix}}$, suppose the external assignment is (w, w, x, y). If the internal edge is assigned w, then b^2 is contributed to the sum. If the internal edge is assigned x, then bc is contributed to the sum. If the internal edge is assigned y, then bc is contributed to the sum. If the internal edge is not assigned w, x or y, then c^2 is contributed to the sum for a total contribution of $(\kappa - 3)c^2$.
5. To compute $f_{\begin{smallmatrix}1&3\\1&2\end{smallmatrix}}$, suppose the external assignment is (w, x, w, x). This entry is the same as that for (w, w, x, x). The reason is that the signature is unchanged if the two external edges of the lower vertex are swapped since $\langle a, b, c\rangle$ is symmetric.
6. To compute $f_{\begin{smallmatrix}1&3\\2&1\end{smallmatrix}}$, suppose the external assignment is (w, x, w, y). This entry is the same as that for (w, w, x, y) for the same reason as the previous entry.
7. To compute $f_{\begin{smallmatrix}1&1\\2&2\end{smallmatrix}}$, suppose the external assignment is (w, x, x, w). If the internal edge is assigned w, then ab is contributed to the sum. If the internal edge is assigned x, then ab is contributed to the sum. If the internal edge is not assigned w or x, then b^2 is contributed to the sum for a total contribution of $(\kappa - 2)b^2$.
8. To compute $f_{\begin{smallmatrix}1&1\\2&3\end{smallmatrix}}$, suppose the external assignment is (w, x, y, w). If the internal edge is assigned w, then ac is contributed to the sum. If the internal edge is assigned x, then b^2 is contributed to the sum. If the internal edge is assigned y, then b^2 is contributed to the sum. If the internal edge is not assigned w, x or y, then bc is contributed to the sum for a total contribution of $(\kappa - 3)c^2$.
9. To compute $f_{\begin{smallmatrix}1&4\\2&3\end{smallmatrix}}$, suppose the external assignment is (w, x, y, z). If the internal edge is assigned w, x, y, or z, then bc is contributed to the sum for a total contribution of $4bc$. If the internal edge is not assigned w, x, y or z, then c^2 is contributed to the sum for a total contribution of $(\kappa - 4)c^2$.

These total contributions each sum to their corresponding entry of f given in the statement of Lemma 11.3. □

Although possible, it would be difficult to compute the signature of the gadget in Fig. 24c through partitioning of the internal edge assignments alone. To simplify matters, we utilize the calculations from Lemma 11.3. Since composing the gadget in Fig. 24a with the one in Fig. 24b gives a symmetric signature, we refrain from distinguishing the external edges of the gadget in Fig. 24b.

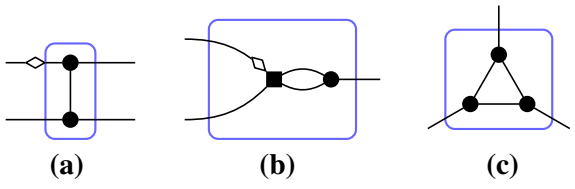


Fig. 24 Decomposition of a ternary gadget. All *circle* vertices are assigned $\langle a, b, c\rangle$, and the *square* vertex in (**b**) is assigned the signature of the gadget in (**a**). **a** Inner structure **b** outer structure **c** entire binary gadget

Lemma 11.4 *Suppose $\kappa \geq 3$ is the domain size and $a, b, c \in \mathbb{C}$. Let $\langle a, b, c\rangle$ be a succinct ternary signature of type τ_3. If we assign $\langle a, b, c\rangle$ to all vertices of the gadget in Fig. 24c, then the succinct ternary signature of type τ_3 of the resulting gadget is $\langle a', b', c'\rangle$, where*

$$\begin{aligned}
a' &= a^3 + 3(\kappa-1)ab^2 + 4(\kappa-1)b^3 + 3(\kappa-1)(\kappa-2)(b^2c + bc^2)\\
&\quad + (\kappa-1)(\kappa-2)(\kappa-3)c^3,\\
b' &= a^2b + 4ab^2 + 2(\kappa-2)abc + (\kappa-2)ac^2 + (5\kappa-7)b^3 + (\kappa-2)(\kappa+5)b^2c\\
&\quad + (\kappa-2)(7\kappa-18)bc^2 + (\kappa-2)(\kappa-3)^2c^3, \qquad \textit{and}\\
c' &= 3ab^2 + 6abc + 3(\kappa-3)ac^2 + (\kappa+5)b^3 + 3(7\kappa-18)b^2c + 9(\kappa-3)^2bc^2\\
&\quad + (\kappa^3 - 9\kappa^2 + 29\kappa - 32)c^3.
\end{aligned}$$

Furthermore, if $\mathfrak{A} = 0$, then

$$\begin{aligned}
a' &= 3b' - 2c',\\
b' &= (5\kappa+14)b^3 + (\kappa^2+9\kappa-42)b^2c + (7\kappa^2-33\kappa+42)bc^2\\
&\quad + (\kappa-2)(\kappa^2-6\kappa+7)c^3, \qquad \textit{and}\\
c' &= (\kappa+14)b^3 + 21(\kappa-2)b^2c + 3(3\kappa^2-15\kappa+14)bc^2 + (\kappa^3-9\kappa^2+23\kappa-14)c^3.
\end{aligned}$$

Proof Since $\langle a, b, c\rangle$ is domain invariant, the signature of this gadget is also domain invariant. As a ternary signature, the rotational symmetry of this gadget implies the symmetry of the signature. Any symmetric domain invariant ternary signature has a succinct signature of type τ_3.

Consider the gadget in Fig. 24a. We assign $\langle a, b, c\rangle$ to both vertices. Then by Lemma 11.3, the succinct quaternary signature of this gadget is the signature f given in Lemma 11.3.

Now consider the gadget in Fig. 24b. We assign $\langle a, b, c\rangle$ to the circle vertex and f to the square vertex. The resulting gadget is the one in Fig. 24c, which is symmetric. Thus, there is no need to distinguish the external edges. We have three entries to compute.

Let $g, r, y \in [\kappa]$ be distinct edge assignments. To compute a', suppose that all external edges are assigned g. We begin with the case where both internal edges have the same assignment. If this assignment is g, then $af_{\begin{smallmatrix}1&1\\1&1\end{smallmatrix}}$ is contributed to the sum. If this assignment is not g, then $bf_{\begin{smallmatrix}1&2\\1&2\end{smallmatrix}}$ is contributed to the sum for a total contribution of $(\kappa-1)bf_{\begin{smallmatrix}1&2\\1&2\end{smallmatrix}}$. Now consider the case that the two internal edges have a different assignment. If one of these assignments is g, then $bf_{\begin{smallmatrix}1&2\\1&1\end{smallmatrix}}$ is contributed to the sum for a total contribution of $2(\kappa-1)bf_{\begin{smallmatrix}1&2\\1&1\end{smallmatrix}}$. If neither assignment is g, then $cf_{\begin{smallmatrix}1&3\\1&2\end{smallmatrix}}$ is contributed to the sum for a total contribution of $(\kappa-1)(\kappa-2)cf_{\begin{smallmatrix}1&3\\1&2\end{smallmatrix}}$. After substituting for the entries of f, these total contributions sum to the value for a' given in Lemma 11.4.

To compute b', suppose the left external edges are assigned g and the right external edge is assigned r. We begin with the case where both internal edges have the same assignment. If this assignment is g, then $bf_{\begin{smallmatrix}1&1\\1&1\end{smallmatrix}}$ is contributed to the sum. If this assignment is r, then $af_{\begin{smallmatrix}1&2\\1&2\end{smallmatrix}}$ is contributed to the sum. If this assignment is not g or r, then $bf_{\begin{smallmatrix}1&2\\1&2\end{smallmatrix}}$ is contributed to the sum for a total contribution of $(\kappa-2)bf_{\begin{smallmatrix}1&2\\1&2\end{smallmatrix}}$. Now consider the case that the two internal edges have a different assignments. If both are assigned g or r, then $bf_{\begin{smallmatrix}1&2\\1&1\end{smallmatrix}}$ is contributed to the sum for a total contribution of $2bf_{\begin{smallmatrix}1&2\\1&1\end{smallmatrix}}$. If one is assigned g and the other is not assigned r, then $cf_{\begin{smallmatrix}1&2\\1&1\end{smallmatrix}}$ is contributed to the sum for a total contribution of $2(\kappa-2)cf_{\begin{smallmatrix}1&2\\1&1\end{smallmatrix}}$. If one is assigned r and the other is not assigned g, then $bf_{\begin{smallmatrix}1&3\\1&2\end{smallmatrix}}$ is contributed to the sum for a total

contribution of $2(\kappa-2)bf_{\begin{smallmatrix}1&3\\1&2\end{smallmatrix}}$. If neither is assigned g or r, then $cf_{\begin{smallmatrix}1&3\\1&2\end{smallmatrix}}$ is contributed to the sum for a total contribution of $(\kappa-2)(\kappa-3)cf_{\begin{smallmatrix}1&3\\1&2\end{smallmatrix}}$. After substituting for the entries of f, these total contributions sum to the value for b' given in Lemma 11.4.

To compute c', suppose the upper-left external edge is assigned g, the lower-left external edge is assigned r, and the right external edge is assigned y. We begin with the case where both internal edges have the same assignment. If this assignment is g, then $bf_{\begin{smallmatrix}1&2\\1&1\end{smallmatrix}}$ is contributed to the sum. If this assignment is r, then $bf_{\begin{smallmatrix}1&2\\1&1\end{smallmatrix}}$ is contributed to the sum. If this assignment is y, then $af_{\begin{smallmatrix}1&3\\1&2\end{smallmatrix}}$ is contributed to the sum. If this assignment is not g, r, or y, then $bf_{\begin{smallmatrix}1&3\\1&2\end{smallmatrix}}$ is contributed to the sum for a total contribution of $(\kappa-3)bf_{\begin{smallmatrix}1&3\\1&2\end{smallmatrix}}$. Now consider the case that the two internal edges have a different assignments. If the top internal edge is assigned g and the bottom one is assigned r, then $cf_{\begin{smallmatrix}1&1\\2&2\end{smallmatrix}}$ is contributed to the sum. If the top internal edge is assigned r and the bottom one is assigned g, then $cf_{\begin{smallmatrix}1&2\\2&1\end{smallmatrix}}$ is contributed to the sum. If the top internal edge is assigned g and the bottom one is assigned y, then $bf_{\begin{smallmatrix}1&1\\2&3\end{smallmatrix}}$ is contributed to the sum. If the top internal edge is assigned y and the bottom one is assigned g, then $bf_{\begin{smallmatrix}1&3\\2&1\end{smallmatrix}}$ is contributed to the sum. If the top internal edge is assigned r and the bottom one is assigned y, then $bf_{\begin{smallmatrix}1&3\\2&1\end{smallmatrix}}$ is contributed to the sum. If the top internal edge is assigned y and the bottom one is assigned r, then $bf_{\begin{smallmatrix}1&1\\2&3\end{smallmatrix}}$ is contributed to the sum. If the top internal edge is assigned g and the bottom one not assigned r or y, then $cf_{\begin{smallmatrix}1&1\\2&3\end{smallmatrix}}$ is contributed to the sum for a total contribution of $(\kappa-3)cf_{\begin{smallmatrix}1&1\\2&3\end{smallmatrix}}$. If the bottom internal edge is assigned g and the top one not assigned r or y, then $cf_{\begin{smallmatrix}1&3\\2&1\end{smallmatrix}}$ is contributed to the sum for a total contribution of $(\kappa-3)cf_{\begin{smallmatrix}1&3\\2&1\end{smallmatrix}}$. If the top internal edge is assigned r and the bottom one not assigned g or y, then $cf_{\begin{smallmatrix}1&3\\2&1\end{smallmatrix}}$ is contributed to the sum for a total contribution of $(\kappa-3)cf_{\begin{smallmatrix}1&3\\2&1\end{smallmatrix}}$. If the bottom internal edge is assigned r and the top one not assigned g or y, then $cf_{\begin{smallmatrix}1&1\\2&3\end{smallmatrix}}$ is contributed to the sum for a total contribution of $(\kappa-3)cf_{\begin{smallmatrix}1&1\\2&3\end{smallmatrix}}$. If the one internal edge is assigned y and the other is not assigned g or r, then $bf_{\begin{smallmatrix}1&4\\2&3\end{smallmatrix}}$ is contributed to the sum for a total contribution of $2(\kappa-3)bf_{\begin{smallmatrix}1&4\\2&3\end{smallmatrix}}$. If neither internal edge is assigned g r, or y, then $cf_{\begin{smallmatrix}1&4\\2&3\end{smallmatrix}}$ is contributed to the sum for a total contribution of $(\kappa-3)(\kappa-4)cf_{\begin{smallmatrix}1&4\\2&3\end{smallmatrix}}$. After substituting for the entries of f, these total contributions sum to the value for c' given in Lemma 11.4. □

The signature of the gadget in Fig. 25 is difficult to compute using gadget compositions and partitioning of internal edge assignments as we have been doing. Instead, we compute this signature using matrix product, trace, and polynomial interpolation.

Lemma 11.5 *Suppose $\kappa \geq 3$ is the domain size and $a, b, c, x_1, y_1, x_2, y_2 \in \mathbb{C}$. Let $\langle a, b, c\rangle$ be a succinct ternary signature of type τ_3 and $\langle x_1, y_1\rangle$ and $\langle x_2, y_2\rangle$ be succinct binary signatures of type τ_2. If to the gadget in Fig. 25 we assign $\langle a, b, c\rangle$ to the circle vertices, $\langle x_1, y_1\rangle$ to the square vertex, and $\langle x_2, y_2\rangle$ to the triangle vertex, then the succinct binary signature of type τ_2 of the resulting gadget is $\langle x, y\rangle$, where*

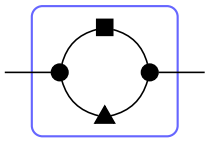


Fig. 25 A more complicated binary gadget

$$
\begin{aligned}
x = {} & x_1x_2a^2 + 2(\kappa-1)(x_1y_2 + x_2y_1 + y_1y_2)ab + 2(\kappa-1)(\kappa-2)y_1y_2ac \\
& + (\kappa-1)[3x_1x_2 + \kappa(x_1y_2 + x_2y_1) + (7\kappa-12)y_1y_2]b^2 \\
& + 2(\kappa-1)(\kappa-2)[2(x_1y_2 + x_2y_1) + (3\kappa-7)y_1y_2]bc \\
& + (\kappa-1)(\kappa-2)[x_1x_2 + (\kappa-3)(x_1y_2 + x_2y_1) + (\kappa^2-5\kappa+7)y_1y_2]c^2 \qquad \textit{and} \\
y = {} & y_1y_2a^2 + 2[x_1x_2 + x_1y_2 + x_2y_1 + 3(\kappa-2)y_1y_2]ab \\
& + 2(\kappa-2)[x_1y_2 + x_2y_1 + (\kappa-3)y_1y_2]ac \\
& + [\kappa x_1x_2 + (7\kappa-12)(x_1y_2 + x_2y_1) + 3(3\kappa^2-11\kappa+11)y_1y_2]b^2 \\
& + 2(\kappa-2)[2x_1x_2 + (3\kappa-7)(x_1y_2 + x_2y_1) + 3(\kappa^2-4\kappa+5)y_1y_2]bc \\
& + (\kappa-2)[(\kappa-3)x_1x_2 + (\kappa^2-5\kappa+7)(x_1y_2 + x_2y_1) + (\kappa^3-6\kappa^2+14\kappa-13)]c^2.
\end{aligned}
$$

Furthermore, if $\langle x_1, y_1\rangle = \frac{1}{\kappa}\langle \omega^r + \kappa - 1, \omega^r - 1\rangle$ *and* $\langle x_2, y_2\rangle = \frac{1}{\kappa}\langle \omega^s + \kappa - 1, \omega^s - 1\rangle$, *then*

$$
\begin{aligned}
x &= \frac{\mathfrak{B}^2}{\kappa^2}\left[\Phi\omega^{r+s} + (\kappa-1)(\omega^r + \omega^s + \Psi + 1)\right] \quad \textit{and} \\
y &= \frac{\mathfrak{B}^2}{\kappa^2}\left[\Phi\omega^{r+s} - (\omega^r + \omega^s + \Psi + 1) + \kappa\right],
\end{aligned}
$$

where $\Phi = \frac{\mathfrak{C}^2}{\mathfrak{B}^2}$ *and* $\Psi = \frac{(\kappa-2)\mathfrak{A}^2}{\mathfrak{B}^2}$.

Proof Since $\langle a, b, c\rangle$, $\langle x_1, y_1\rangle$, and $\langle x_2, y_2\rangle$ are domain invariant, the signature of this gadget is also domain invariant. Any domain invariant binary signature has a succinct signature of type τ_2.

We compute a', b', and c' using the algorithm for Holant($\mathcal{F}$) when every nondegenerate signature in $\mathcal{F}$ is of arity at most 2, which is to use matrix product and trace. Then we finish with polynomial interpolation. Let $M_\kappa(t)$ be a κ-by-κ matrix such that

$$
(M_\kappa(t))_{i,j} = \begin{cases} a & i = j = t, \\ b & i = j \neq t, \\ b & i \neq j \text{ and } (i = t \text{ or } j = t), \\ c & \text{otherwise.} \end{cases}
$$

For example, $M_4(1) = \begin{bmatrix} a & b & b & b \\ b & b & c & c \\ b & c & b & c \\ b & c & c & b \end{bmatrix}$. If we fix an input of $\langle a, b, c\rangle$ to $t \in [\kappa]$, then the resulting binary signature (which is no longer domain invariant) has the signature matrix $M_\kappa(t)$.

Consider x and y as polynomials in κ with coefficients in $\mathbb{Z}[a, b, c, x_1, y_1, x_2, y_2]$. Then

$$
\begin{aligned}
x(\kappa) &= \operatorname{tr}\left(M_\kappa(1)[y_1J_\kappa + (x_1 - y_1)I_\kappa]M_\kappa(1)[y_2J_\kappa + (x_2 - y_2)I_\kappa]\right) \qquad \text{and} \\
y(\kappa) &= \operatorname{tr}\left(M_\kappa(1)[y_1J_\kappa + (x_1 - y_1)I_\kappa]M_\kappa(2)[y_2J_\kappa + (x_2 - y_2)I_\kappa]\right).
\end{aligned}
$$

Since there are just four internal edges in this gadget, both of $x(\kappa)$ and $y(\kappa)$ are of degree at most 4 in κ. Therefore, we interpolate each of these polynomials using their evaluations at $3 \le \kappa \le 7$ and obtain the expressions for x and y given in Lemma 11.5. □

Remark Lemma 11.2 is the special case of Lemma 11.5 with $\langle x_1, y_1\rangle = \langle x_2, y_2\rangle = \langle 1, 0\rangle$.

In order to apply a holographic transformation on a particular signature, it is convenient to express the signature as a sum of degenerate signatures. Let $e_{\kappa,i}$ be the standard basis vector of length κ with a 1 at location i and 0 elsewhere. Also let $\mathbf{1}_\kappa$ be the all 1's vector of length κ. Then the succinct ternary signature $\langle a, b, c\rangle$ on domain size κ can be expressed as

$$\langle a,b,c\rangle = c\mathbf{1}_\kappa^{\otimes 3} + (a-c)\sum_{i=1}^{\kappa} e_{\kappa,i}^{\otimes 3} + (b-c)\sum_{\substack{i,j\in[\kappa]\\ i\neq j}} \begin{pmatrix} e_{\kappa,i}\otimes e_{\kappa,i}\otimes e_{\kappa,j} \\ + e_{\kappa,i}\otimes e_{\kappa,j}\otimes e_{\kappa,i} \\ + e_{\kappa,j}\otimes e_{\kappa,i}\otimes e_{\kappa,i}\end{pmatrix} \tag{21}$$

$$= b\mathbf{1}_\kappa^{\otimes 3} + (a-b)\sum_{i=1}^{\kappa} e_{\kappa,i}^{\otimes 3} + (c-b)\sum_{\substack{\sigma:1,2,3\to[\kappa]\\ \sigma \text{ injective}}} e_{\kappa,\sigma(1)}\otimes e_{\kappa,\sigma(2)}\otimes e_{\kappa,\sigma(3)}. \tag{22}$$

The expression in (21) contains $1+\kappa+3\kappa(\kappa-1) = 3\kappa^2-2\kappa+1$ summands. In general, this is smaller than the one in (22), which contains $1+\kappa+\kappa(\kappa-1)(\kappa-2) = \kappa^3-3\kappa^2+3\kappa+1$ summands. It is advantageous to find an expression that minimizes the number of summands. This leads to less computation in the proof of Lemma 11.6. However, determining the fewest number of summands for a given signature is exactly the problem of determining tensor rank, which is a problem well known to be difficult [38].

There is a gadget construction that mimics the behavior of a holographic transformation. This construction is called a local holographic transformation [24]. For $x, y \in \mathbb{C}$, let $\langle x, y\rangle$ be a succinct binary signature of type τ_2. Consider the gadget in Fig. 26. If we assign $\langle a, b, c\rangle$ to the circle vertex and $\langle x, y\rangle$ to the square vertex, then the resulting signature of this gadget is the same as applying a holographic transformation on $\langle a, b, c\rangle$ with basis $T = yJ_\kappa + (x-y)I_\kappa$. We use this fact in the following proof.

Lemma 11.6 *Suppose $\kappa \geq 3$ is the domain size and $a, b, c, x, y \in \mathbb{C}$. Let $\langle a, b, c\rangle$ be a succinct signature of type τ_3 and let $T = yJ_\kappa + (x-y)I_\kappa$. Then $T^{\otimes 3}\langle a, b, c\rangle = \langle a', b', c'\rangle$, where*

$$\begin{aligned} a' &= a\left[x^3+(\kappa-1)y^3\right] \\ &\quad + 3b(\kappa-1)\left[x^2y+xy^2+(\kappa-2)y^3\right] \\ &\quad + c(\kappa-1)(\kappa-2)\left[3xy^2+(\kappa-3)y^3\right], \\ b' &= a\left[x^2y+xy^2+(\kappa-2)y^3\right] \\ &\quad + b\left[x^3+\kappa x^2y+(7\kappa-12)xy^2+(3\kappa^2-11\kappa+11)y^3\right] \\ &\quad + c(\kappa-2)\left[2x^2y+(3\kappa-7)xy^2+(\kappa^2-4\kappa+5)y^3\right], \text{ and} \end{aligned}$$

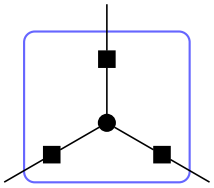


Fig. 26 Local holographic transformation gadget construction for a ternary signature

$$\begin{aligned} c' = {} & a\left[3xy^2 + (\kappa-3)y^3\right] \\ & + 3b\left[2x^2y + (3\kappa-7)xy^2 + (\kappa^2-4\kappa+5)y^3\right] \\ & + c\left[x^3 + 3(\kappa-3)x^2y + 3(\kappa^2-5\kappa+7)xy^2 + (\kappa^3-6\kappa^2+14\kappa-13)y^3\right]. \end{aligned}$$

In particular,

$$a' - b' = (x-y)^2[2\mathfrak{D} + \mathfrak{A}(x-y)] \qquad \textit{and} \qquad b' - c' = (x-y)^2\mathfrak{D},$$

where $\mathfrak{D} = (b-c)(x-y) + \mathfrak{B}y$. *Furthermore, if* $\mathfrak{A} = 0$, *then*

$$\begin{aligned} a' &= 3b' - 2c', \\ b' &= [x + (\kappa-1)y]\left\{bx^2 + 2[2b + (\kappa-3)c]xy + [(3\kappa-5)b + (\kappa^2-5\kappa+6)c]y^2\right\} \quad \textit{and} \\ c' &= [x + (\kappa-1)y]\left\{cx^2 + 2[3b + (\kappa-4)c]xy + [(3\kappa-6)b + (\kappa^2-5\kappa+7)c]y^2\right\}. \end{aligned}$$

If $\kappa = 3$, $x = -1$, *and* $y = 2$, *then*

$$a' = -3(5a + 12b - 8c), \qquad b' = -3(2a + 3b + 4c), \qquad \textit{and} \qquad c' = 3(4a - 12b - c).$$

Proof Let $\widehat{f} = T^{\otimes 3}\langle a, b, c\rangle$. Since $\langle a, b, c\rangle$ and $\langle x, y\rangle$ are domain invariant, the signature of the gadget in Fig. 26, which is the same signature $\widehat{f}$, is also domain invariant. As a ternary signature, the rotational symmetry of this gadget implies the symmetry of the signature. Any symmetric domain invariant ternary signature has a succinct signature of type τ_3.

The entries of $\widehat{f}$ are polynomials in κ with coefficients from $\mathbb{Z}[a, b, c, x, y]$. The degree of these polynomials is at most 3 since the arity of $\langle a, b, c\rangle$ is 3. We compute the entries of $\widehat{f} = T^{\otimes 3}\langle a, b, c\rangle$ as elements in $\mathbb{Z}[a, b, c, x, y]$ for domain sizes $3 \le \kappa \le 6$ by replacing $\langle a, b, c\rangle$ with an equivalent expression from either (21) or (22). Then we interpolate the entries of $\widehat{f}$ as elements in $(\mathbb{Z}[a, b, c, x, y])[\kappa]$. The resulting expressions for the signature entries are as given in the statement of Lemma 11.6.

It is straightforward to verify the expressions for $a' - b'$ and $b' - c'$ given those for a', b', and c'. Recall that $\mathfrak{A} = a - 3b + 2c$. If $\mathfrak{A} = 0$, then it follows that $a' - 3b' + 2c' = 0$ as well since

$$\begin{aligned} a' - 3b' + 2c' &= a' - b' - 2(b' - c') \\ &= (x-y)^2[2\mathfrak{D} + \mathfrak{A}(x-y)] - 2(x-y)^2\mathfrak{D} \\ &= \mathfrak{A}(x-y)^3 = 0. \end{aligned}$$

The expressions for b' and c' when $\mathfrak{A} = 0$ directly follow from their general expressions above. □

By composing smaller gadgets, we can easily compute the signatures of rather large gadgets.

Lemma 11.7 *Suppose* $\kappa \ge 3$ *is the domain size and* $a, b, c \in \mathbb{C}$. *Let* $\langle a, b, c\rangle$ *be a succinct ternary signature of type* τ_3. *If* $\langle a, b, c\rangle$ *is assigned to every vertex of the gadget in Fig. 27c, then the resulting signature is the succinct binary signature* $\langle x, y\rangle$ *of type* τ_2, *where*

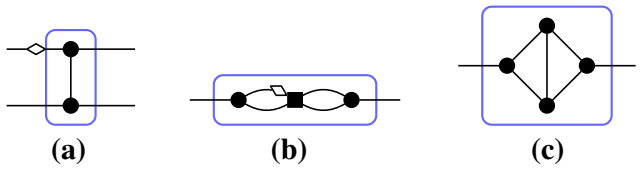


Fig. 27 Decomposition of a binary gadget. All *circle* vertices are assigned $\langle a, b, c\rangle$, and the square vertex in (**b**) is assigned the signature of the gadget in (**a**). **a** Inner structure **b** Outer structure **c** Entire binary gadget

$$
\begin{aligned}
x = {} & a^4 + 6(\kappa-1)a^2b^2 + 16(\kappa-1)ab^3 + 12(\kappa-1)(\kappa-2)ab^2c \\
& + 12(\kappa-1)(\kappa-2)abc^2 + 4(\kappa-1)(\kappa-2)(\kappa-3)ac^3 + 3(\kappa-1)(5\kappa-7)b^4 \\
& + 4(\kappa-1)(\kappa-2)(\kappa+5)b^3c + 6(\kappa-1)(\kappa-2)(7\kappa-18)b^2c^2 \\
& + 12(\kappa-3)^2(\kappa-1)(\kappa-2)bc^3 + (\kappa-1)(\kappa-2)(\kappa^3-9\kappa^2+29\kappa-32)c^4 \qquad \textit{and} \\
y = {} & 2a^3b + (\kappa+4)a^2b^2 + 4(\kappa-2)a^2bc + (\kappa-2)a^2c^2 + 2(9\kappa-11)ab^3 \\
& + 2(\kappa-2)(3\kappa+8)ab^2c + 2(\kappa-2)(12\kappa-31)abc^2 + 2(\kappa-2)(2\kappa^2-11\kappa+16)ac^3 \\
& + (7\kappa^2+3\kappa-24)b^4 + 2(\kappa-2)(\kappa^2+31\kappa-70)b^3c + (\kappa-2)(48\kappa^2-234\kappa+301)b^2c^2 \\
& + 2(\kappa-2)(6\kappa^3-45\kappa^2+121\kappa-116)bc^3 \\
& + (\kappa-2)(\kappa-3)(\kappa^3-7\kappa^2+19\kappa-20)c^4.
\end{aligned}
$$

Proof Since $\langle a, b, c\rangle$ is domain invariant, the signature of this gadget is also domain invariant. Any domain invariant binary signature has a succinct signature of type τ_2.

Consider the gadget in Fig. 27a. We assign $\langle a, b, c\rangle$ to both vertices. By Lemma 11.3, this gadget has the succinct quaternary signature f of type τ_4, where f is given in Lemma 11.3.

Now consider the gadget in Fig. 27b. We assign $\langle a, b, c\rangle$ the circle vertices and f to the square vertex. By partitioning the internal edge assignments into parts with the same contribution to the sum, one can verify that this gadget has the succinct binary signature $\langle x, y\rangle$ of type τ_2, where

$$
\begin{aligned}
x = {} & f_{\substack{1\,1\\1\,1}}\left[a^2 + (\kappa-1)b^2\right] \\
& + 4(\kappa-1)f_{\substack{1\,2\\1\,1}}\left[ab + b^2 + (\kappa-2)bc\right] \\
& + (\kappa-1)f_{\substack{1\,2\\1\,2}}\left[2ab + (\kappa-2)b^2\right] \\
& + 2(\kappa^2-3\kappa+2)f_{\substack{1\,3\\1\,2}}\left[ac + 2b^2 + (\kappa-3)bc\right] \\
& + (\kappa-1)f_{\substack{1\,2\\2\,1}}\left[2b^2 + (\kappa-2)c^2\right] \\
& + 2(\kappa^2-3\kappa+2)f_{\substack{1\,3\\2\,1}}\left[b^2 + 2bc + (\kappa-3)c^2\right] \\
& + (\kappa-1)f_{\substack{1\,1\\2\,2}}\left[2b^2 + (\kappa-2)c^2\right] \\
& + 2(\kappa^2-3\kappa+2)f_{\substack{1\,1\\2\,3}}\left[b^2 + 2bc + (\kappa-3)c^2\right] \\
& + (\kappa^3-6\kappa^2+11\kappa-6)f_{\substack{1\,4\\2\,3}}\left[4bc + (\kappa-4)c^2\right] \qquad \text{and} \\
y = {} & f_{\substack{1\,1\\1\,1}}\left[2ab + (\kappa-2)b^2\right] \\
& + 4f_{\substack{1\,2\\1\,1}}\left[ab + (\kappa-2)ac + (2\kappa-3)b^2 + (\kappa-2)^2bc\right] \\
& + f_{\substack{1\,2\\1\,2}}\left[a^2 + 2(\kappa-2)ab + (\kappa^2-3\kappa+3)b^2\right] \\
& + 2(\kappa-2)f_{\substack{1\,3\\1\,2}}\left[2ab + (\kappa-3)ac + 2(\kappa-2)b^2 + (\kappa^2-4\kappa+5)bc\right] \\
& + f_{\substack{1\,2\\2\,1}}\left[2b^2 + 4(\kappa-2)bc + (\kappa^2-5\kappa+6)c^2\right] \\
& + 2(\kappa-2)f_{\substack{1\,3\\2\,1}}\left[3b^2 + 2(2\kappa-5)bc + (\kappa^2-5\kappa+7)c^2\right]
\end{aligned}
$$

$$+ f_{\begin{smallmatrix}1&1\\2&2\end{smallmatrix}}\left[2b^2 + 4(\kappa-2)bc + (\kappa^2-5\kappa+6)c^2\right]$$
$$+ 2(\kappa-2)f_{\begin{smallmatrix}1&1\\2&3\end{smallmatrix}}\left[3b^2 + 2(2\kappa-5)bc + (\kappa^2-5\kappa+7)c^2\right]$$
$$+ (\kappa^2-5\kappa+6)f_{\begin{smallmatrix}1&4\\2&3\end{smallmatrix}}\left[4b^2 + 4(\kappa-3)bc + (\kappa^2-5\kappa+8)c^2\right].$$

Substituting for the entries of f gives the result stated in Lemma 11.7. □

Lemma 11.8 *Suppose $\kappa \geq 3$ is the domain size and $a, b, c \in \mathbb{C}$. Let $\langle a, b, c\rangle$ be a succinct ternary signature of type τ_3. If $\langle a, b, c\rangle$ is assigned to every vertex of the gadget in Fig. 28c, then the resulting signature is the binary succinct signature $\langle x, y\rangle$ of type τ_2, where x and y are given in Table 4.*

Proof Since $\langle a, b, c\rangle$ is domain invariant, the signature of this gadget is also domain invariant. Any domain invariant binary signature has a succinct signature of type τ_2.

Consider the gadget in Fig. 28a. We assign $\langle a, b, c\rangle$ to all vertices. By Lemma 11.4, this gadget has the succinct ternary signature $f = \langle a_0, b_0, c_0\rangle$ of type τ_4, where a_0, b_0, and c_0 are given in the statement of Lemma 11.4 as a', b', and c', respectively.

Now consider the gadget in Fig. 28b. We assign f to the vertices. By Lemma 11.2, the resulting gadget has the binary succinct signature $\langle x, y\rangle$ of type τ_2, where

$$x = a_0^2 + 3(\kappa-1)b_0^2 + (\kappa-1)(\kappa-2)c_0^2 \qquad \text{and}$$
$$y = 2a_0b_0 + \kappa b_0^2 + 4(\kappa-2)b_0c_0 + (\kappa-2)(\kappa-3)c_0^2.$$

Substituting for a_0, b_0, and c_0 gives the result in Table 4. □

Beyond the gadgets in this section, there are two 9-by-9 recurrence matrices that appear in our proofs (see Table 1 and Table 2). No entry in those recurrence matrices is any harder to compute than any signature entry appearing in this section. The difficulty with these recurrence matrices is the sheer number of terms that must be computed.

12 Appendix 2: More binary interpolation

For some settings of $a, b, c \in \mathbb{C}$, Lemma 9.3 and Lemma 9.11 do not apply. However, these settings are easily handled on a case-by-case basis.

Lemma 12.1 *Suppose $\kappa \geq 3$ is the domain size. Let $\mathcal{F}$ be a signature set containing the succinct unary signature $\langle 1\rangle$ of type τ_1 and any of the following succinct ternary signatures of type τ_3:*

1. $\langle \kappa - 2 \pm i\kappa\sqrt{2(\kappa-2)}, \kappa-2, -2\rangle$;
2. $\langle (\kappa-2)^2 \pm i\kappa\sqrt{\kappa^2-4}, -2(\kappa-2), 4\rangle$;

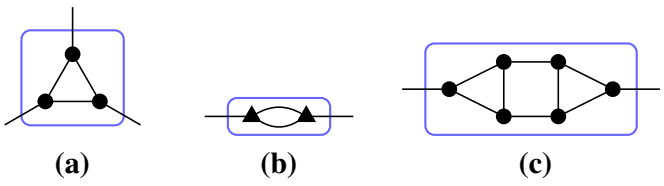


Fig. 28 Decomposition of a binary gadget. All *circle* vertices are assigned $\langle a, b, c\rangle$, and the triangle vertices in (**b**) is assigned the signature of the gadget in (**a**). **a** Inner structure, **b** outer structure **c** entire binary gadget

Table 4 The signature of the gadget in Fig. 28c is $\langle x, y\rangle$ for the x and y above

$$
\begin{aligned}
x = {} & a^6 + 9(\kappa-1)a^4b^2 + 32(\kappa-1)a^3b^3 + 18(\kappa-1)(\kappa-2)a^3b^2c + 12(\kappa-1)(\kappa-2)a^3bc^2 \\
& + 2(\kappa-1)(\kappa-2)(\kappa-3)a^3c^3 + 3(\kappa-1)(16\kappa-7)a^2b^4 + 6(\kappa-1)(\kappa-2)(\kappa+19)a^2b^3c \\
& + 18(\kappa-1)(\kappa-2)(4\kappa-7)a^2b^2c^2 + 6(\kappa-1)(\kappa-2)(\kappa^2+2\kappa-13)a^2bc^3 \\
& + 3(\kappa-1)(\kappa-2)(3\kappa^2-17\kappa+25)a^2c^4 + 6(\kappa-1)(\kappa^2+27\kappa-42)ab^5 \\
& + 6(\kappa-1)(\kappa-2)(40\kappa-41)ab^4c + 24(\kappa-1)(\kappa-2)(3\kappa^2+8\kappa-36)ab^3c^2 \\
& + 6(\kappa-1)(\kappa-2)(\kappa^3+50\kappa^2-285\kappa+393)ab^2c^3 \\
& + 6(\kappa-1)(\kappa-2)(13\kappa^3-108\kappa^2+311\kappa-307)abc^4 \\
& + 6(\kappa-1)(\kappa-2)(\kappa-3)(\kappa^3-8\kappa^2+24\kappa-26)ac^5 \\
& + (\kappa-1)(\kappa^3+83\kappa^2-189\kappa+81)b^6 + 18(\kappa-1)(\kappa-2)(4\kappa^2+13\kappa-43)b^5c \\
& + 3(\kappa-1)(\kappa-2)(7\kappa^3+222\kappa^2-1156\kappa+1442)b^4c^2 \\
& + 2(\kappa-1)(\kappa-2)(\kappa^4+221\kappa^3-1725\kappa^2+4576\kappa-4153)b^3c^3 \\
& + 3(\kappa-1)(\kappa-2)(43\kappa^4-441\kappa^3+1791\kappa^2-3393\kappa+2505)b^2c^4 \\
& + 6(\kappa-1)(\kappa-2)(\kappa-3)(3\kappa^4-29\kappa^3+116\kappa^2-228\kappa+182)bc^5 \\
& + (\kappa-1)(\kappa-2)(\kappa^6-15\kappa^5+98\kappa^4-361\kappa^3+798\kappa^2-1004\kappa+556)c^6
\end{aligned}
$$

and

$$
\begin{aligned}
y = {} & 2a^5b + (\kappa+8)a^4b^2 + 4(\kappa-2)a^4bc + 2(\kappa-2)a^4c^2 + 4(9\kappa-11)a^3b^3 + 2(\kappa-2)(3\kappa+17)a^3b^2c \\
& + 4(\kappa-2)(7\kappa-18)a^3bc^2 + 2(\kappa-3)^2(\kappa-2)a^3c^3 + (23\kappa^2+49\kappa-114)a^2b^4 \\
& + 2(\kappa-2)(\kappa^2+94\kappa-147)a^2b^3c + 6(\kappa-2)(12\kappa^2-34\kappa+17)a^2b^2c^2 \\
& + 2(\kappa-2)(3\kappa^3+9\kappa^2-97\kappa+149)a^2bc^3 + (\kappa-2)(9\kappa^3-68\kappa^2+181\kappa-171)a^2c^4 \\
& + 2(3\kappa^3+73\kappa^2-183\kappa+99)ab^5 + 2(\kappa-2)(96\kappa^2-43\kappa-255)ab^4c \\
& + 4(\kappa-2)(16\kappa^3+94\kappa^2-655\kappa+855)ab^3c^2 \\
& + 2(\kappa-2)(3\kappa^4+159\kappa^3-1233\kappa^2+3164\kappa-2809)ab^2c^3 \\
& + 2(\kappa-2)(39\kappa^4-375\kappa^3+1425\kappa^2-2555\kappa+1825)abc^4 \\
& + 2(\kappa-2)(3\kappa^5-36\kappa^4+181\kappa^3-482\kappa^2+686\kappa-418)ac^5 \\
& + (\kappa^4+50\kappa^3-17\kappa^2-396\kappa+486)b^6 \\
& + 2(\kappa-2)(28\kappa^3+251\kappa^2-1302\kappa+1467)b^5c \\
& + (\kappa-2)(19\kappa^4+745\kappa^3-5374\kappa^2+12664\kappa-10320)b^4c^2 \\
& + 2(\kappa-2)(\kappa^5+224\kappa^4-2062\kappa^3+7371\kappa^2-12357\kappa+8227)b^3c^3 \\
& + (\kappa-2)(129\kappa^5-1464\kappa^4+6952\kappa^3-17464\kappa^2+23397\kappa-13387)b^2c^4 \\
& + 2(\kappa-2)(9\kappa^6-123\kappa^5+727\kappa^4-2405\kappa^3+4754\kappa^2-5374\kappa+2718)bc^5 \\
& + (\kappa-3)(\kappa-2)(\kappa^6-13\kappa^5+74\kappa^4-239\kappa^3+470\kappa^2-544\kappa+292)c^6.
\end{aligned}
$$

3. $\langle -(2\kappa-3)\big[2(\kappa-2) \pm i\kappa\sqrt{2(\kappa-2)}\big], -2(\kappa-3)(\kappa-2) \pm i\kappa\sqrt{2(\kappa-2)}, 4(2\kappa-3)\rangle$ *with* $\kappa \neq 4$;
4. $\langle -\kappa^2+2, 2, 2\rangle$;
5. $\langle \kappa^2-6\kappa+6, -2(\kappa-3), 6\rangle$;
6. $\langle (\kappa-3)(\kappa-2)^2 \pm i\kappa(2\kappa-3)\sqrt{\kappa^2-4}, -3(\kappa-2)^2 \mp i\kappa\sqrt{\kappa^2-4}, 2(5\kappa-6)\rangle$;
7. $\langle -(\kappa-1)\big[5(\kappa-2) \pm 3i\kappa\sqrt{2(\kappa-2)}\big], -(\kappa-2)(3\kappa-5) \pm i\kappa\sqrt{2(\kappa-2)}, 9\kappa-10\rangle$;
8. $\langle (\kappa-1)\big[(\kappa-2)(2\kappa+3) \pm 3\kappa\sqrt{\kappa^2-5\kappa+6}\big], (\kappa-3)(\kappa-2) \mp \kappa\sqrt{\kappa^2-5\kappa+6}, -5\kappa+6\rangle$;
9. $\langle (\kappa-1)\big[(\kappa-2)(2\kappa-7) \pm 3i\kappa\sqrt{\kappa^2-\kappa-2}\big], -(\kappa-2)(5\kappa-7) \mp i\kappa\sqrt{\kappa^2-\kappa-2}, 13\kappa-14\rangle$;
10. $\langle 1, 0, -2\rangle$ *with* $\kappa = 3$;
11. $\langle \pm i\sqrt{2}, 0, 1\rangle$ *with* $\kappa = 3$;
12. $\langle -1 \pm i\sqrt{2}, 0, 1\rangle$ *with* $\kappa = 3$;
13. $\langle -1 \pm 3i\sqrt{3}, 0, 2\rangle$ *with* $\kappa = 3$;

Then

$$\text{Pl-Holant}(\mathcal{F} \cup \{\langle x, y\rangle\}) \leq_T \text{Pl-Holant}(\mathcal{F})$$

for any $x, y \in \mathbb{C}$, *where* $\langle x, y\rangle$ *is a succinct binary signature of type* τ_2.

Proof In each case, we use the recursive construction in Fig. 7. We simply state which gadget we use, the signature of that gadget, and the eigenvalues of its associated recurrence matrix (cf. Lemma 4.11). Then the result easily follows from Corollary 4.13 as the eigenvalues have distinct complex norms.

We use three possible gadgets, which are in Figs. 19a, 27c, and 28c. The signatures for the last two gadgets are given by Lemmas 11.7 and 11.8, respectively.

1. For $\langle \kappa - 2 \pm i\kappa\sqrt{2(\kappa-2)}, \kappa - 2, -2\rangle$, we first use the gadget in Fig. 27c. Let $\gamma = \pm i\sqrt{2(\kappa-2)}$. Up to a nonzero factor of $\frac{(\gamma-2)^7\gamma^2(\gamma+2)^3}{64}$, the signature of the gadget is $\langle -1, 1\rangle$, which means the eigenvalues are $\kappa - 2$ and -2. If $\kappa \neq 4$, then these eigenvalues have distinct complex norms. Otherwise, $\kappa = 4$ and we use the gadget in Fig. 28c. Up to a factor of $\pm 65536i$, the signature of this gadget is $\langle 1, -3\rangle$, which means the eigenvalues are -8 and 4.
2. For $\langle (\kappa-2)^2 \pm i\kappa\sqrt{\kappa^2-4}, -2(\kappa-2), 4\rangle$, we first use the gadget in Fig. 27c. Let $\gamma = \pm i\sqrt{\kappa^2-4}$. Up to a nonzero factor of $-4(\kappa-2)\kappa^3(\kappa^2-4\gamma-8)$, the signature of this gadget is $\langle \kappa^2-6\kappa+4, -2(\kappa-4)\rangle$, which means the eigenvalues are $-(\kappa-2)^2$ and $\kappa^2 - 4k - 4$. If $\kappa \geq 5$, then these eigenvalues have opposite signs but cannot be the negative of each other. Thus, they have distinct complex norms. The same conclusion holds for $\kappa = 3$ by direct inspection. Otherwise, $\kappa = 4$ and we use the gadget in Fig. 28c. Up to a factor of 2097152, the signature of this gadget is $\langle 5, 1\rangle$, which means the eigenvalues are 8 and 4.
3. For $\langle -(2\kappa-3)\big[2(\kappa-2) \pm i\kappa\sqrt{2(\kappa-2)}\big], -2(\kappa-3)(\kappa-2) \pm i\kappa\sqrt{2(\kappa-2)}, 4(2\kappa-3)\rangle$, we have $\kappa \neq 4$. We use the gadget in Fig. 27c. Let $\gamma = \pm i\sqrt{2(\kappa-2)}$. Up to a nonzero factor of $-4(\kappa-2)\kappa^6(3\kappa-4)(4\kappa^2-28\kappa+41-4\gamma(2\kappa-5))$, the signature of the gadget is $\frac{1}{\kappa}\langle 3\kappa-4, \kappa-4\rangle$, which means the eigenvalues are $\kappa - 2$ and 2.
4. For $\langle -\kappa^2+2, 2, 2\rangle$, we use the gadget in Fig. 27c. Up to a nonzero factor of $(\kappa-2)\kappa^5$, the signature for this gadget is $\langle \kappa^2+2\kappa-4, -4\rangle$, which means the eigenvalues are $(\kappa-2)\kappa$ and $\kappa(\kappa+2)$.
5. For $\langle \kappa^2 - 6\kappa + 6, -2(\kappa-3), 6\rangle$, we use the gadget in Fig. 27c. Up to a nonzero factor of $(\kappa-2)\kappa^5$, the signature for this gadget is $\langle \kappa^2+2\kappa-4, -4\rangle$, which means the eigenvalues are $(\kappa-2)\kappa$ and $\kappa(\kappa+2)$.
6. For $\langle (\kappa-3)(\kappa-2)^2 \pm i\kappa(2\kappa-3)\sqrt{\kappa^2-4}, -3(\kappa-2)^2 \mp i\kappa\sqrt{\kappa^2-4}, 2(5\kappa-6)\rangle$, we use the gadget in Fig. 27c. Let $\gamma = \pm i\sqrt{\kappa^2-4}$. Up to a nonzero factor of $(\gamma-2)^2(\gamma+2)^2(\kappa-2)\kappa[7\kappa^2+60\kappa-164+8\gamma(3\kappa-10)]$, the signature of the gadget is $\langle -\kappa^4+6\kappa^3+4\kappa^2-24\kappa+16, 2(\kappa^3-2\kappa^2-8\kappa+8)\rangle$, which means the eigenvalues are $\lambda_1 = (\kappa-2)\kappa(\kappa^2+2\kappa-4)$ and $\lambda_2 = -\kappa(\kappa+2)(\kappa^2-6\kappa+4)$. For $3 \leq \kappa \leq 5$, one can directly check that these eigenvalues have distinct complex norms. For $\kappa \geq 6$, we have $\lambda_2 < 0$, so these eigenvalues have the same complex norm preciously when $\lambda_1 = -\lambda_2$. However, $\lambda_1 + \lambda_2 = 4\kappa^3 \neq 0$, so the eigenvalues have distinct complex norms.
7. For $\langle -(\kappa-1)\big[5(\kappa-2) \pm 3i\kappa\sqrt{2(\kappa-2)}\big], -(\kappa-2)(3\kappa-5) \pm i\kappa\sqrt{2(\kappa-2)}, 9\kappa-10\rangle$, we first use the gadget in Fig. 27c. Let $\gamma = \pm i\sqrt{2(\kappa-2)}$. Up to a nonzero factor of $-(\kappa-2)(\kappa-1)\kappa^5[81\kappa^2-756\kappa+1252-24(9\kappa-26)\gamma]$, the signature of this gadget is $\langle 5\kappa-6, \kappa-6\rangle$, which means the eigenvalues are $\kappa-2$ and 4. If $\kappa \neq 6$, then these eigenvalues have distinct complex norms. Otherwise, $\kappa = 6$ and we use the gadget

in Fig. 28c. Up to a factor of $-17199267840(1169 \pm 450i\sqrt{2})$, the signature of this gadget is $\langle 7, 13\rangle$, which means the eigenvalues are 72 and -6.

8. For $\langle(\kappa-1)\big[(\kappa-2)(2\kappa+3)\pm 3\kappa\sqrt{\kappa^2-5\kappa+6}\big], (\kappa-3)(\kappa-2)\mp\kappa\sqrt{\kappa^2-5\kappa+6}, -5\kappa+6\rangle$, we first use the gadget in Fig. 27c. Let $\gamma = \pm\sqrt{\kappa^2-5\kappa+6}$. Up to a factor of $(\kappa-2)(\kappa-1)\kappa^5[313\kappa^2-1500\kappa+1764-24(13\kappa-30)\gamma]$, the signature of this gadget is $\langle\kappa^3-3\kappa^2+3, -\kappa+3\rangle$, which means the eigenvalues are $\lambda_1 = (\kappa-2)^2\kappa$ and $\lambda_2 = \kappa(\kappa^2-3\kappa+1)$. If $\kappa \geq 4$, these eigenvalues are positive, so they have the same complex norm preciously when $\lambda_1 = \lambda_2$. However, $\lambda_1 - \lambda_2 = -(\kappa-3)\kappa \neq 0$, so the eigenvalues have distinct complex norms. Otherwise, $\kappa = 3$ and we use the gadget in Fig. 28c. Up to a factor of 9565938, the signature of this gadget is $\langle 5, 2\rangle$, which means the eigenvalues are 9 and 3.
9. For $\langle(\kappa-1)\big[(\kappa-2)(2\kappa-7)\pm 3i\kappa\sqrt{\kappa^2-\kappa-2}\big], -(\kappa-2)(5\kappa-7)\mp i\kappa\sqrt{\kappa^2-\kappa-2}, 13\kappa-14\rangle$, we use the gadget in Fig. 27c. Let $\gamma = \pm i\sqrt{\kappa^2-\kappa-2}$. Up to a nonzero factor of $(\kappa-2)(\kappa-1)\kappa^5[119\kappa^2+76\kappa-772+24(5\kappa-22)\gamma]$, the signature of this gadget is $\langle-\kappa^3+7\kappa^2-4\kappa-3, 2\kappa^2-7\kappa-3\rangle$, which means the eigenvalues are $\lambda_1 = (\kappa-2)\kappa^2$ and $\lambda_2 = -\kappa(k^2-5\kappa-3)$. For $3 \leq \kappa \leq 5$, one can directly check that these eigenvalues have distinct complex norms. For $\kappa \geq 6$, we have $\lambda_2 < 0$, so these eigenvalues have the same complex norm preciously when $\lambda_1 = -\lambda_2$. However, $\lambda_1 + \lambda_2 = 3\kappa(\kappa+1) \neq 0$, so the eigenvalues have distinct complex norms.
10. For $\langle 1, 0, -2\rangle$ with $\kappa = 3$, we use the gadget in Fig. 27c. Up to a factor of 3, the signature of this gadget is $\langle 11, -4\rangle$, which means the eigenvalues are 3 and 15.
11. For $\langle\pm i\sqrt{2}, 0, 1\rangle$ with $\kappa = 3$, we use the gadget in Fig. 19a. The signature of this gadget is $\langle\pm i\sqrt{2}, 1\rangle$, which means the eigenvalues are $2 \pm i\sqrt{2}$ and $-1 \pm i\sqrt{2}$.
12. For $\langle-1 \pm i\sqrt{2}, 0, 1\rangle$ with $\kappa = 3$, we use the gadget in Fig. 19a. The signature of this gadget is $\langle-1 \pm i\sqrt{2}, 1\rangle$, which means the eigenvalues are $1 \pm i\sqrt{2}$ and $-2 \pm i\sqrt{2}$.
13. For $\langle-1 \pm 3i\sqrt{3}, 0, 2\rangle$ with $\kappa = 3$, we use the gadget in Fig. 27c. Up to a factor of 72, the signature of this gadget is $\frac{1}{3}\langle 25 \pm 13\sqrt{3}, -5 \pm i\sqrt{3}\rangle$, which means the eigenvalues are $5(1 \pm i\sqrt{3})$ and $2(5 \pm 2\sqrt{3})$. □

13 Appendix 3: Invariance properties from row eigenvectors

The purpose of this section is to show how a recursive construction in an interpolation proof can be used to form a hypothesis about possible invariance properties. We often find that no matter what constructions one considers, all signatures they produce satisfy certain invariance. Instead of defining this notion formally, we prove the following lemma as an example. After this lemma and its proof, we explain that this invariance can be suggested by certain recursive constructions in an alternative proof of Theorem 4.8, that it is #P-hard to count edge κ-coloring over planar κ-regular graphs for all $\kappa \geq 3$. This alternative proof uses the interpolation techniques that we developed in Sect. 6.

Lemma 13.1 *Suppose $\kappa \geq 3$ is the domain size. If F is a planar $\{\mathrm{AD}_{\kappa,\kappa}\}$-gate with succinct quaternary signature $\langle a, b, c, d, e\rangle$ of type τ_{color}, then $a + c = b + d$.*

Proof Fix two distinct colors $g, y \in [\kappa]$. We define the *swap* of an edge colored g or y to be the opposite of these two colors. That is, swapping the color of an edge colored g (resp. y) gives the same edge colored y (resp. g). The ith external edge of F is the external edge that corresponds to the ith input of F. Recall that the input edges of F are ordered cyclically.

For $1 \leq i \leq 4$, let S_i (resp. S'_i) be the set of colorings of the edges (both internal and external) of F with an external coloring in the partition P_i of the succinct signature type τ_{color} such that the first external edge of F is colored g (resp. y) and the remaining external edges are either colored g or y (as dictated by P_i). Note that $|S_i| = |S'_i|$ for $1 \leq i \leq 4$. Furthermore, the sizes of these sets do not depend on the choice of $g, y \in [\kappa]$. Thus, it suffices to show that

$$|S_1 \cup S'_1 \cup S_3 \cup S'_3| = |S_2 \cup S'_2 \cup S_4 \cup S'_4|. \tag{23}$$

Let $\sigma \in S_1 \cup S'_1 \cup S_3 \cup S'_3$ be a coloring of F. Starting at the first external edge of F, there is a unique path π_1 that alternates in edge colors between g and y and terminates at another external edge of F. Suppose for a contradiction that this path terminates at the third external edge of F. Also consider the unique path π_2 that starts at the second external edge of F, alternates in edge colors between g and y, and must terminate at the fourth external edge of F. These two paths must cross somewhere since their ends are crossed. By planarity, they must cross at a vertex, and yet they must be vertex disjoint. This is a contradiction. Therefore, the path π_1 either terminates at the second or fourth external edge of F.

Suppose π_1 terminates at the second external edge of F. If $\sigma \in S_1$ (resp. $\sigma \in S'_1$), then swapping the colors of every edge in π_1 gives a new coloring $\pi'_1 \in S'_2$ (resp. $\pi'_1 \in S_2$). Similarly, if $\sigma \in S_3$ (resp. $\sigma \in S'_3$), then swapping the colors of every edge in π_1 gives a new coloring $\pi'_1 \in S'_4$ (resp. $\pi'_1 \in S_4$).

Otherwise, π_1 terminates at the fourth external edge of F. If $\sigma \in S_1$ (resp. $\sigma \in S'_1$), then swapping the colors of every edge in π_1 gives a new coloring $\pi'_1 \in S'_4$ (resp. $\pi'_1 \in S_4$). Similarly, if $\sigma \in S_3$ (resp. $\sigma \in S'_3$), then swapping the colors of every edge in π_1 gives a new coloring $\pi'_1 \in S'_2$ (resp. $\pi'_1 \in S_2$).

Furthermore, this mapping from $S_1 \cup S'_1 \cup S_3 \cup S'_3$ to $S_2 \cup S'_2 \cup S_4 \cup S'_4$ is invertible. Therefore, we have established (23), as desired. □

Now we give an alternative proof of Theorem 4.8. The recursive construction in this proof will suggest the invariance in Lemma 13.1.

Let $q(x, \kappa) = x^3 - x^2 + x - (\kappa - 1)$. First we determine the nature of the roots of $q(x, \kappa)$.

Lemma 13.2 *For all $\kappa \in \mathbb{Z}$, the polynomial $q(x, \kappa)$ in x has one real root $r \in \mathbb{R}$ and two nonreal complex conjugate roots $\alpha, \overline{\alpha} \in \mathbb{C}$, such that $\alpha + \overline{\alpha} = 1 - r$ and $\alpha\overline{\alpha} = r^2 - r + 1$.*

Furthermore, if $q(x, \kappa)$ is reducible in $\mathbb{Q}[x]$ and $\kappa \geq 3$, then $r \geq 2$ is an integer.

Proof The discriminant of $q(x, \kappa)$ with respect to x is $\text{disc}_x(q(x, \kappa)) = -27\kappa^2 + 68\kappa - 44 < 0$, so $q(x, \kappa)$ has one real root $r \in \mathbb{R}$ and two nonreal complex conjugate roots $\alpha, \overline{\alpha} \in \mathbb{C}$. We have

$$\begin{aligned} \alpha + \overline{\alpha} + r &= 1 \\ \alpha\overline{\alpha} + (\alpha + \overline{\alpha})r &= 1 \\ \alpha\overline{\alpha}r &= \kappa - 1. \end{aligned}$$

It follows that $\alpha + \overline{\alpha} = 1 - r$, $\alpha\overline{\alpha} = r^2 - r + 1$, and

$$\kappa = r^3 - r^2 + r + 1. \tag{24}$$

If $q(x,\kappa)$ is reducible in $\mathbb{Q}[x]$ with $\kappa \geq 3$, then $r \in \mathbb{Z}$ by Gauss's Lemma and so $r \geq 2$ by (24). □

Lemma 13.3 *If $\kappa \geq 3$ is an integer, then the roots of $x^3 - x^2 + x - (\kappa - 1)$ satisfy the lattice condition.*

Proof If $q(x,\kappa)$ is irreducible in $\mathbb{Q}[x]$, then its roots satisfy the lattice condition by Lemma 6.4.

Otherwise, $q(x,\kappa)$ is reducible in $\mathbb{Q}[x]$. By Lemma 13.2, $q(x,\kappa)$ has one real root $r \in \mathbb{Z}$ satisfying $r \geq 2$ and two nonreal complex conjugate roots $\alpha, \overline{\alpha} \in \mathbb{C}$ satisfying $\alpha + \overline{\alpha} = 1 - r$ and $\alpha\overline{\alpha} = r^2 - r + 1$. Suppose there exist $i, j, k \in \mathbb{Z}$ such that $\alpha^i \overline{\alpha}^j = r^k$ and $i + j = k$. We want to show that $i = j = k = 0$.

There is an element in the Galois group of $q(x,\kappa)$ that fixes $\mathbb{Q}$ pointwise and swaps α and $\overline{\alpha}$. Thus, $\alpha^j \overline{\alpha}^i = r^k$. Dividing these two equations gives $(\alpha/\overline{\alpha})^{i-j} = 1$. We claim that $\omega = \alpha/\overline{\alpha}$ cannot be a root of unity and hence $i = j$. For a contradiction, suppose ω is a dth primitive root of unity. Let $f(x) = (x - \alpha)(x - \overline{\alpha}) = x^2 + (r - 1)x + (r^2 - r + 1) \in \mathbb{Z}[x]$. Then ω belongs to the splitting field of f over $\mathbb{Q}$, which is a degree 2 extension over $\mathbb{Q}$. This implies that the Euler totient function $\phi(d) \mid 2$. Therefore, $d \in \{1, 2, 3, 4, 6\}$. Let $\rho = \frac{\alpha+\overline{\alpha}}{\alpha\overline{\alpha}} = \frac{1+\omega}{\omega\overline{\alpha}} = \frac{1-r}{r^2-r+1} \in \mathbb{Q}$. Since $r \geq 2$, we have $\rho \neq 0$ and hence $d \neq 2$. Moreover, $f(x) = x^2 - (2+\omega+\omega^{-1})\rho^{-1}x + (2+\omega+\omega^{-1})\rho^{-2}$. Notice that the quantity $2+\omega+\omega^{-1}$ is 4, 1, 2, 3, respectively, when $d = 1, 3, 4, 6$. As $(2+\omega+\omega^{-1})\rho^{-2} \in \mathbb{Z}$, we get that ρ^{-1} must be an integer when $d = 3, 4, 6$ and half an integer when $d = 1$. However $\rho^{-1} = -r + \frac{1}{r-1}$. The only possibility is $r = 3$ and $d = 1$, yet it is easy to check that $\omega \neq 1$ when this holds. This proves the claim.

From $\alpha\overline{\alpha} = r^2 - r + 1$, we have $(r^2 - r + 1)^i = (\alpha\overline{\alpha})^i = r^k$. Since r and $r^2 - r + 1$ are relatively prime and $r \geq 2$, we must have $i = k = 0$. □

Alternative proof of Theorem 4.8 As before, let κ be the domain size of all Holant problems in this proof and let $\langle 2, 1, 0, 1, 0\rangle$ be a succinct quaternary signature of type τ_{color}. We reduce from Pl-Holant($\langle 2, 1, 0, 1, 0\rangle$) to Pl-Holant($\text{AD}_{\kappa,\kappa}$), which denotes the problem of counting edge κ-colorings in planar κ-regular graphs as a Holant problem. Then by Corollary 4.7, we conclude that Pl-Holant($\text{AD}_{\kappa,\kappa}$) is #P-hard.

Consider the gadget in Fig. 5, where the bold edge represents $\kappa - 2$ parallel edges. We assign $\text{AD}_{\kappa,\kappa}$ to both vertices. Up to a nonzero factor of $(\kappa - 2)!$, this gadget has the succinct quaternary signature $f = \langle 0, 1, 1, 0, 0\rangle$ of type τ_{color}. Now consider the recursive construction in Fig. 29. All vertices are assigned the signature f. Let f_s be the succinct

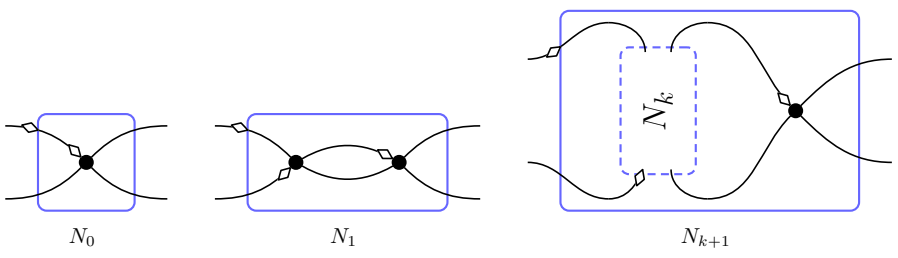


Fig. 29 Alternate recursive construction to interpolate $\langle 2, 1, 0, 1, 0\rangle$. The vertices are assigned the signature of the gadget in Fig. 5

quaternary signature of type τ_{color} for the sth gadget of the recursive construction. Then $f_0 = f$ and $f_s = M^s f_0$, where

$$M = \begin{bmatrix} 0 & 0 & 0 & \kappa-1 & 0 \\ 1 & 0 & 0 & \kappa-2 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

The row vectors

$$(1, -1, 1, -1, 0) \qquad \text{and} \qquad (0, 0, 0, 0, 1)$$

are linearly independent row eigenvectors of M, with eigenvalues -1 and 1, respectively, that are orthogonal to the initial signature f_0. Note that our target signature $\langle 2, 1, 0, 1, 0\rangle$ is also orthogonal to these two row eigenvectors.

Up to a factor of $(x-1)(x+1)$, the characteristic polynomial of M is $x^3 - x^2 + x - (\kappa - 1)$. The roots of this polynomial satisfy the lattice condition by Lemma 13.3. In particular, these three roots are distinct. By Lemma 13.2, the only real root is at least 2. Thus, all five eigenvalues of M are distinct, so M is diagonalizable.

The 3-by-3 matrix in the upper-left corner of $[f_0\, Mf_0\, \ldots\, M^4 f_0]$ is $\left[\begin{smallmatrix} 0 & 0 & \kappa-1 \\ 1 & 0 & \kappa-2 \\ 1 & 1 & 0 \end{smallmatrix}\right]$. Its determinant is $\kappa - 1 \neq 0$. Thus, $[f_0\, Mf_0\, \ldots\, M^4 f_0]$ has rank at least 3, so by Lemma 6.2, f_0 is not orthogonal to the three remaining row eigenvectors of M.

Therefore, by Lemma 6.6, we can interpolate $\langle 2, 1, 0, 1, 0\rangle$, which completes the proof. □

Notice that the row eigenvector $(1, -1, 1, -1, 0)$ suggests that $a - b + c - d = 0$ is an invariance shared by all signatures of symmetric ternary constructions. Some row eigenvectors, like $(0, 0, 0, 0, 1)$, only indicate an invariance present in some recursive constructions. (When $\kappa = 4$, there are recursive constructions for which $(0, 0, 0, 0, 1)$ is not a row eigenvector of the recurrence matrix.) The row eigenvector $(1, -1, 1, -1, 0)$ is more intrinsic; it must appear because of the invariance present in all constructions as shown in Lemma 13.1.

This suggests an approach to discover new invariance properties. Given a set $\mathcal{F}$ of signatures, create some recursive construction and inspect the row eigenvectors of the resulting recurrence matrix. For example, consider the set $\mathcal{F}_{\mathfrak{A}} = \{\langle a, b, c\rangle \mid a, b, c \in \mathbb{C} \text{ and } \mathfrak{A} = 0\}$, where $\mathfrak{A} = a - 3b + 2c$. It seems that $\mathcal{F}_{\mathfrak{A}}$ is closed under symmetric ternary constructions, such as those in Sect. 7.1. In particular, $(1, -3, 2)$ is a row eigenvector of the recurrence matrix for every recursive ternary construction with symmetric signatures that we tried. However, we do not know how to prove this closure property.

Received: 9 June 2015 Accepted: 5 May 2016
Published online: 01 September 2016

References

1. Arratia, R., Bollobás, B., Sorkin, G.B.: The interlace polynomial: a new graph polynomial. In: SODA, pp. 237–245. Society for Industrial and Applied Mathematics (2000)
2. Borgs, C., Chayes, J., Lovász, L., Sós, V.T., Vesztergombi, K.: Counting graph homomorphisms. In: Klazar, M., Kratochvíl, J., Loebl, M., Matoušek, J., Valtr, P., Thomas, R. (eds.) Topics in Discrete Mathematics, volume 26 of Algorithms and Combinatorics, pp. 315–371. Springer, Berlin (2006)
3. Brylawski, T., Oxley, J.: The Tutte polynomial and its applications. In: White, N. (ed.) Matriod Applications, pp. 123–225. Cambridge University Press, Cambridge (1992)

4. Bulatov, A., Dyer, M., Ann Goldberg, L., Jalsenius, M., Richerby, D.: The complexity of weighted Boolean #CSP with mixed signs. Theor. Comput. Sci. **410**(38–40), 3949–3961 (2009)
5. Bulatov, A., Grohe, M.: The complexity of partition functions. Theor. Comput. Sci. **348**(2), 148–186 (2005)
6. Bulatov, A.A.: A dichotomy theorem for constraint satisfaction problems on a 3-element set. J. ACM **53**(1), 66–120 (2006)
7. Bulatov, A.A.: The complexity of the counting constraint satisfaction problem. J. ACM **60**(5), 34:1–34:41 (2013)
8. Bulatov, A.A., Dalmau, V.: Towards a dichotomy theorem for the counting constraint satisfaction problem. Inform. Comput. **205**(5), 651–678 (2007)
9. Cai, J.-Y., Chen, X.: Complexity of counting CSP with complex weights. In: STOC, pp. 909–920. ACM (2012)
10. Cai, J.-Y., Chen, X., Lipton, R.J., Lu, P.: On tractable exponential sums. In: FAW, pp. 148–159. Springer, Berlin (2010)
11. Cai, J.-Y., Chen, X., Lu, P.: Non-negatively weighted #CSP: an effective complexity dichotomy. In: CCC, pp. 45–54. IEEE Computer Society (2011)
12. Cai, J.-Y., Chen, X., Lu, P.: Graph homomorphisms with complex values: a dichotomy theorem. SIAM J. Comput. **42**(3), 924–1029 (2013)
13. Cai, J.-Y., Choudhary, V.: Valiant's Holant theorem and matchgate tensors. Theor. Comput. Sci. **384**(1), 22–32 (2007)
14. Cai, J.-Y., Guo, H., Williams, T.: A complete dichotomy rises from the capture of vanishing signatures (extended abstract). In: STOC, pp. 635–644. ACM (2013)
15. Cai, J.-Y., Huang, S., Pinyan, L.: From Holant to #CSP and back: Dichotomy for Holant^c problems. Algorithmica **64**(3), 511–533 (2012)
16. Cai, J.-Y., Kowalczyk, M.: Spin systems on k-regular graphs with complex edge functions. Theor. Comput. Sci. **461**, 2–16 (2012)
17. Cai, J.-Y., Kowalczyk, M., Williams, T.: Gadgets and anti-gadgets leading to a complexity dichotomy. In: ITCS, pp. 452–467. ACM (2012)
18. Cai, J.-Y., Lu, P., Xia, M.: Holant problems and counting CSP. In: STOC, pp. 715–724. ACM (2009)
19. Cai, J.-Y., Lu, P., Xia, M.: Holographic algorithms with matchgates capture precisely tractable planar #CSP. In: FOCS, pp. 427–436. IEEE Computer Society (2010)
20. Cai, J.-Y., Pinyan, L., Xia, M.: Computational complexity of Holant problems. SIAM J. Comput. **40**(4), 1101–1132 (2011)
21. Cai, J.-Y., Pinyan, L., Xia, M.: Holographic reduction, interpolation and hardness. Comput. Complex. **21**(4), 573–604 (2012)
22. Cai, J.-Y., Lu, P., Xia, M.: Dichotomy for Holant* problems with domain size 3. In: SODA, pp. 1278–1295. SIAM (2013)
23. Cai, J.-Y., Pinyan, L., Xia, M.: Holographic algorithms by Fibonacci gates. Linear Algebra Appl. **438**(2), 690–707 (2013)
24. Cai, J.-Y., Pinyan, L., Xia, M.: The complexity of complex weighted Boolean #CSP. J. Comput. Syst. Sci. **80**(1), 217–236 (2014)
25. Dodson, C.T.J., Poston, T.: Tensor Geometry. Graduate Texts in Mathematics, vol. 130, 2nd edn. Springer, Berlin (1991)
26. Dyer, M., Ann Goldberg, L., Jerrum, M.: The complexity of weighted Boolean #CSP. SIAM J. Comput. **38**(5), 1970–1986 (2009)
27. Dyer, M., Greenhill, C.: The complexity of counting graph homomorphisms. Random Struct. Algorithms **17**(3–4), 260–289 (2000)
28. Dyer, M., Richerby, D.: On the complexity of #CSP. In: STOC, pp. 725–734. ACM (2010)
29. Ellis-Monaghan, J.A.: New results for the Martin polynomial. J. Comb. Theory Ser. B **74**(2), 326–352 (1998)
30. Ellis-Monaghan, J.A.: Identities for circuit partition polynomials, with applications to the Tutte polynomial. Adv. Appl. Math. **32**(1–2), 188–197 (2004)
31. Faltings, G.: Endlichkeitssätze für abelsche Varietäten über Zahlkörpern. Invent. Math. **73**(3), 349–366 (1983)
32. Feder, T., Vardi, M.Y.: The computational structure of monotone monadic SNP and constraint satisfaction: a study through Datalog and group theory. SIAM J. Comput. **28**(1), 57–104 (1998)
33. Gallagher, P.X.: The large sieve and probabilistic Galois theory. In: Proc. Symp. Pure Math., volume 24 of Analytic Number Theory, pp. 91–101. American Mathematical Society (1973)
34. Goldberg, L.A., Grohe, M., Jerrum, M., Thurley, M.: A complexity dichotomy for partition functions with mixed signs. SIAM J. Comput. **39**(7), 3336–3402 (2010)
35. Guo, H., Huang, S., Lu, P., Xia, M.: The complexity of weighted Boolean #CSP modulo k. In: STACS, pp. 249–260. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik (2011)
36. Guo, H., Williams, T.: The complexity of planar Boolean #CSP with complex weights. CoRR, abs/1212.2284 (2012)
37. Guo, H., Williams, T.: The complexity of planar Boolean #CSP with complex weights. In: ICALP, pp. 516–527. Springer, Berlin (2013)
38. Håstad, J.: Tensor rank is NP-complete. J. Algorithm. **11**(4), 644–654 (1990)
39. Holyer, I.: The NP-completeness of edge-coloring. SIAM J. Comput. **10**(4), 718–720 (1981)
40. Jacobson, N.: Basic Algebra I, 2nd edn. W. H. Freeman & Co., San Francisco (1985)
41. Joshi, A.W.: Matrices and Tensors in Physics. New Age International, revised third edition(1995)
42. David Forney Jr., G.: Codes on graphs: normal realizations. IEEE Trans. Inf. Theory **47**(2), 520–548 (2001)
43. Kowalczyk, M., Cai, J.-Y.: Holant problems for regular graphs with complex edge functions. In: STACS, pp. 525–536. Schloss Dagstuhl—Leibniz-Zentrum fuer Informatik (2010)
44. Kowalczyk, M., Cai, J.-Y.: Holant problems for regular graphs with complex edge functions. CoRR. arXiv:1001.0464 (2010)
45. Leven, D., Galil, Z.: NP completeness of finding the chromatic index of regular graphs. J. Algorithm **4**(1), 35–44 (1983)
46. Levin, A.: Private communication (2013)
47. Loeliger, H.-A.: An introduction to factor graphs. IEEE Signal Process. Mag. **21**(1), 28–41 (2004)
48. Markov, I.L., Shi, Y.: Simulating quantum computation by contracting tensor networks. SIAM J. Comput. **38**(3), 963–981 (2008)
49. Martin, P.:. Enumérations Eulériennes dans les multigraphes et invariants de Tutte-Grothendieck. PhD thesis, Joseph Fourier University (1977). http://tel.archives-ouvertes.fr/tel-00287330

50. Müller, P.: Hilbert's irreducibility theorem for prime degree and general polynomials. Israel J. Math. **109**(1), 319–337 (1999)
51. Poonen, B.: Private communication (2013)
52. Siegel, C.L.: Über einige anwendungen diophantischer approximationen. Abh. Pruess. Akad. Wiss. Phys. Math. Kl., pp. 41–69 (1929)
53. Stewart, I.: Galois Theory, 3rd edn. Chapman Hall/CRC Mathematics Series. Taylor & Francis, London (2003)
54. Stiebitz, M., Scheide, D., Toft, B., Favrholdt, L.M.: Graph Edge Coloring: Vizing's Theorem and Goldberg's Conjecture. Wiley, New York (2012)
55. Tait, P.: Remarks on the colourings of maps. Proc. R. Soc. Edinb. **10**, 729 (1880)
56. Valiant, L.G.: Quantum circuits that can be simulated classically in polynomial time. SIAM J. Comput. **31**(4), 1229–1254 (2002)
57. Leslie Valiant, G.: Holographic algorithms. SIAM J. Comput. **37**(5), 1565–1594 (2008)
58. Michel Las Vergnas: Le polynôme de Martin d'un graphe Eulerien. Ann. Discrete Math. **17**, 397–411 (1983)
59. Vertigan, D.: The computational complexity of Tutte invariants for planar graphs. SIAM J. Comput. **35**(3), 690–712 (2005)
60. Vizing, V.G.: Critical graphs with given chromatic class. Metody Diskret. Analiz. **5**, 9–17 (1965)
61. Walsh, P.G.: A quantitative version of Runge's theorem on Diophantine equations. Acta Arith. **62**(2), 157–172 (1992)
62. Welsh, D.: Complexity: Knots. Colourings and Countings. London Mathematical Society Lecture Note Series. Cambridge University Press (1993)

Paper 5

Complexity of Counting CSP with Complex Weights

JIN-YI CAI, University of Wisconsin–Madison and Beijing University
XI CHEN, Columbia University

We give a complexity dichotomy theorem for the counting constraint satisfaction problem (#CSP in short) with algebraic complex weights. To this end, we give three conditions for its tractability. Let $\mathcal{F}$ be any finite set of algebraic complex-valued functions defined on an arbitrary finite domain. We show that #CSP($\mathcal{F}$) is solvable in polynomial time if all three conditions are satisfied and is #P-hard otherwise.

Our dichotomy theorem generalizes a long series of important results on counting problems and reaches a natural culmination: (a) the problem of counting graph homomorphisms is the special case when $\mathcal{F}$ has a single symmetric binary function [Dyer and Greenhill 2000; Bulatov and Grohe 2005; Goldberg et al. 2010; Cai et al. 2013]; (b) the problem of counting directed graph homomorphisms is the special case when $\mathcal{F}$ has a single but not necessarily symmetric binary function [Dyer et al. 2007; Cai and Chen 2010]; (c) the unweighted form of #CSP is when all functions in $\mathcal{F}$ take values in $\{0, 1\}$ [Bulatov 2008; Dyer and Richerby 2013].

Categories and Subject Descriptors: F.2 [**Theory of Computation**]: Analysis of algorithms and problem complexity

General Terms: Theory, Algorithms

Additional Key Words and Phrases: Constraint satisfaction problem, counting problems, complexity dichotomy

ACM Reference Format:
Jin-Yi Cai and Xi Chen. 2017. Complexity of counting CSP with complex weights. J. ACM 64, 3, Article 19 (June 2017), 39 pages.
DOI: http://dx.doi.org/10.1145/2822891

1. INTRODUCTION

It is well known that if NP $\neq$ P, then there is an infinite hierarchy of complexity classes between them, a theorem due to Ladner [1975]. However, for some broad classes of problems a *complexity dichotomy* exists: Every problem in the class is either solvable in polynomial time or NP-hard. Such results include Schaefer's theorem [1978], the dichotomy of Hell and Nešetřil [1990] for H-coloring, a.k.a., graph homomorphism, and some subclasses of the constraint satisfaction problem (CSP in short) [Creignou et al. 2001]. Recent developments include a dichotomy for CSP with a three-element domain [Bulatov 2006] and a dichotomy for CSP over digraphs with no sources or sinks [Barto et al. 2009].

This work is supported by NSF CCF-0914969, NSF CCF-1217549, NSF CCF-1149257, NSF CCF-1423100 and start-up funds from Columbia University.
Authors' addresses: J.-Y. Cai, Computer Sciences Department, University of Wisconsin-Madison, 1210 West Dayton Street, Madison, WI 53706-1685; email: jyc@cs.wisc.edu; X. Chen, Computer Science Department, 500 West 120 Street, Room 450, New York, NY 10027; email: xichen@cs.columbia.edu. A preliminary version appeared in the proceedings of the 44th Symposium on Theory of Computing, 2012.

DOI: http://dx.doi.org/10.1145/2822891

These dichotomy theorems can be seen as providing support to the intuitive notion that most problems studied in computer science are either in P or NP-hard, Ladner's theorem [1975] notwithstanding. However, there are some exceptions. For example, Integer Factoring and Graph Isomorphism are suspected to be neither in P nor NP-hard. A question of foundational importance in complexity theory is this: For how broad a class of problems can one hope to prove a complexity dichotomy theorem? Given a class of problems, what is the criterion that distinguishes the tractable problems from the intractable ones?

CSP provides a sufficiently broad framework to address a large class of problems for which one can hope to prove dichotomies. The famous CSP dichotomy conjecture by Feder and Vardi [1999] on decision CSP motivated much of the subsequent work, but remains open to date (see Hell and Nešetřil [2008] for a recent survey). For counting problems, the natural corresponding framework is called the counting constraint satisfaction problem or #CSP in short, and one can hope to prove dichotomy theorems that give a broad classification of counting problems to be either in P or #P-hard. This naturally leads to the sum-of-products-type computations, or partition functions, which also have a deep root in statistical physics and other fields. For example, the ferromagnetic two-dimensional Ising model consists of a set of variables s_i on each lattice point, called spins, that can be assigned one of two states $\{+1, -1\}$. The Hamiltonian is

$$E(s) = -\sum_{\text{edge } \{i,j\}} s_i s_j.$$

The partition function is $Z = \sum_s e^{-E(s)/kT}$, where k is Boltzmann's constant and T is the (absolute) temperature. Note that the exponential $e^{-E(s)/kT}$ turns this into a sum-of-products function exactly as we discussed in #CSP. In Baxter's classical book on "Exactly solved models in statistical mechanics" [1982], after defining partition functions in Equation (1.4.1) on page 8, he states on page 9 that "The basic problem of equilibrium statistical mechanics is therefore to calculate the sum-over-states in Equation (1.4.1)..."

In this article, we study the complexity of #CSP with algebraic and complex weights. Let $D = \{1, \ldots, d\}$ denote a finite set, called a *domain*, where d is arbitrary. A *weighted constraint language* $\mathcal{F}$ over the domain D is an arbitrary finite set of algebraic complex-valued functions $\{F_1, \ldots, F_h\}$, where $F_i : D^{r_i} \to \mathbb{C}$ for some $r_i \geq 1$. The language $\mathcal{F}$ defines the following counting constraint satisfaction problem, denoted by #CSP($\mathcal{F}$). The input of #CSP($\mathcal{F}$) consists of a tuple $\mathbf{x} = (x_1, \ldots, x_n)$ of variables over D and a finite multiset I of tuples $(F, i_1, \ldots, i_r)$ in which F is an r-ary function from $\mathcal{F}$ and $i_1, \ldots, i_r \in [n] = \{1, \ldots, n\}$. It then defines the following n-ary function F_I over $\mathbf{x} \in D^n$:

$$F_I(\mathbf{x}) = \prod_{(f, i_1, \ldots, i_r) \in I} F(x_{i_1}, \ldots, x_{i_r}).$$

The output of the problem is the following sum, called the partition function:

$$Z(I) = \sum_{\mathbf{x} \in D^n} F_I(\mathbf{x}).$$

Many well-studied counting problems can be formulated as a #CSP. For example, if $D = \{1, 2\}$ and $\mathcal{F}$ consists of a single binary function with $F(1, 1) = F(1, 2) = F(2, 1) = 1$ and $F(2, 2) = 0$, then #CSP($\mathcal{F}$) is exactly the counting version of the vertex cover problem. When $\mathcal{F}$ consists of a single binary function over $D = \{1, \ldots, d\}$, where $F(i, j) = 1$ if $i \neq j$ and 0 otherwise, #CSP($\mathcal{F}$) is the counting version of the d-coloring problem. In this article, we study #CSP under the most general setting with complex and algebraic weights. In the presence of complex numbers, cancellations

may, in fact, be the source of surprisingly efficient algorithms for computing $Z(I)$ for certain constraint languages $\mathcal{F}$. For example, viewing $D = \{0, 1, \ldots, d-1\}$ as $\mathbb{Z}_d$, if $\mathcal{F} = \{F_1, \ldots, F_h\}$ and each F_j is an r_j-ary function of the form[1]

$$F_j(x_1, \ldots, x_{r_j}) = e^{\frac{2\pi \mathrm{i}}{d} \cdot f_j(x_1, \ldots, x_{r_j})},$$

with f_j being a quadratic polynomial over $\mathbb{Z}_d$, then Cai, Chen, Lipton, and Lu in [2010] showed that #CSP($\mathcal{F}$) is solvable in polynomial time.

Various subclasses of #CSP have been studied intensively:

The partition function of graph homomorphisms to a fixed graph: This is the special case when the language $\mathcal{F}$ has a single symmetric binary function. A series of dichotomies of increasing generality has been discovered, starting with $\{0, 1\}$-valued functions by Dyer and Greenhill [2000], nonnegative functions by Bulatov and Grohe [2005], real-valued functions by Goldberg et al. [2010], and complex-valued functions by Cai et al. [2013].

The partition function of directed graph homomorphisms to a fixed graph: This is the special case when $\mathcal{F}$ has a single not necessarily symmetric binary function. Dyer et al. [2007] gave a dichotomy theorem for $\{0, 1\}$-valued functions that induce an acyclic graph when viewed as the adjacency matrix of a directed graph. Then, Cai and Chen [2010] gave a dichotomy for all nonnegative binary functions. Before the present work, the case of directed graph homomorphisms with a single real-valued (but not necessarily symmetric) binary function remained open.

Unweighted #CSP: This is the special case when every function in $\mathcal{F}$ is $\{0, 1\}$-valued. Creignou and Hermann [1996] first obtained a dichotomy for unweighted #CSP with a two-element domain, the counting analog of Schaefer's theorem. Bulatov [2008] made a breakthrough and gave a dichotomy theorem for all unweighted #CSP (also see Bulatov [2013]). Later Dyer and Richerby [2013] gave an alternative proof of Bulatov's theorem and also proved the decidability of the dichotomy criterion. It was extended to #CSP with nonnegative and rational weights by Bulatov et al. [2012], and then to #CSP with nonnegative weights by Cai et al. [2016].

In this article, we prove a general dichotomy theorem for #CSP with algebraic complex weights, whereby all these previous dichotomies are special cases:

THEOREM 1.1. *Given any constraint language $\mathcal{F}$ with algebraic complex weights, the problem* #CSP($\mathcal{F}$) *defined by $\mathcal{F}$ is either solvable in polynomial time or #P-hard.*

The restriction to algebraic complex numbers in the theorem statement is primarily due to considerations of models of computation; see Section 2.2 for a brief discussion.

To prove our main theorem, we introduce three conditions on a given language $\mathcal{F}$: the Block Orthogonality condition, the Type Partition condition, and the Mal'tsev condition. We then show that

(1) #CSP($\mathcal{F}$) is #P-hard if $\mathcal{F}$ violates any of these three conditions; and
(2) #CSP($\mathcal{F}$) can be solved in polynomial-time when $\mathcal{F}$ satisfies all three conditions.

For example, viewing $D = \{0, 1, \ldots, d-1\}$ as $\mathbb{Z}_d$ for some prime power d, let

$$F(x_1, x_2, x_3) = e^{\frac{2\pi \mathrm{i}}{d} \cdot x_1 x_2 x_3}. \tag{1}$$

[1]We use i to denote the imaginary unit with $\mathrm{i}^2 = -1$, and use i to denote an integer index.

One can check that $\mathcal{F} = \{F\}$ violates the Block Orthogonality condition (indeed, $F^{[2]}$ as defined in Equation (2) violates the condition). Thus, the proof of Theorem 1.1 implies that #CSP($\mathcal{F}$) is #P-hard.

Even at 38 pages, the proof of this dichotomy theorem is significantly shorter than one might expect for general #CSP with complex, algebraic weights. For example, even for the special case of counting graph homomorphisms (i.e., #CSP($\mathcal{F}$) when $\mathcal{F}$ consists of a single binary symmetric function), it takes 66 pages and 106 pages, respectively, to prove a dichotomy theorem for real [Goldberg et al. 2010] and complex weights [Cai et al. 2013]). The reason is because the tractability criterion of our dichotomy is much less explicit compared to those of Goldberg et al. [2010] and Cai et al. [2013]. While one can follow the proofs of Goldberg et al. [2010] and Cai et al. [2013] to check in polynomial time whether a given problem is in P or #P-hard, checking whether a general language $\mathcal{F}$ satisfies each of the three conditions listed above requires one to verify a condition on an infinitary object defined from $\mathcal{F}$ (see details in Section 3). At this time, it remains an open problem as whether the tractability criterion of our dichotomy is decidable.[2]

Proof Sketch

The main idea starts with the following approach for solving #CSP($\mathcal{F}$). Let I be an instance of #CSP($\mathcal{F}$) and F be the n-ary function it defines. For each $t \in [n]$, we use $F^{[t]}$ to denote the following t-ary function:

$$F^{[t]}(x_1, \ldots, x_t) = \sum_{x_{t+1}, \ldots, x_n \in D} F(x_1, \ldots, x_t, x_{t+1}, \ldots, x_n). \tag{2}$$

For the discussion below, it is more convenient to view the function $F^{[t]}$ as a $d^{t-1} \times d$ matrix when $t \geq 2$: the rows are indexed by $\mathbf{x} = (x_1, \ldots, x_{t-1}) \in D^{t-1}$; the columns are indexed by $i \in D$; the $(\mathbf{x}, i)$th entry of the matrix is $F^{[t]}(\mathbf{x}, i)$. We let $F^{[t]}(\mathbf{x}, *)$ denote the d-dimensional row vector indexed by $\mathbf{x} \in D^{t-1}$:

$$F^{[t]}(\mathbf{x}, *) = \big(F^{[t]}(\mathbf{x}, 1), \ldots, F^{[t]}(\mathbf{x}, d)\big).$$

To compute $Z(I)$, for now assume that we have access to the following *oracle*. We can send any $\mathbf{x} \in D^{t-1}$, $t \geq 2$, to the oracle, and it returns a d-dimensional vector $\mathbf{v}$ that is linearly dependent with $F^{[t]}(\mathbf{x}, *)$. Here, either $\mathbf{v} = \mathbf{0}$ if $F^{[t]}(\mathbf{x}, *) = \mathbf{0}$, or $\mathbf{v}$ has its first non-zero entry normalized to 1 so $\mathbf{v}$ is uniquely defined.

Temporarily suspending disbelief that such a helpful oracle might exist, we show that $Z(I)$ can be computed efficiently given access to this oracle as follows. From $Z(I) = \sum_{a \in D} F^{[1]}(a)$, it suffices to compute $F^{[1]}(a)$ for each $a \in D$. Now pick any $a_1 \in D$ and send it to the oracle. The oracle returns a d-dimensional vector $\mathbf{v}$ that is linearly dependent with $F^{[2]}(a_1, *)$. If $\mathbf{v} = \mathbf{0}$, then we have $F^{[1]}(a_1) = \sum_{b \in D} F^{[2]}(a_1, b) = 0$. Otherwise, let $a_2 \in D$ denote the index of the first non-zero entry of $\mathbf{v}$, with $v_{a_2} = 1$. Then,

$$F^{[1]}(a_1) = \sum_{b \in D} F^{[2]}(a_1, b) = F^{[2]}(a_1, a_2) \cdot \sum_{b \in D} v_b,$$

where the last equation follows from the assumption that $\mathbf{v}$ and $F^{[2]}(a_1, *)$ are linearly dependent. This reduces the computation of $F^{[1]}(a_1)$ to that of $F^{[2]}(a_1, a_2)$.

Next, we send (a_1, a_2) to the oracle. Either the vector $\mathbf{w}$ we receive is $\mathbf{0}$ for which case $F^{[2]}(a_1, a_2) = 0$, or we can use $\mathbf{w}$ to further reduce the computation of $F^{[1]}(a_1)$ to that

[2]This indeed makes it difficult for us to come up with an example of #CSP($\mathcal{F}$) that is tractable according to our dichotomy theorem but is not known to be tractable before our work.

of $F^{[3]}(a_1, a_2, a_3)$, for some appropriate a_3. Repeating this process for $n-1$ rounds, it suffices to compute $F^{[n]}(a_1, a_2, \ldots, a_n)$ for some appropriate $a_2, \ldots, a_n \in D$. This gives an efficient algorithm for computing $F^{[1]}(a_1)$, since $F = F^{[n]}$ can be evaluated efficiently using the input instance I.

As a result, we can solve #CSP($\mathcal{F}$) efficiently using this oracle. It turns out that almost the whole proof of Theorem 1.1 is trying to understand *how* and *when* we can efficiently implement this oracle. Notice that we need to "collect" the following huge amount of information: For each $t \in [n]$, we need to compute a set of pairwise linearly independent (and normalized) d-dimensional vectors $\mathbf{v}^{[t,1]}, \ldots, \mathbf{v}^{[t,s_t]}$, for some $s_t \geq 0$, so every nonzero row vector $F^{[t]}(\mathbf{x}, *)$ is linearly dependent with one of them. Moreover, for each vector $\mathbf{v}^{[t,j]}$, we need to know the set of $\mathbf{x} \in D^{t-1}$, denoted by $S^{[t,j]} \subseteq D^{t-1}$, such that $F^{[t]}(\mathbf{x}, *)$ is nonzero and linearly dependent with $\mathbf{v}^{[t,j]}$. Two difficulties arise. First, note that in general an $m \times d$ matrix may have m pairwise linearly independent row vectors. So, in general, we may need to keep track of exponentially many vectors $\mathbf{v}^{[t,j]}$. Second, for each $\mathbf{v}^{[t,j]}$, the size of $S^{[t,j]}$ can in general be exponential in t.

To overcome the first difficulty, we drew inspiration from the recent dichotomy theorems for counting graph homomorphisms with real [Goldberg et al. 2010] and complex weights [Cai et al. 2013]. In both dichotomies, the tractable cases are closely related to matrices in which every two row vectors are either linearly dependent or orthogonal, for example, the Hadamard matrices and the so-called discrete unitary matrices [Cai et al. 2013]. This inspires us to introduce the first necessary condition for tractability: the Block Orthogonality condition. It requires that for any F defined by an instance of #CSP($\mathcal{F}$) and for any $t \in [n]$, every two row vectors of $F^{[t]}$ are either linearly dependent or orthogonal; otherwise, we show that #CSP($\mathcal{F}$) is #P-hard. Indeed, a requirement that is more stringent than the orthogonality must hold (as the word "block" suggests); otherwise, we show that the problem is #P-hard. See the formal definition in Section 3.1. Assume that $\mathcal{F}$ satisfies the Block Orthogonality condition. Then, we know for sure that each $F^{[t]}$ has at most d pairwise linearly independent (and indeed pairwise orthogonal) row vectors.

To overcome the second difficulty, we need some of the powerful techniques developed for unweighted #CSP [Bulatov 2013; Dyer and Richerby 2013]. One of the tools used there is the notion of Mal'tsev polymorphism from Universal Algebra (see Section 2.8). For any set $\Phi \subseteq D^n$ that has a Mal'tsev polymorphism φ. Dyer and Richerby [2013] introduce a succinct representation called a *witness function*, which is of *linear size* in n, the arity of Φ, and essentially contains all the information about Φ. In particular, with a witness function one can decide whether a given tuple $\mathbf{x} \in D^n$ belongs to Φ efficiently. From here, it is only natural to ask whether the sets $S^{[t,j]}$ associated with each $\mathbf{v}^{[t,j]}$ have a Mal'tsev polymorphism. This is where we introduce the second necessary condition, which we call simply the Mal'tsev condition. Roughly speaking, it requires all the sets $S^{[t,j]} \subseteq D^{t-1}$, defined from all F, t, and j, to share a common Mal'tsev polymorphism φ; otherwise, we prove that the problem #CSP($\mathcal{F}$) is #P-hard.

Assume that $\mathcal{F}$ satisfies both the Block Orthogonality condition and Mal'tsev condition; otherwise, we already know that #CSP($\mathcal{F}$) is #P-hard. We can now refine the plan of implementing the oracle as follows. Given an input instance I of #CSP($\mathcal{F}$) that defines an n-ary function F, we compute for each $t : 2 \leq t \leq n$,

(a) A set of (at most d) pairwise orthogonal and normalized d-dimensional vectors $\mathbf{v}^{[t,1]}, \ldots, \mathbf{v}^{[t,s_t]}$, for some $s_t : 0 \leq s_t \leq d$, such that every nonzero row $F^{[t]}(\mathbf{x}, *)$ is linearly dependent with one of them.
(b) A witness function $\omega^{[t,j]}$ for each set $S^{[t,j]} \subseteq D^{t-1}$, which can be used to decide membership efficiently.

So the algorithmic problem left is, how and when can we compute the objects in (a) and (b) efficiently?

To this end, we start with $t = n$ and $F = F^{[n]}$. First, by using the Mal'tsev condition and an elegant algorithm from Dyer and Richerby [2013], we can construct efficiently a witness function ω for $R \subseteq D^n$ where $\mathbf{x} \in R$ if and only if $F(\mathbf{x}) \neq 0$. Given ω, it is also easy to construct a witness function ω' for $R' \subseteq D^{n-1}$, the projection of R on its first $n-1$ coordinates. We are getting closer, since according to the definition of $S^{[n,j]}$, R' is exactly the union of the s_n pairwise disjoint sets $S^{[n,1]}, \ldots, S^{[n,s_n]} \subseteq D^{n-1}$. It turns out that a key algorithmic step we need is the following operation over witness functions:

> **Splitting:** Let $\Phi \subseteq D^n$ be a nonempty set, and let $\Psi_1, \ldots, \Psi_s$ be an s-way partition of Φ, for some $s \in [d]$: The Ψ_i's are nonempty, pairwise disjoint, and satisfy $\Phi = \Psi_1 \cup \cdots \cup \Psi_s$. Assume that we are given φ, a Mal'tsev polymorphism of Φ as well as all the Ψ_i's. At the beginning, we have absolutely no information about the Ψ_i's, not even the number s of the Ψ_i's, though we do know that $s \in [d]$. The only resources we have are a witness function ω for Φ and a *black box* to query: We can send any $\mathbf{x} \in \Phi$ to the black box and it returns the unique index $k \in [s]$, such that $\mathbf{x} \in \Psi_k$. The question is: Can we use ω and the black box to compute the value of s as well as a witness function ω_k for each Ψ_k in polynomial time and using polynomially many queries?

In general, it is not clear how to implement the splitting operation efficiently. However, if the sets Φ and $\Psi_1, \ldots, \Psi_s$ (as well as their permutations, see Lemma 7.3) satisfy the so-called partition condition (see Definition 2.26), then we give an algorithm that computes $s \in [d]$ and a witness function ω_k for each Ψ_k in polynomial time and using polynomially many queries. This brings us to the third and last condition: the Type Partition condition. It turns out that this condition is necessary for tractability as well: #CSP($\mathcal{F}$) must be #P-hard if $\mathcal{F}$ violates it. Roughly speaking, the Type Partition condition requires that whenever we need to apply the splitting operation, the sets Φ and $\Psi_1, \ldots, \Psi_s$ (and their permutations) must satisfy the partition condition so our algorithm applies. In particular, it allows us to apply the splitting operation on R' and $S^{[n,1]}, \ldots, S^{[n,s_n]}$ to (1) compute the value of s_n, and (2) construct a witness function for each $S^{[n,j]}$, $j \in [s_n]$, using ω'. The proof showing that the Type Partition condition is actually necessary for tractability, and the polynomial-time algorithm for the splitting operation assuming the partition condition (Section 5 and Section 7.3) are among the most challenging in the article. Using the splitting operation and the Type Partition condition, we can inductively compute a witness function for each $S^{[t,j]}$ from $t = n$ to 2. This gives an efficient implementation of the oracle and thus, a polynomial-time algorithm for #CSP($\mathcal{F}$) when $\mathcal{F}$ satisfies all three necessary conditions. This finishes the proof of the dichotomy theorem.

2. PRELIMINARIES

2.1. Notation

We use $\mathbb{C}$ to denote the set of algebraic complex numbers throughout the article. Given a positive integer n, we let $[n] = \{1, \ldots, n\}$.

Let $D = [d]$ be a finite set, for some $d \geq 1$. Given an n-ary algebraic complex-valued function $F : D^n \to \mathbb{C}$, we use $\mathsf{Im}(F)$ to denote the *image* of F, that is,

$$\mathsf{Im}(F) = \{c \in \mathbb{C} : c = F(\mathbf{x}) \text{ for some } \mathbf{x} \in D^n\}.$$

Given a finite set $\mathcal{F} = \{F_1, \ldots, F_h\}$ of functions, we use $\mathsf{Im}(\mathcal{F})$ to denote the image of $\mathcal{F}$:

$$\mathsf{Im}(\mathcal{F}) = \mathsf{Im}(F_1) \cup \cdots \cup \mathsf{Im}(F_h).$$

Given $F : D^n \to \mathbb{C}$, we use $|F|$ to denote the real and nonnegative function that maps $\mathbf{x}$ to $|F(\mathbf{x})|$ for all $\mathbf{x} \in D^n$, where $|\cdot|$ in $|F(\mathbf{x})|$ denotes the complex norm. When $n \geq 2$, we sometimes consider F as a matrix with exponentially many rows but only d columns. We use $\mathbf{M}_F$ to denote the following $d^{n-1} \times d$ matrix: its rows and columns are indexed by $\mathbf{x} = (x_1, \ldots, x_{n-1}) \in D^{n-1}$ and $x_n \in D$, respectively; the $(\mathbf{x}, x_n)$th entry of $\mathbf{M}_F$ is

$$M_F(\mathbf{x}, x_n) = F(\mathbf{x}, x_n) = F(x_1, \ldots, x_{n-1}, x_n).$$

We use $F(\mathbf{x}, *)$, where $\mathbf{x} \in D^{n-1}$, to denote the d-dimensional vector whose ith entry is $F(\mathbf{x}, i)$ and $|F(\mathbf{x}, *)|$ to denote the d-dimensional real and non-negative vector whose ith entry is $|F(\mathbf{x}, i)|$. Given a matrix $\mathbf{M}$ we use $\mathbf{M}(i, *)$ to denote its ith row vector. We also write $|\mathbf{M}|$ to denote the matrix of the same size as $\mathbf{M}$, with its (i, j)th entry being $|M(i, j)|$, the complex norm of $M(i, j)$. (Determinant is never used in this article, so the notation should be clear from the context.)

Two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{C}^d$ are said to be orthogonal if they satisfy $\sum_{i \in [d]} x_i \overline{y_i} = 0$, where $\overline{y_i}$ denotes the conjugate of y_i.

Given $\mathbf{x} \in D^n$ and $\ell \in [n]$, we use $\mathsf{Pr}_{[\ell]}\mathbf{x}$ to denote its prefix of length ℓ. Let $\Phi \subseteq D^n$ be an n-ary relation. For each $\ell \in [n]$, we use $\mathsf{Pr}_\ell \Phi \subseteq D$ to denote the *projection* of Φ on the ℓth coordinate: $a \in \mathsf{Pr}_\ell \Phi$ if and only if there is an $\mathbf{x} \in \Phi$ such that $x_\ell = a$. We call $\mathbf{x}$ a *witness* for a at the ℓth coordinate, or simply a witness for the pair (ℓ, a). We also use $\mathsf{Pr}_{[\ell]}\Phi \subseteq D^\ell$ to denote the *projection* of Φ on the first ℓ coordinates: $\mathbf{y} \in \mathsf{Pr}_{[\ell]}\Phi$ if and only if there exists an $\mathbf{x} \in \Phi$ such that $\mathbf{y} = \mathsf{Pr}_{[\ell]}\mathbf{x}$.

Given $\mathbf{a} \in D^\ell$ for some $\ell \in [n]$, we use $\Phi(\mathbf{a}, *) = \Phi(a_1, \ldots, a_\ell, *)$ to denote the relation on $n - \ell$ variables with the first ℓ variables fixed to $\mathbf{a}$: $\mathbf{y} \in \Phi(\mathbf{a}, *)$ iff $\mathbf{a} \circ \mathbf{y} \in \Phi$, where $\mathbf{a} \circ \mathbf{y}$ denotes the concatenation of $\mathbf{a}$ and $\mathbf{y}$.

Given a permutation π of $[n]$, let $\pi(\Phi)$ be the n-ary relation such that $\mathbf{x} \in \pi(\Phi)$ iff

$$\left(x_{\pi(1)}, \ldots, x_{\pi(n)}\right) \in \Phi.$$

Finally, we use $\leq_T$ to denote polynomial-time Turing reductions between problems, and $\equiv_T$ to denote equivalence under polynomial-time Turing reductions.

2.2. Counting CSP with Algebraic Weights

Let $D = [d]$ be a domain, and let $\mathcal{F} = \{F_1, \ldots, F_h\}$ be a finite set of algebraic complex-valued functions over D. Recall the definition of $\#\mathsf{CSP}(\mathcal{F})$ in the introduction. When $\mathcal{F} = \{F\}$ has only one function, we denote $\#\mathsf{CSP}(\mathcal{F})$ by $\#\mathsf{CSP}(F)$ for convenience. Sometimes we write $\#\mathsf{CSP}(\mathcal{F})$ simply as $\#\mathsf{CSP}(F_1, \ldots, F_h)$ to list the functions explicitly.

To complete the definition of $\#\mathsf{CSP}(\mathcal{F})$ as a computational problem, we need to specify the model of computation for algebraic numbers, that is, how the numbers in $\mathcal{F}$ and the output $Z(I)$ are encoded. We can take any reasonable model, for example, the one used earlier in Lenstra [1992], Thurley [2009], and Cai et al. [2013]. Note that functions in $\mathcal{F}$ are constants when $\mathcal{F}$ is fixed and the complexity of $\#\mathsf{CSP}(\mathcal{F})$ is concerned. The input size only depends on the number of variables, and $|I|$ depends on the the number of tuples in I.

Given D and $\mathcal{F}$, we define the following problem denoted by $\mathsf{COUNT}(\mathcal{F})$: the input is a pair (I, c), where I is an input instance of $\#\mathsf{CSP}(\mathcal{F})$ and c is an algebraic complex number. Let $x_1, \ldots, x_n$ denote the variables over D in I. The output is then the number of $\mathbf{x} = (x_1, \ldots, x_n) \in D^n$ such that $F_I(\mathbf{x}) = c$, where F_I is the function defined by I. It turns out that $\mathsf{COUNT}(\mathcal{F})$ and $\#\mathsf{CSP}(\mathcal{F})$ are equivalent under polynomial-time Turing reductions. The proofs of Lemmas 2.1 and 2.2 use the technique of interpolations, which was first used in this context by Dyer and Greenhill [2000].

LEMMA 2.1. $\mathsf{COUNT}(\mathcal{F}) \equiv_T \#\mathsf{CSP}(\mathcal{F})$.

PROOF. Let $\mathsf{Im}(\mathcal{F}) = \{c_1, \ldots, c_k\}$, where $k = |\mathsf{Im}(\mathcal{F})|$ is a constant for fixed $\mathcal{F}$. Let I be an input instance of #CSP($\mathcal{F}$) over n variables $\mathbf{x} \in D^n$ with $m = |I|$, and let F be the n-ary function that I defines. First, we can compute the following set of numbers in time polynomial in m:

$$C_m = \left\{c_1^{\ell_1} \cdots c_k^{\ell_k} : \ell_1, \ldots, \ell_k \text{ are non-negative integers and } \ell_1 + \cdots + \ell_k = m\right\}, \quad (3)$$

since k is a constant. It follows from the definition of F that $F(\mathbf{x}) \in C_m$ for all $\mathbf{x} \in D^n$.

For each $c \in C_m$, we let N_c denote the number of $\mathbf{x} \in D^n$ such that $F(\mathbf{x}) = c$. Then,

$$Z(I) = \sum_{\mathbf{x} \in D^n} F(\mathbf{x}) = \sum_{c \in C_m} c \cdot N_c.$$

This immediately gives us a polynomial-time reduction from #CSP($\mathcal{F}$) to COUNT($\mathcal{F}$).

We prove the other direction: Given any I, we use a subroutine for #CSP($\mathcal{F}$) to compute N_c for all $c \in C_m$. For this purpose, we let $C'_m = C_m - \{0\}$ and let $s = |C'_m|$, which is polynomial in m. We build from I the following instances $I_1, \ldots, I_s$: to get I_ℓ, $\ell \in [s]$, we make ℓ copies of each tuple in I (and thus, $I_1 = I$ and $|I_\ell| = \ell \cdot |I|$). We also let F_ℓ denote the n-ary function defined by I_ℓ.

By the construction of I_ℓ, it is easy to see that $F_\ell(\mathbf{x}) = (F(\mathbf{x}))^\ell$ for all $\mathbf{x} \in D^n$. Thus,

$$Z(I_\ell) = \sum_{\mathbf{x} \in D^n} F_\ell(\mathbf{x}) = \sum_{c \in C_m} c^\ell \cdot N_c = \sum_{c \in C'_m} c^\ell \cdot N_c, \quad \text{for each } \ell = 1, \ldots, s.$$

The left-hand side of the equations can be obtained by calling a subroutine for #CSP($\mathcal{F}$) on I_ℓ. We can then solve the Vandermonde system above to get N_c for each $c \in C'_m$. If $0 \in C_m$, then we can also derive N_0 using the fact that the sum of all the N_c's, $c \in C_m$, is d^n. This finishes the proof of the lemma. □

In certain situations the problem COUNT($\mathcal{F}$) is easier to deal with. For example, we can use the connection above to prove the following lemma.

LEMMA 2.2. #CSP($|F_1|, \ldots, |F_h|$) $\leq_T$ #CSP($\mathcal{F}$).

PROOF. It suffices to show COUNT($|F_1|, \ldots, |F_h|$) $\leq_T$ COUNT($\mathcal{F}$) by Lemma 2.1. We let $\mathsf{Im}(\mathcal{F}) = \{c_1, \ldots, c_k\}$ where k is a constant for fixed $\mathcal{F}$.

Let I be an instance of #CSP($|F_1|, \ldots, |F_h|$) and F be the n-ary non-negative function it defines. Let a be a non-negative number, and we are asked to compute the number of $\mathbf{x} \in D^n$ such that $F(\mathbf{x}) = a$.

From I it is natural to construct an input instance I' of #CSP($\mathcal{F}$) by simply replacing the function $|F_i|$ in each tuple of I with its corresponding function F_i in $\mathcal{F}$. Let F' denote the function that I' defines. Then it is clear that $F(\mathbf{x}) = |F'(\mathbf{x})|$ for all $\mathbf{x} \in D^n$.

Let $m = |I| = |I'|$. Then, we can compute C_m as defined in Equation (3) in time polynomial in m, because k is a constant.

From the definitions of C_m and F', we have $F'(\mathbf{x}) \in C_m$ for all $\mathbf{x} \in D^n$. As a result,

$$[\text{number of } \mathbf{x} \text{ such that } F(\mathbf{x}) = a] = \sum_{c \in C_m:\, |c| = a} [\text{number of } \mathbf{x} \text{ such that } F'(\mathbf{x}) = c],$$

and the right-hand side can be computed efficiently, because the number of such c can be no more than $|C_m|$ and the term for each c can be evaluated by calling a subroutine for COUNT($\mathcal{F}$). This finishes the proof of the lemma. □

2.3. Row Representation

Let $\mathbf{M}$ be an $m \times n$ complex matrix. It induces the following equivalence relation $\sim_{\mathbf{M}}$ over $\{\ell \in [m] : \mathbf{M}(\ell, *) \neq \mathbf{0}\}$, that is, the set of nonzero rows of $\mathbf{M}$:

$$\ell \sim_{\mathbf{M}} \ell' \iff \mathbf{M}(\ell, *) \text{ and } \mathbf{M}(\ell', *) \text{ are linearly dependent over } \mathbb{C}.$$

We say $\mathcal{S} = \{(S_1, \mathbf{v}_1), \ldots, (S_k, \mathbf{v}_k)\}$, for some $k \geq 0$, is the *row representation* of $\mathbf{M}$ if

(1) $S_1, \ldots, S_k \subseteq [m]$ are the equivalence classes of the equivalence relation $\sim_{\mathbf{M}}$; and
(2) For each $i \in [k]$, $\mathbf{v}_i$ is a nonzero n-dimensional vector with its first nonzero entry being 1, and is linearly dependent with $\mathbf{M}(\ell, *)$, for all $\ell \in S_i$. (By the definition of $\sim_{\mathbf{M}}$, $\mathbf{v}_i$ exists and is unique.)

We will refer to $\mathbf{v}_i$ as the *representative row vector* for the equivalence class S_i.

For example, the row representation of the matrix

$$\mathbf{M} = \begin{pmatrix} 1 & 1 & & \\ 1 & 1 & & \\ & & 1 & 2 \\ & & 3 & 6 \\ 3 & 1 & & \\ 9 & 3 & & \end{pmatrix} \tag{4}$$

has $S_1 = \{1, 2\}$, $S_2 = \{3, 4\}$, $S_3 = \{5, 6\}$, and

$$\mathbf{v}_1 = (1, 0, 1, 0, 0, 0), \quad \mathbf{v}_2 = (0, 0, 0, 0, 1, 2), \quad \text{and} \quad \mathbf{v}_3 = (0, 1, 0, 1/3, 0, 0).$$

From the definition, we have S_i is nonempty for all i; the S_i's are pairwise disjoint;

$$S_1 \cup \cdots \cup S_k = \{\ell \in [m] : \mathbf{M}(\ell, *) \neq \mathbf{0}\};$$

for all $i \neq j$, $\mathbf{v}_i$ and $\mathbf{v}_j$ are linearly independent. Clearly, every matrix has a unique row representation.

In general, the row representation $\mathcal{S}$ of an $m \times n$ matrix $\mathbf{M}$ may consist of as many as m pairs. But if it is known that every two rows of $\mathbf{M}$ are either linearly dependent or orthogonal, then the number of pairs in its row representation cannot exceed n, the number of its columns.

We say that a real, *non-negative* matrix $\mathbf{M}$ is *block-rank*-1, if its row representation $\mathcal{S} = \{(S_1, \mathbf{v}_1), \ldots, (S_k, \mathbf{v}_k)\}$ has the property that for all $i \neq j \in [k]$, $\mathbf{v}_i$ and $\mathbf{v}_j$ share no common positive entry, that is, there exists no index $t \in [n]$ such that the tth entries of $\mathbf{v}_i$ and $\mathbf{v}_j$ are both positive.

For example, the 6×6 non-negative matrix in Equation (4) is block-rank-1. Given a block-rank-1 matrix, one can permute its rows and columns (with two different permutations in general) to get a block-diagonal matrix, where each of its blocks is of rank 1.

2.4. The Block-Rank-1 Condition

We also extend the notion of row representations to functions. Given $F : D^n \to \mathbb{C}$ with $n \geq 2$, we have the following equivalence relation $\sim_F$ over $\{\mathbf{x} \in D^{n-1} : F(\mathbf{x}, *) \neq \mathbf{0}\}$:

$$\mathbf{x} \sim_F \mathbf{y} \iff F(\mathbf{x}, *) \text{ and } F(\mathbf{y}, *) \text{ are linearly dependent over } \mathbb{C}.$$

Similarly, we say that $\mathcal{S} = \{(S_1, \mathbf{v}_1), \ldots, (S_k, \mathbf{v}_k)\}$, where $S_i \subseteq D^{n-1}$ for each $i \in [k]$, is the row representation of F if $\mathcal{S}$ is the row representation of the $d^{n-1} \times d$ matrix $\mathbf{M}_F$. (Explicitly, we have $S_1, \ldots, S_k$ are the equivalence classes of $\sim_F$, and for each $i \in [k]$, $\mathbf{v}_i$ is a non-zero d-dimensional vector with its first non-zero entry being 1, and is linearly dependent with $F(\mathbf{x}, *)$, $\mathbf{x} \in S_i$.)

Finally, we call F a *block-rank-1 function* if the real, non-negative matrix $\mathbf{M}_{|F|}$ is a block-rank-1 matrix. Note that the notion of F being a block-rank-1 function is defined in terms of the real and nonnegative function $|F|$. (Explicitly, for all $\mathbf{x}, \mathbf{y} \in D^{n-1}$ with $F(\mathbf{x}, *)$ and $F(\mathbf{y}, *)$ being nonzero, the two non-negative vectors $|F(\mathbf{x}, *)|$ and $|F(\mathbf{y}, *)|$ either are linearly dependent or share no common positive entry.)

We introduce block-rank-1 matrices/functions to apply the following sweeping dichotomy of Bulatov and Grohe [2005]. Given any symmetric $d \times d$ non-negative matrix $\mathbf{A}$ with algebraic entries, we define a counting graph homomorphism problem $\mathsf{EVAL}(\mathbf{A})$: the input is an undirected graph $G = (V, E)$ with $V = [n]$, and the output is

$$Z_{\mathbf{A}}(G) = \sum_{x_1,\ldots,x_n \in [d]} \left(\prod_{ij \in E} A(x_i, x_j) \right).$$

In the language of #CSP, $\mathsf{EVAL}(\mathbf{A})$ is the same as $\#\mathsf{CSP}(F)$ with $F(i, j) = A(i, j)$.

THEOREM 2.3 ([BULATOV AND GROHE 2005]). *Let* $\mathbf{A}$ *be a symmetric, nonnegative square matrix with algebraic entries. Then* $\mathsf{EVAL}(\mathbf{A})$ *is solvable in polynomial time if* $\mathbf{A}$ *is* block-rank-1*, and is* #P-hard *otherwise.*

We extend the definitions of $\mathsf{EVAL}(\mathbf{A})$ and $Z_{\mathbf{A}}(\cdot)$ to any square (but not necessarily symmetric) matrix $\mathbf{A}$ over $\mathbb{C}$. The input of $\mathsf{EVAL}(\mathbf{A})$ is a *directed* graph $G = (V, E)$, and

$$Z_{\mathbf{A}}(G) = \sum_{x_1,\ldots,x_n \in [d]} \left(\prod_{\overrightarrow{ij} \in E} A(x_i, x_j) \right)$$

is the desired output. The following lemma will be useful later in the proof:

LEMMA 2.4. *Let* $\mathbf{A}$ *be a square (though not necessarily symmetric) matrix with algebraic complex entries. If* $|\mathbf{A}|$ *is not block-rank-*1*, then* $\mathsf{EVAL}(\mathbf{A})$ *is* #P-hard*.*

PROOF. By Lemma 2.2, it suffices to show that $\mathsf{EVAL}(|\mathbf{A}|)$ is #P-hard.

To this end, we use $\mathbf{B}$ to denote the *symmetric* and non-negative $d \times d$ matrix $|\mathbf{A}|\,|\mathbf{A}|^T$, where the (i, j)th entry of $\mathbf{B}$ is

$$B(i, j) = \sum_{k \in [d]} |A(i, k)| \cdot |A(j, k)|.$$

We claim that $\mathsf{EVAL}(\mathbf{B}) \leq_T \mathsf{EVAL}(|\mathbf{A}|)$. This is because, given an undirected graph $G = (V, E)$ of $\mathsf{EVAL}(\mathbf{B})$, we can construct a new directed graph $G' = (V', E')$ with

$$V' = \{x_v, x_e : v \in V \text{ and } e \in E\} \quad \text{and} \quad E' = \{\overrightarrow{x_u x_e}, \overrightarrow{x_v x_e} : e = uv \in E\}.$$

It is easy to verify that $Z_{\mathbf{B}}(G) = Z_{|\mathbf{A}|}(G')$ from which the reduction follows.

On the other hand, if $|\mathbf{A}|$ is not block-rank-1 neither is $\mathbf{B}$. To see this, assume that $|A(i, *)|$ and $|A(j, *)|$ are not linearly dependent but share at least one positive entry. The latter implies that $B(i, j) = B(j, i) > 0$. Given that $B(i, i), B(j, j) > 0$, $B(i, *)$ and $B(j, *)$ share at least two common positive entries. However, given that they are not

linearly dependent, by Cauchy–Schwarz, we have

$$B(i,j)\cdot B(j,i)=\left(\sum_{k\in[d]}|A(i,k)|\cdot|A(j,k)|\right)^2 \tag{5}$$
$$<\left(\sum_{k\in[d]}|A(i,k)|^2\right)\cdot\left(\sum_{k\in[d]}|A(j,k)|^2\right)=B(i,i)\cdot B(j,j).$$

This implies that $B(i,*)$, $B(j,*)$ cannot be linearly dependent and, thus, $\mathbf{B}$ is not block-rank-1. It follows from Theorem 2.3 that $\mathsf{EVAL}(\mathbf{B})$ is #P-hard, and so is $\mathsf{EVAL}(|\mathbf{A}|)$. This finishes the proof of the lemma. □

Next, we use Theorem 2.3 to prove a useful #P-hardness lemma for $\#\mathsf{CSP}(F)$ with a single algebraic complex-valued function F. The idea is similar to the proof of Lemma 2.4 above. Let $D=[d]$ be a domain.

LEMMA 2.5 (THE BLOCK-RANK-1 CONDITION). *Let $F:D^r\to\mathbb{C}$ be an algebraic complex-valued function. If $r\geq 2$ but F is not a block-rank-1, then $\#\mathsf{CSP}(F)$ is* #P-hard.

PROOF. By Lemma 2.2 it suffices to show that $\#\mathsf{CSP}(|F|)$ is #P-hard.

To this end, we construct a symmetric and non-negative matrix $\mathbf{A}$ from $|F|$, such that

$$\mathsf{EVAL}(\mathbf{A})\leq_T\#\mathsf{CSP}(|F|), \tag{6}$$

and then use Theorem 2.3 to show that $\mathsf{EVAL}(\mathbf{A})$ is #P-hard.

The rows and columns of the matrix $\mathbf{A}$ are indexed by $\mathbf{x}\in D^{r-1}$, and its entries are

$$A(\mathbf{x},\mathbf{y})=\sum_{i\in D}|F(\mathbf{x},i)|\cdot|F(\mathbf{y},i)|.$$

It is clear that $\mathbf{A}$ is both symmetric and non-negative.

Next, given an undirected graph $G=(V,E)$ as an instance of $\mathsf{EVAL}(\mathbf{A})$, we construct the following instance I of $\#\mathsf{CSP}(|F|)$. It has $(r-1)|V|+|E|$ variables

$$x_{v,1},\ldots,x_{v,r-1},y_e,\quad \text{for each } v\in V \text{ and } e\in E.$$

For each edge $e=uv\in E$, we add the following two tuples to I:

$$(|F|,x_{u,1},\ldots,x_{u,r-1},y_e)\quad\text{and}\quad(|F|,x_{v,1},\ldots,x_{v,r-1},y_e).$$

From the construction of I and the definition of $\mathbf{A}$ from $|F|$, it is easy to check that

$$Z_{\mathbf{A}}(G)=Z(I),\quad \text{where } F_I \text{ is the function that } I \text{ defines.}$$

This gives us a polynomial-time reduction from $\mathsf{EVAL}(\mathbf{A})$ to $\#\mathsf{CSP}(|F|)$.

Finally, we show that if F is not block-rank-1, then $\mathbf{A}$ is not a block-rank-1 matrix, and by Theorem 2.3, $\mathsf{EVAL}(\mathbf{A})$ is #P-hard. As F is not block-rank-1, we know there are two vectors $\mathbf{x},\mathbf{y}\in D^{r-1}$, such that $|F(\mathbf{x},*)|$ and $|F(\mathbf{y},*)|$ are not linearly dependent but share at least one common positive entry. This implies that all the following four entries are positive: $A(\mathbf{x},\mathbf{x})$, $A(\mathbf{x},\mathbf{y})=A(\mathbf{y},\mathbf{x})$, $A(\mathbf{y},\mathbf{x})>0$. By Cauchy–Schwarz (similar to Equation (5) in the proof of Lemma 2.4), we have $A(\mathbf{x},\mathbf{y})\cdot A(\mathbf{y},\mathbf{x})<A(\mathbf{x},\mathbf{x})\cdot A(\mathbf{y},\mathbf{y})$. Therefore, $A(\mathbf{x},*)$ and $A(\mathbf{y},*)$ share at least two positive entries but are not linearly dependent. Thus, $\mathbf{A}$ is not block-rank-1 and this finishes the proof of the lemma. □

2.5. Block Orthogonality

Orthogonality played an important role in previous work on counting graph homomorphisms with real [Goldberg et al. 2010] and complex weights [Cai et al. 2013]. Here, we generalize it and introduce the notion of block orthogonality.

Let $\mathbf{x}, \mathbf{y} \in \mathbb{C}^d$ be two nonzero d-dimensional vectors and $\mathbf{x}', \mathbf{y}'$ be two real, nonnegative vectors with $x'_i = |x_i|$ and $y'_i = |y_i|$ for all $i \in [d]$. Assume that $\mathbf{x}'$ and $\mathbf{y}'$ are linearly dependent. As a result, these four vectors are nonzero at the same indices and we use $T \subseteq [d]$ to denote the set of such indices. Let $\{\mu_1, \ldots, \mu_\ell\} = \{x'_i : i \in T\}$, for some $\ell \geq 1$, such that $\mu_1 > \cdots > \mu_\ell > 0$. This further partitions T into $T_1, \ldots, T_\ell$, where $x'_i = \mu_k$ for all $i \in T_k$ and $k \in [\ell]$. It is also clear that $\mathbf{y}'$ would yield the same partition, because it is linearly dependent with $\mathbf{x}'$.

Now, we say $\mathbf{x}$ and $\mathbf{y}$ are *block-orthogonal* if for every $k \in [\ell]$ we have

$$\sum_{i \in T_k} x_i \cdot \overline{y_i} = 0. \tag{7}$$

By definition, we have that $\mathbf{x}$ and $\mathbf{y}$ are *orthogonal* if they are block-orthogonal:

$$\sum_{i \in [d]} x_i \cdot \overline{y_i} = \sum_{i \in T} x_i \cdot \overline{y_i} = \sum_{k \in [\ell]} \sum_{i \in T_k} x_i \cdot \overline{y_i} = 0.$$

On the other hand, two orthogonal vectors $\mathbf{x}$ and $\mathbf{y}$ are not block-orthogonal in general even when $\mathbf{x}'$ and $\mathbf{y}'$ are linearly dependent. For example, the vectors

$$\mathbf{x} = (2, 1, 1, 1, 1) \quad \text{and} \quad \mathbf{y} = (2, -1, -1, -1, -1)$$

are orthogonal and satisfy $\mathbf{x}' = \mathbf{y}'$, but they are not block-orthogonal: we have $T_1 = \{1\}$ and $T_2 = \{2, 3, 4, 5\}$, but $\sum_{i \in T_1} x_i \cdot \overline{y_i} = 4$.

The following property holds if two vectors are block-orthogonal:

LEMMA 2.6. *If* $\mathbf{x}$ *and* $\mathbf{y}$ *are block-orthogonal and for some integer* $K \geq 1$*, the* K*th power of all nonzero entries of these two vectors are real and positive, then we have*

$$\sum_{i \in D} {x_i}^{sK+1} \cdot y_i^{rK-1} = 0, \quad \textit{for any integers } s \geq 0 \textit{ and } r \geq 1.$$

PROOF. We use the same notation as in the definition of block orthogonality above. For each $i \in T$, let $z_i = x_i/|x_i|$ and $w_i = y_i/|y_i|$. By the assumption, both z_i and w_i are roots of unity whose orders divide K. Since $\mathbf{x}'$ and $\mathbf{y}'$ are linearly dependent, there are $\nu_1 > \cdots > \nu_\ell > 0$ such that $|y_i| = \nu_k$ for all $i \in T_k$ and $k \in [\ell]$. Now, we can rewrite Equation (7) as

$$0 = \sum_{i \in T_k} x_i \cdot \overline{y_i} = \mu_k \cdot \nu_k \sum_{i \in T_k} z_i \cdot \overline{w_i}.$$

Then the lemma follows from

$$\begin{aligned} \sum_{i \in D} x_i^{sK+1} \cdot y_i^{rK-1} = \sum_{k \in [\ell]} \sum_{i \in T_k} x_i^{sK+1} \cdot y_i^{rK-1} &= \sum_{k \in [\ell]} \mu_k^{sK+1} \cdot \nu_k^{rK-1} \sum_{i \in T_k} z_i^{sK+1} \cdot w_i^{rK-1} \\ &= \sum_{k \in [\ell]} \mu_k^{sK+1} \cdot \nu_k^{rK-1} \sum_{i \in T_k} z_i \cdot \overline{w_i} = 0. \end{aligned}$$

The second to the last equation uses the fact that z_i, w_i are roots of unity whose orders divide K. This finishes the proof of the lemma. □

We are now ready to define *block-orthogonal functions*:

Definition 2.7 (*Block-Orthogonal Function*). Let $F : D^n \to \mathbb{C}$ be a block-rank-1 function with $n \geq 2$. We call it a *block-orthogonal* function if for all $\mathbf{x}, \mathbf{y} \in D^{n-1}$ such that $F(\mathbf{x}, *), F(\mathbf{y}, *) \neq \mathbf{0}$ and $\mathbf{x} \sim_{|F|} \mathbf{y}$, the two vectors $F(\mathbf{x}, *)$ and $F(\mathbf{y}, *)$ are *either linearly dependent or block-orthogonal*.

2.6. Unweighted Counting CSP

We need the following connection between weighted and *unweighted* #CSP. The latter is the special case when all the functions in $\mathcal{F}$ take values in $\{0, 1\}$, for which we adopt the following notation. Let $D = [d]$ be a domain. An unweighted constraint language Γ over domain D is a finite set of relations $\{\Phi_1, \ldots, \Phi_h\}$ in which each Φ_i is an r_i-ary relation over D^{r_i}, for some $r_i \geq 1$. D and Γ define the following problem, denoted by #CSP(Γ). Let $\mathbf{x} = (x_1, \ldots, x_n) \in D^n$ be a set of n variables over D. The input is a finite set I of tuples $(\Phi, i_1, \ldots, i_r)$ in which Φ is an r-ary relation in Γ and $i_1, \ldots, i_r \in [n]$. The input I defines the following relation R_I over D^n:

$$\mathbf{x} \in R_I \iff (x_{i_1}, \ldots, x_{i_r}) \in \Phi \text{ for every tuple } (\Phi, i_1, \ldots, i_r) \in I.$$

Given I, the output of the problem is $|R_I|$.

Given $F : D^n \to \mathbb{C}$, let $\Phi_F = \mathsf{Boolean}(F)$ denote the relation over n variables where

$$\mathbf{x} \in \Phi_F \iff F(\mathbf{x}) \neq 0, \qquad \text{for all } \mathbf{x} \in D^n.$$

The following lemma is a corollary of Lemma 2.1:

LEMMA 2.8. *Given a set $\mathcal{F} = \{F_1, \ldots, F_h\}$ of algebraic complex-valued functions,*

$$\#\mathsf{CSP}(\Gamma) \leq_{\mathsf{T}} \#\mathsf{CSP}(\mathcal{F}),$$

where $\Gamma = \{\Phi_1, \ldots, \Phi_h\}$ and $\Phi_i = \mathsf{Boolean}(F_i)$ for each $i \in [h]$.

PROOF. By Lemma 2.1, it suffices to show that #CSP(Γ) $\leq_T$ COUNT($\mathcal{F}$).

Let I be an input instance of #CSP(Γ) over n variables, and let R be the relation it defines. We then construct an instance I' of #CSP($\mathcal{F}$) in polynomial time, by replacing the relation Φ_i in each tuple of I with its corresponding function $F_i \in \mathcal{F}$. Let F denote the function that I' defines. Then, we have $\mathbf{x} \in R$ if and only if $F(\mathbf{x}) \neq 0$, for all $\mathbf{x} \in D^n$. Thus, $|R| = d^n - |\{\mathbf{x} \in D^n : F(\mathbf{x}) = 0\}|$. The right-hand side can be obtained by calling a subroutine for COUNT($\mathcal{F}$). □

2.7. The Purification Lemma

As it will become clear later, it is much easier to work with functions that take complex values with rational arguments (i.e., arguments that are rational multiples of π). We need the following definition:

Definition 2.9 (*Pure Functions*). We call $F : D^n \to \mathbb{C}$ a *pure* complex function if $F(\mathbf{x})$ is the product of a non-negative rational number and a root of unity, for every $\mathbf{x} \in D^n$. Given a pure function F, we use order(F) to denote the smallest positive integer K such that $(F(\mathbf{x}))^K$ is real and positive for all $\mathbf{x} \in D^n$ with $F(\mathbf{x}) \neq 0$.

A useful tool in proving the hardness part of our dichotomy is the following Purification Lemma (Lemma 2.12). It was introduced in the study of complex graph homomorphisms in Cai et al. [2013] and gives us a connection between pure and general functions (which can take values with irrational arguments). In Sections 4 and 5, we will see two instances where the Purification Lemma is used to extend two hardness lemmas from pure to general functions.

We start with the following definition of *generating sets*:

Definition 2.10. Let $C = \{c_1, \ldots, c_n\}$ be a finite set of nonzero algebraic numbers, for some $n \geq 1$. We say a finite set $\{g_1, \ldots, g_s\}$, for some $s \geq 0$, is a *generating set* of C if

(1) Every g_i is a nonzero algebraic number in $\mathbb{Q}(C)$, that is, the extension of the rational field by adjoining the elements of C;
(2) For all $(k_1, \ldots, k_s) \in \mathbb{Z}^s - \{\mathbf{0}\}$, $g_1^{k_1} \cdots g_s^{k_s}$ is not a root of unity; and
(3) For every $c \in C$, there exists a unique tuple $(k_1, \ldots, k_s) \in \mathbb{Z}^s$, such that

$$\frac{c}{g_1^{k_1} \cdots g_s^{k_s}} \text{ is a root of unity.}$$

Note that $s = 0$ happens if and only if all the c_i's in C are roots of unity.

The following lemma was proved in Cai et al. [2013] (Lemma 7.3):

LEMMA 2.11. *Every finite set C of nonzero algebraic numbers has a generating set.*

We now state and prove the Purification Lemma:

LEMMA 2.12 (PURIFICATION LEMMA). *There is a mapping* Pure *that, given any finite tuple $(F_1, \ldots, F_h)$ of algebraic complex-valued functions together with a generating set $\{g_1, \ldots, g_s\}$ of $\mathsf{Im}(F_1, \ldots, F_h) - \{0\}$, produces a tuple of pure functions*

$$\big(F_1', \ldots, F_h'\big) = \mathsf{Pure}\big(F_1, \ldots, F_h; \{g_1, \ldots, g_s\}\big) \tag{8}$$

in which each F_i' has the same arity $r_i \geq 1$ as F_i, such that

(1) $\#\mathsf{CSP}(F_1', \ldots, F_h') \equiv_T \#\mathsf{CSP}(F_1, \ldots, F_h)$*;*
(2) For every $i \in [h]$, we have $\mathsf{Boolean}(F_i') = \mathsf{Boolean}(F_i)$*;*
(3) For every $i \in [h]$ with $r_i \geq 2$, if F_i' is block-rank-1 then F_i is block-rank-1; and
(4) If F_i' is block-rank-1, then for any $\mathbf{x}, \mathbf{y} \in D^{n-1}$ such that $F_i'(\mathbf{x}, *)$ and $F_i'(\mathbf{y}, *)$ share at least one common nonzero entry, we have
  (a) $F_i'(\mathbf{x}, *)$ and $F_i'(\mathbf{y}, *)$ are linearly dependent if and only if $F_i(\mathbf{x}, *)$ and $F_i(\mathbf{y}, *)$ are linearly dependent;
  (b) If $F_i'(\mathbf{x}, *)$ and $F_i'(\mathbf{y}, *)$ are block-orthogonal, then $F_i(\mathbf{x}, *)$ and $F_i(\mathbf{y}, *)$ are also block-orthogonal.

PROOF. We start with some intuition. Assume for now that $h = 1$ and

$$F_1 = \begin{pmatrix} g_1^2 g_2^4 e_1 & g_2^5 e_2 \\ g_1^3 g_3^2 e_3 & g_4^6 e_4 \end{pmatrix},$$

where e_i's are roots of unity. Then Pure replaces each g_i by the ith smallest prime and

$$F_1' = \begin{pmatrix} 2^2 3^4 e_1 & 3^5 e_2 \\ 2^3 5^2 e_3 & 7^6 e_4 \end{pmatrix},$$

which by definition is a pure function. The equivalence of these two problems follows from Lemma 2.1, and the following property: For any instance I, let F and F' denote the functions it defines using F_1 and F_1', respectively. Then $F(\mathbf{x}) = g_1^{k_1} g_2^{k_2} g_3^{k_3} g_4^{k_4} e$ for some integers k_i and root of unity e if and only if $F'(\mathbf{x}) = 2^{k_1} 3^{k_2} 5^{k_3} 7^{k_4} e$.

We now describe formally the mapping Pure and then prove its properties.

We need the following notation. Given any tuple $\mathbf{k} = (k_1, \ldots, k_s) \in \mathbb{Z}^s$, we write $\mathbf{g}^{\mathbf{k}}$ to denote $g_1^{k_1} \cdots g_s^{k_s}$. As $\{g_1, \ldots, g_s\}$ is a generating set, there is a unique tuple $\mathbf{k} \in \mathbb{Z}^s$ for each $c \in \mathsf{Im}(F_1, \ldots, F_h) - \{0\}$ such that $c/\mathbf{g}^{\mathbf{k}}$ is a root of unity. Since $F_1, \ldots, F_h$ are fixed, all integers in $\mathbf{k}$ are constants in the problem $\mathsf{COUNT}(F_1, \ldots, F_h)$.

We define F_i' from F_i as follows. For each $\mathbf{x} \in D^{r_i}$, $F_i'(\mathbf{x}) = 0$ if $F_i(\mathbf{x}) = 0$. If $F_i(\mathbf{x}) \neq 0$, then there exists a unique tuple $\mathbf{k} \in \mathbb{Z}^s$, such that $F_i(\mathbf{x})/\mathbf{g}^{\mathbf{k}}$ is a root of unity, and we set

$$F_i'(\mathbf{x}) = p_1^{k_1} \cdots p_s^{k_s} \cdot \frac{F_i(\mathbf{x})}{\mathbf{g}^{\mathbf{k}}}, \tag{9}$$

where p_i denotes the ith smallest prime. F_i' is pure by definition, and property 2 of the lemma is satisfied. In the rest of the proof, we will use $\mathbf{p}^{\mathbf{k}}$ to denote $p_1^{k_1} \cdots p_s^{k_s}$.

Next, we show the equivalence of the two #CSP problems in property 1. By Lemma 2.1 it suffices to show that

$$\mathsf{COUNT}(F_1, \ldots, F_h) \equiv_T \mathsf{COUNT}(F_1', \ldots, F_h'). \tag{10}$$

We start with the reduction from $\mathsf{COUNT}(F_1, \ldots, F_h)$ to $\mathsf{COUNT}(F_1', \ldots, F_h')$.

Given an instance I of $\#\mathsf{CSP}(F_1, \ldots, F_h)$ over n variables, we write I' to denote the instance of $\#\mathsf{CSP}(F_1', \ldots, F_h')$ obtained by replacing the F_i in each tuple of I with its corresponding function F_i'. Also, let $m = |I| = |I'|$ and let F and F' denote the functions that I and I' define, respectively. By property 2, we have $F(\mathbf{x}) \neq 0$ iff $F'(\mathbf{x}) \neq 0$, and thus, the number of $\mathbf{x}$ such that $F(\mathbf{x}) = 0$ is the same as the number of $\mathbf{x}$ such that $F'(\mathbf{x}) = 0$. The latter can be obtained by calling a subroutine for $\mathsf{COUNT}(F_1', \ldots, F_h')$.

Let $\{c_1, \ldots, c_t\} = \mathsf{Im}(F_1, \ldots, F_h) - \{0\}$, with t being a constant as the set of functions is fixed. We then compute the following set C_m in time polynomial in m:

$$C_m = \{c_1^{\ell_1} \cdots c_t^{\ell_t} : \ell_1, \ldots, \ell_t \text{ are non-negative integers and } \ell_1 + \cdots + \ell_t = m\}.$$

For each $c \in C_m$, we also compute the unique tuple $\mathbf{k} \in \mathbb{Z}^s$ such that $c/\mathbf{g}^{\mathbf{k}}$ is a root of unity, using the known tuples for $\{c_1, \ldots, c_t\}$. By the definition of F_i' from F_i and by the assumption that $\{g_1, \ldots, g_s\}$ is a generating set, we have

$$F(\mathbf{x}) = c \iff F'(\mathbf{x}) = \mathbf{p}^{\mathbf{k}} \cdot \frac{c}{\mathbf{g}^{\mathbf{k}}}, \quad \text{for all } \mathbf{x} \in D^n.$$

As a result, the number of $\mathbf{x}$ with $F(\mathbf{x}) = c$ can be obtained by calling a subroutine for $\mathsf{COUNT}(F_1', \ldots, F_h')$. The other direction of reduction can be proved similarly.

Now, we check property 3. In the rest of the proof, we use F to denote F_i, F' to denote F_i', and r to denote r_i, the arity of F_i, for convenience. Assume $r \geq 2$ and F' is block-rank-1. Let $\mathbf{x}, \mathbf{y} \in D^{r-1}$ be two vectors such that $F(\mathbf{x}, *)$ and $F(\mathbf{y}, *)$ share at least one common nonzero entry. From property 2 and the assumption that F' is block-rank-1, we know that $|F'(\mathbf{x}, *)|$ and $|F'(\mathbf{y}, *)|$ must be nonzero and linearly dependent.

To prove that $|F(\mathbf{x}, *)|$ and $|F(\mathbf{y}, *)|$ are linearly dependent, it suffices to show for all indices $i, j \in D$ of nonzero entries of $F(\mathbf{x}, *)$ (which are also indices of nonzero entries of $F(\mathbf{y}, *)$, $F'(\mathbf{x}, *)$, $F'(\mathbf{y}, *)$),

$$|F(\mathbf{x}, i)| \cdot |F(\mathbf{y}, j)| = |F(\mathbf{y}, i)| \cdot |F(\mathbf{x}, j)|. \tag{11}$$

To this end, we let $\mathbf{u}, \mathbf{v}, \mathbf{w}, \mathbf{z} \in \mathbb{Z}^s$ denote the four vectors, such that

$$\frac{F'(\mathbf{x}, i)}{\mathbf{p}^{\mathbf{u}}}, \quad \frac{F'(\mathbf{y}, j)}{\mathbf{p}^{\mathbf{v}}}, \quad \frac{F'(\mathbf{y}, i)}{\mathbf{p}^{\mathbf{w}}}, \quad \frac{F'(\mathbf{x}, j)}{\mathbf{p}^{\mathbf{z}}}$$

are all roots of unity. Because $|F'(\mathbf{x}, *)|$ and $|F'(\mathbf{y}, *)|$ are linearly dependent, we have

$$p_1^{u_1+v_1} \cdots p_s^{u_s+v_s} = |F'(\mathbf{x}, i)| \cdot |F'(\mathbf{y}, j)| = |F'(\mathbf{y}, i)| \cdot |F'(\mathbf{x}, j)| = p_1^{w_1+z_1} \cdots p_s^{w_s+z_s},$$

and thus, $u_k + v_k = w_k + z_k$ for all $k \in [s]$. Equation (11) follows from the definition of F' from F.

Next, we prove property 4(a). Assume that $F(\mathbf{x}, *)$ and $F(\mathbf{y}, *)$ are linearly dependent. Then, we use $i, j \in D$ to denote two indices of nonzero entries of $F'(\mathbf{x}, *)$, which must be indices of nonzero entries of $F'(\mathbf{y}, *)$, $F(\mathbf{x}, *)$, and $F(\mathbf{y}, *)$ as well. Similarly, let $\mathbf{u}, \mathbf{v}, \mathbf{w}, \mathbf{z} \in \mathbb{Z}^s$ be the vectors, such that

$$c_1 = \frac{F(\mathbf{x}, i)}{\mathbf{g}^{\mathbf{u}}}, \quad c_2 = \frac{F(\mathbf{y}, j)}{\mathbf{g}^{\mathbf{v}}}, \quad c_3 = \frac{F(\mathbf{y}, i)}{\mathbf{g}^{\mathbf{w}}}, \quad c_4 = \frac{F(\mathbf{x}, j)}{\mathbf{g}^{\mathbf{z}}},$$

and $c_1, \ldots, c_4$ are all roots of unity. As $F(\mathbf{x}, *)$ and $F(\mathbf{y}, *)$ are linearly dependent,

$$c_1 \cdot c_2 \cdot g_1^{u_1+v_1} \cdots g_s^{u_s+v_s} = F(\mathbf{x}, i) \cdot F(\mathbf{y}, j) = F(\mathbf{y}, i) \cdot F(\mathbf{x}, j) = c_3 \cdot c_4 \cdot g_1^{w_1+z_1} \cdots g_s^{w_s+z_s}.$$

By the definition of generating sets, we must have $c_1 \cdot c_2 = c_3 \cdot c_4$ and $u_k + v_k = w_k + z_k$ for all $k \in [s]$. On the other hand, by the construction of F', we have

$$F'(\mathbf{x}, i) \cdot F'(\mathbf{y}, j) = c_1 \cdot c_2 \cdot p_1^{u_1+v_1} \cdots p_s^{u_s+v_s} = c_3 \cdot c_4 \cdot p_1^{w_1+z_1} \cdots p_s^{w_s+z_s} = F'(\mathbf{y}, i) \cdot F'(\mathbf{x}, j).$$

So, $F'(\mathbf{x}, *)$ and $F'(\mathbf{y}, *)$ are also linearly dependent. The other direction is similar.

For property 4(b), assume that $F'(\mathbf{x}, *)$ and $F'(\mathbf{y}, *)$ are block-orthogonal. Note that when F' is block-rank-1, F is also block-rank-1 by property 3. We then use $T \subseteq D$ to denote the set of indices $j \in D$ such that $F(\mathbf{x}, j) \neq 0$ (and $F(\mathbf{y}, j), F'(\mathbf{x}, j), F'(\mathbf{y}, j) \neq 0$ as both functions are block-rank-1). We use $F'(\mathbf{x}, *)$ to further partition T into $T_1, \ldots, T_t$, for some $t \geq 1$: there are positive integers $\mu_1 > \cdots > \mu_t > 0$, such that $|F'(\mathbf{x}, j)| = \mu_k$ for all $j \in T_k$ and $k \in [t]$. (Here, μ_i's are integers because of the definition of F' from F in Equation (9).) Since F' is block-rank-1, we know $|F'(\mathbf{x}, *)|, |F'(\mathbf{y}, *)|$ are linearly dependent and, thus, there are positive integers $\nu_1 > \cdots > \nu_t > 0$, such that $(\mu_1, \ldots, \mu_t)$ and $(\nu_1, \ldots, \nu_t)$ are linearly dependent and $|F'(\mathbf{y}, j)| = \nu_k$ for all $j \in T_k$ and $k \in [t]$.

We also use $c(\mathbf{x}, j)$ and $c(\mathbf{y}, j)$ to denote the roots of unity, such that

$$F'(\mathbf{x}, j) = \mu_k \cdot c(\mathbf{x}, j) \quad \text{and} \quad F'(\mathbf{y}, j) = \nu_k \cdot c(\mathbf{y}, j).$$

Because $F'(\mathbf{x}, *)$ and $F'(\mathbf{y}, *)$ are block-orthogonal, by definition, we have

$$\sum_{j \in T_k} F'(\mathbf{x}, j) \cdot \overline{F'(\mathbf{y}, j)} = \mu_k \cdot \nu_k \sum_{j \in T_k} c(\mathbf{x}, j) \cdot \overline{c(\mathbf{y}, j)} = 0, \quad \text{for all } k \in [t]. \tag{12}$$

For each $k \in [t]$, we write $\mathbf{u}_k \in \mathbb{Z}^s$ and $\mathbf{v}_k \in \mathbb{Z}^s$ to denote the two unique vectors, such that $\mu_k = \mathbf{p}^{\mathbf{u}_k}$ and $\nu_k = \mathbf{p}^{\mathbf{v}_k}$. Then by the construction of F', we have for all $j \in T_k$,

$$\begin{aligned} F(\mathbf{x}, j) &= \mathbf{g}^{\mathbf{u}_k} \cdot c(\mathbf{x}, j) \quad \text{and} \quad |F(\mathbf{x}, j)| = |\mathbf{g}^{\mathbf{u}_k}|, \\ F(\mathbf{y}, j) &= \mathbf{g}^{\mathbf{v}_k} \cdot c(\mathbf{y}, j) \quad \text{and} \quad |F(\mathbf{y}, j)| = |\mathbf{g}^{\mathbf{v}_k}|. \end{aligned} \tag{13}$$

Now, we are ready to show that $F(\mathbf{x}, *)$ and $F(\mathbf{y}, *)$ are indeed block-orthogonal. Let $\upsilon = |F(\mathbf{x}, j)| > 0$ for some $j \in T$ and let $S_\upsilon \subseteq T$ denote the set of indices ℓ such that $|F(\mathbf{x}, \ell)| = \upsilon$. Then, by Equation (13), S_υ must be the union of some of the T_k's. Without loss of generality, let $S_\upsilon = T_1 \cup \cdots \cup T_q$ for some $q \leq t$. Then, we have

$$\begin{aligned} \sum_{\ell \in S_\upsilon} F(\mathbf{x}, \ell) \cdot \overline{F(\mathbf{y}, \ell)} &= \sum_{k \in [q]} \sum_{\ell \in T_k} \mathbf{g}^{\mathbf{u}_k} \cdot c(\mathbf{x}, \ell) \cdot \overline{\mathbf{g}^{\mathbf{v}_k} \cdot c(\mathbf{y}, \ell)} \\ &= \sum_{k \in [q]} \mathbf{g}^{\mathbf{u}_k} \cdot \overline{\mathbf{g}^{\mathbf{v}_k}} \sum_{\ell \in T_k} c(\mathbf{x}, \ell) \cdot \overline{c(\mathbf{y}, \ell)} = 0. \end{aligned}$$

The first equation uses Equation (13) and the last equation uses Equation (12). This finishes the proof. □

We remark that in both property 3 and property 4(b) of the lemma, the statement only holds in one direction. For example, when F_i is block-rank-1, it is not clear how to

prove that F_i' is block-rank-1 as well. However, it turns out that the directions that we can prove are the ones that we will actually need later in proving the hardness lemmas for general functions.

Using properties 2, 3, and 4 of the Purification Lemma, we have

COROLLARY 2.13. *As defined in Equation* (8)*, if F_i' is block-orthogonal, then so is F_i. Moreover, the equivalence relations $\sim_{F_i}$ and $\sim_{F_i'}$ defined by F_i and F_i', respectively, are the same.*

2.8. Mal'tsev Polymorphisms and Witness Functions

Our dichotomy theorem needs the following concept of *Mal'tsev polymorphisms*:

Definition 2.14. Let $\Phi \subseteq D^n$ be an n-ary relation and $\varphi : D^3 \to D$ be a map. If

$$(\varphi(u_1, v_1, w_1), \ldots, \varphi(u_n, v_n, w_n)) \in \Phi, \quad \text{for all } \mathbf{u}, \mathbf{v}, \mathbf{w} \in \Phi,$$

then we say Φ is *closed* under φ, and call φ a ternary *polymorphism* of Φ.

Given $\Phi \subseteq D^n$ and $\varphi : D^3 \to D$, we use $\mathsf{cl}_\varphi \Phi$ to denote the closure of Φ under φ, that is, the smallest relation in D^n (in terms of set containment) that contains Φ and is closed under φ.

Definition 2.15 (*Mal'tsev Polymorphism*). Let $\Phi \subseteq D^n$ be an n-ary relation. Then, we say $\varphi : D^3 \to D$ is a *Mal'tsev polymorphism* of Φ if φ is a polymorphism of Φ and in addition satisfies

$$\varphi(a, b, b) = \varphi(b, b, a) = a, \quad \text{for all } a, b \in D. \tag{14}$$

Let $\Gamma = \{\Phi_1, \ldots, \Phi_h\}$ be a finite set of relations. We say φ is a *Mal'tsev polymorphism* of Γ if it is a Mal'tsev polymorphism of Φ_i for all $i \in [h]$.

Let $\Gamma = \{\Phi_1, \ldots, \Phi_h\}$ be a finite set of relations. Let I denote an instance of #CSP(Γ) and R denote the relation it defines. If Γ has a Mal'tsev polymorphism φ, then φ is a Mal'tsev polymorphism of R as well. On the other hand, Bulatov and Dalmau [2007] gave the following #P-hardness theorem. Also see Dyer and Richerby [2013].

THEOREM 2.16 ([BULATOV AND DALMAU 2007]). *Let Γ be a finite set of relations. If Γ does not have a Mal'tsev polymorphism, then* #CSP(Γ) *is* #P-complete.

Theorem 2.16 has the following simple corollary.

COROLLARY 2.17. *Let Λ be a collection of (possibly infinitely many) relations. Then either all relations in Λ share a common Mal'tsev polymorphism φ; or there is a finite subset $\Gamma \subset \Lambda$ such that* #CSP(Γ) *is #P-hard.*

PROOF. Note that, given d, there are only finitely many maps $\varphi : D^3 \to D$. We let P denote the set of all such maps. Now assume the relations in Λ do not share a common Mal'tsev polymorphism, then for any $\varphi \in P$, there is a relation $\Phi_\varphi \in \Lambda$, which does not have φ as a Mal'tsev polymorphism. Then from Theorem 2.16, we know that #CSP(Γ) is #P-hard, where $\Gamma = \{\Phi_\varphi : \varphi \in P\}$ is a finite subset of Λ. □

Let Φ be an n-ary relation with variables $x_1, \ldots, x_n$ ranging over D. In general, $|\Phi|$ could be exponentially large in n. But when Φ is known to have a Mal'tsev polymorphism and such a polymorphism φ is also given, Dyer and Richerby [2013] introduced the following elegant succinct representation for Φ. (See Bulatov and Dalmau [2006] for a similar notion called a "compact representation.") We start with some notation.

For each $i \in [n]$, we define the following relation $\sim_i$ on $\mathsf{Pr}_i \Phi$, the projection of Φ on its ith coordinate: $a \sim_i b$ if there exist tuples $\mathbf{x} \in D^{i-1}$ and $\mathbf{y}_a, \mathbf{y}_b \in D^{n-i}$, such that

$$\mathbf{x} \circ a \circ \mathbf{y}_a \in \Phi \quad \text{and} \quad \mathbf{x} \circ b \circ \mathbf{y}_b \in \Phi.$$

For the special case when $i = 1$, we have $a \sim_1 b$ for all $a, b \in \mathsf{Pr}_1\Phi$, because they share the common empty prefix ϵ. It was then shown in Dyer and Richerby [2013] that if Φ has a Mal'tsev polymorphism, $\sim_i$ must be an equivalence relation:

LEMMA 2.18. *If Φ has a Mal'tsev polymorphism, then $\sim_i$ is an equivalence relation for all $i \in [n]$.*

When Φ has a Mal'tsev polymorphism, we let $\mathcal{E}_{i,k} \subseteq \mathsf{Pr}_i\Phi$, where $k = 1, 2, \ldots$, denote the equivalence classes of $\sim_i$. The following lemma can be found in Dyer and Richerby [2013]. We include its short proof as an example of the use of a Mal'tsev polymorphism.

LEMMA 2.19. *If $a \sim_i b$ and $\mathbf{x} \in \Phi$ with $x_i = a$, then there is a $\mathbf{y} \in \Phi$ with $y_i = b$ and $\mathsf{Pr}_{[i-1]}\mathbf{x} = \mathsf{Pr}_{[i-1]}\mathbf{y}$.*

PROOF. As $a \sim_i b$, by definition there exist $\mathbf{z} \in D^{i-1}$ and $\mathbf{u}_1, \mathbf{u}_2 \in D^{n-i}$, such that

$$\mathbf{z} \circ b \circ \mathbf{u}_2 \in \Phi \quad \text{and} \quad \mathbf{z} \circ a \circ \mathbf{u}_1 \in \Phi.$$

By applying a Mal'tsev polymorphism φ of Φ on these two vectors together with $\mathbf{x} \in \Phi$, we get a new vector $\mathbf{y} \in \Phi$. It is easy to check that $\mathbf{y}$ satisfies both properties, and the lemma is proven. □

Next, we define the succinct representation called witness functions from Dyer and Richerby [2013].

Definition 2.20 (*Witness Function*). Let $\Phi \subseteq D^n$ denote a relation that has a Mal'tsev polymorphism. Then $\omega : [n] \times D \to D^n \cup \{\perp\}$ is called a *witness function* of Φ if it satisfies the following conditions:

(1) For any $i \in [n]$ and $a \notin \mathsf{Pr}_i\Phi$, $\omega(i, a) = \perp$;
(2) For any $i \in [n]$ and $a \in \mathsf{Pr}_i\Phi$, $\omega(i, a) \in \Phi$ is a witness for (i, a), that is, its ith entry is a;
(3) For any $i \in [n]$ and $a, b \in \mathsf{Pr}_i\Phi$ with $a \sim_i b$, we have $\mathsf{Pr}_{[i-1]}\omega(i, a) = \mathsf{Pr}_{[i-1]}\omega(i, b)$.

In Dyer and Richerby [2013] a subset of Φ that contains the image of a witness function of Φ is called a *frame* of Φ. But in this article, we will only use witness functions. The following lemma from Dyer and Richerby [2013] is the reason why a witness function is considered as a succinct (and linear-size) representation of Φ:

LEMMA 2.21 (MEMBERSHIP). *Let $\Phi \subseteq D^n$ be an n-ary relation that has a Mal'tsev polymorphism. With ω, a witness function of Φ, and φ, a Mal'tsev polymorphism of Φ, we can solve the following problem in time polynomial in n: given an $\mathbf{x} \in D^n$, decide if $\mathbf{x} \in \Phi$ or not.*

For completeness, we include the proof of Lemma 2.21 (and those of Lemmas 2.22–2.24) in the Appendix. Readers may find them helpful in becoming more familiar with the notions of Mal'tsev polymorphisms and witness functions.

Next, if φ is a Mal'tsev polymorphism of $\Phi \subseteq D^n$, then all three operations on Φ as described in Section 2.1, that is, projection, pinning (i.e., $\Phi(\mathbf{a}, *)$ for some prefix $\mathbf{a}$) and permutation, would result in a relation of which φ remains a Mal'tsev polymorphism.

LEMMA 2.22. *Let φ be a Mal'tsev polymorphism of $\Phi \subseteq D^n$, $\ell \in [n]$, $\mathbf{a} \in D^\ell$, and π be a permutation of $[n]$. Then φ is a Mal'tsev polymorphism of $\mathsf{Pr}_{[\ell]}\Phi$, $\Phi(\mathbf{a}, *)$ and $\pi(\Phi)$.*

Furthermore, given a witness function ω of Φ, we can construct witness functions of $\mathsf{Pr}_{[\ell]}\Phi$, $\Phi(\mathbf{a}, *)$, and $\pi(\Phi)$, respectively, in time polynomial in n. (By "construct a witness function" or "compute a witness function," we mean to list the table of values of the function on the domain $[n] \times D$, that is, a list of nd values.) For pinning and

projection, the following two lemmas can be found in Dyer and Richerby [2013]. In Section 7, we will discuss permutation and two other polynomial-time operations on witness functions, *union* and *splitting*. They play a key role in the algorithmic part of our dichotomy.

LEMMA 2.23 (PROJECTION). *Let φ be a Mal'tsev polymorphism and ω be a witness function of $\Phi \subseteq D^n$. Given an $\ell \in [n]$, we can construct a witness function for $\mathsf{Pr}_{[\ell]}\,\Phi$ in time polynomial in n. When ℓ is bounded by a constant, we can use ω to compute the projection $\mathsf{Pr}_{[\ell]}\,\Phi$ itself in polynomial time.*

Moreover, given any $\mathbf{x} \in \mathsf{Pr}_{[\ell]}\,\Phi$ for some $\ell \in [n]$, we can compute a vector $\mathbf{y} \in \Phi$ with $\mathbf{x} = \mathsf{Pr}_{[\ell]}\,\mathbf{y}$ in polynomial time.

LEMMA 2.24 (PINNING). *Let φ be a Mal'tsev polymorphism and ω be a witness function of $\Phi \subseteq D^n$. Given any $\mathbf{a} \in D^\ell$ for some $\ell \in [n]$, we can construct a witness function for $\Phi(\mathbf{a}, *)$ in time polynomial in n.*

Let $\Gamma = \{\Phi_1, \ldots, \Phi_h\}$ be an unweighted constraint language over D. It follows from Theorem 2.16 that #CSP(Γ) is #P-complete when Γ does not have a Mal'tsev polymorphism. Dyer and Richerby [2013] showed that when Γ has a Mal'tsev polymorphism, then given an instance I of #CSP(Γ), a witness function for the relation R_I defined by I (which also has φ as a Mal'tsev polymorphism) can be computed efficiently.

THEOREM 2.25. *Let φ be a Mal'tsev polymorphism shared by all the relations in Γ. Then given any input instance I of #CSP(Γ), one can compute a witness function of R_I in polynomial time.*

In addition, Dyer and Richerby [2013] gave a polynomial-time algorithm that, given a witness function of R_I, computes $|R_I|$ (though we will not use this algorithm here).

2.9. The Partition Condition

Let $S \subseteq D^n$ be a nonempty set and $S_1, \ldots, S_k$ be a partition of S, for some $k \geq 1$: the S_i's are nonempty and pairwise disjoint subsets of S, with $S = S_1 \cup \cdots \cup S_k$. The pair $(S, (S_1, \ldots, S_k))$ defines the following map $\mathsf{type}(\cdot)$: given any $\ell \in [n]$ and $\mathbf{x} \in D^\ell$,

$$\mathsf{type}(\mathbf{x}) = \left\{ j \in [k] : \exists \mathbf{y} \in S_j \text{ such that } \mathbf{x} = \mathsf{Pr}_{[\ell]}\,\mathbf{y} \right\} \subseteq [k]. \tag{15}$$

We also set $\mathsf{type}(\epsilon) = [k]$, where ϵ is the empty tuple. We will refer to $\mathsf{type}(\cdot)$ as the *type map* of $(S, (S_1, \ldots, S_k))$, and $\mathsf{type}(\mathbf{x})$ as the *type* of $\mathbf{x}$ (with respect to $(S, (S_1, \ldots, S_k))$). When $\ell = n$, $\mathsf{type}(\mathbf{x})$ is either $\emptyset$ or a singleton. When $\ell = n$ and $\mathsf{type}(\mathbf{x})$ is a singleton, we refer to the element in $\mathsf{type}(\mathbf{x})$ simply as the type of $\mathbf{x}$ for convenience.

Definition 2.26. Let $\mathsf{type}(\cdot)$ denote the type map of $(S, (S_1, \ldots, S_k))$. Then, we say $(S, (S_1, \ldots, S_k))$ satisfies the *partition condition* if for all $\ell \in [n]$ and $\mathbf{x}, \mathbf{y} \in D^\ell$, $\mathsf{type}(\mathbf{x})$ and $\mathsf{type}(\mathbf{y})$ are *either the same or disjoint*.

Given $(S, (S_1, \ldots, S_k))$, we refer to the $(n+1)$-tuple

$$\mathfrak{T} = (\mathcal{T}_0, \mathcal{T}_1, \ldots, \mathcal{T}_n), \quad \text{where } \mathcal{T}_\ell = \left\{ \mathsf{type}(\mathbf{x}) \subseteq [k] : \mathbf{x} \in D^\ell \text{ and } \mathsf{type}(\mathbf{x}) \neq \emptyset \right\} \subset 2^{[k]},$$

as its *list of types*. For the special case of $\ell = 0$, we have $\mathcal{T}_0 = \{[k]\}$. It is clear from the definition that all the sets in $\mathcal{T}_\ell$ are nonempty, since we are only interested in $\mathbf{x} \in D^\ell$ with $\mathsf{type}(\mathbf{x}) \neq \emptyset$, and their union must be $[k]$, since S_i's are nonempty.

The next lemma follows directly from the definition of the partition condition:

LEMMA 2.27. *If $(S, (S_1, \ldots, S_k))$ satisfies the partition condition, then $|\mathcal{T}_\ell| \leq k$ for all ℓ. Moreover, for any $i, j : 0 \leq i < j \leq n$ and $U \in \mathcal{T}_i$, $V \in \mathcal{T}_j$, either $V \subseteq U$ or $U \cap V = \emptyset$.*

One way to better understand the list $\mathfrak{T}$ is to consider it as a tree of height n: $[k] \in \mathcal{T}_0$ is the root, and the sets of $\mathcal{T}_\ell$ are nodes at level ℓ of the tree; $U \in \mathcal{T}_\ell$ and $V \in \mathcal{T}_{\ell+1}$ are adjacent if $V \subseteq U$. The tree has the property that the leaves are singletons and every other node is the union of its children.

3. A COMPLEXITY DICHOTOMY FOR #CSP WITH COMPLEX WEIGHTS

We prove Theorem 1.1 in this section. The rest of the article consists of proofs of lemmas stated here. We start by describing the necessary conditions for tractability.

Let $D = [d]$ be a domain. Let $\mathcal{F}$ be a finite set of algebraic complex functions over D. Recall the definition of $F^{[t]}$ in Equation (2). We use $\mathcal{W}_\mathcal{F}$ to denote the following set of infinitely many (though countable) algebraic complex-valued functions:

$$\mathcal{W}_\mathcal{F} = \left\{F^{[t]} : F \text{ is a function defined by an instance of } \#\mathsf{CSP}(\mathcal{F}) \text{ and } 1 \le t \le \text{arity of } F\right\}.$$

The following lemma concerning $\mathcal{W}_\mathcal{F}$ is easy to prove:

LEMMA 3.1. *For any finite subset $\mathcal{F}' \subset \mathcal{W}_\mathcal{F}$, we have* $\#\mathsf{CSP}(\mathcal{F}') \le_\mathsf{T} \#\mathsf{CSP}(\mathcal{F})$.

3.1. Hardness Part of the Dichotomy

The hardness part of the dichotomy theorem consists of three necessary conditions on $\mathcal{W}_\mathcal{F}$. The violation of any of these conditions implies that $\#\mathsf{CSP}(\mathcal{F})$ is #P-hard.

First, we impose the following condition:

> **Block Orthogonality:** Let $\{F_1, \ldots, F_k\}$ be any finite subset of $\mathcal{W}_\mathcal{F}$ and $\{g_1, \ldots, g_s\}$ be any generating set of $\mathsf{Im}(F_1, \ldots, F_k) - \{0\}$. Let
>
> $$\left(F'_1, \ldots, F'_k\right) = \mathsf{Pure}\left(F_1, \ldots, F_k; \{g_1, \ldots, g_s\}\right).$$
>
> Then, every F'_i of arity ≥ 2 is *block-orthogonal* (and in particular, *block-rank*-1).

We prove the following lemma in Section 4:

LEMMA 3.2. *If $\mathcal{F}$ does not satisfy the* Block Orthogonality condition*, then* $\#\mathsf{CSP}(\mathcal{F})$ *is #P-hard.*

Remark 3.3. For the special case of languages $\mathcal{F}$ of non-negative weights, the Block Orthogonality condition above trivially implies the condition of weak balance [Cai et al. 2016], which played an important role in their dichotomy for #CSP with non-negative weights. It remains unclear as to whether it implies the condition of balance as well [Cai et al. 2016] (see also Bulatov [2013] and Dyer and Richerby [2013] for the condition of balance in the unweighted case).

Assume $\mathcal{F}$ satisfies the Block Orthogonality condition. By Corollary 2.13, every F in $\mathcal{W}_\mathcal{F}$ with arity ≥ 2 is block-orthogonal (and in particular, block-rank-1). Let $n \ge 2$ be the arity of a function $F \in \mathcal{W}_\mathcal{F}$ that is not identically zero, and let

$$\left\{(S_1, \mathbf{v}_1), \ldots, (S_k, \mathbf{v}_k)\right\} \tag{16}$$

be the row representation of F for some $k \ge 1$, where S_j's are nonempty and disjoint. We note that $k \le d$, because F is block-orthogonal. Let $\Psi_F = S_1 \cup \cdots \cup S_k$.

In addition to Block Orthogonality, here is the second condition on $\mathcal{W}_\mathcal{F}$:

> **Type Partition:** For any function $F \in \mathcal{W}_\mathcal{F}$ that has arity $n \ge 2$ and is not identically zero, the pair $(\Psi_F, (S_1, \ldots, S_k))$ satisfies the partition condition.

It is worth pointing out that the partition condition is trivially satisfied if the arity of F is 2 and $\Psi_F \subseteq D$. We prove the following hardness lemma in Section 5.

LEMMA 3.4. *If $\mathcal{F}$ does not satisfy the* Type Partition condition*, then* #CSP($\mathcal{F}$) *is #P-hard.*

Finally, we need a condition on relations defined from $\mathcal{W}_{\mathcal{F}}$. Assume $\mathcal{F}$ satisfies the Block Orthogonality condition. If F has arity $n \geq 2$, then we denote its row representation by Equation (16) and define the following (equivalence) relation Ω_F over $2(n-1)$ variables $\mathbf{x} = (x_1, \ldots, x_{n-1})$ and $\mathbf{y} = (y_1, \ldots, y_{n-1})$:

$$\begin{aligned}(\mathbf{x}, \mathbf{y}) \in \Omega_F &\iff \mathbf{x}, \mathbf{y} \in S_j \text{ for some } j \in [k] \\ &\iff F(\mathbf{x}, *), F(\mathbf{y}, *) \text{ are nonzero and linearly dependent.}\end{aligned} \tag{17}$$

This gives us the following set $\Lambda_{\mathcal{F}}$ of infinitely many (though countable) relations:

$$\Lambda_{\mathcal{F}} = \{\text{Boolean}(F) : F \in \mathcal{W}_{\mathcal{F}}\} \cup \{\Omega_F : F \in \mathcal{W}_{\mathcal{F}} \text{ of arity} \geq 2\}. \tag{18}$$

We now impose the last condition on $\Lambda_{\mathcal{F}}$ (derived from $\mathcal{W}_{\mathcal{F}}$):

Mal'tsev: All relations in $\Lambda_{\mathcal{F}}$ share a common Mal'tsev polymorphism $\varphi : D^3 \to D$.

To finish the hardness part, we prove the following hardness lemma in Section 6:

LEMMA 3.5. *If $\mathcal{F}$ does not satisfy the* Mal'tsev condition*, then* #CSP($\mathcal{F}$) *is* #P-hard.

Remark 3.6. [3]Note that the Mal'tsev condition directly implies that relations in the unweighted version Γ of $\mathcal{F}$ share a common Mal'tsev polymorphism, which is known to be equivalent to Γ satisfying the condition of strong rectangularity [Bulatov 2013; Dyer and Richerby 2013] (though we do not need it in our proof here). This equivalence makes it easy to check (in NP) whether Γ is strongly rectangular. However, checking whether a language $\mathcal{F}$ of complex weights satisfies the Mal'tsev condition above seems to be much more challenging. While Γ having a Mal'tsev polymorphism implies that the same holds for any relation defined by an instance of #CSP(Γ), this is not the case for our Mal'tsev condition, due to the presence of cancellations in sums behind $F^{[t]}$.

3.2. Algorithmic Part of the Dichotomy

We show that if a finite set $\mathcal{F}$ of algebraic complex-valued functions satisfies all three conditions:

(a) the Block Orthogonality condition,
(b) the Type Partition condition,
(c) the Mal'tsev condition,

then there is a polynomial-time algorithm for #CSP($\mathcal{F}$). Theorem 1.1 then follows.

First, from the Mal'tsev condition, all the relations in $\Lambda_{\mathcal{F}}$ share a common Mal'tsev polymorphism. We may assume that such a polymorphism φ is given (since $\mathcal{F}$ is considered as a constant[4]) and will use it later in the algorithm.

Let I be an instance of #CSP($\mathcal{F}$), and let $F : D^n \to \mathbb{C}$ denote the function it defines. To compute $Z(I)$, we examine functions $F = F^{[n]}, \ldots, F^{[2]}$. For each $F^{[t]}$, $2 \leq t \leq n$, let

$$\mathcal{S}^{[t]} = \left\{\left(S^{[t,j]}, \mathbf{v}^{[t,j]}\right) : j \in [s_t]\right\} \tag{19}$$

denote the row representation of $F^{[t]}$. At this moment, we do not know what exactly s_t is, though by the Block Orthogonality condition, we know $F^{[t]} \in \mathcal{W}_{\mathcal{F}}$ is block-orthogonal and, thus, $0 \leq s_t \leq d$ for all t.

[3]The reader may feel free to skip the remark, since it is about the decidability of the Mal'tsev condition and is not related to the proof of the dichotomy theorem.
[4]Indeed, one can enumerate all possible maps from D^3 to D one by one, and there are only constant many, since $|D|$ is a constant.

Next, by using the Mal'tsev condition, we know that φ is a Mal'tsev polymorphism of $\Omega_{F^{[t]}}$, a relation over $2(t-1)$ variables. The following lemma shows that φ must also be a Mal'tsev polymorphism of the $S^{[t,j]}$'s when viewed as relations over $t-1$ variables:

LEMMA 3.7. *If φ is a Mal'tsev polymorphism of $\Omega_{F^{[t]}}$, then it is a Mal'tsev polymorphism of the $S^{[t,j]}$'s.*

PROOF. The claim is trivial if $s_t = 0$. Otherwise, let $\mathbf{u} \in D^{t-1}$ be a vector in $S^{[t,j]}$ for some $j \in [s_t]$. By the definition of $\Omega_{F^{[t]}}$, we have $\mathbf{y} \in S^{[t,j]}$ iff $(\mathbf{u}, \mathbf{y}) \in \Omega_{F^{[t]}}$. As a result, $S^{[t,j]} = \Omega_{F^{[t]}}(\mathbf{u}, *)$, and the lemma follows directly from Lemma 2.22. □

It makes sense now to talk about witness functions for $S^{[t,j]}$'s (though we still do not know s_t at this moment). We prove the following key lemma in Sections 7 and 8:

LEMMA 3.8. *Suppose $\mathcal{F}$ satisfies all three conditions, with φ being a Mal'tsev polymorphism of $\Lambda_{\mathcal{F}}$. Given an input instance I of $\#\mathsf{CSP}(\mathcal{F})$, letting $F : D^n \to \mathbb{C}$ denote the function it defines, we can compute in polynomial time a sequence of $n-1$ nonnegative integers $s_n, \ldots, s_2 \le d$, such that s_t is the number of pairs in the row representation of $F^{[t]}$ in Equation* (19)*. Moreover, we can compute in polynomial time s_t pairs for each $2 \le t \le n$:*

$$\left\{\left(\omega^{[t,j]}, \mathbf{v}^{[t,j]}\right) : j \in [s_t]\right\}, \tag{20}$$

where $\omega^{[t,j]} : [t-1] \times D \to D^{t-1} \cup \{\perp\}$ and $\mathbf{v}^{[t,j]} \neq \mathbf{0}$ is a d-dimensional vector, such that

(1) $\{\mathbf{v}^{[t,j]} : j \in [s_t]\}$ are exactly the s_t vectors in the row representation of $F^{[t]}$ in (19)*;*
(2) $\omega^{[t,j]}$ is a witness function of $S^{[t,j]} \subseteq D^{t-1}$, the set paired with $\mathbf{v}^{[t,j]}$ in the row representation of $F^{[t]}$ in Equation (19)*.*

Once we have obtained s_t and the pairs in Equation (20), $Z(F)$ can be computed efficiently:

LEMMA 3.9 (COMPUTATION OF $Z(I)$). *Given s_t and Equation* (20) *for each $2 \le t \le n$, $Z(F)$ can be computed in polynomial time.*

PROOF. For each $t : 2 \le t \le n$, we use Equation (19) to denote the row representation of $F^{[t]}$. By Lemma 3.8, all vectors $\mathbf{v}^{[t,j]}$ in Equation (19) have been computed and for each set $S^{[t,j]}$, we have computed one of its witness functions $\omega^{[t,j]}$.

For any $a_1 \in D$, we show how to compute $F^{[1]}(a_1)$ efficiently. The lemma follows as

$$Z(I) = \sum_{a_1 \in D} F^{[1]}(a_1).$$

We start with an informal description of the algorithm.

We first check whether $a_1 \in S^{[2,j]}$ for some $j \in [s_2]$. This can be done efficiently, since $s_2 \le d$ is bounded by a constant and for each $j \in [s_2]$, whether $a_1 \in S^{[2,j]}$ or not can be checked efficiently using the witness function $\omega^{[2,j]}$ of $S^{[2,j]}$. By definition, if $a_1 \notin S^{[2,j]}$ for all $j \in [s_2]$, we must have $F^{[2]}(a_1, *) = \mathbf{0}$, and thus,

$$F^{[1]}(a_1) = \sum_{b \in D} F^{[2]}(a_1, b) = 0.$$

Otherwise, we let $j \in [s_2]$ be the unique index such that $a_1 \in S^{[2,j]}$ and $a_2 \in D$ be the smallest nonzero index of $\mathbf{v}^{[2,j]}$. By the definition of row representations, $v_{a_2}^{[2,j]} = 1$ and $F^{[2]}(a_1, *)$ is a nonzero vector that is linearly dependent with $\mathbf{v}^{[2,j]}$. Therefore,

$$F^{[1]}(a_1) = \sum_{b \in D} F^{[2]}(a_1, b) = F^{[2]}(a_1, a_2) \cdot \sum_{b \in D} v_b^{[2,j]}.$$

ComputeF$(t, \mathbf{a})$, where $t \in [n]$ and $\mathbf{a} \in D^t$

1. if $t = n$ then
2. use the input instance I to evaluate $F^{[n]}(\mathbf{a}) = F(\mathbf{a})$; output $F(\mathbf{a})$ and exit
3. end if
4. use $\omega^{[t+1,j]}$, $j \in [s_{t+1}]$, to check if there is a $j \in [s_{t+1}]$ such that $\mathbf{a} \in S^{[t+1,j]}$
5. if no such $j \in [s_{t+1}]$ exists then
6. output 0 and exit
7. else
8. let $j \in [s_{t+1}]$ denote the unique index such that $\mathbf{a} \in S^{[t+1,j]}$
9. let $a_{t+1} \in D$ denote the smallest nonzero index of $\mathbf{v}^{[t+1,j]}$
10. recursively compute $F^{[t+1]}(\mathbf{a}, a_{t+1}) = \text{ComputeF}(t+1, \mathbf{a} \circ a_{t+1})$
11. output the following and exit

$$F^{[t]}(\mathbf{a}) = \sum_{b \in D} F^{[t+1]}(\mathbf{a}, b) = F^{[t+1]}(\mathbf{a}, a_{t+1}) \cdot \sum_{b \in D} v_b^{[t+1,j]}$$

12. end if

Fig. 1. The recursive procedure ComputeF.

This reduces the computation of $F^{[1]}(a_1)$ to that of $F^{[2]}(a_1, a_2)$. If $n = 2$, then we are already done, because $F^{[2]}(a_1, a_2)$ can be evaluated efficiently using the input instance I. Otherwise, we continue and reduce the computation of $F^{[2]}(a_1, a_2)$ to that of $F^{[3]}(a_1, a_2, a_3)$ for some appropriate $a_3 \in D$.

As $F^{[2]}(a_1, *)$ is nonzero and is linearly dependent with $\mathbf{v}^{[2,j]}$, we have $F^{[2]}(a_1, a_2) \neq 0$, and thus, $(a_1, a_2) \in S^{[3,j]}$ for some $j \in [s_3]$. Using witness functions $\omega^{[3,j]}$, we can find this $j \in [s_3]$ efficiently. By the definition of row representations, $F^{[3]}(a_1, a_2, *)$ is linearly dependent with $\mathbf{v}^{[3,j]}$. Let a_3 denote the smallest nonzero index of $\mathbf{v}^{[3,j]}$. Then,

$$v_{a_3}^{[3,j]} = 1 \quad \text{and} \quad F^{[2]}(a_1, a_2) = \sum_{b \in D} F^{[3]}(a_1, a_2, b) = F^{[3]}(a_1, a_2, a_3) \cdot \sum_{b \in D} v_b^{[3,j]}.$$

This further reduces the computation of $F^{[1]}(a_1)$ to that of $F^{[3]}(a_1, a_2, a_3)$.

After $n - 1$ rounds of such reductions, it suffices to compute $F^{[n]}(a_1, \ldots, a_n)$ for some appropriate $a_2, a_3, \ldots, a_n \in D$, to get $F^{[1]}(a_1)$. This gives an efficient algorithm for $F^{[1]}(a_1)$ as $F^{[n]}$ can be evaluated efficiently using the input instance I.

A formal recursive procedure called ComputeF is described in Figure 1. It takes two inputs: t and $\mathbf{a}$, where $t \in [n]$ and $\mathbf{a} \in D^t$, and outputs $F^{[t]}(\mathbf{a})$. Its correctness can be easily proved by induction on t, and its running time is polynomial, because the total number of recursive calls is at most $n - 1$ and in each call, the only non-trivial part is line 4, which has an efficient implementation by Lemma 2.21. □

4. THE BLOCK ORTHOGONALITY CONDITION

We prove Lemma 3.2 in this section. We start with a hardness lemma for *pure* functions and then use the Purification Lemma to extend it to general functions.

LEMMA 4.1. *Let $F : D^n \to \mathbb{C}$ be a pure function with arity $n \geq 2$. If F is not block-orthogonal, then* #CSP(F) *is* #P-hard.

PROOF. As F is pure, we let K denote the constant order(F). Without loss of generality, we assume that F is block-rank-1; otherwise, #CSP(F) is #P-hard by Lemma 2.5.

Now suppose F is not block-orthogonal. Then, by definition, there exist $\mathbf{x}, \mathbf{y} \in D^{n-1}$ such that the two vectors $F(\mathbf{x}, *)$ and $F(\mathbf{y}, *)$ have at least one common non-zero entry but are neither linearly dependent nor block-orthogonal. As F is block-rank-1, we have $F(\mathbf{x}, i) = 0$ if and only if $F(\mathbf{y}, i) = 0$. We use $T \subseteq D$ to denote the nonempty set of $i \in D$ such that $F(\mathbf{x}, i)$ is non-zero. Since F is pure and block-rank-1, we can partition T into $T_1, \ldots, T_\ell$ for some $\ell \geq 1$ and there are real and positive a and b, and $\mu_1 > \cdots > \mu_\ell > 0$, such that

$$F(\mathbf{x}, i) = a \cdot \mu_j \cdot c(\mathbf{x}, i) \quad \text{and} \quad F(\mathbf{y}, i) = b \cdot \mu_j \cdot c(\mathbf{y}, i), \quad \text{for all } j \in [\ell] \text{ and } i \in T_j,$$

where $c(\mathbf{x}, i)$ and $c(\mathbf{y}, i)$ are all roots of unity whose orders divide K.

To show that #CSP(F) is #P-hard, we write $\mathbf{A}_r$, for each integer $r \geq 1$, to denote the following $d^{n-1} \times d^{n-1}$ matrix with its rows and columns indexed by D^{n-1}:

$$A_r(\mathbf{w}, \mathbf{w}') = \sum_{i \in D} F(\mathbf{w}, i) \cdot \big(F(\mathbf{w}', i)\big)^{rK-1}, \quad \text{for all } \mathbf{w}, \mathbf{w}' \in D^{n-1}.$$

Note that $\mathbf{A}_r$ is not necessarily symmetric. We prove that, for every $r \geq 1$,

$$\text{EVAL}(\mathbf{A}_r) \leq_T \#\text{CSP}(F).$$

Given any directed graph $G = (V, E)$ as an input instance of EVAL($\mathbf{A}_r$), we construct I, an instance of #CSP(F), with the following set of variables:

$$z_{v,1}, \ldots, z_{v,n-1}, w_e, \quad \text{for all } v \in V \text{ and } e \in E,$$

ranging over D. Then for each edge $e = uv \in E$, we apply

$$F \text{ over } (z_{u,1}, \ldots, z_{u,n-1}, w_e) \text{ and } (rK-1) \text{ copies of } F \text{ over } (z_{v,1}, \ldots, z_{v,n-1}, w_e).$$

The reduction follows from $Z_{\mathbf{A}_r}(G) = Z(I)$. Now to prove that #CSP(F) is #P-hard, it suffices to show that EVAL($\mathbf{A}_r$) is #P-hard for some integer $r \geq 1$.

Focusing on the 2×2 sub-matrix of $\mathbf{A}_r$ indexed by $\mathbf{x}$ and $\mathbf{y}$, we have

$$A_r(\mathbf{x}, \mathbf{x}) = \sum_{j \in [\ell]} \sum_{i \in T_j} \big(a \cdot \mu_j \cdot c(\mathbf{x}, i)\big)^{rK} = a^{rK} \sum_{j \in [\ell]} |T_j| \cdot (\mu_j)^{rK},$$

$$A_r(\mathbf{y}, \mathbf{y}) = \sum_{j \in [\ell]} \sum_{i \in T_j} \big(b \cdot \mu_j \cdot c(\mathbf{y}, i)\big)^{rK} = b^{rK} \sum_{j \in [\ell]} |T_j| \cdot (\mu_j)^{rK},$$

while

$$A_r(\mathbf{x}, \mathbf{y}) = \sum_{j \in [\ell]} \sum_{i \in T_j} a\mu_j \cdot c(\mathbf{x}, i) \cdot \big(b\mu_j \cdot c(\mathbf{y}, i)\big)^{rK-1} = ab^{rK-1} \sum_{j \in [\ell]} (\mu_j)^{rK} \sum_{i \in T_j} c(\mathbf{x}, i)\overline{c(\mathbf{y}, i)},$$

$$A_r(\mathbf{y}, \mathbf{x}) = \sum_{j \in [\ell]} \sum_{i \in T_j} b\mu_j \cdot c(\mathbf{y}, i) \cdot \big(a\mu_j \cdot c(\mathbf{x}, i)\big)^{rK-1} = a^{rK-1}b \sum_{j \in [\ell]} (\mu_j)^{rK} \sum_{i \in T_j} c(\mathbf{y}, i)\overline{c(\mathbf{x}, i)}.$$

We use L_r to denote

$$\sum_{j \in [\ell]} (\mu_j)^{rK} \sum_{i \in T_j} c(\mathbf{x}, i)\overline{c(\mathbf{y}, i)}. \tag{21}$$

Since all the μ_j's are real and positive, we have

$$A_r(\mathbf{x}, \mathbf{y}) \cdot A_r(\mathbf{y}, \mathbf{x}) = a^{rK}b^{rK} \cdot |L_r|^2.$$

We discuss the following three cases. First, if

$$|L_r| = \sum_{j \in [\ell]} |T_j| \cdot (\mu_j)^{rK}, \quad \text{for some } r \geq 1,$$

then by Cauchy–Schwarz, it must be the case that $c(\mathbf{x}, *)$ and $c(\mathbf{y}, *)$, as two $|T|$-dimensional vectors, are linearly dependent and thus, $F(\mathbf{x}, *)$ and $F(\mathbf{y}, *)$ are linearly dependent, contradicting the assumption.

Second, if $L_r = 0$ for all $r \geq 1$, then by solving a Vandermonde system in which Equation (21) is 0 for r from 1 to ℓ, we must have

$$\sum_{i \in T_j} c(\mathbf{x}, i)\overline{c(\mathbf{y}, i)} = 0, \quad \text{for all } j \in [\ell].$$

As a result, these two rows are block-orthogonal, contradicting the assumption again.

Otherwise, we must have

$$0 < |L_r| < \sum_{j \in [\ell]} |T_j| \cdot (\mu_j)^{rK}, \quad \text{for some } r \geq 1.$$

So, all four entries of this sub-matrix of $|\mathbf{A}_r|$ are positive but its rank is 2. This implies that $|\mathbf{A}_r|$ is not a block-rank-1 matrix. From Lemma 2.4, we have $\mathsf{EVAL}(\mathbf{A}_r)$ is #P-hard and so is $\#\mathsf{CSP}(F)$. This finishes the proof of the lemma. □

Lemma 3.2 now follows from Lemmas 4.1, 3.1, and 2.12.

Proof Lemma 3.2 Assume for contradiction that $\mathcal{F}$ does not satisfy the Block Orthogonality condition. Let $\{F_1, \ldots, F_k\} \subset \mathcal{W}_{\mathcal{F}}$ be a finite set that violates the Block Orthogonality condition with a generating set $\{g_1, \ldots, g_s\}$ of $\mathsf{Im}(F_1, \ldots, F_k) - \{0\}$. Then, let $(F'_1, \ldots, F'_k) = \mathsf{Pure}(F_1, \ldots, F_k; \{g_1, \ldots, g_s\})$. By Lemma 2.12 and Lemma 3.1,

$$\#\mathsf{CSP}(F'_i) \leq_T \#\mathsf{CSP}(F'_1, \ldots, F'_k) \equiv_T \#\mathsf{CSP}(F_1, \ldots, F_k) \leq_T \#\mathsf{CSP}(\mathcal{F}).$$

If F'_i is not block-orthogonal, by Lemma 4.1 $\#\mathsf{CSP}(F'_i)$ is #P-hard and so is $\#\mathsf{CSP}(\mathcal{F})$. □

5. THE TYPE PARTITION CONDITION

We prove Lemma 3.4 in this section. Again, we start by working on pure functions and then extend it to general functions. Let $F : D^n \to \mathbb{C}$ be a pure function of arity $n \geq 2$. Also assume that F is block-orthogonal (and in particular, block-rank-1 as well).

Let $\mathcal{S} = \{(S_1, \mathbf{v}_1), \ldots, (S_k, \mathbf{v}_k)\}$ be the row representation of F, with $S_1, \ldots, S_k$ being nonempty and pairwise disjoint subsets of D^{n-1}. Let $\Psi = S_1 \cup \cdots \cup S_k$. Given the pair $(\Psi, (S_1, \ldots, S_k))$, recall its type map $\mathsf{type}(\cdot)$ as in Equation (15): for any $\ell \in [n-1]$ and $\mathbf{x} \in D^\ell$,

$$\mathsf{type}(\mathbf{x}) = \left\{ j \in [k] : \exists\, \mathbf{y} \in S_j \text{ such that } \mathbf{x} = \mathsf{Pr}_{[\ell]}\mathbf{y} \right\}.$$

We show that $\#\mathsf{CSP}(F)$ is #P-hard if $(\Psi, (S_1, \ldots, S_k))$ violates the partition condition.

Lemma 5.1. *Let $F : D^n \to \mathbb{C}$ be a pure and block-orthogonal function with $n \geq 2$. Then $\#\mathsf{CSP}(F)$ is #P-hard if there exist an $\ell \in [n-1]$ and $\mathbf{x}, \mathbf{y} \in D^\ell$, such that*

$$\textit{neither } \mathsf{type}(\mathbf{x}) \cap \mathsf{type}(\mathbf{y}) = \emptyset \textit{ nor } \mathsf{type}(\mathbf{x}) = \mathsf{type}(\mathbf{y}). \tag{22}$$

Proof. We start with some notation. Let K be the constant $\mathsf{order}(F)$.

Because $\mathcal{S}$ is the row representation of F, there is a function $g : \Psi \to \mathbb{C}$, such that

$$F(\mathbf{x}, *) = g(\mathbf{x}) \cdot \mathbf{v}_j, \quad \text{for all } j \in [k] \text{ and } \mathbf{x} \in S_j.$$

By the definition of row representations, $g(\mathbf{x})$ is the first non-zero entry of $F(\mathbf{x}, *)$. As F is pure, $g(\mathbf{x})$ is the product of a positive rational number and a root of unity whose order divides K, for all $\mathbf{x} \in \Psi$. This then implies that all the nonzero entries of $\mathbf{v}_1, \ldots, \mathbf{v}_k$ are products of a positive rational number and a root of unity whose order divides K.

Moreover, we know that for any $i \neq j \in [k]$, it follows from Lemma 2.6 that

$$\sum_{a \in D} v_{i,a} \cdot (v_{j,a})^{K-1} = 0, \tag{23}$$

because they are block-orthogonal. For each $j \in [k]$, let

$$c_j = \sum_{a \in D} v_{j,a} \cdot (v_{j,a})^{K-1} = \sum_{a \in D} |v_{j,a}|^K > 0.$$

Now, we start the proof. Let $\ell \in [n-1]$ and let $\mathbf{x}, \mathbf{y} \in D^\ell$ be two vectors that satisfy Equation (22). Note that when $\ell = n-1$, $\mathsf{type}(\mathbf{x})$ is either $\emptyset$ or a singleton set. So for Equation (22) to hold, ℓ must be smaller than $n-1$. (Note that this implies that the hardness condition of the lemma never occurs for binary functions.) Without loss of generality, let

$$\mathsf{type}(\mathbf{x}) = L_1 \cup L_2 \quad \text{and} \quad \mathsf{type}(\mathbf{y}) = L_1 \cup L_3,$$

where L_i's are pairwise disjoint and L_1 and at least one of L_2, L_3 are nonempty.

Let $\mathbf{A}$ denote the following $d^\ell \times d^\ell$ matrix: for $\mathbf{z}, \mathbf{w} \in D^\ell$, the $(\mathbf{z}, \mathbf{w})$th entry of $\mathbf{A}$ is

$$A(\mathbf{z}, \mathbf{w}) = \sum_{\mathbf{z}', \mathbf{w}' \in D^{n-1-\ell}} \left(\sum_{p \in D} F(\mathbf{z}, \mathbf{z}', p) \cdot (F(\mathbf{w}, \mathbf{w}', p))^{K-1} \right) \left(\sum_{q \in D} (F(\mathbf{z}, \mathbf{z}', q))^{K-1} \cdot F(\mathbf{w}, \mathbf{w}', q) \right).$$

It is easy to see that $\mathbf{A}$ is symmetric. We use the following construction to show that

$$\mathsf{EVAL}(\mathbf{A}) \leq_{\mathsf{T}} \#\mathsf{CSP}(F). \tag{24}$$

Given any undirected graph $G = (V, E)$ as an instance of $\mathsf{EVAL}(\mathbf{A})$, we construct I, an instance of $\#\mathsf{CSP}(F)$, with the following variables:

$$v_1, \ldots, v_\ell \text{ for each } v \in V \text{ and } p_e, q_e, s_{e,\ell+1}, \ldots, s_{e,n-1}, r_{e,\ell+1}, \ldots, r_{e,n-1} \text{ for each } e \in E.$$

For each edge $e = uv \in E$, we apply one copy of F over each of

$$(u_1, \ldots, u_\ell, s_{e,\ell+1}, \ldots, s_{e,n-1}, p_e) \quad \text{and} \quad (v_1, \ldots, v_\ell, r_{e,\ell+1}, \ldots, r_{e,n-1}, q_e),$$

and we apply $(K-1)$ copies of F over each of

$$(u_1, \ldots, u_\ell, s_{e,\ell+1}, \ldots, s_{e,n-1}, q_e) \quad \text{and} \quad (v_1, \ldots, v_\ell, r_{e,\ell+1}, \ldots, r_{e,n-1}, p_e).$$

(Technically the construction of I above chooses an orientation for each edge, but the value $Z(I)$ is independent of this choice.) It then follows from the definition of $\mathbf{A}$ from F that $Z_{\mathbf{A}}(G) = Z(I)$, and Equation (24) follows.

Now to finish the proof, it suffices to show that $\mathsf{EVAL}(\mathbf{A})$ is #P-hard. To this end, we analyze the four entries of $\mathbf{A}$ with $\mathbf{z}, \mathbf{w} \in \{\mathbf{x}, \mathbf{y}\}$.

For each $i \in \mathsf{type}(\mathbf{x}) = L_1 \cup L_2$, we use U_i to denote the nonempty set of vectors $\mathbf{x}' \in D^{n-\ell-1}$ such that $\mathbf{x} \circ \mathbf{x}' \in S_i$. And we define V_i similarly for $\mathbf{y}$. Then, for $i \neq j \in L_1 \cup L_2$ and $\mathbf{z}' \in U_i, \mathbf{w}' \in U_j$, we have

$$\sum_{p \in D} F(\mathbf{x}, \mathbf{z}', p) \cdot (F(\mathbf{x}, \mathbf{w}', p))^{K-1} = 0$$

by Equation (23). This can be used to simplify the sum in $A(\mathbf{x}, \mathbf{x})$ as follows:

$$\begin{aligned} A(\mathbf{x}, \mathbf{x}) &= \sum_{i \in L_1 \cup L_2} \sum_{\mathbf{x}', \mathbf{x}'' \in U_i} \left(\sum_{p \in D} g(\mathbf{x}, \mathbf{x}') \cdot v_{i,p} \cdot \left(g(\mathbf{x}, \mathbf{x}'') \cdot v_{i,p} \right)^{K-1} \right) \\ &\quad \times \left(\sum_{q \in D} \left(g(\mathbf{x}, \mathbf{x}') \cdot v_{i,q} \right)^{K-1} \cdot g(\mathbf{x}, \mathbf{x}'') \cdot v_{i,q} \right) \\ &= \sum_{i \in L_1 \cup L_2} \sum_{\mathbf{x}', \mathbf{x}'' \in U_i} |g(\mathbf{x}, \mathbf{x}')|^K \cdot |g(\mathbf{x}, \mathbf{x}'')|^K \cdot (c_i)^2 = \sum_{i \in L_1 \cup L_2} \left(\sum_{\mathbf{x}' \in U_i} |g(\mathbf{x}, \mathbf{x}')|^K \cdot c_i \right)^2. \end{aligned}$$

Similarly, we have

$$A(\mathbf{y}, \mathbf{y}) = \sum_{i \in L_1 \cup L_3} \sum_{\mathbf{y}', \mathbf{y}'' \in V_i} |g(\mathbf{y}, \mathbf{y}')|^K \cdot |g(\mathbf{y}, \mathbf{y}'')|^K \cdot (c_i)^2 = \sum_{i \in L_1 \cup L_3} \left(\sum_{\mathbf{y}' \in V_i} |g(\mathbf{y}, \mathbf{y}')|^K \cdot c_i \right)^2.$$

On the other hand, by a similar proof, we also have

$$\begin{aligned} A(\mathbf{x}, \mathbf{y}) &= \sum_{i \in L_1} \sum_{\mathbf{x}' \in U_i,\, \mathbf{y}' \in V_i} \left(\sum_{p \in D} g(\mathbf{x}, \mathbf{x}') \cdot v_{i,p} \cdot \left(g(\mathbf{y}, \mathbf{y}') \cdot v_{i,p} \right)^{K-1} \right) \\ &\quad \times \left(\sum_{q \in D} \left(g(\mathbf{x}, \mathbf{x}') \cdot v_{i,q} \right)^{K-1} \cdot g(\mathbf{y}, \mathbf{y}') \cdot v_{i,q} \right) \\ &= \sum_{i \in L_1} \sum_{\mathbf{x}' \in U_i,\, \mathbf{y}' \in V_i} |g(\mathbf{x}, \mathbf{x}')|^K \cdot |g(\mathbf{y}, \mathbf{y}')|^K \cdot (c_i)^2 \\ &= \sum_{i \in L_1} \left(\sum_{\mathbf{x}' \in U_i} |g(\mathbf{x}, \mathbf{x}')|^K \cdot c_i \right) \left(\sum_{\mathbf{y}' \in V_i} |g(\mathbf{y}, \mathbf{y}')|^K \cdot c_i \right), \end{aligned}$$

and $A(\mathbf{y}, \mathbf{x}) = A(\mathbf{x}, \mathbf{y})$ (as the definition of $\mathbf{A}$ is symmetric). Using the same argument, we can also see that all entries of $\mathbf{A}$ are non-negative.

Since L_1 is nonempty, we have $A(\mathbf{x}, \mathbf{y}) = A(\mathbf{y}, \mathbf{x}) > 0$. It is now easy to see that if at least one of the L_2, L_3 is nonempty, then $A(\mathbf{x}, \mathbf{x}) \cdot A(\mathbf{y}, \mathbf{y}) > A(\mathbf{x}, \mathbf{y}) \cdot A(\mathbf{y}, \mathbf{x})$. By Theorem 2.3, we have that $\mathsf{EVAL}(\mathbf{A})$ is #P-hard and so is $\#\mathsf{CSP}(F)$. This proves the lemma. □

Finally, we use the Purification Lemma to prove Lemma 3.4.

Proof of Lemma 3.4. Without loss of generality, we may assume that $\mathcal{F}$ satisfies the Block Orthogonality condition; otherwise, $\#\mathsf{CSP}(\mathcal{F})$ is #P-hard by Lemma 3.2.

Let $F \in \mathcal{W}_{\mathcal{F}}$ be a function of arity at least 2 and let $F' = \mathsf{Pure}(F; \{g_1, \ldots, g_s\})$, where $\{g_1, \ldots, g_s\}$ is a generating set of $\mathsf{Im}(F) - \{0\}$. By Lemma 3.1, we have

$$\#\mathsf{CSP}(F') \equiv_T \#\mathsf{CSP}(F) \leq_T \#\mathsf{CSP}(\mathcal{F}).$$

As $\mathcal{F}$ satisfies the Block Orthogonality condition, F' is block-orthogonal. Using Corollary 2.13 of the Purification Lemma, F and F' must induce the same equivalence relation $\sim_F$ and $\sim_{F'}$ and therefore, the type maps $\mathsf{type}_F(\cdot)$ and $\mathsf{type}_{F'}(\cdot)$, induced by F

and F', respectively, are the same. If $\mathsf{type}_F(\cdot)$ violates the partition condition, then so does $\mathsf{type}_{F'}(\cdot)$. By Lemma 5.1, #CSP($F'$) is #P-hard and so is #CSP($\mathcal{F}$). □

6. THE MAL'TSEV CONDITION

We prove Lemma 3.5 in this section. It follows directly from the following lemma:

LEMMA 6.1. *If $\mathcal{F}$ satisfies the Block Orthogonality condition, then for any finite set $\Gamma \subset \Lambda_{\mathcal{F}}$, we have* #CSP($\Gamma$) $\leq_{\mathsf{T}}$ #CSP($\mathcal{F}$).

PROOF OF LEMMA 3.5. If $\mathcal{F}$ does not satisfy the Block Orthogonality condition, then we are done by Lemma 3.2. Assume $\mathcal{F}$ satisfies the Block Orthogonality condition but does not satisfy the Mal'tsev condition.

Recall that $\Lambda_{\mathcal{F}}$ is a set of relations defined using $\mathcal{W}_{\mathcal{F}}$ in Equation (18). By Corollary 2.17 there exists a finite set $\Gamma \subset \Lambda_{\mathcal{F}}$ with #CSP(Γ) being #P-hard. By Lemma 6.1, #CSP($\mathcal{F}$) is also #P-hard, and the lemma is proven. □

We prove Lemma 6.1 in the rest of the section.

PROOF OF LEMMA 6.1. Given Γ, we can find a finite subset $\{F_1, \ldots, F_k\} \subset \mathcal{W}_{\mathcal{F}}$, such that $\Gamma \subseteq \Delta$, where we use Ω_i to denote the relation defined using F_i as in Equation (17), and

$$\Delta = \{\mathsf{Boolean}(F_i) : i \in [k]\} \cup \{\Omega_i : i \in [k] \text{ and the arity of } F_i \text{ is} \geq 2\}.$$

Let r_i denote the arity of F_i. Recall that Ω_i is a relation over $2(r_i - 1)$ variables:

$$(\mathbf{x}, \mathbf{y}) \in \Omega_i \iff F_i(\mathbf{x}, *) \text{ and } F_i(\mathbf{y}, *) \text{ are non-zero and linearly dependent.}$$

Using Lemma 3.1, we have

$$\#\mathsf{CSP}\big(F_1, \ldots, F_k\big) \leq_{\mathsf{T}} \#\mathsf{CSP}(\mathcal{F}), \tag{25}$$

so it suffices to give a polynomial-time reduction from #CSP(Δ) to the former.

For this purpose, we first apply the Purification Lemma to get

$$\big(F'_1, \ldots, F'_k\big) = \mathsf{Pure}\big(F_1, \ldots, F_k; \{g_1, \ldots, g_s\}\big), \tag{26}$$

using an arbitrary generating set $\{g_1, \ldots, g_s\}$ of $\mathsf{Im}\{F_1, \ldots, F_k\} - \{0\}$. Now, all k functions $F'_1, \ldots, F'_k$ are pure, and we have

$$\#\mathsf{CSP}\big(F'_1, \ldots, F'_k\big) \equiv_T \#\mathsf{CSP}\big(F_1, \ldots, F_k\big). \tag{27}$$

We use K to denote the least common multiple of the orders of all the pure F'_i's.

The plan of the proof is the following. For each $i \in [k]$ with $r_i \geq 2$, we use a construction to define, from F'_i, a $2(r_i - 1)$-ary function H_i and prove that

$$\#\mathsf{CSP}\Big(\big\{F'_i : i \in [k]\big\} \cup \big\{H_i : i \in [k] \text{ and } r_i \geq 2\big\}\Big) \leq_{\mathsf{T}} \#\mathsf{CSP}\big(F'_1, \ldots, F'_k\big). \tag{28}$$

We will show that for each $i \in [k]$ with $r_i \geq 2$, $\Omega_i = \mathsf{Boolean}(H_i)$. On the other hand, by property 2 of the Purification Lemma, we have that

$$\mathsf{Boolean}(F_i) = \mathsf{Boolean}(F'_i), \quad \text{for all } i \in [k].$$

As a result, by Lemma 2.8, we have

$$\#\mathsf{CSP}(\Delta) \leq_{\mathsf{T}} \#\mathsf{CSP}\Big(\big\{F'_i : i \in [k]\big\} \cup \big\{H_i : i \in [k] \text{ and } r_i \geq 2\big\}\Big), \tag{29}$$

and the lemma follows by combining Equations (29), (28), (27), and (25).

For each $i \in [k]$ with $r_i \geq 2$, we use H_i to denote the following function:

$$H_i(\mathbf{x}, \mathbf{y}) = \sum_{z \in D} F_i'(\mathbf{x}, z) \cdot \left(F_i'(\mathbf{y}, z)\right)^{K-1}, \quad \text{for all } \mathbf{x}, \mathbf{y} \in D^{r_i - 1}.$$

We use the following construction to show the reduction in Equation (28). Given an instance I of the first problem in Equation (28), we construct an instance I' of the second problem as follows. We start with the same set of variables as I and add to I' all the constraints in I whose function is some F_i'. For each constraint $(H_i, x_1, \ldots, x_{r_i-1}, y_1, \ldots, y_{r_i-1})$ in I, we create a new variable z and add the following K constraints to I': $(F_i', x_1, \ldots, x_{r_i-1}, z)$ and $(K-1)$ copies of $(F_i', y_1, \ldots, y_{r_i-1}, z)$. We have $Z(I) = Z(I')$, and Equation (28) follows.

Finally, we show $\Omega_i = \mathsf{Boolean}(H_i)$. As $\mathcal{F}$ satisfies the Block Orthogonality condition, Equation (26) implies that F_i' is block-orthogonal. From this, it follows from Lemma 2.6 that

$$H_i(\mathbf{x}, \mathbf{y}) \neq 0 \iff F_i'(\mathbf{x}, *) \text{ and } F_i'(\mathbf{y}, *) \text{ are nonzero and linearly dependent.}$$

It follows from Corollary 2.13 that the two equivalence relations $\sim_{F_i}$ and $\sim_{F_i'}$, induced by F_i and F_i', respectively, are the same. As a result, $F_i'(\mathbf{x}, *)$, $F_i'(\mathbf{y}, *)$ are nonzero and linearly dependent if and only if $F_i(\mathbf{x}, *)$, $F_i(\mathbf{y}, *)$ are nonzero and linearly dependent. This proves $\Omega_i = \mathsf{Boolean}(H_i)$ and finishes the proof of the lemma. □

7. POLYNOMIAL-TIME OPERATIONS ON WITNESS FUNCTIONS

In this section, we present three polynomial-time operations on witness functions of relations that share a common Mal'tsev polymorphism. They will be used in Section 8 to prove Lemma 3.8.

7.1. Variable Permutation of Witness Functions

LEMMA 7.1 (VARIABLE PERMUTATION). *Let $\Phi \subseteq D^n$ be an n-ary relation. Let φ be a Mal'tsev polymorphism of Φ and ω be a witness function of Φ. Given any permutation π of $[n]$, we can compute a witness function ω' for $\pi(\Phi)$ in time polynomial in n.*

PROOF. It suffices to show that given any $i \in [n-1]$, we can compute a new witness function ω', that is, to give the full table of its values, for

$$\Phi' = \left\{(a_1, \ldots, a_{i-1}, a_i, a_{i+1}, a_{i+2}, \ldots, a_n) : (a_1, \ldots, a_{i-1}, a_{i+1}, a_i, a_{i+2}, \ldots, a_n) \in \Phi\right\}.$$

in time polynomial in n. For each $j \in [n]$ let $\sim_j$ and $\sim_j'$ denote the equivalence relations defined by Φ and Φ', respectively. Clearly, for $j \notin \{i, i+1\}$, $\sim_j'$ is the same as $\sim_j$. Thus, we can set $\omega'(j, a)$ to be $\omega(j, a)$, after transposing the ith and $(i+1)$th coordinates.

Next, we compute $\sim_i'$. Let $b \in \mathsf{Pr}_i \Phi' = \mathsf{Pr}_{i+1} \Phi$. Note that the latter can be computed efficiently from ω. We want to compute the class $\mathcal{E}$ of b in $\sim_i'$ and, in addition, a witness for each $b' \in \mathcal{E}$ that shares a common $(i-1)$-prefix. We are then done for $\mathcal{E}$ by setting $\omega'(i, b')$ to be this witness for every $b' \in \mathcal{E}$.

To this end, we denote $\omega(i+1, b)$, a witness for $(i+1, b)$ in Φ, by

$$\mathbf{x} \circ a \circ b \circ \mathbf{u} \in \Phi, \quad \text{where } \mathbf{x} \in D^{i-1}, a \in D \text{ and } \mathbf{u} \in D^{n-i-1}. \tag{30}$$

We use Lemma 2.24 to compute a witness function for $\Phi(\mathbf{x}, *)$ on $n-(i-1)$ variables and use it to project $\Phi(\mathbf{x}, *)$ on its second coordinate: $\mathsf{Pr}_2 \Phi(\mathbf{x}, *)$. We claim $\mathcal{E} = \mathsf{Pr}_2 \Phi(\mathbf{x}, *)$.

Clearly every $b' \in \mathsf{Pr}_2 \Phi(\mathbf{x}, *)$ satisfies $b' \sim_i' b$, since $b' \in \mathsf{Pr}_2 \Phi(\mathbf{x}, *)$ implies that there is a witness for $(i+1, b')$ in Φ with the same prefix $\mathbf{x}$. Now, suppose $b' \sim_i' b$. Then, by the definition of $\sim_i'$, there exist a $\mathbf{y} \in D^{i-1}$ and $a_1, a_2 \in D$, $\mathbf{u}_1, \mathbf{u}_2 \in D^{n-i-1}$, such that

$$\mathbf{y} \circ a_1 \circ b' \circ \mathbf{u}_1 \in \Phi \quad \text{and} \quad \mathbf{y} \circ a_2 \circ b \circ \mathbf{u}_2 \in \Phi.$$

Applying the Mal'tsev polymorphism φ on these two vectors and the one in Equation (30) gives us a witness for $(i+1, b')$ in Φ with $\mathbf{x}$ as its prefix. Thus, $b' \in \mathsf{Pr}_2\, \Phi(\mathbf{x}, *)$.

Now, we have computed $\mathcal{E}$. We can use the witness function of $\Phi(\mathbf{x}, *)$ to get a witness for $(i+1, b')$ in Φ with $\mathbf{x}$ being its prefix. Transposing the ith and $(i+1)$th coordinates gives a witness for (i, b') in Φ'. This finishes the construction of $\omega'(i, b)$ for all b.

Finally, we work on $\sim'_{i+1}$. Let $a \in \mathsf{Pr}_{i+1}\, \Phi' = \mathsf{Pr}_i\, \Phi$. We need to compute the equivalence class $\mathcal{E}$ of a in $\sim'_{i+1}$. We denote the vector $\omega(i, a)$ by

$$\mathbf{x} \circ a \circ b \circ \mathbf{u} \in \Phi, \quad \text{where } \mathbf{x} \in D^{i-1}, b \in D \text{ and } \mathbf{u} \in D^{n-i-1}. \tag{31}$$

We use Lemma 2.24 to compute a witness function of $\Phi(\mathbf{x}, *)$ and then $\mathsf{Pr}_{[2]}\, \Phi(\mathbf{x}, *)$ by Lemma 2.23. For each pair in $(a', b') \in \mathsf{Pr}_{[2]}\, \Phi(\mathbf{x}, *)$, we also compute a vector in $\Phi(\mathbf{x}, *)$, which starts with the 2-prefix (a', b'). We then collect all the $a' \in D$ such that for some $b' \in D$, both $(a', b'), (a, b') \in \mathsf{Pr}_{[2]}\, \Phi(\mathbf{x}, *)$, and claim that this is exactly $\mathcal{E}$.

First, it is easy to check that $(a', b'), (a, b') \in \mathsf{Pr}_{[2]}\, \Phi(\mathbf{x}, *)$ for some b' implies $a \sim'_{i+1} a'$. Conversely, if $a \sim'_{i+1} a'$, then there are $\mathbf{y} \in D^{i-1}$, $c \in D$, $\mathbf{u}_1, \mathbf{u}_2 \in D^{n-i-1}$, such that

$$\mathbf{y} \circ a' \circ c \circ \mathbf{u}_2 \in \Phi \quad \text{and} \quad \mathbf{y} \circ a \circ c \circ \mathbf{u}_1 \in \Phi.$$

Applying the Mal'tsev polymorphism φ on these two vectors together with the one in Equation (31) gives a vector in Φ with prefix $\mathbf{x} \circ a' \circ b$. This implies that $(a', b) \in \mathsf{Pr}_{[2]}\, \Phi(\mathbf{x}, *)$.

We have computed the class $\mathcal{E}$ of a in $\sim'_{i+1}$. Now, for each $a' \in \mathcal{E}$ with $(a', b'), (a, b') \in \mathsf{Pr}_{[2]}\, \Phi(\mathbf{x}, *)$, we can compute two vectors in Φ with prefixes $\mathbf{x} \circ a' \circ b'$ and $\mathbf{x} \circ a \circ b'$. By applying the Mal'tsev polymorphism φ on these two vectors together with the one in Equation (31) gives a vector in Φ with prefix $\mathbf{x} \circ a' \circ b$. As a result, we obtain a witness of $(i+1, a')$ in Φ', for every $a' \in \mathcal{E}$, which shares the same prefix $\mathbf{x} \circ b$. We set $\omega'(i+1, a')$ to be this witness for each $a' \in \mathcal{E}$. This finishes the construction of ω'. □

7.2. Union of Witness Functions

Let $\Psi_1, \ldots, \Psi_s \subseteq D^n$ be s pairwise disjoint relations over n variables $x_1, \ldots, x_n \in D$ for some $s \geq 1$. Assume that they share a Mal'tsev polymorphism φ. Let $\Phi = \Psi_1 \cup \cdots \cup \Psi_s$.

In general φ might not be a Mal'tsev polymorphism of Φ. The following lemma shows that, if it is guaranteed that φ is a Mal'tsev polymorphism of Φ as well, then we can efficiently construct a witness function of Φ from witness functions of the Ψ_k's.

LEMMA 7.2. *Let $\Psi_1, \ldots, \Psi_s$ be s pairwise disjoint and nonempty subsets of D^n, and let $\Phi = \Psi_1 \cup \cdots \cup \Psi_s$. Also assume that φ is a Mal'tsev polymorphism of both Φ and the Ψ_k's. Given a witness function ω_k of Ψ_k for each $k \in [s]$, we can construct a witness function ω of Φ in polynomial time (in s and n).*

PROOF. Pick any pair $(i, a) \in [n] \times D$. We first decide whether there is a vector $\mathbf{x} \in \Phi$ such that $x_i = a$. Since Φ is the union of the Ψ_k's, it suffices to check if $\omega_k(i, a) \neq \perp$ for some $k \in [s]$. If $\omega_k(i, a) = \perp$ for every $k \in [s]$, then we simply set $\omega(i, a) = \perp$; otherwise, we have found a tuple $\mathbf{z} \in \Phi$ such that $z_i = a$.

Next, for each $i \in [n]$, we compute the equivalence relation $\sim_i$ of Φ as follows. Pick any $a \neq b \in D$ for which we have already found witnesses $\mathbf{x}, \mathbf{y}$ in Φ, with $x_i = a$ and $y_i = b$. By Lemma 2.19, we have

$$a \sim_i b \iff \exists\, \mathbf{z} \in \Phi \text{ such that } \mathsf{Pr}_{[i]}\mathbf{z} = (x_1, \ldots, x_{i-1}, b).$$

As Φ is the union of the Ψ_k's, this happens if there exists such a $\mathbf{z} \in \Psi_k$ for some $k \in [s]$. To check whether Ψ_k has such a $\mathbf{z}$, by Lemma 2.24, we can use ω_k and φ to construct a witness function for $\Psi_k(x_1, \ldots, x_{i-1}, b, *)$. Then, Ψ_k has a $\mathbf{z}$ with $\mathsf{Pr}_{[i]}\mathbf{z} = (x_1, \ldots, x_{i-1}, b)$ if and only if the witness function we get is nonempty.

It is clear that the computation of $\sim_i$, $i \in [n]$ gives us a witness function ω for Φ. □

7.3. Splitting a Witness Function

Here, we describe the *inverse* of the union operation described above. The setting is the following. Let $\Phi \subseteq D^n$ be a nonempty relation over n variables, and let $\Psi_1, \ldots, \Psi_s$ be an s-way partition of Φ, for some $s \in [d]$: the Ψ_i's are nonempty, pairwise disjoint, and satisfy $\Phi = \Psi_1 \cup \cdots \cup \Psi_s$. Assume that φ is a Mal'tsev polymorphism of Φ and the Ψ_i's.

At the beginning, we have no information about the Ψ_i's. Even the number s of sets is not given, though we do know that $s \in [d]$. In addition to the Mal'tsev polymorphism φ, the only resources we have are a witness function ω for Φ as well as a black box to query: We can send any $\mathbf{x} \in \Phi$ to the black box and it returns the unique $k \in [s]$ such that $\mathbf{x} \in \Psi_k$. The question is: Can we use ω and the black box to compute the value of s and a witness function ω_k for each Ψ_k in polynomial time and only using polynomially many queries?

In general, we do not know how to solve this problem efficiently. But if the following condition holds then there is an efficient algorithm. Given any permutation π of $[n]$, we use type_π to denote the type map of $(\pi(\Phi), (\pi(\Psi_1), \ldots, \pi(\Psi_s)))$. Recall that

$$\mathsf{type}_\pi(\mathbf{x}) = \left\{k \in [s] : \exists\, \mathbf{y} \in \pi(\Psi_k) \text{ such that } \mathsf{Pr}_{[\ell]}\mathbf{y} = \mathbf{x}\right\}, \quad \text{for all } \mathbf{x} \in D^\ell \text{ with } \ell \in [n].$$

We also have $\mathsf{type}_\pi(\epsilon) = [s]$, where ϵ denotes the empty tuple. Our main lemma of this subsection shows that if $(\pi(\Phi), (\pi(\Psi_1), \ldots, \pi(\Psi_s)))$ satisfies the partition condition for all π, then there is an efficient algorithm to compute the value of s and a witness function for each set Ψ_k.

LEMMA 7.3. *Let $(\Psi_1, \ldots, \Psi_s)$ denote an s-way partition of $\Phi \subseteq D^n$, for some $s \in [d]$. Assume φ is a Mal'tsev polymorphism of Φ and the Ψ_k's, and $(\pi(\Phi), (\pi(\Psi_1), \ldots, \pi(\Psi_s)))$ satisfies the partition condition for all permutations π of $[n]$. Given φ, a witness function ω of Φ, and a black box specified above, we can compute the value of s and a witness function ω_k of each Ψ_k in polynomial time and using polynomially many queries in n.*

We start with some notation and definitions. We use $\mathsf{type}(\cdot)$ to denote the type map of $(\Phi, (\Psi_1, \ldots, \Psi_s))$ and use

$$\mathfrak{T} = (\mathcal{T}_0, \mathcal{T}_1, \ldots, \mathcal{T}_n), \quad \text{where } \mathcal{T}_j = \left\{\mathsf{type}(\mathbf{x}) \subseteq [s] : \mathbf{x} \in \mathsf{Pr}_{[j]}\Phi\right\},$$

to denote its list of types. As $(\Phi, (\Psi_1, \ldots, \Psi_s))$ satisfies the partition condition, we have $|\mathcal{T}_j| \le s \le d$ for all j. It is clear that $\mathcal{T}_j$'s are nonempty as Φ is nonempty; every set in $\mathcal{T}_j$ is nonempty, because we are only interested in $\mathbf{x} \in \mathsf{Pr}_{[j]}\Phi$ in the definition of $\mathcal{T}_j$.

We need the following definition in the algorithm:

Definition 7.4. We say $\mathfrak{S} = (\mathcal{S}_0, \mathcal{S}_1, \ldots, \mathcal{S}_n)$ is a *partial* list of $\mathfrak{T}$ if $\mathcal{S}_j \subseteq \mathcal{T}_j$ for all $0 \le j \le n$. Given $U \in \mathcal{T}_\ell$ for some $0 \le \ell \le n$, we say $\mathfrak{S}$ is *closedwith respect to U at level ℓ* if $U \in \mathcal{S}_\ell$ and for every $j > \ell$, we have $V \in \mathcal{S}_j$ for every $V \in \mathcal{T}_j$ with $V \subseteq U$. Finally, we say $\mathfrak{S}$ is *closed* if it is closed with respect to every $U \in \mathcal{S}_j$ at level j, for all $0 \le j \le n$.

In particular, $\mathfrak{S}$ is closed if $\mathcal{S}_j = \emptyset$ for all j. The next lemma follows from the definition of closed partial lists of $\mathfrak{T}$.

LEMMA 7.5. *If $\mathfrak{S}$ is a closed partial list of $\mathfrak{T}$ and $\mathcal{S}_0$ has the set $[s]$, then $\mathfrak{S} = \mathfrak{T}$.*

PROOF. By definition $\mathfrak{S}$ is closed with respect to $[s]$ at level 0. For each $V \in \mathcal{T}_j$ with $j \ge 1$, we have $V \in \mathcal{S}_j$, since $V \subseteq [s]$. It follows that $\mathcal{T}_j \subseteq \mathcal{S}_j$ for each j and thus, $\mathfrak{S} = \mathfrak{T}$, since $\mathfrak{S}$ is a partial list of $\mathfrak{T}$. □

We present a recursive procedure ComputeType for computing s and the list of types $\mathfrak{T}$, using the witness function ω of Φ and the black box. The procedure is presented in Figure 2. It takes two inputs:

ComputeType$(\mathbf{x}, \mathfrak{S})$, where $\mathbf{x} \in \mathsf{Pr}_{[\ell]}\Phi$ and $0 \le \ell \le n$

0. set $\mathfrak{S}^* = \mathfrak{S}$, where $\mathfrak{S}^* = (\mathcal{S}_0^*, \mathcal{S}_1^*, \ldots, \mathcal{S}_n^*)$
1. if $\ell = n$ then
2. query the black box to get the index $k \in [d]$ such that $\mathbf{x} \in \Psi_k$
3. add $\{k\}$ to $\mathcal{S}_n^*$ if $\{k\} \notin \mathcal{S}_n^*$; output $(\{k\}, \mathfrak{S}^*)$ and exit
4. end if
5. compute a witness function ω' of $\Phi' = \Phi(\mathbf{x}, *) \subseteq D^{n-\ell}$ ($\Phi' = \Phi$ if $\mathbf{x} = \epsilon$ and $\ell = 0$)
6. use ω' to find a vector $\mathbf{y} \in D^{n-\ell}$ such that $\mathbf{x} \circ \mathbf{y} \in \Phi$
7. query the black box to get the index $k \in [d]$ such that $\mathbf{x} \circ \mathbf{y} \in \Psi_k$
8. if k belongs to one of the subsets U in $\mathcal{S}_\ell^*$ then
9. output $(U, \mathfrak{S}^*)$ and exit
10. else
11. use ω' to compute $\mathsf{Pr}_1\Phi' = \{b \in D : \omega'(1, b) \neq \perp\}$
12. for each $a \in \mathsf{Pr}_1\Phi'$
13. let $\mathbf{z} = \omega'(1, a) \in D^{n-\ell}$ (and we have $z_1 = a$ and $\mathbf{x} \circ \mathbf{z} \in \Phi$)
14. query the black box to get the index $k \in [d]$ such that $\mathbf{x} \circ \mathbf{z} \in \Psi_k$
15. if k belongs to one of the subsets in $\mathcal{S}_{\ell+1}^*$ then
16. denote this subset of $\mathcal{S}_{\ell+1}^*$ by U_a (and we have $\mathsf{type}(\mathbf{x} \circ a) = U_a$)
17. else
18. let $(U_a, \mathfrak{S}^*) = \mathsf{ComputeType}(\mathbf{x} \circ a, \mathfrak{S}^*)$
19. end if
20. end for
21. let U be the union of U_a's over $a \in \mathsf{Pr}_1\Phi'$
22. add U to $\mathcal{S}_\ell^*$ if $U \notin \mathcal{S}_\ell^*$; output $(U, \mathfrak{S}^*)$ and exit
23. end if

Fig. 2. The recursive procedure ComputeType.

(i) a vector $\mathbf{x} \in \mathsf{Pr}_{[\ell]}\Phi$, where $0 \le \ell \le n$ (and $\mathbf{x} = \epsilon$ when $\ell = 0$); and
(ii) a *closed partial list* $\mathfrak{S} = (\mathcal{S}_0, \mathcal{S}_1, \ldots, \mathcal{S}_n)$ of $\mathfrak{T}$.

The output of ComputeType is a pair $(U, \mathfrak{S}')$ in which $U \subseteq [s]$ and $\mathfrak{S}'$ is a closed partial list of $\mathfrak{T}$. We first prove a lemma concerning the output of ComputeType and analyze its running time later in the proof of Lemma 7.3.

LEMMA 7.6. *Let* $\mathbf{x} \in \mathsf{Pr}_{[\ell]}\Phi$ *for some* $\ell : 0 \le \ell \le n$ *and let* $\mathfrak{S} = (\mathcal{S}_0, \mathcal{S}_1, \ldots, \mathcal{S}_n)$ *be a closed partial list of* $\mathfrak{T}$. *Then* ComputeType$(\mathbf{x}, \mathfrak{S})$ *outputs* $(U, \mathfrak{S}')$ *such that* $U = \mathsf{type}(\mathbf{x})$ *and* $\mathfrak{S}' = (\mathcal{S}'_0, \mathcal{S}'_1, \ldots, \mathcal{S}'_n)$ *is a closed partial list of* $\mathfrak{T}$ *that satisfies*

$$\mathsf{type}(\mathbf{x}) \in \mathcal{S}'_\ell \ \text{ and } \ \mathcal{S}_j \subseteq \mathcal{S}'_j, \quad \text{for all } j : 0 \le j \le n. \tag{32}$$

PROOF. We prove the lemma by induction on $\ell = n, n-1, \ldots, 1, 0$.

The base case when $\ell = n$ is trivial. This is because, if $\mathfrak{S}$ is closed at the beginning, then $\mathfrak{S}^*$ is also a closed partial list of $\mathfrak{T}$ after adding a singleton set $\{k\} \in \mathcal{T}_n$ to $\mathcal{S}_n^*$.

Now assume that the lemma holds for all calls to ComputeType with an $\mathbf{x}$ of length at least $\ell + 1$ and any closed partial list $\mathfrak{S}$. We show that if $\mathbf{x} \in \mathsf{Pr}_{[\ell]}\Phi$ and $\mathfrak{S}$ is a closed partial list of $\mathfrak{T}$, then ComputeType$(\mathbf{x}, \mathfrak{S})$ outputs $(U, \mathfrak{S}')$, where $\mathsf{type}(\mathbf{x}) = U$ and $\mathfrak{S}'$ is a closed partial list of $\mathfrak{T}$ that satisfies Equation (32).

There are two cases to discuss. First, if the algorithm reaches line 9 then we clearly have $\mathsf{type}(\mathbf{x}) = U$ as $(\Phi, (\Psi_1, \ldots, \Psi_s))$ satisfies the partition condition and $\mathfrak{S}$ is assumed to be a partial list of $\mathfrak{T}$. Properties about $\mathfrak{S}'$ hold, because $\mathfrak{S}' = \mathfrak{S}$ in this case.

Otherwise the algorithm uses a for-loop to get U_a for each $a \in \mathsf{Pr}_1\Phi'$. By the inductive hypothesis, we know at the end of each iteration of lines 12–20, $\mathfrak{S}^*$ remains a closed partial list of $\mathfrak{T}$ and satisfies $\mathcal{S}_j \subseteq \mathcal{S}_j^*$ for all j. After the for-loop, we have $U_a = \mathsf{type}(\mathbf{x} \circ a)$ and $\mathfrak{S}^*$ is a closed partial list with $\mathsf{type}(\mathbf{x} \circ a) \in \mathcal{S}_{\ell+1}^*$ for all $a \in \mathsf{Pr}_1\Phi'$ and $\mathcal{S}_j \subseteq \mathcal{S}_j^*$ for all j.

Let $(U, \mathfrak{S}')$ denote the output of $\mathsf{ComputeType}(\mathbf{x}, \mathfrak{S})$. By line 18 and line 21, we have

$$U = \bigcup_{a \in \mathsf{Pr}_1\Phi'} U_a = \bigcup_{a \in \mathsf{Pr}_1\Phi'} \mathsf{type}(\mathbf{x} \circ a) = \mathsf{type}(\mathbf{x}).$$

It is easy to show that $\mathfrak{S}'$ is a partial list of $\mathfrak{T}$ that satisfies Equation (32) (since $U \in \mathcal{S}_\ell^*$ by line 22). To see that $\mathfrak{S}'$ is closed, note that by the inductive hypothesis $\mathfrak{S}^*$ in the procedure remains closed before line 22 and we have $\mathsf{type}(\mathbf{x} \circ a) \in \mathcal{S}_{\ell+1}^*$ for all $a \in \mathsf{Pr}_1\Phi'$. Therefore before and after line 22, $\mathfrak{S}^*$ is closed with respect to $\mathsf{type}(\mathbf{x} \circ a)$ at level $\ell + 1$, for all such a. Note that these are all the subsets of $\mathsf{type}(\mathbf{x})$ in $\mathcal{T}_{\ell+1}$. It follows that $\mathfrak{S}^*$ remains closed after line 22, since it is closed with respect to $\mathsf{type}(\mathbf{x})$ at level ℓ.

This finishes the induction and the proof of the lemma. □

Proof of Lemma 7.3. Using Lemma 7.6, we can call $\mathsf{ComputeType}(\epsilon, \mathfrak{S}) = (U, \mathfrak{S}')$, with $\mathcal{S}_j = \emptyset$ in $\mathfrak{S}$ for all j, to get the number $s \in [d]$ of Ψ_k's, since $U = \mathsf{type}(\epsilon) = [s]$. By Lemma 7.6, we also have $\mathsf{type}(\epsilon) \in \mathcal{S}_0'$ and $\mathfrak{S}'$ remains a closed partial list of $\mathfrak{T}$. It then follows from Lemma 7.5 that $\mathfrak{S}' = \mathfrak{T}$.

Next, we show that $\mathsf{ComputeType}(\epsilon, \mathfrak{S})$ runs in polynomial time, and only uses polynomially many queries to the black box. Notice that the running time and number of queries used in each call to ComputeType, excluding those spent in the recursive calls in line 18, are bounded by a polynomial in n.

We now prove the following claim: at the end of each recursive call to ComputeType in line 18, at least one new set is added to $\mathcal{S}_{\ell+1}^*$ in $\mathfrak{S}^*$. This is because each recursive call to ComputeType in line 18 has the following property: the index k obtained in line 14 belongs to $\mathsf{type}(\mathbf{x} \circ a)$ by the choice of $\mathbf{z}$ in line 13 and the definition of k in line 14. The fact that we reach line 18 means that the condition in line 15 fails and thus, k is not in any set in $\mathcal{S}_{\ell+1}^*$ before the execution of ComputeType. After the recursive call, $\mathsf{type}(\mathbf{x} \circ a)$, which contains k, is in the updated $\mathcal{S}_{\ell+1}^*$ by Lemma 7.6. The claim follows.

As a result, each recursive call of ComputeType in line 18 strictly increases the cardinality of $\mathcal{S}_\ell^*$ in $\mathfrak{S}^*$ if its first input is a tuple of length ℓ. But we also have

$$\sum_{i=0}^{n} |\mathcal{T}_i| \leq 1 + dn = O(n),$$

since $|\mathcal{T}_i| \leq d$ for every $i \in [n]$ and $|\mathcal{T}_0| = 1$. Hence, there can be at most $O(n)$ recursive calls in every execution of $\mathsf{ComputeType}(\mathbf{x}, \mathfrak{S})$. We conclude that the running time as well as the number of queries to the black box used by $\mathsf{ComputeType}(\epsilon, \mathfrak{S})$, where $\mathcal{S}_j$ in $\mathfrak{S}$ is $\emptyset$ for all j, are both polynomial in n.

We have computed $s \in [d]$ and $\mathfrak{T}$. Given $\mathfrak{T}$, we can compute $\mathsf{type}(\mathbf{x})$ for any $\mathbf{x} \in \mathsf{Pr}_{[\ell]}\Phi$ in polynomial time. The algorithm is presented in Figure 3. As $(\Phi, (\Psi_1, \ldots, \Psi_s))$ satisfies the partition condition, by the definition of $\mathfrak{T}$, we know, in line 3, there is a unique $U \in \mathcal{T}_\ell$ such that $k \in U$, and we have $\mathsf{type}(\mathbf{x}) = U$. Furthermore, given any $\mathbf{x} \in \mathsf{Pr}_{[\ell]}\Phi$ and $k \in \mathsf{type}(\mathbf{x})$, we can find recursively a $\mathbf{y}$ such that $\mathbf{x} \circ \mathbf{y} \in \Psi_k$ in polynomial time. The algorithm is described in Figure 4.

Given $\mathbf{x} \in \mathsf{Pr}_{[\ell]}\Phi$ and $\mathfrak{T}$, compute $\mathsf{type}(\mathbf{x})$:

1. use ω to find a vector $\mathbf{y} \in D^{n-\ell}$ such that $\mathbf{x} \circ \mathbf{y} \in \Phi$
2. query the black box to get the index $k \in [s]$ such that $\mathbf{x} \circ \mathbf{y} \in \Psi_k$
3. use k to find a (unique) set U in $\mathcal{T}_\ell$ that contains k; output U and exit

Fig. 3. Computation of $\mathsf{type}(\mathbf{x})$ using $\mathfrak{T}$.

Given $\mathbf{x} \in \mathsf{Pr}_{[\ell]}\Phi$ and $k \in \mathsf{type}(\mathbf{x})$, find a $\mathbf{y}$ such that $\mathbf{x} \circ \mathbf{y} \in \Psi_k$:

1. for every $a \in D$ such that $\mathbf{x} \circ a \in \mathsf{Pr}_{[\ell+1]}\Phi$
2. compute $\mathsf{type}(\mathbf{x} \circ a)$
3. find an $a \in D$ such that $k \in \mathsf{type}(\mathbf{x} \circ a)$
4. recursively find a $\mathbf{z}$ with $\mathbf{x} \circ a \circ \mathbf{z} \in \Psi_k$; output $a \circ \mathbf{z}$ and exit

Fig. 4. Finding a $\mathbf{y}$ such that $\mathbf{x} \circ \mathbf{y} \in \Psi_k$, where $k \in \mathsf{type}(\mathbf{x})$.

Let π be any permutation of $[n]$. We use $\mathsf{type}_\pi(\cdot)$ to denote the type map of the pair $(\pi(\Phi), (\pi(\Psi_1), \ldots, \pi(\Psi_s)))$, which also satisfies the partition condition. We note that all the algorithms in Figures 2, 3, and 4 still work after we replace $\mathsf{type}(\cdot)$ by $\mathsf{type}_\pi(\cdot)$ and replace the witness function ω of Φ by a witness function ω_π of $\pi(\Phi)$. Also note that ω_π can be computed from ω efficiently using Lemma 7.1.

Now, for any $k \in [s]$, we show how to compute a witness function ω_k for Ψ_k as follows. Pick a pair (i, a) with $i \in [n]$ and $a \in D$. Let π denote a permutation of $[n]$ with $\pi(i) = 1$. Using algorithms in Figures 2 and 3, we can compute $\mathsf{type}_\pi(a)$. We then use $\mathsf{type}_\pi(a)$ to decide if $a \in \mathsf{Pr}_i\, \Psi_k$ as follows. If $k \in \mathsf{type}_\pi(a)$, then $a \in \mathsf{Pr}_i\, \Psi_k$, and we use the algorithm in Figure 4 to find a witness in Ψ_k for (i, a); otherwise, we know no such witness exists in Ψ_k and set $\omega_k(i, a) = \perp$.

To derive the equivalence relation $\sim_{i,k}$ defined by Ψ_k for the ith coordinate, we pick $a, b \in \mathsf{Pr}_i \Psi_k$ and then use $\mathbf{x}, \mathbf{y} \in \Psi_k$ to denote the witnesses in Ψ_k that we have found for (i, a) and (i, b). We follow the algorithm in Figure 3 to check if

$$k \in \mathsf{type}\big(\big(\mathsf{Pr}_{[i-1]}\,\mathbf{x}\big) \circ b\big). \tag{33}$$

We show that $a \sim_{i,k} b$ if and only if Equation (33) holds, and this gives us the relation $\sim_{i,k}$. Here the "if" part is trivial, and the "only if" part follows from Lemma 2.19.

Finally, for each $b \sim_{i,k} a$, we can also use the algorithm in Figure 4 to find a vector $\mathbf{x}'$ such that $(\mathsf{Pr}_{[i-1]}\mathbf{x}) \circ b \circ \mathbf{x}' \in \Psi_k$. This finishes the construction of ω_k and the proof. □

8. PROOF OF LEMMA 3.8

We prove Lemma 3.8 in this section.

Recall that I is an input instance of $\#\mathsf{CSP}(\mathcal{F})$ and F is the n-ary function it defines. For each $\ell : 2 \le \ell \le n$, let $\Phi_\ell = \mathsf{Boolean}(F^{[\ell]})$. We also use

$$\big\{(S^{[\ell,j]}, \mathbf{v}^{[\ell,j]}) : j \in [s_\ell]\big\}$$

to denote the row representation of $F^{[\ell]}$, for some $s_\ell \le d$.

We show how to compute the value of s_ℓ, a witness function ω_ℓ for Φ_ℓ, and

$$\big\{(\omega^{[\ell,j]}, \mathbf{v}^{[\ell,j]}) : j \in [s_\ell]\big\}$$

in polynomial time for all $2 \le \ell \le n$, such that $\omega^{[\ell,j]}$ is a witness function of $S^{[\ell,j]}$ for all ℓ and j. Here it makes sense to talk about witness functions for Φ_ℓ and $S^{[\ell,j]}$, since by the Mal'tsev condition and Lemma 3.7, they share φ as a Mal'tsev polymorphism.

We use induction on ℓ from n to 2, and we start with the base case when $\ell = n$. Let $\mathcal{F} = \{F_1, \ldots, F_h\}$ and let φ denote a Mal'tsev polymorphism shared by relations in $\Lambda_{\mathcal{F}}$ and thus, φ is a Mal'tsev polymorphism of $\{\mathsf{Boolean}(F_1), \ldots, \mathsf{Boolean}(F_h)\}$. Therefore, by Theorem 2.25, we can construct a witness function ω_n in polynomial time for

$$\Phi_n = \mathsf{Boolean}(F^{[n]}) = \mathsf{Boolean}(F),$$

since $\Phi_n = \mathsf{Boolean}(F)$ is the relation defined by an input instance of the unweighted

$$\#\mathsf{CSP}(\mathsf{Boolean}(F_1), \ldots, \mathsf{Boolean}(F_h))$$

obtained from I by replacing each F_i by $\mathsf{Boolean}(F_i)$.

Given ω_n, we use Lemma 2.23 to construct a witness function ω'_n for $\Psi_n = \mathsf{Pr}_{[n-1]}\Phi_n$. Using the row representation of $F = F^{[n]}$, we have

$$\Psi_n = \bigcup_{j \in [s_n]} S^{[n,j]}.$$

Hence, $S^{[n,1]}, \ldots, S^{[n,s_n]}$ form an s_n-way partition of Ψ_n, and they share φ as a Mal'tsev polymorphism. By the Type Partition condition, we have that

$$\left(\pi(\Psi_n), \left(\pi\left(S^{[n,1]}\right), \ldots, \pi\left(S^{[n,s_n]}\right)\right)\right)$$

satisfies the partition condition for any permutation π of $[n-1]$. This follows from the fact that, given any function in $\mathcal{W}_{\mathcal{F}}$, we can arbitrarily permute its variables and the new function still belongs to $\mathcal{W}_{\mathcal{F}}$.

As a result, we can now apply Lemma 7.3 to compute the value of s_n and a witness function $\omega^{[n,j]}$ for each $S^{[n,j]}$, $j \in [s_n]$. Notice that the black box that Lemma 7.3 needs to query can be implemented quite trivially here: given any $\mathbf{x} \in D^{n-1}$, we can evaluate the vector $F(\mathbf{x}, *)$ efficiently, entry by entry, using the input instance I. The black box keeps all the linearly independent vectors $F(\mathbf{x}, *)$ evaluated so far and associates each of them with a unique label $j \in [s_n]$. With $\omega^{[n,j]}$ computed, we can next use it to obtain a vector $\mathbf{x} \in S^{[n,j]}$ and then evaluate $F(\mathbf{x}, *)$ to get the representative vector $\mathbf{v}^{[n,j]}$.

Assume for induction that for some $\ell : 2 \le \ell < n$, we have already computed $s_t \in [d]$, a witness function of Φ_t, and

$$\left\{\left(\omega^{[t,j]}, \mathbf{v}^{[t,j]}\right) : j \in [s_t]\right\}, \quad \text{for all } t = \ell + 1, \ldots, n,$$

such that $\omega^{[t,j]}$ is a witness function for $S^{[t,j]}$. To work on $F^{[\ell]}$, we first notice that

$$F^{[\ell]}(\mathbf{x}) = \sum_{a \in D} F^{[\ell+1]}(\mathbf{x}, a).$$

As a result, we have $F^{[\ell]}(\mathbf{x}) \neq 0$ if and only if $\mathbf{x} \in S^{[\ell+1,j]}$ for some $j \in [s_{\ell+1}]$ and

$$\sum_{a \in D} v_a^{[\ell+1,j]} \neq 0.$$

Let L denote the subset of $[s_{\ell+1}]$ such that $j \in L$ if the sum above is nonzero. Then,

$$\Phi_\ell = \mathsf{Boolean}\left(F^{[\ell]}\right) = \bigcup_{j \in L} S^{[\ell+1,j]}.$$

Using the Mal'tsev condition and Lemma 3.7, we also know that φ is a Mal'tsev polymorphism of Φ_ℓ and the $S^{[\ell+1,j]}$'s. By using Lemma 7.2 as well as the witness functions $\omega^{[\ell+1,j]}$ for $S^{[\ell+1,j]}$, we can compute a witness function ω_ℓ of Φ_ℓ efficiently.

Next, we use ω_ℓ and Lemma 2.23 to compute a witness function ω'_ℓ for $\Psi_\ell = \mathsf{Pr}_{[\ell-1]}\Phi_\ell$, a relation over $\ell - 1$ variables. Because

$$\left\{\left(S^{[\ell,j]}, \mathbf{v}^{[\ell,j]}\right) : j \in [s_\ell]\right\}$$

is the row representation of $F^{[\ell]}$, we have

$$\Psi_\ell = \bigcup_{j \in [s_\ell]} S^{[\ell,j]}.$$

Similarly, $S^{[\ell,1]}, \ldots, S^{[\ell,s_\ell]}$ form an s_ℓ-way partition of Ψ_ℓ and they share φ as a Mal'tsev polymorphism. By the Type Partition condition, we have that

$$\left(\pi(\Psi_\ell), \left(\pi\left(S^{[\ell,1]}\right), \ldots, \pi\left(S^{[\ell,s_\ell]}\right)\right)\right)$$

satisfies the partition condition for any permutation π of $[\ell - 1]$.

But before we can apply the algorithm of Lemma 7.3 to compute the value of s_ℓ and a witness function $\omega^{[\ell,j]}$ for each $S^{[\ell,j]}$ in the row representation, we need to first show how to implement the black box efficiently. To this end, it suffices to give an efficient algorithm for computing $F^{[\ell]}(\mathbf{x})$ given $\mathbf{x} \in D^\ell$.

This can be done by calling ComputeF$(\ell, \mathbf{x})$, the polynomial-time algorithm described in the proof of Lemma 3.9 in Figure 1. Note that the execution of ComputeF$(\ell, \mathbf{x})$ only uses $s_{\ell+1}, \ldots, s_n$ and the pairs

$$\left\{\left(\omega^{[t,j]}, \mathbf{v}^{[t,j]}\right) : \ell + 1 \le t \le n \text{ and } j \in [s_t]\right\},$$

all of which have already been computed by the inductive hypothesis. Now, we can use the algorithm in Lemma 7.3 to compute the value of s_ℓ and the pairs $(\omega^{[\ell,j]}, \mathbf{v}^{[\ell,j]})$.

This finishes the induction and Lemma 3.8 is proven.

9. CONCLUSIONS

We proved a complexity dichotomy theorem for #CSP with algebraic complex weights. To this end, we introduced three criteria over the language $\mathcal{F}$: the Block Orthogonality condition, the Type Partition condition, and the Mal'tsev condition. We show that #CSP$(\mathcal{F})$ is #P-hard if $\mathcal{F}$ violates any of these three conditions and give a polynomial-time algorithm for #CSP$(\mathcal{F})$ when all three conditions are satisfied. This is the culmination of a long series of important results by many researchers in the field.

One open question is the decidability of these dichotomy criteria. Note that all the dichotomies discussed in the introduction are known to be decidable in NP, with many of them decidable in polynomial time. From the definitions of our dichotomy criterion, each of the three conditions requires one to check a property over an infinitary object. While it is often the case that in certain related problems, properties stated for $\mathcal{F}$ can be shown to automatically carry over to its "closure" $\mathcal{W}_\mathcal{F}$, this does not seem to be the case for our dichotomy criterion, due to the nature of cancellations in the presence of complex weights. (For example, $\Lambda_\mathcal{F}$ in general may not satisfy the Mal'tsev condition even if the unweighted version Γ of $\mathcal{F}$ has a Mal'tsev polymorphism, making it siganificantly different from the unweighted case.) Given a finite language $\mathcal{F}$ as the input, can we decide whether $\mathcal{F}$ satisfies our tractability criterion or not in finite time? If so, then can we further put the decision problem in NP?

APPENDIXES: BASIC OPERATIONS ON WITNESS FUNCTIONS

We include below proofs of Lemma 2.21–2.24 for completeness, which may help readers get more familiar with notions of Mal'tsev polymorphisms and witness functions.

LEMMA A.1 (MEMBERSHIP). *Let $\Phi \subseteq D^n$ be an n-ary relation that has a Mal'tsev polymorphism. With ω, a witness function of Φ, and φ, a Mal'tsev polymorphism of Φ, we can solve the following problem in time polynomial in n: given an $\mathbf{x} \in D^n$, decide if $\mathbf{x} \in \Phi$ or not.*

PROOF. To decide if $\mathbf{x} = (x_1, \ldots, x_n) \in \Phi$, we first check if $\omega(1, x_1) = \perp$. If so, then by the definition of witness functions, we have $\mathbf{x} \notin \Phi$; otherwise, we get a vector $\omega(1, x_1) \in \Phi$ that starts with x_1.

Assume for induction that we have found a vector $\mathbf{y} \in \Phi$ that has the same k-prefix as $\mathbf{x}$, for some $k : 1 \leq k < n$. We show how to use it to either find a vector in Φ that has the same $(k+1)$-prefix as $\mathbf{x}$ or prove that $\mathbf{x} \notin \Phi$. By repeating the procedure described below, we can decide if $\mathbf{x} \in \Phi$ by the end.

For this purpose, we check if $y_{k+1} \sim_{k+1} x_{k+1}$ using ω, that is, check if $\omega(k+1, y_{k+1}) \neq \perp$ and $\omega(k+1, x_{k+1}) \neq \perp$ share the same k-prefix. If $y_{k+1} \not\sim_{k+1} x_{k+1}$, then by the definition of witness functions and $\mathbf{y} \in \Phi$, we have $\mathbf{x} \notin \Phi$ and we are done; otherwise, applying the Mal'tsev polymorphism φ on $\omega(k+1, x_{k+1})$, $\omega(k+1, y_{k+1})$, and $\mathbf{y}$ gives us a vector in Φ that has the same $(k+1)$-prefix as $\mathbf{x}$. This finishes the proof of the lemma, since the algorithm described above is polynomial-time. □

Let $S = \{\omega(i, a) : \omega(i, a) \neq \perp, i \in [n] \text{ and } a \in D\}$ denote the image of ω. The proof of Lemma A.1 implies that every $\mathbf{x} \in \Phi$ is in the closure of S under φ, and thus,

COROLLARY A.2. $\Phi = \mathsf{cl}_\varphi S$.

Let $I \subseteq [n]$ denote a set of indices and let Pr_I denote the projection on coordinates in I. Then Corollary A.2 implies that

$$\mathsf{Pr}_I \Phi = \mathsf{Pr}_I \mathsf{cl}_\varphi S = \mathsf{cl}_\varphi \mathsf{Pr}_I S,$$

which in turns gives us the following useful corollary:

COROLLARY A.3. *Let φ be a Mal'tsev polymorphism and ω be a witness function of $\Phi \subseteq D^n$. Given φ, ω and $I \subseteq [n]$, where $|I|$ is bounded by a constant, we can compute the set $\mathsf{Pr}_I \Phi$ itself as well as a witness for each vector in $\mathsf{Pr}_I \Phi$, that is, a vector $\mathbf{y} \in \Phi$ for each $\mathbf{x} \in \mathsf{Pr}_I \Phi$ such that $\mathbf{x} = \mathsf{Pr}_I \mathbf{y}$, in time polynomial in n.*

PROOF. We first use ω to obtain $S = \{\omega(i, a) : \omega(i, a) \neq \perp, i \in [n] \text{ and } a \in D\}$. We then enumerate all triples of vectors from S to see if applying the Mal'tsev polymorphism φ on them gives a new vector $\mathbf{y} \in \Phi$ such that $\mathsf{Pr}_I \mathbf{y} \notin \mathsf{Pr}_I S$. If so, then add $\mathbf{y}$ to S and repeat. When this process stops, we must have $\mathsf{Pr}_I S = \mathsf{Pr}_I \Phi$ by Corollary A.2, and every vector in $\mathsf{Pr}_I \Phi$ has a witness in S by the end. As $|\mathsf{Pr}_I \Phi| \leq d^{|I|}$ is bounded by a constant, this process takes only a constant number of rounds to stop. This finishes the proof. □

LEMMA A.4. *Let φ be a Mal'tsev polymorphism of $\Phi \subseteq D^n$. Let $\ell \in [n]$, $\mathbf{a} \in D^\ell$, and π be a permutation of $[n]$. Then φ is a Mal'tsev polymorphism of $\mathsf{Pr}_{[\ell]} \Phi$, $\Phi(\mathbf{a}, *)$, and $\pi(\Phi)$.*

PROOF. We start with the projection $\mathsf{Pr}_{[\ell]} \Phi$. Let $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathsf{Pr}_{[\ell]} \Phi$. By definition, there exist $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \Phi$ such that $\mathbf{u} = \mathsf{Pr}_{[\ell]} \mathbf{x}$, $\mathbf{v} = \mathsf{Pr}_{[\ell]} \mathbf{y}$, and $\mathbf{w} = \mathsf{Pr}_{[\ell]} \mathbf{z}$.

Applying the Mal'tsev polymorphism φ of Φ on $\mathbf{x}, \mathbf{y}$, and $\mathbf{z}$ gives us a vector in Φ, with its prefix being the same as what one gets from applying φ on $\mathbf{u}, \mathbf{v}$, and $\mathbf{w}$. This shows that φ is a Mal'tsev polymorphism of $\mathsf{Pr}_{[\ell]} \Phi$ as well.

Next, for $\Phi(\mathbf{a}, *)$, let $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \Phi(\mathbf{a}, *)$. By definition, there exist $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \Phi$ such that $\mathbf{x} = \mathbf{a} \circ \mathbf{u}$, $\mathbf{y} = \mathbf{a} \circ \mathbf{v}$, and $\mathbf{z} = \mathbf{a} \circ \mathbf{w}$. Applying the Mal'tsev polymorphism φ on $\mathbf{x}, \mathbf{y}$ and $\mathbf{z}$ gives a vector in Φ with the same prefix $\mathbf{a}$ (since $\varphi(a, a, a) = a$). This shows that φ is also a Mal'tsev polymorphism of $\Phi(\mathbf{a}, *)$.

The proof for the permutation operation $\pi(\Phi)$ is trivial. □

LEMMA A.5 (PROJECTION). *Let φ be a Mal'tsev polymorphism, and ω be a witness function of $\Phi \subseteq D^n$. Given an $\ell \in [n]$, we can construct a witness function for $\mathsf{Pr}_{[\ell]} \Phi$ in time polynomial in n. Moreover, given an $\mathbf{x} \in \mathsf{Pr}_{[\ell]} \Phi$ for some $\ell \in [n]$, we can compute a*

vector $\mathbf{y} \in \Phi$ *with* $\mathbf{x} = \mathsf{Pr}_{[\ell]}\mathbf{y}$ *in polynomial time. When* ℓ *is bounded by a constant, we can use* ω *to compute the projection* $\mathsf{Pr}_{[\ell]}\Phi$ *itself in polynomial time.*

PROOF. Let $\ell \in [n]$. We set $\omega'(i,a) = \perp$ if $\omega(i,a) = \perp$, and $\omega'(i,a) = \mathsf{Pr}_{[\ell]}\omega(i,a) \in D^\ell$, otherwise, for all $i \in [\ell]$ and $a \in D$. It then follows from definition that ω' is a witness function of $\mathsf{Pr}_{[\ell]}\Phi$. When ℓ is bounded by a constant, one can use the algorithm given in Corollary A.3 to compute the set $\mathsf{Pr}_{[\ell]}\Phi$ itself, by setting $I = [\ell]$.

Finally, given $\mathbf{x} \in \mathsf{Pr}_{[\ell]}\Phi$ (which can be verified efficiently using a witness function of $\mathsf{Pr}_{[\ell]}\Phi$), we inductively compute a vector in $\mathsf{Pr}_{[k]}\Phi$ that has prefix $\mathbf{x}$, for $k = \ell, \ldots, n$, as follows. Given $\mathbf{y} \in \mathsf{Pr}_{[k]}\Phi$ with prefix $\mathbf{x}$ for some $k : \ell \le k < n$, we first compute a witness function ω^* of $\mathsf{Pr}_{[k+1]}\Phi$. Then for each $a \in D$, we use ω^* to check if $\mathbf{y} \circ a \in \mathsf{Pr}_{[k+1]}\Phi$, by using Lemma 2.21. Because $\mathbf{y} \in \mathsf{Pr}_{[k]}\Phi$, there must exist at least one $a \in D$ such that $\mathbf{y} \circ a \in \mathsf{Pr}_{[k+1]}\Phi$. This finishes the induction step, and proof of the lemma. □

LEMMA A.6 (PINNING). *Let* φ *be a Mal'tsev polymorphism and* ω *be a witness function of* $\Phi \subseteq D^n$. *Given any* $\mathbf{a} \in D^\ell$ *for some* $\ell \in [n]$, *we can construct a witness function for* $\Phi(\mathbf{a}, *)$ *in time polynomial in* n.

PROOF. First, it suffices to give an algorithm for computing a witness function ω' of $\Phi(a, *)$ for a given $a \in D$. To this end, we first decide for each $k \in [n-1]$ and $b \in D$ whether $b \in \mathsf{Pr}_k\Phi(a, *)$. This can be done by using the algorithm given in Corollary A.3 (setting $I = \{1, k+1\}$). When $b \in \mathsf{Pr}_k\Phi(a, *)$ the algorithm of Corollary A.3 also finds a witness vector $\mathbf{x} \in \Phi$ with $x_1 = a$ and $x_{k+1} = b$.

Let $\sim_k$ and $\sim'_k$ denote the equivalence relations induced by Φ and $\Phi(a, *)$, for the kth component, respectively. It is easy to show that for any $k \in [n-1]$ and $b, c \in \mathsf{Pr}_k\Phi(a, *)$, $b \sim'_k c$ iff $b \sim_{k+1} c$. Given $b, c \in \mathsf{Pr}_k\Phi(a, *)$ with $b \sim'_k c$ and a vector $\mathbf{x} \in \Phi$ with $x_1 = a$ and $x_{k+1} = b$, applying the Mal'tsev polymorphism φ on $\mathbf{x}$, $\omega(k+1, b)$ and $\omega(k+1, c)$ gives us a vector $\mathbf{y} \in \Phi$ such that $\mathbf{y}$ has the same k-prefix as $\mathbf{x}$ and $y_{k+1} = c$. We can use this procedure to compute a witness function ω' of $\Phi(a, *)$ using ω. □

ACKNOWLEDGMENTS

We thank the anonymous referees for many insightful suggestions that greatly helped us improve the presentation of the article.

REFERENCES

L. Barto, M. Kozik, and T. Niven. 2009. The CSP dichotomy holds for digraphs with no sources and no sinks (A positive answer to a conjecture of Bang-Jensen and Hell). *SIAM J. Comput.* 38, 5 (2009), 1782–1802.

R. J. Baxter. 1982. *Exactly Solved Models in Statistical Mechanics*. Academic Press, London.

A. A. Bulatov. 2006. A dichotomy theorem for constraints on a three-element set. *J. ACM* 53, 1 (2006), 66–120.

A. A. Bulatov. 2008. The complexity of the counting constraint satisfaction problem. In *Proceedings of the 35th International Colloquium on Automata, Languages and Programming*. 646–661.

A. A. Bulatov. 2013. The complexity of the counting constraint satisfaction problem. *J. ACM* 60, 5 (2013), 34:1–34:41.

A. A. Bulatov and V. Dalmau. 2006. A simple algorithm for Mal'tsev constraints. *SIAM J. Comput.* 36, 1 (2006), 16–27.

A. A. Bulatov and V. Dalmau. 2007. Towards a dichotomy theorem for the counting constraint satisfaction problem. *Inform. Comput.* 205, 5 (2007), 651–678.

A. A. Bulatov, M. E. Dyer, L. A. Goldberg, M. Jalsenius, M. R. Jerrum, and D. Richerby. 2012. The complexity of weighted and unweighted #CSP. *J. Comput. System Sci.* 78, 2 (2012), 681–688.

A. A. Bulatov and M. Grohe. 2005. The complexity of partition functions. *Theor. Comput. Sci.* 348, 2 (2005), 148–186.

J.-Y. Cai and X. Chen. 2010. A decidable dichotomy theorem on directed graph homomorphisms with non-negative weights. In *Proceedings of the 51st Annual IEEE Symposium on Foundations of Computer Science*. 437–446.

J.-Y. Cai, X. Chen, R. Lipton, and P. Lu. 2010. On tractable exponential sums. In *Proceedings of the 4th International Frontiers of Algorithmics Workshop*. 148–159.

J.-Y. Cai, X. Chen, and P. Lu. 2016. Nonnegative weighted #CSP: An effective complexity dichotomy. *SIAM J. Comput.* 45, 6 (2016), 2177–2198.

J.-Y. Cai, X. Chen, and P. Lu. 2013. Graph homomorphisms with complex values: A dichotomy theorem. *SIAM J. Comput.* 42, 3 (2013), 924–1029.

N. Creignou and M. Hermann. 1996. Complexity of generalized satisfiability counting problems. *Inform. Comput.* 125, 1 (1996), 1–12.

N. Creignou, S. Khanna, and M. Sudan. 2001. *Complexity Classications of Boolean Constraint Satisfaction Problems*. SIAM Monographs on Discrete Mathematics and Applications.

M. E. Dyer, L. A. Goldberg, and M. Paterson. 2007. On counting homomorphisms to directed acyclic graphs. *J. ACM* 54, 6 (2007), Article 27.

M. Dyer and C. Greenhill. 2000. The complexity of counting graph homomorphisms. *Random Structures and Algorithms* 17, 3–4 (2000), 260–289.

M. E. Dyer and D. Richerby. 2013. An effective dichotomy for the counting constraint satisfaction problem. *SIAM J. Comput.* 42, 3 (2013), 1245–1274.

T. Feder and M. Vardi. 1999. The computational structure of monotone monadic SNP and constraint satisfaction: A study through Datalog and group theory. *SIAM J. Comput.* 28, 1 (1999), 57–104.

L. A. Goldberg, M. Grohe, M. Jerrum, and M. Thurley. 2010. A complexity dichotomy for partition functions with mixed signs. *SIAM J. Comput.* 39, 7 (2010), 3336–3402.

P. Hell and J. Nešetřil. 1990. On the complexity of H-coloring. *J. Combinat. Theory, Ser. B* 48, 1 (1990), 92–110.

P. Hell and J. Nešetřil. 2008. Colouring, constraint satisfaction, and complexity. *Comput. Sci. Rev.* 2, 3 (2008), 143–163.

R. E. Ladner. 1975. On the structure of polynomial time reducibility. *J. ACM* 22, 1 (1975), 155–171.

H. W. Lenstra. 1992. Algorithms in algebraic number theory. *Bull. Amer. Math. Soc.* 26, 2 (1992).

T. J. Schaefer. 1978. The complexity of satisfiability problems. In *Proceedings of the 10th Annual ACM Symposium on Theory of Computing*. 216–226.

M. Thurley. 2009. *The Complexity of Partition Functions. PhD Thesis, Humboldt Universitat zu Berlin.*

Received February 2014; revised August 2015; accepted Septemebr 2015

Paper 6

SIAM J. COMPUT.
Vol. 42, No. 3, pp. 924–1029

GRAPH HOMOMORPHISMS WITH COMPLEX VALUES: A DICHOTOMY THEOREM*

JIN-YI CAI†, XI CHEN‡, AND PINYAN LU§

Abstract. Each symmetric matrix $\mathbf{A}$ over $\mathbb{C}$ defines a graph homomorphism function $Z_{\mathbf{A}}(\cdot)$ on undirected graphs. The function $Z_{\mathbf{A}}(\cdot)$ is also called the partition function from statistical physics, and can encode many interesting graph properties, including counting vertex covers and k-colorings. We study the computational complexity of $Z_{\mathbf{A}}(\cdot)$ for arbitrary symmetric matrices $\mathbf{A}$ with algebraic complex values. Building on work by Dyer and Greenhill [*Random Structures and Algorithms*, 17 (2000), pp. 260–289], Bulatov and Grohe [*Theoretical Computer Science*, 348 (2005), pp. 148–186], and especially the recent beautiful work by Goldberg et al. [*SIAM J. Comput.*, 39 (2010), pp. 3336–3402], we prove a complete dichotomy theorem for this problem. We show that $Z_{\mathbf{A}}(\cdot)$ is either computable in polynomial-time or #P-hard, depending explicitly on the matrix $\mathbf{A}$. We further prove that the tractability criterion on $\mathbf{A}$ is polynomial-time decidable.

Key words. computational complexity, counting complexity, graph homomorphisms, partition functions

AMS subject classifications. 68Q17, 68Q25, 68R05, 68R10, 05C31

DOI. 10.1137/110840194

1. Introduction.

Graph homomorphism has been studied intensely over the years [28, 23, 13, 18, 4, 12, 21]. Given two graphs G and H, a graph homomorphism from G to H is a map f from the vertex set $V(G)$ to $V(H)$ such that, whenever (u, v) is an edge in G, $(f(u), f(v))$ is an edge in H. The counting problem for graph homomorphism is to compute the number of homomorphisms from G to H. For a fixed graph H, this problem is also known as the $\#H$-coloring problem. In 1967, Lovász [28] proved that H and H' are isomorphic iff for all G, the number of homomorphisms from G to H and from G to H' are the same. Graph homomorphisms and the associated partition function defined below provide us an elegant and wide-ranging notion of *graph properties* [23].

In this paper, all graphs considered are undirected. We follow standard definitions: G is allowed to have multiple edges; H can have loops, multiple edges, and, more generally, edge weights. (The standard definition of graph homomorphism does not allow self-loops for G. However, our result is stronger: We prove polynomial-time tractability even for input graphs G with self-loops; at the same time, our hardness results hold for the more restricted case of G with no self-loops.) Formally, we use $\mathbf{A}$ to denote an $m \times m$ symmetric matrix with entries $(A_{i,j})$, $i, j \in [m] = \{1, 2, \ldots, m\}$. Given any undirected graph $G = (V, E)$, we define the graph homomorphism function

*Received by the editors July 11, 2011; accepted for publication (in revised form) February 21, 2013; published electronically May 21, 2013.
http://www.siam.org/journals/sicomp/42-3/84019.html

†Department of Computer Sciences, University of Wisconsin–Madison, Madison, WI 53706 (jyc@cs.wisc.edu). This work was supported by NSF CCF-0914969.

‡Department of Computer Science, Columbia University, New York, NY 10027 (xichen@cs.columbia.edu). This work was supported by NSF grants CCF-0832797 and DMS-0635607 when the author was a postdoc at the Institute for Advanced Study and Princeton University, by a USC Viterbi School of Engineering startup fund to Shang-Hua Teng, and by CCF-1149257 and a Sloan research fellowship.

§Microsoft Research Asia, Beijing 100080, China (pinyanl@microsoft.com).

$$Z_{\mathbf{A}}(G) = \sum_{\xi: V \to [m]} \prod_{(u,v) \in E} A_{\xi(u),\xi(v)}. \tag{1.1}$$

This is also called the *partition function* from statistical physics. It is clear from the definition that $Z_{\mathbf{A}}(G)$ is exactly the number of homomorphisms from G to H, when $\mathbf{A}$ is the adjacency matrix of H.

Graph homomorphism can express many natural graph properties. For example, if we take H to be the graph over two vertices $\{0,1\}$ with an edge $(0,1)$ and a loop at 1, then the set of vertices mapped to 1 in a graph homomorphism from G to H corresponds to a vertex cover of G, and the counting problem simply counts the number of vertex covers. As another example, if H is the complete graph over k vertices (without self-loops), then the problem is exactly the k-coloring problem for G. Many additional graph invariants can be expressed as $Z_{\mathbf{A}}(G)$ for appropriate $\mathbf{A}$. Consider the Hadamard matrix

$$\mathbf{H} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}. \tag{1.2}$$

We index its rows and columns by $\{0,1\}$. In the sum $Z_{\mathbf{H}}(G)$, each term is either 1 or -1 and equals -1 precisely when the induced subgraph of G on $\xi^{-1}(1)$ has an odd number of edges. Therefore, $(2^n - Z_{\mathbf{H}}(G))/2$ is the number of induced subgraphs of G with an odd number of edges. Also expressible as $Z_{\mathbf{A}}(\cdot)$ are S-flows, where S is a subset of a finite Abelian group closed under inversion [18], and a scaled version of the Tutte polynomial $\hat{T}(x,y)$, where $(x-1)(y-1)$ is a positive integer. In [18], Freedman, Lovász and Schrijver characterized the graph functions that can be expressed as $Z_{\mathbf{A}}(\cdot)$.

In this paper, we study the complexity of the partition function $Z_{\mathbf{A}}(\cdot)$, where $\mathbf{A}$ is an *arbitrary fixed symmetric matrix over the algebraic complex numbers*. Throughout the paper, we let $\mathbb{C}$ denote the set of algebraic complex numbers and refer to them simply as complex numbers when it is clear from the context. More discussion on the model of computation can be found in section 2.2.

The complexity question of $Z_{\mathbf{A}}(\cdot)$ has been intensely studied. Hell and Nešetřil first studied the H-coloring problem [22, 23] (i.e., given an undirected graph G, decide whether there exists a graph homomorphism from G to H) and proved that for any fixed undirected graph H, the problem is either in polynomial time or NP-complete. Results of this type are called *complexity dichotomy theorems*. Such theorems state that every member of the class of problems concerned is either tractable (i.e., solvable in P) or intractable (i.e., NP-hard or #P-hard depending on whether it is a decision or a counting problem). This includes the well-known Schaefer's dichotomy theorem [31]. The famous complexity dichotomy conjecture made by Feder and Vardi [16] on decision constraint satisfaction problems [11] motivated much of the subsequent work.

In [13], Dyer and Greenhill studied the counting version of the H-coloring problem. They proved that for any fixed symmetric $\{0,1\}$-matrix $\mathbf{A}$, $Z_{\mathbf{A}}(\cdot)$ is either computable in polynomial time or #P-hard. (In this paper, for a function computable in polynomial time we will simply say "in P.") Then in [4], Bulatov and Grohe gave a sweeping generalization of this theorem to all nonnegative symmetric matrices $\mathbf{A}$. (See Theorem 2.5 for the precise statement.) They obtained an elegant dichotomy theorem, which basically says that $Z_{\mathbf{A}}(\cdot)$ is computable in P if each *block* of $\mathbf{A}$ has rank at most one, and is #P-hard otherwise. More precisely, decompose $\mathbf{A}$ as a direct sum of $\mathbf{A}_i$ which correspond to the connected components H_i of the undirected graph H defined by the nonzero entries of $\mathbf{A}$. Then, $Z_{\mathbf{A}}(\cdot)$ is computable in P if every $Z_{\mathbf{A}_i}(\cdot)$ is and is #P-hard otherwise. For each nonbipartite graph H_i, the corresponding $Z_{\mathbf{A}_i}(\cdot)$

is computable in P if $\mathbf{A}_i$ has rank at most one and is #P-hard otherwise. For each bipartite H_i, the corresponding $Z_{\mathbf{A}_i}(\cdot)$ is in P if $\mathbf{A}_i$ has the form

$$\mathbf{A}_i = \begin{pmatrix} \mathbf{0} & \mathbf{B}_i \\ \mathbf{B}_i^{\mathrm{T}} & \mathbf{0} \end{pmatrix},$$

where $\mathbf{B}_i$ has rank one, and is #P-hard otherwise.

The result of Bulatov and Grohe is both sweeping and enormously applicable. It completely solves the problem for all nonnegative symmetric matrices. However, when we are dealing with nonnegative matrices, there are no cancellations in the exponential sum $Z_{\mathbf{A}}(\cdot)$. These potential cancellations, when $\mathbf{A}$ is either a real or a complex matrix, may in fact be the source of surprisingly efficient algorithms for computing $Z_{\mathbf{A}}(\cdot)$. The occurrence of these cancellations, or the mere possibility of such occurrence, makes proving any complexity dichotomies more difficult. Such a proof must identify all polynomial-time decidable problems utilizing the potential cancellations, such as those found in holographic algorithms [36, 37, 8], and at the same time carve out exactly what is left. This situation is similar to *monotone* versus *nonmonotone* circuit complexity. It turns out that indeed there are more interesting tractable cases over the reals, and in particular, the 2×2 Hadamard matrix $\mathbf{H}$ in (1.2) turns out to be one such case. This is the starting point for the next great chapter on the complexity of $Z_{\mathbf{A}}(\cdot)$.

In a paper [21] comprising 67 pages of beautiful proofs of both exceptional depth and conceptual vision, Goldberg et al. proved a complexity dichotomy theorem for algebraic real-valued symmetric matrices $\mathbf{A}$. Their result is too intricate to give a short and accurate summary here. It states that the problem of computing $Z_{\mathbf{A}}(G)$ for any algebraic real $\mathbf{A}$ is either in P or #P-hard. Which case it is depends on the connected components of $\mathbf{A}$. The overall statement remains that $Z_{\mathbf{A}}(G)$ is tractable if every connected component of $\mathbf{A}$ is and is #P-hard otherwise. However, the exact description of tractability for connected $\mathbf{A}$ is much more technical and involved. The Hadamard matrix $\mathbf{H}$ and its tensor products $\mathbf{H} \otimes \cdots \otimes \mathbf{H}$ play a major role in the tractable case. If we index rows and columns of $\mathbf{H}$ by the finite field $\mathbb{Z}_2$, then its (x, y) entry is $(-1)^{xy}$. For the nonbipartite case, there is another 4×4 symmetric matrix $\mathbf{H}_4$, different from $\mathbf{H} \otimes \mathbf{H}$, where the rows and columns are indexed by $(\mathbb{Z}_2)^2$ and the entry at $((x_1, x_2), (y_1, y_2))$ is $(-1)^{x_1 y_2 + x_2 y_1}$. These two matrices, and their arbitrary tensor products, all correspond to new tractable $Z_{\mathbf{A}}(\cdot)$. In fact, there are some more tractable cases, starting with what can be roughly described as certain rank one modifications on these tensor products.

The proof of [21] proceeds by establishing a long sequence of successively more stringent properties that a tractable $\mathbf{A}$ must satisfy. Ultimately, it arrives at a point where satisfaction of these properties implies that $Z_{\mathbf{A}}(G)$ can be computed as

$$\sum_{x_1, x_2, \ldots, x_n \in \mathbb{Z}_2} (-1)^{f_G(x_1, x_2, \ldots, x_n)},$$

where f_G is a quadratic polynomial over $\mathbb{Z}_2$. This sum is known to be computable in polynomial time in n [10] [27, Theorem 6.30], the number of variables. In hindsight, the case with the simplest Hadamard matrix $\mathbf{H}$ which was an obstacle to the Bulatov–Grohe dichotomy theorem and was left open for some time could have been directly solved if one had adopted the polynomial viewpoint of [21].

While positive and negative real numbers provide the possibility of cancellations, there is a significantly richer variety of possible cancellations over the complex domain. We independently came to the tractability of $Z_{\mathbf{H}}(\cdot)$, with $\mathbf{H}$ being the 2×2

Hadamard matrix, from a slightly different angle. In [9], the authors studied a certain type of constraint satisfaction problem. This is motivated by investigations of a class of counting problems called Holant problems, and it is connected with the technique called holographic reductions introduced by Valiant [35, 36]. Let us briefly describe this framework. A *signature grid* $\Omega = (G, \mathcal{F})$ is a tuple in which $G = (V, E)$ is a graph and each $v \in V$ is attached a function $F_v \in \mathcal{F}$. An edge assignment σ for every $e \in E$ gives an evaluation $\prod_{v \in V} F_v(\sigma|_{E(v)})$, where $E(v)$ denotes the set of incident edges of v. The counting problem on an input instance Ω is to compute

$$\text{Holant}(\Omega) = \sum_{\sigma} \prod_{v \in V} F_v\big(\sigma|_{E(v)}\big).$$

For example, if we take $\sigma\colon E \to \{0, 1\}$ and attach the exact-one function at every vertex $v \in V$, then Holant(Ω) is the number of perfect matchings of G. Incidentally, Freedman, Lovász, and Schrijver showed [18] that counting perfect matchings *cannot* be expressed as $Z_{\mathbf{A}}(\cdot)$ for any matrix $\mathbf{A}$ over $\mathbb{R}$. However, every function $Z_{\mathbf{A}}(\cdot)$ (vertex assignment) *can* be simulated by Holant$(\cdot)$ (edge assignment) as follows: $\mathbf{A}$ defines a function of arity 2 for every edge of G. Consider the bipartite vertex-edge incidence graph $G' = (V(G), E(G), E')$ of G, where $(v, e) \in E'$ if e is incident to v in G. Then attach the equality function at every $v \in V(G)$ and the function defined by $\mathbf{A}$ at every $e \in E(G)$. This defines a signature grid Ω with the underlying graph G'. Then $Z_{\mathbf{A}}(G) = \text{Holant}(\Omega)$.

Denote a symmetric function on n boolean variables by $[f_0, f_1, \ldots, f_n]$, where f_j is the value on inputs of Hamming weight j. For example, the exact-one function is $[0, 1, 0, \ldots, 0]$ and $\mathbf{H}$ is just $[1, 1, -1]$. The authors of [9] discovered that the three families of functions (listing the values of a function lexicographically as in a truth table on k boolean variables)

$$\mathcal{F}_1 = \big\{\, \lambda([1,0]^{\otimes k} + i^r[0,\ \ 1]^{\otimes k}) \;\big|\; \lambda \in \mathbb{C},\ k = 1, 2, \ldots,\ \text{and } r = 0, 1, 2, 3 \,\big\},$$
$$\mathcal{F}_2 = \big\{\, \lambda([1,1]^{\otimes k} + i^r[1,-1]^{\otimes k}) \;\big|\; \lambda \in \mathbb{C},\ k = 1, 2, \ldots,\ \text{and } r = 0, 1, 2, 3 \,\big\},$$
$$\mathcal{F}_3 = \big\{\, \lambda([1,\ i]^{\otimes k} + i^r[1,-i]^{\otimes k}) \;\big|\; \lambda \in \mathbb{C},\ k = 1, 2, \ldots,\ \text{and } r = 0, 1, 2, 3 \,\big\}$$

all give rise to tractable problems: Holant(Ω) for any $\Omega = (G, \mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3)$ can be solved in P. In particular, by taking $r = 1$, $k = 2$, and $\lambda = (1+i)^{-1}$ in $\mathcal{F}_3$, we recover the binary function $[1, 1, -1]$ that corresponds to the Hadamard matrix $\mathbf{H}$ in (1.2). If we take $r = 0$, $\lambda = 1$ in $\mathcal{F}_1$, we get the equality function $[1, 0, \ldots, 0, 1]$ on k bits. This shows that $Z_{\mathbf{H}}(\cdot)$, as a special case, can be computed in P.

However, more instructive for us is the natural way in which complex numbers appear in such counting problems, especially when applying holographic reductions. One can say that the presence of powers of $i = \sqrt{-1}$ in these three families "reveals" the true nature of $\mathbf{H}$ as belonging to a family of tractable counting problems, where complex numbers are the correct language. In fact, the tractability of Holant(Ω) for $\Omega = (G, \mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3)$ all boils down to an exponential sum of the form

$$\sum_{x_1, x_2, \ldots, x_n \in \{0,1\}} i^{L_1 + L_2 + \cdots + L_s}, \tag{1.3}$$

where each L_j is an indicator function of an affine form of $x_1, x_2, \ldots, x_n$ over $\mathbb{Z}_2$ (and thus, the exponent of i in the equation above is a mod 4 sum of mod 2 sums). From here it is only natural to investigate the complexity of $Z_{\mathbf{A}}(\cdot)$ for symmetric complex

matrices, since it not only is a natural generalization but also can reveal the inner unity and some deeper structural properties. Interested readers can find more details in [9]. Also see Remark 12.10 at the end of section 12.

Our investigation of complex-valued graph homomorphisms is also motivated by the partition function in quantum physics. In classical statistical physics, the partition function is always real-valued. But in a generic quantum system, for which complex numbers are the right language, the partition function is in general complex-valued [17]. In particular, if the physics model is over a discrete graph and is nonorientable, then the edge weights are given by a symmetric complex matrix.

Our main result is the following complexity dichotomy theorem, though its criterion is too complicated to explain here.

THEOREM 1.1. *Let* $\mathbf{A}$ *be a symmetric and algebraic complex matrix. Then* $Z_{\mathbf{A}}(\cdot)$ *either can be computed in polynomial time or is #P-hard.*

Furthermore, under the model of computation described in section 2.2, we show that the following decision problem is solvable in polynomial time.

THEOREM 1.2 (polynomial-time decidability). *Given a symmetric and algebraic complex matrix* $\mathbf{A}$*, there is a polynomial-time algorithm that decides whether* $Z_{\mathbf{A}}(\cdot)$ *is in polynomial time or is #P-hard.*

Recent developments. In [34], Thurley announced a dichotomy theorem[1] for $Z_{\mathbf{A}}(\cdot)$, where $\mathbf{A}$ is a complex Hermitian matrix. The tractability result of the present paper (in section 12) was used in [34]. Cai and Chen proved a dichotomy theorem for $Z_{\mathbf{A}}(\cdot)$ for directed graph homomorphisms, where $\mathbf{A}$ is a nonnegative but not necessarily symmetric matrix [5]. A dichotomy theorem is also proved for the more general counting constraint satisfaction problem when the constraint functions take values in $\{0, 1\}$ [1, 2] (with an alternative proof given in [14] that also shows the decidability of the dichotomy criterion), when the functions take nonnegative and rational values [3], and when they are nonnegative and algebraic [7]. Finally, built on the methods and results of [1, 14, 21] and the present paper, Cai and Chen proved a dichotomy theorem for all algebraic complex-valued counting constraint satisfaction problems [6].

Organization. Due to the complexity of the proof of Theorem 1.1, both in terms of its overall structure and in terms of technical difficulty, we first give a high-level description of the proof for the bipartite case in section 3. We prove the first and second pinning lemmas in section 4. A more detailed outline of the proof for the two cases, bipartite and nonbipartite, is presented in sections 5 and 6, respectively, with formal definitions and theorems. We then prove all the lemmas and theorems used in sections 5 and 6, as well as Theorem 1.2, in the rest of the paper. An index of conditions and problem definitions is given in Figure 1.1.

2. Preliminaries. In the paper, we let $\mathbb{Q}$ denote the set of rational numbers and let $\mathbb{R}$ and $\mathbb{C}$ denote the set of algebraic real and algebraic complex numbers, respectively, for convenience (even though many of the supporting lemmas and theorems actually hold for general real or complex numbers, especially when computation or polynomial-time reduction is not concerned in the statement).

2.1. Notation. For a positive integer n, we use $[n]$ to denote the set $\{1, \ldots, n\}$ (when $n = 0$, $[0] = \emptyset$). We use $[m : n]$, where $m \leq n$, to denote $\{m, m+1, \ldots, n\}$. We

[1] However, the following is a counter example to Claim 3 on p. 50 of [34]: $D_{11}^{[c];1} = D_{22}^{[c];1} = 1$, $D_{11}^{[c];2} = i$ (the imaginary unit), and $D_{22}^{[c];2} = -i$. We believe that this minor deficiency in the proof probably can be overcome using the techniques in this paper, in particular those from section 8.4.

FIG. 1.1. *Index of conditions and problem definitions.*

use $\mathbf{1}_n$ to denote the all-one vector of dimension n. Sometimes we omit n when the dimension is clear from the context. For a positive integer N, we let $\omega_N = e^{2\pi i/N}$, a primitive Nth root of unity.

Let $\mathbf{x}, \mathbf{y}$ be two vectors in $\mathbb{C}^n$. Then we use $\langle \mathbf{x}, \mathbf{y} \rangle$ to denote their inner product,

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^{n} x_i \cdot \overline{y_i},$$

and $\mathbf{x} \circ \mathbf{y} \in \mathbb{C}^n$ to denote their Hadamard product, $(\mathbf{x} \circ \mathbf{y})_i = x_i \cdot y_i$ for all $i \in [n]$.

Let $\mathbf{A} = (A_{i,j})$ be a $k \times \ell$ matrix and $\mathbf{B} = (B_{i,j})$ be an $m \times n$ matrix. We use $\mathbf{A}_{i,*}$, $i \in [k]$, to denote the ith row vector and $\mathbf{A}_{*,j}$, $j \in [\ell]$, to denote the jth column vector of $\mathbf{A}$. We let $\mathbf{C} = \mathbf{A} \otimes \mathbf{B}$ denote their tensor product: $\mathbf{C}$ is a $km \times \ell n$ matrix whose rows and columns are indexed by $[k] \times [m]$ and $[\ell] \times [n]$, respectively, such that

$$C_{(i_1,i_2),(j_1,j_2)} = A_{i_1,j_1} \cdot B_{i_2,j_2} \quad \text{for all } i_1 \in [k],\ i_2 \in [m],\ j_1 \in [\ell], \text{ and } j_2 \in [n].$$

Given an $n \times n$ symmetric complex matrix $\mathbf{A}$, we use $G = (V, E)$ to denote the following undirected graph: $V = [n]$ and $ij \in E$ iff $A_{i,j} \neq 0$. We say $\mathbf{A}$ is *connected* if G is connected, and we say $\mathbf{A}$ has connected components $\mathbf{A}_1, \ldots, \mathbf{A}_s$ if the connected components of G are $V_1, \ldots, V_s$ and $\mathbf{A}_i$ is the $|V_i| \times |V_i|$ submatrix of $\mathbf{A}$ restricted by $V_i \subseteq [n]$ for all $i \in [s]$. Moreover, we say $\mathbf{A}$ is *bipartite* if G is bipartite; otherwise, $\mathbf{A}$ is *nonbipartite*. Let Σ and Π be two permutations of $[n]$. Then we use $\mathbf{A}_{\Sigma,\Pi}$ to denote the $n \times n$ matrix whose (i,j)th entry is $A_{\Sigma(i),\Pi(j)}$, $i, j \in [n]$.

We say $\mathbf{C}$ is the *bipartization* of a matrix $\mathbf{F}$ if

$$\mathbf{C} = \begin{pmatrix} \mathbf{0} & \mathbf{F} \\ \mathbf{F}^{\mathrm{T}} & \mathbf{0} \end{pmatrix}.$$

We usually use D_i to denote the (i,i)th entry of a diagonal matrix $\mathbf{D}$.

We say a problem is tractable if it can be solved in polynomial time. Given two problems $\mathcal{P}$ and $\mathcal{Q}$, we say $\mathcal{P}$ is polynomial-time reducible to $\mathcal{Q}$, or $\mathcal{P} \leq \mathcal{Q}$, if there is a polynomial-time algorithm that solves $\mathcal{P}$ using an oracle for $\mathcal{Q}$. These reductions are known as Cook reductions. We also say $\mathcal{P}$ is polynomial-time equivalent to $\mathcal{Q}$, or $\mathcal{P} \equiv \mathcal{Q}$, if $\mathcal{P} \leq \mathcal{Q}$ and $\mathcal{Q} \leq \mathcal{P}$.

2.2. Model of computation.[2] One technical issue is the *model of computation* with algebraic numbers. We adopt a standard model from [26] for computation in an algebraic number field. We start with some notation.

Let $\mathbf{A}$ be a fixed symmetric matrix where every entry $A_{i,j}$ is an algebraic number. We let $\mathscr{A}$ denote the finite set of algebraic numbers consisting of entries $A_{i,j}$ of $\mathbf{A}$. Then it is easy to see that $Z_{\mathbf{A}}(G)$, for any undirected graph G, is a number in $\mathbb{Q}(\mathscr{A})$, the algebraic extension of $\mathbb{Q}$ by $\mathscr{A}$. By the primitive element theorem [30], there exists an algebraic number $\alpha \in \mathbb{Q}(\mathscr{A})$ such that $\mathbb{Q}(\mathscr{A}) = \mathbb{Q}(\alpha)$. (Essentially, $\mathbb{Q}$ has characteristic 0, and therefore the field extension $\mathbb{Q}(\mathscr{A})$ is separable. We can take the normal closure of $\mathbb{Q}(\mathscr{A})$, which is a finite-dimensional separable and normal extension of $\mathbb{Q}$, and thus Galois [24]. By Galois correspondence, there are only a finite number of intermediate fields between $\mathbb{Q}$ and this Galois extension field and thus a fortiori only a finite number of intermediate fields between $\mathbb{Q}$ and $\mathbb{Q}(\mathscr{A})$. Then Artin's theorem on primitive elements implies that $\mathbb{Q}(\mathscr{A})$ is a simple extension $\mathbb{Q}(\alpha)$.) In the proof of Theorem 1.1 when the complexity of a partition function $Z_{\mathbf{A}}(\cdot)$ is concerned, the matrix $\mathbf{A}$ is considered fixed. Thus, we may assume we are given, as part of the problem description, such a number α, encoded by a minimal polynomial $F(x) \in \mathbb{Q}[x]$ of α. In addition to F, we are given a sufficiently good rational approximation $\widehat{\alpha}$ of α which uniquely determines α as a root of $F(x)$.[3]

Let $d = \deg(F)$. Then every number c in $\mathbb{Q}(\mathscr{A})$, including the $A_{i,j}$'s and $Z_{\mathbf{A}}(G)$ for any G, has a unique representation as a polynomial of α:

$$c_0 + c_1 \cdot \alpha + \cdots + c_{d-1} \cdot \alpha^{d-1}, \qquad \text{where every } c_i \text{ is a rational number.}$$

We will refer to this polynomial as the *standard representation* of c. Given a number $c \in \mathbb{Q}(\mathscr{A})$ in the standard representation, its input size is the sum of the binary lengths of all the rational coefficients. It is easy to see that all the field operations over $\mathbb{Q}(\mathscr{A})$ in this representation can be computed in polynomial time in the input size.

We emphasize that when the complexity of $Z_{\mathbf{A}}(\cdot)$ is concerned in the proof of Theorem 1.1, all the following are considered as constants since they are part of the problem description and not part of the input: the size of $\mathbf{A}$, the minimal polynomial $F(x)$ of α, the approximation $\widehat{\alpha}$ of α, as well as the entries $A_{i,j}$ of $\mathbf{A}$ encoded in the standard representation. Given an undirected graph G, the problem is then to output $Z_{\mathbf{A}}(G) \in \mathbb{Q}(\mathscr{A})$ encoded in the standard representation. We remark that the same model applies to the problem of computing $Z_{\mathbf{C},\mathfrak{D}}(\cdot)$, to be defined in section 2.3.

However, for most of the proof of Theorem 1.1 this issue of computation model seems not to be central, because our proof starts with a preprocessing step using the purification lemma (see section 3 for a high-level description of the proof, and see section 7 for the purification lemma), after which the matrix concerned becomes a *pure* one, meaning that every entry is the product of a nonnegative integer and a root of unity. So throughout the proof, we let $\mathbb{C}$ denote the set of algebraic numbers and refer to them simply as complex numbers, except in the proof of the purification lemma in section 7, where we will be more careful about the model of computation.

[2] For readers who are not particularly concerned with details of the model of computation with complex numbers, this section can be skipped initially.

[3] This is a slight modification to the model of [26] and of [34, 33]. It will come in handy later in one step of the proof in section 7, in which it allows us to avoid certain technical subtleties.

After the proof of Theorem 1.1, we consider the decidability of the dichotomy theorem and prove Theorem 1.2. The input of the problem is the full description of $\mathbf{A}$, including the minimal polynomial $F(x)$ of α, the approximation $\widehat{\alpha}$ of α, as well as the standard representation of the entries $A_{i,j}$ of $\mathbf{A}$. We refer to the binary length of all the components above as the input size of $\mathbf{A}$. To prove Theorem 1.2, we give an algorithm that runs in polynomial time in the binary length of $\mathbf{A}$ and decides whether the problem of computing $Z_{\mathbf{A}}(\cdot)$ is in polynomial time or #P-hard.

2.3. Definitions of $\mathsf{EVAL}(\mathbf{A})$ and $\mathsf{EVAL}(\mathbf{C}, \mathfrak{D})$. Let $\mathbf{A} \in \mathbb{C}^{m\times m}$ be a symmetric matrix with entries $(A_{i,j})$. It defines a graph homomorphism problem $\mathsf{EVAL}(\mathbf{A})$ as follows: Given an undirected graph $G = (V, E)$, compute

$$Z_{\mathbf{A}}(G) = \sum_{\xi: V\to[m]} \mathrm{wt}_{\mathbf{A}}(\xi), \quad \text{where} \quad \mathrm{wt}_{\mathbf{A}}(\xi) = \prod_{(u,v)\in E} A_{\xi(u),\xi(v)}.$$

We call ξ an *assignment* to the vertices of G and $\mathrm{wt}_{\mathbf{A}}(\xi)$ the *weight* of ξ.

To study the complexity of $\mathsf{EVAL}(\mathbf{A})$, we introduce a much larger class of EVAL problems with not only edge weights but also vertex weights. Moreover, the vertex weights depend on the degrees of vertices of G, modulo some integer modulus. It is a generalization of the edge-vertex weight problems introduced in [21]. See also [29].

Definition 2.1. *Let $\mathbf{C} \in \mathbb{C}^{m\times m}$ be a symmetric matrix and*

$$\mathfrak{D} = \left(\mathbf{D}^{[0]}, \mathbf{D}^{[1]}, \ldots, \mathbf{D}^{[N-1]}\right)$$

be a sequence of diagonal matrices in $\mathbb{C}^{m\times m}$ for some $N \geq 1$. We define the following problem $\mathsf{EVAL}(\mathbf{C}, \mathfrak{D})$: Given an undirected graph $G = (V, E)$, compute

$$Z_{\mathbf{C},\mathfrak{D}}(G) = \sum_{\xi: V\to[m]} \mathrm{wt}_{\mathbf{C},\mathfrak{D}}(\xi), \tag{2.1}$$

where

$$\mathrm{wt}_{\mathbf{C},\mathfrak{D}}(\xi) = \left(\prod_{(u,v)\in E} C_{\xi(u),\xi(v)}\right)\left(\prod_{v\in V} D^{[\deg(v) \bmod N]}_{\xi(v)}\right)$$

and $\deg(v)$ *denotes the degree of v in G.*

Let G be an undirected graph with connected components $G_1, \ldots, G_s$.

Property 2.2. $Z_{\mathbf{C},\mathfrak{D}}(G) = Z_{\mathbf{C},\mathfrak{D}}(G_1) \times \cdots \times Z_{\mathbf{C},\mathfrak{D}}(G_s)$.

Property 2.2 implies that whether we need to design an algorithm for $\mathsf{EVAL}(\mathbf{C}, \mathfrak{D})$ or reduce $\mathsf{EVAL}(\mathbf{C}, \mathfrak{D})$ to another problem $\mathsf{EVAL}(\mathbf{C}', \mathfrak{D}')$, it suffices to consider connected input graphs. Also note that since $\mathsf{EVAL}(\mathbf{A})$ is a special case of $\mathsf{EVAL}(\mathbf{C}, \mathfrak{D})$ in which every $\mathbf{D}^{[i]}$ is an identity matrix, Property 2.2 and the remarks above apply to $\mathsf{EVAL}(\mathbf{A})$ as well.

Next, suppose $\mathbf{C}$ is the bipartization of an $m \times n$ $\mathbf{F}$, so $\mathbf{C}$ is $(m + n) \times (m + n)$. Given a graph G and a vertex u in G, we use Ξ_1 to denote the set of $\xi\colon V \to [m + n]$ with $\xi(u) \in [m]$ and Ξ_2 to denote the set of ξ with $\xi(u) \in [m + 1 : m + n]$. Then let

$$Z^{\rightarrow}_{\mathbf{C},\mathfrak{D}}(G, u) = \sum_{\xi\in\Xi_1} \mathrm{wt}_{\mathbf{C},\mathfrak{D}}(\xi) \quad \text{and} \quad Z^{\leftarrow}_{\mathbf{C},\mathfrak{D}}(G, u) = \sum_{\xi\in\Xi_2} \mathrm{wt}_{\mathbf{C},\mathfrak{D}}(\xi).$$

The next property follows from the definitions.

Property 2.3. $Z_{\mathbf{C},\mathfrak{D}}(G) = Z^{\rightarrow}_{\mathbf{C},\mathfrak{D}}(G, u) + Z^{\leftarrow}_{\mathbf{C},\mathfrak{D}}(G, u)$.

We introduce these two new functions because of the following lemma.

LEMMA 2.4. *For each $i \in \{0,1,2\}$, let $\mathbf{F}^{[i]}$ be an $m_i \times n_i$ complex matrix, where $m_0 = m_1 m_2$ and $n_0 = n_1 n_2$; let $\mathbf{C}^{[i]}$ be the bipartization of $\mathbf{F}^{[i]}$; and let*

$$\mathfrak{D}^{[i]} = \left(\mathbf{D}^{[i,0]}, \ldots, \mathbf{D}^{[i,N-1]}\right)$$

be a sequence of $(m_i + n_i) \times (m_i + n_i)$ diagonal matrices for some $N \geq 1$, where

$$\mathbf{D}^{[i,r]} = \begin{pmatrix} \mathbf{P}^{[i,r]} & \\ & \mathbf{Q}^{[i,r]} \end{pmatrix}$$

and $\mathbf{P}^{[i,r]}$, $\mathbf{Q}^{[i,r]}$ are $m_i \times m_i$, $n_i \times n_i$ diagonal matrices, respectively. Assume

$$\mathbf{F}^{[0]} = \mathbf{F}^{[1]} \otimes \mathbf{F}^{[2]}, \quad \mathbf{P}^{[0,r]} = \mathbf{P}^{[1,r]} \otimes \mathbf{P}^{[2,r]}, \quad \text{and} \quad \mathbf{Q}^{[0,r]} = \mathbf{Q}^{[1,r]} \otimes \mathbf{Q}^{[2,r]}$$

for all $r \in [0 : N-1]$. Then for any connected *graph G and any vertex u^* in G,*

$$Z^{\rightarrow}_{\mathbf{C}^{[0]},\mathfrak{D}^{[0]}}(G,u^*) = Z^{\rightarrow}_{\mathbf{C}^{[1]},\mathfrak{D}^{[1]}}(G,u^*) \cdot Z^{\rightarrow}_{\mathbf{C}^{[2]},\mathfrak{D}^{[2]}}(G,u^*) \quad \text{and} \tag{2.2}$$

$$Z^{\leftarrow}_{\mathbf{C}^{[0]},\mathfrak{D}^{[0]}}(G,u^*) = Z^{\leftarrow}_{\mathbf{C}^{[1]},\mathfrak{D}^{[1]}}(G,u^*) \cdot Z^{\leftarrow}_{\mathbf{C}^{[2]},\mathfrak{D}^{[2]}}(G,u^*). \tag{2.3}$$

Proof. We only prove (2.2) about $Z^{\rightarrow}$. The proof of (2.3) is similar. First, if G is not bipartite, then $Z^{\rightarrow}_{\mathbf{C}^{[i]},\mathfrak{D}^{[i]}}(G,u^*) = 0$ for all $i \in \{0,1,2\}$, and (2.2) holds trivially.

Now assume $G = (U \cup V, E)$ is a bipartite graph, $u^* \in U$, and every edge $uv \in E$ has one vertex u from U and one vertex v from V. We let Ξ_i, $i \in \{0,1,2\}$, denote the set of assignments ξ_i from $U \cup V$ to $[m_i + n_i]$ such that $\xi_i(u) \in [m_i]$ for all $u \in U$ and $\xi_i(v) \in [m_i + 1 : m_i + n_i]$ for all $v \in V$. Since G is connected, we have

$$Z^{\rightarrow}_{\mathbf{C}^{[i]},\mathfrak{D}^{[i]}}(G,u^*) = \sum_{\xi_i \in \Xi_i} \mathrm{wt}_{\mathbf{C}^{[i]},\mathfrak{D}^{[i]}}(\xi_i) \quad \text{for } i \in \{0,1,2\}.$$

We define a map ρ from $\Xi_1 \times \Xi_2$ to Ξ_0 as follows: $\rho(\xi_1,\xi_2) = \xi_0$, where for every $u \in U$, $\xi_0(u)$ is the row index of $\mathbf{F}^{[0]}$ that corresponds to row $\xi_1(u)$ of $\mathbf{F}^{[1]}$ and row $\xi_2(u)$ of $\mathbf{F}^{[2]}$ in the tensor product $\mathbf{F}^{[0]} = \mathbf{F}^{[1]} \otimes \mathbf{F}^{[2]}$; and for every $v \in V$, $\xi_0(v) - m_0$ is the column index of $\mathbf{F}^{[0]}$ that corresponds to column $\xi_1(v) - m_1$ of $\mathbf{F}^{[1]}$ and column $\xi_2(v) - m_2$ of $\mathbf{F}^{[2]}$ in the tensor product. It is clear that ρ is a bijection, and

$$\mathrm{wt}_{\mathbf{C}^{[0]},\mathfrak{D}^{[0]}}(\xi_0) = \mathrm{wt}_{\mathbf{C}^{[1]},\mathfrak{D}^{[1]}}(\xi_1) \cdot \mathrm{wt}_{\mathbf{C}^{[2]},\mathfrak{D}^{[2]}}(\xi_2),$$

if $\rho(\xi_1,\xi_2) = \xi_0$. Equation (2.2) then follows, and the lemma is proved. ∎

2.4. Basic #P-hardness. We state the dichotomy of Bulatov and Grohe.

THEOREM 2.5 (Bulatov and Grohe [4]). *Let $\mathbf{A}$ be a symmetric and connected matrix with nonnegative algebraic entries. Then $\mathsf{EVAL}(\mathbf{A})$ is either in polynomial time or #P-hard. Moreover, we have the following two cases:*

1. *If $\mathbf{A}$ is bipartite, then $\mathsf{EVAL}(\mathbf{A})$ is in polynomial time if the rank of $\mathbf{A}$ is 2; otherwise $\mathsf{EVAL}(\mathbf{A})$ is #P-hard.*
2. *If $\mathbf{A}$ is not bipartite, then $\mathsf{EVAL}(\mathbf{A})$ is in polynomial time if the rank of $\mathbf{A}$ is at most 1; otherwise $\mathsf{EVAL}(\mathbf{A})$ is #P-hard.*

Theorem 2.5 gives us the following useful corollary.

COROLLARY 2.6. *Let $\mathbf{A}$ be a symmetric and connected matrix with nonnegative algebraic entries. If $\mathbf{A}$ has a 2×2 submatrix $\mathbf{B}$ such that all four entries of $\mathbf{B}$ are nonzero and $\det(\mathbf{B}) \neq 0$, then the problem $\mathsf{EVAL}(\mathbf{A})$ is* #P-hard.

3. A high-level description of the proof. The first step in the proof of Theorem 1.1 is to reduce the problem to connected graphs and matrices.

Let $\mathbf{A}$ be an $m \times m$ symmetric complex matrix. If G has connected components $\{G_i\}$, then $Z_{\mathbf{A}}(G) = \prod_i Z_{\mathbf{A}}(G_i)$; if G is connected and $\mathbf{A}$ has connected components $\{\mathbf{A}_j\}$, then $Z_{\mathbf{A}}(G) = \sum_j Z_{\mathbf{A}_j}(G)$. Thus, if every $Z_{\mathbf{A}_j}(\cdot)$ is computable in polynomial time, then so is $Z_{\mathbf{A}}(\cdot)$. The hardness direction is less obvious. Assume that $\mathsf{EVAL}(\mathbf{A}_j)$ is #P-hard for some j; we want to show that $\mathsf{EVAL}(\mathbf{A})$ is also #P-hard by giving a polynomial-time reduction from $\mathsf{EVAL}(\mathbf{A}_j)$ to $\mathsf{EVAL}(\mathbf{A})$.

Now let G be an undirected graph. To compute $Z_{\mathbf{A}_j}(G)$, it suffices to compute $Z_{\mathbf{A}_j}(G_i)$ for all connected components G_i of G. Therefore, we may just assume that G is connected. Define a *pinning* version of $Z_{\mathbf{A}}(\cdot)$ as follows. For any chosen vertex $w \in V(G)$ and any $k \in [m]$, we let

$$Z_{\mathbf{A}}(G, w, k) = \sum_{\xi: V \to [m],\, \xi(w) = k} \prod_{(u,v) \in E} A_{\xi(u),\xi(v)}.$$

Then we can prove a *pinning lemma* (Lemma 4.1) which states that the problem of computing $Z_{\mathbf{A}}(\cdot)$ is polynomial-time equivalent to computing $Z_{\mathbf{A}}(\cdot,\cdot,\cdot)$. Note that if V_j denotes the subset of $[m]$ where $\mathbf{A}_j$ is the submatrix of $\mathbf{A}$ restricted by V_j, then for a connected graph G, we have

$$Z_{\mathbf{A}_j}(G) = \sum_{k \in V_j} Z_{\mathbf{A}}(G, w, k),$$

which gives us the desired polynomial-time reduction from $\mathsf{EVAL}(\mathbf{A}_j)$ to $\mathsf{EVAL}(\mathbf{A})$.

The proof of this pinning lemma (Lemma 4.1) is a standard adaptation to the complex numbers of the one proved in [21]. For technical reasons we indeed need a total of three pinning lemmas (Lemmas 4.1, 4.3, and 8.4), and the proofs of the other two are a bit more involved. We remark that all three pinning lemmas show only the *existence* of polynomial-time reductions between $Z_{\mathbf{A}}(\cdot)$ and $Z_{\mathbf{A}}(\cdot,\cdot,\cdot)$ but do not *constructively* produce such a reduction, given $\mathbf{A}$. The proof of the pinning lemma in [21] used a result by Lovász [29] for real matrices. It is possible to use a new result of Schrijver [32] in the complex case. However, we give direct and self-contained proofs of our three lemmas without using [29] or [32].

After this preliminary step, we restrict to *connected* and symmetric $\mathbf{A}$. As indicated, for our work the two most influential predecessor papers are those by Bulatov and Grohe [4] and Goldberg et al. [21]. In both papers, the polynomial-time algorithms for the tractable cases are relatively straightforward or are previously known. The difficult part of the proof is to show that, in all other cases, the problem is #P-hard. Our proof follows a conceptual framework similar to that of Goldberg et al. [21]. However, over the complex numbers, new difficulties arise in both the tractability and the hardness part of the proof. Therefore, both the overall organization and the substantive part of the proof have to be done separately.

First, the complex numbers afford a richer variety of cancellations, which could lead to surprisingly efficient algorithms for $\mathsf{EVAL}(\mathbf{A})$ when the complex matrix $\mathbf{A}$ satisfies certain nice conditions. This turns out to be the case, and we obtain additional nontrivial tractable cases. These boil down to the following class of problems called $\mathsf{EVAL}(q)$. Let q be a fixed prime power. The input of $\mathsf{EVAL}(q)$ is a quadratic polynomial $f(x_1, x_2, \ldots, x_n)$ with integer coefficients; the output is

$$Z_q(f) = \sum_{x_1, x_2, \ldots, x_n \in \mathbb{Z}_q} \omega_q^{f(x_1, x_2, \ldots, x_n)}.$$

We show that for any fixed prime power q, $\mathsf{EVAL}(q)$ is in polynomial time. In the algorithm (see section 12), Gauss sums play a crucial role. The tractability part of our dichotomy theorem is then done by reducing $\mathsf{EVAL}(\mathbf{A})$, assuming $\mathbf{A}$ satisfies a set of nice structural conditions (to be described in the rest of this section) imposed by the hardness part, to $\mathsf{EVAL}(q)$ for some appropriate prime power q. While the corresponding sums for finite fields (when q is a prime) are known to be in polynomial time [10, 15], [27, Theorem 6.30] and, in particular, this includes the special case of $\mathbb{Z}_2$ used in [21], our algorithm over rings $\mathbb{Z}_q$ is new and should be of independent interest.

Next we briefly describe the proof structure of the hardness part of the dichotomy theorem. Let $\mathbf{A}$ be a connected and symmetric matrix. The difficulty starts with the most basic proof technique, called gadget constructions. With a graph gadget, one can take any input undirected graph G and produce a modified graph G^* by replacing each edge of G with the gadget. Moreover, one can define a suitable modified matrix $\mathbf{A}^*$ from the fixed matrix $\mathbf{A}$ and the gadget such that $Z_{\mathbf{A}^*}(G) = Z_{\mathbf{A}}(G^*)$ for all undirected graphs G.

A simple example of this maneuver is called *thickening*, where one replaces each edge in the input G by t parallel edges to get G^*. It is easy to see that if $\mathbf{A}^*$ is obtained from $\mathbf{A}$ by replacing each entry $A_{i,j}$ by its tth power $(A_{i,j})^t$, then the equation above holds and we get a reduction from $\mathsf{EVAL}(\mathbf{A}^*)$ to $\mathsf{EVAL}(\mathbf{A})$. In particular, if $\mathbf{A}$ is real (as in the case of [21]) and t is even, this produces a nonnegative matrix $\mathbf{A}^*$, to which one may apply the Bulatov–Grohe result:

1. If $\mathbf{A}^*$, as a symmetric and nonnegative matrix, does not satisfy the tractability criteria of Bulatov and Grohe as described in Theorem 2.5, then both $\mathsf{EVAL}(\mathbf{A}^*)$ and $\mathsf{EVAL}(\mathbf{A})$ are #P-hard and we are done.
2. Otherwise, $\mathbf{A}^*$ satisfies the Bulatov–Grohe tractability criteria, from which $\mathbf{A}$ must satisfy certain necessary structural properties since $\mathbf{A}^*$ is derived from $\mathbf{A}$.

The big picture of the proof of the dichotomy theorem is then to design various graph gadgets to show that, assuming $\mathsf{EVAL}(\mathbf{A})$ is not #P-hard, the matrix $\mathbf{A}$ must satisfy a collection of strong necessary conditions over its complex entries $A_{i,j}$. (The exact proof structure, however, is different from this very-high-level description, which will become clear in the rest of this section.) To finish the proof, we show that for every $\mathbf{A}$ that satisfies all these structural conditions, one can reduce $\mathsf{EVAL}(\mathbf{A})$ to $\mathsf{EVAL}(q)$ for some appropriate prime power q (which depends only on $\mathbf{A}$), and thus $\mathsf{EVAL}(\mathbf{A})$ is tractable.

For complex matrices $\mathbf{A}$, we immediately encountered the following difficulty. Any graph gadget will only produce a matrix $\mathbf{A}^*$ whose entries are obtained from entries of $\mathbf{A}$ by arithmetic operations $+$ and $\times$. While for real numbers any even power guarantees a nonnegative quantity, as was done in [21], no obvious arithmetic operations on the complex numbers have this property. Pointedly, *conjugation* is not an arithmetic operation. However, it is clear that for roots of unity, one *can* produce conjugation by multiplication.

Thus, our proof starts with a process of replacing an arbitrary complex matrix by a *purified* complex matrix with a special form. It turns out that we must separate out the cases where $\mathbf{A}$ is bipartite or nonbipartite. A purified bipartite (and symmetric, connected) matrix is the bipartization of a matrix $\mathbf{B}$, where

$$\mathbf{B} = \begin{pmatrix} \mu_1 & & & \\ & \mu_2 & & \\ & & \ddots & \\ & & & \mu_k \end{pmatrix} \begin{pmatrix} \zeta_{1,1} & \zeta_{1,2} & \cdots & \zeta_{1,m-k} \\ \zeta_{2,1} & \zeta_{2,2} & \cdots & \zeta_{2,m-k} \\ \vdots & \vdots & \ddots & \vdots \\ \zeta_{k,1} & \zeta_{k,2} & \cdots & \zeta_{k,m-k} \end{pmatrix} \begin{pmatrix} \mu_{k+1} & & & \\ & \mu_{k+2} & & \\ & & \ddots & \\ & & & \mu_m \end{pmatrix}$$

for some $1 \le k < m$, in which every μ_i is a positive rational number and every $\zeta_{i,j}$ is a root of unity. The claim is that for every symmetric, connected, and bipartite matrix $\mathbf{A} \in \mathbb{C}^{m\times m}$, either we can already prove the #P-hardness of $\mathsf{EVAL}(\mathbf{A})$ or there exists a purified bipartite matrix $\mathbf{A}' \in \mathbb{C}^{m\times m}$ such that $\mathsf{EVAL}(\mathbf{A}')$ is polynomial-time equivalent to $\mathsf{EVAL}(\mathbf{A})$ (Theorem 5.2). For nonbipartite matrices $\mathbf{A}$, a corresponding statement holds (Theorem 6.2). For convenience, we only consider the bipartite case in the discussion below.

Continuing now with a purified bipartite matrix $\mathbf{A}'$, the next step is to *further regularize* its entries. In particular we need to combine those rows and columns of the matrix where they are essentially the same, apart from a multiple of a root of unity. This process is called *cyclotomic reduction*. To carry out this process, we need to use the more general problem $\mathsf{EVAL}(\mathbf{C}, \mathfrak{D})$ defined earlier in section 2.3. We also need to introduce the following type of matrices, called discrete unitary matrices.

DEFINITION 3.1 (discrete unitary matrix). *Let $\mathbf{F} \in \mathbb{C}^{m\times m}$ be a (not necessarily symmetric) matrix with entries $(F_{i,j})$. We call $\mathbf{F}$ an M-*discrete unitary matrix*, for some positive integer M, if it satisfies the following conditions:*

1. *Every entry $F_{i,j}$ of $\mathbf{F}$ is a root of unity, and $F_{1,i} = F_{i,1} = 1$ for all $i \in [m]$.*
2. *M is the least common multiple (lcm) of orders of all the entries $F_{i,j}$ of $\mathbf{F}$.*
3. *For all $i \ne j \in [m]$, we have $\langle \mathbf{F}_{i,*}, \mathbf{F}_{j,*} \rangle = 0$ and $\langle \mathbf{F}_{*,i}, \mathbf{F}_{*,j} \rangle = 0$.*

Some of the simplest examples of discrete unitary matrices are as follows:

$$\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 1 & 1 \\ 1 & \omega & \omega^2 \\ 1 & \omega^2 & \omega \end{pmatrix}, \quad \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & \zeta & \zeta^{-1} & \zeta^2 & \zeta^{-2} \\ 1 & \zeta^2 & \zeta^{-2} & \zeta^{-1} & \zeta \\ 1 & \zeta^{-1} & \zeta & \zeta^{-2} & \zeta^2 \\ 1 & \zeta^{-2} & \zeta^2 & \zeta & \zeta^{-1} \end{pmatrix},$$

where $\omega = e^{2\pi i/3}$ and $\zeta = e^{2\pi i/5}$. Tensor products of discrete unitary matrices are also discrete unitary matrices. These matrices play a major role in our proof.

Now we come back to the proof outline. We show that $\mathsf{EVAL}(\mathbf{A}')$ is either #P-hard or polynomial-time equivalent to $\mathsf{EVAL}(\mathbf{C}, \mathfrak{D})$ for some matrix $\mathbf{C} \in \mathbb{C}^{2n\times 2n}$ and some $\mathfrak{D}$ of diagonal matrices from $\mathbb{C}^{2n\times 2n}$, where $n \le m$ and $\mathbf{C}$ is the bipartization of a discrete unitary matrix, denoted by $\mathbf{F}$. In addition, there are further stringent requirements for $\mathfrak{D}$; otherwise $\mathsf{EVAL}(\mathbf{A}')$ is #P-hard. The detailed statements can be found in Theorems 5.3 and 5.4, summarized in properties $(\mathcal{U}_1)$ to $(\mathcal{U}_5)$. Roughly speaking, the first matrix $\mathbf{D}^{[0]}$ in $\mathfrak{D}$ must be the identity matrix, and for any matrix $\mathbf{D}^{[r]}$ in $\mathfrak{D}$, each entry of $\mathbf{D}^{[r]}$ is either zero or a root of unity. We call these conditions, with some abuse of terminology, the discrete unitary requirements. The proof that these requirements are necessary is demanding and among the most difficult in the paper.

Next, assume that we have a problem $\mathsf{EVAL}(\mathbf{C}, \mathfrak{D})$ satisfying the discrete unitary requirements with $\mathbf{C}$ being the bipartization of $\mathbf{F}$. Recall that $\omega_q = e^{2\pi i/q}$.

DEFINITION 3.2. *Let $q > 1$ be a prime power. The following $q \times q$ matrix $\boldsymbol{\mathcal{F}}_q$ is called the q-Fourier matrix: The (x, y)th entry of $\boldsymbol{\mathcal{F}}_q$ is ω_q^{xy}, $x, y \in [0 : q-1]$.*

We show that either $\mathsf{EVAL}(\mathbf{C}, \mathfrak{D})$ is #P-hard or, after a permutation of rows and columns, $\mathbf{F}$ becomes the *tensor product* of a collection of suitable Fourier matrices:

$$\boldsymbol{\mathcal{F}}_{q_1} \otimes \boldsymbol{\mathcal{F}}_{q_2} \otimes \cdots \otimes \boldsymbol{\mathcal{F}}_{q_d}, \qquad \text{where } d \ge 1 \text{ and every } q_i \text{ is a prime power.}$$

Basically, we show that even with the stringent conditions imposed on the pair $(\mathbf{C}, \mathfrak{D})$ by the discrete unitary requirements, most of $\mathsf{EVAL}(\mathbf{C}, \mathfrak{D})$ are still #P-hard, unless

$\mathbf{F}$ is the tensor product of Fourier matrices. On the other hand, the tensor product decomposition into Fourier matrices finally brings in group theory and Gauss sums. It gives us a canonical way of writing the entries of $\mathbf{F}$ in a closed form. More exactly, we index the rows and columns of $\mathbf{F}$ using $\mathbf{x} = (x_1, \ldots, x_d) \in \mathbb{Z}_{q_1} \times \cdots \times \mathbb{Z}_{q_d}$ so that

$$F_{\mathbf{x},\mathbf{y}} = \prod_{i \in [d]} \omega_{q_i}^{x_i y_i} \quad \text{for any } \mathbf{x}, \mathbf{y} \in \mathbb{Z}_{q_1} \times \cdots \times \mathbb{Z}_{q_d}.$$

Assume $q_1, \ldots, q_d$ are powers of $s \leq d$ distinct primes $p_1, \ldots, p_s$. We can also view the set of indices as $\mathbb{Z}_{q_1} \times \cdots \times \mathbb{Z}_{q_d} = G_1 \times \cdots \times G_s$, where G_i is the finite Abelian group which is the product of all the groups $\mathbb{Z}_{q_j}$ with q_j being a power of p_i.

This canonical tensor product decomposition of $\mathbf{F}$ gives us a natural way to index the rows and columns of $\mathbf{C}$ and the diagonal matrices in $\mathfrak{D}$ using $\mathbf{x}$. More exactly, we index the first half of the rows and columns of $\mathbf{C}$ and every $\mathbf{D}^{[r]}$ in $\mathfrak{D}$ using $(0, \mathbf{x})$ and index the second half of the rows and columns using $(1, \mathbf{x})$, $\mathbf{x} \in \mathbb{Z}_{q_1} \times \cdots \times \mathbb{Z}_{q_d}$.

With this canonical expression of $\mathbf{F}$ and $\mathbf{C}$, we further inquire into the structure of $\mathfrak{D}$. Here one more substantial difficulty awaits us. There are two more properties that we must demand of those diagonal matrices in $\mathfrak{D}$. If $\mathfrak{D}$ does not satisfy these additional properties, then $\mathsf{EVAL}(\mathbf{C}, \mathfrak{D})$ is #P-hard.

First, for each r, we define Λ_r and Δ_r to be the support of $\mathbf{D}^{[r]}$, where Λ_r refers to the first half of the entries and Δ_r refers to the second half of the entries (here we follow the convention of using D_i to denote the (i,i)th entry of a diagonal matrix $\mathbf{D}$):

$$\Lambda_r = \left\{\mathbf{x} : D^{[r]}_{(0,\mathbf{x})} \neq 0\right\} \quad \text{and} \quad \Delta_r = \left\{\mathbf{x} : D^{[r]}_{(1,\mathbf{x})} \neq 0\right\}.$$

We let $\mathcal{S}$ denote the set of subscripts r such that $\Lambda_r \neq \emptyset$ and let $\mathcal{T}$ denote the set of r such that $\Delta_r \neq \emptyset$. We can prove that for each $r \in \mathcal{S}$, $\Lambda_r = \prod_{i=1}^{s} \Lambda_{r,i}$ must be a direct product of cosets $\Lambda_{r,i}$ in the Abelian groups G_i, where $i = 1, \ldots, s$ correspond to the constituent prime powers of the group, and for each $r \in \mathcal{T}$, $\Delta_r = \prod_{i=1}^{s} \Delta_{r,i}$ is a direct product of cosets in the same Abelian groups. Otherwise, $\mathsf{EVAL}(\mathbf{C}, \mathfrak{D})$ is #P-hard.

Second, we show that for each $r \in \mathcal{S}$ and $r \in \mathcal{T}$, respectively, $\mathbf{D}^{[r]}$ on its support Λ_r for the first half of its entries and on Δ_r for the second half of its entries, respectively, possesses a *quadratic* structure; otherwise $Z_{\mathbf{C},\mathfrak{D}}(\cdot)$ is #P-hard. We can express the quadratic structure as *a set of exponential difference equations* over bases which are appropriate roots of unity of orders equal to various prime powers. The constructions used in this part of the proof are the most demanding in the paper.

After all these necessary conditions, we finally show that if $\mathbf{C}$ and $\mathfrak{D}$ satisfy all these requirements, there is a polynomial-time algorithm for $\mathsf{EVAL}(\mathbf{C}, \mathfrak{D})$ and thus, $\mathsf{EVAL}(\mathbf{A})$ is also in polynomial time. To this end, we reduce $\mathsf{EVAL}(\mathbf{C}, \mathfrak{D})$ to $\mathsf{EVAL}(q)$ for some appropriate prime power q (which depends only on $\mathbf{C}$ and $\mathfrak{D}$). As noted earlier, the tractability of $\mathsf{EVAL}(q)$ is new and is of independent interest.

4. Pinning lemmas and preliminary reductions. We prove two pinning lemmas in this section, one for $\mathsf{EVAL}(\mathbf{A})$ and one for $\mathsf{EVAL}(\mathbf{C}, \mathfrak{D})$. The proof of the first lemma is very similar to that of the pinning lemma from [21], but the second one has some complications. We will prove a third pinning lemma in section 8.1.

4.1. A pinning lemma for $\mathsf{EVAL}(\mathbf{A})$. Let $\mathbf{A}$ be an $m \times m$ symmetric complex matrix. We define a new problem $\mathsf{EVALP}(\mathbf{A})$: The input is a triple (G, w, i), where $G = (V, E)$ is an undirected graph, $w \in V$ is a vertex, and $i \in [m]$; the output is

$$Z_{\mathbf{A}}(G,w,i) = \sum_{\substack{\xi:V\to[m]\\ \xi(w)=i}} \mathrm{wt}_{\mathbf{A}}(\xi).$$

It is easy to see that $\mathsf{EVAL}(\mathbf{A}) \le \mathsf{EVALP}(\mathbf{A})$. The other direction also holds.

LEMMA 4.1 (first pinning lemma). $\mathsf{EVALP}(\mathbf{A}) \equiv \mathsf{EVAL}(\mathbf{A})$.

Proof. We define an equivalence relation $\sim$ over $[m]$: $i \sim j$ if for any undirected graph $G=(V,E)$ and $w \in V$, $Z_{\mathbf{A}}(G,w,i) = Z_{\mathbf{A}}(G,w,j)$. Note that we do not know, given $\mathbf{A}$, how to compute $\sim$ efficiently, although this is possible using the new results of Schrijver [32]. Instead, the lemma only proves, nonconstructively, the existence of a polynomial-time reduction, which is sufficient for our purposes.

This relation divides the set $[m]$ into s equivalence classes $\mathcal{A}_1,\dots,\mathcal{A}_s$ for some positive integer s. For any distinct $t,t' \in [s]$, there exists a pair $P_{t,t'} = (G,w)$, where G is an undirected graph and w is a vertex of G, such that

$$Z_{\mathbf{A}}(G,w,i) = Z_{\mathbf{A}}(G,w,j) \neq Z_{\mathbf{A}}(G,w,i') = Z_{\mathbf{A}}(G,w,j')$$

for all $i,j \in \mathcal{A}_t$ and $i',j' \in \mathcal{A}_{t'}$. Again, we do not know how to compute such a pair efficiently, but it always exists by the definition of the equivalence relation $\sim$.

Now given any subset $S \subseteq [s]$, we define a problem $\mathsf{EVAL}(\mathbf{A},S)$. The input is a pair (G,w), where $G=(V,E)$ is an undirected graph and $w \in V$; the output is

$$Z_{\mathbf{A}}(G,w,S) = \sum_{\substack{\xi:V\to[m]\\ \xi(w)\in\bigcup_{t\in S}\mathcal{A}_t}} \mathrm{wt}_{\mathbf{A}}(\xi).$$

When $S=[s]$, $\mathsf{EVAL}(\mathbf{A},S)$ is exactly $\mathsf{EVAL}(\mathbf{A})$. We make the following claim.

CLAIM 4.2. *If $S \subseteq [s]$ and $|S| \ge 2$, then there exists a partition $\{S_1,\dots,S_k\}$ of S for some $k>1$ such that $\mathsf{EVAL}(\mathbf{A},S_d) \le \mathsf{EVAL}(\mathbf{A},S)$ for all $d \in [k]$.*

We use Claim 4.2 to prove Lemma 4.1. Let (G,w,i) be an input of $\mathsf{EVALP}(\mathbf{A})$, and let $i \in \mathcal{A}_t$ for some $t \in [s]$. We will use Claim 4.2 to prove that $\mathsf{EVAL}(\mathbf{A},\{t\}) \le \mathsf{EVAL}(\mathbf{A})$. If this is the case, then we are done because

$$Z_{\mathbf{A}}(G,w,i) = \frac{1}{|\mathcal{A}_t|} \cdot Z_{\mathbf{A}}(G,w,\{t\}).$$

Finally we show that $\mathsf{EVAL}(\mathbf{A},\{t\}) \le \mathsf{EVAL}(\mathbf{A})$. It is trivially true when $s=1$. When $s \ge 2$, by Claim 4.2 there exists a partition $\{S_1,\dots,S_k\}$ of S for some $k>1$, such that $\mathsf{EVAL}(\mathbf{A},S_d) \le \mathsf{EVAL}(\mathbf{A},S) \equiv \mathsf{EVAL}(\mathbf{A})$, for all $d \in [k]$. Without loss of generality, assume $t \in S_1$. If $S_1 = \{t\}$, then we are done; otherwise, $|S_1| \ge 2$, and we just rename S_1 to be S and repeat the process above. As $|S|$ is strictly decreasing after each iteration, this procedure will stop at some time. The lemma is proved. ∎

Proof of Claim 4.2. Let t,t' be two distinct integers in S. We let $P_{t,t'} = (G^*,w^*)$, where $G^* = (V^*,E^*)$. It defines the following equivalence relation $\sim^*$ over S: For $a,b \in S$, $a \sim^* b$ if $Z_{\mathbf{A}}(G^*,w^*,i) = Z_{\mathbf{A}}(G^*,w^*,j)$, where $i \in \mathcal{A}_a$ and $j \in \mathcal{A}_b$.

This equivalence relation $\sim^*$ is well-defined, being independent of our choices of $i \in \mathcal{A}_a, j \in \mathcal{A}_b$. It gives us equivalence classes $\{S_1,\dots,S_k\}$, a partition of S. Because $(G^*,w^*) = P_{t,t'}$, by the definition of $\sim^*$, t and t' belong to different classes and thus $k \ge 2$. For each $d \in [k]$, we let $X_d = Z_{\mathbf{A}}(G^*,w^*,i)$, where $i \in \mathcal{A}_a$ and $a \in S_d$. This number X_d is well-defined and is independent of the choices of $a \in S_d$ and $i \in \mathcal{A}_a$. Moreover, the definition of $\sim^*$ implies that $X_d \neq X_{d'}$ for all $d \neq d' \in [k]$.

Next, let G be an undirected graph and w be a vertex. We show that by querying $\mathsf{EVAL}(\mathbf{A}, S)$ as an oracle, one can compute $Z_{\mathbf{A}}(G, w, S_d)$ efficiently for all d. To this end, for each $p \in [0 : k-1]$ we construct a graph $G^{[p]} = (V^{[p]}, E^{[p]})$ as follows. $G^{[p]}$ is the disjoint union of G and p independent copies of G^*, except that the w in G and the w^*'s in all copies of G^* are identified as one single vertex $w' \in V^{[p]}$. Thus, we have $|V^{[p]}| = |V| + p \cdot |V^*| - p$. In particular, $G^{[0]} = G$.

From the construction of these graphs, we get the following equations:

$$Z_{\mathbf{A}}(G^{[p]}, w', S) = \sum_{d \in [k]} (X_d)^p \cdot Z_{\mathbf{A}}(G, w, S_d) \quad \text{for every } p \in [0 : k-1].$$

Since $X_d \neq X_{d'}$ for all $d \neq d'$, this is a Vandermonde system. We can solve it to get $Z_{\mathbf{A}}(G, w, S_d)$ for all d. As k and the size of G^* are constants that are independent of G, we get a polynomial-time reduction from $\mathsf{EVAL}(\mathbf{A}, S_d)$ to $\mathsf{EVAL}(\mathbf{A}, S)$. ∎

4.2. A pinning lemma for $\mathsf{EVAL}(\mathbf{C}, \mathfrak{D})$. Let $\mathbf{C} \in \mathbb{C}^{2m \times 2m}$ be the bipartization of $\mathbf{F} \in \mathbb{C}^{m \times m}$. Let $\mathfrak{D} = (\mathbf{D}^{[0]}, \ldots, \mathbf{D}^{[N-1]})$ be a sequence of N $2m \times 2m$ diagonal matrices. We define a problem $\mathsf{EVALP}(\mathbf{C}, \mathfrak{D})$: The input is a triple (G, w, i), where $G = (V, E)$ is an undirected graph, $w \in V$, and $i \in [2m]$; the output is

$$Z_{\mathbf{C},\mathfrak{D}}(G, w, i) = \sum_{\substack{\xi : V \to [2m] \\ \xi(w) = i}} \mathrm{wt}_{\mathbf{C},\mathfrak{D}}(\xi).$$

Clearly, $\mathsf{EVAL}(\mathbf{C}, \mathfrak{D}) \leq \mathsf{EVALP}(\mathbf{C}, \mathfrak{D})$. However, unlike $\mathsf{EVALP}(\mathbf{A})$ and $\mathsf{EVAL}(\mathbf{A})$, we can prove the other direction only when $(\mathbf{C}, \mathfrak{D})$ satisfies the following condition:

(*Pinning*) Every entry of $\mathbf{F}$ is a power of ω_N, where N denotes the number of matrices in $\mathfrak{D}$; $\mathbf{F}/\sqrt{m}$ is a unitary matrix, and $\mathbf{D}^{[0]}$ is the $2m \times 2m$ identity matrix.

LEMMA 4.3 (second pinning lemma). *If $(\mathbf{C}, \mathfrak{D})$ satisfies the condition* (*Pinning*) *above, then* $\mathsf{EVALP}(\mathbf{C}, \mathfrak{D}) \equiv \mathsf{EVAL}(\mathbf{C}, \mathfrak{D})$.

COROLLARY 4.4. *If $(\mathbf{C}, \mathfrak{D})$ satisfies the condition* (*Pinning*)*, then the problem of computing $Z^{\rightarrow}_{\mathbf{C},\mathfrak{D}}$ as well as $Z^{\leftarrow}_{\mathbf{C},\mathfrak{D}}$ is polynomial-time reducible to* $\mathsf{EVAL}(\mathbf{C}, \mathfrak{D})$.

Proof of Lemma 4.3. The proof structure is similar to that of Lemma 4.1. We start by introducing the following equivalence relation over $[2m]$: $i \sim j$ if for any undirected $G = (V, E)$ and $w \in V$, $Z_{\mathbf{C},\mathfrak{D}}(G, w, i) = Z_{\mathbf{C},\mathfrak{D}}(G, w, j)$. It partitions $[2m]$ into s equivalence classes $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_s$ for some $s \geq 1$. For any distinct $t, t' \in [s]$, there exists a pair $P_{t,t'} = (G, w)$, where G is an undirected graph and w is a vertex, such that for all $i, j \in \mathcal{A}_t$ and $i', j' \in \mathcal{A}_{t'}$,

$$Z_{\mathbf{C},\mathfrak{D}}(G, w, i) = Z_{\mathbf{C},\mathfrak{D}}(G, w, j) \neq Z_{\mathbf{C},\mathfrak{D}}(G, w, i') = Z_{\mathbf{C},\mathfrak{D}}(G, w, j').$$

Now for any subset $S \subseteq [s]$, we define $\mathsf{EVAL}(\mathbf{C}, \mathfrak{D}, S)$. The input is a pair (G, w), where $G = (V, E)$ is an undirected graph and w is a vertex in G; and the output is

$$Z_{\mathbf{C},\mathfrak{D}}(G, w, S) = \sum_{\substack{\xi : V \to [2m] \\ \xi(w) \in \bigcup_{t \in S} \mathcal{A}_t}} \mathrm{wt}_{\mathbf{C},\mathfrak{D}}(\xi).$$

When $S = [s]$, $\mathsf{EVAL}(\mathbf{C}, \mathfrak{D}, S)$ is exactly $\mathsf{EVAL}(\mathbf{C}, \mathfrak{D})$. We make the following claim.

CLAIM 4.5. *If $S \subseteq [s]$ and $|S| \geq 2$, there exists a partition $\{S_1, \ldots, S_k\}$ of S for some $k > 1$ such that $\mathsf{EVAL}(\mathbf{C}, \mathfrak{D}, S_d) \leq \mathsf{EVAL}(\mathbf{C}, \mathfrak{D}, S)$ for all $d \in [k]$.*

Lemma 4.3 then follows from Claim 4.5. The rest of the proof is exactly the same as that of Lemma 4.1 using Claim 4.2, so we omit it here. ∎

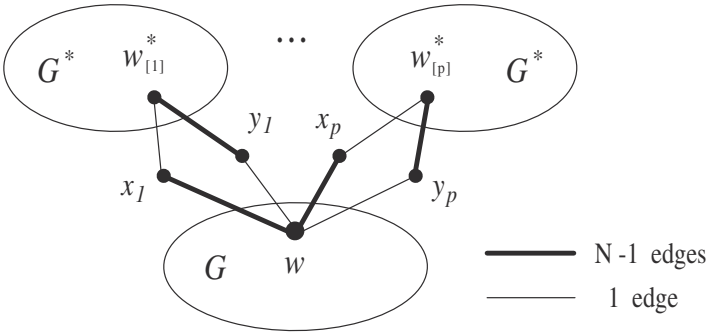


FIG. 4.1. *Graph* $G^{[p]}$, $p \in [0 : k-1]$.

Proof of Claim 4.5. Let t, t' be two distinct integers in S (as $|S| \geq 2$). Let $P_{t,t'} = (G^*, w^*)$, where $G^* = (V^*, E^*)$. It defines the following equivalence relation. For $a, b \in S$, $a \sim^* b$ if $Z_{\mathbf{C},\mathfrak{D}}(G^*, w^*, i) = Z_{\mathbf{C},\mathfrak{D}}(G^*, w^*, j)$, where $i \in \mathcal{A}_a$ and $j \in \mathcal{A}_b$.

This partitions S into equivalence classes $\{S_1, \ldots, S_k\}$. Because $(G^*, w^*) = P_{t,t'}$, t and t' must belong to different classes and thus we have $k \geq 2$. For each $d \in [k]$, we let $Y_d = Z_{\mathbf{C},\mathfrak{D}}(G^*, w^*, i)$, where $i \in \mathcal{A}_a$ and $a \in S_d$. The definition of the equivalence relation implies that $Y_d \neq Y_{d'}$ for all distinct $d, d' \in [k]$.

Now let G be an undirected graph and w be a vertex. We show that by querying $\mathsf{EVAL}(\mathbf{C}, \mathfrak{D}, S)$ as an oracle, one can compute $Z_{\mathbf{C},\mathfrak{D}}(G, w, S_d)$ efficiently for all d. To this end, for each integer $p \in [0 : k-1]$, we construct a graph $G^{[p]} = (V^{[p]}, E^{[p]})$ as follows: $G^{[p]}$ contains G and p independent copies of G^*. The vertex w in G is then *connected appropriately* to the w^* of each G^* (see Figure 4.1). More precisely,

$$V^{[p]} = V \cup \{v_i : i \in [p] \text{ and } v \in V^*\} \cup \{x_1, \ldots, x_p, y_1, \ldots, y_p\},$$

where $x_1, \ldots, x_p, y_1, \ldots, y_p$ are new vertices, and $E^{[p]}$ contains the following edges:

1. if $uv \in E$, then $uv \in E^{[p]}$; if $uv \in E^*$, then $u_i v_i \in E^{[p]}$ for all $i \in [p]$;
2. one edge between (w_i^*, x_i) and (y_i, w) for each $i \in [p]$; and
3. $N-1$ edges between (x_i, w) and (w_i^*, y_i) for each $i \in [p]$.

In particular, we have $G^{[0]} = G$.

We get the following equations. For $p \in [0 : k-1]$, $Z_{\mathbf{C},\mathfrak{D}}(G^{[p]}, w, S)$ is equal to

$$\sum_{\substack{i \in \bigcup_{a \in S} \mathcal{A}_a \\ i_1, \ldots, i_p \in [2m]}} Z_{\mathbf{C},\mathfrak{D}}(G, w, i) \left(\prod_{j=1}^{p} Z_{\mathbf{C},\mathfrak{D}}(G^*, w^*, i_j) \right) \prod_{j=1}^{p} \left(\sum_{x \in [2m]} C_{i_j,x} \overline{C_{i,x}} \sum_{y \in [2m]} \overline{C_{i_j,y}} C_{i,y} \right).$$

Note that $\deg(x_i) = \deg(y_i) = N$ and the changes to the degrees of w and w_i^* are all multiples of N. By (*Pinning*), there are no new vertex weight contributions from $\mathfrak{D}$.

Also by (*Pinning*), $\sum_{x \in [2m]} C_{i_j,x} \overline{C_{i,x}} = \langle \mathbf{F}_{i_j,*}, \mathbf{F}_{i,*} \rangle = 0$ unless $i = i_j$. Therefore,

$$\begin{aligned} Z_{\mathbf{C},\mathfrak{D}}(G^{[p]}, w, S) &= m^{2p} \cdot \sum_{i \in \bigcup_{a \in S} \mathcal{A}_a} Z_{\mathbf{C},\mathfrak{D}}(G, w, i) \cdot \left(Z_{\mathbf{C},\mathfrak{D}}(G^*, w^*, i) \right)^p \\ &= m^{2p} \cdot \sum_{d \in [k]} (Y_d)^p \cdot Z_{\mathbf{C},\mathfrak{D}}(G, w, S_d). \end{aligned}$$

Because $Y_d \neq Y_{d'}$ for all $d \neq d'$, this is a Vandermonde system and we can solve it to get $Z_{\mathbf{C},\mathfrak{D}}(G, w, S_d)$ for all d. As both k and the size of G^* are constants independent of G, this gives a reduction from $\mathsf{EVAL}(\mathbf{C}, \mathfrak{D}, S_d)$ to $\mathsf{EVAL}(\mathbf{C}, \mathfrak{D}, S)$ for every d. □

4.3. Reduction to connected matrices. The following lemma allows us to focus on the connected components of $\mathbf{A}$.

LEMMA 4.6. *Let $\mathbf{A} \in \mathbb{C}^{m\times m}$ be a symmetric matrix with components $\{\mathbf{A}_i\}$.*

1. *If $\mathsf{EVAL}(\mathbf{A}_i)$ is #P-hard for some $i \in [s]$, then $\mathsf{EVAL}(\mathbf{A})$ is #P-hard.*
2. *If $\mathsf{EVAL}(\mathbf{A}_i)$ is polynomial-time computable for every i, then so is $\mathsf{EVAL}(\mathbf{A})$.*

Proof. Lemma 4.6 follows from the first pinning lemma (Lemma 4.1). □

The main dichotomy, Theorem 1.1, will be proved by showing that for every connected $\mathbf{A} \in \mathbb{C}^{m\times m}$, $\mathsf{EVAL}(\mathbf{A})$ is either solvable in polynomial time or #P-hard.

5. Proof outline of the case: A is bipartite. We now give an overview of the proof of Theorem 1.1 for the case when $\mathbf{A}$ is connected and bipartite. The proof consists of two parts: a hardness part and a tractability part. The hardness part is further divided into three major steps in which we gradually "simplify" the problem being considered. In each of the three steps, we consider an EVAL problem passed down by the previous step (Step 1 starts with $\mathsf{EVAL}(\mathbf{A})$ itself) and show that

1. either the problem is #P-hard, or
2. the matrix that defines the problem satisfies certain structural properties, or
3. the problem is polynomial-time equivalent to a new EVAL problem, and the matrix that defines the new problem satisfies certain structural properties.

One can view these three steps as three filters that remove #P-hard $\mathsf{EVAL}(\mathbf{A})$ using different arguments. Finally, in the tractability part, we show that all the EVAL problems that survive the three filters are indeed polynomial-time solvable.

5.1. Step 1: Purification of matrix A. We start with $\mathsf{EVAL}(\mathbf{A})$, where $\mathbf{A} \in \mathbb{C}^{m\times m}$ is a fixed symmetric, connected, and bipartite matrix with *algebraic* entries. It is easy to see that if $m = 1$, then $\mathsf{EVAL}(\mathbf{A})$ is tractable. So in the discussion below, we always assume $m > 1$. In this step, we show that $\mathsf{EVAL}(\mathbf{A})$ is either #P-hard or polynomial-time equivalent to $\mathsf{EVAL}(\mathbf{A}')$, in which $\mathbf{A}'$ is also an $m \times m$ matrix but has a very nice structure.

DEFINITION 5.1. *Let $\mathbf{A} \in \mathbb{C}^{m\times m}$ be a symmetric, connected, and bipartite matrix. We say it is a* purified bipartite matrix *if there exist positive rational numbers $\mu_1, \ldots, \mu_m$ and an integer $1 \leq k < m$ such that*

1. *$A_{i,j} = 0$ for all $i, j \in [k]$; $A_{i,j} = 0$ for all $i, j \in [k+1 : m]$; and*
2. *$A_{i,j}/(\mu_i\mu_j) = A_{j,i}/(\mu_i\mu_j)$ is a root of unity for all $i \in [k]$, $j \in [k+1 : m]$.*

In other words, there exists a $k \times (m-k)$ matrix $\mathbf{B}$ of the form

$$\mathbf{B} = \begin{pmatrix} \mu_1 & & & \\ & \mu_2 & & \\ & & \ddots & \\ & & & \mu_k \end{pmatrix} \begin{pmatrix} \zeta_{1,1} & \zeta_{1,2} & \cdots & \zeta_{1,m-k} \\ \zeta_{2,1} & \zeta_{2,2} & \cdots & \zeta_{2,m-k} \\ \vdots & \vdots & \ddots & \vdots \\ \zeta_{k,1} & \zeta_{k,2} & \cdots & \zeta_{k,m-k} \end{pmatrix} \begin{pmatrix} \mu_{k+1} & & & \\ & \mu_{k+2} & & \\ & & \ddots & \\ & & & \mu_m \end{pmatrix},$$

where every μ_i is a positive rational number and every $\zeta_{i,j}$ is a root of unity, and $\mathbf{A}$ is the bipartization of $\mathbf{B}$.

THEOREM 5.2. *Let $\mathbf{A} \in \mathbb{C}^{m\times m}$ be a symmetric, connected, and bipartite matrix with algebraic entries. Then either $\mathsf{EVAL}(\mathbf{A})$ is #P-hard or there exists an $m \times m$ purified bipartite matrix $\mathbf{A}'$ such that $\mathsf{EVAL}(\mathbf{A}) \equiv \mathsf{EVAL}(\mathbf{A}')$. (By Definition 5.1, $\mathbf{A}'$ is symmetric and thus $\mathsf{EVAL}(\mathbf{A}')$ is well-defined.)*

5.2. Step 2: Reduction to discrete unitary matrix. Now let $\mathbf{A} \in \mathbb{C}^{m\times m}$ denote a purified bipartite matrix. Note that we renamed the $\mathbf{A}'$ passed down from Step 1 to $\mathbf{A}$ for convenience. We show that $\mathsf{EVAL}(\mathbf{A})$ is either #P-hard or polynomial-time equivalent to $\mathsf{EVAL}(\mathbf{C}, \mathfrak{D})$ for some $\mathbf{C}$ and $\mathfrak{D}$, where the matrix $\mathbf{C}$ is the bipartization of a discrete unitary matrix. (See section 3 for the definition.) Also note that the tensor product of two discrete unitary matrices is also discrete unitary.

THEOREM 5.3. *Given a purified bipartite matrix $\mathbf{A} \in \mathbb{C}^{m\times m}$, either* 1. $\mathsf{EVAL}(\mathbf{A})$ *is tractable; or* 2. $\mathsf{EVAL}(\mathbf{A})$ *is #P-hard; or* 3. *there exists a triple $((M, N), \mathbf{C}, \mathfrak{D})$ such that $\mathsf{EVAL}(\mathbf{A}) \equiv \mathsf{EVAL}(\mathbf{C}, \mathfrak{D})$, and $((M, N), \mathbf{C}, \mathfrak{D})$ satisfies the following conditions:*

($\mathcal{U}_1$) *$\mathbf{C} \in \mathbb{C}^{2n\times 2n}$ for some $n \geq 1$, and*

$$\mathfrak{D} = \left(\mathbf{D}^{[0]}, \mathbf{D}^{[1]}, \ldots, \mathbf{D}^{[N-1]}\right)$$

is a sequence of N $2n \times 2n$ diagonal matrices over $\mathbb{C}$ for some even *$N > 1$.*

($\mathcal{U}_2$) *$\mathbf{C}$ is the bipartization of an M-*discrete unitary matrix *$\mathbf{F} \in \mathbb{C}^{n\times n}$, where $M \geq 1$ and $M \mid N$. (Note that $\mathbf{C}$ and $\mathbf{F}$ uniquely determine each other.)*

($\mathcal{U}_3$) *$\mathbf{D}^{[0]}$ is the $2n \times 2n$ identity matrix, and for every $r \in [N-1]$ we have*

$$\exists i \in [n],\ D_i^{[r]} \neq 0 \implies \exists i' \in [n],\ D_{i'}^{[r]} = 1, \qquad \text{and}$$

$$\exists i \in [n+1:2n],\ D_i^{[r]} \neq 0 \implies \exists i' \in [n+1:2n],\ D_{i'}^{[r]} = 1.$$

($\mathcal{U}_4$) *For all $r \in [N-1]$ and all $i \in [2n]$, $D_i^{[r]} \in \mathbb{Q}(\omega_N)$ and $|D_i^{[r]}| \in \{0, 1\}$.*

5.3. Step 3: Canonical form of C, F, and $\mathfrak{D}$. After the first two steps, the original problem $\mathsf{EVAL}(\mathbf{A})$ is shown to be either tractable or #P-hard or polynomial-time equivalent to a new problem $\mathsf{EVAL}(\mathbf{C}, \mathfrak{D})$. There are also positive integers M and N such that $((M, N), \mathbf{C}, \mathfrak{D})$ satisfies conditions ($\mathcal{U}_1$)–($\mathcal{U}_4$).

For convenience, we still use $2m$ to denote the number of rows of $\mathbf{C}$ and $\mathbf{D}^{[r]}$, though it should be noted that this new m is indeed the n in Theorem 5.3, which is different from the m used in the first two steps. We also denote the upper-right $m \times m$ block of $\mathbf{C}$ by $\mathbf{F}$.

In this step, we adopt the following convention: Given an $n \times n$ matrix, we use $[0 : n-1]$, instead of $[n]$, to index its rows and columns. For example, we index the rows of $\mathbf{F}$ using $[0 : m-1]$ and index the rows of $\mathbf{C}$ using $[0 : 2m-1]$.

We start with the special case when $M = 1$. As $\mathbf{F}$ is M-discrete unitary, we must have $m = 1$. It is easy to check that $\mathsf{EVAL}(\mathbf{C}, \mathfrak{D})$ is tractable: $\mathbf{C}$ is a 2×2 matrix

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix};$$

$Z_{\mathbf{C},\mathfrak{D}}(G)$ is 0 unless G is bipartite; for connected and bipartite G, there are at most two assignments $\xi\colon V \to \{0, 1\}$ which could yield nonzero values; finally, for a graph G with connected components G_i $Z_{\mathbf{C},\mathfrak{D}}(G)$ is the product of $Z_{\mathbf{C},\mathfrak{D}}(G_i)$'s.

For the general case when the parameter $M > 1$ we further investigate the structure of $\mathbf{F}$ as well as the diagonal matrices in $\mathfrak{D}$ and derive three necessary conditions on them for $\mathsf{EVAL}(\mathbf{C}, \mathfrak{D})$ to not be #P-hard. In the tractability part, we prove that these conditions are actually sufficient for it to be polynomial-time computable.

5.3.1. Step 3.1: Entries of $\mathbf{D}^{[r]}$ are either 0 or powers of ω_N. In the first step, we prove the following theorem.

THEOREM 5.4. *Suppose $((M, N), \mathbf{C}, \mathfrak{D})$ satisfies ($\mathcal{U}_1$)–($\mathcal{U}_4$) with $M > 1$. Then either $\mathsf{EVAL}(\mathbf{C}, \mathfrak{D})$ is #P-hard or $((M, N), \mathbf{C}, \mathfrak{D})$ satisfies the following condition ($\mathcal{U}_5$):*

($\mathcal{U}_5$) *For all $r \in [N-1]$ and $i \in [0 : 2n-1]$, $D_i^{[r]}$ is either 0 or a power of ω_N.*

5.3.2. Step 3.2: Fourier decomposition. Second, we show that either problem $\mathsf{EVAL}(\mathbf{C}, \mathfrak{D})$ is #P-hard, or we can permute the rows and columns of $\mathbf{F}$, so that the new $\mathbf{F}$ is the tensor product of a collection of *Fourier matrices* defined below.

DEFINITION 5.5. *Let $q > 1$ be a prime power, and $k \geq 1$ be an integer such that* $\gcd(k, q) = 1$. *We call the following $q \times q$ matrix $\boldsymbol{\mathcal{F}}_{q,k}$ a* (q, k)-Fourier matrix*: The (x, y)th entry of $\boldsymbol{\mathcal{F}}_{q,k}$, where $x, y \in [0 : q-1]$, is*

$$\omega_q^{kxy} = e^{2\pi i\left(kxy/q\right)}.$$

In particular, when $k = 1$, we use $\boldsymbol{\mathcal{F}}_q$ to denote $\boldsymbol{\mathcal{F}}_{q,1}$ for short.

THEOREM 5.6. *Assume $((M, N), \mathbf{C}, \mathfrak{D})$ satisfies conditions $(\mathcal{U}_1)$–$(\mathcal{U}_5)$ and $M > 1$. Then either $\mathsf{EVAL}(\mathbf{C}, \mathfrak{D})$ is #P-hard or there exist permutations Σ and Π of $[0 : m-1]$ and a sequence $q_1, q_2, \ldots, q_d$ of d prime powers, for some $d \geq 1$, such that*

$$\mathbf{F}_{\Sigma,\Pi} = \bigotimes_{i \in [d]} \boldsymbol{\mathcal{F}}_{q_i}. \tag{5.1}$$

Suppose there do exist permutations Σ, Π and prime powers $q_1, \ldots, q_d$ such that $\mathbf{F}_{\Sigma,\Pi}$ satisfies (5.1). Then we let $\mathbf{C}_{\Sigma,\Pi}$ denote the bipartization of $\mathbf{F}_{\Sigma,\Pi}$ and let $\mathfrak{D}_{\Sigma,\Pi}$ denote a sequence of N $2m \times 2m$ diagonal matrices in which the rth matrix is

$$\begin{pmatrix} D^{[r]}_{\Sigma(0)} & & & & & & \\ & \ddots & & & & & \\ & & D^{[r]}_{\Sigma(m-1)} & & & & \\ & & & D^{[r]}_{\Pi(0)+m} & & & \\ & & & & \ddots & & \\ & & & & & D^{[r]}_{\Pi(m-1)+m} \end{pmatrix}, \quad r \in [0 : N-1].$$

Since permuting the rows and columns of $\mathbf{C}$ and $\mathbf{D}^{[r]}$ by the same permutation pair does not affect the complexity of $\mathsf{EVAL}(\mathbf{C}, \mathfrak{D})$, $\mathsf{EVAL}(\mathbf{C}_{\Sigma,\Pi}, \mathfrak{D}_{\Sigma,\Pi}) \equiv \mathsf{EVAL}(\mathbf{C}, \mathfrak{D})$. From now on, we let $\mathbf{F}, \mathbf{C}$, and $\mathfrak{D}$ denote $\mathbf{F}_{\Sigma,\Pi}, \mathbf{C}_{\Sigma,\Pi}$, and $\mathfrak{D}_{\Sigma,\Pi}$, respectively, with

$$\mathbf{F} = \bigotimes_{i \in [d]} \boldsymbol{\mathcal{F}}_{q_i}. \tag{5.2}$$

Before moving forward, we rearrange the prime powers $q_1, q_2, \ldots, q_d$ and divide them into groups according to different primes. We need the following notation. Let $\mathbf{p} = (p_1, \ldots, p_s)$ be a strictly increasing sequence of primes and $\mathbf{t} = (t_1, \ldots, t_s)$ be a sequence of positive integers. Let $\mathcal{Q} = \{\mathbf{q}_i : i \in [s]\}$ be a set of s sequences in which each $\mathbf{q}_i$ is a nonincreasing sequence $(q_{i,1}, \ldots, q_{i,t_i})$ of powers of p_i. We let q_i denote $q_{i,1}$ for all $i \in [s]$, let

$$\mathbb{Z}_{\mathbf{q}_i} = \prod_{j \in [t_i]} \mathbb{Z}_{q_{i,j}} = \mathbb{Z}_{q_{i,1}} \times \cdots \times \mathbb{Z}_{q_{i,t_i}}$$

for all $i \in [s]$, and let

$$\mathbb{Z}_{\mathcal{Q}} = \prod_{i \in [s], j \in [t_i]} \mathbb{Z}_{q_{i,j}} = \prod_{i \in [s]} \mathbb{Z}_{\mathbf{q}_i} = \mathbb{Z}_{q_{1,1}} \times \cdots \times \mathbb{Z}_{q_{1,t_1}} \times \cdots \times \mathbb{Z}_{q_{s,1}} \times \cdots \times \mathbb{Z}_{q_{s,t_s}}$$

be the Cartesian products of the respective finite Abelian groups. Both $\mathbb{Z}_{\mathcal{Q}}$ and $\mathbb{Z}_{\mathbf{q}_i}$ are finite Abelian groups under componentwise operations. This implies that both $\mathbb{Z}_{\mathcal{Q}}$ and $\mathbb{Z}_{\mathbf{q}_i}$ are $\mathbb{Z}$-modules and thus $k\mathbf{x}$ is well-defined for all $k \in \mathbb{Z}$ and $\mathbf{x}$ in $\mathbb{Z}_{\mathcal{Q}}$ or $\mathbb{Z}_{\mathbf{q}_i}$. As $\mathbb{Z}$-modules, we can also refer to their members as "vectors." When we use $\mathbf{x}$ to denote a vector in $\mathbb{Z}_{\mathcal{Q}}$, we denote its (i,j)th entry by $x_{i,j} \in \mathbb{Z}_{q_{i,j}}$. We use $\mathbf{x}_i$ to denote $(x_{i,j} : j \in [t_i]) \in \mathbb{Z}_{\mathbf{q}_i}$, so $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_s)$. Given $\mathbf{x}, \mathbf{y} \in \mathbb{Z}_{\mathcal{Q}}$, we let $\mathbf{x} \pm \mathbf{y}$ denote the vector in $\mathbb{Z}_{\mathcal{Q}}$ whose (i,j)th entry is $x_{i,j} \pm y_{i,j} \pmod{q_{i,j}}$. Similarly, for each $i \in [s]$, we can define $\mathbf{x} \pm \mathbf{y}$ for vectors $\mathbf{x}, \mathbf{y} \in \mathbb{Z}_{\mathbf{q}_i}$.

From (5.2), there exist $\mathbf{p}, \mathbf{t}, \mathcal{Q}$ such that $((M,N), \mathbf{C}, \mathfrak{D}, (\mathbf{p}, \mathbf{t}, \mathcal{Q}))$ satisfies the following three conditions $(\mathcal{R}_1)$–$(\mathcal{R}_3)$, which we refer to combined as $(\mathcal{R})$.

$(\mathcal{R}_1)$ $\mathbf{p} = (p_1, \ldots, p_s)$ is a strictly increasing sequence of primes; $\mathbf{t} = (t_1, \ldots, t_s)$ is a sequence of positive integers; $\mathcal{Q} = \{\mathbf{q}_i : i \in [s]\}$ is a collection of s sequences, in which each $\mathbf{q}_i = (q_{i,1}, \ldots, q_{i,t_i})$ is a nonincreasing sequence of powers of p_i.

$(\mathcal{R}_2)$ $\mathbf{C}$ is the bipartization of $\mathbf{F} \in \mathbb{C}^{m \times m}$ and $((M,N), \mathbf{C}, \mathfrak{D})$ satisfies $(\mathcal{U}_1)$–$(\mathcal{U}_5)$.

$(\mathcal{R}_3)$ There is a bijection $\rho\colon [0:m-1] \to \mathbb{Z}_{\mathcal{Q}}$ (so $m = \prod_{i,j} q_{i,j}$) such that

$$F_{a,b} = \prod_{i \in [s], j \in [t_i]} \omega_{q_{i,j}}^{x_{i,j}\, y_{i,j}} \quad \text{for all } a, b \in [0 : m-1], \tag{5.3}$$

where $(x_{i,j} : i \in [s], j \in [t_i]) = \mathbf{x} = \rho(a)$ and $(y_{i,j} : i \in [s], j \in [t_i]) = \mathbf{y} = \rho(b)$. Note that (5.3) also gives us an expression of M using $\mathcal{Q}$. It is the product of the largest prime powers $q_i = q_{i,1}$ for each distinct prime p_i: $M = q_1 q_2 \cdots q_s$.

For convenience, from now on we use $\mathbf{x} \in \mathbb{Z}_{\mathcal{Q}}$ to index rows and columns of $\mathbf{F}$:

$$F_{\mathbf{x},\mathbf{y}} = F_{\rho^{-1}(\mathbf{x}),\rho^{-1}(\mathbf{y})} = \prod_{i \in [s], j \in [t_i]} \omega_{q_{i,j}}^{x_{i,j}\, y_{i,j}} \quad \text{for all } \mathbf{x}, \mathbf{y} \in \mathbb{Z}_{\mathcal{Q}}, \tag{5.4}$$

whenever we have a tuple $((M,N), \mathbf{C}, \mathfrak{D}, (\mathbf{p}, \mathbf{t}, \mathcal{Q}))$ that is known to satisfy condition $(\mathcal{R})$. We assume that $\mathbf{F}$ is indexed by $(\mathbf{x}, \mathbf{y}) \in \mathbb{Z}_{\mathcal{Q}}^2$ rather than $(a,b) \in [0 : m-1]^2$ and that $(\mathcal{R}_3)$ refers to (5.4). Correspondingly, we use $\{0,1\} \times \mathbb{Z}_{\mathcal{Q}}$ to index the entries of matrices $\mathbf{C}$ and $\mathbf{D}^{[r]}$: $(0, \mathbf{x})$ refers to the $(\rho^{-1}(\mathbf{x}))$th row or column, and $(1, \mathbf{x})$ refers to the $(m + \rho^{-1}(\mathbf{x}))$th row or column.

5.3.3. Step 3.3: Affine support for $\mathfrak{D}$. Now we have a 4-tuple $((M,N), \mathbf{C}, \mathfrak{D}, (\mathbf{p}, \mathbf{t}, \mathcal{Q}))$ that satisfies $(\mathcal{R})$. In this step, we prove for every $r \in [N-1]$ (recall that $\mathbf{D}^{[0]}$ is already known to be the identity matrix), the nonzero entries of the rth matrix $\mathbf{D}^{[r]}$ in $\mathfrak{D}$ must have a very nice *coset* structure; otherwise $\mathsf{EVAL}(\mathbf{C}, \mathfrak{D})$ is #P-hard.

For every $r \in [N-1]$, we define $\Lambda_r \subseteq \mathbb{Z}_{\mathcal{Q}}$ and $\Delta_r \subseteq \mathbb{Z}_{\mathcal{Q}}$ as

$$\Lambda_r = \left\{\mathbf{x} \in \mathbb{Z}_{\mathcal{Q}} : D^{[r]}_{(0,\mathbf{x})} \neq 0\right\} \quad \text{and} \quad \Delta_r = \left\{\mathbf{x} \in \mathbb{Z}_{\mathcal{Q}} : D^{[r]}_{(1,\mathbf{x})} \neq 0\right\}.$$

We use $\mathcal{S}$ to denote the set of $r \in [N-1]$ such that $\Lambda_r \neq \emptyset$ and $\mathcal{T}$ to denote the set of $r \in [N-1]$ such that $\Delta_r \neq \emptyset$. We recall the following standard definition of a coset of a group, specialized to our situation.

DEFINITION 5.7. *Let Φ be a nonempty subset of $\mathbb{Z}_{\mathcal{Q}}$ (or $\mathbb{Z}_{\mathbf{q}_i}$ for some $i \in [s]$). We say Φ is a* coset *in $\mathbb{Z}_{\mathcal{Q}}$ (or $\mathbb{Z}_{\mathbf{q}_i}$) if there is a vector $\mathbf{x}_0 \in \Phi$ such that $\{\mathbf{x} - \mathbf{x}_0 \,|\, \mathbf{x} \in \Phi\}$ is a subgroup of $\mathbb{Z}_{\mathcal{Q}}$ (or $\mathbb{Z}_{\mathbf{q}_i}$). Given a coset Φ (in $\mathbb{Z}_{\mathcal{Q}}$ or $\mathbb{Z}_{\mathbf{q}_i}$), we use Φ^{lin} to denote its corresponding subgroup $\{\mathbf{x} - \mathbf{x}' \,|\, \mathbf{x}, \mathbf{x}' \in \Phi\}$.*

THEOREM 5.8. *Let $((M,N), \mathbf{C}, \mathfrak{D}, (\mathbf{p}, \mathbf{t}, \mathcal{Q}))$ be a 4-tuple that satisfies $(\mathcal{R})$. Then either $\mathsf{EVAL}(\mathbf{C}, \mathfrak{D})$ is #P-hard or $\Lambda_r, \Delta_r \subseteq \mathbb{Z}_{\mathcal{Q}}$ satisfy the following condition $(\mathcal{L})$:*

$(\mathcal{L}_1)$ For every $r \in \mathcal{S}$, $\Lambda_r = \prod_{i=1}^{s} \Lambda_{r,i}$, where $\Lambda_{r,i}$ is a coset in $\mathbb{Z}_{\mathbf{q}_i}$, $i \in [s]$.

$(\mathcal{L}_2)$ For every $r \in \mathcal{T}$, $\Delta_r = \prod_{i=1}^{s} \Delta_{r,i}$, where $\Delta_{r,i}$ is a coset in $\mathbb{Z}_{\mathbf{q}_i}$, $i \in [s]$.

Suppose $\mathsf{EVAL}(\mathbf{C}, \mathfrak{D})$ is not #P-hard. By Theorem 5.8, $((M,N), \mathbf{C}, \mathfrak{D}, (\mathbf{p}, \mathbf{t}, \mathcal{Q}))$ satisfies not only $(\mathcal{R})$ but also $(\mathcal{L})$. Actually, by $(\mathcal{U}_3)$, $\mathfrak{D}$ also satisfies the following:

$(\mathcal{L}_3)$ There exists an $\mathfrak{a}^{[r]} \in \Lambda_r$ for each $r \in \mathcal{S}$, a $\mathfrak{b}^{[r]} \in \Delta_r$ for each $r \in \mathcal{T}$ such that

$$D^{[r]}_{(0,\mathfrak{a}^{[r]})} = D^{[r]}_{(1,\mathfrak{b}^{[r]})} = 1.$$

From now on, when we say condition $(\mathcal{L})$, we mean all three conditions $(\mathcal{L}_1)$–$(\mathcal{L}_3)$.

5.3.4. Step 3.4: Quadratic structure. In this final step within Step 3, we prove that for every $r \in [N-1]$, the nonzero entries of $\mathbf{D}^{[r]}$ must have a *quadratic* structure; otherwise $\mathsf{EVAL}(\mathbf{C}, \mathfrak{D})$ is #P-hard. We start with some notation.

Given $\mathbf{x}$ in $\mathbb{Z}_{\mathbf{q}_i}$ for some $i \in [s]$, we use $\mathbf{ext}_r(\mathbf{x})$ (extension of $\mathbf{x}$ for short), where $r \in \mathcal{S}$, to denote the following unique vector:

$$\left(\mathfrak{a}^{[r]}_1, \ldots, \mathfrak{a}^{[r]}_{i-1}, \mathbf{x}, \mathfrak{a}^{[r]}_{i+1}, \ldots, \mathfrak{a}^{[r]}_s\right) \in \mathbb{Z}_{\mathcal{Q}}.$$

Similarly we let $\mathbf{ext}'_r(\mathbf{x})$, where $r \in \mathcal{T}$, denote the following unique vector:

$$\left(\mathfrak{b}^{[r]}_1, \ldots, \mathfrak{b}^{[r]}_{i-1}, \mathbf{x}, \mathfrak{b}^{[r]}_{i+1}, \ldots, \mathfrak{b}^{[r]}_s\right) \in \mathbb{Z}_{\mathcal{Q}}.$$

Let $\mathbf{a}$ be a vector in $\mathbb{Z}_{\mathbf{q}_i}$ for some $i \in [s]$. Then we use $\widetilde{\mathbf{a}}$ to denote the vector $\mathbf{b} \in \mathbb{Z}_{\mathcal{Q}}$ such that $\mathbf{b}_i = \mathbf{a}$ and $\mathbf{b}_j = \mathbf{0}$ for all other $j \neq i$. Also recall that $q_k = q_{k,1}$.

THEOREM 5.9. *Let $((M,N), \mathbf{C}, \mathfrak{D}, (\mathbf{p}, \mathbf{t}, \mathcal{Q}))$ be a tuple that satisfies both $(\mathcal{R})$ and $(\mathcal{L})$. Then either $\mathsf{EVAL}(\mathbf{C}, \mathfrak{D})$ is #P-hard, or $\mathfrak{D}$ satisfies the following condition $(\mathcal{D})$:*

$(\mathcal{D}_1)$ *For all $r \in \mathcal{S}$ and $\mathbf{x} \in \Lambda_r$, we have*

$$D^{[r]}_{(0,\mathbf{x})} = D^{[r]}_{(0,\mathbf{ext}_r(\mathbf{x}_1))} D^{[r]}_{(0,\mathbf{ext}_r(\mathbf{x}_2))} \cdots D^{[r]}_{(0,\mathbf{ext}_r(\mathbf{x}_s))}. \tag{5.5}$$

$(\mathcal{D}_2)$ *For all $r \in \mathcal{T}$ and $\mathbf{x} \in \Delta_r$, we have*

$$D^{[r]}_{(1,\mathbf{x})} = D^{[r]}_{(1,\mathbf{ext}'_r(\mathbf{x}_1))} D^{[r]}_{(1,\mathbf{ext}'_r(\mathbf{x}_2))} \cdots D^{[r]}_{(1,\mathbf{ext}'_r(\mathbf{x}_s))}. \tag{5.6}$$

$(\mathcal{D}_3)$ *For all $r \in \mathcal{S}$, $k \in [s]$, and $\mathbf{a} \in \Lambda^{\mathrm{lin}}_{r,k}$, there are $\mathbf{b} \in \mathbb{Z}_{\mathbf{q}_k}$ and $\alpha \in \mathbb{Z}_N$ such that*

$$\omega_N^{\alpha} \cdot F_{\mathbf{x},\widetilde{\mathbf{b}}} = D^{[r]}_{(0,\mathbf{x}+\widetilde{\mathbf{a}})} \cdot \overline{D^{[r]}_{(0,\mathbf{x})}} \quad \text{for all } \mathbf{x} \in \Lambda_r. \tag{5.7}$$

$(\mathcal{D}_4)$ *For all $r \in \mathcal{T}$, $k \in [s]$, and $\mathbf{a} \in \Delta^{\mathrm{lin}}_{r,k}$, there are $\mathbf{b} \in \mathbb{Z}_{\mathbf{q}_k}$ and $\alpha \in \mathbb{Z}_N$ such that*

$$\omega_N^{\alpha} \cdot F_{\widetilde{\mathbf{b}},\mathbf{x}} = D^{[r]}_{(1,\mathbf{x}+\widetilde{\mathbf{a}})} \cdot \overline{D^{[r]}_{(1,\mathbf{x})}} \quad \text{for all } \mathbf{x} \in \Delta_r. \tag{5.8}$$

Note that in $(\mathcal{D}_3)$ and $(\mathcal{D}_4)$, the expressions on the left-hand side do not depend on all other components of $\mathbf{x}$ except the kth component $\mathbf{x}_k$, since all other components of $\widetilde{\mathbf{b}}$ are $\mathbf{0}$. The statements in conditions $(\mathcal{D}_3)$–$(\mathcal{D}_4)$ are a technically precise way to express the idea that there is a quadratic structure on the support of each diagonal matrix $\mathbf{D}^{[r]}$. We express it in terms of an exponential difference equation.

5.4. Tractability. Now we can state a theorem of tractability.

THEOREM 5.10. *Suppose that $((M,N), \mathbf{C}, \mathfrak{D}, (\mathbf{p}, \mathbf{t}, \mathcal{Q}))$ satisfies $(\mathcal{R}), (\mathcal{L})$, and $(\mathcal{D})$. Then the problem $\mathsf{EVAL}(\mathbf{C}, \mathfrak{D})$ can be solved in polynomial time.*

6. Proof outline of the case: A is not bipartite. Both the definitions and the theorems of the case when the fixed matrix $\mathbf{A}$ is not bipartite are similar to, but also have significant differences from, those of the bipartite case.

6.1. Step 1: Purification of matrix A. We start with $\mathbf{A} \in \mathbb{C}^{m\times m}$, a symmetric, connected, and nonbipartite matrix with algebraic entries. In the discussion below, we assume $m > 1$; $\mathsf{EVAL}(\mathbf{A})$ is clearly tractable if $m = 1$.

DEFINITION 6.1. *Let $\mathbf{A} \in \mathbb{C}^{m\times m}$ be a symmetric matrix. We say $\mathbf{A}$ is a* purified *nonbipartite matrix if there exist positive rational numbers $\mu_1, \mu_2, \ldots, \mu_m$ such that $A_{i,j}/(\mu_i\mu_j)$ is a root of unity for all $i, j \in [m]$.*

In other words, $\mathbf{A}$ has the form

$$\mathbf{A} = \begin{pmatrix} \mu_1 & & & \\ & \mu_2 & & \\ & & \ddots & \\ & & & \mu_m \end{pmatrix} \begin{pmatrix} \zeta_{1,1} & \zeta_{1,2} & \cdots & \zeta_{1,m} \\ \zeta_{2,1} & \zeta_{2,2} & \cdots & \zeta_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ \zeta_{m,1} & \zeta_{m,2} & \cdots & \zeta_{m,m} \end{pmatrix} \begin{pmatrix} \mu_1 & & & \\ & \mu_2 & & \\ & & \ddots & \\ & & & \mu_m \end{pmatrix},$$

where $\zeta_{i,j} = \zeta_{j,i}$ are all roots of unity. We prove the following theorem.

THEOREM 6.2. *Let $\mathbf{A} \in \mathbb{C}^{m\times m}$ be a symmetric, connected, and nonbipartite matrix, where $m > 1$. Then either $\mathsf{EVAL}(\mathbf{A})$ is #P-hard or there exists a purified nonbipartite matrix $\mathbf{A}' \in \mathbb{C}^{m\times m}$ such that $\mathsf{EVAL}(\mathbf{A}) \equiv \mathsf{EVAL}(\mathbf{A}')$.*

6.2. Step 2: Reduction to discrete unitary matrix.

THEOREM 6.3. *Let $\mathbf{A} \in \mathbb{C}^{m\times m}$ be a purified nonbipartite matrix. Then either* (1) *$\mathsf{EVAL}(\mathbf{A})$ is tractable or* (2) *$\mathsf{EVAL}(\mathbf{A})$ is #P-hard or* (3) *there exists a triple $((M, N), \mathbf{F}, \mathfrak{D})$ such that $\mathsf{EVAL}(\mathbf{A}) \equiv \mathsf{EVAL}(\mathbf{F}, \mathfrak{D})$ and $((M, N), \mathbf{F}, \mathfrak{D})$ satisfies $(\mathcal{U}'_1)$–$(\mathcal{U}'_4)$:*

$(\mathcal{U}'_1)$ $\mathbf{F} \in \mathbb{C}^{n\times n}$ for some $n \geq 1$, and $\mathfrak{D} = (\mathbf{D}^{[0]}, \ldots, \mathbf{D}^{[N-1]})$ is a sequence of N $n \times n$ diagonal matrices for some even *$N > 1$.*

$(\mathcal{U}'_2)$ $\mathbf{F}$ is a symmetric M-discrete unitary matrix, where $M \geq 1$ and $M \,|\, N$.

$(\mathcal{U}'_3)$ $\mathbf{D}^{[0]}$ is the identity matrix. For each $r \in [N-1]$, either $\mathbf{D}^{[r]} = \mathbf{0}$ or $\mathbf{D}^{[r]}$ has an entry equal to 1*.*

$(\mathcal{U}'_4)$ For all $r \in [N-1]$ and $i \in [n]$, $D_i^{[r]} \in \mathbb{Q}(\omega_N)$ and $|D_i^{[r]}| \in \{0, 1\}$.

6.3. Step 3: Canonical form of F and $\mathfrak{D}$. Now suppose we have a tuple $((M, N), \mathbf{F}, \mathfrak{D})$ that satisfies $(\mathcal{U}'_1)$–$(\mathcal{U}'_4)$. For convenience we still use m to denote the number of rows and columns of $\mathbf{F}$ and each $\mathbf{D}^{[r]}$ in $\mathfrak{D}$, though it should be noted that this new m is indeed the n in Theorem 6.3, which is different from the m used in the first two steps. Similar to the bipartite case, we adopt the following convention in this step: given an $n \times n$ matrix, we use $[0 : n-1]$, instead of $[n]$, to index its rows and columns.

We start with the special case when $M = 1$. Since $\mathbf{F}$ is M-discrete unitary, we must have $m = 1$ and $\mathbf{F} = (1)$. In this case, it is clear that the problem $\mathsf{EVAL}(\mathbf{C}, \mathfrak{D})$ is tractable. So in the rest of this section, we always assume that $M > 1$.

6.3.1. Step 3.1: Entries of $\mathbf{D}^{[r]}$ are either 0 or powers of ω_N.

THEOREM 6.4. *Suppose $((M, N), \mathbf{F}, \mathfrak{D})$ satisfies $(\mathcal{U}'_1)$–$(\mathcal{U}'_4)$ and $M > 1$. Then either $\mathsf{EVAL}(\mathbf{F}, \mathfrak{D})$ is #P-hard or $((M, N), \mathbf{F}, \mathfrak{D})$ satisfies the following condition $(\mathcal{U}'_5)$:*

$(\mathcal{U}'_5)$ For all $r \in [N-1]$, entries of $\mathbf{D}^{[r]}$ are either zero or powers of ω_N.

6.3.2. Step 3.2: Fourier decomposition. Let q be a prime power. We say $\mathbf{W}$ is a *nondegenerate* matrix in $\mathbb{Z}_q^{2\times 2}$ if $\mathbf{W}\mathbf{x} \neq \mathbf{0}$ for all $\mathbf{x} \neq \mathbf{0} \in \mathbb{Z}_q^2$. The following

lemma gives some equivalent characterizations of nondegenerate matrices. The proof is elementary, so we omit it here.

LEMMA 6.5. *Let q be a prime power and $\mathbf{W} \in \mathbb{Z}_q^{2\times 2}$. The following statements are equivalent:* (1) $\mathbf{W}$ *is nondegenerate;* (2) $\mathbf{x} \mapsto \mathbf{W}\mathbf{x}$ *is a bijection from $\mathbb{Z}_q^2$ to itself; and* (3) $\det(\mathbf{W})$ *is invertible in $\mathbb{Z}_q$.*

DEFINITION 6.6 (generalized Fourier matrix). *Let q be a prime power and $\mathbf{W} = (W_{ij})$ be a symmetric nondegenerate matrix in $\mathbb{Z}_q^{2\times 2}$. We say a $q^2 \times q^2$ matrix $\boldsymbol{\mathcal{F}}_{q,\mathbf{W}}$ is a $(q, \mathbf{W})$*-generalized Fourier matrix *if there exists a bijection ρ from $[0 : q^2 - 1]$ to $[0 : q - 1]^2$ such that*

$$(\boldsymbol{\mathcal{F}}_{q,\mathbf{W}})_{i,j} = \omega_q^{W_{11}x_1y_1+W_{12}x_1y_2+W_{21}x_2y_1+W_{22}x_2y_2} \quad \text{for all } i, j \in [0 : q^2 - 1],$$

where $\mathbf{x} = (x_1, x_2) = \rho(i)$ and $\mathbf{y} = (y_1, y_2) = \rho(j)$.

THEOREM 6.7. *Suppose $((M, N), \mathbf{F}, \mathfrak{D})$ satisfies conditions $(\mathcal{U}_1')$–$(\mathcal{U}_5')$. Then either $\mathsf{EVAL}(\mathbf{F}, \mathfrak{D})$ is #P-hard or there exists a permutation Σ of $[0 : m - 1]$ such that*

$$\mathbf{F}_{\Sigma,\Sigma} = \left(\bigotimes_{i=1}^{g} \boldsymbol{\mathcal{F}}_{d_i,\mathbf{W}^{[i]}}\right) \otimes \left(\bigotimes_{i=1}^{\ell} \boldsymbol{\mathcal{F}}_{q_i,k_i}\right),$$

where $\mathbf{d} = (d_1, \ldots, d_g)$ and $\mathcal{W} = (\mathbf{W}^{[1]}, \ldots, \mathbf{W}^{[g]})$ are two sequences, for some $g \geq 0$. (Note that the g here can be 0, in which case $\mathbf{d}$ and $\mathcal{W}$ are empty.) For each $i \in [g]$, $d_i > 1$ is a power of 2 and $\mathbf{W}^{[i]}$ is a 2×2 symmetric nondegenerate matrix over $\mathbb{Z}_{d_i}$; $\mathbf{q} = (q_1, \ldots, q_\ell)$ and $\mathbf{k} = (k_1, \ldots, k_\ell)$ are two sequences for some $\ell \geq 0$ (again ℓ can be 0). For each $i \in [\ell]$, q_i is a prime power, $k_i \in \mathbb{Z}_{q_i}$, and $\gcd(q_i, k_i) = 1$.

Assume there does exist a permutation Σ, together with the four sequences, such that $\mathbf{F}_{\Sigma,\Sigma}$ satisfies the equation above; otherwise, $\mathsf{EVAL}(\mathbf{F}, \mathfrak{D})$ is #P-hard. Then we apply Σ to $\mathbf{D}^{[r]}$, $r \in [0 : N - 1]$, to get a new sequence $\mathfrak{D}_\Sigma$ of N diagonal matrices in which the rth matrix of $\mathfrak{D}_\Sigma$ is

$$\begin{pmatrix} D^{[r]}_{\Sigma(0)} & & \\ & \ddots & \\ & & D^{[r]}_{\Sigma(m-1)} \end{pmatrix}.$$

It is clear that $\mathsf{EVAL}(\mathbf{F}_{\Sigma,\Sigma}, \mathfrak{D}_\Sigma) \equiv \mathsf{EVAL}(\mathbf{F}, \mathfrak{D})$. From now on, we simply let $\mathbf{F}$ and $\mathfrak{D}$ denote $\mathbf{F}_{\Sigma,\Sigma}$ and $\mathfrak{D}_\Sigma$, respectively. Thus, we have

$$\mathbf{F} = \left(\bigotimes_{i=1}^{g} \boldsymbol{\mathcal{F}}_{d_i,\mathbf{W}^{[i]}}\right) \otimes \left(\bigotimes_{i=1}^{\ell} \boldsymbol{\mathcal{F}}_{q_i,k_i}\right). \tag{6.1}$$

Before moving forward to Step 3.3, we rearrange the prime powers in $\mathbf{d}$ and $\mathbf{q}$ and divide them into groups according to different primes.

By (6.1), there exist $\mathbf{d}, \mathcal{W}, \mathbf{p}, \mathbf{t}, \mathcal{Q}$, and $\mathcal{K}$ such that tuple $((M, N), \mathbf{F}, \mathfrak{D}, (\mathbf{d}, \mathcal{W}, \mathbf{p}, \mathbf{t}, \mathcal{Q}, \mathcal{K}))$ satisfies the following condition $(\mathcal{R}')$:

$(\mathcal{R}_1')$ $\mathbf{d} = (d_1, \ldots, d_g)$ is a nonincreasing sequence of powers of 2 for some $g \geq 0$; $\mathcal{W} = (\mathbf{W}^{[1]}, \ldots, \mathbf{W}^{[g]})$ is a sequence of symmetric nondegenerate 2×2 matrices over $\mathbb{Z}_{d_i}$ (note that $\mathbf{d}$ and $\mathcal{W}$ can be empty); $\mathbf{p} = (p_1, \ldots, p_s)$ is a strictly increasing sequence of s primes for some $s \geq 1$, starting with $p_1 = 2$; $\mathbf{t} = (t_1, \ldots, t_s)$ is a sequence of integers with $t_1 \geq 0$ and $t_i \geq 1$ for all $i > 1$; $\mathcal{Q} = \{\mathbf{q}_i : i \in [s]\}$ is a collection of

sequences in which each $\mathbf{q}_i = (q_{i,1}, \ldots, q_{i,t_i})$ is a nonincreasing sequence of powers of p_i (only $\mathbf{q}_1$ can be empty as we always fix $p_1 = 2$ even when no powers of 2 occur in $\mathcal{Q}$); $\mathcal{K} = \{\mathbf{k}_i : i \in [s]\}$ is a collection of sequences in which each $\mathbf{k}_i = (k_{i,1}, \ldots, k_{i,t_i})$ is a sequence of length t_i. Finally, for all $i \in [s]$ and $j \in [t_i]$, $k_{i,j} \in [0 : q_{i,j} - 1]$ and satisfies $\gcd(k_{i,j}, q_{i,j}) = \gcd(k_{i,j}, p_i) = 1$.

$(\mathcal{R}'_2)$ $((M, N), \mathbf{F}, \mathfrak{D})$ satisfies conditions $(\mathcal{U}'_1)$–$(\mathcal{U}'_5)$, and

$$m = \prod_{i \in [g]} (d_i)^2 \times \prod_{i \in [s], j \in [t_i]} q_{i,j}\,.$$

$(\mathcal{R}'_3)$ There is a bijection ρ from $[0 : m - 1]$ to $\mathbb{Z}^2_{\mathbf{d}} \times \mathbb{Z}_{\mathcal{Q}}$, where

$$\mathbb{Z}^2_{\mathbf{d}} = \prod_{i \in [g]} (\mathbb{Z}_{d_i})^2 \quad \text{and} \quad \mathbb{Z}_{\mathcal{Q}} = \prod_{i \in [s], j \in [t_i]} \mathbb{Z}_{q_{i,j}},$$

such that (for each $a \in [0 : m - 1]$, we use

$$\big(x_{0,i,j} : i \in [g], j \in \{1, 2\}\big) \in \mathbb{Z}^2_{\mathbf{d}} \quad \text{and} \quad \big(x_{1,i,j} : i \in [s], j \in [t_i]\big) \in \mathbb{Z}_{\mathcal{Q}}$$

to denote the components of $\mathbf{x} = \rho(a)$, where $x_{0,i,j} \in \mathbb{Z}_{d_i}$ and $x_{1,i,j} \in \mathbb{Z}_{q_{i,j}}$)

$$F_{a,b} = \prod_{i \in [g]} \omega_{d_i}^{(x_{0,i,1}\ x_{0,i,2}) \cdot \mathbf{W}^{[i]} \cdot (y_{0,i,1}\ y_{0,i,2})^{\mathrm{T}}} \prod_{i \in [s], j \in [t_i]} \omega_{q_{i,j}}^{k_{i,j} \cdot x_{1,i,j} y_{1,i,j}}$$

for all $a, b \in [0 : m - 1]$, where $((x_{0,i,j}), (x_{1,i,j})) = \mathbf{x} = \rho(a)$ and $\mathbf{y} = \rho(b)$.

For convenience, from now on we will directly use $\mathbf{x} \in \mathbb{Z}^2_{\mathbf{d}} \times \mathbb{Z}_{\mathcal{Q}}$ to index the rows and columns of $\mathbf{F}$, i.e., $F_{\mathbf{x},\mathbf{y}} \equiv F_{\rho^{-1}(\mathbf{x}), \rho^{-1}(\mathbf{y})}$.

6.3.3. Step 3.3: Affine support for $\mathfrak{D}$. Now we have a tuple $((M, N), \mathbf{F}, \mathfrak{D}, (\mathbf{d}, \mathcal{W}, \mathbf{p}, \mathbf{t}, \mathcal{Q}, \mathcal{K}))$ that satisfies $(\mathcal{R}')$. In the next step, we show for every $r \in [N - 1]$ ($\mathbf{D}^{[0]}$ is already known to be the identity matrix) the nonzero entries of $\mathbf{D}^{[r]}$ (in $\mathfrak{D}$) must have a coset structure; otherwise $\mathsf{EVAL}(\mathbf{F}, \mathfrak{D})$ is #P-hard.

For each $r \in [N - 1]$, let $\Gamma_r \subseteq \mathbb{Z}^2_{\mathbf{d}} \times \mathbb{Z}_{\mathcal{Q}}$ denote the set of $\mathbf{x}$ such that the entry of $\mathbf{D}^{[r]}$ indexed by $\mathbf{x}$ is nonzero. We also use $\mathcal{Z}$ to denote the set of $r \in [N - 1]$ such that $\Gamma_r \neq \emptyset$. For convenience, we let $\tilde{\mathbb{Z}}_{\mathbf{q}_i}$, $i \in [s]$, denote the following set (or group):

$$\tilde{\mathbb{Z}}_{\mathbf{q}_i} = \begin{cases} \mathbb{Z}_{\mathbf{q}_i} & \text{if } i > 1, \\ \mathbb{Z}^2_{\mathbf{d}} \times \mathbb{Z}_{\mathbf{q}_1} & \text{if } i = 1. \end{cases}$$

This gives us a new way to denote the components of

$$\mathbf{x} \in \mathbb{Z}^2_{\mathbf{d}} \times \mathbb{Z}_{\mathcal{Q}} = \tilde{\mathbb{Z}}_{\mathbf{q}_1} \times \tilde{\mathbb{Z}}_{\mathbf{q}_2} \times \cdots \times \tilde{\mathbb{Z}}_{\mathbf{q}_s},$$

i.e., $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_s)$, where $\mathbf{x}_i \in \tilde{\mathbb{Z}}_{\mathbf{q}_i}$ for each $i \in [s]$.

THEOREM 6.8. *Assume that* $((M, N), \mathbf{F}, \mathfrak{D}, (\mathbf{d}, \mathcal{W}, \mathbf{p}, \mathbf{t}, \mathcal{Q}, \mathcal{K}))$ *satisfies condition* $(\mathcal{R}')$. *Then either* $\mathsf{EVAL}(\mathbf{F}, \mathfrak{D})$ *is #P-hard or* $\mathfrak{D}$ *satisfies the following condition:*

$(\mathcal{L}'_1)$ *For every* $r \in \mathcal{Z}$, $\Gamma_r = \prod_{i=1}^{s} \Gamma_{r,i}$, *where* $\Gamma_{r,i}$ *is a coset in* $\tilde{\mathbb{Z}}_{\mathbf{q}_i}$ *for all* $i \in [s]$.

Suppose $\mathsf{EVAL}(\mathbf{F}, \mathfrak{D})$ is not #P-hard. Then by Theorem 6.8, tuple $((M, N), \mathbf{F}, \mathfrak{D}, (\mathbf{d}, \mathcal{W}, \mathbf{p}, \mathbf{t}, \mathcal{Q}, \mathcal{K}))$ satisfies not only $(\mathcal{R}')$ but also $(\mathcal{L}'_1)$. By $(\mathcal{U}'_3)$, $\mathfrak{D}$ also satisfies the following:

$(\mathcal{L}'_2)$ For every $r \in \mathcal{Z}$, there exists an $\mathfrak{a}^{[r]} \in \Gamma_r \subseteq \mathbb{Z}^2_{\mathbf{d}} \times \mathbb{Z}_{\mathcal{Q}}$ such that the entry of $\mathbf{D}^{[r]}$ indexed by $\mathfrak{a}^{[r]}$ is equal to 1.

From now on, we refer to conditions $(\mathcal{L}'_1)$ and $(\mathcal{L}'_2)$ as condition $(\mathcal{L}')$.

6.3.4. Step 3.4: Quadratic structure. In this final step within Step 3 for the nonbipartite case, we show that for any index $r \in [N-1]$, the nonzero entries of $\mathbf{D}^{[r]}$ must have a quadratic structure; otherwise $\mathsf{EVAL}(\mathbf{F}, \mathfrak{D})$ is #P-hard.

We need the following notation. Given $\mathbf{x}$ in $\tilde{\mathbb{Z}}_{\mathbf{q}_i}$ for some $i \in [s]$, we let $\mathbf{ext}_r(\mathbf{x})$, where $r \in \mathcal{Z}$, denote the following unique vector:

$$\left(\mathfrak{a}_1^{[r]}, \ldots, \mathfrak{a}_{i-1}^{[r]}, \mathbf{x}, \mathfrak{a}_{i+1}^{[r]}, \ldots, \mathfrak{a}_s^{[r]}\right) \in \prod_{j \in [s]} \tilde{\mathbb{Z}}_{\mathbf{q}_j}.$$

Given $\mathbf{a} \in \tilde{\mathbb{Z}}_{\mathbf{q}_i}$ for some $i \in [s]$, we let $\widetilde{\mathbf{a}} = (\widetilde{\mathbf{a}}_1, \ldots, \widetilde{\mathbf{a}}_s) \in \prod_{j \in [s]} \tilde{\mathbb{Z}}_{\mathbf{q}_j}$ such that $\widetilde{\mathbf{a}}_i = \mathbf{a}$ and all other components are $\mathbf{0}$.

THEOREM 6.9. *Suppose* $((M, N), \mathbf{F}, \mathfrak{D}, (\mathbf{d}, \mathcal{W}, \mathbf{p}, \mathbf{t}, \mathcal{Q}, \mathcal{K}))$ *satisfies* $(\mathcal{R}')$ *and* $(\mathcal{L})$. *Then either* $\mathsf{EVAL}(\mathbf{F}, \mathfrak{D})$ *is #P-hard or* $\mathfrak{D}$ *satisfies the following condition* $(\mathcal{D}')$*:*

$(\mathcal{D}'_1)$ *For all* $r \in \mathcal{Z}$ *and* $\mathbf{x} \in \Gamma_r$*, we have*

$$D^{[r]}_{\mathbf{x}} = D^{[r]}_{\mathbf{ext}_r(\mathbf{x}_1)} D^{[r]}_{\mathbf{ext}_r(\mathbf{x}_2)} \cdots D^{[r]}_{\mathbf{ext}_r(\mathbf{x}_s)}. \tag{6.2}$$

$(\mathcal{D}'_2)$ *For all* $r \in \mathcal{Z}$*,* $k \in [s]$*, and* $\mathbf{a} \in \Gamma^{\mathrm{lin}}_{r,k}$*, there are* $\mathbf{b} \in \tilde{\mathbb{Z}}_{\mathbf{q}_k}$ *and* $\alpha \in \mathbb{Z}_N$ *such that*

$$\omega_N^{\alpha} \cdot F^{\sim}_{\mathbf{b},\mathbf{x}} = D^{[r]}_{\mathbf{x}+\widetilde{\mathbf{a}}} \cdot \overline{D^{[r]}_{\mathbf{x}}} \quad \text{for all } \mathbf{x} \in \Gamma_r. \tag{6.3}$$

Note that in (6.3), the expression on the left-hand side does not depend on other components of $\mathbf{x}$ except the kth component $\mathbf{x}_k \in \tilde{\mathbb{Z}}_{\mathbf{q}_k}$.

6.4. Tractability.

THEOREM 6.10. *Let* $((M, N), \mathbf{F}, \mathfrak{D}, (\mathbf{d}, \mathcal{W}, \mathbf{p}, \mathbf{t}, \mathcal{Q}, \mathcal{K}))$ *be a tuple that satisfies all conditions* $(\mathcal{R}'), (\mathcal{L}')$*, and* $(\mathcal{D}')$*. Then* $\mathsf{EVAL}(\mathbf{F}, \mathfrak{D})$ *can be solved in polynomial time.*

7. Proofs of Theorems 5.2 and 6.2. In this section, we prove Theorems 5.2 and 6.2. Let $\mathbf{A} = (A_{i.j})$ denote a connected, symmetric $m \times m$ algebraic matrix. (At this moment, we do not make any assumptions about whether $\mathbf{A}$ is bipartite.) We also let $\mathscr{A} = \{A_{i,j} : i, j \in [m]\}$ denote the finite set of algebraic numbers from the entries of $\mathbf{A}$. In the first step, we construct a new $m \times m$ matrix $\mathbf{B}$ from $\mathbf{A}$, which satisfies the following conditions:

1. $\mathbf{B}$ is also connected and symmetric (so that $\mathsf{EVAL}(\mathbf{B})$ is well-defined);
2. $\mathsf{EVAL}(\mathbf{B}) \equiv \mathsf{EVAL}(\mathbf{A})$; and
3. each entry of $\mathbf{B}$ is the product of a nonnegative integer and a root of unity.

We let $\mathbf{B}'$ be the nonnegative matrix such that $B'_{i,j} = |B_{i,j}|$. In the second step, we show that $\mathsf{EVAL}(\mathbf{B}') \leq \mathsf{EVAL}(\mathbf{B})$. Because $\mathbf{B}'$ is a connected, symmetric, and nonnegative (integer) matrix, we can apply the dichotomy of Bulatov and Grohe [4] (see Theorem 2.5) to $\mathbf{B}'$ and show that either $\mathsf{EVAL}(\mathbf{B}')$ is #P-hard or $\mathbf{B}$ is a (bipartite or nonbipartite, depending on $\mathbf{A}$) *purified* matrix. When $\mathsf{EVAL}(\mathbf{B}')$ is #P-hard, we have $\mathsf{EVAL}(\mathbf{B}') \leq \mathsf{EVAL}(\mathbf{B}) \equiv \mathsf{EVAL}(\mathbf{A})$ and thus $\mathsf{EVAL}(\mathbf{A})$ is #P-hard as well. This proves both Theorems 5.2 and 6.2.

7.1. Equivalence between $\mathsf{EVAL}(\mathbf{A})$ and $\mathsf{COUNT}(\mathbf{A})$. Before the construction of $\mathbf{B}$, we define a class of counting problems closely related to $\mathsf{EVAL}(\mathbf{A})$. It has been used in previous work [21] for establishing polynomial-time reductions between different EVAL problems.

Let $\mathbf{A} \in \mathbb{C}^{m \times m}$ be any fixed symmetric matrix with algebraic entries. The input of the problem $\mathsf{COUNT}(\mathbf{A})$ is a pair (G, x), where $G = (V, E)$ is an undirected graph and $x \in \mathbb{Q}(\mathscr{A})$. The output is

$$\#_{\mathbf{A}}(G,x) = \left|\left\{\text{assignment } \xi : V \to [m] \;\middle|\; \mathrm{wt}_{\mathbf{A}}(\xi) = x\right\}\right|,$$

a nonnegative integer. We prove the following lemma.

LEMMA 7.1. $\mathsf{EVAL}(\mathbf{A}) \equiv \mathsf{COUNT}(\mathbf{A})$.

Proof. To prove $\mathsf{EVAL}(\mathbf{A}) \leq \mathsf{COUNT}(\mathbf{A})$, recall that the matrix $\mathbf{A}$ is considered fixed with m being a constant. Let $G = (V, E)$ and $n = |E|$. We use X to denote the following set of complex numbers:

$$X = \left\{ \prod_{i,j\in[m]} A_{i,j}^{k_{i,j}} \;\middle|\; \text{integers } k_{i,j} \geq 0 \text{ and } \sum_{i,j\in[m]} k_{i,j} = n \right\}. \tag{7.1}$$

It is clear that $|X|$ is polynomial in n, being $\binom{n+m^2-1}{m^2-1}$ counting multiplicity, and X can be enumerated in polynomial time (in n). It follows from the expression in the definition of $\mathrm{wt}_{\mathbf{A}}(\xi)$ that for any $x \notin X$, $\#_{\mathbf{A}}(G,x) = 0$. This implies that

$$Z_{\mathbf{A}}(G) = \sum_{x\in X} x \cdot \#_{\mathbf{A}}(G,x)$$

for any undirected graph G and thus $\mathsf{EVAL}(\mathbf{A}) \leq \mathsf{COUNT}(\mathbf{A})$.

For the other direction, we construct for any $p \in [|X|]$ (recall that $|X|$ is polynomial in n) a new undirected graph $G^{[p]}$ from G by replacing every edge uv of G with p parallel edges between u and v. It is easy to check that any assignment ξ that has weight x over G has weight x^p over $G^{[p]}$. This gives us the following collection of equations: For every $p \in [|X|]$,

$$Z_{\mathbf{A}}(G^{[p]}) = \sum_{x\in X} x^p \cdot \#_{\mathbf{A}}(G,x).$$

Note that this is a Vandermonde system. Since we can query $\mathsf{EVAL}(\mathbf{A})$ for the values of $Z_{\mathbf{A}}(G^{[p]})$, we can solve it and get $\#_{\mathbf{A}}(G,x)$ for every nonzero $x \in X$. We can also derive $\#_{\mathbf{A}}(G,0)$, if $0 \in X$, using the fact that the $\#_{\mathbf{A}}(G,x)$'s sum to $m^{|V|}$. ∎

7.2. Step 1.1. We now construct the desired matrix $\mathbf{B}$ from $\mathbf{A}$. We need the following notion of a *generating set*.

DEFINITION 7.2. *Let $\mathscr{A} = \{a_1, \ldots, a_n\}$ be a set of n nonzero algebraic numbers for some $n \geq 1$. We say $\{g_1, \ldots, g_d\}$ for some $d \geq 0$ is a* generating set *of $\mathscr{A}$ if*

1. *every g_i is a nonzero algebraic number in $\mathbb{Q}(\mathscr{A})$, and*
2. *for every $a \in \mathscr{A}$, there exists a* unique *tuple $(k_1, \ldots, k_d) \in \mathbb{Z}^d$ such that*

$$\frac{a}{g_1^{k_1} \cdots g_d^{k_d}} \text{ is a root of unity.}$$

Clearly $d = 0$ iff the set $\mathscr{A}$ consists of roots of unity only. It can also be derived from the definition that $g_1^{k_1} \cdots g_d^{k_d}$ of any nonzero $(k_1, \ldots, k_d) \in \mathbb{Z}^d$ cannot be a root of unity. We prove the following lemma.

LEMMA 7.3. *Every set $\mathscr{A}$ of nonzero algebraic numbers has a generating set.*

Lemma 7.3 follows directly from Theorem 17.1. Actually the statement of Theorem 17.1 is stronger: A generating set $\{g_1, g_2, \ldots, g_d\}$ can be computed from $\mathscr{A}$ in polynomial time. More precisely, following the model of computation discussed in section 2.2, we let α be a primitive element of $\mathbb{Q}(\mathscr{A})$ so that $\mathbb{Q}(\mathscr{A}) = \mathbb{Q}(\alpha)$ and let $F(x)$ be a minimal polynomial of α. Then Theorem 17.1 shows that given the

standard representation of the a_j's, one can compute the standard representation of $g_1 \ldots, g_d \in \mathbb{Q}(\alpha)$ in polynomial time in the input size of the a_j's with $\{g_1, \ldots, g_d\}$ being a generating set of $\mathscr{A}$. Moreover, for each element $a \in \mathscr{A}$ one can also compute in polynomial time the unique tuple of integers $(k_1, \ldots, k_d)$ such that $a/(g_1^{k_1} \cdots g_d^{k_d})$ is a root of unity. In addition, if we are given an approximation $\hat{\alpha}$ of α that uniquely determines α as a root of $F(x)$, then we can use it to determine which root of unity it is in polynomial time. Note that in Lemma 7.3 we only need the existence of a generating set $\{g_1, \ldots, g_d\}$. But later in section 17, the polynomial-time computability of a generating set will be critical to the proof of Theorem 1.2, the polynomial-time decidability of the dichotomy criterion.

Now we return to the construction of $\mathbf{B}$. Letting $\mathscr{A}$ denote the set of nonzero entries of $\mathbf{A}$, by Lemma 7.3, $\mathscr{A}$ has a generating set $\mathscr{G} = \{g_1, \ldots, g_d\}$. The matrix $\mathbf{B} = (B_{i,j})$ is constructed as follows. Let $p_1 < \cdots < p_d$ denote the d smallest primes. For every $i, j \in [m]$, $B_{i,j} = 0$ if $A_{i,j} = 0$. Suppose $A_{i,j} \neq 0$. Since $\mathscr{G}$ is a generating set, we know there exists a unique tuple of integers $(k_1, \ldots, k_d)$ such that

$$\zeta_{i,j} = \frac{A_{i,j}}{g_1^{k_1} \cdots g_d^{k_d}}$$

is a root of unity. Then we set $B_{i,j} = p_1^{k_1} \cdots p_d^{k_d} \cdot \zeta_{i,j}$.

What we did in constructing $\mathbf{B}$ is just replace each g_i in $\mathscr{G}$ with a prime p_i. $B_{i,j}$ is well-defined by the uniqueness of $(k_1, \ldots, k_d) \in \mathbb{Z}^d$; conversely by taking the prime factorization of $|B_{i,j}|$ we can recover $(k_1, \ldots, k_d)$ uniquely and recover $A_{i,j}$ by

$$A_{i,j} = g_1^{k_1} \cdots g_d^{k_d} \cdot \frac{B_{i,j}}{p_1^{k_1} \cdots p_d^{k_d}}.$$

The next lemma shows that such a replacement does not affect the complexity.

LEMMA 7.4. *Let $\mathbf{A} \in \mathbb{C}^{m \times m}$ be a symmetric and connected matrix with algebraic entries and let $\mathbf{B}$ be the $m \times m$ matrix constructed above. Then* $\mathsf{EVAL}(\mathbf{A}) \equiv \mathsf{EVAL}(\mathbf{B})$.

Proof. By Lemma 7.1, it suffices to show that $\mathsf{COUNT}(\mathbf{A}) \equiv \mathsf{COUNT}(\mathbf{B})$. Here we only prove one of the two directions: $\mathsf{COUNT}(\mathbf{A}) \leq \mathsf{COUNT}(\mathbf{B})$. The other direction can be proved similarly.

Let (G, x) be an input pair of $\mathsf{COUNT}(\mathbf{A})$, where $G = (V, E)$ and $n = |E|$. We use X to denote the set of algebraic numbers defined earlier in (7.1). Recall that $|X|$ is polynomial in n since m is a constant and can be enumerated in polynomial time. Furthermore, if $x \notin X$, then $\#_{\mathbf{A}}(G, x)$ must be zero.

Suppose $x \in X$. Then we can find a particular sequence of nonnegative integers $(k^*_{i,j} : i, j \in [m])$ in polynomial time such that $\sum_{i,j} k^*_{i,j} = n$ and

$$x = \prod_{i,j \in [m]} A_{i,j}^{k^*_{i,j}}. \tag{7.2}$$

Note that $(k^*_{i,j})$ is in general *not unique* for the given x. Using $(k^*_{i,j})$, we define y by

$$y = \prod_{i,j \in [m]} B_{i,j}^{k^*_{i,j}}. \tag{7.3}$$

It is clear that $x = 0$ iff $y = 0$. This happens precisely when some $k^*_{i,j} > 0$ for some entry $A_{i,j} = 0$.

The reduction $\mathsf{COUNT}(\mathbf{A}) \le \mathsf{COUNT}(\mathbf{B})$ then follows from the following claim:

$$\#_{\mathbf{A}}(G,x) = \#_{\mathbf{B}}(G,y). \tag{7.4}$$

To prove this claim, it suffices to show that for any assignment $\xi\colon V \to [m]$, $\mathrm{wt}_{\mathbf{A}}(\xi) = x$ iff $\mathrm{wt}_{\mathbf{B}}(\xi) = y$. Here we only show that $\mathrm{wt}_{\mathbf{A}}(\xi) = x$ implies $\mathrm{wt}_{\mathbf{B}}(\xi) = y$. The other direction can be proved similarly.

Let $\xi\colon V \to [m]$ denote an assignment. For every $i, j \in [m]$, we use $k_{i,j}$ to denote the number of edges $uv \in E$ such that $(\xi(u), \xi(v)) = (i,j)$ or (j,i). Then

$$\mathrm{wt}_{\mathbf{A}}(\xi) = \prod_{i,j\in[m]} A_{i,j}^{k_{i,j}} \quad \text{and} \quad \mathrm{wt}_{\mathbf{B}}(\xi) = \prod_{i,j\in[m]} B_{i,j}^{k_{i,j}}. \tag{7.5}$$

For $x = 0$, we note that the weight $\mathrm{wt}_{\mathbf{A}}(\xi)$ is 0 iff for some zero entry $A_{i,j} = 0$ we have $k_{i,j} > 0$. By the construction of $\mathbf{B}$, $A_{i,j} = 0$ iff $B_{i,j} = 0$, so $\mathrm{wt}_{\mathbf{B}}(\xi) = 0$.

In the following, we assume both $x, y \neq 0$. We only consider assignments ξ such that its $k_{i,j} = 0$ for any $A_{i,j} = 0$ (equivalently $k_{i,j} = 0$ for any $B_{i,j} = 0$). Thus we may consider the products in (7.5) are over nonzero entries $A_{i,j}$ and $B_{i,j}$, respectively.

Now we use the generating set $\mathscr{G} = \{g_1, \ldots, g_d\}$ chosen for $\mathscr{A}$. There are integer exponents $e_{1,i,j}, e_{2,i,j}, \ldots, e_{d,i,j}$ and roots of unity $\zeta_{i,j}$ such that for all $A_{i,j} \neq 0$,

$$A_{i,j} = \prod_{\ell=1}^{d} g_\ell^{e_{\ell,i,j}} \cdot \zeta_{i,j} \quad \text{and} \quad B_{i,j} = \prod_{\ell=1}^{d} p_\ell^{e_{\ell,i,j}} \cdot \zeta_{i,j}.$$

The expression of $B_{i,j}$ here follows from the construction of $\mathbf{B}$. By (7.2) and (7.5),

$$\mathrm{wt}_{\mathbf{A}}(\xi) = x \implies \prod_{\ell=1}^{d} g_\ell^{\sum_{i,j}(k_{i,j}-k^*_{i,j})e_{\ell,i,j}} = \text{a root of unity}.$$

The sum in the exponent is over $i, j \in [m]$ where the corresponding $A_{i,j}$ is nonzero. This last equation is equivalent to (since $\mathscr{G}$ is a generating set)

$$\sum_{i,j} (k_{i,j} - k^*_{i,j}) \cdot e_{\ell,i,j} = 0 \quad \text{for all } \ell \in [d], \tag{7.6}$$

which in turn implies that

$$\prod_{i,j} \left(\zeta_{i,j}\right)^{k_{i,j}} = \prod_{i,j} \left(\zeta_{i,j}\right)^{k^*_{i,j}}. \tag{7.7}$$

It then follows from (7.3), (7.5), (7.6), and (7.7) that $\mathrm{wt}_{\mathbf{B}}(\xi) = y$. ∎

7.3. Step 1.2. The following lemma holds for any symmetric $\mathbf{B} \in \mathbb{C}^{m\times m}$.

LEMMA 7.5. *If $B'_{i,j} = |B_{i,j}|$ for all $i, j \in [m]$, then $\mathsf{EVAL}(\mathbf{B}') \le \mathsf{EVAL}(\mathbf{B})$.*

Proof. From Lemma 7.1, it suffices to show that $\mathsf{COUNT}(\mathbf{B}') \le \mathsf{COUNT}(\mathbf{B})$. Let (G,x) be an input of $\mathsf{COUNT}(\mathbf{B}')$. As $\mathbf{B}'$ is nonnegative, we have $\#_{\mathbf{B}'}(G,x) = 0$ if x is not real or $x < 0$. Now suppose $x \ge 0$, $G = (V,E)$, and $n = |E|$. We let

$$Y = \left\{ \prod_{i,j\in[m]} B_{i,j}^{k_{i,j}} \;\middle|\; \text{integers } k_{i,j} \ge 0 \text{ and } \sum_{i,j\in[m]} k_{i,j} = n \right\}.$$

We know that $|Y|$ is polynomial in n, and Y can be enumerated in polynomial time in n. Let Y_x denote the set of elements of Y whose complex norm is x.

The lemma then follows directly from the equation

$$\#_{\mathbf{B}'}(G,x) = \sum_{y \in Y_x} \#_{\mathbf{B}}(G,y),$$

because for every assignment $\xi\colon V \to [m]$, $\mathrm{wt}_{\mathbf{B}'}(\xi) = x$ iff $|\mathrm{wt}_{\mathbf{B}}(\xi)| = x$. This gives us a polynomial reduction since $Y_x \subseteq Y$, $|Y_x|$ is polynomially bounded in n, and Y_x can be enumerated in polynomial time. □

Finally we prove Theorems 5.2 and 6.2.

Proof of Theorem 5.2. Let $\mathbf{A} \in \mathbb{C}^{m\times m}$ be a symmetric, connected, and bipartite matrix. We construct matrices $\mathbf{B}$ and $\mathbf{B}'$ as above. Since we assumed $\mathbf{A}$ to be connected and bipartite, both matrices $\mathbf{B}$ and $\mathbf{B}'$ are connected and bipartite. Thus, we know there is a permutation Π of $[m]$ such that $\mathbf{B}_{\Pi,\Pi}$ is the bipartization of a $k \times (m-k)$ matrix $\mathbf{F}$ for some $k \in [m-1]$, and $\mathbf{B}'_{\Pi,\Pi}$ is the bipartization of $\mathbf{F}'$, where $F'_{i,j} = |F_{i,j}|$ for all $i \in [k]$ and $j \in [m-k]$. Since permuting the rows and columns of $\mathbf{B}$ does not affect the complexity of $\mathsf{EVAL}(\mathbf{B})$, we have

$$\mathsf{EVAL}(\mathbf{B}'_{\Pi,\Pi}) \le \mathsf{EVAL}(\mathbf{B}_{\Pi,\Pi}) \equiv \mathsf{EVAL}(\mathbf{B}) \equiv \mathsf{EVAL}(\mathbf{A}). \tag{7.8}$$

As $\mathbf{B}'_{\Pi,\Pi}$ is nonnegative, by Bulatov and Grohe we have the following cases:

1. If $\mathsf{EVAL}(\mathbf{B}'_{\Pi,\Pi})$ is #P-hard, then by (7.8), $\mathsf{EVAL}(\mathbf{A})$ is also #P-hard.
2. If $\mathsf{EVAL}(\mathbf{B}'_{\Pi,\Pi})$ is not #P-hard, then the rank of $\mathbf{F}'$ must be 1. (It cannot be 0 since $\mathbf{A}$ is assumed to be connected and bipartite.) Thus, there exist nonnegative rational numbers $\mu_1, \ldots, \mu_m$ such that $F'_{i,j} = \mu_i \mu_{j+k}$ for all $i \in [k]$ and $j \in [m-k]$. Moreover, $\mu_i \neq 0$ for all $i \in [m]$ since otherwise $\mathbf{B}'_{\Pi,\Pi}$ is not connected.

As every entry of $\mathbf{B}_{\Pi,\Pi}$ is the product of the corresponding entry of $\mathbf{B}'_{\Pi,\Pi}$ and some root of unity, $\mathbf{B}_{\Pi,\Pi}$ is a purified bipartite matrix. The theorem is proved. □

Proof of Theorem 6.2. Let $\mathbf{A} \in \mathbb{C}^{m\times m}$ be a symmetric, connected, and nonbipartite matrix. We construct $\mathbf{B}$ and $\mathbf{B}'$ as above. Since $\mathbf{A}$ is connected and non-bipartite, $\mathbf{B}$ and $\mathbf{B}'$ are connected and nonbipartite. Also, $\mathbf{B}'$ is nonnegative. Consider the following cases. If $\mathbf{B}'$ is #P-hard, then $\mathsf{EVAL}(\mathbf{B}') \le \mathsf{EVAL}(\mathbf{B}) \equiv \mathsf{EVAL}(\mathbf{A})$ implies that $\mathsf{EVAL}(\mathbf{A})$ must also be #P-hard. If $\mathbf{B}'$ is not #P-hard, then by Bulatov and Grohe, the rank of $\mathbf{B}$ is 1. (It cannot be 0 as we assumed $m > 1$, and $\mathbf{B}$ is connected.) Because $\mathbf{B}$ is symmetric, it is a purified nonbipartite matrix. The theorem follows. □

8. Proof of Theorem 5.3. We start the section by introducing a technique for establishing reductions between problems $\mathsf{EVAL}(\mathbf{A})$ and $\mathsf{EVAL}(\mathbf{C}, \mathfrak{D})$. It was inspired by the twin reduction lemma proved in [21].

8.1. Cyclotomic reduction and inverse cyclotomic reduction. Let $\mathbf{A}$ be an $m \times m$ symmetric (but not necessarily bipartite) complex matrix, and let $(\mathbf{C}, \mathfrak{D})$ be a pair that satisfies the following condition $(\mathcal{T})$:

$(\mathcal{T}_1)$ $\mathbf{C}$ is an $n \times n$ symmetric complex matrix.

$(\mathcal{T}_2)$ $\mathfrak{D} = (\mathbf{D}^{[0]}, \ldots, \mathbf{D}^{[N-1]})$ is a sequence of N $n \times n$ diagonal complex matrices for some $N \ge 1$.

$(\mathcal{T}_3)$ Every diagonal entry in $\mathbf{D}^{[0]}$ is a positive integer. Moreover, for each $a \in [n]$, there exist nonnegative integers $\alpha_{a,0}, \ldots, \alpha_{a,N-1}$ such that

$$D^{[0]}_a = \sum_{b=0}^{N-1} \alpha_{a,b} \quad \text{and} \quad D^{[r]}_a = \sum_{b=0}^{N-1} \alpha_{a,b} \cdot \omega_N^{br} \quad \text{for all } r \in [N-1].$$

In particular, we say that the tuple $(\alpha_{a,0},\dots,\alpha_{a,N-1})$ *generates* the ath entries of $\mathfrak{D}$.

We need the following definition.

DEFINITION 8.1. *Let $\mathscr{R}=\{R_{a,b}: a\in[n], b\in[0:N-1]\}$ be a partition of $[m]$ (note that any $R_{a,b}$ here may be empty) such that for every $a\in[n]$,*

$$\bigcup_{b=0}^{N-1} R_{a,b}\neq\emptyset.$$

We say $\mathbf{A}$ *can be* generated *by* $\mathbf{C}$ *using $\mathscr{R}$ if for all $i,j\in[m]$,*

$$A_{i,j}=C_{a,a'}\cdot\omega_N^{b+b'},\quad \text{where } i\in R_{a,b} \text{ and } j\in R_{a',b'}. \tag{8.1}$$

Given any pair $(\mathbf{C},\mathfrak{D})$ that satisfies $(\mathcal{T})$, we prove the following lemma.

LEMMA 8.2 (cyclotomic reduction lemma). *Assume that $(\mathbf{C},\mathfrak{D})$ satisfies $(\mathcal{T})$ with nonnegative integers $\alpha_{a,b}$. Let $\mathscr{R}=\{R_{a,b}\}$ be a partition of $[m]$ satisfying*

$$|R_{a,b}|=\alpha_{a,b}\quad \text{and}\quad m=\sum_{a=1}^{n}\sum_{b=0}^{N-1}\alpha_{a,b}\geq n,$$

and let $\mathbf{A}$ denote the matrix generated by $\mathbf{C}$ using $\mathscr{R}$. Then $\mathsf{EVAL}(\mathbf{A})\equiv\mathsf{EVAL}(\mathbf{C},\mathfrak{D})$.

Proof. It suffices to prove for any undirected graph $G=(V,E)$,

$$Z_{\mathbf{A}}(G)=\sum_{\xi:V\to[m]}\mathrm{wt}_{\mathbf{A}}(\xi)\quad\text{and}\quad Z_{\mathbf{C},\mathfrak{D}}(G)=\sum_{\eta:V\to[n]}\mathrm{wt}_{\mathbf{C},\mathfrak{D}}(\eta)$$

are exactly the same. To this end, we define a surjective map ρ from $\{\xi\}$, the set of all assignments from V to $[m]$, to $\{\eta\}$, the set of all assignments from V to $[n]$. Then we show that for every $\eta\colon V\to[n]$,

$$\mathrm{wt}_{\mathbf{C},\mathfrak{D}}(\eta)=\sum_{\xi:\rho(\xi)=\eta}\mathrm{wt}_{\mathbf{A}}(\xi). \tag{8.2}$$

We define $\rho(\xi)$ as follows. As $\mathscr{R}$ is a partition of $[m]$, for each $v\in V$ there exists a unique pair $(a(v),b(v))$ such that $\xi(v)\in R_{a(v),b(v)}$. Let $\eta(v)=a(v)$ for each v, and let $\rho(\xi)=\eta$. It is easy to check that ρ is surjective. To prove (8.2), we write $\mathrm{wt}_{\mathbf{A}}(\xi)$ as

$$\mathrm{wt}_{\mathbf{A}}(\xi)=\prod_{uv\in E}A_{\xi(u),\xi(v)}=\prod_{uv\in E}C_{\eta(u),\eta(v)}\times\omega_N^{\xi_2(u)+\xi_2(v)}.$$

It follows that

$$\begin{aligned}\sum_{\xi:\rho(\xi)=\eta}\mathrm{wt}_{\mathbf{A}}(\xi)&=\prod_{uv\in E}C_{\eta(u),\eta(v)}\times\sum_{\xi:\rho(\xi)=\eta}\prod_{v\in V}\omega_N^{\xi_2(v)\cdot\deg(v)}\\&=\prod_{uv\in E}C_{\eta(u),\eta(v)}\times\prod_{v\in V}\left(\sum_{b=0}^{N-1}\left|R_{\eta(v),b}\right|\cdot\omega_N^{b\cdot\deg(v)}\right)\\&=\prod_{uv\in E}C_{\eta(u),\eta(v)}\times\prod_{v\in V}D_{\eta(v)}^{[\deg(v)\bmod N]}\;=\;\mathrm{wt}_{\mathbf{C},\mathfrak{D}}(\eta),\end{aligned}$$

and the lemma follows. ∎

By combining Lemmas 8.2 and 7.5, as well as the dichotomy theorem of Bulatov and Grohe, we have the following handy corollary for dealing with $\mathsf{EVAL}(\mathbf{C}, \mathfrak{D})$.

COROLLARY 8.3 (inverse cyclotomic reduction lemma). *Let $(\mathbf{C}, \mathfrak{D})$ be a pair that satisfies condition $(\mathcal{T})$. If $\mathbf{C}$ has a 2×2 submatrix*

$$\begin{pmatrix} C_{i,k} & C_{i,\ell} \\ C_{j,k} & C_{j,\ell} \end{pmatrix}$$

such that all four entries are nonzero and $|C_{i,k}C_{j,\ell}| \neq |C_{i,\ell}C_{j,k}|$, then $\mathsf{EVAL}(\mathbf{C}, \mathfrak{D})$ is #P-hard.

Proof. By the cyclotomic reduction lemma, there is a symmetric $m \times m$ matrix $\mathbf{A}$ for some positive integer m and a partition $\mathscr{R}$ of $[m]$, where

$$\mathscr{R} = \left\{ R_{a,b} \;\middle|\; a \in [n], b \in [0 : N-1] \right\} \quad \text{and} \quad \bigcup_{b=0}^{N-1} R_{a,b} \neq \emptyset \quad \text{for all } a \in [n], \tag{8.3}$$

such that $\mathsf{EVAL}(\mathbf{A}) \equiv \mathsf{EVAL}(\mathbf{C}, \mathfrak{D})$. Moreover, $\mathbf{A}$ and $\mathbf{C}$ satisfy (8.1).

Now suppose there exist $i \neq j, k \neq \ell \in [n]$ such that $C_{i,k}, C_{i,\ell}, C_{j,k}$, and $C_{j,\ell}$ are nonzero and $|C_{i,k}C_{j,\ell}| \neq |C_{i,\ell}C_{j,k}|$. We arbitrarily pick an i' from $\cup_b R_{i,b}$ (known to be nonempty), a j' from $\cup_b R_{j,b}$, a k' from $\cup_b R_{k,b}$, and an ℓ' from $\cup_b R_{\ell,b}$. Then from (8.1), we have $|A_{i',k'}| = |C_{i,k}|$, $|A_{i',\ell'}| = |C_{i,\ell}|$, $|A_{j',k'}| = |C_{j,k}|$, $|A_{j',\ell'}| = |C_{j,\ell}|$, and

$$|A_{i',k'}A_{j',\ell'}| \neq |A_{i',\ell'}A_{j',k'}|.$$

Let $\mathbf{A}' = (|A_{i,j}|)$ for all $i, j \in [m]$. Then $\mathbf{A}'$ has a 2×2 submatrix of rank 2 and all its four entries are nonzero. By the dichotomy of Bulatov and Grohe (Corollary 2.6), $\mathsf{EVAL}(\mathbf{A}')$ is #P-hard. It follows that $\mathsf{EVAL}(\mathbf{C}, \mathfrak{D})$ is #P-hard, since $\mathsf{EVAL}(\mathbf{C}, \mathfrak{D}) \equiv \mathsf{EVAL}(\mathbf{A})$ and by Lemma 7.5, $\mathsf{EVAL}(\mathbf{A}') \leq \mathsf{EVAL}(\mathbf{A})$. ∎

Combining Lemma 8.2, (8.2), and the first pinning lemma (Lemma 4.1), we get the following.

COROLLARY 8.4 (third pinning lemma). *Let $(\mathbf{C}, \mathfrak{D})$ be a pair that satisfies $(\mathcal{T})$. Then we have $\mathsf{EVALP}(\mathbf{C}, \mathfrak{D}) \equiv \mathsf{EVAL}(\mathbf{C}, \mathfrak{D})$. In particular, the problem of computing $Z^{\rightarrow}_{\mathbf{C},\mathfrak{D}}$ (or $Z^{\leftarrow}_{\mathbf{C},\mathfrak{D}}$) is polynomial-time reducible to $\mathsf{EVAL}(\mathbf{C}, \mathfrak{D})$.*

Proof. It suffices to show that $\mathsf{EVALP}(\mathbf{C}, \mathfrak{D}) \leq \mathsf{EVAL}(\mathbf{C}, \mathfrak{D})$. By the cyclotomic reduction lemma, there exist a symmetric $m \times m$ matrix $\mathbf{A}$ for some $m \geq 1$ and a partition $\mathscr{R}$ of $[m]$ such that $\mathscr{R}$ satisfies (8.3) and $\mathsf{EVAL}(\mathbf{A}) \equiv \mathsf{EVAL}(\mathbf{C}, \mathfrak{D})$. $\mathbf{A}, \mathbf{C}$, and $\mathscr{R}$ also satisfy (8.1). By the first pinning lemma, we have $\mathsf{EVALP}(\mathbf{A}) \equiv \mathsf{EVAL}(\mathbf{A}) \equiv \mathsf{EVAL}(\mathbf{C}, \mathfrak{D})$. So we only need to reduce $\mathsf{EVALP}(\mathbf{C}, \mathfrak{D})$ to $\mathsf{EVALP}(\mathbf{A})$.

Now let (G, w, i) be an input of $\mathsf{EVALP}(\mathbf{C}, \mathfrak{D})$, where G is an undirected graph, w is a vertex in G, and $i \in [n]$. By (8.2), we have

$$Z_{\mathbf{C},\mathfrak{D}}(G, w, i) = \sum_{\eta : \eta(w) = i} \mathrm{wt}_{\mathbf{C},\mathfrak{D}}(\eta) = \sum_{\xi : \xi_1(w) = i} \mathrm{wt}_{\mathbf{A}}(\xi) = \sum_{j \in \cup_b R_{i,b}} Z_{\mathbf{A}}(G, w, j).$$

This gives us a polynomial-time reduction from $\mathsf{EVALP}(\mathbf{C}, \mathfrak{D})$ to $\mathsf{EVALP}(\mathbf{A})$. ∎

Note that compared to the second pinning lemma, the third pinning lemma does not require $\mathbf{C}$ to be the bipartization of a unitary matrix. It only requires $(\mathcal{T})$.

8.2. Step 2.1. Let $\mathbf{A}$ be a purified bipartite matrix. After collecting its entries of equal norm in decreasing order by permuting its rows and columns, there exist a positive integer N and four sequences $\boldsymbol{\mu}$, $\boldsymbol{\nu}$, $\mathbf{m}$, and $\mathbf{n}$ such that $(\mathbf{A}, (N, \boldsymbol{\mu}, \boldsymbol{\nu}, \mathbf{m}, \mathbf{n}))$ satisfies the following condition:

($\mathcal{S}_1$) $\mathbf{A}$ is the bipartization of an $m \times n$ matrix $\mathbf{B}$, so $\mathbf{A}$ is $(m+n) \times (m+n)$. $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_s)$ and $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_t)$ are two strictly decreasing sequences of positive rational numbers where $s \geq 1$ and $t \geq 1$. $\mathbf{m} = (m_1, \ldots, m_s)$ and $\mathbf{n} = (n_1, \ldots, n_t)$ are two sequences of positive integers such that $m = \sum m_i$ and $n = \sum n_i$. The rows of $\mathbf{B}$ are indexed by $\mathbf{x} = (x_1, x_2)$, where $x_1 \in [s]$ and $x_2 \in [m_{x_1}]$; the columns of $\mathbf{B}$ are indexed by $\mathbf{y} = (y_1, y_2)$, where $y_1 \in [t]$ and $y_2 \in [n_{y_1}]$. We have, for all $\mathbf{x}, \mathbf{y}$,

$$B_{\mathbf{x},\mathbf{y}} = B_{(x_1,x_2),(y_1,y_2)} = \mu_{x_1}\nu_{y_1}S_{\mathbf{x},\mathbf{y}},$$

where $\mathbf{S} = \{S_{\mathbf{x},\mathbf{y}}\}$ is an $m \times n$ matrix in which every entry is a power of ω_N:

$$\mathbf{B} = \begin{pmatrix} \mu_1\mathbf{I}_{m_1} & & & \\ & \mu_2\mathbf{I}_{m_2} & & \\ & & \ddots & \\ & & & \mu_s\mathbf{I}_{m_s} \end{pmatrix} \begin{pmatrix} \mathbf{S}_{(1,*),(1,*)} & \mathbf{S}_{(1,*),(2,*)} & \cdots & \mathbf{S}_{(1,*),(t,*)} \\ \mathbf{S}_{(2,*),(1,*)} & \mathbf{S}_{(2,*),(2,*)} & \cdots & \mathbf{S}_{(2,*),(t,*)} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{S}_{(s,*),(1,*)} & \mathbf{S}_{(s,*),(2,*)} & \cdots & \mathbf{S}_{(s,*),(t,*)} \end{pmatrix} \begin{pmatrix} \nu_1\mathbf{I}_{n_1} & & & \\ & \nu_2\mathbf{I}_{n_2} & & \\ & & \ddots & \\ & & & \nu_t\mathbf{I}_{n_t} \end{pmatrix},$$

where $\mathbf{I}_k$ denotes the $k \times k$ identity matrix.

We let

$$I = \bigcup_{i \in [s]} \big\{(i,j) : j \in [m_i]\big\} \quad \text{and} \quad J = \bigcup_{i \in [t]} \big\{(i,j) : j \in [n_i]\big\},$$

respectively. We use $\{0\} \times I$ to index the first m rows (or columns) of $\mathbf{A}$ and $\{1\} \times J$ to index the last n rows (or columns) of $\mathbf{A}$. Given $\mathbf{x} \in I$ and $j \in [t]$, we let

$$\mathbf{S}_{\mathbf{x},(j,*)} = \big(S_{\mathbf{x},(j,1)}, \ldots, S_{\mathbf{x},(j,n_j)}\big) \in \mathbb{C}^{n_j}$$

denote the jth block of the $\mathbf{x}$th row vector of $\mathbf{S}$. Similarly, given $\mathbf{y} \in J$ and $i \in [s]$,

$$\mathbf{S}_{(i,*),\mathbf{y}} = \big(S_{(i,1),\mathbf{y}}, \ldots, S_{(i,m_i),\mathbf{y}}\big) \in \mathbb{C}^{m_i}$$

denotes the ith block of the $\mathbf{y}$th column vector of $\mathbf{S}$.

LEMMA 8.5. *Suppose* $(\mathbf{A}, (N, \boldsymbol{\mu}, \boldsymbol{\nu}, \mathbf{m}, \mathbf{n}))$ *satisfies* $(\mathcal{S}_1)$. *Then either* $\mathsf{EVAL}(\mathbf{A})$ *is #P-hard, or* $(\mathbf{A}, (N, \boldsymbol{\mu}, \boldsymbol{\nu}, \mathbf{m}, \mathbf{n}))$ *satisfies the following two conditions:*

($\mathcal{S}_2$) *For all* $\mathbf{x}, \mathbf{x}' \in I$, *either there exists an integer* k *such that* $\mathbf{S}_{\mathbf{x},*} = \omega_N^k \cdot \mathbf{S}_{\mathbf{x}',*}$ *or for every* $j \in [t]$, $\langle \mathbf{S}_{\mathbf{x},(j,*)}, \mathbf{S}_{\mathbf{x}',(j,*)} \rangle = 0$.

($\mathcal{S}_3$) *For all* $\mathbf{y}, \mathbf{y}' \in J$, *either there exists an integer* k *such that* $\mathbf{S}_{*,\mathbf{y}} = \omega_N^k \cdot \mathbf{S}_{*,\mathbf{y}'}$ *or for every* $i \in [s]$, $\langle \mathbf{S}_{(i,*),\mathbf{y}}, \mathbf{S}_{(i,*),\mathbf{y}'} \rangle = 0$.

Proof. Assume $\mathsf{EVAL}(\mathbf{A})$ is not #P-hard. We prove ($\mathcal{S}_2$) here. ($\mathcal{S}_3$) is similar.

Let $G = (V, E)$ be an undirected graph. We construct a new graph $G^{[p]}$ for each $p \geq 1$ by replacing every edge uv in E with a gadget shown in Figure 8.1. Formally we define graph $G^{[p]} = (V^{[p]}, E^{[p]})$ as

$$V^{[p]} = V \cup \big\{a_e, b_e : e \in E\big\},$$

and $E^{[p]}$ contains the following edges: For each $e = uv \in E$, add

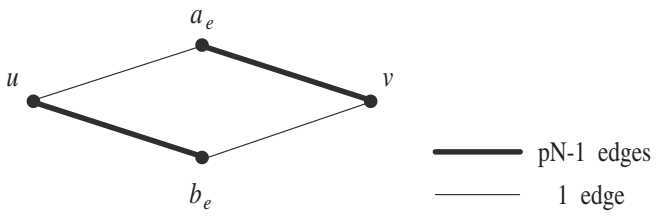


FIG. 8.1. *Gadget for constructing graph* $G^{[p]}$, $p \geq 1$.

1. one edge (u, a_e) and (b_e, v) and
2. $(pN-1)$ parallel edges (a_e, v) and (u, b_e).

The construction of $G^{[p]}$ gives us an $(m+n)\times(m+n)$ matrix $\mathbf{A}^{[p]}$ such that

$$Z_{\mathbf{A}^{[p]}}(G) = Z_{\mathbf{A}}(G^{[p]}) \quad \text{for all undirected graphs } G.$$

Thus, we have $\mathsf{EVAL}(\mathbf{A}^{[p]}) \le \mathsf{EVAL}(\mathbf{A})$, and $\mathsf{EVAL}(\mathbf{A}^{[p]})$ is also not #P-hard.

The entries of $\mathbf{A}^{[p]}$ are as follows. First,

$$A^{[p]}_{(0,\mathbf{u}),(1,\mathbf{v})} = A^{[p]}_{(1,\mathbf{v}),(0,\mathbf{u})} = 0 \quad \text{for all } \mathbf{u}\in I \text{ and } \mathbf{v}\in J.$$

So $\mathbf{A}^{[p]}$ is a block diagonal matrix with two blocks of $m\times m$ and $n\times n$, respectively. The entries in the upper-left $m\times m$ block are

$$\begin{aligned} A^{[p]}_{(0,\mathbf{u}),(0,\mathbf{v})} &= \left(\sum_{\mathbf{a}\in J} A_{(0,\mathbf{u}),(1,\mathbf{a})}(A_{(0,\mathbf{v}),(1,\mathbf{a})})^{pN-1}\right)\left(\sum_{\mathbf{b}\in J}(A_{(0,\mathbf{u}),(1,\mathbf{b})})^{pN-1}A_{(0,\mathbf{v}),(1,\mathbf{b})}\right)\\ &= \left(\sum_{\mathbf{a}\in J} B_{\mathbf{u},\mathbf{a}}(B_{\mathbf{v},\mathbf{a}})^{pN-1}\right)\left(\sum_{\mathbf{b}\in J}(B_{\mathbf{u},\mathbf{b}})^{pN-1}B_{\mathbf{v},\mathbf{b}}\right) \end{aligned}$$

for all $\mathbf{u},\mathbf{v}\in I$. The first factor of the last expression is

$$\sum_{\mathbf{a}\in J}\mu_{u_1}\nu_{a_1}S_{\mathbf{u},\mathbf{a}}(\mu_{v_1}\nu_{a_1})^{pN-1}\overline{S_{\mathbf{v},\mathbf{a}}} = \mu_{u_1}\mu_{v_1}^{pN-1}\sum_{i\in[t]}\nu_i^{pN}\langle \mathbf{S}_{\mathbf{u},(i,*)}, \mathbf{S}_{\mathbf{v},(i,*)}\rangle.$$

Similarly, we have for the second factor

$$\sum_{\mathbf{b}\in J}(B_{\mathbf{u},\mathbf{b}})^{pN-1}B_{\mathbf{v},\mathbf{b}} = \mu_{u_1}^{pN-1}\mu_{v_1}\sum_{i\in[t]}\nu_i^{pN}\overline{\langle \mathbf{S}_{\mathbf{u},(i,*)}, \mathbf{S}_{\mathbf{v},(i,*)}\rangle}.$$

As a result, we have

$$A^{[p]}_{(0,\mathbf{u}),(0,\mathbf{v})} = (\mu_{u_1}\mu_{v_1})^{pN}\left|\sum_{i\in[t]}\nu_i^{pN}\langle \mathbf{S}_{\mathbf{u},(i,*)}, \mathbf{S}_{\mathbf{v},(i,*)}\rangle\right|^2.$$

It is clear that the upper-left $m\times m$ block of $\mathbf{A}^{[p]}$ is nonnegative. This holds for its lower-right $n\times n$ block as well, so $\mathbf{A}^{[p]}$ is a nonnegative matrix.

Now let $\mathbf{u}\neq\mathbf{v}$ be two indices in I (if $|I|=1$, $(\mathcal{S}_2)$ is trivially true); then we have

$$A^{[p]}_{(0,\mathbf{u}),(0,\mathbf{u})}A^{[p]}_{(0,\mathbf{v}),(0,\mathbf{v})} = (\mu_{u_1}\mu_{v_1})^{2pN}\left(\sum_{i\in[t]} n_i\cdot\nu_i^{pN}\right)^4,$$

which is positive, and

$$A^{[p]}_{(0,\mathbf{u}),(0,\mathbf{v})}A^{[p]}_{(0,\mathbf{v}),(0,\mathbf{u})} = (\mu_{u_1}\mu_{v_1})^{2pN}\left|\sum_{i\in[t]}\nu_i^{pN}\langle \mathbf{S}_{\mathbf{u},(i,*)}, \mathbf{S}_{\mathbf{v},(i,*)}\rangle\right|^4.$$

Since $\mathsf{EVAL}(\mathbf{A}^{[p]})$ is not #P-hard, by Bulatov and Grohe (Corollary 2.6),

$$\left|\sum_{i\in[t]}\nu_i^{pN}\langle \mathbf{S}_{\mathbf{u},(i,*)}, \mathbf{S}_{\mathbf{v},(i,*)}\rangle\right| \in \left\{0, \sum_{i\in[t]} n_i\cdot\nu_i^{pN}\right\}. \tag{8.4}$$

On the other hand, the following inequality always holds: For any $p \geq 1$,

$$\left| \sum_{i \in [t]} \nu_i^{pN} \cdot \langle \mathbf{S}_{\mathbf{u},(i,*)}, \mathbf{S}_{\mathbf{v},(i,*)} \rangle \right| \leq \sum_{i \in [t]} n_i \cdot \nu_i^{pN}. \tag{8.5}$$

For the equality of (8.5) to hold, $\mathbf{S}$ must satisfy $|\langle \mathbf{S}_{\mathbf{u},(i,*)}, \mathbf{S}_{\mathbf{v},(i,*)} \rangle| = n_i$ for all $i \in [t]$ and thus $\mathbf{S}_{\mathbf{u},(i,*)} = (\omega_N)^{k_i} \cdot \mathbf{S}_{\mathbf{v},(i,*)}$ for some $k_i \in [0 : N-1]$. Furthermore, these k_i's must be the same. As a result, $\mathbf{S}_{\mathbf{u},*}$ and $\mathbf{S}_{\mathbf{v},*}$ are linearly dependent, which contradicts our assumption. It then follows from (8.4) that

$$\sum_{i \in [t]} \nu_i^{pN} \langle \mathbf{S}_{\mathbf{u},(i,*)}, \mathbf{S}_{\mathbf{v},(i,*)} \rangle = 0 \quad \text{for all } p \geq 1.$$

As $\nu_1, \ldots, \nu_t$ is strictly decreasing, these equations form a Vandermonde system. It follows that $\langle \mathbf{S}_{\mathbf{u},(i,*)}, \mathbf{S}_{\mathbf{v},(i,*)} \rangle = 0$ for all $i \in [t]$. This finishes the proof of $(\mathcal{S}_2)$. ∎

We have the following corollary.

COROLLARY 8.6. *For all $i \in [s]$ and $j \in [t]$, the rank of the (i,j)th block matrix $\mathbf{S}_{(i,*),(j,*)}$ of $\mathbf{S}$ has the same rank as $\mathbf{S}$.*

Proof. Without loss of generality, we prove $\operatorname{rank}(\mathbf{S}_{(1,*),(1,*)}) = \operatorname{rank}(\mathbf{S})$.

First, we use Lemma 8.5 to show that

$$\operatorname{rank} \begin{pmatrix} \mathbf{S}_{(1,*),(1,*)} \\ \vdots \\ \mathbf{S}_{(s,*),(1,*)} \end{pmatrix} = \operatorname{rank}(\mathbf{S}).$$

To see this, we take any $h = \operatorname{rank}(\mathbf{S})$ rows of $\mathbf{S}$ which are linearly independent. Since any two of them $\mathbf{S}_{\mathbf{x},(*,*)}$ and $\mathbf{S}_{\mathbf{y},(*,*)}$ are linearly independent, by condition $(\mathcal{S}_2)$, the two subvectors $\mathbf{S}_{\mathbf{x},(1,*)}$ and $\mathbf{S}_{\mathbf{y},(1,*)}$ are orthogonal. Therefore, the corresponding h rows of the matrix on the left-hand side are pairwise orthogonal, and the left-hand side is at least h. Of course it cannot be larger than h, so it is equal to h.

By using condition $(\mathcal{S}_3)$, we can similarly show that

$$\operatorname{rank}(\mathbf{S}_{(1,*),(1,*)}) = \operatorname{rank} \begin{pmatrix} \mathbf{S}_{(1,*),(1,*)} \\ \vdots \\ \mathbf{S}_{(s,*),(1,*)} \end{pmatrix}.$$

As a result, we have $\operatorname{rank}(\mathbf{S}_{(1,*),(1,*)}) = \operatorname{rank}(\mathbf{S})$. ∎

Now suppose $h = \operatorname{rank}(\mathbf{S})$. Then by Corollary 8.6, there must exist indices $1 \leq i_1 < \cdots < i_h \leq m_1$ and $1 \leq j_1 < \cdots < j_h \leq n_1$ such that the $\{(1,i_1), \ldots, (1,i_h)\} \times \{(1,j_1), \ldots, (1,j_h)\}$ submatrix of $\mathbf{S}$ has full rank h. Without loss of generality (if this is not true, we can apply an appropriate permutation Π to the rows and columns of $\mathbf{A}$ so that the new $\mathbf{S}$ has this property) we assume $i_k = k$ and $j_k = k$ for all $k \in [h]$. We use $\mathbf{H}$ to denote this $h \times h$ matrix: $H_{i,j} = S_{(1,i),(1,j)}$.

By Corollary 8.6 and Lemma 8.5, for every index $\mathbf{x} \in I$, there exists a unique pair of integers $j \in [h]$ and $k \in [0 : N-1]$ such that

$$\mathbf{S}_{\mathbf{x},*} = \omega_N^k \cdot \mathbf{S}_{(1,j),*}. \tag{8.6}$$

This gives us a partition of index set $\{0\} \times I$:

$$\mathscr{R}_0 = \big\{ R_{(0,i,j),k} : i \in [s], j \in [h], k \in [0 : N-1] \big\}.$$

For every $\mathbf{x} \in I$, $(0, \mathbf{x}) \in R_{(0,i,j),k}$ if $i = x_1$ and $\mathbf{x}, j, k$ satisfy (8.6). By Corollary 8.6,

$$\bigcup_{k \in [0:N-1]} R_{(0,i,j),k} \neq \emptyset \quad \text{for all } i \in [s] \text{ and } j \in [h].$$

Similarly, for every index $\mathbf{y} \in J$ there exists a unique pair of integers $j \in [h]$ and $k \in [0 : N-1]$ such that

$$\mathbf{S}_{*,\mathbf{y}} = \omega_N^k \cdot \mathbf{S}_{*,(1,j)}, \tag{8.7}$$

and we partition $\{1\} \times J$ into

$$\mathscr{R}_1 = \big\{ R_{(1,i,j),k} : i \in [t], j \in [h], k \in [0 : N-1] \big\}.$$

For every $\mathbf{y} \in J$, $(1, \mathbf{y}) \in R_{(1,i,j),k}$ if $i = y_1$ and $\mathbf{y}, j, k$ satisfy (8.7). By Corollary 8.6,

$$\bigcup_{k \in [0:N-1]} R_{(1,i,j),k} \neq \emptyset \quad \text{for all } i \in [t] \text{ and } j \in [h].$$

Now we define $(\mathbf{C}, \mathfrak{D})$ and use the cyclotomic reduction lemma (Lemma 8.2) to show that $\mathsf{EVAL}(\mathbf{C}, \mathfrak{D}) \equiv \mathsf{EVAL}(\mathbf{A})$. First, $\mathbf{C}$ is an $(s+t)h \times (s+t)h$ matrix which is the bipartization of an $sh \times th$ matrix $\mathbf{F}$. We use the set $I' \equiv [s] \times [h]$ to index the rows of $\mathbf{F}$ and $J' \equiv [t] \times [h]$ to index the columns of $\mathbf{F}$. We have

$$F_{\mathbf{x},\mathbf{y}} = \mu_{x_1} \nu_{y_1} H_{x_2,y_2} = \mu_{x_1} \nu_{y_1} S_{(1,x_2),(1,y_2)} \quad \text{for all } \mathbf{x} \in I', \mathbf{y} \in J',$$

or equivalently,

$$\mathbf{F} = \begin{pmatrix} \mu_1 \mathbf{I} & & & \\ & \mu_2 \mathbf{I} & & \\ & & \ddots & \\ & & & \mu_s \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{H} & \mathbf{H} & \dots & \mathbf{H} \\ \mathbf{H} & \mathbf{H} & \dots & \mathbf{H} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{H} & \mathbf{H} & \dots & \mathbf{H} \end{pmatrix} \begin{pmatrix} \nu_1 \mathbf{I} & & & \\ & \nu_2 \mathbf{I} & & \\ & & \ddots & \\ & & & \nu_t \mathbf{I} \end{pmatrix},$$

where $\mathbf{I}$ is the $h \times h$ identity matrix. We use $(\{0\} \times I') \cup (\{1\} \times J')$ to index the rows and columns of $\mathbf{C}$.

Second, $\mathfrak{D} = (\mathbf{D}^{[0]}, \dots, \mathbf{D}^{[N-1]})$ is a sequence of N diagonal matrices of the same size as $\mathbf{C}$. We use $\{0\} \times I'$ to index the first sh entries and $\{1\} \times J'$ to index the last th entries. The $(0, \mathbf{x})$th entries of $\mathfrak{D}$ are generated by $(|R_{(0,x_1,x_2),0}|, \dots, |R_{(0,x_1,x_2),N-1}|)$, and the $(1, \mathbf{y})$th entries of $\mathfrak{D}$ are generated by $(|R_{(1,y_1,y_2),0}|, \dots, |R_{(1,y_1,y_2),N-1}|)$:

$$D^{[r]}_{(0,\mathbf{x})} = \sum_{k=0}^{N-1} \big|R_{(0,x_1,x_2),k}\big| \cdot \omega_N^{kr} \quad \text{and} \quad D^{[r]}_{(1,\mathbf{y})} = \sum_{k=0}^{N-1} \big|R_{(1,y_1,y_2),k}\big| \cdot \omega_N^{kr}$$

for all $r \in [0 : N-1]$, $\mathbf{x} = (x_1, x_2) \in I'$, and $\mathbf{y} = (y_1, y_2) \in J'$.

This finishes the construction of $(\mathbf{C}, \mathfrak{D})$, and we prove the following lemma.

LEMMA 8.7. $\mathsf{EVAL}(\mathbf{A}) \equiv \mathsf{EVAL}(\mathbf{C}, \mathfrak{D})$.

Proof. First we show that $\mathbf{A}$ can be generated from $\mathbf{C}$ using $\mathscr{R}_0 \cup \mathscr{R}_1$.

Let $\mathbf{x}, \mathbf{x}' \in I$, $(0, \mathbf{x}) \in R_{(0,x_1,j),k}$, and $(0, \mathbf{x}') \in R_{(0,x_1',j'),k'}$. Then we have

$$A_{(0,\mathbf{x}),(0,\mathbf{x}')} = C_{(0,x_1,j),(0,x_1',j')} = 0,$$

since $\mathbf{A}$ and $\mathbf{C}$ are the bipartizations of $\mathbf{B}$ and $\mathbf{F}$, respectively. Therefore,

$$A_{(0,\mathbf{x}),(0,\mathbf{x}')} = C_{(0,x_1,j),(0,x_1',j')} \cdot \omega_N^{k+k'}$$

holds trivially. Clearly, this also holds for the lower-right $n \times n$ block of $\mathbf{A}$.

Let $\mathbf{x} \in I$, $(0,\mathbf{x}) \in R_{(0,x_1,j),k}$, $\mathbf{y} \in J$, and $(1,\mathbf{y}) \in R_{(1,y_1,j'),k'}$ for some j, k, j', k'. By (8.6) and (8.7), we have

$$\begin{aligned} A_{(0,\mathbf{x}),(1,\mathbf{y})} &= \mu_{x_1}\nu_{y_1} S_{\mathbf{x},\mathbf{y}} = \mu_{x_1}\nu_{y_1} S_{(1,j),\mathbf{y}} \cdot \omega_N^k \\ &= \mu_{x_1}\nu_{y_1} S_{(1,j),(1,j')} \cdot \omega_N^{k+k'} = C_{(0,x_1,j),(1,y_1,j')} \cdot \omega_N^{k+k'}. \end{aligned}$$

A similar equation also holds for the lower-left block. Thus, $\mathbf{A}$ can be generated from $\mathbf{C}$ using $\mathscr{R}_0 \cup \mathscr{R}_1$. Moreover, the construction of $\mathfrak{D}$ implies that $\mathfrak{D}$ can be generated from the partition $\mathscr{R}_0 \cup \mathscr{R}_1$. The lemma then follows directly from the cyclotomic reduction lemma. ∎

Before moving forward to the next step, we summarize our progress so far. We showed that $\mathsf{EVAL}(\mathbf{A})$ is either #P-hard or equivalent to $\mathsf{EVAL}(\mathbf{C},\mathfrak{D})$, where the pair $(\mathbf{C},\mathfrak{D})$ satisfies the following conditions (*Shape*$_1$)–(*Shape*$_3$):

(*Shape*$_1$) $\mathbf{C} \in \mathbb{C}^{m\times m}$ (note that the m here is different from the m used at the beginning of Step 2.1) is the bipartization of an $sh \times th$ matrix $\mathbf{F}$ (so $m = (s+t)h$). $\mathbf{F}$ is an $s \times t$ block matrix. We use $I = [s] \times [h]$ and $J = [t] \times [h]$ to index the rows and columns of $\mathbf{F}$, respectively.

(*Shape*$_2$) There are two strictly decreasing sequences $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_s)$ and $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_t)$ of positive rational numbers. There is also an $h \times h$ full-rank matrix $\mathbf{H}$ whose entries are all powers of ω_N for some positive integer N. Entries of $\mathbf{F}$ can be expressed using $\boldsymbol{\mu}, \boldsymbol{\nu}$, and $\mathbf{H}$ explicitly as follows:

$$F_{\mathbf{x},\mathbf{y}} = \mu_{x_1}\nu_{y_1} H_{x_2,y_2} \quad \text{for all } \mathbf{x} \in I \text{ and } \mathbf{y} \in J.$$

(*Shape*$_3$) $\mathfrak{D} = (\mathbf{D}^{[0]}, \ldots, \mathbf{D}^{[N-1]})$ is a sequence of $m \times m$ diagonal matrices. We use $(\{0\} \times I) \cup (\{1\} \times J)$ to index the rows and columns of matrices $\mathbf{C}$ and $\mathbf{D}^{[r]}$. $\mathfrak{D}$ satisfies $(\mathcal{T}_3)$, so for all $r \in [N-1]$, $\mathbf{x} \in [s] \times [h]$, and $\mathbf{y} \in [t] \times [h]$,

$$D^{[r]}_{(0,\mathbf{x})} = \overline{D^{[N-r]}_{(0,\mathbf{x})}} \quad \text{and} \quad D^{[r]}_{(1,\mathbf{y})} = \overline{D^{[N-r]}_{(1,\mathbf{y})}}.$$

8.3. Step 2.2. In Step 2.2, we prove the following lemma.

LEMMA 8.8. *Either* $\mathsf{EVAL}(\mathbf{C},\mathfrak{D})$ *is #P-hard or* $\mathbf{H}$ *and* $\mathbf{D}^{[0]}$ *satisfy the following two conditions:*

(*Shape*$_4$) $(1/\sqrt{h}) \cdot \mathbf{H}$ *is a unitary matrix, i.e.,*

$$\langle \mathbf{H}_{i,*}, \mathbf{H}_{j,*} \rangle = \langle \mathbf{H}_{*,i}, \mathbf{H}_{*,j} \rangle = 0 \quad \textit{for all } i \neq j \in [h].$$

(*Shape*$_5$) $\mathbf{D}^{[0]}$ *satisfies, for all* $\mathbf{x} \in I$ *and for all* $\mathbf{y} \in J$,

$$D^{[0]}_{(0,\mathbf{x})} = D^{[0]}_{(0,(x_1,1))} \quad \textit{and} \quad D^{[0]}_{(1,\mathbf{y})} = D^{[0]}_{(1,(y_1,1))}.$$

Proof. We rearrange the entries of $\mathbf{D}^{[0]}$ indexed by $\{1\} \times J$ into a $t \times h$ matrix

$$X_{i,j} = D^{[0]}_{(1,(i,j))} \quad \text{for all } i \in [t] \text{ and } j \in [h]$$

and rearrange its entries indexed by $\{0\} \times I$ into an $s \times h$ matrix

$$Y_{i,j} = D^{[0]}_{(0,(i,j))} \quad \text{for all } i \in [s] \text{ and } j \in [h].$$

Note that by condition $(\mathcal{T}_3)$, all entries of $\mathbf{X}$ and $\mathbf{Y}$ are positive integers.

The proof has two stages. First, we show in Lemma 8.9 that either $\mathsf{EVAL}(\mathbf{C}, \mathfrak{D})$ is #P-hard or $\mathbf{H}, \mathbf{X}$, and $\mathbf{Y}$ must satisfy

$$\langle \mathbf{H}_{i,*} \circ \overline{\mathbf{H}_{j,*}}, \mathbf{X}_{k,*} \rangle = 0 \quad \text{for all } k \in [t] \text{ and } i \neq j \in [h] \quad \text{and} \tag{8.8}$$

$$\langle \mathbf{H}_{*,i} \circ \overline{\mathbf{H}_{*,j}}, \mathbf{Y}_{k,*} \rangle = 0 \quad \text{for all } k \in [s] \text{ and } i \neq j \in [h]. \tag{8.9}$$

We use U to denote the set of h-dimensional vectors that are orthogonal to

$$\mathbf{H}_{1,*} \circ \overline{\mathbf{H}_{2,*}},\ \mathbf{H}_{1,*} \circ \overline{\mathbf{H}_{3,*}},\ \ldots,\ \mathbf{H}_{1,*} \circ \overline{\mathbf{H}_{h,*}}.$$

The above set of $h-1$ vectors is linearly independent. This is because

$$\sum_{i=2}^{h} a_i \left(\mathbf{H}_{1,*} \circ \overline{\mathbf{H}_{i,*}}\right) = \mathbf{H}_{1,*} \circ \left(\sum_{i=2}^{h} a_i \overline{\mathbf{H}_{i,*}}\right),$$

and if $\sum_{i=2}^{h} a_i (\mathbf{H}_{1,*} \circ \overline{\mathbf{H}_{i,*}}) = \mathbf{0}$, then $\sum_{i=2}^{h} a_i \overline{\mathbf{H}_{i,*}} = \mathbf{0}$ since all entries of $\mathbf{H}_{1,*}$ are nonzero. Because $\mathbf{H}$ has full rank, we have $a_i = 0$, $i = 2, \ldots, h$. As a result, U is a linear space of dimension 1 over $\mathbb{C}$.

Second, we show in Lemma 8.10 that, assuming (8.8) and (8.9), either

$$\langle \mathbf{H}_{i,*} \circ \overline{\mathbf{H}_{j,*}}, (\mathbf{X}_{k,*})^2 \rangle = 0 \quad \text{for all } k \in [t] \text{ and } i \neq j \in [h] \quad \text{and} \tag{8.10}$$

$$\langle \mathbf{H}_{*,i} \circ \overline{\mathbf{H}_{*,j}}, (\mathbf{Y}_{k,*})^2 \rangle = 0 \quad \text{for all } k \in [s] \text{ and } i \neq j \in [h], \tag{8.11}$$

or $\mathsf{EVAL}(\mathbf{C}, \mathfrak{D})$ is #P-hard. Here we use $(\mathbf{X}_{k,*})^2$ to denote $\mathbf{X}_{k,*} \circ \mathbf{X}_{k,*}$.

Equations (8.8) and (8.10) then imply that both $\mathbf{X}_{k,*}$ and $(\mathbf{X}_{k,*})^2$ are in U and thus they are linearly dependent (since the dimension of U is 1). On the other hand, by $(\mathcal{T}_3)$, every entry in $\mathbf{X}_{k,*}$ is a positive integer. Therefore, $\mathbf{X}_{k,*}$ must have the form $u \cdot \mathbf{1}$ for some positive integer u. The same argument works for $\mathbf{Y}_{k,*}$ and the latter must also have the form $u' \cdot \mathbf{1}$. By (8.8) and (8.9), this further implies that

$$\langle \mathbf{H}_{i,*}, \mathbf{H}_{j,*} \rangle = 0 \ \text{ and } \ \langle \mathbf{H}_{*,i}, \mathbf{H}_{*,j} \rangle = 0 \quad \text{for all } i \neq j \in [h].$$

This finishes the proof of Lemma 8.8. ∎

Now we proceed to the two stages of the proof. In the first stage, we prove the following.

LEMMA 8.9. *Either* $\mathbf{H}, \mathbf{X}, \mathbf{Y}$ *satisfy* (8.8) *and* (8.9)*, or* $\mathsf{EVAL}(\mathbf{C}, \mathfrak{D})$ *is #P-hard.*

Proof. Suppose $\mathsf{EVAL}(\mathbf{C}, \mathfrak{D})$ is not #P-hard; otherwise we are done.

We let $\mathfrak{D}^* = (\mathbf{D}^{[0]}, \ldots, \mathbf{D}^{[0]})$, a sequence of N $m \times m$ diagonal matrices in which every matrix is a copy of $\mathbf{D}^{[0]}$ (as in $\mathfrak{D}$). It is easy to check that $\mathfrak{D}^*$ satisfies condition $(\mathcal{T}_3)$. Let $G = (V, E)$ be an undirected graph. For each $p \geq 1$, we build a new graph $G^{[p]} = (V^{[p]}, E^{[p]})$ in the same way as we did in the proof of Lemma 8.5. This gives us an $m \times m$ matrix $\mathbf{C}^{[p]}$ such that $Z_{\mathbf{C}^{[p]}, \mathfrak{D}^*}(G) = Z_{\mathbf{C}, \mathfrak{D}}(G^{[p]})$ for all undirected graphs G. Thus, $\mathsf{EVAL}(\mathbf{C}^{[p]}, \mathfrak{D}^*) \leq \mathsf{EVAL}(\mathbf{C}, \mathfrak{D})$, and $\mathsf{EVAL}(\mathbf{C}^{[p]}, \mathfrak{D}^*)$ is also not #P-hard.

Matrix $\mathbf{C}^{[p]}$ is a block matrix with the same block structure as $\mathbf{C}$. The upper-right and lower-left blocks of $\mathbf{C}^{[p]}$ are zero matrices. For $\mathbf{x}, \mathbf{y} \in I$, we have

$$C^{[p]}_{(0,\mathbf{x}),(0,\mathbf{y})} = \left(\sum_{\mathbf{a} \in J} F_{\mathbf{x},\mathbf{a}} (F_{\mathbf{y},\mathbf{a}})^{pN-1} X_{a_1,a_2}\right) \left(\sum_{\mathbf{b} \in J} (F_{\mathbf{x},\mathbf{b}})^{pN-1} F_{\mathbf{y},\mathbf{b}} X_{b_1,b_2}\right).$$

From ($Shape_2$) and the fact that all entries of $\mathbf{X}$ are positive integers, we can rewrite the first factor as

$$\begin{aligned}&\mu_{x_1}(\mu_{y_1})^{pN-1}\sum_{\mathbf{a}\in J}(\nu_{a_1})^{pN}H_{x_2,a_2}\overline{H_{y_2,a_2}}X_{a_1,a_2}\\&\quad=\mu_{x_1}(\mu_{y_1})^{pN-1}\sum_{a\in[t]}(\nu_a)^{pN}\langle\mathbf{H}_{x_2,*}\circ\overline{\mathbf{H}_{y_2,*}},\mathbf{X}_{a,*}\rangle.\end{aligned}$$

Similarly, we can rewrite the second factor as

$$(\mu_{x_1})^{pN-1}\mu_{y_1}\sum_{a\in[t]}(\nu_a)^{pN}\overline{\langle\mathbf{H}_{x_2,*}\circ\overline{\mathbf{H}_{y_2,*}},\mathbf{X}_{a,*}\rangle}.$$

Since $\nu_a>0$ for all a, we have

$$C^{[p]}_{(0,\mathbf{x}),(0,\mathbf{y})}=(\mu_{x_1}\mu_{y_1})^{pN}\left|\sum_{a\in[t]}(\nu_a)^{pN}\langle\mathbf{H}_{x_2,*}\circ\overline{\mathbf{H}_{y_2,*}},\mathbf{X}_{a,*}\rangle\right|^2, \tag{8.12}$$

so the upper-left block of $\mathbf{C}^{[p]}$ is nonnegative. Similarly one can show that the same holds for its lower-right block. Thus, $\mathbf{C}^{[p]}$ is a nonnegative matrix.

Now for any $\mathbf{x}\in I$, we have

$$C^{[p]}_{(0,\mathbf{x}),(0,\mathbf{x})}=(\mu_{x_1})^{2pN}\left(\sum_{a\in[t]}(\nu_a)^{pN}\sum_{b\in[h]}X_{a,b}\right)^2,$$

which is positive, and for any $\mathbf{x}\neq\mathbf{y}\in I$, we have

$$C^{[p]}_{(0,\mathbf{x}),(0,\mathbf{x})}C^{[p]}_{(0,\mathbf{y}),(0,\mathbf{y})}=(\mu_{x_1}\mu_{y_1})^{2pN}\left(\sum_{a\in[t]}(\nu_a)^{pN}\sum_{b\in[h]}X_{a,b}\right)^4>0.$$

Since $\mathsf{EVAL}(\mathbf{C}^{[p]},\mathfrak{D}^*)$ is not #P-hard and $(\mathbf{C}^{[p]},\mathfrak{D}^*)$ satisfies $(\mathcal{T})$, by the inverse cyclotomic reduction lemma (Corollary 8.3), we have either

$$\left(C^{[p]}_{(0,\mathbf{x}),(0,\mathbf{y})}\right)^2=C^{[p]}_{(0,\mathbf{x}),(0,\mathbf{x})}C^{[p]}_{(0,\mathbf{y}),(0,\mathbf{y})}\ \text{ or }\ C^{[p]}_{(0,\mathbf{x}),(0,\mathbf{y})}=0. \tag{8.13}$$

We claim that if the former is true, then $x_2=y_2$. This is because, in this case,

$$\left|\sum_{a\in[t]}(\nu_a)^{pN}\langle\mathbf{H}_{x_2,*}\circ\overline{\mathbf{H}_{y_2,*}},\mathbf{X}_{a,*}\rangle\right|=\sum_{a\in[t]}(\nu_a)^{pN}\sum_{b\in[h]}X_{a,b},$$

and the norm of $\langle\mathbf{H}_{x_2,*}\circ\overline{\mathbf{H}_{y_2,*}},\mathbf{X}_{a,*}\rangle$ must be $\sum_{b\in[h]}X_{a,b}$. The inner product, however, is a sum of $X_{a,b}$'s weighted by roots of unity, so the entries of $\mathbf{H}_{x_2,*}\circ\overline{\mathbf{H}_{y_2,*}}$ must be the same root of unity. Thus, $\mathbf{H}_{x_2,*}$ and $\mathbf{H}_{y_2,*}$ are linearly dependent. Since $\mathbf{H}$ is a matrix of full rank, we conclude that $x_2=y_2$. Together with (8.13), we have

$$\sum_{a\in[t]}(\nu_a)^{pN}\langle\mathbf{H}_{x_2,*}\circ\overline{\mathbf{H}_{y_2,*}},\mathbf{X}_{a,*}\rangle=0\quad\text{for all } p\geq1 \text{ and all } x_2\neq y_2,$$

since the argument is independent of the value of p. These equations form a Vandermonde system, and we conclude that $\langle\mathbf{H}_{x_2,*}\circ\overline{\mathbf{H}_{y_2,*}},\mathbf{X}_{a,*}\rangle=0$ for all $a\in[t]$ and all $x_2\neq y_2$. This finishes the proof of (8.8). Equation (8.9) can be proved similarly. ∎

In the second stage, we prove the following lemma.

LEMMA 8.10. *Suppose matrices* $\mathbf{H}$, $\mathbf{X}$, *and* $\mathbf{Y}$ *satisfy both* (8.8) *and* (8.9). *Then either they also satisfy* (8.10) *and* (8.11) *or* $\mathsf{EVAL}(\mathbf{C}, \mathfrak{D})$ *is #P-hard.*

Proof. We will only prove (8.11). Equation (8.10) can be proved similarly.

Again, we let $\mathfrak{D}^*$ denote a sequence of N $m \times m$ diagonal matrices in which each matrix is a copy of $\mathbf{D}^{[0]}$ (so $\mathfrak{D}^*$ satisfies $(\mathcal{T}_3)$). Note that the matrix $\mathbf{C}^{[1]}$ we used in the proof of Lemma 8.9 satisfies the following property: When $x_2 = y_2$, by (8.12),

$$C^{[1]}_{(0,\mathbf{x}),(0,\mathbf{y})} = (\mu_{x_1}\mu_{y_1})^N \left(\sum_{a\in[t]} (\nu_a)^N \sum_{b\in[h]} X_{a,b} \right)^2,$$

and this is equal to 0 when $x_2 \neq y_2$. Let L denote the second factor on the right-hand side, which is independent of $\mathbf{x}$ and $\mathbf{y}$, so the right-hand side becomes $(\mu_{x_1}\mu_{y_1})^N L$.

Additionally, because of (8.9), we have that $\mathbf{Y}_{k,*}$ and $\mathbf{Y}_{1,*}$ are linearly dependent for every k. Thus, for every $k \in [s]$, there exists a positive, rational λ_k such that

$$\mathbf{Y}_{k,*} = \lambda_k \cdot \mathbf{Y}_{1,*}. \tag{8.14}$$

Because of this, we only need to prove (8.11) for the case when $k = 1$.

Now we start the proof of (8.11). Suppose $\mathsf{EVAL}(\mathbf{C}, \mathfrak{D})$ is not #P-hard. We use $G = (V, E)$ to denote an undirected graph; then for each $p \geq 1$, we build a new graph $G^{(p)} = (V^{(p)}, E^{(p)})$ by replacing every edge $e = uv \in E$ with a gadget that is shown in Figure 8.2. More exactly, we define $G^{(p)} = (V^{(p)}, E^{(p)})$ as

$$V^{(p)} = V \cup \{a_e, b_e, c_e, d_e, a'_e, b'_e, c'_e, d'_e : e \in E\},$$

and $E^{(p)}$ contains exactly the following edges: For every edge $e = uv \in E$, add

1. one edge $(u, a_e), (a'_e, v), (c_e, b_e), (d_e, a_e), (c'_e, b'_e)$, and (d'_e, a'_e);
2. $pN - 1$ parallel edges between (a_e, v) and (u, a'_e);
3. $N - 1$ parallel edges between $(a_e, c_e), (b_e, d_e), (a'_e, c'_e)$, and (b'_e, d'_e).

It is easy to check that the degree of every vertex in $G^{(p)}$ is a multiple of N.

Moreover, the construction of $G^{(p)}$ gives us a new $m \times m$ matrix $\mathbf{R}^{(p)}$, which is symmetric since the gadget is symmetric, such that $Z_{\mathbf{R}^{(p)},\mathfrak{D}^*}(G) = Z_{\mathbf{C},\mathfrak{D}}(G^{(p)})$ for all G. Thus, $\mathsf{EVAL}(\mathbf{R}^{(p)}, \mathfrak{D}^*) \leq \mathsf{EVAL}(\mathbf{C}, \mathfrak{D})$, and $\mathsf{EVAL}(\mathbf{R}^{(p)}, \mathfrak{D}^*)$ is also not #P-hard.

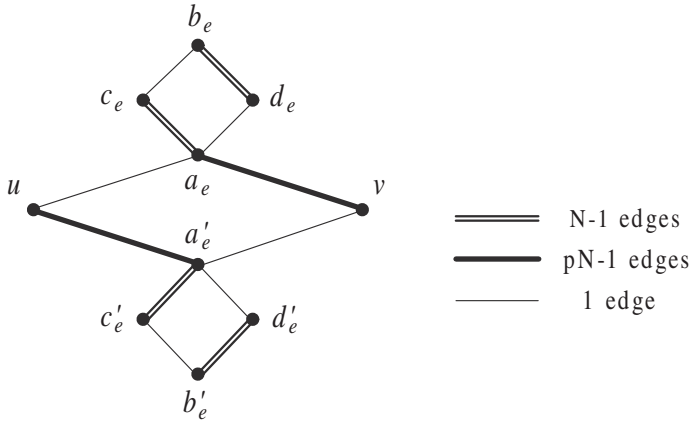


FIG. 8.2. *Gadget for constructing* $G^{(p)}$, $p \geq 1$.

Moreover, $\mathbf{R}^{(p)}$ is a block matrix which has the same block structure as $\mathbf{C}$. The upper-right and lower-left blocks of $\mathbf{R}^{(p)}$ are zero matrices. The entries in its lower-right block are as follows: For $\mathbf{x}, \mathbf{y} \in J$,

$$R^{(p)}_{(1,\mathbf{x}),(1,\mathbf{y})} = \left(\sum_{\mathbf{a},\mathbf{b} \in I} F_{\mathbf{a},\mathbf{x}} (F_{\mathbf{a},\mathbf{y}})^{pN-1} C^{[1]}_{(0,\mathbf{a}),(0,\mathbf{b})} Y_{a_1,a_2} Y_{b_1,b_2} \right) \times \left(\sum_{\mathbf{a},\mathbf{b} \in I} (F_{\mathbf{a},\mathbf{x}})^{pN-1} F_{\mathbf{a},\mathbf{y}} C^{[1]}_{(0,\mathbf{a}),(0,\mathbf{b})} Y_{a_1,a_2} Y_{b_1,b_2} \right).$$

Recall that $C^{[1]}_{(0,\mathbf{a}),(0,\mathbf{b})} = 0$ when $a_2 \neq b_2$. From (8.14), $Y_{a_1,a_2} Y_{b_1,b_2} = \lambda_{a_1} \lambda_{b_1} Y_{1,a_2} Y_{1,b_2}$. As a result, we can simplify the first factor to be

$$\begin{aligned} &\nu_{x_1} (\nu_{y_1})^{pN-1} L \sum_{\mathbf{a},\mathbf{b} \in I, a_2 = b_2} (\mu_{a_1})^{pN} H_{a_2,x_2} \overline{H_{a_2,y_2}} (\mu_{a_1} \mu_{b_1})^N \lambda_{a_1} \lambda_{b_1} Y_{1,a_2} Y_{1,b_2} \\ &= \nu_{x_1} (\nu_{y_1})^{pN-1} L \sum_{a_1,b_1 \in [s]} (\mu_{a_1})^{(p+1)N} (\mu_{b_1})^N \lambda_{a_1} \lambda_{b_1} \sum_{a_2 \in [h]} H_{a_2,x_2} \overline{H_{a_2,y_2}} (Y_{1,a_2})^2 \\ &= \nu_{x_1} (\nu_{y_1})^{pN-1} L' \cdot \langle \mathbf{H}_{*,x_2} \circ \overline{\mathbf{H}_{*,y_2}}, (\mathbf{Y}_{1,*})^2 \rangle, \end{aligned}$$

where

$$L' = L \sum_{a_1,b_1 \in [s]} (\mu_{a_1})^{(p+1)N} (\mu_{b_1})^N \lambda_{a_1} \lambda_{b_1}$$

is positive and is independent of $\mathbf{x}, \mathbf{y}$. Similarly, the second factor becomes

$$(\nu_{x_1})^{pN-1} \nu_{y_1} L' \cdot \overline{\langle \mathbf{H}_{*,x_2} \circ \overline{\mathbf{H}_{*,y_2}}, (\mathbf{Y}_{1,*})^2 \rangle}.$$

As a result, we have

$$R^{(p)}_{(1,\mathbf{x}),(1,\mathbf{y})} = (L')^2 \cdot (\nu_{x_1} \nu_{y_1})^{pN} \cdot \left| \langle \mathbf{H}_{*,x_2} \circ \overline{\mathbf{H}_{*,y_2}}, (\mathbf{Y}_{1,*})^2 \rangle \right|^2.$$

Thus the lower-right block of $\mathbf{R}^{(p)}$ is nonnegative. Similarly, one can prove that the same holds for its upper-left block, so $\mathbf{R}^{(p)}$ is nonnegative.

We apply Corollary 8.3 to $(\mathbf{R}^{(p)}, \mathfrak{D}^*)$. As $\mathsf{EVAL}(\mathbf{R}^{(p)}, \mathfrak{D}^*)$ is not #P-hard, either

$$\left(R^{(p)}_{(1,\mathbf{x}),(1,\mathbf{y})}\right)^2 = R^{(p)}_{(1,\mathbf{x}),(1,\mathbf{x})} R^{(p)}_{(1,\mathbf{y}),(1,\mathbf{y})} \quad \text{or} \quad R^{(p)}_{(1,\mathbf{x}),(1,\mathbf{y})} = 0 \quad \text{for any } \mathbf{x} \neq \mathbf{y} \in J.$$

We claim that if the former is true, then $x_2 = y_2$. This is because, in this case,

$$\left| \langle \mathbf{H}_{*,x_2} \circ \overline{\mathbf{H}_{*,y_2}}, (\mathbf{Y}_{1,*})^2 \rangle \right| = \sum_{i \in [h]} Y_{1,i}^2.$$

However, the left-hand side is a sum of $(Y_{1,i})^2$, which are positive integers, weighted by roots of unity. To sum to a number of norm $\sum_{i \in [h]} Y_{1,i}^2$ the entries of $\mathbf{H}_{*,x_2} \circ \overline{\mathbf{H}_{*,y_2}}$ must be the same root of unity. As a result, $\mathbf{H}_{*,x_2}$ and $\mathbf{H}_{*,y_2}$ are linearly dependent. Since $\mathbf{H}$ is of full rank, we conclude that $x_2 = y_2$. In other words, we have shown that

$$\langle \mathbf{H}_{*,x_2} \circ \overline{\mathbf{H}_{*,y_2}}, (\mathbf{Y}_{1,*})^2 \rangle = 0 \quad \text{for all } x_2 \neq y_2.$$

By combining it with (8.14), we have finished the proof of (8.11). ∎

8.4. Step 2.3. Now we get a pair $(\mathbf{C}, \mathfrak{D})$ that satisfies $(Shape_1)$–$(Shape_5)$ since otherwise, by Lemma 8.8, $\mathsf{EVAL}(\mathbf{C}, \mathfrak{D})$ is #P-hard and we are done.

In particular, by using $(Shape_5)$, we define two diagonal matrices $\mathbf{K}^{[0]}$ and $\mathbf{L}^{[0]}$ as follows. $\mathbf{K}^{[0]}$ is an $(s+t) \times (s+t)$ diagonal matrix. We use $(0, i)$, $i \in [s]$, to index its first s rows and $(1, j)$, $j \in [t]$, to index its last t rows. Its diagonal entries are

$$K^{[0]}_{(0,i)} = D^{[0]}_{(0,(i,1))} \quad \text{and} \quad K^{[0]}_{(1,j)} = D^{[0]}_{(1,(j,1))} \quad \text{for all } i \in [s] \text{ and } j \in [t].$$

$\mathbf{L}^{[0]}$ is the $2h \times 2h$ identity matrix. We use $(0, i)$, $i \in [h]$, to index its first h rows and $(1, j)$, $j \in [h]$, to index its last h rows. By $(Shape_5)$, we have

$$D^{[0]}_{(0,\mathbf{x})} = K^{[0]}_{(0,x_1)} \cdot L^{[0]}_{(0,x_2)} \quad \text{and} \quad D^{[0]}_{(1,\mathbf{y})} = K^{[0]}_{(1,y_1)} \cdot L^{[0]}_{(1,y_2)} \tag{8.15}$$

for all $\mathbf{x} \in I$ and $\mathbf{y} \in J$, or equivalently,

$$\mathbf{D}^{[0]} = \begin{pmatrix} \mathbf{D}^{[0]}_{(0,*)} & \\ & \mathbf{D}^{[0]}_{(1,*)} \end{pmatrix} = \begin{pmatrix} \mathbf{K}^{[0]}_{(0,*)} \otimes \mathbf{L}^{[0]}_{(0,*)} & \\ & \mathbf{K}^{[0]}_{(1,*)} \otimes \mathbf{L}^{[0]}_{(1,*)} \end{pmatrix}. \tag{8.16}$$

The goal of Step 2.3 is to prove a similar statement for $\mathbf{D}^{[r]}$, $r \in [N-1]$, and these equations will allow us in Step 2.4 to decompose $\mathsf{EVAL}(\mathbf{C}, \mathfrak{D})$ into two subproblems.

In the proof of Lemma 8.8, we crucially used the property (from $(\mathcal{T}_3)$) that all the diagonal entries of $\mathbf{D}^{[0]}$ are positive integers. However, for $r \geq 1$, $(\mathcal{T}_3)$ only gives us some very weak properties about $\mathbf{D}^{[r]}$. For example, the entries are not guaranteed to be real numbers. So the proof that we are going to present here is more difficult. We prove the following lemma.

LEMMA 8.11. *Let $(\mathbf{C}, \mathfrak{D})$ be a pair that satisfies $(Shape_1)$–$(Shape_5)$. Then either $\mathsf{EVAL}(\mathbf{C}, \mathfrak{D})$ is #P-hard or it satisfies the following additional condition:*

$(Shape_6)$ There exist diagonal matrices $\mathbf{K}^{[0]}$ and $\mathbf{L}^{[0]}$ such that $\mathbf{D}^{[0]}, \mathbf{K}^{[0]}$, and $\mathbf{L}^{[0]}$ satisfy (8.16). Every entry of $\mathbf{K}^{[0]}$ is a positive integer, and $\mathbf{L}^{[0]}$ is the $2h \times 2h$ identity matrix. For each $r \in [N-1]$, there exist two diagonal matrices $\mathbf{K}^{[r]}$ and $\mathbf{L}^{[r]}$. $\mathbf{K}^{[r]}$ is an $(s+t) \times (s+t)$ matrix, and $\mathbf{L}^{[r]}$ is a $2h \times 2h$ matrix. We index $\mathbf{K}^{[r]}$ and $\mathbf{L}^{[r]}$ in the same way we index $\mathbf{K}^{[0]}$ and $\mathbf{L}^{[0]}$, respectively. Then

$$\mathbf{D}^{[r]} = \begin{pmatrix} \mathbf{D}^{[r]}_{(0,*)} & \\ & \mathbf{D}^{[r]}_{(1,*)} \end{pmatrix} = \begin{pmatrix} \mathbf{K}^{[r]}_{(0,*)} \otimes \mathbf{L}^{[r]}_{(0,*)} & \\ & \mathbf{K}^{[r]}_{(1,*)} \otimes \mathbf{L}^{[r]}_{(1,*)} \end{pmatrix}.$$

Moreover, the norm of every entry in $\mathbf{L}^{[r]}$ is either 0 or 1, and for any $r \in [N-1]$,

$$\mathbf{K}^{[r]}_{(0,*)} = \mathbf{0} \iff \mathbf{L}^{[r]}_{(0,*)} = \mathbf{0} \quad \text{and} \quad \mathbf{K}^{[r]}_{(1,*)} = \mathbf{0} \iff \mathbf{L}^{[r]}_{(1,*)} = \mathbf{0};$$

$$\mathbf{L}^{[r]}_{(0,*)} \neq \mathbf{0} \implies \exists\, i \in [h],\ L^{[r]}_{(0,i)} = 1 \quad \text{and} \quad \mathbf{L}^{[r]}_{(1,*)} \neq \mathbf{0} \implies \exists\, i \in [h],\ L^{[r]}_{(1,i)} = 1.$$

We now present the proof of Lemma 8.11. Fix an $r \in [N-1]$ to be any index. We use the following notation. Consider the diagonal matrix $\mathbf{D}^{[r]}$. It has two parts:

$$\mathbf{D}^{[r]}_{(0,*)} \in \mathbb{C}^{sh \times sh} \quad \text{and} \quad \mathbf{D}^{[r]}_{(1,*)} \in \mathbb{C}^{th \times th}.$$

The first part has s blocks, where each block is a diagonal matrix with h entries. We will rearrange the entries indexed by $(0, *)$ into another $s \times h$ matrix, which we denote as $\mathbf{D}$ (just as we did with $\mathbf{D}^{[0]}$ in the proof of Lemma 8.8), where

$$D_{i,j} = D^{[r]}_{(0,(i,j))} \quad \text{for all } i \in [s] \text{ and } j \in [h].$$

We prove the following lemma in section 8.4.2.

LEMMA 8.12. *Either problem* $\mathsf{EVAL}(\mathbf{C}, \mathfrak{D})$ *is #P-hard, or we have* (1) $\operatorname{rank}(\mathbf{D}) \le 1$ *and* (2) *for each* $i \in [s]$, *all nonzero entries of* $\mathbf{D}_{i,*}$ *have the same norm.*

Proof of Lemma 8.11. We start with the first half, that is,

$$\mathbf{D}^{[r]}_{(0,*)} = \mathbf{K}^{[r]}_{(0,*)} \otimes \mathbf{L}^{[r]}_{(0,*)}. \tag{8.17}$$

Assume $\mathbf{D}^{[r]}_{(0,*)}$ is nonzero; otherwise the lemma is true by setting $\mathbf{K}^{[r]}_{(0,*)} = \mathbf{L}^{[r]}_{(0,*)} = \mathbf{0}$. As a result, we know that $\mathbf{D} \neq \mathbf{0}$.

Let $D_{a,b}$ be a nonzero entry of $\mathbf{D}$, where $a \in [s]$ and $b \in [h]$. From Lemma 8.12, the rank of $\mathbf{D}$ is 1, so $\mathbf{D}_{i,*} = (D_{i,b}/D_{a,b}) \cdot \mathbf{D}_{a,*}$ for any $i \in [s]$. By setting

$$K^{[r]}_{(0,i)} = D_{i,b} \quad \text{and} \quad L^{[r]}_{(0,j)} = \frac{D_{a,j}}{D_{a,b}},$$

we have

$$D^{[r]}_{(0,(i,j))} = D_{i,j} = K^{[r]}_{(0,i)} \cdot L^{[r]}_{(0,j)} \quad \text{for all } i \in [s] \text{ and } j \in [h],$$

and (8.17) follows. The second half can be proved similarly.

One can also check that $\mathbf{K}^{[r]}$ and $\mathbf{L}^{[r]}$ satisfy all the properties stated in ($Shape_6$). This finishes the proof of Lemma 8.11 (assuming Lemma 8.12). ∎

8.4.1. The vanishing lemma. We need the following lemma in the proof of Lemma 8.12.

LEMMA 8.13 (vanishing lemma). *Let* k *be a positive integer and let* $(x_{i,n})_{n\ge1}$, *for* $1 \le i \le k$, *be* k *infinite sequences of nonzero real numbers. For notational uniformity we also denote by* $(x_{0,n})_{n\ge1}$ *the sequence where* $x_{0,n} = 1$ *for all* $n \ge 1$. *Suppose*

$$\lim_{n\to\infty} \frac{x_{i+1,n}}{x_{i,n}} = 0 \quad \textit{for } 0 \le i < k.$$

Part A. *Let* a_i *and* $b_i \in \mathbb{C}$ *for* $0 \le i \le k$. *Suppose for some* $\ell \in [k]$, $a_i = b_i$ *for all* $0 \le i < \ell$; $a_0 = b_0 = 1$; *and* $\operatorname{Im}(a_\ell) = \operatorname{Im}(b_\ell)$. *If for infinitely many* n,

$$\left|\sum_{i=0}^{k} a_i x_{i,n}\right| = \left|\sum_{i=0}^{k} b_i x_{i,n}\right|,$$

then we have $a_\ell = b_\ell$.

Part B. *Let* $a_i \in \mathbb{C}$ *for* $0 \le i \le k$. *If for infinitely many* n,

$$\left|\sum_{i=0}^{k} a_i x_{i,n}\right| = 0,$$

then we have $a_i = 0$ *for all* $0 \le i \le k$.

Proof. We first prove Part B, which is simpler. Taking $n \to \infty$ (technically we take a subsequence of n approaching ∞ where the equality holds, and the same below), we get $a_0 = 0$. Since $x_{1,n} \neq 0$, we can divide out $|x_{1,n}|$ and get for infinitely many n,

$$\left|\sum_{i=1}^{k} a_i x_{i,n}/x_{1,n}\right| = 0.$$

Now the result follows by induction.

Next we prove Part A. Multiplying by its conjugate, we get

$$\left(\sum_{i=0}^{k} a_i x_{i,n}\right)\left(\sum_{j=0}^{k} \overline{a_j} x_{j,n}\right) = \left(\sum_{i=0}^{k} b_i x_{i,n}\right)\left(\sum_{j=0}^{k} \overline{b_j} x_{j,n}\right).$$

Every term involves a product $x_{i,n}x_{j,n}$. If $\max\{i,j\} < \ell$, then the terms

$$a_i\overline{a_j}x_{i,n}x_{j,n} = b_i\overline{b_j}x_{i,n}x_{j,n}$$

and they cancel (since $a_i = b_i$ and $a_j = b_j$). If $\max\{i,j\} > \ell$, then both $a_i\overline{a_j}x_{i,n}x_{j,n}$ and $b_i\overline{b_j}x_{i,n}x_{j,n}$ are $o(|x_{\ell,n}|)$ as $n \to \infty$. This is also true when $\max\{i,j\} = \ell$ and $\min\{i,j\} > 0$. The remaining terms correspond to $\max\{i,j\} = \ell$ and $\min\{i,j\} = 0$. After canceling out identical terms, we get

$$(a_\ell + \overline{a_\ell})x_{\ell,n} + o(|x_{\ell,n}|) = (b_\ell + \overline{b_\ell})x_{\ell,n} + o(|x_{\ell,n}|)$$

as $n \to \infty$. Dividing out $x_{\ell,n}$ and then taking limit $n \to \infty$, we get $\mathrm{Re}(a_\ell) = \mathrm{Re}(b_\ell)$. It follows that $a_\ell = b_\ell$ since $\mathrm{Im}(a_\ell) = \mathrm{Im}(b_\ell)$. □

We also remark that Part A of the vanishing lemma above cannot be extended to arbitrary sequences $\{a_i\}$ and $\{b_i\}$ without the condition that $\mathrm{Im}(a_\ell) = \mathrm{Im}(b_\ell)$, as shown by the following example: Let

$$a_1 = 3 + \sqrt{3}i, \quad a_2 = 3\left(\frac{1}{2} + \frac{\sqrt{3}}{2}i\right), \quad \text{and} \quad b_1 = b_2 = 3.$$

Then $|1 + a_1x + a_2x^2| = |1 + b_1x + b_2x^2|$ is an identity for all real values x. In particular this holds when $x \to 0$. We note that $a_1 \neq b_1$.

8.4.2. Proof of Lemma 8.12. Without loss of generality, we assume $1 = \mu_1 > \cdots > \mu_s > 0$ and $1 = \nu_1 > \cdots > \nu_t > 0$. (Otherwise, we can multiply $\mathbf{C}$ by an appropriate scalar so that the new $\mathbf{C}$ has this property. This operation clearly does not affect the complexity of $\mathsf{EVAL}(\mathbf{C}, \mathfrak{D})$.) We assume $\mathsf{EVAL}(\mathbf{C}, \mathfrak{D})$ is not #P-hard.

Again we let $\mathfrak{D}^*$ denote a sequence of N $m \times m$ diagonal matrices in which every matrix is a copy of the matrix $\mathbf{D}^{[0]}$ in $\mathfrak{D}$ (so $\mathfrak{D}^*$ satisfies $(\mathcal{T}_3)$). Recall that r is a fixed index in $[N-1]$, and recall the definition of the $s \times h$ matrix $\mathbf{D}$ from $\mathbf{D}^{[r]}$.

Let $G = (V, E)$ be an undirected graph. For each $n \geq 1$, we build a new graph $G^{[n]}$ by replacing each edge $uv \in E$ with a gadget shown in Figure 8.3. More exactly, we define $G^{[n]}$ as follows. Let $p_n = n^2N + 1$ and $q_n = nN - 1$. (When $n \to \infty$, q_n will be arbitrarily large, and for a given q_n, p_n will be arbitrarily *larger*.) Then

$$V^{[n]} = V \cup \left\{a_e, x_{e,i}, y_{e,i}, b_e, c_e, a'_e, x'_{e,i}, y'_{e,i}, b'_e, c'_e : e \in E, i \in [r]\right\},$$

and $E^{[n]}$ contains exactly the following edges: For every edge $e = uv \in E$, add

1. one edge $(u, a_e), (v, a'_e), (a_e, y_{e,i})$, and $(a'_e, y'_{e,i})$ for all $i \in [r]$;
2. $N - 1$ parallel edges $(v, a_e), (u, a'_e), (a_e, x_{e,i})$, and $(a'_e, x'_{e,i})$ for all $i \in [r]$;
3. p_n parallel edges $(b_e, x_{e,i})$ and $(b'_e, x'_{e,i})$ for all $i \in [r]$;
4. q_n parallel edges $(c_e, y_{e,i})$ and $(c'_e, y'_{e,i})$ for all $i \in [r]$.

It is easy to check that the degree of every vertex in $G^{[n]}$ is a multiple of N except for b_e and b'_e, which have degree r mod N, and c_e and c'_e, which have degree $N - r$ mod N.

Since the gadget is symmetric with respect to vertices u and v, the construction of $G^{[n]}$ gives us a symmetric $m \times m$ matrix $\mathbf{R}^{[n]}$ (recall that $m = (s + t)h$) such that

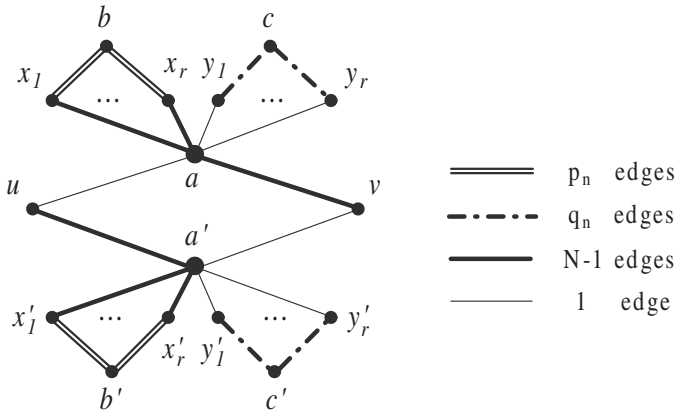


FIG. 8.3. *Gadget for constructing* $G^{[n]}$, $n \geq 1$. *(Note that the subscript e is suppressed.)*

$Z_{\mathbf{R}^{[n]},\mathfrak{D}^*}(G) = Z_{\mathbf{C},\mathfrak{D}}(G^{[n]})$ for all G. As a result, $\mathsf{EVAL}(\mathbf{R}^{[n]},\mathfrak{D}^*) \leq \mathsf{EVAL}(\mathbf{C},\mathfrak{D})$, and we know that $\mathsf{EVAL}(\mathbf{R}^{[n]},\mathfrak{D}^*)$ is also not #P-hard.

The entries of $\mathbf{R}^{[n]}$ are as follows: For $\mathbf{u} \in I$ and $\mathbf{v} \in J$, the $((0,\mathbf{u}),(1,\mathbf{v}))$th and $((1,\mathbf{u}),(0,\mathbf{v}))$th entries of $\mathbf{R}^{[n]}$ are zero. For $\mathbf{u},\mathbf{v} \in J$, $R^{[n]}_{(1,\mathbf{u}),(1,\mathbf{v})}$ is the product of

$$\sum_{\mathbf{a},\mathbf{b},\mathbf{c}\in I}\left(\sum_{\mathbf{x}\in J} F^{N-1}_{\mathbf{a},\mathbf{x}} F^{p_n}_{\mathbf{b},\mathbf{x}} D^{[0]}_{(1,\mathbf{x})}\right)^r \left(\sum_{\mathbf{y}\in J} F_{\mathbf{a},\mathbf{y}} F^{q_n}_{\mathbf{c},\mathbf{y}} D^{[0]}_{(1,\mathbf{y})}\right)^r F_{\mathbf{a},\mathbf{u}} F^{N-1}_{\mathbf{a},\mathbf{v}} D^{[0]}_{(0,\mathbf{a})} D^{[r]}_{(0,\mathbf{b})} D^{[N-r]}_{(0,\mathbf{c})}$$

and

$$\sum_{\mathbf{a},\mathbf{b},\mathbf{c}\in I}\left(\sum_{\mathbf{x}\in J} F^{N-1}_{\mathbf{a},\mathbf{x}} F^{p_n}_{\mathbf{b},\mathbf{x}} D^{[0]}_{(1,\mathbf{x})}\right)^r \left(\sum_{\mathbf{y}\in J} F_{\mathbf{a},\mathbf{y}} F^{q_n}_{\mathbf{c},\mathbf{y}} D^{[0]}_{(1,\mathbf{y})}\right)^r F^{N-1}_{\mathbf{a},\mathbf{u}} F_{\mathbf{a},\mathbf{v}} D^{[0]}_{(0,\mathbf{a})} D^{[r]}_{(0,\mathbf{b})} D^{[N-r]}_{(0,\mathbf{c})}.$$

We simplify the first sum. By using ($Shape_2$) and ($Shape_5$), we have

$$\begin{aligned}\sum_{\mathbf{x}\in J} F^{N-1}_{\mathbf{a},\mathbf{x}} F^{p_n}_{\mathbf{b},\mathbf{x}} D^{[0]}_{(1,\mathbf{x})} &= \mu^{N-1}_{a_1}\mu^{p_n}_{b_1} \sum_{\mathbf{x}\in J} (\nu_{x_1})^{N-1+p_n} \overline{H_{a_2,x_2}} H_{b_2,x_2} D^{[0]}_{(1,(x_1,1))} \\ &= \mu^{N-1}_{a_1}\mu^{p_n}_{b_1} \sum_{x_1\in[t]} (\nu_{x_1})^{N-1+p_n} D^{[0]}_{(1,(x_1,1))} \langle \mathbf{H}_{b_2,*}, \mathbf{H}_{a_2,*}\rangle. \end{aligned} \tag{8.18}$$

Let L denote the following positive number that is independent of $\mathbf{u},\mathbf{v},\mathbf{a},\mathbf{b}$, and $\mathbf{c}$:

$$L = h \cdot \sum_{x_1\in[t]} (\nu_{x_1})^{N-1+p_n} \cdot D^{[0]}_{(1,(x_1,1))}.$$

By ($Shape_4$), (8.18) is equal to $L \cdot \mu^{N-1}_{a_1}\mu^{p_n}_{b_1}$ if $a_2 = b_2$ and 0 otherwise. Similarly,

$$\sum_{\mathbf{y}\in J} F_{\mathbf{a},\mathbf{y}} F^{q_n}_{\mathbf{c},\mathbf{y}} D^{[0]}_{(1,\mathbf{y})} = L' \cdot \mu_{a_1}\mu^{q_n}_{c_1} \quad \text{if } a_2 = c_2$$

and 0 otherwise, where L' is a positive number independent of $\mathbf{u},\mathbf{v},\mathbf{a},\mathbf{b}$, and $\mathbf{c}$.

By ($Shape_3$), we have

$$D^{[N-r]}_{(0,\mathbf{c})} = \overline{D^{[r]}_{(0,\mathbf{c})}} = \overline{D_{c_1,c_2}}.$$

Combining these equations, the first factor of $R^{[n]}_{(1,\mathbf{u}),(1,\mathbf{v})}$ becomes

$$\nu_{u_1}\nu_{v_1}^{N-1}\sum_{\mathbf{a}\in I,b,c\in[s]}\left(L\cdot\mu_{a_1}^{N-1}\mu_b^{p_n}\right)^r\left(L'\cdot\mu_{a_1}\mu_c^{q_n}\right)^r\mu_{a_1}^N H_{a_2,u_2}\overline{H_{a_2,v_2}}D^{[0]}_{(0,(a_1,1))}D_{b,a_2}\overline{D_{c,a_2}}.$$

Let Z denote the following positive number that is independent of $\mathbf{u}$ and $\mathbf{v}$:

$$Z=\sum_{a_1\in[s]}\left(L\cdot\mu_{a_1}^{N-1}\right)^r\left(L'\cdot\mu_{a_1}\right)^r\mu_{a_1}^N D^{[0]}_{(0,(a_1,1))}.$$

Let $P_n=rp_n$ and $Q_n=rq_n$; then the first factor becomes

$$Z\cdot\nu_{u_1}\nu_{v_1}^{N-1}\sum_{b,c\in[s]}\mu_b^{P_n}\mu_c^{Q_n}\sum_{a\in[h]}D_{b,a}\overline{D_{c,a}}H_{a,u_2}\overline{H_{a,v_2}}.$$

We can also simplify the second factor so that

$$R^{[n]}_{(1,\mathbf{u}),(1,\mathbf{v})}=Z^2(\nu_{u_1}\nu_{v_1})^N\left(\sum_{b,c\in[s]}\mu_b^{P_n}\mu_c^{Q_n}\sum_{a\in[h]}D_{b,a}\overline{D_{c,a}}H_{a,u_2}\overline{H_{a,v_2}}\right)\times\left(\sum_{b',c'\in[s]}\mu_{b'}^{P_n}\mu_{c'}^{Q_n}\sum_{a\in[h]}D_{b',a}\overline{D_{c',a}H_{a,u_2}}H_{a,v_2}\right).$$

As $\mathsf{EVAL}(\mathbf{R}^{[n]},\mathfrak{D}^*)$ is not #P-hard and $(\mathbf{R}^{[n]},\mathfrak{D}^*)$ satisfies $(\mathcal{T})$ for all $n\geq 1$, the necessary condition of the inverse cyclotomic reduction lemma (Corollary 8.3) applies to $\mathbf{R}^{[n]}$.

In the proof below, for notational convenience we suppress the index $n\geq 1$ and use P,Q, and $\mathbf{R}$ to represent sequences $\{P_n\},\{Q_n\}$, and $\{\mathbf{R}^{[n]}\}$, respectively. Whenever we state or prove a property about $\mathbf{R}$, we mean $\mathbf{R}^{[n]}$ has this property for any large enough n (sometimes it holds for all $n\geq 1$). Moreover, since we only use the entries of $\mathbf{R}^{[n]}$ indexed by $((1,\mathbf{u}),(1,\mathbf{v}))$ with $u_1=v_1=1$, we let $R_{u,v}\equiv R_{(1,(1,u)),(1,(1,v))}$ for all $u,v\in[h]$. As a result, we have (note that $\nu_1=1$)

(8.19)
$$R_{u,v}=Z^2\left(\sum_{b,c\in[s]}\mu_b^P\mu_c^Q\sum_{a\in[h]}D_{b,a}\overline{D_{c,a}}H_{a,u}\overline{H_{a,v}}\right)\left(\sum_{b',c'\in[s]}\mu_{b'}^P\mu_{c'}^Q\sum_{a\in[h]}D_{b',a}\overline{D_{c',a}H_{a,u}}H_{a,v}\right).$$

We will consider the above expression for $R_{u,v}$ *stratified* according to the order of magnitude of $\mu_b^P\mu_c^Q\mu_{b'}^P\mu_{c'}^Q=(\mu_b\mu_{b'})^P(\mu_c\mu_{c'})^Q$. Because $P=\Theta(n^2)$ and $Q=\Theta(n)$, when $n\to\infty$, Q is arbitrarily and sufficiently large, and P is further arbitrarily and sufficiently large compared to Q. Thus, terms are ordered strictly first by $\mu_b\mu_{b'}$ and then by $\mu_c\mu_{c'}$. Inspired by this, we define the following total order $\leq_\mu$ over

$$\mathcal{T}=\left\{\begin{pmatrix}b & c\\ b' & c'\end{pmatrix}:b,b',c,c'\in[s]\right\}.$$

For T_1 and T_2 in $\mathcal{T}$, where

$$T_1=\begin{pmatrix}b_1 & c_1\\ b_1' & c_1'\end{pmatrix}\quad\text{and}\quad T_2=\begin{pmatrix}b_2 & c_2\\ b_2' & c_2'\end{pmatrix},$$

we have $T_1 \le_\mu T_2$ if either $\mu_{b_1}\mu_{b_1'} < \mu_{b_2}\mu_{b_2'}$, or $\mu_{b_1}\mu_{b_1'} = \mu_{b_2}\mu_{b_2'}$ and $\mu_{c_1}\mu_{c_1'} \le \mu_{c_2}\mu_{c_2'}$. For convenience, we denote the entries of a 2×2 matrix T_i or T in $\mathcal{T}$ by

$$\begin{pmatrix} b_i & c_i \\ b_i' & c_i' \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} b & c \\ b' & c' \end{pmatrix},$$

respectively. Using $\le_\mu$, we divide $\mathcal{T}$ into classes $\mathcal{T}_1, \mathcal{T}_2, \ldots, \mathcal{T}_d$ ordered from the largest to the smallest, for some $d \ge 1$, such that the following hold:

1. If $T_1, T_2 \in \mathcal{T}_i$, for some $i \in [d]$, then $\mu_{b_1}\mu_{b_1'} = \mu_{b_2}\mu_{b_2'}$ and $\mu_{c_1}\mu_{c_1'} = \mu_{c_2}\mu_{c_2'}$. Note that this is an equivalence relation which we denote by $\equiv_\mu$.

2. If $T_1 \in \mathcal{T}_i, T_2 \in \mathcal{T}_j$ and $i < j$, then either $\mu_{b_1}\mu_{b_1'} > \mu_{b_2}\mu_{b_2'}$ or $\mu_{b_1}\mu_{b_1'} = \mu_{b_2}\mu_{b_2'}$ and $\mu_{c_1}\mu_{c_1'} > \mu_{c_2}\mu_{c_2'}$.

For each $i \in [d]$, we arbitrarily pick a $T \in \mathcal{T}_i$ and use U_i to denote $\mu_b\mu_{b'}$ and W_i to denote $\mu_c\mu_{c'}$. (Note that U_i and W_i are independent of the choice of T.) It is clear that there is exactly one matrix, $\left(\begin{smallmatrix} 1 & 1 \\ 1 & 1 \end{smallmatrix}\right)$, in $\mathcal{T}_1$.

Now we can rewrite (8.19) as follows:

$$R_{u,v} = Z^2 \sum_{i \in [d]} U_i^P W_i^Q \sum_{T \in \mathcal{T}_i} X_{u,v,T}, \tag{8.20}$$

where

$$X_{u,v,T} = \left(\sum_{a \in [h]} D_{b,a} \overline{D_{c,a}} H_{a,u} \overline{H_{a,v}} \right) \left(\sum_{a \in [h]} D_{b',a} \overline{D_{c',a} H_{a,u}} H_{a,v} \right) \quad \text{for } T = \begin{pmatrix} b & c \\ b' & c' \end{pmatrix}.$$

Clearly the term with the maximum possible order in the sum (8.20) corresponds to the choice of $T = \left(\begin{smallmatrix} 1 & 1 \\ 1 & 1 \end{smallmatrix}\right) \in \mathcal{T}_1$, since μ_1 is strictly maximum among all $\mu_1, \ldots, \mu_s$. This is true for every (u, v), and it will be the actual leading term of the sum, provided the coefficient of $U_1^P W_1^Q = \mu_1^{2P+2Q}$ is nonzero.

Consider the diagonal entries where $u = v$. First notice that from (8.19), we have $R_{u,u} = R_{1,1}$ for all $u \in [h]$; second, the coefficient of the leading term $U_1^P W_1^Q$ is

$$X_{u,u,\left(\begin{smallmatrix} 1 & 1 \\ 1 & 1 \end{smallmatrix}\right)} = \left(\sum_{a \in [h]} |D_{1,a}|^2 \right)^2 = \|\mathbf{D}_{1,*}\|^4,$$

which is, again, independent of u. Without loss of generality, we may assume $\mathbf{D}_{1,*} \neq \mathbf{0}$; otherwise, we can remove all terms involving μ_1 in (8.19) and μ_2 will take its place, and the proof is completed by induction. (If all $\mathbf{D}_{i,*} = \mathbf{0}$, then the statement that $\mathbf{D}$ has rank at most one is trivial.)

Assuming that $\mathbf{D}_{1,*} \neq \mathbf{0}$, we have $R_{u,u} = R_{1,1} \neq 0$ for all $u \in [h]$ (and sufficiently large n). This is because, ignoring the positive factor Z^2, the coefficient $\|\mathbf{D}_{1,*}\|^4$ of the leading term $U_1^P W_1^Q$ is positive. By using Corollary 8.3, we have the following.

PROPERTY 8.14. *For all sufficiently large n, $|R_{1,1}| > 0$ and $|R_{u,v}| \in \{0, |R_{1,1}|\}$ for all $u, v \in [h]$.*

From now on, we focus on $u = 1$ and let $\boldsymbol{\mathcal{H}}_{*,v} = \mathbf{H}_{*,1} \circ \overline{\mathbf{H}_{*,v}}$. $\{\boldsymbol{\mathcal{H}}_{*,v}\}_{v \in [h]}$ forms an orthogonal basis with each $\|\boldsymbol{\mathcal{H}}_{*,v}\|^2 = h$. We also denote $X_{1,v,T}$ by $X_{v,T}$, so

$$X_{v,T} = \left(\sum_{a \in [h]} D_{b,a} \overline{D_{c,a}} \mathcal{H}_{a,v} \right) \left(\sum_{a \in [h]} D_{b',a} \overline{D_{c',a} \mathcal{H}_{a,v}} \right) \quad \text{for } T = \begin{pmatrix} b & c \\ b' & c' \end{pmatrix}. \tag{8.21}$$

We make three more definitions. Let $K = \{i \in [h] : D_{1,i} \neq 0\}$. By our assumption $K \neq \emptyset$. Let $A = \{v \in [h] : \text{ for all } i, j \in K, \mathcal{H}_{i,v} = \mathcal{H}_{j,v}\}$ and $B = [h] - A$. If $|K| = 1$, then $A = [h]$. The converse is also true, which follows from the fact that $\{\boldsymbol{\mathcal{H}}_{*,v}\}_{v\in[h]}$ forms an orthogonal basis. Also since $\boldsymbol{\mathcal{H}}_{*,1}$ is the all-one vector, $1 \in A$ and A is nonempty. Moreover, if $K = [h]$, then $A = \{1\}$. This again follows from the fact that $\{\boldsymbol{\mathcal{H}}_{*,v}\}$ forms an orthogonal basis.

Now we consider the coefficient $X_{v,T}$ of $U_1^P W_1^Q$ in $R_{1,v}$, where $T = \left(\begin{smallmatrix}1 & 1\\ 1 & 1\end{smallmatrix}\right)$. For every $v \in A$, it has norm $\|\mathbf{D}_{1,*}\|^4 > 0$. Then from Property 8.14 and Part B of the vanishing lemma the next property follows.

PROPERTY 8.15. *For any $v \in A$ and sufficiently large n, $|R_{1,v}| = |R_{1,1}|$.*

If $B \neq \emptyset$, then for any $v \in B$, the coefficient of $T = \left(\begin{smallmatrix}1 & 1\\ 1 & 1\end{smallmatrix}\right)$ in $R_{1,v}$ is

$$X_{v,T} = \left(\sum_{a\in K} |D_{1,a}|^2 \mathcal{H}_{a,v}\right)\left(\sum_{a\in K} |D_{1,a}|^2 \overline{\mathcal{H}_{a,v}}\right) = \left|\sum_{a\in K} |D_{1,a}|^2 \mathcal{H}_{a,v}\right|^2 \in \mathbb{R}.$$

Since we assumed $v \in B$, $\sum_{a\in K} |D_{1,a}|^2 \mathcal{H}_{a,v}$ is a sum of positive terms $|D_{1,a}|^2$ weighted by nonconstant $\mathcal{H}_{a,v}$, for $a \in K$, each with complex norm 1. Thus its absolute value must be strictly less than $\|\mathbf{D}_{1,*}\|^2$, which is only achieved when all $\mathcal{H}_{a,v}$, for $a \in K$, are equal to a constant. It follows that $X_{v,T} < \|\mathbf{D}_{1,*}\|^4$. Therefore, for $v \in B$ (and n sufficiently large), we have $|R_{1,v}| < |R_{1,1}|$. By using Property 8.14 and Part B of the vanishing lemma, we have the following property.

PROPERTY 8.16. *If $v \in B$, then for all sufficiently large n, $R_{1,v} = 0$ and thus,*

$$\sum_{T\in\mathcal{T}_i} X_{v,T} = 0 \quad \textit{for all } i \in [d].$$

In particular, by applying Property 8.16 to $\mathcal{T}_1 = \{\left(\begin{smallmatrix}1 & 1\\ 1 & 1\end{smallmatrix}\right)\}$, we have

$$\sum_{a\in K} |D_{1,a}|^2 \mathcal{H}_{a,v} = \sum_{a\in K} |D_{1,a}|^2 \overline{\mathcal{H}_{a,v}} = \langle |\mathbf{D}_{1,*}|^2, \boldsymbol{\mathcal{H}}_{*,v}\rangle = 0 \quad \text{for every } v \in B,$$

because $|D_{1,a}|$ is real. Here we use $|\mathbf{D}_{1,*}|^2$ to denote the vector $(|D_{1,1}|^2, |D_{1,2}|^2, \ldots)$. Furthermore, because $\{\boldsymbol{\mathcal{H}}_{*,v}\}$ forms an orthogonal basis, $|\mathbf{D}_{1,*}|^2$ must be expressible as a linear combination of $\{\boldsymbol{\mathcal{H}}_{*,v} : v \in A\}$ over $\mathbb{C}$. From such an expression, we have $|D_{1,i}|^2 = |D_{1,j}|^2$ for all $i, j \in K$, by the definition of K. Since $\mathbf{D}_{1,*}$ is only nonzero on K, $|D_{1,i}|$ is a constant on K and $D_{1,i} = 0$ for any $i \in [h] - K$. (The above proof does not actually assume $B \neq \emptyset$; if $B = \emptyset$, then $A = [h]$ and by $\{\boldsymbol{\mathcal{H}}_{*,v}\}$ being an orthogonal basis, $|K| = 1$. Then the above statement about $\mathbf{D}_{1,*}$ is still valid, namely, $\mathbf{D}_{1,*}$ has a unique nonzero entry and is zero elsewhere.) We summarize as follows.

PROPERTY 8.17. *$|\mathbf{D}_{1,*}|^2 \perp \boldsymbol{\mathcal{H}}_{*,v}$ for all $v \in B$. $|\mathbf{D}_{1,*}|^2$ is constant on K and 0 elsewhere. In particular, the vector χ_K, which is 1 on K and 0 elsewhere, is in the span of $\{\boldsymbol{\mathcal{H}}_{*,v} : v \in A\}$ and is orthogonal to all $\{\boldsymbol{\mathcal{H}}_{*,v} : v \in B\}$.*

Our next goal is to show that on K, $\mathbf{D}_{2,*}$ is a constant multiple of $\mathbf{D}_{1,*}$. Clearly if $B = \emptyset$, then we have $|K| = 1$ as noted above and thus it is trivially true that $\mathbf{D}_{2,*}$ is a constant multiple of $\mathbf{D}_{1,*}$ on K. So we assume $B \neq \emptyset$. We now consider

$$T_1 = \begin{pmatrix}2 & 1\\ 1 & 2\end{pmatrix} \quad \text{and} \quad T_2 = \begin{pmatrix}1 & 2\\ 2 & 1\end{pmatrix}.$$

T_1 and T_2 belong to the same $\mathcal{T}_g$ for some $g \in [d]$. By Property 8.16, $\sum_{T\in\mathcal{T}_g} X_{v,T} = 0$ for every $v \in B$. So we focus on terms $X_{v,T}$, where $T \in \mathcal{T}_g$ (i.e., $T \equiv_\mu T_1$). Suppose $T \equiv_\mu T_1$; then $\mu_b\mu_{b'} = \mu_1\mu_2$ and $\mu_c\mu_{c'} = \mu_1\mu_2$. Thus, $\{b, b'\} = \{c, c'\} = \{1, 2\}$, so

$$\mathcal{T}_g = \left\{ T_1, T_2, T_3 = \begin{pmatrix} 1 & 1 \\ 2 & 2 \end{pmatrix}, T_4 = \begin{pmatrix} 2 & 2 \\ 1 & 1 \end{pmatrix} \right\}.$$

However, due to the presence of a row (1 1), the sum

$$\sum_{a=1}^{h} |D_{1,a}|^2 \mathcal{H}_{a,v} = \sum_{a=1}^{h} |D_{1,a}|^2 \overline{\mathcal{H}_{a,v}} = 0$$

for any $v \in B$ as shown above. Therefore, the coefficients X_{v,T_3}, X_{v,T_4} corresponding to T_3 and T_4 are both 0.

We need one more definition: T is of a *conjugate-pair* form if it is of the form

$$T = \begin{pmatrix} b & c \\ c & b \end{pmatrix}.$$

For a matrix T in conjugate-pair form, the corresponding coefficient

$$X_{v,T} = \left| \sum_{a=1}^{h} D_{b,a} \overline{D_{c,a}} \mathcal{H}_{a,v} \right|^2 \geq 0.$$

The remaining two matrices T_1 and T_2 in $\mathcal{T}_g$ both have this form, so both X_{v,T_1} and X_{v,T_2} are nonnegative. Since $X_{v,T_1} + X_{v,T_2} = 0$, $X_{v,T_1} = X_{v,T_2} = 0$. This gives

$$\sum_{a \in [h]} \overline{D_{1,a}} D_{2,a} \overline{\mathcal{H}_{a,v}} = 0 \quad \text{for all } v \in B.$$

Hence $\overline{\mathbf{D}_{1,*}} \circ \mathbf{D}_{2,*} \perp \boldsymbol{\mathcal{H}}_{*,v}$ for all $v \in B$. It follows that $\overline{\mathbf{D}_{1,*}} \circ \mathbf{D}_{2,*}$ can be expressed as a linear combination of $\boldsymbol{\mathcal{H}}_{*,v}$ over $v \in A$. By the definition of A, this expression has a constant value on entries indexed by $a \in K$, where $|D_{1,a}|$ is a positive constant. Therefore, over K, $\mathbf{D}_{2,*}$ is a constant multiple of $\mathbf{D}_{1,*}$. This accomplishes our goal stated above, which we summarize as follows.

PROPERTY 8.18. *There exists some complex number λ, such that $D_{2,a} = \lambda D_{1,a}$, for all $a \in K$.*

Let $K_2 = \{i \in [h] : D_{2,i} \neq 0\}$. Note that the λ above could be 0, so it is possible that $K \not\subset K_2$. Our next goal is to show that for every $v \in A$, $\boldsymbol{\mathcal{H}}_{*,v}$ takes a constant value on K_2. This means that for all $v \in A$, $\mathcal{H}_{i,v} = \mathcal{H}_{j,v}$, for all $i, j \in K_2$. Without loss of generality, we assume $\mathbf{D}_{2,*} \neq \mathbf{0}$ since otherwise $K_2 = \emptyset$ and everything below regarding $\mathbf{D}_{2,*}$ and regarding $\boldsymbol{\mathcal{H}}_{*,v}$ on K_2 is trivially true.

To this end, we consider the matrices in $\mathcal{T}_g$ and their corresponding coefficients X_{v,T_i} for any $v \in A$. We will apply the more delicate Part A of the vanishing lemma on $R_{1,v}$ and $R_{1,1}$ for an arbitrary $v \in A$. Our target is to show that

$$\sum_{T \in \mathcal{T}_g} X_{v,T} = \sum_{T \in \mathcal{T}_g} X_{1,T} \quad \text{for any } v \in A. \tag{8.22}$$

By Property 8.15, $|R_{1,v}| = |R_{1,1}|$ for any sufficiently large n. To apply the vanishing lemma, we first show that terms that have a higher order of magnitude satisfy

$$\sum_{T \in \mathcal{T}_{g'}} X_{v,T} = \sum_{T \in \mathcal{T}_{g'}} X_{1,T} \quad \text{for all } 1 \leq g' < g \text{ and } v \in A. \tag{8.23}$$

We also need to show that

$$\operatorname{Im}\left(\sum_{T\in\mathcal{T}_g} X_{v,T}\right) = \operatorname{Im}\left(\sum_{T\in\mathcal{T}_g} X_{1,T}\right). \tag{8.24}$$

By definition, every $T \geq_\mu T_1$ satisfies $\mu_b\mu_{b'} \geq \mu_1\mu_2$. Thus, the first column of T is either $(1\ 1)^{\mathrm{T}}$, $(1\ 2)^{\mathrm{T}}$, or $(2\ 1)^{\mathrm{T}}$.

First, consider those matrices $T \geq_\mu T_1$ where each row of T has at least one 1. For every $v \in A$, the two inner product factors in (8.21), namely,

$$\sum_{a=1}^{h} D_{b,a}\overline{D_{c,a}}\mathcal{H}_{a,v} \quad \text{and} \quad \sum_{a=1}^{h} D_{b',a}\overline{D_{c',a}}\,\overline{\mathcal{H}_{a,v}}$$

must be actually a sum over $a \in K$, since $\mathbf{D}_{1,*}$ is zero elsewhere. But for $a \in K$, $\mathcal{H}_{a,v}$ is just a constant α_v of norm 1 (a root of unity), independent of $a \in K$. Thus

$$\sum_{a=1}^{h} D_{b,a}\overline{D_{c,a}}\mathcal{H}_{a,v} = \alpha_v \sum_{a\in K} D_{b,a}\overline{D_{c,a}} \quad \text{and} \quad \sum_{a=1}^{h} D_{b',a}\overline{D_{c',a}}\,\overline{\mathcal{H}_{a,v}} = \overline{\alpha_v} \sum_{a\in K} D_{b',a}\overline{D_{c',a}}.$$

Since $\alpha_v\overline{\alpha_v} = |\alpha_v|^2 = 1$, it follows that their product is

$$\left(\sum_{a=1}^{h} D_{b,a}\overline{D_{c,a}}\mathcal{H}_{a,v}\right)\left(\sum_{a=1}^{h} D_{b',a}\overline{D_{c',a}\mathcal{H}_{a,v}}\right) = \left(\sum_{a\in K} D_{b,a}\overline{D_{c,a}}\right)\left(\sum_{a\in K} D_{b',a}\overline{D_{c',a}}\right),$$

which is the same as the coefficient $X_{1,T}$ corresponding to T for $v_0 = 1 \in A$. So for all such T, their contributions to $R_{1,v}$ and to $R_{1,1}$ are the same for any $v \in A$.

Such $T \geq_\mu T_1$ with at least one 1 in each row include any matrix of the form

$$\begin{pmatrix} 1 & c \\ 1 & c' \end{pmatrix},\ \begin{pmatrix} 1 & 1 \\ 2 & 1 \end{pmatrix},\ \text{or}\ \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}.$$

These exhaust all $T >_\mu T_1$, and (8.23) follows.

Such $T \geq_\mu T_1$ also include T_1 and T_2 in $\mathcal{T}_g$. So $X_{v,T_1} = X_{1,T_1}$ and $X_{v,T_2} = X_{1,T_2}$ for any $v \in A$. Now we deal with matrices T_3 and T_4. We note that the sum of X_{v,T_3} and X_{v,T_4}, at any v, is

$$\left(\sum_{a\in K} |D_{1,a}|^2\mathcal{H}_{a,v}\right)\left(\sum_{a=1}^{h} |D_{2,a}|^2\overline{\mathcal{H}_{a,v}}\right) + \left(\sum_{a=1}^{h} |D_{2,a}|^2\mathcal{H}_{a,v}\right)\left(\sum_{a\in K} |D_{1,a}|^2\overline{\mathcal{H}_{a,v}}\right), \tag{8.25}$$

which is a real number. Equation (8.24) then follows.

Now we can apply Part A of the vanishing lemma, which gives us (8.22). Since $X_{v,T_1} = X_{1,T_1}$ and $X_{v,T_2} = X_{1,T_2}$, we have

$$X_{v,T_3} + X_{v,T_4} = X_{1,T_3} + X_{1,T_4} = 2 \cdot \|\mathbf{D}_{1,*}\|^2\|\mathbf{D}_{2,*}\|^2.$$

However, this is clearly the maximum possible value of (8.25). (By our assumption, $\|\mathbf{D}_{1,*}\|^2\|\mathbf{D}_{2,*}\|^2 > 0$.) The only way the sum in (8.25) can achieve this maximum at $v \in A$ is for $\mathcal{H}_{a,v}$ to take a constant value β_v for all $a \in K_2$, and $\mathcal{H}_{a,v}$ to take a constant value α_v for all $a \in K$, for some pair of complex numbers α_v and β_v of norm 1. Moreover, by (8.25) we have $\alpha_v\overline{\beta_v} + \overline{\alpha_v}\beta_v = 2$. It follows that $\alpha_v = \beta_v$. Therefore, $\mathcal{H}_{a,v}$ is constant on $a \in K \cup K_2$ for each $v \in A$. We summarize it as follows.

PROPERTY 8.19. *For every $v \in A$, there exists a complex number α_v of norm 1 such that $\mathcal{H}_{a,v} = \alpha_v$ for all a in $K \cup K_2$.*

We eventually want to prove $K_2 = K$. Our next goal is to prove that $|\mathbf{D}_{2,*}|^2 \perp \boldsymbol{\mathcal{H}}_{*,v}$ for all $v \in B$. Of course if $B = \emptyset$, then this is vacuously true. We assume $B \neq \emptyset$.

For this purpose we will examine

$$T^* = \begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix}$$

and the class $\mathcal{T}_g$ it belongs to. By Property 8.16, we have

$$\sum_{T \in \mathcal{T}_g} X_{v,T} = 0 \quad \text{for any } v \in B.$$

Thus we will examine $T \in \mathcal{T}_g$, namely, $\mu_b \mu_{b'} = \mu_c \mu_{c'} = \mu_2^2$.

Now there might be some other pair $(b, b') \neq (2, 2)$ such that $\mu_b \mu_{b'} = \mu_2^2$. If such a pair exists, it is essentially unique and is of the form $(1, s)$ or $(s, 1)$, where $s > 2$. Then $\mathcal{T}_g$ consists of precisely the following matrices, namely, each column must be either $(2\ 2)^{\mathrm{T}}$, $(s\ 1)^{\mathrm{T}}$, or $(1\ s)^{\mathrm{T}}$. Let's examine such a matrix T in more detail. Suppose $T \in \mathcal{T}_g$ has a row that is either (1 1) or (1 2) or (2 1). Then,

$$X_{v,T} = \left(\sum_{a=1}^{h} D_{b,a} \overline{D_{c,a}} \mathcal{H}_{a,v} \right) \left(\sum_{a=1}^{h} D_{b',a} \overline{D_{c',a}}\, \overline{\mathcal{H}_{a,v}} \right) = 0 \quad \text{for any } v \in B.$$

This is because of the following: The presence of $\mathbf{D}_{1,*}$ restricts the sum to $a \in K$. By Property 8.17, we know that for every $v \in B$, $|\mathbf{D}_{1,*}|^2 \perp \boldsymbol{\mathcal{H}}_{*,v}$. Moreover, on set K, we know from Property 8.18 that both vectors $\overline{\mathbf{D}_{1,*}} \circ \mathbf{D}_{2,*}$ and $\mathbf{D}_{1,*} \circ \overline{\mathbf{D}_{2,*}}$ can be replaced by a constant multiple of the vector $|\mathbf{D}_{1,*}|^2$ (the constant could be 0) and thus also perpendicular to $\boldsymbol{\mathcal{H}}_{*,v}$ (and to $\overline{\boldsymbol{\mathcal{H}}_{*,v}}$).

Now suppose T is a matrix in $\mathcal{T}_g$, and yet it does not have a row which is either (1 1) or (1 2) or (2 1). It is easy to check that the only cases are

$$T^* = \begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix}, \quad T_1 = \begin{pmatrix} 1 & s \\ s & 1 \end{pmatrix}, \quad \text{and} \quad T_2 = \begin{pmatrix} s & 1 \\ 1 & s \end{pmatrix}.$$

Thus, $X_{v,T^*} + X_{v,T_1} + X_{v,T_2} = 0$ for all $v \in B$. However, as noted above, all three matrices T^*, T_1, and T_2 have the conjugate-pair form, so their contributions

$$\left| \sum_{a=1}^{h} D_{2,a} \overline{D_{2,a}} \mathcal{H}_{a,v} \right|^2, \quad \left| \sum_{a=1}^{h} D_{1,a} \overline{D_{s,a}} \mathcal{H}_{a,v} \right|^2, \quad \text{and} \quad \left| \sum_{a=1}^{h} D_{s,a} \overline{D_{1,a}} \mathcal{H}_{a,v} \right|^2$$

are all nonnegative. It follows that all three sums are zero. In particular, from X_{v,T^*} we get $|\mathbf{D}_{2,*}|^2 \perp \boldsymbol{\mathcal{H}}_{*,v}$ for all $v \in B$.

It follows that the vector $|\mathbf{D}_{2,*}|^2$ is in the span of $\{\boldsymbol{\mathcal{H}}_{*,v} : v \in A\}$. This linear combination produces a constant value at any entry $|D_{2,a}|^2$ for $a \in K \cup K_2$. This is because each vector $\boldsymbol{\mathcal{H}}_{*,v}$ for $v \in A$ has this property by Property 8.19.

As we assumed $\mathbf{D}_{2,*} \neq 0$, and $\mathbf{D}_{2,*}$ is 0 outside of K_2 (by the definition of K_2), this constant value produced at each entry $|D_{2,a}|^2$ for $a \in K \cup K_2$ must be nonzero. In particular, $D_{2,a} \neq 0$ at $a \in K$. It follows that $K \subseteq K_2$. It also implies that the vector, which is 1 on $K \cup K_2 = K_2$ and 0 elsewhere, is in the span of $\{\boldsymbol{\mathcal{H}}_{*,v} : v \in A\}$.

Next we prove that $K = K_2$, by showing that $|K| = |K_2|$ (since we already know $K \subseteq K_2$). Let χ_K denote the h-dimensional characteristic vector for K, which is 1 for any index $a \in K$ and 0 elsewhere. Similarly, we denote by χ_{K_2} the characteristic vector for K_2. Both vectors χ_K and χ_{K_2} are in the linear span of $\{\boldsymbol{\mathcal{H}}_{*,v} : v \in A\}$. Write $\chi_K = \sum_{v \in A} x_v \boldsymbol{\mathcal{H}}_{*,v}$, where $x_v \in \mathbb{C}$; then

$$x_v \|\boldsymbol{\mathcal{H}}_{*,v}\|^2 = \langle \chi_K, \boldsymbol{\mathcal{H}}_{*,v} \rangle = \sum_{a=1}^{h} \chi_K(a) \overline{\mathcal{H}_{a,v}} = \sum_{a \in K} \overline{\mathcal{H}_{a,v}} = |K| \overline{\alpha_v}$$

by Property 8.19. It follows that $|x_v| h = |K|$ for each $v \in A$. Thus

$$|K| = \|\chi_K\|^2 = \sum_{v \in A} |x_v|^2 \cdot \|\boldsymbol{\mathcal{H}}_{*,v}\|^2 = |A| \left(\frac{|K|}{h}\right)^2 h = \frac{|A||K|^2}{h},$$

and it follows that $|K| = h/|A|$. Exactly the same argument gives $|K_2| = h/|A|$. Hence $|K| = |K_2|$ and $K = K_2$. At this point the statement in Property 8.18 can be strengthened to the following.

PROPERTY 8.20. *There exists some complex number λ such that $\mathbf{D}_{2,*} = \lambda \mathbf{D}_{1,*}$.*

Our final goal is to generalize this proof to all $\mathbf{D}_{\ell,*}$ for $\ell = 1, 2, \ldots, s$. We prove this by induction.

> Inductive hypothesis: For some $\ell \geq 2$, the $(\ell - 1)$ rows $\mathbf{D}_{1,*}, \ldots, \mathbf{D}_{\ell-1,*}$ satisfy that $\mathbf{D}_{i,*} = \lambda_i \cdot \mathbf{D}_{1,*}$ for some λ_i and $1 \leq i < \ell$.

The proof mainly follow that of the case $\ell = 2$ above, except for one crucial argument at the end. We presented the special case $\ell = 2$ alone for ease of understanding.

We prove that $\mathbf{D}_{\ell,*} = \lambda_\ell \cdot \mathbf{D}_{1,*}$ for some λ_ℓ. Clearly we may assume $\mathbf{D}_{\ell,*} \neq \mathbf{0}$, for otherwise the inductive step is trivial. To start, consider the matrices

$$T_1 = \begin{pmatrix} \ell & 1 \\ 1 & \ell \end{pmatrix} \quad \text{and} \quad T_2 = \begin{pmatrix} 1 & \ell \\ \ell & 1 \end{pmatrix}$$

and the corresponding class $\mathcal{T}_g$ they belong to. By Property 8.16, we have for every $v \in B$, $\sum_{T \in \mathcal{T}_g} X_{v,T} = 0$. We only need to examine those $T \in \mathcal{T}_g$ with exactly the same order as that of T_1, T_2: $\mu_b \mu_{b'} = \mu_c \mu_{c'} = \mu_1 \mu_\ell$. To satisfy this condition, both columns of T must have entries $\{1, \ell\}$ or have both entries $< \ell$. No entry in $\{b, b', c, c'\}$ can be $> \ell$. There are two cases now: Case 1—There is a row $(b\ c)$ or $(b'\ c')$ (or both) which has both entries $< \ell$; Case 2—Both rows have an entry $= \ell$.

In Case 1, at least one of the inner product sums in the product

$$X_{v,T} = \left(\sum_{a=1}^{h} D_{b,a} \overline{D_{c,a}} \mathcal{H}_{a,v} \right) \left(\sum_{a=1}^{h} D_{b',a} \overline{D_{c',a}}\, \overline{\mathcal{H}_{a,v}} \right)$$

takes place over $a \in K$. This follows from the inductive hypothesis. In fact that inner product is a constant multiple of $\sum_{a \in K} |D_{1,a}|^2 \mathcal{H}_{a,v}$ or its conjugate $\sum_{a \in K} |D_{1,a}|^2 \overline{\mathcal{H}_{a,v}}$ which are 0 according to Property 8.17 for all $v \in B$.

In Case 2, it is easy to check that to have the same order $\mu_1 \mu_\ell$, T can only be T_1 or T_2. Now observe that both T_1 and T_2 have the conjugate-pair form. Thus, their contributions X_{v,T_1} and X_{v,T_2} are both nonnegative. Since $X_{v,T_1} + X_{v,T_2} = 0$, both of them have to vanish:

$$\sum_{a \in [h]} \overline{D_{1,a}} D_{\ell,a} \overline{\mathcal{H}_{a,v}} = 0 \quad \text{and} \quad \sum_{a \in [h]} D_{1,a} \overline{D_{\ell,a}}\, \overline{\mathcal{H}_{a,v}} = 0 \quad \text{for all } v \in B.$$

Hence $\overline{\mathbf{D}_{1,*}} \circ \mathbf{D}_{\ell,*} \perp \boldsymbol{\mathcal{H}}_{*,v}$ for all $v \in B$. It follows that the vector $\overline{\mathbf{D}_{1,*}} \circ \mathbf{D}_{\ell,*}$ belongs to the linear span of $\{\boldsymbol{\mathcal{H}}_{*,v} : v \in A\}$. From the definition of A, this expression has a constant value on entries indexed by $a \in K$. Therefore, on K, $\mathbf{D}_{\ell,*}$ is a constant multiple of $\mathbf{D}_{1,*}$. We summarize this as follows.

PROPERTY 8.21. *There exists some complex number λ_ℓ such that $D_{\ell,a} = \lambda_\ell D_{1,a}$ for all $a \in K$.*

Let $K_\ell = \{i \in [r] : D_{\ell,i} \neq 0\}$. Next, we prove that for every $v \in A$, $\boldsymbol{\mathcal{H}}_{*,v}$ takes a constant value on K_ℓ, i.e., $\mathcal{H}_{i,v} = \mathcal{H}_{j,v}$, for all indices $i, j \in K_\ell$. We had assumed $\mathbf{D}_{\ell,*} \neq 0$, since otherwise the induction is completed for ℓ. Then $K_\ell \neq \emptyset$.

To show that $\boldsymbol{\mathcal{H}}_{*,v}$ is a constant on K_ℓ, we consider

$$T_3 = \begin{pmatrix} \ell & \ell \\ 1 & 1 \end{pmatrix} \quad \text{and} \quad T_4 = \begin{pmatrix} 1 & 1 \\ \ell & \ell \end{pmatrix}$$

and the class $\mathcal{T}_g$ they belong to. We want to apply Part A of the vanishing lemma to show that

$$\sum_{T \in \mathcal{T}_g} X_{v,T} = \sum_{T \in \mathcal{T}_g} X_{1,T} \quad \text{for any } v \in A. \tag{8.26}$$

For this purpose, we need to compare the respective terms of the sum (8.20) for an arbitrary $v \in A$ and for the particular $v_0 = 1 \in A$. More exactly, we will show that

$$\sum_{T \in \mathcal{T}_{g'}} X_{v,T} = \sum_{T \in \mathcal{T}_{g'}} X_{1,T} \quad \text{and} \quad \operatorname{Im}\left(\sum_{T \in \mathcal{T}_g} X_{v,T}\right) = \operatorname{Im}\left(\sum_{T \in \mathcal{T}_g} X_{1,T}\right) \tag{8.27}$$

for all $v \in A$ and $g' < g$. Then (8.26) follows from Part A of the vanishing lemma.

To this end, we first consider matrices T that have an order of magnitude strictly larger than that of T_3 and T_4. We have either $\mu_b \mu_{b'} > \mu_1 \mu_\ell$ or $\mu_b \mu_{b'} = \mu_1 \mu_\ell$ and $\mu_c \mu_{c'} > \mu_1 \mu_\ell$. The first alternative implies $b, b' < \ell$. The second implies $c, c' < \ell$.

In both cases, each row of T has at least one entry $< \ell$. By the inductive hypothesis, both inner products in (8.21), namely,

$$\sum_{a=1}^{h} D_{b,a} \overline{D_{c,a}} \mathcal{H}_{a,v} \quad \text{and} \quad \sum_{a=1}^{h} D_{b',a} \overline{D_{c',a}}\, \overline{\mathcal{H}_{a,v}}$$

must be a sum over K since $\mathbf{D}_{1,*}$ is zero elsewhere. However, for any $a \in K$, $\mathcal{H}_{a,v}$ is a constant α_v of norm 1 (a root of unity), independent of $a \in K$. Thus

$$\sum_{a \in [h]} D_{b,a} \overline{D_{c,a}} \mathcal{H}_{a,v} = \alpha_v \sum_{a \in K} D_{b,a} \overline{D_{c,a}} \quad \text{and} \quad \sum_{a \in [h]} D_{b',a} \overline{D_{c',a}}\, \overline{\mathcal{H}_{a,v}} = \overline{\alpha_v} \sum_{a \in K} D_{b',a} \overline{D_{c',a}}.$$

Since $\alpha_v \overline{\alpha_v} = |\alpha_v|^2 = 1$, it follows that their product is

$$X_{v,T} = \left(\sum_{a \in K} D_{b,a} \overline{D_{c,a}}\right)\left(\sum_{a \in K} D_{b',a} \overline{D_{c',a}}\right),$$

which is exactly the same as the coefficient $X_{1,T}$ for $v_0 = 1 \in A$. Thus for any T, where each row has at least one entry $< \ell$, $X_{v,T} = X_{1,T}$, for any $v \in A$. This includes *all* matrices $T >_\mu T_3$ (as well as some matrices $T \equiv_\mu T_3 \in \mathcal{T}_g$), and the first part of (8.27) follows.

Now we consider any matrix $T \in \mathcal{T}_g$. If each row of T has at least one entry $< \ell$, then by the proof above, we know $X_{v,T} = X_{1,T}$ for any $v \in A$. Suppose $T \in \mathcal{T}_g$ does not have this property. Then each column of such a matrix must consist of $\{1, \ell\}$. We have four such matrices: T_1, T_2, T_3, and T_4. But the former two matrices already belong to the case covered above. So we have

$$\sum_{T \in \mathcal{T}_g} X_{v,T} - \sum_{T \in \mathcal{T}_g} X_{1,T} = X_{v,T_3} + X_{v,T_4} - (X_{1,T_3} + X_{1,T_4}) \quad \text{for any } v \in A.$$

Now to the matrices T_3, T_4 themselves. We note that the sum of their coefficients $X_{v,T_3} + X_{v,T_4}$, at any $v \in A$, is

(8.28)
$$\left(\sum_{a \in K} |D_{1,a}|^2 \mathcal{H}_{a,v}\right)\left(\sum_{a=1}^{h} |D_{\ell,a}|^2 \overline{\mathcal{H}_{a,v}}\right) + \left(\sum_{a=1}^{h} |D_{\ell,a}|^2 \mathcal{H}_{a,v}\right)\left(\sum_{a \in K} |D_{1,a}|^2 \overline{\mathcal{H}_{a,v}}\right).$$

This is a real number, and the second part of (8.27) follows.

Now we can apply Part A of the vanishing lemma to conclude that

$$X_{v,T_3} + X_{v,T_4} = X_{1,T_3} + X_{1,T_4} = 2 \cdot \|\mathbf{D}_{1,*}\|^2 \|\mathbf{D}_{\ell,*}\|^2 \quad \text{for any } v \in A.$$

This is the maximum possible value of (8.28). By assumption, $\|\mathbf{D}_{1,*}\|^2 \|\mathbf{D}_{\ell,*}\|^2 > 0$. The only way the sum in (8.28) achieves this maximum at $v \in A$ is for $\mathcal{H}_{a,v}$ to take a constant value γ_v for all $a \in K_\ell$ (and we already know that $\mathcal{H}_{a,v}$ takes a constant value α_v for all $a \in K$), where α_v and γ_v are of norm 1. Moreover, by (8.28), we have $\alpha_v \overline{\gamma_v} + \overline{\alpha_v} \gamma_v = 2$. It follows that $\alpha_v = \gamma_v$. Thus $\boldsymbol{\mathcal{H}}_{*,v}$ is a constant on $K \cup K_\ell$ for each $v \in A$. We summarize it as the next property.

PROPERTY 8.22. *For every $v \in A$, there exists a complex number α_v of norm 1 such that $\mathcal{H}_{v,a} = \alpha_v$ for all $a \in K \cup K_\ell$.*

Our next goal is to prove that $|\mathbf{D}_{\ell,*}|^2 \perp \boldsymbol{\mathcal{H}}_{*,v}$ for all $v \in B$. Of course, if $B = \emptyset$, then this is trivially true. We assume $B \neq \emptyset$. For this purpose, we examine T^*, the matrix with all four entries being ℓ, and the class $\mathcal{T}_g$ it belongs to. By Property 8.16, we have $\sum_{T \in \mathcal{T}_g} X_{v,T} = 0$ for any $v \in B$, and our target is to show that $X_{v,T^*} = 0$. To prove this, we need to examine terms $X_{v,T}$ for all $T \equiv_\mu T^* \in \mathcal{T}_g$.

It is now possible to have a number of pairs, $(a_1, b_1), (a_2, b_2), \ldots, (a_k, b_k)$, for some $k \geq 0$, such that $\mu_{a_i} \mu_{b_i} = \mu_\ell^2$ for $1 \leq i \leq k$. (When $\ell = 2$, such a pair, if it exists, is essentially unique, but for $\ell > 2$ there could be many such pairs. This is a complication for $\ell > 2$.) Every matrix $T \in \mathcal{T}_g$ must have each column chosen from either $(\ell\ \ell)^{\mathrm{T}}$ or one of the pairs $(a_i\ b_i)^{\mathrm{T}}$ or $(b_i\ a_i)^{\mathrm{T}}$. Note that if such pairs do not exist, i.e., $k = 0$, then $\mathcal{T}_g = \{T^*\}$ and we have

$$X_{v,T^*} = \left(\sum_{a=1}^{h} |D_{\ell,a}|^2 \mathcal{H}_{a,v}\right)\left(\sum_{a=1}^{h} |D_{\ell,a}|^2 \overline{\mathcal{H}_{a,v}}\right) = 0 \quad \text{at any } v \in B.$$

The following proof is to show that even when such pairs exist ($k \geq 1$), we still have $X_{v,T^*} = 0$. For this purpose, we show that $\sum_{T \in \mathcal{T}_g, T \neq T^*} X_{v,T} \geq 0$.

Suppose $k \geq 1$. We may assume $a_i < \ell < b_i$ for all $i \in [k]$. We examine all the $T \in \mathcal{T}_g$ other than T^*. If T has at least one row, say, $(b\ \ c)$, with $\max\{b, c\} \leq \ell$ and $\min\{b, c\} < \ell$, then by the inductive hypothesis and Property 8.21, the corresponding inner product actually takes place over K. In fact, the inner product is a constant

multiple of the projection of $|\mathbf{D}_{1,*}|^2$ on either $\boldsymbol{\mathcal{H}}_{*,v}$ or $\overline{\boldsymbol{\mathcal{H}}_{*,v}}$. But we already know that this projection is zero for all $v \in B$.

For the remaining T where both rows satisfy $[\max\{b,c\} > \ell$ or $\min\{b,c\} \geq \ell]$, if T is not T^*, then one of its two columns is not $(\ell\ \ell)^{\mathrm{T}}$, and one entry of this column is $a_i < \ell$ for some $i \in [k]$. It follows that the other entry in the same row as a_i must be $b_j > \ell$ for some $j \in [k]$. As a result, the only matrices remaining are of two types:

$$\begin{pmatrix} a_i & b_j \\ b_i & a_j \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} b_i & a_j \\ a_i & b_j \end{pmatrix} \quad \text{for some } 1 \leq i,j \leq k.$$

We consider the first type. The total contribution of these matrices is

$$\begin{aligned}
&\sum_{i,j=1}^{k} \left(\sum_{a=1}^{h} D_{a_i,a} \overline{D_{b_j,a}} \mathcal{H}_{a,v} \right) \left(\sum_{a'=1}^{h} D_{b_i,a'} \overline{D_{a_j,a'}}\, \overline{\mathcal{H}_{a',v}} \right) \\
&= \sum_{i,j=1}^{k} \left(\sum_{a=1}^{h} \lambda_{a_i} D_{1,a} \overline{D_{b_j,a}} \mathcal{H}_{a,v} \right) \left(\sum_{a'=1}^{h} D_{b_i,a'} \overline{\lambda_{a_j}}\, \overline{D_{1,a'}}\, \overline{\mathcal{H}_{a',v}} \right) \\
&= \sum_{i,j=1}^{k} \sum_{a,a'=1}^{h} \overline{\lambda_{a_j}} D_{1,a} \overline{D_{b_j,a}} \mathcal{H}_{a,v} \cdot \lambda_{a_i} D_{b_i,a'} \overline{D_{1,a'}}\, \overline{\mathcal{H}_{a',v}} \\
&= \left[\sum_{a=1}^{h} D_{1,a} \mathcal{H}_{a,v} \left(\sum_{j=1}^{k} \overline{\lambda_{a_j}}\, \overline{D_{b_j,a}} \right) \right] \cdot \left[\sum_{a'=1}^{h} \overline{D_{1,a'}}\, \overline{\mathcal{H}_{a',v}} \left(\sum_{i=1}^{k} \lambda_{a_i} D_{b_i,a'} \right) \right] \\
&= \left| \sum_{a=1}^{h} D_{1,a} \mathcal{H}_{a,v} \left(\sum_{j=1}^{k} \overline{\lambda_{a_j}}\, \overline{D_{b_j,a}} \right) \right|^2 \geq 0.
\end{aligned}$$

Here in the first equality we used the inductive hypothesis for $a_i, a_j < \ell$.

The argument for the second type of matrices is symmetric. Note also that T^* has the conjugate-pair form, and therefore its contribution X_{v,T^*} at any $v \in B$ is nonnegative. It follows from $\sum_{T \in \mathcal{T}_g} X_{v,T} = 0$ (Property 8.16) that $X_{v,T^*} = 0$ and

$$\left| \sum_{a=1}^{h} |D_{\ell,a}|^2 \overline{\mathcal{H}_{a,v}} \right|^2 = 0 \quad \text{for all } v \in B.$$

This means that $|\mathbf{D}_{\ell,*}|^2 \perp \boldsymbol{\mathcal{H}}_{*,v}$ for all $v \in B$ and thus $|\mathbf{D}_{\ell,*}|^2$ is in the linear span of $\{\boldsymbol{\mathcal{H}}_{*,v} : v \in A\}$. Then by the same argument used for $\ell = 2$, we obtain $K = K_\ell$, and summarize as follows.

PROPERTY 8.23. *There exists a complex number λ_ℓ such that $\mathbf{D}_{\ell,*} = \lambda_\ell \mathbf{D}_{1,*}$.*

This completes the proof by induction that $\mathbf{D}$ has rank at most one.

8.5. Step 2.4. After Step 2.3, we obtain a pair $(\mathbf{C}, \mathfrak{D})$ that satisfies conditions (*Shape*$_1$)–(*Shape*$_6$). By (*Shape*$_2$), we have

$$\mathbf{C} = \begin{pmatrix} \mathbf{0} & \mathbf{F} \\ \mathbf{F}^T & \mathbf{0} \end{pmatrix} = \begin{pmatrix} \mathbf{0} & \mathbf{M} \otimes \mathbf{H} \\ (\mathbf{M} \otimes \mathbf{H})^T & \mathbf{0} \end{pmatrix},$$

where $\mathbf{M}$ is an $s \times t$ matrix of rank 1, $M_{i,j} = \mu_i \nu_j$, and $\mathbf{H}$ is the $h \times h$ matrix defined in (*Shape*$_2$). By (*Shape*$_5$) and (*Shape*$_6$), we have for every $r \in [0 : N-1]$

$$\mathbf{D}^{[r]} = \begin{pmatrix} \mathbf{D}^{[r]}_{(0,*)} & \\ & \mathbf{D}^{[r]}_{(1,*)} \end{pmatrix} = \begin{pmatrix} \mathbf{K}^{[r]}_{(0,*)} \otimes \mathbf{L}^{[r]}_{(0,*)} & \\ & \mathbf{K}^{[r]}_{(1,*)} \otimes \mathbf{L}^{[r]}_{(1,*)} \end{pmatrix}.$$

Every entry in $\mathbf{L}^{[r]}$ either is 0 or has norm 1 and $\mathbf{L}^{[0]}$ is the $2h \times 2h$ identity matrix.

Using these matrices, we define two new pairs $(\mathbf{C}', \mathfrak{K})$ and $(\mathbf{C}'', \mathfrak{L})$, which give rise to two problems, $\mathsf{EVAL}(\mathbf{C}', \mathfrak{K})$ and $\mathsf{EVAL}(\mathbf{C}'', \mathfrak{L})$. First, $\mathbf{C}'$ is the bipartization of $\mathbf{M}$, so it is $(s+t) \times (s+t)$, and $\mathfrak{K}$ is a sequence of N diagonal matrices also of this size: $(\mathbf{K}^{[0]}, \ldots, \mathbf{K}^{[N-1]})$. Second, $\mathbf{C}''$ is the bipartization of $\mathbf{H}$, and it is $2h \times 2h$, and $\mathfrak{L}$ is the sequence of N diagonal matrices: $(\mathbf{L}^{[0]}, \ldots, \mathbf{L}^{[N-1]})$. The following lemma shows that $\mathsf{EVAL}(\mathbf{C}, \mathfrak{D})$ has the same complexity as $\mathsf{EVAL}(\mathbf{C}'', \mathfrak{L})$.

LEMMA 8.24. $\mathsf{EVAL}(\mathbf{C}, \mathfrak{D}) \equiv \mathsf{EVAL}(\mathbf{C}'', \mathfrak{L})$.

Proof. Let G be a connected undirected graph and let u^* be one of its vertices. Then by Lemmas 2.3 and 2.4, we have

$$\begin{aligned} Z_{\mathbf{C},\mathfrak{D}}(G) &= Z^{\rightarrow}_{\mathbf{C},\mathfrak{D}}(G, u^*) + Z^{\leftarrow}_{\mathbf{C},\mathfrak{D}}(G, u^*), \\ Z^{\rightarrow}_{\mathbf{C},\mathfrak{D}}(G, u^*) &= Z^{\rightarrow}_{\mathbf{C}',\mathfrak{K}}(G, u^*) \cdot Z^{\rightarrow}_{\mathbf{C}'',\mathfrak{L}}(G, u^*), \quad \text{and} \\ Z^{\leftarrow}_{\mathbf{C},\mathfrak{D}}(G, u^*) &= Z^{\leftarrow}_{\mathbf{C}',\mathfrak{K}}(G, u^*) \cdot Z^{\leftarrow}_{\mathbf{C}'',\mathfrak{L}}(G, u^*). \end{aligned}$$

As $\mathbf{M}$ is of rank 1, both $Z^{\rightarrow}_{\mathbf{C}',\mathfrak{K}}$ and $Z^{\leftarrow}_{\mathbf{C}',\mathfrak{K}}$ can be computed in polynomial time. We only prove for $Z^{\rightarrow}_{\mathbf{C}',\mathfrak{K}}$ here. If G is not bipartite, $Z^{\rightarrow}_{\mathbf{C}',\mathfrak{K}}(G, u^*)$ is trivially 0; otherwise let $U \cup V$ be the vertex set of G, $u^* \in U$, and every edge $uv \in E$ has one vertex u from U and one vertex v from V. Let Ξ denote the set of assignments ξ which map U to $[s]$ and V to $[t]$. Then (note that we use $\mathbf{K}^{[r]}$ to denote $\mathbf{K}^{[r \bmod N]}$ for any $r \geq N$)

$$\begin{aligned} Z^{\rightarrow}_{\mathbf{C}',\mathfrak{K}}(G, u^*) &= \sum_{\xi \in \Xi} \left(\prod_{uv \in E} \mu_{\xi(u)} \cdot \nu_{\xi(v)} \right) \left(\prod_{u \in U} K^{[\deg(u)]}_{(0,\xi(u))} \right) \left(\prod_{v \in V} K^{[\deg(v)]}_{(1,\xi(v))} \right) \\ &= \prod_{u \in U} \left(\sum_{i \in [s]} (\mu_i)^{\deg(u)} \cdot K^{[\deg(u)]}_{(0,i)} \right) \times \prod_{v \in V} \left(\sum_{j \in [t]} (\nu_j)^{\deg(v)} \cdot K^{[\deg(v)]}_{(1,j)} \right), \end{aligned}$$

which can be computed in polynomial time.

Moreover, since $(\mathbf{C}'', \mathfrak{L})$ satisfies (*Pinning*), by the second pinning lemma (Lemma 4.3), the problem of computing $Z^{\rightarrow}_{\mathbf{C}'',\mathfrak{L}}$ and $Z^{\leftarrow}_{\mathbf{C}'',\mathfrak{L}}$ is reducible to $\mathsf{EVAL}(\mathbf{C}'', \mathfrak{L})$. It then follows that $\mathsf{EVAL}(\mathbf{C}, \mathfrak{D}) \leq \mathsf{EVAL}(\mathbf{C}'', \mathfrak{L})$.

We next prove the reverse direction. First note that by the third pinning lemma (Corollary 8.4), computing $Z^{\rightarrow}_{\mathbf{C},\mathfrak{D}}$ and $Z^{\leftarrow}_{\mathbf{C},\mathfrak{D}}$ is reducible to $\mathsf{EVAL}(\mathbf{C}, \mathfrak{D})$. However, this does not finish the proof because $Z^{\rightarrow}_{\mathbf{C}',\mathfrak{K}}$ (or $Z^{\leftarrow}_{\mathbf{C}',\mathfrak{K}}$) could be 0 at (G, u^*). To deal with this case, we prove the following claim.

CLAIM 8.25. *Given a connected, bipartite $G = (U \cup V, E)$ and vertex $u^* \in U$, either we can construct a new connected, bipartite $G' = (U' \cup V', E')$ in polynomial time such that $u^* \in U \subset U'$,*

$$Z^{\rightarrow}_{\mathbf{C}'',\mathfrak{L}}(G', u^*) = h^{|U \cup V|} \cdot Z^{\rightarrow}_{\mathbf{C}'',\mathfrak{L}}(G, u^*), \tag{8.29}$$

and $Z^{\rightarrow}_{\mathbf{C}',\mathfrak{K}}(G', u^*) \neq 0$, or we can show that $Z^{\rightarrow}_{\mathbf{C}'',\mathfrak{L}}(G, u^*) = 0$.

Claim 8.25 gives us a polynomial-time reduction from $Z^{\rightarrow}_{\mathbf{C}'',\mathfrak{L}}$ to $Z^{\rightarrow}_{\mathbf{C},\mathfrak{D}}$. A similar claim can be proved for $Z^{\leftarrow}$, and Lemma 8.24 follows. We now prove Claim 8.25.

For each $u \in U$ (and $v \in V$), we use r_u (and r_v) to denote its degree in G. To get G', we need an $\ell_u \in [s]$ for each $u \in U$ and an $\ell_v \in [t]$ for each $v \in V$ such that

$$\sum_{i \in [s]} \mu_i^{\ell_u N + r_u} \cdot K^{[r_u]}_{(0,i)} \neq 0 \quad \text{and} \quad \sum_{i \in [t]} \nu_i^{\ell_v N + r_v} \cdot K^{[r_v]}_{(1,i)} \neq 0. \tag{8.30}$$

Assume there exists a $u \in U$ such that no $\ell_u \in [s]$ satisfies (8.30). In this case, note that the s equations for $\ell_u = 1, \ldots, s$ form a Vandermonde system since $\mu_1 > \cdots > \mu_s > 0$. Therefore, the $(0,*)$-block of $\mathbf{K}^{[r_u]}$ is $\mathbf{0}$ and thus the $(0,*)$-block of $\mathbf{L}^{[r_u]}$ is also $\mathbf{0}$ by ($Shape_6$). It follows that $Z^{\rightarrow}_{\mathbf{C}'',\mathfrak{L}}(G,u^*) = 0$, and we are done. Similarly, we have $Z^{\rightarrow}_{\mathbf{C}'',\mathfrak{L}}(G,u^*) = 0$ if there exists a $v \in V$ such that no $\ell_v \in [t]$ satisfies (8.30).

Otherwise, suppose there do exist an $\ell_u \in [s]$ for each $u \in U$ and an $\ell_v \in [t]$ for each $v \in V$, which satisfy (8.30). We construct a bipartite $G' = (U' \cup V', E')$. First, $U' = U \cup \widehat{V}$ and $V' = V \cup \widehat{U}$, where $\widehat{V} = \{\widehat{v} : v \in V\}$ and $\widehat{U} = \{\widehat{u} : u \in U\}$. Edge set E' contains E over $U \cup V$ and the following edges: $\ell_u N$ parallel edges between u and $\widehat{u}$, for every $u \in U$, and $\ell_v N$ parallel edges between v and $\widehat{v}$, for every $v \in V$.

Clearly, G' is a connected and bipartite graph. The degree of $u \in U$ (or $v \in V$) is $r_u + \ell_u N$ (or $r_v + \ell_v N$), and the degree of $\widehat{u}$ (or $\widehat{v}$) is $\ell_u N$ (or $\ell_v N$). We now use G' to prove Claim 8.25.

First, we have (the sum is over all ξ that map U' to $[s]$, V' to $[t]$)

$$\begin{aligned}
Z^{\rightarrow}_{\mathbf{C}',\mathfrak{K}}(G',u^*) &= \sum_{\xi} \left(\prod_{uv \in E} M_{\xi(u),\xi(v)} \prod_{u \in U} M^{\ell_u N}_{\xi(u),\xi(\widehat{u})} \prod_{v \in V} M^{\ell_v N}_{\xi(\widehat{v}),\xi(v)} \right) \\
&\quad \times \left(\prod_{u \in U} K^{[r_u]}_{(0,\xi(u))} K^{[0]}_{(1,\xi(\widehat{u}))} \right) \left(\prod_{v \in V} K^{[r_v]}_{(1,\xi(v))} K^{[0]}_{(0,\xi(\widehat{v}))} \right) \\
&= \prod_{u \in U} \left(\sum_{i \in [s]} \mu_i^{\ell_u N + r_u} \cdot K^{[r_u]}_{(0,i)} \right) \prod_{v \in V} \left(\sum_{i \in [t]} \nu_i^{\ell_v N + r_v} \cdot K^{[r_v]}_{(1,i)} \right) \\
&\quad \times \prod_{\widehat{u} \in \widehat{U}} \left(\sum_{i \in [t]} \nu_i^{\ell_u N} \cdot K^{[0]}_{(1,i)} \right) \prod_{\widehat{v} \in \widehat{V}} \left(\sum_{i \in [s]} \mu_i^{\ell_v N} \cdot K^{[0]}_{(0,i)} \right).
\end{aligned}$$

It is nonzero: The first two factors are nonzero because of the way we pick ℓ_u and ℓ_v; the latter two factors are nonzero because $\mu_i, \nu_i > 0$, and by ($Shape_6$), every entry of $\mathbf{K}^{[0]}$ is a positive integer.

It now suffices to prove (8.29). Let η be an assignment that maps U to $[s]$ and V to $[t]$. Given η, let Ξ denote the set of assignments ξ over $U' \cup V'$ that map U' to $[s]$ and V' to $[t]$ and that satisfy $\xi(u) = \eta(u)$, $\xi(v) = \eta(v)$ for all $u \in U$ and $v \in V$. We have

$$\begin{aligned}
\sum_{\xi \in \Xi} \mathrm{wt}_{\mathbf{C}'',\mathfrak{L}}(\xi) &= \sum_{\xi \in \Xi} \left(\prod_{uv \in E} H_{\eta(u),\eta(v)} \prod_{u \in U} (H_{\eta(u),\xi(\widehat{u})})^{\ell_u N} \prod_{v \in V} (H_{\xi(\widehat{v}),\eta(v)})^{\ell_v N} \right) \\
&\quad \times \left(\prod_{u \in U} L^{[r_u]}_{(0,\eta(u))} L^{[0]}_{(1,\xi(\widehat{u}))} \right) \left(\prod_{v \in V} L^{[r_v]}_{(1,\eta(v))} L^{[0]}_{(0,\xi(\widehat{v}))} \right) \\
&= \sum_{\xi \in \Xi} \mathrm{wt}_{\mathbf{C}'',\mathfrak{L}}(\eta) \;=\; h^{|\widehat{U} \cup \widehat{V}|} \cdot \mathrm{wt}_{\mathbf{C}'',\mathfrak{L}}(\eta).
\end{aligned}$$

The second equation uses the fact that entries of $\mathbf{H}$ are powers of ω_N (thus $(H_{i,j})^N = 1$) and $\mathbf{L}^{[0]}$ is the identity matrix. Equation (8.29) then follows. ∎

8.6. Step 2.5. We are almost done with Step 2. The only conditions $(\mathcal{U}_i)$ that are possibly violated by $(\mathbf{C}'', \mathfrak{L})$ are $(\mathcal{U}_1)$ (N might be odd) and $(\mathcal{U}_2)$ ($H_{i,1}$ and $H_{1,j}$ might not be 1). We deal with $(\mathcal{U}_2)$ first.

What we will do below is to normalize $\mathbf{H}$ (in $\mathbf{C}''$) so that it becomes a discrete unitary matrix for some positive integer M that divides N, while not changing the complexity of $\mathsf{EVAL}(\mathbf{C}'', \mathfrak{L})$.

First, without loss of generality, we may assume $\mathbf{H}$ satisfies $H_{1,1} = 1$ since otherwise we can divide $\mathbf{H}$ by $H_{1,1}$, which does not affect the complexity of $\mathsf{EVAL}(\mathbf{C}'', \mathfrak{L})$. Then we construct the following pair: $(\mathbf{X}, \mathfrak{Y})$. $\mathbf{X}$ is the bipartization of an $h \times h$ matrix over $\mathbb{C}$, whose (i,j)th entry is $H_{i,j}\overline{H_{1,j}H_{i,1}}$; $\mathfrak{Y}$ is a sequence $(\mathbf{Y}^{[0]}, \ldots, \mathbf{Y}^{[N-1]})$ of $2h \times 2h$ diagonal matrices; $\mathbf{Y}^{[0]}$ is the identity matrix. Let

$$\mathcal{S} = \{r \in [0 : N-1] : \mathbf{L}^{[r]}_{(0,*)} \neq \mathbf{0}\} \quad \text{and} \quad \mathcal{T} = \{r \in [0 : N-1] : \mathbf{L}^{[r]}_{(1,*)} \neq \mathbf{0}\};$$

then we have

$$\mathbf{Y}^{[r]}_{(0,*)} = \mathbf{0} \ \text{ for all } r \notin \mathcal{S} \quad \text{and} \quad \mathbf{Y}^{[r]}_{(1,*)} = \mathbf{0} \ \text{ for all } r \notin \mathcal{T}.$$

For each $r \in \mathcal{S}$ (or $r \in \mathcal{T}$), by $(Shape_6)$ there must be an $a_r \in [h]$ (or $b_r \in [h]$, resp.) such that the $(0, a_r)$th entry of $\mathbf{L}^{[r]}$ is 1 (or the $(1, b_r)$th entry of $\mathbf{L}^{[r]}$ is 1, resp.). Set

$$Y^{[r]}_{(0,i)} = L^{[r]}_{(0,i)} \left(\frac{H_{i,1}}{H_{a_r,1}}\right)^r \text{ for all } i \in [h]; \ \ Y^{[r]}_{(1,j)} = L^{[r]}_{(1,j)} \left(\frac{H_{1,j}}{H_{1,b_r}}\right)^r \text{ for all } j \in [h].$$

We show that $\mathsf{EVAL}(\mathbf{C}'', \mathfrak{L}) \equiv \mathsf{EVAL}(\mathbf{X}, \mathfrak{Y})$. For $\mathsf{EVAL}(\mathbf{X}, \mathfrak{Y}) \leq \mathsf{EVAL}(\mathbf{C}'', \mathfrak{L})$, we let $G = (U \cup V, E)$ be a connected undirected graph and u^* be a vertex in U. For every $r \in \mathcal{S}$ (and $r \in \mathcal{T}$), we use $U_r \subseteq U$ (and $V_r \subseteq V$, resp.) to denote the set of vertices with degree r mod N. It is clear that if $U_r \neq \emptyset$ for some $r \notin \mathcal{S}$ or if $V_r \neq \emptyset$ for some $r \notin \mathcal{T}$, both $Z^{\rightarrow}_{\mathbf{C}'',\mathfrak{L}}(G, u^*)$ and $Z^{\rightarrow}_{\mathbf{X},\mathfrak{Y}}(G, u^*)$ are trivially zero. Otherwise, we have

$$Z^{\rightarrow}_{\mathbf{C}'',\mathfrak{L}}(G, u^*) = \left(\prod_{r \in \mathcal{S}} (H_{a_r,1})^{r|U_r|}\right) \left(\prod_{r \in \mathcal{T}} (H_{1,b_r})^{r|V_r|}\right) \cdot Z^{\rightarrow}_{\mathbf{X},\mathfrak{Y}}(G, u^*).$$

So the problem of computing $Z^{\rightarrow}_{\mathbf{X},\mathfrak{Y}}$ is reducible to computing $Z^{\rightarrow}_{\mathbf{C}'',\mathfrak{L}}$. By combining it with the second pinning lemma (Lemma 4.3), we know that computing $Z^{\rightarrow}_{\mathbf{X},\mathfrak{Y}}$ is reducible to $\mathsf{EVAL}(\mathbf{C}'', \mathfrak{L})$. A similar statement can be proved for $Z^{\leftarrow}_{\mathbf{X},\mathfrak{Y}}$, and it follows that $\mathsf{EVAL}(\mathbf{X}, \mathfrak{Y}) \leq \mathsf{EVAL}(\mathbf{C}'', \mathfrak{L})$. The other direction, $\mathsf{EVAL}(\mathbf{C}'', \mathfrak{L}) \leq \mathsf{EVAL}(\mathbf{X}, \mathfrak{Y})$, can be proved similarly.

One can verify that $(\mathbf{X}, \mathfrak{Y})$ satisfies $(\mathcal{U}_1)$–$(\mathcal{U}_4)$, except that N might be odd. In particular the upper-right $h \times h$ block of $\mathbf{X}$ is an M-discrete unitary matrix for some positive integer $M \mid N$, and $\mathfrak{Y}$ satisfies both $(\mathcal{U}_3)$ and $(\mathcal{U}_4)$ (which follows from the fact that every entry of $\mathbf{H}$ is a power of ω_N).

If N is even, then we are done with Step 2; otherwise we extend $\mathfrak{Y}$ to be

$$\mathfrak{Y}' = \{\mathbf{Y}^{[0]}, \ldots, \mathbf{Y}^{[N-1]}, \mathbf{Y}^{[N]}, \ldots, \mathbf{Y}^{[2N-1]}\},$$

where $\mathbf{Y}^{[r]} = \mathbf{Y}^{[r-N]}$, for all $r \in [N : 2N-1]$. We have $\mathsf{EVAL}(\mathbf{X}, \mathfrak{Y}) \equiv \mathsf{EVAL}(\mathbf{X}, \mathfrak{Y}')$, since $Z_{\mathbf{X},\mathfrak{Y}}(G) = Z_{\mathbf{X},\mathfrak{Y}'}(G)$, for all undirected G, and the new tuple $((M, 2N), \mathbf{X}, \mathfrak{Y}')$ now satisfies conditions $(\mathcal{U}_1)$–$(\mathcal{U}_4)$.

9. Proofs of Theorems 5.4 and 5.6. Let $((M, N), \mathbf{C}, \mathfrak{D})$ be a tuple that satisfies $(\mathcal{U}_1)$–$(\mathcal{U}_4)$ and let $\mathbf{F} \in \mathbb{C}^{m \times m}$ be the upper-right block of $\mathbf{C}$. In this section, we index the rows and columns of an $n \times n$ matrix with $[0 : n-1]$.

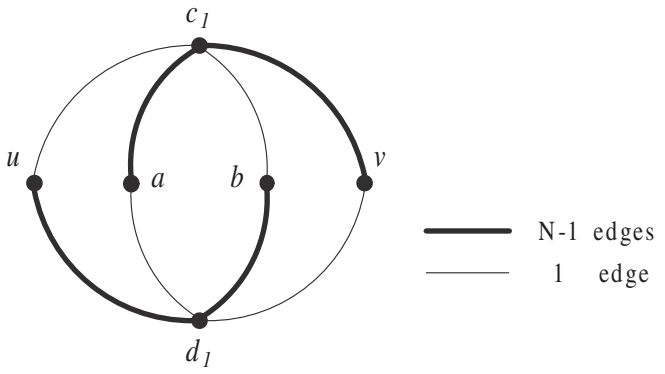


FIG. 9.1. *The gadget for $p = 1$. (Note that the subscript e is suppressed.)*

9.1. The group condition. We first show that either $\mathbf{F}$ satisfies the following condition or $\mathsf{EVAL}(\mathbf{C}, \mathfrak{D})$ is #P-hard.

LEMMA 9.1. *Let $((M, N), \mathbf{C}, \mathfrak{D})$ be a tuple that satisfies $(\mathcal{U}_1)$–$(\mathcal{U}_4)$. Then either $\mathbf{F}$ satisfies the group condition $(\mathcal{GC})$,*

(row-$\mathcal{GC}$) for all $i, j \in [0 : m-1]$, $\exists\, k \in [0 : m-1]$ such that $\mathbf{F}_{k,*} = \mathbf{F}_{i,*} \circ \mathbf{F}_{j,*}$;

(column-$\mathcal{GC}$) for all $i, j \in [0 : m-1]$, $\exists\, k \in [0 : m-1]$ such that $\mathbf{F}_{*,k} = \mathbf{F}_{*,i} \circ \mathbf{F}_{*,j}$,

or $\mathsf{EVAL}(\mathbf{C}, \mathfrak{D})$ is #P-hard.

Proof. Suppose $\mathsf{EVAL}(\mathbf{C}, \mathfrak{D})$ is not #P-hard.

Let $G = (V, E)$ be an undirected graph. For every integer $p \geq 1$, we construct a new graph $G^{[p]}$ by replacing every edge $uv \in E$ with a gadget. The gadget for $p = 1$ is shown in Figure 9.1. More exactly, we define $G^{[p]} = (V^{[p]}, E^{[p]})$ as

$$V^{[p]} = V \cup \big\{a_e, b_e, c_{e,1}, \ldots, c_{e,p}, d_{e,1}, \ldots, d_{e,p} : e \in E\big\},$$

and $E^{[p]}$ contains the following edges: For every $e = uv \in E$ and $i \in [p]$, add

1. one edge $(u, c_{e,i})$, $(c_{e,i}, b_e)$, $(d_{e,i}, a_e)$, and $(d_{e,i}, v)$;
2. $N - 1$ parallel edges $(c_{e,i}, v)$, $(c_{e,i}, a_e)$, $(d_{e,i}, b_e)$, and $(d_{e,i}, u)$.

It is easy to verify that the degree of every vertex in $G^{[p]}$ is a multiple of N. Thus, we have $Z_{\mathbf{C},\mathfrak{D}}(G^{[p]}) = Z_{\mathbf{C}}(G^{[p]})$ because $\mathfrak{D}$ satisfies $(\mathcal{U}_3)$. On the other hand, the way we construct $G^{[p]}$ gives us, for each $p \geq 1$, a symmetric matrix $\mathbf{A}^{[p]} \in \mathbb{C}^{2m \times 2m}$ which only depends on $\mathbf{C}$, such that $Z_{\mathbf{A}^{[p]}}(G) = Z_{\mathbf{C}}(G^{[p]}) = Z_{\mathbf{C},\mathfrak{D}}(G^{[p]})$ for all G. It follows that $\mathsf{EVAL}(\mathbf{A}^{[p]}) \leq \mathsf{EVAL}(\mathbf{C}, \mathfrak{D})$ and thus $\mathsf{EVAL}(\mathbf{A}^{[p]})$ is not #P-hard for all $p \geq 1$.

The (i, j)th entry of $\mathbf{A}^{[p]}$, where $i, j \in [0 : 2m-1]$, is

$$\begin{aligned} A^{[p]}_{i,j} &= \sum_{a=0}^{2m-1} \sum_{b=0}^{2m-1} \left(\sum_{c=0}^{2m-1} C_{i,c} \overline{C_{a,c}} C_{b,c} \overline{C_{j,c}} \right)^p \left(\sum_{d=0}^{2m-1} \overline{C_{i,d}} C_{a,d} \overline{C_{b,d}} C_{j,d} \right)^p \\ &= \sum_{a=0}^{2m-1} \sum_{b=0}^{2m-1} \left| \sum_{c=0}^{2m-1} C_{i,c} \overline{C_{a,c}} C_{b,c} \overline{C_{j,c}} \right|^{2p}. \end{aligned}$$

For the first equality, we used the fact that $M \,|\, N$ and thus, e.g., $(C_{a,c})^{N-1} = \overline{C_{a,c}}$ as $C_{a,c}$ is a power of ω_M. Note that $\mathbf{A}^{[p]}$ is symmetric and nonnegative and satisfies

$$A^{[p]}_{i,j} = A^{[p]}_{j,i} = 0 \quad \text{for all } i \in [0 : m-1] \text{ and } j \in [m, 2m-1].$$

For $i, j \in [0 : m-1]$, we have

$$A^{[p]}_{i,j} = \sum_{a=0}^{m-1} \sum_{b=0}^{m-1} \left|\langle \mathbf{F}_{i,*} \circ \overline{\mathbf{F}_{j,*}}, \mathbf{F}_{a,*} \circ \overline{\mathbf{F}_{b,*}} \rangle\right|^{2p} \quad \text{and}$$

$$A^{[p]}_{i+m,j+m} = \sum_{a=0}^{m-1} \sum_{b=0}^{m-1} \left|\langle \mathbf{F}_{*,i} \circ \overline{\mathbf{F}_{*,j}}, \mathbf{F}_{*,a} \circ \overline{\mathbf{F}_{*,b}} \rangle\right|^{2p}. \tag{9.1}$$

It is clear that all these entries are positive real numbers (by taking $a = i$ and $b = j$). Now let us focus on the upper-left $m \times m$ block of $\mathbf{A}^{[p]}$. Since it is a nonnegative symmetric matrix, we can apply the dichotomy theorem of Bulatov and Grohe.

On the one hand, for the special case when $j = i \in [0 : m-1]$, we have

$$A^{[p]}_{i,i} = \sum_{a=0}^{m-1} \sum_{b=0}^{m-1} \left|\langle \mathbf{1}, \mathbf{F}_{a,*} \circ \overline{\mathbf{F}_{b,*}} \rangle\right|^{2p} = \sum_{a=0}^{m-1} \sum_{b=0}^{m-1} \left|\langle \mathbf{F}_{a,*}, \mathbf{F}_{b,*} \rangle\right|^{2p}.$$

As $\mathbf{F}$ is discrete unitary, $A^{[p]}_{i,i} = m \cdot m^{2p}$. On the other hand, assuming $\mathsf{EVAL}(\mathbf{C}, \mathfrak{D})$ is not #P-hard, by using the Bulatov–Grohe dichotomy theorem (Corollary 2.6),

$$A^{[p]}_{i,i} \cdot A^{[p]}_{j,j} = A^{[p]}_{i,j} \cdot A^{[p]}_{j,i} = (A^{[p]}_{i,j})^2 \quad \text{for all } i \neq j \in [0 : m-1],$$

and thus $A^{[p]}_{i,j} = m^{2p+1}$ for all $i, j \in [0 : m-1]$.

Now we use this condition to prove that $\mathbf{F}$ satisfies (row-$\mathcal{GC}$). We introduce the following notation. For $i, j \in [0 : m-1]$, let

$$X_{i,j} = \left\{ |\langle \mathbf{F}_{i,*} \circ \overline{\mathbf{F}_{j,*}}, \mathbf{F}_{a,*} \circ \overline{\mathbf{F}_{b,*}} \rangle| \;\middle|\; a, b \in [0 : m-1] \right\}.$$

Clearly $X_{i,j}$ is finite for all i, j, with $|X_{i,j}| \leq m^2$. Each $x \in X_{i,j}$ satisfies $0 \leq x \leq m$. For each $x \in X_{i,j}$, let $s_{i,j}(x)$ denote the number of pairs $(a, b) \in [0 : m-1]^2$ such that

$$|\langle \mathbf{F}_{i,*} \circ \overline{\mathbf{F}_{j,*}}, \mathbf{F}_{a,*} \circ \overline{\mathbf{F}_{b,*}} \rangle| = x.$$

We can now rewrite $A^{[p]}_{i,j}$ as the sum

$$A^{[p]}_{i,j} = \sum_{x \in X_{i,j}} s_{i,j}(x) \cdot x^{2p}, \tag{9.2}$$

which is equal to m^{2p+1} for all $p \geq 1$. Note that $s_{i,j}(x)$ does not depend on p, and

$$\sum_{x \in X_{i,j}} s_{i,j}(x) = m^2. \tag{9.3}$$

We can view (9.2) and (9.3) as a linear system of equations in the unknowns $s_{i,j}(x)$. Fix i, j; then there are $|X_{i,j}|$ many variables $s_{i,j}(x)$, one for each distinct value $x \in X_{i,j}$. Equations in (9.2) are indexed by p. If we choose (9.3) and (9.2) for p from 1 up to $|X_{i,j}| - 1$, this linear system has an $|X_{i,j}| \times |X_{i,j}|$ Vandermonde matrix $((x^2)^p)$, with row index p and column index $x \in X_{i,j}$. It has full rank. Note that by setting $(a, b) = (i, j)$ and (i', j), where $i' \neq i$, respectively, we get $m \in X_{i,j}$ and $0 \in X_{i,j}$, respectively. Moreover, $s_{i,j}(0) = m^2 - m$, $s_{i,j}(m) = m$, and all other $s_{i,j}(x) = 0$ is a solution to the linear system. Therefore this must be the unique solution.

So $X_{i,j} = \{0, m\}$ and thus $|\langle \mathbf{F}_{i,*} \circ \overline{\mathbf{F}_{j,*}}, \mathbf{F}_{a,*} \circ \overline{\mathbf{F}_{b,*}} \rangle| \in \{0, m\}$ for all i, j, a, b.

Finally, we prove (row-$\mathcal{GC}$). Set $j = 0$. As $\mathbf{F}_{0,*} = \mathbf{1}$, the all-1 vector, we have

$$|\langle \mathbf{F}_{i,*} \circ \mathbf{1}, \mathbf{F}_{a,*} \circ \overline{\mathbf{F}_{b,*}} \rangle| = |\langle \mathbf{F}_{i,*} \circ \mathbf{F}_{b,*}, \mathbf{F}_{a,*} \rangle| \in \{0, m\} \quad \text{for all } i, a, b \in [0 : m-1].$$

As $\{\mathbf{F}_{a,*} : a \in [0 : m-1]\}$ is an orthogonal basis with $\|\mathbf{F}_{a,*}\|^2 = m$, by Parseval

$$\sum_a |\langle \mathbf{F}_{i,*} \circ \mathbf{F}_{b,*}, \mathbf{F}_{a,*} \rangle|^2 = m \cdot \|\mathbf{F}_{i,*} \circ \mathbf{F}_{b,*}\|^2.$$

Since every entry of $\mathbf{F}_{i,*} \circ \mathbf{F}_{b,*}$ is a root of unity, $\|\mathbf{F}_{i,*} \circ \mathbf{F}_{b,*}\|^2 = m$. Hence

$$\sum_a |\langle \mathbf{F}_{i,*} \circ \mathbf{F}_{b,*}, \mathbf{F}_{a,*} \rangle|^2 = m^2,$$

and for all $i, b \in [0 : m-1]$, there is a unique a such that $|\langle \mathbf{F}_{i,*} \circ \mathbf{F}_{b,*}, \mathbf{F}_{a,*} \rangle| = m$.

From property ($\mathcal{U}_2$), every entry of $\mathbf{F}_{i,*}$, $\mathbf{F}_{b,*}$, and $\mathbf{F}_{a,*}$ is a root of unity. The inner product $\langle \mathbf{F}_{i,*} \circ \mathbf{F}_{b,*}, \mathbf{F}_{a,*} \rangle$ is a sum of m terms each of complex norm 1. To sum to a complex number of norm m, each term must be a complex number of unit norms with the *same* argument, i.e., they are the same complex number $e^{i\theta}$. Thus, $\mathbf{F}_{i,*} \circ \mathbf{F}_{b,*} = e^{i\theta} \cdot \mathbf{F}_{a,*}$. We assert that in fact $e^{i\theta} = 1$, and $\mathbf{F}_{i,*} \circ \mathbf{F}_{b,*} = \mathbf{F}_{a,*}$. This is because $\mathbf{F}_{i,1} = \mathbf{F}_{a,1} = \mathbf{F}_{b,1} = 1$. This proves the group condition (row-$\mathcal{GC}$). One can prove (column-$\mathcal{GC}$) similarly using (9.1) and the lower-right $m \times m$ block of $\mathbf{A}^{[p]}$. □

Next we prove a property concerning discrete unitary matrices that satisfy ($\mathcal{GC}$). Given an $n \times n$ matrix $\mathbf{A}$, let A^{R} denote the set of its row vectors $\{\mathbf{A}_{i,*}\}$ and A^{C} denote the set of its column vectors $\{\mathbf{A}_{*,j}\}$. For general matrices, it is possible that $|A^{\mathrm{R}}|, |A^{\mathrm{C}}| < n$, since $\mathbf{A}$ may have duplicate rows or columns. But if $\mathbf{A}$ is M-discrete unitary, then it is clear that $|A^{\mathrm{R}}| = |A^{\mathrm{C}}| = n$.

PROPERTY 9.2. *If $\mathbf{A} \in \mathbb{C}^{n \times n}$ is an M-discrete unitary matrix that satisfies ($\mathcal{GC}$), then A^R and A^C are finite Abelian groups (of order n) under the Hadamard product.*

Proof. The Hadamard product $\circ$ gives a binary operation on A^{R} and A^{C}. The group condition ($\mathcal{GC}$) states that both sets A^{R} and A^{C} are closed under this operation, and it is clearly associative and commutative. Being discrete unitary, the all-1 vector $\mathbf{1}$ belongs to both A^{R} and A^{C} and serves as the identity element. This operation also satisfies the cancellation law: if $x \circ y = x \circ z$, then $y = z$. From general group theory, a finite set with these properties already forms a group. But here we can be more specific about the inverse of an element. For each $\mathbf{A}_{i,*}$, the inverse should clearly be $\overline{\mathbf{A}_{i,*}}$. By ($\mathcal{GC}$), there exists a $k \in [0 : m-1]$ such that $\mathbf{A}_{k,*} = (\mathbf{A}_{i,*})^{M-1} = \overline{\mathbf{A}_{i,*}}$. The second equation is because $A_{i,j}$, for all j, is a power of ω_M. □

9.2. Proof of Theorem 5.4. In this section, we prove Theorem 5.4 by showing that ($\mathcal{U}_1$)–($\mathcal{U}_4$) indeed imply ($\mathcal{U}_5$).

Suppose $\mathsf{EVAL}(\mathbf{C}, \mathfrak{D})$ is not #P-hard; otherwise we are already done. By Lemma 9.1, $((M, N), \mathbf{C}, \mathfrak{D})$ satisfies ($\mathcal{GC}$). Fixing r to be any index in $[N-1]$, we will prove ($\mathcal{U}_5$) for the (i,i)th entries of $\mathbf{D}^{[r]}$, where $i \in [m : 2m-1]$. The proof for the first half of $\mathbf{D}^{[r]}$ is similar. For simplicity, let $\mathbf{D}$ be the m-dimensional vector such that

$$D_i = D^{[r]}_{m+i} \quad \text{for all } i \in [0 : m-1].$$

Also let $K = \{i \in [0 : m-1] : D_i \neq 0\}$. If $|K| = 0$, then there is nothing to prove; if $|K| = 1$, then by ($\mathcal{U}_3$), the only nonzero entry in $\mathbf{D}$ must be 1. So we assume $|K| \geq 2$.

We start with a useful lemma. It implies that to prove Theorem 5.4, i.e., $(\mathcal{U}_5)$, it suffices to prove that D_i is a root of unity for every $i \in K$.

LEMMA 9.3. *If $D \in \mathbb{Q}(\omega_N)$ is a root of unity, then D must be a power of ω_N.*

Proof. Assume $D = \omega_M^k$ for some positive integers k and M with $\gcd(k, M) = 1$. Since $D \in \mathbb{Q}(\omega_N)$, we have $\omega_M^k \in \mathbb{Q}(\omega_N)$. By $\gcd(k, M) = 1$, $\omega_M \in \mathbb{Q}(\omega_N)$ and

$$\mathbb{Q}(\omega_N) = \mathbb{Q}(\omega_N, \omega_M) = \mathbb{Q}(\omega_{\mathrm{lcm}(M,N)}).$$

The degree of the field extension is $[\mathbb{Q}(\omega_N) : \mathbb{Q}] = \phi(N)$, the Euler function [25].

When $N \mid N'$ and $\phi(N) = \phi(N')$, by expanding according to the prime factorization for N, we can get (and actually this is all there is to be had) that if N is even, then $N' = N$; if N is odd, then $N' = N$ or $N' = 2N$. As by $(\mathcal{U}_1)$ N is even, we have $\mathrm{lcm}(M, N) = N$, $M \mid N$, and D is a power of ω_N. ∎

Next we show that every D_i, $i \in K$, is a root of unity. Suppose for a contradiction that this is not true. We show the following lemma about $\mathbf{Z} = (Z_0, \ldots, Z_{m-1})$, where $Z_i = (D_i)^N$.

LEMMA 9.4. *Suppose there is a $k \in K$ such that Z_k is not a root of unity. Then there exists an infinite integer sequence $\{P_n\}$ such that when $n \to \infty$, the vector sequence $((Z_k)^{P_n} : k \in K)$ approaches, but never reaches, the all-one vector $\mathbf{1}_{|K|}$.*

Proof. As Z_k has norm 1, $Z_k = e^{2\pi i \theta_k}$ for some real number $\theta_k \in [0, 1)$. We will treat θ_k as a number in the $\mathbb{Z}$-module $\mathbb{R}_{\bmod 1}$, i.e., real numbers modulo 1. By the assumption, we know that at least one of the θ_k's, $k \in K$, is irrational.

This lemma follows from the well-known Dirichlet's box principle. For completeness, we include a proof here. First, for any positive integer P, $((Z_k)^P : k \in K) \neq \mathbf{1}$; otherwise, every θ_k is rational, contradicting the assumption.

Let $n^* = n^{|K|} + 1$ for some integer $n > 1$. We consider $(L \cdot \theta_k : k \in K)$ for all $L \in [n^*]$. We divide the unit cube $[0, 1)^{|K|}$ into $n^* - 1$ subcubes of the following form:

$$\left[\frac{a_1}{n}, \frac{a_1 + 1}{n}\right) \times \cdots \times \left[\frac{a_{|K|}}{n}, \frac{a_{|K|} + 1}{n}\right),$$

where $a_k \in \{0, \ldots, n-1\}$ for all k. By cardinality, there are $L \neq L' \in [n^*]$ such that

$$\big(L \cdot \theta_k \bmod 1 : k \in K\big) \quad \text{and} \quad \big(L' \cdot \theta_k \bmod 1 : k \in K\big)$$

fall in the same subcube. Assume $L > L'$; by setting $P_n = L - L' \geq 1$, we have

$$\big|P_n \cdot \theta_k \bmod 1\big| = |(L - L') \cdot \theta_k \bmod 1| < 1/n \quad \text{for all } k \in K.$$

Repeating the procedure for every n, we get an infinite sequence $\{P_n\}$ such that

$$\Big((Z_k)^{P_n} = e^{2\pi i (P_n \cdot \theta_k)} : k \in K\Big)$$

approaches, but never reaches, the all-one vector of dimension $|K|$. ∎

Let $G = (V, E)$ be an undirected graph. Then for each $p \geq 1$, we build a graph $G^{[p]}$ by replacing every edge $e = uv \in E$ with a gadget that is shown in Figure 9.2. Recall that $r \in [N - 1]$ is fixed. More exactly, we define $G^{[p]} = (V^{[p]}, E^{[p]})$ as follows:

$$V^{[p]} = V \cup \big\{a_e, b_{e,i}, c_{e,i,j}, a'_e, b'_{e,i}, c'_{e,i,j} : e \in E, i \in [pN], j \in [r]\big\},$$

and $E^{[p]}$ contains the following edges: For each edge $e = uv \in E$, add

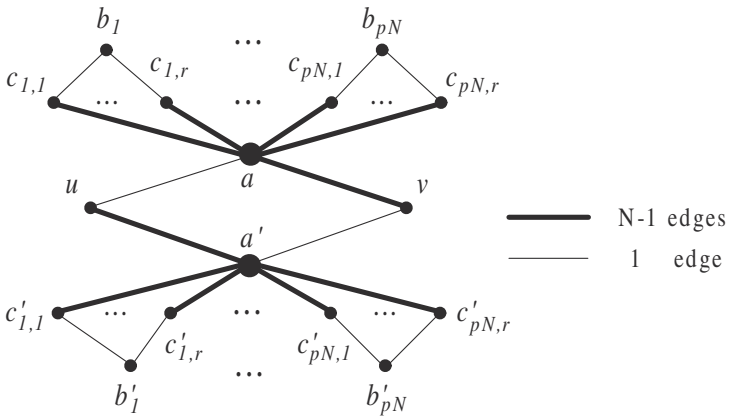


FIG. 9.2. *The gadget for $p = 1$. (Note that the subscript e is suppressed.)*

1. one edge (u, a_e) and (v, a'_e);
2. $N - 1$ parallel edges (a_e, v) and (u, a'_e);
3. one edge $(c_{e,i,j}, b_{e,i})$ and $(c'_{e,i,j}, b'_{e,i})$ for all $i \in [pN]$ and $j \in [r]$;
4. $N - 1$ parallel edges $(a_e, c_{e,i,j})$ and $(a'_e, c'_{e,i,j})$ for all $i \in [pN]$ and $j \in [r]$.

It is easy to verify that the degree of every vertex in $G^{[p]}$ is a multiple of N, except $b_{e,i}$ and $b'_{e,i}$, which have degree $r \bmod N$.

As the gadget is symmetric, the construction gives a symmetric $2m \times 2m$ matrix $\mathbf{A}^{[p]}$ such that $Z_{\mathbf{A}^{[p]}}(G) = Z_{\mathbf{C},\mathfrak{D}}(G^{[p]})$ for all G and thus $\mathsf{EVAL}(\mathbf{A}^{[p]}) \le \mathsf{EVAL}(\mathbf{C}, \mathfrak{D})$, and $\mathsf{EVAL}(\mathbf{A}^{[p]})$ is also not #P-hard.

The entries of $\mathbf{A}^{[p]}$ are as follows. First, for all $u, v \in [0 : m-1]$, the $(u, m+v)$th and $(m+u, v)$th entries of $\mathbf{A}^{[p]}$ are zero. The entries in the upper-left block are

$$A^{[p]}_{u,v} = \left(\sum_{a \in [0:m-1]} F_{u,a} \overline{F_{v,a}} \left(\sum_{b \in [0:m-1]} D^{[r]}_{m+b} \left(\sum_{c \in [0:m-1]} F_{c,b} \overline{F_{c,a}} \right)^r \right)^{pN} \right) \times \left(\sum_{a \in [0:m-1]} \overline{F_{u,a}} F_{v,a} \left(\sum_{b \in [0:m-1]} D^{[r]}_{m+b} \left(\sum_{c \in [0:m-1]} F_{c,b} \overline{F_{c,a}} \right)^r \right)^{pN} \right)$$

for all $u, v \in [0 : m-1]$. Since $\mathbf{F}$ is discrete unitary,

$$\sum_{c \in [0:m-1]} F_{c,b} \overline{F_{c,a}} = \langle \mathbf{F}_{*,b}, \mathbf{F}_{*,a} \rangle = 0,$$

unless $a = b$. As a result, the equation can be simplified to

$$A^{[p]}_{u,v} = L_p \cdot \left(\sum_{k \in K} (D_k)^{pN} F_{u,k} \overline{F_{v,k}} \right) \left(\sum_{k \in K} (D_k)^{pN} \overline{F_{u,k}} F_{v,k} \right)$$

for $u, v \in [0 : m-1]$, where L_p is a positive constant that is independent of u and v.

Assume for a contradiction that some D_k, $k \in K$, is not a root of unity. Then by Lemma 9.4 we know there exists a sequence $\{P_n\}$ such that $((D_k)^{NP_n} : k \in K)$

approaches, but never equals, the all-one vector, when $n \to \infty$. Also by $(\mathcal{U}_3)$ we know there exists an $i \in K$ such that $D_i = 1$. Now consider $G^{[P_n]}$ with parameter $p = P_n$ from this sequence. We have

$$A_{u,u}^{[P_n]} = L_{P_n} \cdot \left(\sum_{k \in K} (D_k)^{NP_n} \right)^2 \quad \text{for any } u \in [0 : m-1].$$

We let T_n denote the second factor on the right-hand side; then $|T_n|$ could be arbitrarily close to $|K|^2$ if we choose n large enough. By using the dichotomy theorem of Bulatov and Grohe and Lemma 7.5 together with the assumption that $\mathsf{EVAL}(\mathbf{A}^{[P_n]})$ is not #P-hard, we know the norm of every entry of $\mathbf{A}^{[P_n]}$ in its upper-left block is either 0 or $L_{P_n}|T_n|$.

Now we focus on the first row by fixing $u = 0$. Since $\mathbf{F}_{0,*} = \mathbf{1}$, we have

$$A_{0,v}^{[P_n]} = L_{P_n} \cdot \left(\sum_{k \in K} (D_k)^{NP_n} \overline{F_{v,k}} \right) \left(\sum_{k \in K} (D_k)^{NP_n} F_{v,k} \right) \quad \text{for any } v \in [0 : m-1].$$

By Property 9.2, $F^{\mathrm{R}} = \{\mathbf{F}_{v,*}\}$ is a group under the Hadamard product. We let

$$S = \{v \in [0 : m-1] : \text{ for all } i, j \in K, F_{v,i} = F_{v,j}\}$$

and denote $\{\mathbf{F}_{v,*} : v \in S\}$ by F^S. F^S is a subgroup of F^{R}, and $0 \in S$ as $\mathbf{F}_{0,*} = \mathbf{1}$.

For any $v \notin S$, when n is sufficiently large, we have

$$\left| A_{0,v}^{[P_n]} \right| < \left| A_{0,0}^{[P_n]} \right|.$$

This is because when $n \to \infty$, $T_n \to |K|^2$ but

$$\left(\sum_{k \in K} (D_k)^{NP_n} \overline{F_{v,k}} \right) \left(\sum_{k \in K} (D_k)^{NP_n} F_{v,k} \right) \to \left(\sum_{k \in K} \overline{F_{v,k}} \right) \left(\sum_{k \in K} F_{v,k} \right),$$

which has norm $< |K|^2$ since $v \notin S$. So when n is sufficiently large, $A_{0,v}^{[P_n]} = 0$ for all $v \notin S$. Denote $((D_k)^{NP_n} : k \in [0 : m-1])$ by $\mathbf{D}^n$; for $v \notin S$ and sufficiently large n,

$$\text{either } \langle \mathbf{D}^n, \mathbf{F}_{v,*} \rangle = 0 \quad \text{or } \langle \mathbf{D}^n, \overline{\mathbf{F}_{v,*}} \rangle = 0. \tag{9.4}$$

Next, we focus on the characteristic vector $\boldsymbol{\chi}$ (of dimension m) of K: $\chi_k = 1$ if $k \in K$ and $\chi_k = 0$ elsewhere. By (9.4) and the definition of S, we have

$$\langle \boldsymbol{\chi}, \mathbf{F}_{v,*} \rangle = 0 \text{ for all } v \notin S \quad \text{and} \quad |\langle \boldsymbol{\chi}, \mathbf{F}_{v,*} \rangle| = |K| \text{ for all } v \in S. \tag{9.5}$$

To prove the first equation, note that by (9.4), either there is an infinite subsequence $(\mathbf{D}^n)$ that satisfies $\langle \mathbf{D}^n, \mathbf{F}_{v,*} \rangle = 0$ or there is an infinite subsequence that satisfies $\langle \mathbf{D}^n, \overline{\mathbf{F}_{v,*}} \rangle = 0$. Since $\mathbf{D}^n \to \boldsymbol{\chi}$ when $n \to \infty$, either $\langle \boldsymbol{\chi}, \mathbf{F}_{v,*} \rangle = 0$ or $\langle \boldsymbol{\chi}, \overline{\mathbf{F}_{v,*}} \rangle = 0$. The second case still gives us $\langle \boldsymbol{\chi}, \mathbf{F}_{v,*} \rangle = 0$ since $\boldsymbol{\chi}$ is real. The second equation in (9.5) follows directly from the definition of S. As a result, we have

$$\boldsymbol{\chi} = \frac{1}{m} \sum_{v \in S} \langle \boldsymbol{\chi}, \mathbf{F}_{v,*} \rangle \cdot \mathbf{F}_{v,*}.$$

Now we assume the expression of $\mathbf{D}^n$, under the orthogonal basis $\{\mathbf{F}_{v,*}\}$, is

$$\mathbf{D}^n = \sum_{i=0}^{m-1} x_{i,n}\mathbf{F}_{i,*}, \quad \text{where } \; x_{i,n} = \frac{1}{m}\langle \mathbf{D}^n, \mathbf{F}_{i,*}\rangle.$$

If for some n we have $x_{i,n} = 0$ for all $i \notin S$, then we are done, because by the definition of S, every $\mathbf{F}_{i,*}$, $i \in S$, is a constant over K and thus the vector $\mathbf{D}^n$ is a constant over K. Since we know there exists an $i \in K$ such that $D_i = 1$, every D_j, $j \in K$, must be a root of unity.

Assume this is not the case. Then (here consider those sufficiently large n so that (9.4) holds),

$$\boldsymbol{\chi} = \mathbf{D}^n \circ \overline{\mathbf{D}^n} = \left(\sum_i x_{i,n}\mathbf{F}_{i,*}\right) \circ \left(\sum_j \overline{x_{j,n}}\,\overline{\mathbf{F}_{j,*}}\right) = \sum_v y_{v,n}\mathbf{F}_{v,*},$$

where

$$y_{v,n} = \sum_{\mathbf{F}_{i,*}\circ\overline{\mathbf{F}_{j,*}}=\mathbf{F}_{v,*}} x_{i,n}\overline{x_{j,n}}.$$

The last equation uses the fact that F^{R} is a group under the Hadamard product (so for any i, j there exists a unique v such that $\mathbf{F}_{v,*} = \mathbf{F}_{i,*} \circ \overline{\mathbf{F}_{j,*}}$).

Since the Fourier expansion of $\boldsymbol{\chi}$ under $\{\mathbf{F}_{v,*}\}$ is unique, we must have $y_{v,n} = 0$ for any $v \notin S$. Because $\mathbf{D}^n \to \boldsymbol{\chi}$, by (9.5), we know that when $n \to \infty$, $x_{i,n}$, for any $i \notin S$ can be arbitrarily close to 0, while $|x_{i,n}|$ can be arbitrarily close to $|K|/m$ for any $i \in S$. So there exists a sufficiently large n such that

$$|x_{i,n}| < \frac{4|K||S|}{5m^2} \quad \text{for all } i \notin S \quad \text{and} \quad |x_{i,n}| > \frac{4|K|}{5m} \quad \text{for all } i \in S.$$

We pick such an n and will use it to reach a contradiction. Since we assumed that for any n (which is of course also true for this particular n we picked here), there exists at least one index $i \notin S$ such that $x_{i,n} \neq 0$, and we can choose a $w \notin S$ that maximizes $|x_{i,n}|$ among all $i \notin S$. Clearly, $|x_{w,n}|$ is positive.

We consider the expression of $y_{w,n}$ using $x_{i,n}$. We divide the summation into two parts: the *main* terms $x_{i,n}\overline{x_{j,n}}$, in which either $i \in S$ or $j \in S$, and the remaining terms, in which $i, j \notin S$. (Note that if $\mathbf{F}_{w,*} = \mathbf{F}_{i,*} \circ \overline{\mathbf{F}_{j,*}}$, then i and j cannot both be in S; otherwise, since F^S is a subgroup, we have $w \in S$, which contradicts the assumption that $w \notin S$.) The main terms of $y_{w,n}$ are given by

$$\frac{1}{m^2}\sum_{j\in S}\langle \mathbf{D}^n, \mathbf{F}_{w,*} \circ \mathbf{F}_{j,*}\rangle\overline{\langle \mathbf{D}^n, \mathbf{F}_{j,*}\rangle} + \frac{1}{m^2}\sum_{i\in S}\langle \mathbf{D}^n, \mathbf{F}_{i,*}\rangle\overline{\langle \mathbf{D}^n, \mathbf{F}_{i,*} \circ \overline{\mathbf{F}_{w,*}}\rangle}.$$

Note that $x_{0,n} = \langle \mathbf{D}^n, \mathbf{F}_{0,*}\rangle/m$ and $\mathbf{F}_{0,*} = \mathbf{1}$. Also note that (by the definition of S) when $j \in S$, $F_{j,k} = \alpha_j$ for all $k \in K$, for some complex number α_j of norm 1. Since $\mathbf{D}^n$ is only nonzero on K, we have

$$\langle \mathbf{D}^n, \mathbf{F}_{w,*} \circ \mathbf{F}_{j,*}\rangle\overline{\langle \mathbf{D}^n, \mathbf{F}_{j,*}\rangle} = \langle \mathbf{D}^n, \alpha_j\mathbf{F}_{w,*}\rangle\overline{\langle \mathbf{D}^n, \alpha_j\mathbf{1}\rangle} = m\overline{x_{0,n}} \cdot \langle \mathbf{D}^n, \mathbf{F}_{w,*}\rangle.$$

Similarly, we can simplify the other sum so that the main terms of $y_{w,n}$ are given by

$$\frac{|S|}{m} \cdot \left(\overline{x_{0,n}}\langle \mathbf{D}^n, \mathbf{F}_{w,*}\rangle + x_{0,n}\langle \overline{\mathbf{D}^n}, \mathbf{F}_{w,*}\rangle\right).$$

By (9.4) we have either $\langle \mathbf{D}^n, \mathbf{F}_{w,*} \rangle$ or $\langle \overline{\mathbf{D}^n}, \mathbf{F}_{w,*} \rangle$ is 0. Since we assumed that $x_{w,n} = \langle \mathbf{D}^n, \mathbf{F}_{w,*} \rangle / m \neq 0$, the latter has to be 0. Therefore, the sum of the main terms of $y_{w,n}$ is equal to $\overline{x_{0,n}} x_{w,n} |S|$. As $0 \in S$, we have

$$\left| \overline{x_{0,n}} x_{w,n} |S| \right| \geq \frac{4|K||S|}{5m} |x_{w,n}|.$$

Consider the remaining terms. Below we prove that the sum of all these terms cannot have a norm as large as $|\overline{x_{0,n}} x_{w,n} |S||$ and thus $y_{w,n}$ is nonzero and we get a contradiction. To see this, it is easy to check that the number of remaining terms is at most m, and the norm of each of them is

$$|x_{i,n} \overline{x_{j,n}}| \leq |x_{w,n}|^2 < \frac{4|K||S|}{5m^2} |x_{w,n}|$$

since $i, j \notin S$. So the norm of their sum is $< |\overline{x_{0,n}} x_{w,n} |S||$. Theorem 5.4 is proved.

9.3. Decomposing F into Fourier matrices. Suppose $((M, N), \mathbf{C}, \mathfrak{D})$ satisfies $(\mathcal{U}_1)$–$(\mathcal{U}_5)$ and $(\mathcal{GC})$; otherwise $\mathsf{EVAL}(\mathbf{C}, \mathfrak{D})$ is #P-hard. We prove Theorem 5.6. To decompose $\mathbf{F}$ into Fourier matrices (recall that $\mathbf{F}$ is the upper-right $m \times m$ block matrix of $\mathbf{C}$), we first show that if $M = pq$ and $\gcd(p, q) = 1$, then up to a permutation of rows and columns, $\mathbf{F}$ is the tensor product of two smaller matrices, both of which are discrete unitary and satisfy $(\mathcal{GC})$. Note that p and q here are not necessarily primes or prime powers.

LEMMA 9.5. *Let $\mathbf{F} \in \mathbb{C}^{m \times m}$ be an M-discrete unitary matrix that satisfies $(\mathcal{GC})$, where $M = pq$, $p, q > 1$, and $\gcd(p, q) = 1$. Then there exist two permutations Π and Σ over $[0 : m-1]$ such that $\mathbf{F}_{\Pi,\Sigma} = \mathbf{F}' \otimes \mathbf{F}''$, where $\mathbf{F}'$ is a p-discrete unitary matrix, $\mathbf{F}''$ is a q-discrete unitary matrix, and both of them satisfy $(\mathcal{GC})$.*

Proof. Using Property 9.2, both F^{R} and F^{C} are finite Abelian groups. Since $\mathbf{F}$ is M-discrete unitary, the order of any vector in F^{R} or F^{C} is a divisor of M.

By the fundamental theorem of Abelian groups, there is a group isomorphism

$$\rho : F^{\mathrm{R}} \to \mathbb{Z}_{g_1} \times \cdots \times \mathbb{Z}_{g_h} \equiv \mathbb{Z}_{\mathbf{g}},$$

where $g_1, \ldots, g_h$ are prime powers, and $g_i \,|\, M$ for all i. As $\gcd(p, q) = 1$, without loss of generality, we may assume there exists an integer h' such that $g_i \,|\, p$ for all $i \in [h']$ and $g_i \,|\, q$ for all other i. We use ρ^{-1} to define the following two subsets of F^{R}:

$$S^p = \{\rho^{-1}(\mathbf{x}) : \mathbf{x} \in \mathbb{Z}_{\mathbf{g}},\ x_i = 0 \text{ for all } i > h'\} \quad \text{and}$$
$$S^q = \{\rho^{-1}(\mathbf{x}) : \mathbf{x} \in \mathbb{Z}_{\mathbf{g}},\ x_i = 0 \text{ for all } i \leq h'\}.$$

It is easy to show the following four properties: Letting $m' = |S^p|$ and $m'' = |S^q|$,

1. both S^p and S^q are subgroups of F^{R};
2. $S^p = \{\mathbf{u} \in F^{\mathrm{R}} : (\mathbf{u})^p = \mathbf{1}\}$ and $S^q = \{\mathbf{v} \in F^{\mathrm{R}} : (\mathbf{v})^q = \mathbf{1}\}$;
3. $m = m' \cdot m''$, $\gcd(m', q) = 1$, $\gcd(m'', p) = 1$, $\gcd(m', m'') = 1$;
4. $(\mathbf{u}, \mathbf{v}) \mapsto \mathbf{u} \circ \mathbf{v}$ is a group isomorphism from $S^p \times S^q$ onto F^{R}.

Let $S^p = \{\mathbf{u}_0 = \mathbf{1}, \mathbf{u}_1, \ldots, \mathbf{u}_{m'-1}\}$ and $S^q = \{\mathbf{v}_0 = \mathbf{1}, \mathbf{v}_1, \ldots, \mathbf{v}_{m''-1}\}$. By 4, there is a bijection $f : i \mapsto (f_1(i), f_2(i))$ from $[0 : m-1]$ to $[0 : m'-1] \times [0 : m''-1]$ such that

$$\mathbf{F}_{i,*} = \mathbf{u}_{f_1(i)} \circ \mathbf{v}_{f_2(i)} \quad \text{for all } i \in [0 : m-1]. \tag{9.6}$$

Next we apply the fundamental theorem to F^{C}. We use the group isomorphism in the same way to define two subgroups T^p and T^q with four corresponding properties:

1. Both T^p and T^q are subgroups of F^{C};
2. $T^p = \{\mathbf{w} \in F^{\mathrm{C}} : (\mathbf{w})^p = \mathbf{1}\}$ and $T^q = \{\mathbf{r} \in F^{\mathrm{C}} : (\mathbf{r})^q = \mathbf{1}\}$;
3. $m = |T^p| \cdot |T^q|$, $\gcd(|T^p|, q) = 1$, $\gcd(|T^q|, p) = 1$, and $\gcd(|T^p|, |T^q|) = 1$;
4. $(\mathbf{w}, \mathbf{r}) \mapsto \mathbf{w} \circ \mathbf{r}$ is a group isomorphism from $T^p \times T^q$ onto F^{C}.

By comparing item 3 in both lists, we have $|T^p| = |S^p| = m'$ and $|T^q| = |S^q| = m''$.

Let $T^p = \{\mathbf{w}_0 = \mathbf{1}, \mathbf{w}_1, \ldots, \mathbf{w}_{m'-1}\}$ and $T^q = \{\mathbf{r}_0 = \mathbf{1}, \mathbf{r}_1, \ldots, \mathbf{r}_{m''-1}\}$. Then by item 4, we have a bijection g from $[0 : m-1]$ to $[0 : m'-1] \times [0 : m''-1]$ and

$$\mathbf{F}_{*,j} = \mathbf{w}_{g_1(j)} \circ \mathbf{r}_{g_2(j)} \quad \text{for all } j \in [0 : m-1]. \tag{9.7}$$

Now we are ready to permute the rows and columns of $\mathbf{F}$ to get a new matrix $\mathbf{G}$ that is the tensor product of two smaller matrices. We use (x_1, x_2), where $x_1 \in [0 : m'-1], x_2 \in [0 : m''-1]$, to index the rows and columns of $\mathbf{G}$. We use $\Pi(x_1, x_2) = f^{-1}(x_1, x_2)$, from $[0 : m'-1] \times [0 : m''-1]$ to $[0 : m-1]$, to permute the rows of $\mathbf{F}$ and $\Sigma(y_1, y_2) = g^{-1}(y_1, y_2)$ to permute the columns of $\mathbf{F}$. We get $\mathbf{G} = \mathbf{F}_{\Pi,\Sigma}$, where

$$G_{(x_1,x_2),(y_1,y_2)} = F_{\Pi(x_1,x_2),\Sigma(y_1,y_2)} \quad \text{for all } x_1, y_1 \in [0 : m'-1], x_2, y_2 \in [0 : m''-1].$$

By (9.6), and using the fact that $\mathbf{u}_0 = \mathbf{1}$ and $\mathbf{v}_0 = \mathbf{1}$, we have

$$\mathbf{G}_{(x_1,x_2),*} = \mathbf{G}_{(x_1,0),*} \circ \mathbf{G}_{(0,x_2),*}.$$

Similarly by (9.7) and $\mathbf{w}_0 = \mathbf{1}$ and $\mathbf{r}_0 = \mathbf{1}$, we have

$$\mathbf{G}_{*,(y_1,y_2)} = \mathbf{G}_{*,(y_1,0)} \circ \mathbf{G}_{*,(0,y_2)}.$$

Therefore, applying both relations, we have

$$G_{(x_1,x_2),(y_1,y_2)} = G_{(x_1,0),(y_1,0)} \cdot G_{(x_1,0),(0,y_2)} \cdot G_{(0,x_2),(y_1,0)} \cdot G_{(0,x_2),(0,y_2)}.$$

We claim

$$G_{(x_1,0),(0,y_2)} = 1 \quad \text{and} \quad G_{(0,x_2),(y_1,0)} = 1. \tag{9.8}$$

Then we have

$$G_{(x_1,x_2),(y_1,y_2)} = G_{(x_1,0),(y_1,0)} \cdot G_{(0,x_2),(0,y_2)}. \tag{9.9}$$

To prove the first equation in (9.8), we realize that it appears as an entry in both $\mathbf{u}_{x_1}$ and $\mathbf{r}_{y_2}$. Then, by item 2 for S^p and T^q, both its pth and qth powers are 1. Thus it has to be 1. The other equation in (9.8) can be proved the same way.

As a result, we have obtained our tensor product decomposition $\mathbf{G} = \mathbf{F}' \otimes \mathbf{F}''$:

$$\mathbf{F}' = \left(F'_{x,y} \equiv G_{(x,0),(y,0)}\right) \quad \text{and} \quad \mathbf{F}'' = \left(F''_{x,y} \equiv G_{(0,x),(0,y)}\right).$$

The only thing left is to show that $\mathbf{F}', \mathbf{F}''$ are both discrete unitary and satisfy $(\mathcal{GC})$. Here we only prove it for $\mathbf{F}'$. The proof for $\mathbf{F}''$ is the same. For all $x \neq y$,

$$\begin{aligned} 0 = \langle \mathbf{G}_{(x,0),*}, \mathbf{G}_{(y,0),*} \rangle &= \sum_{z_1,z_2} G_{(x,0),(z_1,z_2)} \overline{G_{(y,0),(z_1,z_2)}} \\ &= \sum_{z_1,z_2} G_{(x,0),(z_1,0)} G_{(0,0),(0,z_2)} \overline{G_{(y,0),(z_1,0)} G_{(0,0),(0,z_2)}} = m'' \cdot \langle \mathbf{F}'_{x,*}, \mathbf{F}'_{y,*} \rangle. \end{aligned}$$

Here we used the factorization (9.9) and $\mathbf{u}_0 = \mathbf{1}$ and $\mathbf{v}_0 = \mathbf{1}$. Similarly, we can prove that $\mathbf{F}'_{*,x}$ and $\mathbf{F}'_{*,y}$ are orthogonal for all $x \neq y$. $\mathbf{F}'$ also satisfies $(\mathcal{GC})$ because both S^p and T^p are groups and thus closed under the Hadamard product. Finally, $\mathbf{F}'$ is exactly p-discrete unitary. First, by the definition of M and (9.9), we have

$$pq = M = \operatorname{lcm}\big\{\text{order of } G_{(x_1,0),(y_1,0)} \cdot G_{(x_2,0),(y_2,0)} : \mathbf{x}, \mathbf{y}\big\}.$$

Second, the order of $G_{(x_1,0),(y_1,0)}$ divides p and the order of $G_{(x_2,0),(y_2,0)}$ divides q. As a result, p is the least common multiple of orders of entries of $\mathbf{F}'$ and thus $\mathbf{F}'$ is p-discrete unitary. □

Next we prove Lemma 9.7, which deals with the case when M is a prime power.

PROPERTY 9.6. *Let $\mathbf{A}$ be an M-discrete unitary matrix that satisfies the group condition $(\mathcal{GC})$. If M is a prime power, then one of its entries is equal to ω_M.*

Proof. Since M is a prime power, some entry of $\mathbf{A}$ has order exactly M as a root of unity. Hence it has the form ω_M^k for some k relatively prime to M. Then by the group condition $(\mathcal{GC})$ all powers of ω_M^k appear as entries of $\mathbf{A}$, in particular ω_M. □

LEMMA 9.7. *Let $\mathbf{F} \in \mathbb{C}^{m\times m}$ be an M-discrete unitary matrix that satisfies $(\mathcal{GC})$. Moreover, $M = p^k$ is a prime power for some $k \geq 1$. Then there exist two permutations Π and Σ such that $\mathbf{F}_{\Pi,\Sigma} = \boldsymbol{\mathcal{F}}_M \otimes \mathbf{F}'$, where $\mathbf{F}'$ is an M'-discrete unitary matrix, $M' = p^{k'}$ for some $k' \leq k$, and $\mathbf{F}'$ satisfies $(\mathcal{GC})$.*

Proof. By Property 9.6, there exist a and b such that $F_{a,b} = \omega_M$. Thus, both the order of $\mathbf{F}_{a,*}$ (in F^{R}) and the order of $\mathbf{F}_{*,b}$ (in F^{C}) are M. Let

$$S_1 = \big\{\mathbf{1}, \mathbf{F}_{a,*}, (\mathbf{F}_{a,*})^2, \ldots, (\mathbf{F}_{a,*})^{M-1}\big\}$$

denote the subgroup of F^{R} generated by $\mathbf{F}_{a,*}$. As the order of $\mathbf{F}_{a,*}$ is M, $|S_1| = M$.

Let S_2 denote the subset of F^{R} such that $\mathbf{u} \in S_2$ iff $u_b = 1$. Then it is clear that S_2 is a subgroup of F^{R}. Moreover, $(\mathbf{w}_1, \mathbf{w}_2) \mapsto \mathbf{w}_1 \circ \mathbf{w}_2$ is a group isomorphism from $S_1 \times S_2$ onto F^{R}. As a result, $|S_2| = m/M$, which we denote by n.

Let $S_2 = \{\mathbf{u}_0 = \mathbf{1}, \mathbf{u}_1, \ldots, \mathbf{u}_{n-1}\}$. Then there exists a bijection f from $[0 : m-1]$ to $[0 : M-1] \times [0 : n-1]$, where $i \mapsto f(i) = (f_1(i), f_2(i))$, such that

$$\mathbf{F}_{i,*} = (\mathbf{F}_{a,*})^{f_1(i)} \circ \mathbf{u}_{f_2(i)} \quad \text{for all } i \in [0 : m-1]. \tag{9.10}$$

In particular, we have $f(a) = (1, 0)$.

Similarly, we use T_1 to denote the subgroup of F^{C} generated by $\mathbf{F}_{*,b}$ ($|T_1| = M$) and T_2 to denote the subgroup of F^{C} that contains all the $\mathbf{v} \in F^{\mathrm{C}}$ such that $v_a = 1$. $(\mathbf{w}_1, \mathbf{w}_2) \mapsto \mathbf{w}_1 \circ \mathbf{w}_2$ is an isomorphism from $T_1 \times T_2$ onto F^{C}, so $|T_2| = m/M = n$

Let $T_2 = \{\mathbf{v}_0 = \mathbf{1}, \mathbf{v}_1, \ldots, \mathbf{v}_{n-1}\}$. Then there exists a bijection g from $[0 : m-1]$ to $[0 : M-1] \times [0 : n-1]$, where $j \mapsto g(j) = (g_1(j), g_2(j))$, such that

$$\mathbf{F}_{*,j} = (\mathbf{F}_{*,b})^{g_1(j)} \circ \mathbf{v}_{g_2(j)} \quad \text{for all } j \in [0 : m-1]. \tag{9.11}$$

In particular, we have $g(b) = (1, 0)$.

We are ready to permute the rows and columns of $\mathbf{F}$ to get a new $m \times m$ matrix $\mathbf{G}$. We use (x_1, x_2), where $x_1 \in [0 : M-1]$ and $x_2 \in [0 : n-1]$, to index the rows and columns of matrix $\mathbf{G}$. We use $\Pi(x_1, x_2) = f^{-1}(x_1, x_2)$, from $[0 : M-1] \times [0 : n-1]$ to $[0 : m-1]$, to permute the rows and $\Sigma(y_1, y_2) = g^{-1}(y_1, y_2)$ to permute the columns of $\mathbf{F}$, respectively. As a result, we get $\mathbf{G} = \mathbf{F}_{\Pi,\Sigma}$.

By (9.10) and (9.11), and $\mathbf{u}_0 = \mathbf{1}$ and $\mathbf{v}_0 = \mathbf{1}$, we have

$$\mathbf{G}_{(x_1,x_2),*} = (\mathbf{G}_{(1,0),*})^{x_1} \circ \mathbf{G}_{(0,x_2),*} \quad \text{and} \quad \mathbf{G}_{*,(y_1,y_2)} = (\mathbf{G}_{*,(1,0)})^{y_1} \circ \mathbf{G}_{*,(0,y_2)}.$$

Applying them in succession, we get

$$\begin{aligned} G_{(x_1,x_2),(y_1,y_2)} &= (G_{(1,0),(y_1,y_2)})^{x_1} G_{(0,x_2),(y_1,y_2)} \\ &= (G_{(1,0),(1,0)})^{x_1 y_1} (G_{(1,0),(0,y_2)})^{x_1} (G_{(0,x_2),(1,0)})^{y_1} G_{(0,x_2),(0,y_2)}. \end{aligned}$$

By $f(a) = (1,0)$ and $g(b) = (1,0)$, we have

$$G_{(1,0),(1,0)} = F_{\Pi(1,0),\Sigma(1,0)} = F_{f^{-1}(1,0),g^{-1}(1,0)} = F_{a,b} = \omega_M.$$

By (9.11), and similar reasoning, we have

$$G_{(1,0),(0,y_2)} = F_{a,g^{-1}(0,y_2)} = (F_{a,b})^0 \cdot v_{y_2,a} = v_{y_2,a} = 1,$$

where $v_{y_2,a}$ denotes the ath entry of $\mathbf{v}_{y_2}$, which is 1 by the definition of T_2. By (9.10),

$$G_{(0,x_2),(1,0)} = F_{f^{-1}(0,x_2),b} = (F_{a,b})^0 \cdot u_{x_2,b} = u_{x_2,b} = 1,$$

where $u_{x_2,b}$ denotes the bth entry of $\mathbf{u}_{x_2}$, which is 1 by the definition of S_2.

Combining all these equations, we have

$$G_{(x_1,x_2),(y_1,y_2)} = \omega_M^{x_1 y_1} \cdot G_{(0,x_2),(0,y_2)}. \tag{9.12}$$

As a result, $\mathbf{G} = \boldsymbol{\mathcal{F}}_M \otimes \mathbf{F}'$, where $\mathbf{F}' = (F'_{x,y} \equiv G_{(0,x),(0,y)})$ is an $n \times n$ matrix.

To see $\mathbf{F}'$ is discrete unitary, by (9.12), we have

$$0 = \langle \mathbf{G}_{(0,x),*}, \mathbf{G}_{(0,y),*} \rangle = M \cdot \langle \mathbf{F}'_{x,*}, \mathbf{F}'_{y,*} \rangle \quad \text{for any } x \neq y \in [0:n-1].$$

Similarly we can prove that $\mathbf{F}'_{*,x}$ and $\mathbf{F}'_{*,y}$ are orthogonal for $x \neq y$. $\mathbf{F}'$ also satisfies the group condition because both S_2 and T_2 are groups and thus closed under the Hadamard product. More precisely, for (row-$\mathcal{GC}$), suppose $\mathbf{F}'_{x,*}$ and $\mathbf{F}'_{y,*}$ are two rows of $\mathbf{F}'$. The corresponding rows $\mathbf{G}_{(0,x),*}$ and $\mathbf{G}_{(0,y),*}$ in $\mathbf{G}$ are permuted versions of $\mathbf{u}_x$ and $\mathbf{u}_y$, respectively. We have, by (9.6),

$$F'_{x,z} = F_{f^{-1}(0,x),g^{-1}(0,z)} = u_{x,g^{-1}(0,z)} \quad \text{and} \quad F'_{y,z} = F_{f^{-1}(0,y),g^{-1}(0,z)} = u_{y,g^{-1}(0,z)}.$$

Since S_2 is a group, we have some $w \in [0:n-1]$ such that $\mathbf{u}_x \circ \mathbf{u}_y = \mathbf{u}_w$ and thus

$$F'_{x,z} \cdot F'_{y,z} = u_{w,g^{-1}(0,z)} = F'_{w,z}.$$

The proof of (column-$\mathcal{GC}$) is similar. $\mathbf{F}'$ is also $p^{k'}$-discrete unitary for some $k' \leq k$. ∎

Theorem 5.6 then follows from Lemmas 9.5 and 9.7.

10. Proof of Theorem 5.8. Let $((M,N), \mathbf{C}, \mathfrak{D}, (\mathbf{q}, \mathbf{t}, \mathcal{Q}))$ be a 4-tuple that satisfies condition $(\mathcal{R})$. Also assume that $\mathsf{EVAL}(\mathbf{C}, \mathfrak{D})$ is not #P-hard; otherwise, we are done. For every r in $\mathcal{T}$ (recall that $\mathcal{T}$ is the set of $r \in [N-1]$ such that $\Delta_r \neq \emptyset$), we show that Δ_r must be a coset in $\mathbb{Z}_{\mathcal{Q}}$. Condition $(\mathcal{L}_2)$ then follows from the following lemma. Condition $(\mathcal{L}_1)$ about Λ_r can be proved similarly.

LEMMA 10.1. *Let Φ be a coset in $G_1 \times G_2$, where G_1 and G_2 are finite Abelian groups such that* $\gcd(|G_1|, |G_2|) = 1$. *Then for both $i = 1, 2$, there exists a coset Φ_i in G_i such that $\Phi = \Phi_1 \times \Phi_2$.*

Proof. First, we show that if $\mathbf{u} = (u_1, u_2), \mathbf{v} = (v_1, v_2) \in \Phi$, where $u_i, v_i \in G_i$, then $(u_1, v_2) \in \Phi$.

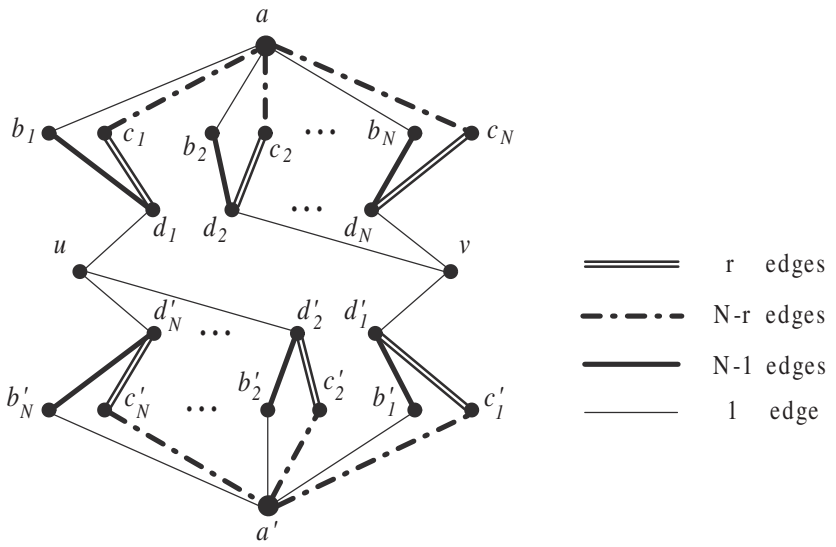


FIG. 10.1. *The gadget for constructing graph G'. (Note that the subscript e is suppressed.)*

Since $\gcd(|G_1|, |G_2|) = 1$, we can pick an integer k such that $|G_1| \,|\, k$ and $k \equiv 1 \pmod{|G_2|}$. As Φ is a coset, we have $\mathbf{u} + k(\mathbf{v} - \mathbf{u}) \in \Phi$. From $u_1 + k(v_1 - u_1) = u_1$ and $u_2 + k(v_2 - u_2) = v_2$, we conclude that $(u_1, v_2) \in \Phi$.

This implies the existence of $\Phi_1 \subseteq G_1$ and $\Phi_2 \subseteq G_2$ such that $\Phi = \Phi_1 \times \Phi_2$: Let

$$\Phi_1 = \big\{x \in G_1 : \exists\, y \in G_2, (x, y) \in \Phi\big\} \quad \text{and} \quad \Phi_2 = \big\{y \in G_2 : \exists\, x \in G_1, (x, y) \in \Phi\big\}.$$

Then both Φ_1 and Φ_2 are cosets (in G_1 and G_2, respectively), and $\Phi = \Phi_1 \times \Phi_2$. □

To prove Theorem 5.8, we need the following construction. Given an undirected graph $G = (V, E)$, we build a new graph G' by replacing every edge $e = uv \in E$ with the gadget shown in Figure 10.1. More exactly, we define $G' = (V', E')$ as

$$V' = V \cup \big\{a_e, b_{e,i}, c_{e,i}, d_{e,i}, a'_e, b'_{e,i}, c'_{e,i}, d'_{e,i} : e \in E \text{ and } i \in [N]\big\}$$

and E' contains exactly the following edges: For each $e = uv \in E$, add

1. one edge $(u, d_{e,1})$, $(v, d'_{e,1})$, $(u, d'_{e,i})$ and $(v, d_{e,i})$ for all $i \in [2 : N]$;
2. one edge $(a_e, b_{e,i})$ and $N - 1$ parallel edges $(b_{e,i}, d_{e,i})$ for all $i \in [N]$;
3. $N - r$ parallel edges $(a_e, c_{e,i})$ and r parallel edges $(c_{e,i}, d_{e,i})$ for all $i \in [N]$;
4. one edge $(a'_e, b'_{e,i})$ and $N - 1$ parallel edges $(b'_{e,i}, d'_{e,i})$ for all $i \in [N]$;
5. $N - r$ parallel edges $(a'_e, c'_{e,i})$ and r parallel edges $(c'_{e,i}, d'_{e,i})$ for all $i \in [N]$.

The degree of $d_{e,i}$ and $d'_{e,i}$ for all $e \in E, i \in [N]$, is $r \pmod N$. All other vertices in V' have degree $0 \pmod N$. It is also noted that the graph fragment that defines the gadget is bipartite, with $u, v, b_{e,i}, c_{e,i}, b'_{e,i}, c'_{e,i}$ on one side and $a_e, a'_e, d_{e,i}, d'_{e,i}$ on the other side. The way we construct G' gives us a $2m \times 2m$ matrix $\mathbf{A}$ such that $Z_{\mathbf{A}}(G) = Z_{\mathbf{C},\mathfrak{D}}(G')$ for all G, and thus $\mathsf{EVAL}(\mathbf{A}) \leq \mathsf{EVAL}(\mathbf{C}, \mathfrak{D})$, and $\mathsf{EVAL}(\mathbf{A})$ is also not #P-hard. We use $\{0, 1\} \times \mathbb{Z}_{\mathcal{Q}}$ to index the rows and columns of $\mathbf{A}$. Then for all $\mathbf{u}, \mathbf{v}$ in $\mathbb{Z}_{\mathcal{Q}}$, $A_{(0,\mathbf{u}),(1,\mathbf{v})} = A_{(1,\mathbf{u}),(0,\mathbf{v})} = 0$, which follows from the gadget being bipartite.

We now analyze the upper-left $m \times m$ block of $\mathbf{A}$. For $\mathbf{u}, \mathbf{v} \in \mathbb{Z}_{\mathcal{Q}}$, $A_{(0,\mathbf{u}),(0,\mathbf{v})}$ is the product of the following two sums:

$$\sum_{\mathbf{a},\mathbf{d}_1,\ldots,\mathbf{d}_N\in\mathbb{Z}_{\mathcal{Q}}} F_{\mathbf{u},\mathbf{d}_1}\prod_{i=2}^{N}F_{\mathbf{v},\mathbf{d}_i}\left(\prod_{i=1}^{N}\left(\sum_{\mathbf{b}_i\in\mathbb{Z}_{\mathcal{Q}}}F_{\mathbf{b}_i,\mathbf{a}}\overline{F_{\mathbf{b}_i,\mathbf{d}_i}}\right)\left(\sum_{\mathbf{c}_i\in\mathbb{Z}_{\mathcal{Q}}}F_{\mathbf{c}_i,\mathbf{a}}^{N-r}F_{\mathbf{c}_i,\mathbf{d}_i}^{r}\right)\right)\prod_{i=1}^{N}D^{[r]}_{(1,\mathbf{d}_i)}$$

and

$$\sum_{\mathbf{a},\mathbf{d}_1,\ldots,\mathbf{d}_N\in\mathbb{Z}_{\mathcal{Q}}} F_{\mathbf{v},\mathbf{d}_1}\prod_{i=2}^{N}F_{\mathbf{u},\mathbf{d}_i}\left(\prod_{i=1}^{N}\left(\sum_{\mathbf{b}_i\in\mathbb{Z}_{\mathcal{Q}}}F_{\mathbf{b}_i,\mathbf{a}}\overline{F_{\mathbf{b}_i,\mathbf{d}_i}}\right)\left(\sum_{\mathbf{c}_i\in\mathbb{Z}_{\mathcal{Q}}}F_{\mathbf{c}_i,\mathbf{a}}^{N-r}F_{\mathbf{c}_i,\mathbf{d}_i}^{r}\right)\right)\prod_{i=1}^{N}D^{[r]}_{(1,\mathbf{d}_i)}.$$

Note that in deriving these sums, we used the fact that $M\,|\,N$ and entries of $\mathbf{F}$ are all powers of ω_M. Next, since $\mathbf{F}$ is discrete unitary,

$$\sum_{\mathbf{b}_i\in\mathbb{Z}_{\mathcal{Q}}} F_{\mathbf{b}_i,\mathbf{a}}\overline{F_{\mathbf{b}_i,\mathbf{d}_i}} = \langle \mathbf{F}_{*,\mathbf{a}}, \mathbf{F}_{*,\mathbf{d}_i}\rangle$$

is m when $\mathbf{d}_i = \mathbf{a}$ and is 0 otherwise. The same thing can be said about those sums over $\mathbf{c}_i$. Assuming $\mathbf{d}_i = \mathbf{a}$ for all i, by $(\mathcal{U}_5)$, we have that

$$\prod_{i\in[N]} D^{[r]}_{(1,\mathbf{d}_i)} = \left(D^{[r]}_{(1,\mathbf{a})}\right)^N$$

is 1 when $\mathbf{a}\in\Delta_r$ and 0 otherwise. As a result, we have

(10.1)
$$A_{(0,\mathbf{u}),(0,\mathbf{v})} = \left(\sum_{\mathbf{a}\in\Delta_r} F_{\mathbf{u},\mathbf{a}}\overline{F_{\mathbf{v},\mathbf{a}}}m^{2N}\right)\left(\sum_{\mathbf{a}\in\Delta_r} F_{\mathbf{v},\mathbf{a}}\overline{F_{\mathbf{u},\mathbf{a}}}m^{2N}\right) = m^{4N}\left|\sum_{\mathbf{a}\in\Delta_r} F_{\mathbf{u},\mathbf{a}}\overline{F_{\mathbf{v},\mathbf{a}}}\right|^2.$$

By using condition $(\mathcal{R}_3)$, we can further simplify (10.1) to be

$$A_{(0,\mathbf{u}),(0,\mathbf{v})} = m^{4N}\left|\sum_{\mathbf{a}\in\Delta_r} F_{\mathbf{u}-\mathbf{v},\mathbf{a}}\right|^2 = m^{4N}\left|\langle\chi, \mathbf{F}_{\mathbf{u}-\mathbf{v},*}\rangle\right|^2, \tag{10.2}$$

where χ is a 0-1 characteristic vector such that $\chi_{\mathbf{a}} = 0$ if $\mathbf{a}\notin\Delta_r$ and $\chi_{\mathbf{a}} = 1$ if $\mathbf{a}\in\Delta_r$, for all $\mathbf{a}\in\mathbb{Z}_{\mathcal{Q}}$. Since $\mathbf{F}$ is discrete unitary, it is easy to show that

$$0 \le A_{(0,\mathbf{u}),(0,\mathbf{v})} \le m^{4N}|\Delta_r|^2 \quad\text{and}\quad A_{(0,\mathbf{u}),(0,\mathbf{u})} = m^{4N}|\Delta_r|^2 \quad \text{for all } \mathbf{u},\mathbf{v}\in\mathbb{Z}_{\mathcal{Q}}.$$

As $r\in\mathcal{T}$, we have $|\Delta_r|\ge 1$, and let n denote $|\Delta_r|$. Using the dichotomy of Bulatov and Grohe (Corollary 11.1) and the assumption that $\mathsf{EVAL}(\mathbf{A})$ is not #P-hard,

$$A_{(0,\mathbf{u}),(0,\mathbf{v})} \in \left\{0, m^{4N}n^2\right\} \text{ for all } \mathbf{u},\mathbf{v}\in\mathbb{Z}_{\mathcal{Q}}.$$

As a result, we have for all $\mathbf{u}\in\mathbb{Z}_{\mathcal{Q}}$,

$$\left|\langle\chi, \mathbf{F}_{\mathbf{u},*}\rangle\right| \in \{0, n\}. \tag{10.3}$$

The inner product $\langle\chi, \mathbf{F}_{\mathbf{u},*}\rangle$ is a sum of n terms, each term a power of ω_M. To sum to a complex number of norm n, each term must have exactly the same argument; any misalignment will result in a complex number of norm $< n$, which is the maximum possible. This implies that

$$\langle\chi, \mathbf{F}_{\mathbf{u},*}\rangle \in \left\{0, n, n\omega_M, n\omega_M^2, \ldots, n\omega_M^{M-1}\right\}. \tag{10.4}$$

Next, let $\mathfrak{a}$ denote a vector in Δ_r. We use Φ to denote $\mathfrak{a} + \langle \Delta_r - \mathfrak{a} \rangle$, where

$$\Delta_r - \mathfrak{a} \equiv \left\{ \mathbf{x} - \mathfrak{a} \,\middle|\, \mathbf{x} \in \Delta_r \right\}$$

and $\langle \Delta_r - \mathfrak{a} \rangle$ is the subgroup generated by $\Delta_r - \mathfrak{a}$. Clearly $\Delta_r \subseteq \Phi$. We want to prove that $\Delta_r = \Phi$, which by definition is a coset in $\mathbb{Z}_{\mathcal{Q}}$. This, combined with Lemma 10.1, will finish the proof of Theorem 5.8.

To this end, we use $\boldsymbol{\kappa}$ to denote the characteristic vector of Φ: $\kappa_{\mathbf{x}} = 0$ if $\mathbf{x} \notin \Phi$ and $\kappa_{\mathbf{x}} = 1$ if $\mathbf{x} \in \Phi$. We will show that for every $\mathbf{u} \in \mathbb{Z}_{\mathcal{Q}}$,

$$\langle \boldsymbol{\kappa}, \mathbf{F}_{\mathbf{u},*} \rangle = \frac{|\Phi|}{|\Delta_r|} \langle \chi, \mathbf{F}_{\mathbf{u},*} \rangle. \tag{10.5}$$

Since $\mathbf{F}$ is discrete unitary, $\{\mathbf{F}_{\mathbf{u},*} : \mathbf{u} \in \mathbb{Z}_{\mathcal{Q}}\}$ is an orthogonal basis. From (10.5),

$$\boldsymbol{\kappa} = \frac{|\Phi|}{|\Delta_r|} \chi,$$

which implies $\boldsymbol{\kappa} = \chi$ (since both are 0-1 vectors) and thus, $\Delta_r = \Phi$ is a coset in $\mathbb{Z}_{\mathcal{Q}}$.

We now prove (10.5). We make the following observations: (1) If $|\langle \chi, \mathbf{F}_{\mathbf{u},*} \rangle| = n$, then there is an $\alpha \in \mathbb{Z}_M$ such that $F_{\mathbf{u},\mathbf{x}} = \omega_M^{\alpha}$ for all $\mathbf{x} \in \Delta_r$. (2) Otherwise (which is equivalent to $\langle \chi, \mathbf{F}_{\mathbf{u},*} \rangle = 0$ from (10.3)), there exist $\mathbf{y}$ and $\mathbf{z}$ in Δ_r such that $F_{\mathbf{u},\mathbf{y}} \neq F_{\mathbf{u},\mathbf{z}}$. Observation (1) has already been noted when we proved (10.4). Observation (2) is obvious since if $F_{\mathbf{u},\mathbf{y}} = F_{\mathbf{u},\mathbf{z}}$ for all $\mathbf{y}, \mathbf{z} \in \Delta_r$, then clearly $\langle \chi, \mathbf{F}_{\mathbf{u},*} \rangle \neq 0$.

Equation (10.5) then follows from the following two lemmas.

LEMMA 10.2. *If there exists an α such that $F_{\mathbf{u},\mathbf{x}} = \omega_M^{\alpha}$ for all $\mathbf{x} \in \Delta_r$, then we have $F_{\mathbf{u},\mathbf{x}} = \omega_M^{\alpha}$ for all $\mathbf{x} \in \Phi$.*

Proof. Let $\mathbf{x}$ be a vector in Φ; then there exist $\mathbf{x}_1, \ldots, \mathbf{x}_k \in \Delta_r$ and $h_1, \ldots, h_k \in \{\pm 1\}$ for some $k \geq 0$ such that $\mathbf{x} = \mathfrak{a} + \sum_{i=1}^{k} h_i(\mathbf{x}_i - \mathfrak{a})$. By using $(\mathcal{R}_3)$ together with the assumption that $F_{\mathbf{u},\mathfrak{a}} = F_{\mathbf{u},\mathbf{x}_i} = \omega_M^{\alpha}$, we have

$$F_{\mathbf{u},\mathbf{x}} = F_{\mathbf{u},\mathfrak{a}+\sum_i h_i(\mathbf{x}_i-\mathfrak{a})} = F_{\mathbf{u},\mathfrak{a}} \prod_i F_{\mathbf{u},h_i(\mathbf{x}_i-\mathfrak{a})} = F_{\mathbf{u},\mathfrak{a}} \prod_i \left(F_{\mathbf{u},\mathbf{x}_i} \overline{F_{\mathbf{u},\mathfrak{a}}}\right)^{h_i} = \omega_M^{\alpha},$$

and the lemma is proved. ∎

LEMMA 10.3. *If there exist $\mathbf{y}, \mathbf{z} \in \Phi$ such that $F_{\mathbf{u},\mathbf{y}} \neq F_{\mathbf{u},\mathbf{z}}$, then $\sum_{\mathbf{x} \in \Phi} F_{\mathbf{u},\mathbf{x}} = 0$.*

Proof. Let ℓ be the smallest positive integer such that $\ell(\mathbf{y} - \mathbf{z}) = \mathbf{0}$; then ℓ exists because $\mathbb{Z}_{\mathcal{Q}}$ is a finite group and $\ell > 1$ because $\mathbf{y} \neq \mathbf{z}$. We use c to denote $F_{\mathbf{u},\mathbf{y}} \overline{F_{\mathbf{u},\mathbf{z}}}$. By $(\mathcal{R}_3)$ together with the assumption, we have $c^{\ell} = F_{\mathbf{u},\ell(\mathbf{y}-\mathbf{z})} = 1$ but $c \neq 1$.

We define the following equivalence relation $\sim$ over Φ. For $\mathbf{x}, \mathbf{x}' \in \Phi$, $\mathbf{x} \sim \mathbf{x}'$ iff there exists an integer k such that $\mathbf{x} - \mathbf{x}' = k(\mathbf{y} - \mathbf{z})$. For each $\mathbf{x} \in \Phi$, its equivalence class contains the following ℓ vectors: $\mathbf{x},\ \mathbf{x} + (\mathbf{y} - \mathbf{z}),\ \ldots,\ \mathbf{x} + (l-1)(\mathbf{y} - \mathbf{z})$, as Φ is a coset in $\mathbb{Z}_{\mathcal{Q}}$. We conclude that $\sum_{\mathbf{x} \in \Phi} F_{\mathbf{u},\mathbf{x}} = 0$ since for every class, by using $(\mathcal{R}_3)$,

$$\sum_{i=0}^{l-1} F_{\mathbf{u},\mathbf{x}+i(\mathbf{y}-\mathbf{z})} = F_{\mathbf{u},\mathbf{x}} \sum_{i=0}^{l-1} c^i = F_{\mathbf{u},\mathbf{x}} \frac{1 - c^l}{1 - c} = 0,$$

and the lemma is proved. ∎

Now (10.5) can be proved as follows. If $|\langle \chi, \mathbf{F}_{\mathbf{u},*} \rangle| = n\ (= |\Delta_r|)$, then by observation (1) and Lemma 10.2, $|\langle \boldsymbol{\kappa}, \mathbf{F}_{\mathbf{u},*} \rangle| = |\Phi|$. If $|\langle \chi, \mathbf{F}_{\mathbf{u},*} \rangle| \neq n$, then $\langle \chi, \mathbf{F}_{\mathbf{u},*} \rangle = 0$. By observation (2) and $\Delta_r \subseteq \Phi$, Lemma 10.3 implies $\langle \boldsymbol{\kappa}, \mathbf{F}_{\mathbf{u},*} \rangle = 0$. Therefore, Δ_r is a coset in $\mathbb{Z}_{\mathcal{Q}}$. To get the decomposition $(\mathcal{L}_2)$ for $\Delta_r = \prod_{i=1}^{s} \Delta_{r,i}$, we use Lemma 10.1.

10.1. A Corollary of Theorem 5.8. Now that we have proved Theorem 5.8, we know that unless the problem is #P-hard, we may assume that $(\mathcal{L})$ holds. Thus, Λ_r and Δ_r are cosets.

COROLLARY 10.4. *Let* $\mathbf{H}$ *be the* $m \times |\Delta_r|$ *submatrix obtained from* $\mathbf{F}$ *by restricting to the columns indexed by* Δ_r*. Then for any two rows* $\mathbf{H}_{\mathbf{u},*}$ *and* $\mathbf{H}_{\mathbf{v},*}$*, where* $\mathbf{u}, \mathbf{v} \in \mathbb{Z}_{\mathcal{Q}}$*, either there exists some* $\alpha \in \mathbb{Z}_M$ *such that* $\mathbf{H}_{\mathbf{u},*} = \omega_M^{\alpha} \cdot \mathbf{H}_{\mathbf{v},*}$ *or* $\langle \mathbf{H}_{\mathbf{u},*}, \mathbf{H}_{\mathbf{v},*} \rangle = 0$.

Similarly we denote by $\mathbf{G}$ *the* $|\Lambda_r| \times m$ *submatrix obtained from* $\mathbf{F}$ *by restricting to the rows indexed by* Λ_r*. Then for any two columns* $\mathbf{G}_{*,\mathbf{u}}$ *and* $\mathbf{G}_{*,\mathbf{v}}$*, where* $\mathbf{u}, \mathbf{v} \in \mathbb{Z}_{\mathcal{Q}}$*, either there exists an* $\alpha \in \mathbb{Z}_M$ *such that* $\mathbf{G}_{*,\mathbf{u}} = \omega_M^{\alpha} \cdot \mathbf{G}_{*,\mathbf{v}}$ *or* $\langle \mathbf{G}_{*,\mathbf{u}}, \mathbf{G}_{*,\mathbf{v}} \rangle = 0$.

Proof. The rows of $\mathbf{H}$ are restrictions of $\mathbf{F}$. Any two rows $\mathbf{H}_{\mathbf{u},*}, \mathbf{H}_{\mathbf{v},*}$ satisfy

$$\mathbf{H}_{\mathbf{u},*} \circ \overline{\mathbf{H}_{\mathbf{v},*}} = \mathbf{F}_{\mathbf{u}-\mathbf{v},*}|_{\Delta_r} = \mathbf{H}_{\mathbf{u}-\mathbf{v},*},$$

which is a row in $\mathbf{H}$. If this $\mathbf{H}_{\mathbf{u}-\mathbf{v},*}$ is a constant, namely, ω_M^{α} for some $\alpha \in \mathbb{Z}_M$, then $\mathbf{H}_{\mathbf{u},*} = \omega_M^{\alpha} \mathbf{H}_{\mathbf{v},*}$; otherwise, Lemma 10.3 says that $\langle \mathbf{H}_{\mathbf{u},*}, \mathbf{H}_{\mathbf{v},*} \rangle = 0$.

The proof for $\mathbf{G}$ is exactly the same. ∎

As part of a discrete unitary matrix $\mathbf{F}$, all columns $\{\mathbf{H}_{*,\mathbf{u}} \mid \mathbf{u} \in \Delta_r\}$ of $\mathbf{H}$ must be orthogonal and thus $\operatorname{rank}(\mathbf{H}) = |\Delta_r|$. We denote by n the cardinality $|\Delta_r|$. There must be n linearly independent rows in $\mathbf{H}$. We may start with $\mathfrak{b}_0 = \mathbf{0}$ and assume the n vectors $\mathfrak{b}_0 = \mathbf{0}, \mathfrak{b}_1, \ldots, \mathfrak{b}_{n-1} \in \mathbb{Z}_{\mathcal{Q}}$ are the indices of a set of linearly independent rows. By Corollary 10.4, these must be orthogonal as row vectors (over $\mathbb{C}$). Since the rank of the matrix $\mathbf{H}$ is exactly n, it is clear that all other rows must be a multiple of these rows, since the only alternative is to be orthogonal to them all, by Corollary 10.4 again, which is absurd. A symmetric statement for $\mathbf{G}$ also holds.

11. Proof of Theorem 5.9. Let $((M, N), \mathbf{C}, \mathfrak{D}, (\mathbf{p}, \mathbf{t}, \mathcal{Q}))$ be a tuple that satisfies $(\mathcal{R})$ and $(\mathcal{L})$ including $(\mathcal{L}_3)$. We also assume that $\mathsf{EVAL}(\mathbf{C}, \mathfrak{D})$ is not #P-hard. By $(\mathcal{L})$, we have $\Lambda_r = \prod_{i=1}^{s} \Lambda_{r,i}$ for every $r \in \mathcal{S}$ and $\Delta_r = \prod_{i=1}^{s} \Delta_{r,i}$ for every $r \in \mathcal{T}$, where both $\Lambda_{r,i}$ and $\Delta_{r,i}$ are cosets in $\mathbb{Z}_{\mathbf{q}_i}$.

Let r be an integer in $\mathcal{S}$. Below we prove $(\mathcal{D}_1)$ and $(\mathcal{D}_3)$ for Λ_r. The other parts of the theorem, that is, $(\mathcal{D}_2)$ and $(\mathcal{D}_4)$, can be proved similarly.

Let $\mathbf{G}$ denote the $|\Lambda_r| \times m$ submatrix of $\mathbf{F}$ whose row set is $\Lambda_r \subseteq \mathbb{Z}_{\mathcal{Q}}$. We start with the following simple lemma about $\mathbf{G}$. In this section, we let $n = |\Lambda_r| \geq 1$. A symmetric statement also holds for the $m \times |\Delta_r|$ submatrix of $\mathbf{F}$ whose column set is Δ_r, where we replace $n = |\Lambda_r|$ by $|\Delta_r|$, which could be different.

LEMMA 11.1. *There exist vectors* $\mathfrak{b}_0 = \mathbf{0}, \mathfrak{b}_1, \ldots, \mathfrak{b}_{n-1} \in \mathbb{Z}_{\mathcal{Q}}$ *such that*

1. $\{\mathbf{G}_{*,\mathfrak{b}_i} : i \in [0 : n-1]\}$ *forms an orthogonal basis;*
2. *for all* $\mathbf{b} \in \mathbb{Z}_{\mathcal{Q}}$, $\exists\, i \in [0 : n-1]$ *and* $\alpha \in \mathbb{Z}_M$ *such that* $\mathbf{G}_{*,\mathbf{b}} = \omega_M^{\alpha} \cdot \mathbf{G}_{*,\mathfrak{b}_i}$*;*
3. *let* A_i *be the set of* $\mathbf{b} \in \mathbb{Z}_{\mathcal{Q}}$ *s.t.* $\mathbf{G}_{*,\mathbf{b}}$ *is linearly dependent on* $\mathbf{G}_{*,\mathfrak{b}_i}$*; then*

$$|A_0| = |A_1| = \cdots = |A_{n-1}| = m/n.$$

Proof. By Corollary 10.4, and the discussion following Corollary 10.4 (the symmetric statements regarding Λ_r and $\mathbf{G}$), there exist vectors $\mathfrak{b}_0 = \mathbf{0}, \mathfrak{b}_1, \ldots, \mathfrak{b}_{n-1} \in \mathbb{Z}_{\mathcal{Q}}$ such that properties 1 and 2 hold. We now prove property 3.

By $(\mathcal{R}_3)$, fixing $\mathfrak{b}_i$ for any i, there is a bijection between A_i and A_0 by $\mathbf{b} \mapsto \mathbf{b} - \mathfrak{b}_i$. This is clear from $\mathbf{G}_{\mathbf{b}-\mathfrak{b}_i,*} = \mathbf{G}_{\mathbf{b},*} \circ \overline{\mathbf{G}_{\mathfrak{b}_i,*}}$. Hence we have $A_0 = \{\mathbf{b} - \mathfrak{b}_i \mid \mathbf{b} \in A_i\}$ for all sets A_i. It then follows that $|A_0| = |A_1| = \cdots = |A_{n-1}| = m/n$. ∎

Now let $G = (V, E)$ be an undirected graph. For each positive integer p we build a new graph $G^{[p]}$ from G by replacing every edge $e = uv \in E$ with a gadget. We need $G^{[2]}$ in the proof but it is more convenient to describe $G^{[1]}$ first and illustrate it only with the case $p = 1$. (The picture for $G^{[2]}$ will be too cumbersome to draw.)

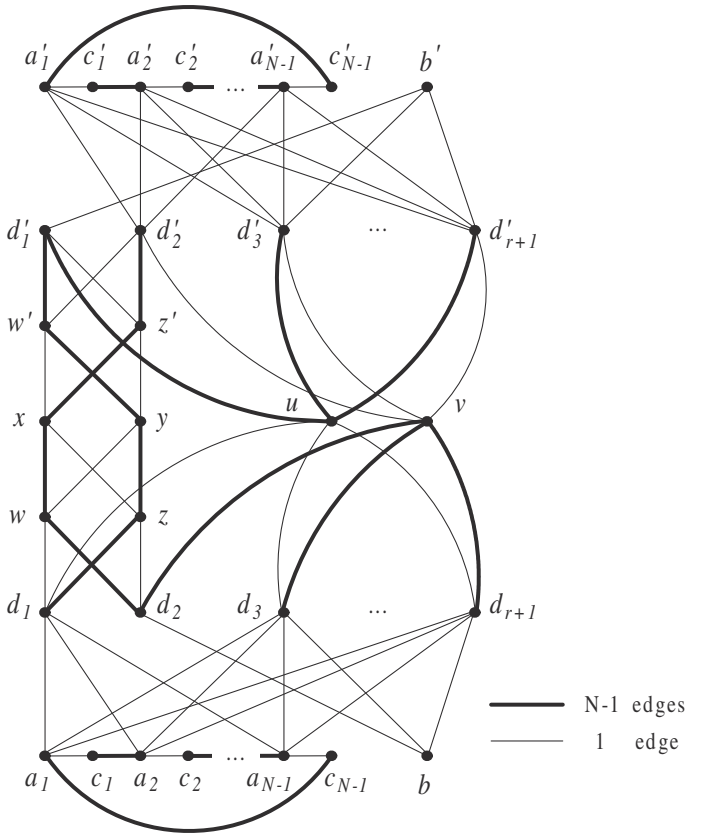


FIG. 11.1. *The gadget for constructing* $G^{[1]}$. *(Note that the subscript* e *is suppressed.)*

The gadget for $G^{[1]}$ is shown in Figure 11.1. Here $G^{[1]} = (V^{[1]}, E^{[1]})$, where

$$V^{[1]} = V \cup \big\{ x_e, y_e, a_{e,i}, a'_{e,i}, b_e, b'_e, c_{e,i}, c'_{e,i}, d_{e,j}, d'_{e,j}, \\ w_e, w'_e, z_e, z'_e : e \in E, i \in [N-1], j \in [r+1] \big\},$$

and $E^{[1]}$ contains exactly the following edges: For every edge $e = uv \in E$, add

1. one edge $(u, d_{e,j})$ for all $j \in [r+1] - \{2\}$;
2. $N-1$ parallel edges $(v, d_{e,j})$ for all $j \in [r+1] - \{1\}$;
3. one edge $(d_{e,1}, w_e)$, $(d_{e,2}, z_e)$, (w_e, y_e), and (z_e, x_e);
4. $N-1$ parallel edges $(d_{e,1}, z_e)$, $(d_{e,2}, w_e)$, (w_e, x_e), and (z_e, y_e);
5. one edge $(a_{e,i}, d_{e,j})$ for all $i \in [N-1]$ and $j \in [r+1] - \{2\}$;
6. one edge $(b_e, d_{e,j})$ for all $j \in [r+1] - \{1\}$;
7. $N-1$ parallel edges $(c_{e,N-1}, a_{e,1})$ and $(c_{e,i}, a_{e,i+1})$ for all $i \in [N-2]$;
8. one edge $(a_{e,i}, c_{e,i})$ for all $i \in [N-1]$;
9. $N-1$ parallel edges $(u, d'_{e,j})$ for all $j \in [r+1] - \{2\}$;
10. one edge $(v, d'_{e,j})$ for all $j \in [r+1] - \{1\}$;
11. one edge $(d'_{e,1}, z'_e)$, $(d'_{e,2}, w'_e)$, (w'_e, x_e), and (z'_e, y_e);
12. $N-1$ parallel edges $(d'_{e,1}, w'_e)$, $(d'_{e,2}, z'_e)$, (w'_e, y_e), and (z'_e, x_e);
13. one edge $(a'_{e,i}, d'_{e,j})$ for all $i \in [N-1]$ and $j \in [r+1] - \{1\}$;
14. one edges $(b'_e, d'_{e,j})$ for all $j \in [r+1] - \{2\}$;

15. $N-1$ parallel edges $(c'_{e,N-1}, a'_{e,1})$ and $(c'_{e,i}, a'_{e,i+1})$ for all $i \in [N-2]$;
16. one edge $(a'_{e,i}, c'_{e,i})$ for all $i \in [N-1]$.

As indicated earlier, the graph we really need in the proof is $G^{[2]}$. The gadget for $G^{[2]}$ consists of two disjoint copies of the gadget for $G^{[1]}$, with the respective copies of the vertices u, v, x, and y in the two copies identified. Given $G = (V, E)$, we use this new gadget to build $G^{[2]}$ by replacing each $e = uv \in E$ with this gadget. The degree of every vertex in $G^{[2]}$ is a 0 (mod N) except both copies of $a_{e,i}, a'_{e,i}, b_e$, and b'_e whose degree is r (mod N).

The construction gives us a $2m \times 2m$ matrix $\mathbf{A}$ such that $Z_{\mathbf{A}}(G) = Z_{\mathbf{C},\mathfrak{D}}(G^{[2]})$ for all G and thus $\mathsf{EVAL}(\mathbf{A})$ ($\leq \mathsf{EVAL}(\mathbf{C}, \mathfrak{D})$) (right now it is not clear whether $\mathbf{A}$ is a symmetric matrix, which we will prove later) is not #P-hard. We index the rows and columns of $\mathbf{A}$ in the same way as we do for $\mathbf{C}$: The first m rows and columns are indexed by $\{0\} \times \mathbb{Z}_{\mathcal{Q}}$ and the last m rows and columns are indexed by $\{1\} \times \mathbb{Z}_{\mathcal{Q}}$. Since $\mathbf{C}$ is the bipartization of $\mathbf{F}$, we have $A_{(0,\mathbf{u}),(1,\mathbf{v})} = A_{(1,\mathbf{u}),(0,\mathbf{v})} = 0$ for all $\mathbf{u}, \mathbf{v} \in \mathbb{Z}_{\mathcal{Q}}$.

Next we analyze the upper-left $m \times m$ block of $\mathbf{A}$. Given $\mathbf{u}, \mathbf{v} \in \mathbb{Z}_{\mathcal{Q}}$, let $A_{\mathbf{u},\mathbf{v},\mathbf{x},\mathbf{y}}$ denote the following sum:

$$\sum_{\substack{\mathbf{a}_1,\ldots,\mathbf{a}_{N-1},\mathbf{b}\in\Lambda_r,\\ \mathbf{d}_1,\mathbf{d}_2\in\mathbb{Z}_{\mathcal{Q}}}} D^{[r]}_{(0,\mathbf{b})} \prod_{i=1}^{N-1} D^{[r]}_{(0,\mathbf{a}_i)} \left(\sum_{\mathbf{w}\in\mathbb{Z}_{\mathcal{Q}}} F_{\mathbf{w},\mathbf{d}_1} F_{\mathbf{w},\mathbf{y}} \overline{F_{\mathbf{w},\mathbf{d}_2} F_{\mathbf{w},\mathbf{x}}} \right) \left(\sum_{\mathbf{z}\in\mathbb{Z}_{\mathcal{Q}}} F_{\mathbf{z},\mathbf{d}_2} F_{\mathbf{z},\mathbf{x}} \overline{F_{\mathbf{z},\mathbf{d}_1} F_{\mathbf{z},\mathbf{y}}} \right)$$
$$\times \left(\prod_{i=1}^{N-2} \sum_{\mathbf{c}_i\in\mathbb{Z}_{\mathcal{Q}}} F_{\mathbf{a}_i,\mathbf{c}_i} \overline{F_{\mathbf{a}_{i+1},\mathbf{c}_i}} \right) \left(\sum_{\mathbf{c}_{N-1}\in\mathbb{Z}_{\mathcal{Q}}} F_{\mathbf{a}_{N-1},\mathbf{c}_{N-1}} \overline{F_{\mathbf{a}_1,\mathbf{c}_{N-1}}} \right)$$
$$\times \left(\prod_{i=3}^{r+1} \sum_{\mathbf{d}_i\in\mathbb{Z}_{\mathcal{Q}}} F_{\mathbf{u},\mathbf{d}_i} F_{\mathbf{b},\mathbf{d}_i} \overline{F_{\mathbf{v},\mathbf{d}_i}} \prod_{j=1}^{N-1} F_{\mathbf{a}_j,\mathbf{d}_i} \right) F_{\mathbf{u},\mathbf{d}_1} \left(\prod_{j=1}^{N-1} F_{\mathbf{a}_j,\mathbf{d}_1} \right) \overline{F_{\mathbf{v},\mathbf{d}_2}} F_{\mathbf{b},\mathbf{d}_2};$$

let $B_{\mathbf{u},\mathbf{v},\mathbf{x},\mathbf{y}}$ denote the following sum:

$$\sum_{\substack{\mathbf{a}_1,\ldots,\mathbf{a}_{N-1},\mathbf{b}\in\Lambda_r,\\ \mathbf{d}_1,\mathbf{d}_2\in\mathbb{Z}_{\mathcal{Q}}}} D^{[r]}_{(0,\mathbf{b})} \prod_{i=1}^{N-1} D^{[r]}_{(0,\mathbf{a}_i)} \left(\sum_{\mathbf{w}\in\mathbb{Z}_{\mathcal{Q}}} F_{\mathbf{w},\mathbf{d}_2} F_{\mathbf{w},\mathbf{x}} \overline{F_{\mathbf{w},\mathbf{d}_1} F_{\mathbf{w},\mathbf{y}}} \right) \left(\sum_{\mathbf{z}\in\mathbb{Z}_{\mathcal{Q}}} F_{\mathbf{z},\mathbf{d}_1} F_{\mathbf{z},\mathbf{y}} \overline{F_{\mathbf{z},\mathbf{d}_2} F_{\mathbf{z},\mathbf{x}}} \right)$$
$$\times \left(\prod_{i=1}^{N-2} \sum_{\mathbf{c}_i\in\mathbb{Z}_{\mathcal{Q}}} F_{\mathbf{a}_i,\mathbf{c}_i} \overline{F_{\mathbf{a}_{i+1},\mathbf{c}_i}} \right) \left(\sum_{\mathbf{c}_{N-1}\in\mathbb{Z}_{\mathcal{Q}}} F_{\mathbf{a}_{N-1},\mathbf{c}_{N-1}} \overline{F_{\mathbf{a}_1,\mathbf{c}_{N-1}}} \right)$$
$$\times \left(\prod_{i=3}^{r+1} \sum_{\mathbf{d}_i\in\mathbb{Z}_{\mathcal{Q}}} F_{\mathbf{v},\mathbf{d}_i} F_{\mathbf{b},\mathbf{d}_i} \overline{F_{\mathbf{u},\mathbf{d}_i}} \prod_{j=1}^{N-1} F_{\mathbf{a}_j,\mathbf{d}_i} \right) F_{\mathbf{v},\mathbf{d}_2} \left(\prod_{j=1}^{N-1} F_{\mathbf{a}_j,\mathbf{d}_2} \right) \overline{F_{\mathbf{u},\mathbf{d}_1}} F_{\mathbf{b},\mathbf{d}_1}.$$

Then we have

$$A_{(0,\mathbf{u}),(0,\mathbf{v})} = \sum_{\mathbf{x},\mathbf{y}\in\mathbb{Z}_{\mathcal{Q}}} A^2_{\mathbf{u},\mathbf{v},\mathbf{x},\mathbf{y}} B^2_{\mathbf{u},\mathbf{v},\mathbf{x},\mathbf{y}}.$$

We simplify $A_{\mathbf{u},\mathbf{v},\mathbf{x},\mathbf{y}}$ first. Since $\mathbf{F}$ is discrete unitary and satisfies $(\mathcal{R}_3)$, we have

$$\sum_{\mathbf{w}\in\mathbb{Z}_{\mathcal{Q}}} F_{\mathbf{w},\mathbf{d}_1} F_{\mathbf{w},\mathbf{y}} \overline{F_{\mathbf{w},\mathbf{d}_2} F_{\mathbf{w},\mathbf{x}}} = \langle \mathbf{F}_{*,\mathbf{d}_1+\mathbf{y}}, \mathbf{F}_{*,\mathbf{d}_2+\mathbf{x}} \rangle = \begin{cases} m & \text{if } \mathbf{d}_1 - \mathbf{d}_2 = \mathbf{x} - \mathbf{y}, \\ 0 & \text{otherwise.} \end{cases}$$

Also when $\mathbf{d}_1 - \mathbf{d}_2 = \mathbf{x} - \mathbf{y}$, we have $\sum_{\mathbf{z}\in\mathbb{Z}_\mathcal{Q}} F_{\mathbf{z},\mathbf{d}_2} F_{\mathbf{z},\mathbf{x}} \overline{F_{\mathbf{z},\mathbf{d}_1} F_{\mathbf{z},\mathbf{y}}} = m$. Similarly,

$$\sum_{\mathbf{c}_i\in\mathbb{Z}_\mathcal{Q}} F_{\mathbf{a}_i,\mathbf{c}_i} \overline{F_{\mathbf{a}_{i+1},\mathbf{c}_i}} = \langle \mathbf{F}_{\mathbf{a}_i,*}, \mathbf{F}_{\mathbf{a}_{i+1},*} \rangle$$

is zero unless $\mathbf{a}_i = \mathbf{a}_{i+1}$ for $i = 1, \ldots, N-2$, and

$$\sum_{\mathbf{c}_{N-1}\in\mathbb{Z}_\mathcal{Q}} F_{\mathbf{a}_{N-1},\mathbf{c}_{N-1}} \overline{F_{\mathbf{a}_1,\mathbf{c}_{N-1}}} = \langle \mathbf{F}_{\mathbf{a}_{N-1},*}, \mathbf{F}_{\mathbf{a}_1,*} \rangle$$

is zero unless $\mathbf{a}_{N-1} = \mathbf{a}_1$. When $\mathbf{a}_1 = \cdots = \mathbf{a}_{N-1}$, all these inner products are equal to m. So now we may assume that $\mathbf{d}_1 - \mathbf{d}_2 = \mathbf{x} - \mathbf{y}$ and all $\mathbf{a}_i$'s are equal, call it $\mathbf{a}$, in the sum for $A_{\mathbf{u},\mathbf{v},\mathbf{x},\mathbf{y}}$. Let $\mathbf{x} - \mathbf{y} = \mathbf{s}$. Then $A_{\mathbf{u},\mathbf{v},\mathbf{x},\mathbf{y}}$ is equal to

(11.1)

$$m^{N+1} \sum_{\substack{\mathbf{a},\mathbf{b}\in\Lambda_r \\ \mathbf{d}_2\in\mathbb{Z}_\mathcal{Q}}} D^{[r]}_{(0,\mathbf{b})} \overline{D^{[r]}_{(0,\mathbf{a})}} \left(\prod_{i=3}^{r+1} \sum_{\mathbf{d}_i\in\mathbb{Z}_\mathcal{Q}} F_{\mathbf{u},\mathbf{d}_i} F_{\mathbf{b},\mathbf{d}_i} \overline{F_{\mathbf{v},\mathbf{d}_i} F_{\mathbf{a},\mathbf{d}_i}} \right) F_{\mathbf{u},\mathbf{d}_2+\mathbf{s}} F_{\mathbf{b},\mathbf{d}_2} \overline{F_{\mathbf{v},\mathbf{d}_2} F_{\mathbf{a},\mathbf{d}_2+\mathbf{s}}}.$$

Again we have

$$\sum_{\mathbf{d}_i\in\mathbb{Z}_\mathcal{Q}} F_{\mathbf{u},\mathbf{d}_i} F_{\mathbf{b},\mathbf{d}_i} \overline{F_{\mathbf{v},\mathbf{d}_i} F_{\mathbf{a},\mathbf{d}_i}} = \langle \mathbf{F}_{\mathbf{u}+\mathbf{b},*}, \mathbf{F}_{\mathbf{v}+\mathbf{a},*} \rangle = \begin{cases} m & \text{if } \mathbf{u}+\mathbf{b} = \mathbf{v}+\mathbf{a}, \\ 0 & \text{otherwise.} \end{cases}$$

If $\mathbf{v} - \mathbf{u} \notin \Lambda_r^{\mathrm{lin}} \equiv \{\mathbf{x} - \mathbf{x}' : \mathbf{x}, \mathbf{x}' \in \Lambda_r\}$, then $A_{\mathbf{u},\mathbf{v},\mathbf{x},\mathbf{y}} = 0$ as $\mathbf{a}, \mathbf{b} \in \Lambda_r$, $\mathbf{b} - \mathbf{a} \in \Lambda_r^{\mathrm{lin}}$.

For every $\mathbf{h} \in \Lambda_r^{\mathrm{lin}}$ (e.g., $\mathbf{h} = \mathbf{v} - \mathbf{u}$), we define a $|\Lambda_r|$-dimensional vector $\mathbf{T}^{[\mathbf{h}]}$:

$$T^{[\mathbf{h}]}_{\mathbf{x}} = D^{[r]}_{(0,\mathbf{x}+\mathbf{h})} \overline{D^{[r]}_{(0,\mathbf{x})}} \quad \text{for all } \mathbf{x} \in \Lambda_r.$$

By $(\mathcal{L})$, Λ_r is a coset in $\mathbb{Z}_\mathcal{Q}$. So for any $\mathbf{x} \in \Lambda_r$, we also have $\mathbf{x} + \mathbf{h} \in \Lambda_r$. Therefore, every entry of $\mathbf{T}^{[\mathbf{h}]}$ is nonzero and is a power of ω_N.

Now we use $\mathbf{T}^{[\mathbf{v}-\mathbf{u}]}$ to express $A_{\mathbf{u},\mathbf{v},\mathbf{x},\mathbf{y}}$. Suppose $\mathbf{v} - \mathbf{u} \in \Lambda_r^{\mathrm{lin}}$; then

$$\begin{aligned} A_{\mathbf{u},\mathbf{v},\mathbf{x},\mathbf{y}} &= m^{N+r} \sum_{\substack{\mathbf{a}\in\Lambda_r, \mathbf{d}_2\in\mathbb{Z}_\mathcal{Q} \\ \mathbf{b}=\mathbf{a}+\mathbf{v}-\mathbf{u}}} D^{[r]}_{(0,\mathbf{b})} \overline{D^{[r]}_{(0,\mathbf{a})}} F_{\mathbf{u},\mathbf{d}_2+\mathbf{s}} F_{\mathbf{b},\mathbf{d}_2} \overline{F_{\mathbf{v},\mathbf{d}_2} F_{\mathbf{a},\mathbf{d}_2+\mathbf{s}}} \\ &= m^{N+r+1} \sum_{\mathbf{a}\in\Lambda_r} D^{[r]}_{(0,\mathbf{a}+\mathbf{v}-\mathbf{u})} \overline{D^{[r]}_{(0,\mathbf{a})}} F_{\mathbf{u},\mathbf{s}} \overline{F_{\mathbf{a},\mathbf{s}}} \\ &= m^{N+r+1} \cdot F_{\mathbf{u},\mathbf{x}-\mathbf{y}} \langle \mathbf{T}^{[\mathbf{v}-\mathbf{u}]}, \mathbf{G}_{*,\mathbf{x}-\mathbf{y}} \rangle. \end{aligned}$$

Here we used $(\mathcal{R}_3)$ in the second equality, and we recall the definition of $\mathbf{s} = \mathbf{x} - \mathbf{y}$.

Similarly, when $\mathbf{v} - \mathbf{u} \notin \Lambda_r^{\mathrm{lin}}$, we have $B_{\mathbf{u},\mathbf{v},\mathbf{x},\mathbf{y}} = 0$, and when $\mathbf{v} - \mathbf{u} \in \Lambda_r^{\mathrm{lin}}$,

$$\begin{aligned} B_{\mathbf{u},\mathbf{v},\mathbf{x},\mathbf{y}} &= m^{N+r} \sum_{\substack{\mathbf{b}\in\Lambda_r, \mathbf{d}_2\in\mathbb{Z}_\mathcal{Q} \\ \mathbf{a}=\mathbf{b}+\mathbf{v}-\mathbf{u}}} D^{[r]}_{(0,\mathbf{b})} \overline{D^{[r]}_{(0,\mathbf{a})}} F_{\mathbf{v},\mathbf{d}_2} F_{\mathbf{b},\mathbf{d}_2+\mathbf{x}-\mathbf{y}} \overline{F_{\mathbf{a},\mathbf{d}_2} F_{\mathbf{u},\mathbf{d}_2+\mathbf{x}-\mathbf{y}}} \\ &= m^{N+r+1} \cdot \overline{F_{\mathbf{u},\mathbf{x}-\mathbf{y}} \langle \mathbf{T}^{[\mathbf{v}-\mathbf{u}]}, \mathbf{G}_{*,\mathbf{x}-\mathbf{y}} \rangle}. \end{aligned}$$

To summarize, when $\mathbf{v} - \mathbf{u} \notin \Lambda_r^{\mathrm{lin}}$, $A_{(0,\mathbf{u}),(0,\mathbf{v})} = 0$, and when $\mathbf{v} - \mathbf{u} \in \Lambda_r^{\mathrm{lin}}$,

(11.2)

$$A_{(0,\mathbf{u}),(0,\mathbf{v})} = m^{4(N+r+1)} \sum_{\mathbf{x},\mathbf{y}\in\mathbb{Z}_\mathcal{Q}} \left| \langle \mathbf{T}^{[\mathbf{v}-\mathbf{u}]}, \mathbf{G}_{*,\mathbf{x}-\mathbf{y}} \rangle \right|^4 = m^{4N+4r+5} \sum_{\mathbf{b}\in\mathbb{Z}_\mathcal{Q}} \left| \langle \mathbf{T}^{[\mathbf{v}-\mathbf{u}]}, \mathbf{G}_{*,\mathbf{b}} \rangle \right|^4.$$

We now show that $\mathbf{A}$ is symmetric. Let $\mathbf{a} = \mathbf{v} - \mathbf{u} \in \Lambda_r^{\mathrm{lin}}$. By $(\mathcal{R}_3)$, for $\mathbf{b} \in \mathbb{Z}_{\mathcal{Q}}$,

$$\left|\langle \mathbf{T}^{[-\mathbf{a}]}, \mathbf{G}_{*,-\mathbf{b}}\rangle\right| = \left|\sum_{\mathbf{x}\in\Lambda_r} D^{[r]}_{(0,\mathbf{x}-\mathbf{a})}\overline{D^{[r]}_{(0,\mathbf{x})}G_{\mathbf{x},-\mathbf{b}}}\right| = \left|\sum_{\mathbf{x}\in\Lambda_r} D^{[r]}_{(0,\mathbf{x})}\overline{D^{[r]}_{(0,\mathbf{x}-\mathbf{a})}G_{\mathbf{x},\mathbf{b}}}\right|$$
$$= \left|\sum_{\mathbf{y}\in\Lambda_r} D^{[r]}_{(0,\mathbf{y}+\mathbf{a})}\overline{D^{[r]}_{(0,\mathbf{y})}G_{\mathbf{y},\mathbf{b}}F_{\mathbf{a},\mathbf{b}}}\right| = \left|\sum_{\mathbf{y}\in\Lambda_r} D^{[r]}_{(0,\mathbf{y}+\mathbf{a})}\overline{D^{[r]}_{(0,\mathbf{y})}G_{\mathbf{y},\mathbf{b}}}\right| = \left|\langle \mathbf{T}^{[\mathbf{a}]}, \mathbf{G}_{*,\mathbf{b}}\rangle\right|,$$

where the second equality is by conjugation, the third equality is by the substitution $\mathbf{x} = \mathbf{y}+\mathbf{a}$, and the fourth equality is because $F_{\mathbf{a},\mathbf{b}}$ is a root of unity. Thus, $A_{(0,\mathbf{u}),(0,\mathbf{v})} = A_{(0,\mathbf{v}),(0,\mathbf{u})}$. The lower-right block can be proved similarly. Hence $\mathbf{A}$ is symmetric.

Next, we further simplify (11.2) using Lemma 11.1:

$$A_{(0,\mathbf{u}),(0,\mathbf{v})} = \frac{m^{4N+4r+6}}{n} \cdot \sum_{i=0}^{n-1} \left|\langle \mathbf{T}^{[\mathbf{v}-\mathbf{u}]}, \mathbf{G}_{*,\mathfrak{b}_i}\rangle\right|^4. \tag{11.3}$$

For the special case of $\mathbf{u} = \mathbf{v}$, since $\mathbf{T}^{[\mathbf{0}]} = \mathbf{1} = \mathbf{G}_{*,\mathfrak{b}_0}$ and $\{\mathbf{G}_{*,\mathfrak{b}_0}, \ldots, \mathbf{G}_{*,\mathfrak{b}_{n-1}}\}$ is an orthogonal basis by Lemma 11.1, we have

$$\sum_{i=0}^{n-1} \left|\langle \mathbf{T}^{[\mathbf{0}]}, \mathbf{G}_{*,\mathfrak{b}_i}\rangle\right|^4 = n^4 \quad \text{and} \quad A_{(0,\mathbf{u}),(0,\mathbf{u})} = L \cdot n^4, \quad \text{where } L \equiv m^{4N+4r+6}/n.$$

Our next goal is to prove for all $\mathbf{a} \in \Lambda_r^{\mathrm{lin}}$ that there exist $\mathbf{b} \in \mathbb{Z}_{\mathcal{Q}}$, $\alpha \in \mathbb{Z}_N$ such that

$$\mathbf{T}^{[\mathbf{a}]} = \omega_N^{\alpha} \cdot \mathbf{G}_{*,\mathbf{b}}. \tag{11.4}$$

If $|\Lambda_r^{\mathrm{lin}}| = 1$, then (11.4) is trivially true. Thus below we assume $|\Lambda_r^{\mathrm{lin}}| > 1$. Because $\mathbf{A}$ is symmetric and nonnegative, we can apply the dichotomy theorem of Bulatov and Grohe. For any pair $\mathbf{u} \neq \mathbf{v}$ such that $\mathbf{u} - \mathbf{v} \in \Lambda_r^{\mathrm{lin}}$, we consider the 2×2 submatrix

$$\begin{pmatrix} A_{(0,\mathbf{u}),(0,\mathbf{u})} & A_{(0,\mathbf{u}),(0,\mathbf{v})} \\ A_{(0,\mathbf{v}),(0,\mathbf{u})} & A_{(0,\mathbf{v}),(0,\mathbf{v})} \end{pmatrix}$$

of $\mathbf{A}$. Since $\mathsf{EVAL}(\mathbf{A})$ is assumed to be not #P-hard, by Corollary 2.6, we have

$$A_{(0,\mathbf{u}),(0,\mathbf{v})} = A_{(0,\mathbf{v}),(0,\mathbf{u})} \in \{0, L \cdot n^4\},$$

and thus from (11.3) we get

$$\sum_{i=0}^{n-1} \left|\langle \mathbf{T}^{[\mathbf{v}-\mathbf{u}]}, \mathbf{G}_{*,\mathfrak{b}_i}\rangle\right|^4 \in \{0, n^4\} \quad \text{for all } \mathbf{u}, \mathbf{v} \text{ such that } \mathbf{u} - \mathbf{v} \in \Lambda_r^{\mathrm{lin}}. \tag{11.5}$$

However, the sum in (11.5) cannot be zero, because by Lemma 11.1, $\{\mathbf{G}_{*,\mathfrak{b}_i} : i \in [0 : n-1]\}$ is an orthogonal basis with each $\|\mathbf{G}_{*,\mathfrak{b}_i}\|^2 = n$. Then by Parseval,

$$\sum_{i=0}^{n-1} \left|\langle \mathbf{T}^{[\mathbf{v}-\mathbf{u}]}, \frac{\mathbf{G}_{*,\mathfrak{b}_i}}{\|\mathbf{G}_{*,\mathfrak{b}_i}\|}\rangle\right|^2 = \|\mathbf{T}^{[\mathbf{v}-\mathbf{u}]}\|^2 = n,$$

as each entry of $\mathbf{T}^{[\mathbf{v}-\mathbf{u}]}$ is a root of unity. Hence $\sum_{i=0}^{n-1} |\langle \mathbf{T}^{[\mathbf{v}-\mathbf{u}]}, \mathbf{G}_{*,\mathfrak{b}_i}\rangle|^2 = n^2$. This shows that for some $0 \le i < n$, $|\langle \mathbf{T}^{[\mathbf{v}-\mathbf{u}]}, \mathbf{G}_{*,\mathfrak{b}_i}\rangle| \neq 0$, and therefore the sum in (11.5) is nonzero, and thus in fact

$$\sum_{i=0}^{n-1} \left|\langle \mathbf{T}^{[\mathbf{v}-\mathbf{u}]}, \mathbf{G}_{*,\mathbf{b}_i}\rangle\right|^4 = n^4 \quad \text{for all } \mathbf{u},\mathbf{v} \text{ such that } \mathbf{u}-\mathbf{v} \in \Lambda_r^{\text{lin}}.$$

If we temporarily denote $x_i = |\langle \mathbf{T}^{[\mathbf{v}-\mathbf{u}]}, \mathbf{G}_{*,\mathbf{b}_i}\rangle|$ for $0 \le i < n$, then each $x_i \ge 0$. We have both $\sum_{i=0}^{n-1} x_i^2 = n^2$ and $\sum_{i=0}^{n-1} x_i^4 = n^4$. By taking the square, we have

$$n^4 = \left(\sum_{i=0}^{n-1} x_i^2\right)^2 = \sum_{i=0}^{n-1} x_i^4 + \text{nonnegative cross terms.}$$

It follows that all cross terms must be zero. Thus, there exists a unique term $x_i \neq 0$. Moreover, this x_i must equal to n, while all other $x_j = 0$. We conclude that for all $\mathbf{u}$ and $\mathbf{v} \in \mathbb{Z}_{\mathcal{Q}}$ such that $\mathbf{u}-\mathbf{v} \in \Lambda_r^{\text{lin}}$, there exists a unique $i \in [0:n-1]$ such that

$$\left|\langle \mathbf{T}^{[\mathbf{v}-\mathbf{u}]}, \mathbf{G}_{*,\mathbf{b}_i}\rangle\right| = n.$$

Applying again the argument that $\langle \mathbf{T}^{[\mathbf{v}-\mathbf{u}]}, \mathbf{G}_{*,\mathbf{b}_i}\rangle$ is a sum of n terms, each of which is a root of unity, (11.4) follows.

Below we use (11.4) to prove $(\mathcal{D}_3)$. Note that if $s = 1$, then $(\mathcal{D}_3)$ follows directly from (11.4). Thus below we assume $s > 1$. First, (11.4) implies the following lemma.

LEMMA 11.2. *Let* $\mathbf{a} \in \Lambda_{r,k}^{\text{lin}}$ *for some* $k \in [s]$. *Then for any* $\ell \neq k$ *and* $\mathbf{c} \in \Lambda_{r,\ell}^{\text{lin}}$,

$$T_{\mathbf{x}+\widetilde{\mathbf{c}}}^{[\widetilde{\mathbf{a}}]} \Big/ T_{\mathbf{x}}^{[\widetilde{\mathbf{a}}]}$$

is a power of ω_{q_ℓ} *for all* $\mathbf{x} \in \Lambda_r$.

Recall that $q_\ell = q_{\ell,1}$. Also note that for every $\mathbf{x} \in \Lambda_r$, the translated point $\mathbf{x}+\widetilde{\mathbf{c}}$ is in Λ_r, so $\mathbf{T}^{[\widetilde{\mathbf{a}}]}$ is defined at both $\mathbf{x}$ and $\mathbf{x}+\widetilde{\mathbf{c}}$. Since they are roots of unity, we can divide one by the other.

Proof. By (11.4), there exists a vector $\mathbf{b} \in \mathbb{Z}_{\mathcal{Q}}$ such that

$$T_{\mathbf{x}+\widetilde{\mathbf{c}}}^{[\widetilde{\mathbf{a}}]} \Big/ T_{\mathbf{x}}^{[\widetilde{\mathbf{a}}]} = G_{\mathbf{x}+\widetilde{\mathbf{c}},\mathbf{b}} \Big/ G_{\mathbf{x},\mathbf{b}} = F_{\widetilde{\mathbf{c}},\mathbf{b}},$$

which, by $(\mathcal{R}_3)$, must be a power of ω_{q_ℓ}. ∎

Let $\mathbf{a} \in \Lambda_{r,k}^{\text{lin}}$ and $\mathbf{c} \in \Lambda_{r,\ell}^{\text{lin}}$, $\ell \neq k \in [s]$. By the definition of $T_{\mathbf{x}}^{[\mathbf{h}]}$ in terms of $D_*^{[r]}$,

$$T_{\mathbf{x}+\widetilde{\mathbf{a}}}^{[\widetilde{\mathbf{c}}]} \cdot T_{\mathbf{x}}^{[\widetilde{\mathbf{a}}]} = T_{\mathbf{x}}^{[\widetilde{\mathbf{a}}+\widetilde{\mathbf{c}}]} = T_{\mathbf{x}+\widetilde{\mathbf{c}}}^{[\widetilde{\mathbf{a}}]} \cdot T_{\mathbf{x}}^{[\widetilde{\mathbf{c}}]},$$

and thus

$$T_{\mathbf{x}+\widetilde{\mathbf{a}}}^{[\widetilde{\mathbf{c}}]} \Big/ T_{\mathbf{x}}^{[\widetilde{\mathbf{c}}]} = T_{\mathbf{x}+\widetilde{\mathbf{c}}}^{[\widetilde{\mathbf{a}}]} \Big/ T_{\mathbf{x}}^{[\widetilde{\mathbf{a}}]}.$$

By Lemma 11.2, the left-hand side of the equation is a power of ω_{q_k}, while the right-hand side of the equation is a power of ω_{q_ℓ}. Since $k \neq \ell$, $\gcd(q_k, q_\ell) = 1$, so

$$T_{\mathbf{x}+\widetilde{\mathbf{c}}}^{[\widetilde{\mathbf{a}}]} \Big/ T_{\mathbf{x}}^{[\widetilde{\mathbf{a}}]} = 1 \quad \text{for all } \mathbf{c} \in \Lambda_{r,\ell}^{\text{lin}} \text{ such that } \ell \neq k. \tag{11.6}$$

This implies that $T_{\mathbf{x}}^{[\widetilde{\mathbf{a}}]}$, as a function of $\mathbf{x}$, only depends on $\mathbf{x}_k \in \Lambda_{r,k}$. By (11.4),

$$T_{\mathbf{x}}^{[\widetilde{\mathbf{a}}]} = T_{\mathbf{ext}_r(\mathbf{x}_k)}^{[\widetilde{\mathbf{a}}]} = \omega_N^{\alpha} \cdot G_{\mathbf{ext}_r(\mathbf{x}_k),\mathbf{b}} = \omega_N^{\alpha+\beta} \cdot F_{\widetilde{\mathbf{x}_k},\widetilde{\mathbf{b}_k}} = \omega_N^{\alpha+\beta} \cdot F_{\mathbf{x},\widetilde{\mathbf{b}_k}}$$

for any $\mathbf{x} \in \Lambda_r$ and for some constants $\alpha, \beta \in \mathbb{Z}_N$, and $\mathbf{b}_k \in \mathbb{Z}_{\mathbf{q}_k}$ that are independent of $\mathbf{x}$. This proves condition $(\mathcal{D}_3)$.

Finally we prove $(\mathcal{D}_1)$ from $(\mathcal{D}_3)$. Let $\mathfrak{a}^{[r]} = (\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_s) \in \Lambda_r$. Then

$$\begin{aligned} D^{[r]}_{(0,\mathbf{x})} &= D^{[r]}_{(0,(\mathbf{x}_1,\mathbf{x}_2,\dots,\mathbf{x}_s))}\overline{D^{[r]}_{(0,(\mathbf{a}_1,\mathbf{a}_2,\dots,\mathbf{a}_s))}} \\ &= \left(D^{[r]}_{(0,(\mathbf{x}_1,\mathbf{x}_2,\dots,\mathbf{x}_{s-1},\mathbf{x}_s))}\overline{D^{[r]}_{(0,(\mathbf{x}_1,\mathbf{x}_2,\dots,\mathbf{x}_{s-1},\mathbf{a}_s))}}\right) \\ &\quad \times \left(D^{[r]}_{(0,(\mathbf{x}_1,\mathbf{x}_2,\dots,\mathbf{x}_{s-1},\mathbf{a}_s))}\overline{D^{[r]}_{(0,(\mathbf{x}_1,\dots,\mathbf{x}_{s-2},\mathbf{a}_{s-1},\mathbf{a}_s))}}\right)\cdots \\ &\quad \times \left(D^{[r]}_{(0,(\mathbf{x}_1,\mathbf{a}_2,\dots,\mathbf{a}_s))}\overline{D^{[r]}_{(0,(\mathbf{a}_1,\mathbf{a}_2,\dots,\mathbf{a}_s))}}\right) \quad \text{for any } \mathbf{x} \in \Lambda_r. \end{aligned}$$

We consider the kth factor

$$D^{[r]}_{(0,(\mathbf{x}_1,\dots,\mathbf{x}_{k-1},\mathbf{x}_k,\mathbf{a}_{k+1},\dots,\mathbf{a}_s))}\overline{D^{[r]}_{(0,(\mathbf{x}_1,\dots,\mathbf{x}_{k-1},\mathbf{a}_k,\mathbf{a}_{k+1},\dots,\mathbf{a}_s))}}. \tag{11.7}$$

From (11.6) this factor is independent of all other components in the starting point $(\mathbf{x}_1, \dots, \mathbf{x}_{k-1}, \mathbf{a}_k, \mathbf{a}_{k+1}, \dots, \mathbf{a}_s)$ except the kth component $\mathbf{a}_k$. In particular, we can replace all other components, as long as we stay within Λ_r. We choose to replace the first $k-1$ components $\mathbf{x}_i$ by $\mathbf{a}_i$. Then (11.7) becomes

$$\begin{aligned} &D^{[r]}_{(0,(\mathbf{a}_1,\dots,\mathbf{a}_{k-1},\mathbf{x}_k,\mathbf{a}_{k+1},\dots,\mathbf{a}_s))}\overline{D^{[r]}_{(0,(\mathbf{a}_1,\dots,\mathbf{a}_{k-1},\mathbf{a}_k,\mathbf{a}_{k+1},\dots,\mathbf{a}_s))}} \\ &= D^{[r]}_{(0,\mathbf{ext}_r(\mathbf{x}_k))}\overline{D^{[r]}_{(0,\mathfrak{a}^{[r]})}} = D^{[r]}_{(0,\mathbf{ext}_r(\mathbf{x}_k))}, \end{aligned}$$

and $(\mathcal{D}_1)$ is now proved.

12. Tractability: Proof of Theorem 5.10. Let $((M,N), \mathbf{C}, \mathfrak{D}, (\mathbf{p}, \mathbf{t}, \mathcal{Q}))$ be a tuple that satisfies $(\mathcal{R}), (\mathcal{L}), (\mathcal{D})$. In this section, we finally show that $\mathsf{EVAL}(\mathbf{C}, \mathfrak{D})$ is tractable by reducing it to the following problem. Let $q = p^k$ be a prime power for some prime p and positive integer k. The input of $\mathsf{EVAL}(q)$ is a quadratic polynomial $f(x_1, x_2, \dots, x_n) = \sum_{i,j\in[n]} a_{i,j} x_i x_j$, where $a_{i,j} \in \mathbb{Z}_q$ for all i, j, and the output is

$$Z_q(f) = \sum_{x_1,\dots,x_n \in \mathbb{Z}_q} \omega_q^{f(x_1,\dots,x_n)}.$$

We postpone the proof of the following theorem to the end of this section.

THEOREM 12.1. *Let q be a prime power. Then $\mathsf{EVAL}(q)$ can be solved in polynomial time (in n, the number of variables).*

The reduction goes as follows. First, we use conditions $(\mathcal{R}), (\mathcal{L})$, and $(\mathcal{D})$ to show that $\mathsf{EVAL}(\mathbf{C}, \mathfrak{D})$ can be decomposed into s smaller problems, where s is the number of primes in the sequence $\mathbf{p}$: $\mathsf{EVAL}(\mathbf{C}^{[1]}, \mathfrak{D}^{[1]}), \dots, \mathsf{EVAL}(\mathbf{C}^{[s]}, \mathfrak{D}^{[s]})$. If each of these s problems is tractable, then so is $\mathsf{EVAL}(\mathbf{C}, \mathfrak{D})$. Second, we reduce each $\mathsf{EVAL}(\mathbf{C}^{[i]}, \mathfrak{D}^{[i]})$ to $\mathsf{EVAL}(q)$ for some appropriate prime power q that will become clear later. It follows from Theorem 12.1 that all $\mathsf{EVAL}(\mathbf{C}^{[i]}, \mathfrak{D}^{[i]})$'s can be solved in polynomial time.

12.1. Step 1. For each integer $i \in [s]$, we define a $2m_i \times 2m_i$ matrix $\mathbf{C}^{[i]}$ where $m_i = |\mathbb{Z}_{\mathbf{q}_i}|$: $\mathbf{C}^{[i]}$ is the bipartization of the following $m_i \times m_i$ matrix $\mathbf{F}^{[i]}$, where

$$F^{[i]}_{\mathbf{x},\mathbf{y}} = \prod_{j\in[t_i]} \omega_{q_{i,j}}^{x_j y_j} \quad \text{for all } \mathbf{x} = (x_1, \dots, x_{t_i}), \mathbf{y} = (y_1, \dots, y_{t_i}) \in \mathbb{Z}_{\mathbf{q}_i}. \tag{12.1}$$

We index the rows and columns of $\mathbf{F}^{[i]}$ by $\mathbf{x} \in \mathbb{Z}_{\mathbf{q}_i}$ and index the rows and columns of $\mathbf{C}^{[i]}$ by $\{0,1\} \times \mathbb{Z}_{\mathbf{q}_i}$. We let x_j, $j \in [t_i]$, denote the jth entry of $\mathbf{x} \in \mathbb{Z}_{q_{i,j}}$. By $(\mathcal{R}_3)$,

$$F_{\mathbf{x},\mathbf{y}} = F^{[1]}_{\mathbf{x}_1,\mathbf{y}_1} \cdot F^{[2]}_{\mathbf{x}_2,\mathbf{y}_2} \cdots F^{[s]}_{\mathbf{x}_s,\mathbf{y}_s} \quad \text{for all } \mathbf{x},\mathbf{y} \in \mathbb{Z}_{\mathcal{Q}}. \tag{12.2}$$

For each integer $i \in [s]$, we define a sequence of N $2m_i \times 2m_i$ diagonal matrices

$$\mathfrak{D}^{[i]} = \left(\mathbf{D}^{[i,0]}, \ldots, \mathbf{D}^{[i,N-1]}\right).$$

$\mathbf{D}^{[i,0]}$ is the $2m_i \times 2m_i$ identity matrix; for every $r \in [N-1]$, we set

$$\mathbf{D}^{[i,r]}_{(0,*)} = \mathbf{0} \text{ if } r \notin \mathcal{S} \quad \text{and} \quad D^{[i,r]}_{(0,\mathbf{x})} = D^{[r]}_{(0,\mathbf{ext}_r(\mathbf{x}))} \text{ for all } \mathbf{x} \in \mathbb{Z}_{\mathbf{q}_i} \text{ if } r \in \mathcal{S};$$

$$\mathbf{D}^{[i,r]}_{(1,*)} = \mathbf{0} \text{ if } r \notin \mathcal{T} \quad \text{and} \quad D^{[i,r]}_{(1,\mathbf{x})} = D^{[r]}_{(1,\mathbf{ext}'_r(\mathbf{x}))} \text{ for all } \mathbf{x} \in \mathbb{Z}_{\mathbf{q}_i} \text{ if } r \in \mathcal{T}.$$

By conditions $(\mathcal{D}_1)$ and $(\mathcal{D}_2)$, we have

$$D^{[r]}_{(b,\mathbf{x})} = D^{[1,r]}_{(b,\mathbf{x}_1)} \cdots D^{[s,r]}_{(b,\mathbf{x}_s)} \quad \text{for all } b \in \{0,1\} \text{ and } \mathbf{x} \in \mathbb{Z}_{\mathcal{Q}}. \tag{12.3}$$

Equation (12.3) is valid for all $\mathbf{x} \in \mathbb{Z}_{\mathcal{Q}}$. For example, for $b = 0$ and $\mathbf{x} \in \mathbb{Z}_{\mathcal{Q}} - \Lambda_r$, the left-hand side is 0 because $\mathbf{x} \notin \Lambda_r$. The right-hand side is also 0, because there exists an index $i \in [s]$ such that $\mathbf{x}_i \notin \Lambda_{r,i}$ and thus $\mathbf{ext}_r(\mathbf{x}_i) \notin \Lambda_r$. It then follows from (12.1), (12.3), and the following lemma that if $\mathsf{EVAL}(\mathbf{C}^{[i]}, \mathfrak{D}^{[i]})$ is in polynomial time for all $i \in [s]$, then $\mathsf{EVAL}(\mathbf{C}, \mathfrak{D})$ is also in polynomial time.

LEMMA 12.2. *Suppose we have the following matrices: for each $i \in \{0,1,2\}$, $\mathbf{C}^{[i]}$ is the bipartization of an $m_i \times m_i$ complex matrix $\mathbf{F}^{[i]}$; $\mathfrak{D}^{[i]} = (\mathbf{D}^{[i,0]}, \ldots, \mathbf{D}^{[i,N-1]})$ is a sequence of N $2m_i \times 2m_i$ diagonal matrices for some $N \geq 1$, where*

$$\mathbf{D}^{[i,r]} = \begin{pmatrix} \mathbf{P}^{[i,r]} & \\ & \mathbf{Q}^{[i,r]} \end{pmatrix}$$

and $\mathbf{P}^{[i,r]}$ and $\mathbf{Q}^{[i,r]}$ are $m_i \times m_i$ diagonal matrices; $(\mathbf{C}^{[i]}, \mathfrak{D}^{[i]})$ satisfies (Pinning);

$$\mathbf{F}^{[0]} = \mathbf{F}^{[1]} \otimes \mathbf{F}^{[2]}, \quad \mathbf{P}^{[0,r]} = \mathbf{P}^{[1,r]} \otimes \mathbf{P}^{[2,r]} \quad \textit{and} \quad \mathbf{Q}^{[0,r]} = \mathbf{Q}^{[1,r]} \otimes \mathbf{Q}^{[2,r]}$$

for all $r \in [0 : N-1]$ (so $m_0 = m_1 m_2$). If $\mathsf{EVAL}(\mathbf{C}^{[1]}, \mathfrak{D}^{[1]})$ and $\mathsf{EVAL}(\mathbf{C}^{[2]}, \mathfrak{D}^{[2]})$ are tractable, then $\mathsf{EVAL}(\mathbf{C}^{[0]}, \mathfrak{D}^{[0]})$ is also tractable.

Proof. By the second pinning lemma (Lemma 4.3), both functions $Z^{\rightarrow}$ and $Z^{\leftarrow}$ of $(\mathbf{C}^{[i]}, \mathfrak{D}^{[i]})$, for both $i = 1$ and 2, can be computed in polynomial time. The lemma then follows from Lemma 2.4. ∎

We now use condition $(\mathcal{D}_4)$ to prove the following lemma.

LEMMA 12.3. *Given $r \in \mathcal{T}, i \in [s]$, and $\mathbf{a} \in \Delta^{\mathrm{lin}}_{r,i}$, there exist $\mathbf{b} \in \mathbb{Z}_{\mathbf{q}_i}$ and $\alpha \in \mathbb{Z}_N$ such that the following equation holds for all $\mathbf{x} \in \Delta_{r,i}$:*

$$D^{[i,r]}_{(1,\mathbf{x}+\mathbf{a})} \cdot \overline{D^{[i,r]}_{(1,\mathbf{x})}} = \omega^{\alpha}_N \cdot F^{[i]}_{\mathbf{b},\mathbf{x}}.$$

Proof. By the definition of $\mathbf{D}^{[i,r]}$, we have

$$D^{[i,r]}_{(1,\mathbf{x}+\mathbf{a})} \cdot \overline{D^{[i,r]}_{(1,\mathbf{x})}} = D^{[r]}_{(1,\mathbf{ext}'_r(\mathbf{x}+\mathbf{a}))} \cdot \overline{D^{[r]}_{(1,\mathbf{ext}'_r(\mathbf{x}))}} = D^{[r]}_{(1,\mathbf{ext}'_r(\mathbf{x})+\widetilde{\mathbf{a}})} \cdot \overline{D^{[r]}_{(1,\mathbf{ext}'_r(\mathbf{x}))}}.$$

Recall that $\widetilde{\mathbf{a}}$ is the vector in $\mathbb{Z}_{\mathcal{Q}}$ such that $\widetilde{\mathbf{a}}_i = \mathbf{a}$ and $\widetilde{\mathbf{a}}_j = \mathbf{0}$ for all other $j \neq i$.

Then by condition $(\mathcal{D}_4)$, we know there exist $\mathbf{b} \in \mathbb{Z}_{\mathbf{q}_i}$ and $\alpha \in \mathbb{Z}_N$ such that

$$D^{[i,r]}_{(1,\mathbf{x}+\mathbf{a})} \cdot \overline{D^{[i,r]}_{(1,\mathbf{x})}} = \omega^{\alpha}_N \cdot F^{[i]}_{\widetilde{\mathbf{b}},\mathbf{ext}'_r(\mathbf{x})} = \omega^{\alpha}_N \cdot F^{[i]}_{\mathbf{b},\mathbf{x}} \quad \text{for all } \mathbf{x} \in \Delta_{r,i},$$

and the lemma is proved. ∎

One can also prove a similar lemma for the other block of $\mathbf{D}^{[i,r]}$, using $(\mathcal{D}_3)$.

12.2. Step 2. For convenience, in this step we abuse the notation slightly and use $\mathsf{EVAL}(\mathbf{C}, \mathfrak{D})$ to denote one of the subproblems $\mathsf{EVAL}(\mathbf{C}^{[i]}, \mathfrak{D}^{[i]})$, $i \in [s]$, defined in the last step. Then by using conditions $(\mathcal{R}), (\mathcal{L})$, and $(\mathcal{D})$, we summarize the properties of this new pair $(\mathbf{C}, \mathfrak{D})$ that we need in the reduction as follows:

$(\mathcal{F}_1)$ There is a prime p and a nonincreasing sequence $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_h)$ of powers of the same p. $\mathbf{F}$ is an $m \times m$ complex matrix, where $m = \pi_1\pi_2 \cdots \pi_h$, and $\mathbf{C}$ is the bipartization of $\mathbf{F}$. We let π denote π_1. We also use $\mathbb{Z}_{\boldsymbol{\pi}} \equiv \mathbb{Z}_{\pi_1} \times \cdots \times \mathbb{Z}_{\pi_h}$ to index the rows and columns of $\mathbf{F}$. Then $\mathbf{F}$ satisfies

$$F_{\mathbf{x},\mathbf{y}} = \prod_{i \in [h]} \omega_{\pi_i}^{x_i y_i} \quad \text{for all } \mathbf{x} = (x_1, \ldots, x_h) \text{ and } \mathbf{y} = (y_1, \ldots, y_h) \in \mathbb{Z}_{\boldsymbol{\pi}},$$

where we use $x_i \in \mathbb{Z}_{\pi_i}$ to denote the ith entry of $\mathbf{x}$, $i \in [h]$.

$(\mathcal{F}_2)$ $\mathfrak{D} = (\mathbf{D}^{[0]}, \ldots, \mathbf{D}^{[N-1]})$ is a sequence of N $2m \times 2m$ diagonal matrices for some $N \geq 1$ with $\pi \mid N$. $\mathbf{D}^{[0]}$ is the identity matrix, and every diagonal entry of $\mathbf{D}^{[r]}$, $r \in [N-1]$, is either 0 or a power of ω_N. We use $\{0,1\} \times \mathbb{Z}_{\boldsymbol{\pi}}$ to index the rows and columns of matrices $\mathbf{C}$ and $\mathbf{D}^{[r]}$. (The condition $\pi \mid N$ is from the condition $M \mid N$ in $(\mathcal{U}_1)$ and the expression of M in terms of the prime powers, stated after $(\mathcal{R}_3)$. The π here is one of the $q_i = q_{i,1}$ there.)

$(\mathcal{F}_3)$ For each $r \in [0 : N-1]$, we use Λ_r and Δ_r to denote

$$\Lambda_r = \{\mathbf{x} \in \mathbb{Z}_{\boldsymbol{\pi}} \mid D^{[r]}_{(0,\mathbf{x})} \neq 0\} \quad \text{and} \quad \Delta_r = \{\mathbf{x} \in \mathbb{Z}_{\boldsymbol{\pi}} \mid D^{[r]}_{(1,\mathbf{x})} \neq 0\}.$$

We use $\mathcal{S}$ to denote the set of r such that $\Lambda_r \neq \emptyset$ and $\mathcal{T}$ to denote the set of r such that $\Delta_r \neq \emptyset$. Then for every $r \in \mathcal{S}$, Λ_r is a coset in $\mathbb{Z}_{\boldsymbol{\pi}}$; for every $r \in \mathcal{T}$, Δ_r is a coset in $\mathbb{Z}_{\boldsymbol{\pi}}$. For each $r \in \mathcal{S}$ (and $r \in \mathcal{T}$), there is an $\mathfrak{a}^{[r]} \in \Lambda_r$ ($\mathfrak{b}^{[r]} \in \Delta_r$, resp.) such that

$$D^{[r]}_{(0,\mathfrak{a}^{[r]})} = 1 \quad \Big(\text{and } D^{[r]}_{(1,\mathfrak{b}^{[r]})} = 1, \text{ resp.}\Big).$$

$(\mathcal{F}_4)$ For all $r \in \mathcal{S}$ and $\mathbf{a} \in \Lambda_r^{\mathrm{lin}}$, there exist $\mathbf{b} \in \mathbb{Z}_{\boldsymbol{\pi}}$ and $\alpha \in \mathbb{Z}_N$ such that

$$D^{[r]}_{(0,\mathbf{x}+\mathbf{a})} \overline{D^{[r]}_{(0,\mathbf{x})}} = \omega_N^{\alpha} \cdot \mathbf{F}_{\mathbf{x},\mathbf{b}} \quad \text{for all } \mathbf{x} \in \Lambda_r;$$

for all $r \in \mathcal{T}$ and $\mathbf{a} \in \Delta_r^{\mathrm{lin}}$, there exist $\mathbf{b} \in \mathbb{Z}_{\boldsymbol{\pi}}$ and $\alpha \in \mathbb{Z}_N$ such that

$$D^{[r]}_{(1,\mathbf{x}+\mathbf{a})} \overline{D^{[r]}_{(1,\mathbf{x})}} = \omega_N^{\alpha} \cdot \mathbf{F}_{\mathbf{b},\mathbf{x}} \quad \text{for all } \mathbf{x} \in \Delta_r.$$

Now let G be a connected graph. Below we reduce the computation of $Z_{\mathbf{C},\mathfrak{D}}(G)$ to $\mathsf{EVAL}(\widehat{\pi})$, where $\widehat{\pi} = \pi$ if $p \neq 2$ and $\widehat{\pi} = 2\pi$ if $p = 2$.

Given $a \in \mathbb{Z}_{\pi_i}$ for some $i \in [h]$, let $\widehat{a}$ denote an element in $\mathbb{Z}_{\widehat{\pi}}$ such that $\widehat{a} \equiv a \pmod{\pi_i}$. As $\pi_i \mid \pi_1 = \pi \mid \widehat{\pi}$, this lifting of a is certainly feasible. For definiteness, we can choose a itself if we consider a to be an integer between 0 and $\pi_i - 1$.

First, if G is not bipartite, then $Z_{\mathbf{C},\mathfrak{D}}(G)$ is trivially 0. From now on we assume $G = (U \cup V, E)$ to be bipartite: every edge has one vertex in U and one vertex in V.

Let u^* be a vertex in U. Then we can decompose $Z_{\mathbf{C},\mathfrak{D}}(G)$ into

$$Z_{\mathbf{C},\mathfrak{D}}(G) = Z^{\rightarrow}_{\mathbf{C},\mathfrak{D}}(G, u^*) + Z^{\leftarrow}_{\mathbf{C},\mathfrak{D}}(G, u^*).$$

We will reduce $Z^{\rightarrow}_{\mathbf{C},\mathfrak{D}}(G, u^*)$ to $\mathsf{EVAL}(\widehat{\pi})$. The $Z^{\leftarrow}$ part can be dealt with similarly.

We use U_r, where $r \in [0 : N-1]$, to denote the set of vertices in U whose degree is $r \pmod N$ and use V_ρ to denote the set of vertices in V whose degree is $\rho \pmod N$. We decompose E into $\bigcup_{i,j} E_{i,j}$, where $E_{i,j}$ contains the edges between U_i and V_j.

If $U_r \neq \emptyset$ for some $r \notin \mathcal{S}$ or if $V_\rho \neq \emptyset$ for some $\rho \notin \mathcal{T}$, then $Z^{\rightarrow}_{\mathbf{C},\mathfrak{D}}(G) = 0$. Thus, we assume that $U_r = \emptyset$ for all $r \notin \mathcal{S}$ and $V_\rho = \emptyset$ for all $\rho \notin \mathcal{T}$. In this case, we have

(12.4)
$$Z^{\rightarrow}_{\mathbf{C},\mathfrak{D}}(G, u^*) = \sum_{(f,g)} \left[\prod_{r \in \mathcal{S}} \left(\prod_{u \in U_r} D^{[r]}_{(0,\mathbf{x}_u)} \right) \prod_{\rho \in \mathcal{T}} \left(\prod_{v \in V_\rho} D^{[r]}_{(1,\mathbf{y}_v)} \right) \right] \left[\prod_{(r,\rho) \in \mathcal{S} \times \mathcal{T}} \prod_{uv \in E_{r,\rho}} F_{\mathbf{x}_u, \mathbf{y}_v} \right].$$

Here the sum ranges over all pairs (f, g), where

$$f = (f_r; r \in \mathcal{S}) \in \prod_{r \in \mathcal{S}} (U_r \to \Lambda_r) \quad \text{and} \quad g = (g_\rho; \rho \in \mathcal{T}) \in \prod_{\rho \in \mathcal{T}} (V_\rho \to \Delta_\rho)$$

such that $f(u) = \mathbf{x}_u$ and $g(v) = \mathbf{y}_v$.

The following lemma gives us a convenient way to do summation over a coset.

LEMMA 12.4. *Let Φ be a coset in $\mathbb{Z}_{\boldsymbol{\pi}}$ and $\mathbf{c} = (c_1, \ldots, c_h)$ be a vector in Φ. Then there exist a positive integer s and an $s \times h$ matrix $\mathbf{A}$ over $\mathbb{Z}_{\widehat{\pi}}$ such that the map $\tau : (\mathbb{Z}_{\widehat{\pi}})^s \to \mathbb{Z}_{\pi_1} \times \cdots \times \mathbb{Z}_{\pi_h}$, where $\tau(\mathbf{x}) = (\tau_1(\mathbf{x}), \ldots, \tau_h(\mathbf{x}))$ and*

$$\tau_j(\mathbf{x}) = \left(\mathbf{x}\mathbf{A}_{*,j} + \widehat{c}_j \pmod{\pi_j} \right) \in \mathbb{Z}_{\pi_j} \quad \text{for all } j \in [h], \tag{12.5}$$

is a uniform *map from $(\mathbb{Z}_{\widehat{\pi}})^s$ onto Φ. This uniformity means that for all $\mathbf{b}, \mathbf{b}' \in \Phi$, the number of $\mathbf{x} \in (\mathbb{Z}_{\widehat{\pi}})^s$ with $\tau(\mathbf{x}) = \mathbf{b}$ is the same as the number of $\mathbf{x}$ with $\tau(\mathbf{x}) = \mathbf{b}'$.*

Proof. Using the fundamental theorem of finite Abelian groups, there is a group isomorphism f from $\mathbb{Z}_{\mathbf{g}}$ onto Φ^{lin}, where $\mathbf{g} = (g_1, \ldots, g_s)$ is a sequence of powers of p and satisfies $\widehat{\pi} \geq \pi = \pi_1 \geq g_1 \geq \cdots \geq g_s$ for some $s \geq 1$. $\mathbb{Z}_{\mathbf{g}} \equiv \mathbb{Z}_{g_1} \times \cdots \times \mathbb{Z}_{g_s}$ is a $\mathbb{Z}_{\widehat{\pi}}$-module. This is clear, since as a $\mathbb{Z}$-module, any multiple of $\widehat{\pi}$ annihilates $\mathbb{Z}_{\mathbf{g}}$. Thus f is also a $\mathbb{Z}_{\widehat{\pi}}$-module isomorphism.

Let $\mathbf{a}_i = f(\mathbf{e}_i) \in \Phi^{\text{lin}}$ for each $i \in [s]$, where $\mathbf{e}_i \in \mathbb{Z}_{\mathbf{g}}$ is the vector whose ith entry is 1 and all other entries are 0. Let $\mathbf{a}_i = (a_{i,1}, \ldots, a_{i,h}) \in \mathbb{Z}_{\boldsymbol{\pi}}$, where $a_{i,j} \in \mathbb{Z}_{\pi_j}$, $i \in [s]$, $j \in [h]$. Let $\widehat{\mathbf{a}}_i = (\widehat{a}_{i,1}, \ldots, \widehat{a}_{i,h}) \in (\mathbb{Z}_{\widehat{\pi}})^h$ be a lifting of $\mathbf{a}_i$ componentwise. Similarly let $\widehat{\mathbf{c}}$ be a lifting of $\mathbf{c}$ componentwise. Then we claim that $\mathbf{A} = (\widehat{a}_{i,j})$ and $\widehat{\mathbf{c}}$ together give us the required *uniform* map τ from $(\mathbb{Z}_{\widehat{\pi}})^s$ to Φ.

To show that τ is uniform, we consider the linear part of $\tau' : (\mathbb{Z}_{\widehat{\pi}})^s \to \Phi^{\text{lin}}$,

$$\tau'(\mathbf{x}) = (\tau'_1(\mathbf{x}), \ldots, \tau'_h(\mathbf{x})), \quad \text{where } \tau'_j(\mathbf{x}) = \left(\mathbf{x}\mathbf{A}_{*,j} \pmod{\pi_j} \right) \in \mathbb{Z}_{\pi_j}$$

for all $j \in [h]$. Clearly we only need to show that τ' is a uniform map.

Let σ be the natural projection from $\mathbb{Z}^s_{\widehat{\pi}}$ to $\mathbb{Z}_{\mathbf{g}}$:

$$\mathbf{x} = (x_1, \ldots, x_s) \mapsto \left(x_1 \pmod{g_1}, \ldots, x_s \pmod{g_s} \right).$$

σ is certainly a uniform map, being a surjective homomorphism. Thus, every vector $\mathbf{b} \in \mathbb{Z}_{\mathbf{g}}$ has $|\ker \sigma| = \widehat{\pi}^s/(g_1 \cdots g_s)$ many preimages. We show that the map τ' factors through σ and f, i.e., $\tau' = f \circ \sigma$. Because f is an isomorphism, this implies that τ' is also a uniform map.

As $g_i \mathbf{e}_i = \mathbf{0}$ in $\mathbb{Z}_{\mathbf{g}}$, the following is a valid expression in the $\mathbb{Z}_{\widehat{\pi}}$-module for $\sigma(\mathbf{x})$:

$$\left(x_1 \pmod{g_1}, \ldots, x_s \pmod{g_s} \right) = \sum_{i=1}^{s} x_i \mathbf{e}_i.$$

Apply f as a $\mathbb{Z}_{\widehat{\pi}}$-module homomorphism $f(\sigma(\mathbf{x})) = \sum_{i=1}^s x_i f(\mathbf{e}_i)$ with its jth entry being $\sum_{i=1}^s x_i a_{i,j}$. This is an expression in the $\mathbb{Z}_{\widehat{\pi}}$-module $\mathbb{Z}_{\pi_j}$, which is the same as

$$\sum_{i=1}^{s} \left(x_i \pmod{\pi_j}\right) \cdot a_{i,j} = \sum_{i=1}^{s} x_i \widehat{a}_{i,j} \pmod{\pi_j} = \tau'_j(\mathbf{x}).$$

The lemma is proved. □

Applying Lemma 12.4 to Λ_r, for every $r \in \mathcal{S}$, there exist a positive integer s_r and an $s_r \times h$ matrix $\mathbf{A}^{[r]}$ over $\mathbb{Z}_{\widehat{\pi}}$ which give us a uniform map $\lambda^{[r]}(\mathbf{x})$ from $\mathbb{Z}_{\widehat{\pi}}^{s_r}$ to Λ_r:

$$\lambda_i^{[r]}(\mathbf{x}) = \left(\mathbf{x}\mathbf{A}_{*,i}^{[r]} + \widehat{\mathfrak{a}}_i^{[r]} \pmod{\pi_i}\right) \quad \text{for all } i \in [h] \text{ and } \mathbf{x} \in \mathbb{Z}_{\widehat{\pi}}^{s_r}. \tag{12.6}$$

Similarly, for every $r \in \mathcal{T}$, there exist a positive integer t_r and an $t_r \times h$ matrix $\mathbf{B}^{[r]}$ over $\mathbb{Z}_{\widehat{\pi}}$ which give us a uniform map $\delta^{[r]}$ from $\mathbb{Z}_{\widehat{\pi}}^{t_r}$ to Δ_r:

$$\delta_i^{[r]}(\mathbf{y}) = \left(\mathbf{y}\mathbf{B}_{*,i}^{[r]} + \widehat{\mathfrak{b}}_i^{[r]} \pmod{\pi_i}\right) \quad \text{for all } i \in [h] \text{ and } \mathbf{y} \in \mathbb{Z}_{\widehat{\pi}}^{t_r}. \tag{12.7}$$

Using $(\mathcal{F}_3)$, we have

$$D^{[r]}_{(0,\lambda^{[r]}(\mathbf{0}))} = 1 \text{ when } r \in \mathcal{S} \quad \text{and} \quad D^{[r]}_{(1,\delta^{[r]}(\mathbf{0}))} = 1 \text{ when } r \in \mathcal{T}. \tag{12.8}$$

Because both $\lambda^{[r]}$ and $\delta^{[r]}$ are uniform, and we know the multiplicity of each map (the cardinality of inverse images), to compute (12.4) it suffices to compute the following:

$$\sum_{(\mathbf{x}_u),(\mathbf{y}_v)} \prod_{r\in\mathcal{S}} \left(\prod_{u\in U_r} D^{[r]}_{(0,\lambda^{[r]}(\mathbf{x}_u))} \right) \prod_{r\in\mathcal{T}} \left(\prod_{v\in V_r} D^{[r]}_{(1,\delta^{[r]}(\mathbf{y}_v))} \right) \prod_{\substack{r_1\in\mathcal{S}, r_2\in\mathcal{T} \\ uv\in E_{r_1,r_2}}} F_{\lambda^{[r_1]}(\mathbf{x}_u),\delta^{[r_2]}(\mathbf{y}_v)}, \tag{12.9}$$

where the sum is over pairs of sequences

$$\left(\mathbf{x}_u; u \in \bigcup_{r\in\mathcal{S}} U_r\right) \in \prod_{r\in\mathcal{S}} \left(\mathbb{Z}_{\widehat{\pi}}^{s_r}\right)^{|U_r|} \quad \text{and} \quad \left(\mathbf{y}_v; v \in \bigcup_{r\in\mathcal{T}} V_r\right) \in \prod_{r\in\mathcal{T}} \left(\mathbb{Z}_{\widehat{\pi}}^{t_r}\right)^{|V_r|}.$$

If (1) for all $r \in \mathcal{S}$, there is a quadratic polynomial $f^{[r]}$ over $\mathbb{Z}_{\widehat{\pi}}$ such that

$$D^{[r]}_{(0,\lambda^{[r]}(\mathbf{x}))} = \omega_{\widehat{\pi}}^{f^{[r]}(\mathbf{x})} \quad \text{for all } \mathbf{x} \in \mathbb{Z}_{\widehat{\pi}}^{s_r}; \tag{12.10}$$

(2) for all $r \in \mathcal{T}$, there is a quadratic polynomial $g^{[r]}$ over $\mathbb{Z}_{\widehat{\pi}}$ such that

$$D^{[r]}_{(1,\delta^{[r]}(\mathbf{y}))} = \omega_{\widehat{\pi}}^{g^{[r]}(\mathbf{y})} \quad \text{for all } \mathbf{y} \in \mathbb{Z}_{\widehat{\pi}}^{t_r}; \tag{12.11}$$

(3) for all $r_1 \in \mathcal{S}, r_2 \in \mathcal{T}$, there is a quadratic polynomial $f^{[r_1,r_2]}$ over $\mathbb{Z}_{\widehat{\pi}}$ such that

$$F_{\lambda^{[r_1]}(\mathbf{x}),\delta^{[r_2]}(\mathbf{y})} = \omega_{\widehat{\pi}}^{f^{[r_1,r_2]}(\mathbf{x},\mathbf{y})} \quad \text{for all } \mathbf{x} \in \mathbb{Z}_{\widehat{\pi}}^{s_{r_1}} \text{ and } \mathbf{y} \in \mathbb{Z}_{\widehat{\pi}}^{t_{r_2}}, \tag{12.12}$$

then we can reduce the computation of the summation in (12.9) to $\mathsf{EVAL}(\widehat{\pi})$.

We start with (3). By $(\mathcal{F}_1)$, the following map $f^{[r_1,r_2]}$ satisfies (12.12):

$$f^{[r_1,r_2]}(\mathbf{x},\mathbf{y}) = \sum_{i\in[h]} \frac{\widehat{\pi}}{\pi_i} \cdot \lambda_i^{[r_1]}(\mathbf{x}) \cdot \delta_i^{[r_2]}(\mathbf{y}) = \sum_{i\in[h]} \frac{\widehat{\pi}}{\pi_i} \left(\mathbf{x}\mathbf{A}_{*,i}^{[r_1]} + \widehat{\mathfrak{a}}_i^{[r_1]}\right) \left(\mathbf{y}\mathbf{B}_{*,i}^{[r_2]} + \widehat{\mathfrak{b}}_i^{[r_2]}\right).$$

Note that the presence of the integer $\widehat{\pi}/\pi_i$ is crucial to be able to substitute the mod π_i expressions in (12.6) and in (12.7), respectively, as if they were mod $\widehat{\pi}$ expressions. It is also clear that $f^{[r_1,r_2]}$ is indeed a quadratic polynomial over $\mathbb{Z}_{\widehat{\pi}}$.

Next we prove (1), which is a little more complicated. The proof of (2) is similar.

Let $r \in S$. Let $\mathbf{e}_i$ denote the vector in $\mathbb{Z}_{\widehat{\pi}}^{s_r}$ whose ith entry is 1 and all other entries are 0. Using $(\mathcal{F}_4)$, for each $i \in [s_r]$, there exist $\alpha_i \in \mathbb{Z}_N$ and $\mathbf{b}_i = (b_{i,1}, \ldots, b_{i,h}) \in \mathbb{Z}_{\boldsymbol{\pi}}$, where $b_{i,j} \in \mathbb{Z}_{\pi_j}$, such that

$$D^{[r]}_{(0,\lambda^{[r]}(\mathbf{x}+\mathbf{e}_i))} \overline{D^{[r]}_{(0,\lambda^{[r]}(\mathbf{x}))}} = \omega_N^{\alpha_i} \prod_{j\in[h]} \omega_{\pi_j}^{b_{i,j}\cdot\lambda_j^{[r]}(\mathbf{x})} \quad \text{for all } \mathbf{x} \in \mathbb{Z}_{\widehat{\pi}}^{s_r}. \tag{12.13}$$

We have this equation because $\lambda^{[r]}(\mathbf{x}+\mathbf{e}_i) - \lambda^{[r]}(\mathbf{x})$ is indeed a vector in $\mathbb{Z}_{\boldsymbol{\pi}}$ that is independent of $\mathbf{x}$. To see this, observe that the jth entry in $\lambda^{[r]}(\mathbf{x}+\mathbf{e}_i) - \lambda^{[r]}(\mathbf{x})$ is

$$\mathbf{e}_i \mathbf{A}^{[r]}_{*,j} = A^{[r]}_{i,j} \pmod{\pi_j},$$

and thus the displacement vector $\lambda^{[r]}(\mathbf{x}+\mathbf{e}_i) - \lambda^{[r]}(\mathbf{x})$ is independent of $\mathbf{x}$ and is in Λ_r^{lin} by definition. This is the $\mathbf{a} \in \Lambda_r^{\mathrm{lin}}$ in the statement of $(\mathcal{F}_4)$ which we applied.

Before moving forward, we show that $\omega_N^{\alpha_i}$ must be a power of $\omega_{\widehat{\pi}}$. This is because

$$1 = \prod_{j=0}^{\widehat{\pi}-1} D^{[r]}_{(0,\lambda^{[r]}((j+1)\mathbf{e}_i))} \overline{D^{[r]}_{(0,\lambda^{[r]}(j\mathbf{e}_i))}} = (\omega_N^{\alpha_i})^{\widehat{\pi}} \prod_{k\in[h]} \omega_{\pi_k}^{b_{i,k}[\lambda_k^{[r]}(0\mathbf{e}_i)+\cdots+\lambda_k^{[r]}((\widehat{\pi}-1)\mathbf{e}_i)]}. \tag{12.14}$$

For each $k \in [h]$, the exponent of ω_{π_k} is $b_{i,k}Q_k \in \mathbb{Z}_{\pi_k}$, where Q_k is the following sum:

$$\sum_{j=0}^{\widehat{\pi}-1} \lambda_k^{[r]}(j\mathbf{e}_i) = \sum_{j=0}^{\widehat{\pi}-1} \left((j\mathbf{e}_i)\mathbf{A}^{[r]}_{*,k} + \widehat{\mathfrak{a}}_k^{[r]} \pmod{\pi_k} \right) = \left(\sum_{j=1}^{\widehat{\pi}-1} j\mathbf{e}_i \right) \mathbf{A}^{[r]}_{*,k} \pmod{\pi_k} = 0. \tag{12.15}$$

The last equality comes from $J \equiv \sum_{j=1}^{\widehat{\pi}-1} j = 0 \pmod{\pi_k}$, and this is due to our definition of $\widehat{\pi}$. When p is odd, J is a multiple of $\widehat{\pi}$ and $\pi_k \mid \widehat{\pi}$, and when $p = 2$, J is a multiple of $\widehat{\pi}/2$. However, in this case, we have $\widehat{\pi}/2 = \pi_1$ and $\pi_k \mid \pi_1$.

As a result, $(\omega_N^{\alpha_i})^{\widehat{\pi}} = 1$. So there exists $\beta_i \in \mathbb{Z}_{\widehat{\pi}}$ for each $i \in [s_r]$ such that

$$D^{[r]}_{(0,\lambda^{[r]}(\mathbf{x}+\mathbf{e}_i))} \overline{D^{[r]}_{(0,\lambda^{[r]}(\mathbf{x}))}} = \omega_{\widehat{\pi}}^{\beta_i} \prod_{j\in[h]} \omega_{\pi_j}^{b_{i,j}\cdot\lambda_j^{[r]}(\mathbf{x})} \quad \text{for all } \mathbf{x} \in \mathbb{Z}_{\widehat{\pi}}^{s_r}. \tag{12.16}$$

It follows that every nonzero entry of $\mathbf{D}^{[r]}$ is a power of $\omega_{\widehat{\pi}}$. This uses $(\mathcal{F}_3)$, that the $(0, \mathfrak{a}^{[r]})$th entry of $\mathbf{D}^{[r]}$ is 1, and the fact that $\lambda^{[r]}$ is surjective to Λ_r: any point in Λ_r is connected to the normalizing point $\mathfrak{a}^{[r]}$ by a sequence of moves $\lambda^{[r]}(\mathbf{x}) \to \lambda^{[r]}(\mathbf{x}+\mathbf{e}_i)$ for $i \in [s_r]$. Now we know there is a function $f^{[r]} \colon \mathbb{Z}_{\widehat{\pi}}^{s_r} \to \mathbb{Z}_{\widehat{\pi}}$ which satisfies (12.10). We want to show that it is indeed a quadratic polynomial. To see this, by (12.16),

$$f^{[r]}(\mathbf{x}+\mathbf{e}_i) - f^{[r]}(\mathbf{x}) = \beta_i + \sum_{j\in[h]} \frac{\widehat{\pi}}{\pi_j} \cdot b_{i,j} \cdot \lambda_j^{[r]}(\mathbf{x}) = \beta_i + \sum_{j\in[h]} \frac{\widehat{\pi}}{\pi_j} \cdot \widehat{b}_{i,j} \cdot \left(\mathbf{x}\mathbf{A}^{[r]}_{*,j} + \widehat{\mathfrak{a}}_j^{[r]} \right) \tag{12.17}$$

for every $i \in [s_r]$. We should remark that originally $b_{i,j}$ is in $\mathbb{Z}_{\pi_j}$; however, with the integer multiplier $(\widehat{\pi}/\pi_j)$, the quantity $(\widehat{\pi}/\pi_j)\cdot b_{i,j}$ is now considered in $\mathbb{Z}_{\widehat{\pi}}$. Moreover,

$$\widehat{b}_{i,j} \equiv b_{i,j} \pmod{\pi_j} \implies \left(\frac{\widehat{\pi}}{\pi_j}\right) \widehat{b}_{i,j} \equiv \left(\frac{\widehat{\pi}}{\pi_j}\right) b_{i,j} \pmod{\widehat{\pi}}.$$

Thus the expression in (12.17) is evaluated in $\mathbb{Z}_{\widehat{\pi}}$, which means that for any $i \in [s_r]$, there exist $c_{i,0}, \ldots, c_{i,s_r} \in \mathbb{Z}_{\widehat{\pi}}$ such that

$$f^{[r]}(\mathbf{x}+\mathbf{e}_i) - f^{[r]}(\mathbf{x}) = c_{i,0} + \sum_{j\in[s_r]} c_{i,j}x_j. \tag{12.18}$$

Since $f^{[r]}(\mathbf{0}) = 0$, the case when p is odd follows from the lemma below.

LEMMA 12.5. *Let π be a power of an odd prime, and let f be a map from $\mathbb{Z}_\pi^s$ to $\mathbb{Z}_\pi$ for some $s \geq 1$. Suppose for every $i \in [s]$, there exist $c_{i,0}, \ldots, c_{i,s} \in \mathbb{Z}_\pi$ such that*

$$f(\mathbf{x}+\mathbf{e}_i) - f(\mathbf{x}) = c_{i,0} + \sum_{j\in[s]} c_{i,j}x_j \quad \textit{for all } \mathbf{x} \in \mathbb{Z}_\pi^s \tag{12.19}$$

and $f(\mathbf{0}) = 0$. Then there exist $a_{i,j}, a_i \in \mathbb{Z}_\pi$ such that

$$f(\mathbf{x}) = \sum_{i\leq j\in[s]} a_{i,j}x_ix_j + \sum_{i\in[s]} a_ix_i \quad \textit{for all } \mathbf{x} \in \mathbb{Z}_\pi^s. \tag{12.20}$$

Proof. First note that f is uniquely determined by the conditions on $f(\mathbf{x}+\mathbf{e}_i) - f(\mathbf{x})$ and $f(\mathbf{0})$. Second, we show that $c_{i,j} = c_{j,i}$ for all $i, j \in [s]$; otherwise f does not exist, contradicting the assumption. On the one hand, we have

$$f(\mathbf{e}_i+\mathbf{e}_j) = f(\mathbf{e}_i+\mathbf{e}_j) - f(\mathbf{e}_j) + f(\mathbf{e}_j) - f(\mathbf{0}) = c_{i,0} + c_{i,j} + c_{j,0}.$$

On the other hand, we have

$$f(\mathbf{e}_i+\mathbf{e}_j) = f(\mathbf{e}_i+\mathbf{e}_j) - f(\mathbf{e}_i) + f(\mathbf{e}_i) - f(\mathbf{0}) = c_{j,0} + c_{j,i} + c_{i,0}.$$

It follows that $c_{i,j} = c_{j,i}$.

Finally, we set $a_{i,j} = c_{i,j}$ for all $i < j \in [s]$; $a_{i,i} = c_{i,i}/2$ for all $i \in [s]$ (here $c_{i,i}/2$ is well defined because π is odd); and $a_i = c_{i,0} - a_{i,i}$ for all $i \in [s]$. We now claim that

$$g(\mathbf{x}) = \sum_{i\leq j\in[s]} a_{i,j}x_ix_j + \sum_{i\in[s]} a_ix_i$$

satisfies both conditions and thus $f = g$. To see this, we check the case when $i = 1$:

$$g(\mathbf{x}+\mathbf{e}_1) - g(\mathbf{x}) = 2a_{1,1}x_1 + \sum_{j>1} a_{1,j}x_j + (a_{1,1}+a_1) = c_{1,1}x_1 + \sum_{j>1} c_{1,j}x_j + c_{1,0}.$$

Other cases are similar, and the lemma is proved. ∎

When $p = 2$, we first claim that the constants $c_{i,i}$ in (12.18) must be even, since

$$0 = f^{[r]}(\widehat{\pi}\mathbf{e}_i) - f^{[r]}((\widehat{\pi}-1)\mathbf{e}_i) + \cdots + f^{[r]}(\mathbf{e}_i) - f^{[r]}(\mathbf{0}) = \widehat{\pi}c_{i,0} + c_{i,i}(\widehat{\pi} - 1 + \cdots + 1 + 0).$$

This equality happens in $\mathbb{Z}_{\widehat{\pi}}$, so $c_{i,i}(\widehat{\pi}(\widehat{\pi}-1)/2) = 0 \pmod{\widehat{\pi}}$. When $\widehat{\pi} - 1$ is odd we have $2 \,|\, c_{i,i}$. It follows from the lemma below that $f^{[r]}$ is a quadratic polynomial.

LEMMA 12.6. *Let π be a power of 2 and let f be a map from $\mathbb{Z}_\pi^s$ to $\mathbb{Z}_\pi$ satisfying $f(\mathbf{0}) = 0$. Suppose for every $i \in [s]$ there exist $c_{i,0}, \ldots, c_{i,s} \in \mathbb{Z}_\pi$, where $2 \,|\, c_{i,i}$, such that* (12.19) *holds. Then there are $a_{i,j}, a_i \in \mathbb{Z}_\pi$ such that f has the form of* (12.20).

Proof. The proof of Lemma 12.6 is essentially the same as that of Lemma 12.5. Because $2 \,|\, c_{i,i}$, $a_{i,i} = c_{i,i}/2$ is well-defined (in particular, when $c_{i,i} = 0$, we set $a_{i,i} = 0$). ∎

12.3. Proof of Theorem 12.1. Finally we turn to the proof of Theorem 12.1, i.e., $\mathsf{EVAL}(q)$ is tractable for any fixed prime power q.

Actually, there is a well-known polynomial-time algorithm for $\mathsf{EVAL}(q)$ when q is a prime [10, 15], [27, Theorem 6.30]. (The algorithm works for any finite field.) Here we present a polynomial-time algorithm that works for any prime power q. We start with the easier case when q is odd.

LEMMA 12.7. *Let p be an odd prime and let $q = p^k$ for some positive integer k. Let $f \in \mathbb{Z}_q[x_1, \ldots, x_n]$ be a quadratic polynomial over n variables $x_1, \ldots, x_n$. Then*

$$Z_q(f) = \sum_{x_1,\ldots,x_n \in \mathbb{Z}_q} \omega_q^{f(x_1,\ldots,x_n)}$$

can be evaluated in polynomial time (in n).

Proof. We assume that $f(x_1, \ldots, x_n)$ has the following form:

$$f(x_1, \ldots, x_n) = \sum_{i \le j \in [n]} c_{i,j} x_i x_j + \sum_{i \in [n]} c_i x_i + c_0, \tag{12.21}$$

where all the $c_{i,j}$ and c_i are elements in $\mathbb{Z}_q$.

First, as a warm up, we give an algorithm and prove its correctness for the case $k = 1$, i.e., $q = p$ is an odd prime. Note that if f is affine, then the evaluation can be trivially done in polynomial time. In fact, it decouples into a product of n sums,

$$\sum_{x_1,\ldots,x_n \in \mathbb{Z}_q} \omega_q^{f(x_1,\ldots,x_n)} = \sum_{x_1,\ldots,x_n \in \mathbb{Z}_q} \omega_q^{\sum_{i=1}^n c_i x_i + c_0} = \omega_q^{c_0} \times \prod_{i=1}^n \sum_{x_i \in \mathbb{Z}_q} \omega_q^{c_i x_i}.$$

This sum is equal to 0 if any $c_i \in \mathbb{Z}_q$ is nonzero and is equal to $q^n \omega_q^{c_0}$ otherwise.

Now assume $f(x_1, \ldots, x_n)$ is not affine linear. Then in each round (which we will describe below), the algorithm will decrease the number of variables by at least one, in polynomial time. Assume f contains some quadratic terms. There are two cases: f has at least one square term or f does not have any square terms.

In the first case, without loss of generality, we assume that $c_{1,1} \neq 0$. There exist an affine function $g \in \mathbb{Z}_q[x_2, \ldots, x_n]$ and a quadratic polynomial $f' \in \mathbb{Z}_q[x_2, \ldots, x_n]$, both over $n-1$ variables $x_2, x_3, \ldots, x_n$, such that

$$f(x_1, x_2, \ldots, x_n) = c_{1,1}\big(x_1 + g(x_2, x_3, \ldots, x_n)\big)^2 + f'(x_2, x_3, \ldots, x_n).$$

Here we used the fact that both 2 and $c_{1,1} \in \mathbb{Z}_q$ are invertible in the field $\mathbb{Z}_q$. (Recall we assumed that $q = p$ is an odd prime.) Thus, we can factor out a coefficient $2c_{1,1}$ from the cross term $x_1 x_i$ for every $i > 1$, and from the linear term x_1, to obtain the expression $c_{1,1}(x_1 + g(x_2, \ldots, x_n))^2$.

For any fixed $x_2, \ldots, x_n$, when x_1 ranges over $\mathbb{Z}_q$, $x_1 + g$ ranges over $\mathbb{Z}_q$. Thus,

$$\sum_{x_1,x_2,\ldots,x_n \in \mathbb{Z}_q} \omega_q^{f(x_1,x_2,\ldots,x_n)} = \sum_{x_2,\ldots,x_n \in \mathbb{Z}_q} \omega_q^{f'} \sum_{x_1 \in \mathbb{Z}_q} \omega_q^{c_{1,1}(x_1+g)^2} = \sum_{x \in \mathbb{Z}_q} \omega_q^{c_{1,1}x^2} \cdot Z_q(f').$$

The first factor can be evaluated in constant time (which is independent of n), and the computation of $Z_q(f)$ is reduced to the computation of $Z_q(f')$ in which f' has at most $n-1$ variables.

Remark 12.8. The claim of $\sum_x \omega_q^{cx^2}$ being "computable in constant time" here is a trivial statement, since we consider $q = p$ to be a fixed constant. However, for a

general prime p, we remark that the sum is the famous Gauss quadratic sum and has the following closed formula: If $c \neq 0$,

$$\sum_{x \in \mathbb{Z}_p} \omega_p^{cx^2} = \left(\frac{c}{p}\right) G, \quad \text{where } G = \sum_{x \in \mathbb{Z}_p} \left(\frac{x}{p}\right) \omega_p^x.$$

Here $(\frac{c}{p})$ is the Legendre symbol. It can be computed in polynomial time in the binary length of c and p. G has the closed form $G = +\sqrt{p}$ if $p \equiv 1 \bmod 4$ and $G = +i\sqrt{p}$ if $p \equiv 3 \bmod 4$.[4]

The second case is that all the quadratic terms in f are cross terms (in particular this implies that $n \geq 2$). In this case we assume, without loss of generality, that $c_{1,2}$ is nonzero. We apply the following transformation: $x_1 = x_1' + x_2'$ and $x_2 = x_1' - x_2'$. As 2 is invertible in $\mathbb{Z}_q$, when x_1' and x_2' go over $\mathbb{Z}_q^2$, x_1 and x_2 also go over $\mathbb{Z}_q^2$. Thus,

$$\sum_{x_1, x_2, \ldots, x_n \in \mathbb{Z}_q} \omega_q^{f(x_1, x_2, \ldots, x_n)} = \sum_{x_1', x_2', , \ldots, x_n \in \mathbb{Z}_q} \omega_q^{f(x_1'+x_2', x_1'-x_2', \ldots, x_n)}.$$

Viewing $f(x_1' + x_2', x_1' - x_2', \ldots, x_n)$ as a new quadratic polynomial f' of $x_1', x_2', \ldots, x_n$ its coefficient of $x_1'^2$ is exactly $c_{1,2} \neq 0$. Thus f' contains at least one square term. This reduces our problem back to the first case. We can use the method described earlier to reduce the number of variables.

Repeating this process we get a polynomial-time algorithm for computing $Z_q(f)$ when $q = p$ is an odd prime. Now we consider the case when $q = p^k$.

We can write any nonzero $a \in \mathbb{Z}_q$ as $a = p^t a'$, where t is a unique nonnegative integer, such that $p \nmid a'$. We call t the order of a (with respect to p). If f is an affine linear function, $Z_q(f)$ is easy to compute, as the sum factors into n sums as before. Now we assume f has nonzero quadratic terms. Let t_0 be the smallest order of all the nonzero quadratic coefficients $c_{i,j}$ of f. We consider the following two cases: there exists at least one square term with coefficient of order t_0 or not.

For the first case, without loss of generality, assume $c_{1,1} = p^{t_0} c$ and $p \nmid c$ (so c is invertible in $\mathbb{Z}_q$). By the minimality of t_0, every nonzero coefficient of a quadratic term has a factor p^{t_0}. Now we factor out $c_{1,1}$ from every quadratic term involving x_1, namely, from $x_1^2, x_1x_2, \ldots, x_1x_n$. (Clearly it does not matter if the coefficient of a term x_1x_i, $i \neq 1$, is 0.) We can write $f(x_1, x_2, \ldots, x_n) = c_{1,1}(x_1 + g(x_2, \ldots, x_n))^2 + c_1x_1 +$ a quadratic polynomial in $(x_2, \ldots, x_n)$, where g is a linear function over $x_2, \ldots, x_n$. By adding and then subtracting $c_1 g(x_2, \ldots, x_n)$, we get

$$f(x_1, x_2, \ldots, x_n) = c_{1,1}\big(x_1 + g(x_2, \ldots, x_n)\big)^2 + c_1\big(x_1 + g(x_2, \ldots, x_n)\big) + f'(x_2, \ldots, x_n),$$

where $f'(x_2, \ldots, x_n) \in \mathbb{Z}_q[x_2, \ldots, x_n]$ is a quadratic polynomial over $x_2, \ldots x_n$.

For any fixed $x_2, \ldots, x_n$, when x_1 ranges over $\mathbb{Z}_q$, $x_1 + g$ also ranges over $\mathbb{Z}_q$. So

$$\sum_{x_1, \ldots, x_n \in \mathbb{Z}_q} \omega_q^f = \left(\sum_{x \in \mathbb{Z}_q} \omega_q^{c_{1,1}x^2 + c_1x}\right)\left(\sum_{x_2, \ldots, x_n \in \mathbb{Z}_q} \omega_q^{f'}\right) = \sum_{x \in \mathbb{Z}_q} \omega_q^{c_{1,1}x^2 + c_1x} \cdot Z_q(f').$$

[4] It had been known to Gauss since 1801 that $G = \pm\sqrt{p}$ if $p \equiv 1 \pmod 4$ and $G = \pm i\sqrt{p}$ if $p \equiv 3 \pmod 4$. The fact that G always takes the plus sign was conjectured by Gauss in his diary in May 1801. He wrote to his friend W. Olbers on September 3, 1805, that seldom had a week passed for four years that he had not tried in vain to prove this very elegant conjecture. Finally, he wrote, "Wie der Blitz einschlägt, hat sich das Räthsel gelöst" (as lightning strikes was the puzzle solved).

The first term can be evaluated in constant time and the problem is reduced to $Z_q(f')$ in which f' has at most $n-1$ variables.

For the second case, all square terms of f either are 0 or have orders larger than t_0. We assume, without loss of generality, that $c_{1,2} = p^{t_0}c$ and $p \nmid c$. We apply the following transformation: $x_1 = x_1' + x_2'$ and $x_2 = x_1' - x_2'$. Since 2 is invertible in $\mathbb{Z}_q$, when x_1' and x_2' go over $\mathbb{Z}_q^2$, x_1 and x_2 also go over $\mathbb{Z}_q^2$. After the transformation, we get a new quadratic polynomial over $x_1', x_2', x_3, \ldots, x_n$ such that $Z_q(f') = Z_q(f)$, and t_0 is still the smallest order of all the quadratic terms of f': The terms x_1^2 and x_2^2 (in f) produce terms with coefficients divisible by p^{t_0+1}, the term x_1x_2 (in f) produces terms $(x_1')^2$ and $(x_2')^2$ with coefficients of order exactly t_0, and terms x_1x_i or x_2x_i for $i \neq 1, 2$ produce terms $x_1'x_i$ and $x_2'x_i$ with coefficients divisible by p^{t_0}. In particular, the coefficient of $(x_1')^2$ in f' has order t_0, so we reduce the problem to the first case.

To sum up, we have a polynomial-time algorithm for every $q = p^k$, when $p \neq 2$. □

Now we deal with the more difficult case when $q = 2^k$, for some $k \geq 1$. We note that the property of an element $c \in \mathbb{Z}_{2^k}$ being even or odd is well-defined.

LEMMA 12.9. *Let $q = 2^k$ for some $k \geq 1$. Let $f \in \mathbb{Z}_q[x_1, \ldots, x_n]$ be a quadratic polynomial over $x_1, \ldots, x_n$. Then $Z_q(f)$ can be evaluated in polynomial time (in n).*

Proof. When $k = 1$, $Z_q(f)$ is computable in polynomial time according to [10], [27, Theorem 6.30] so we assume $k > 1$. We also assume f has the form as in (12.21).

The algorithm goes as follows: For each round, we can, in polynomial time, either

1. output the correct value of $Z_q(f)$, or
2. build a new quadratic $g \in \mathbb{Z}_{q/2}[x_1, \ldots, x_n]$ and reduce $Z_q(f)$ to $Z_{q/2}(g)$, or
3. build a new quadratic $g \in \mathbb{Z}_q[x_1, \ldots, x_{n-1}]$ and reduce $Z_q(f)$ to $Z_q(g)$.

This gives a polynomial-time algorithm for $\mathsf{EVAL}(q)$, because both base cases, when $k = 1$ or $n = 1$, can be solved efficiently.

Suppose we have a quadratic polynomial $f \in \mathbb{Z}_q[x_1, \ldots, x_n]$. Our first step is to transform f so that all the coefficients of its cross terms ($c_{i,j}$, where $i \neq j$) and linear terms (c_i) are divisible by 2. Assume f does not yet have this property. Let t be the smallest index in $[n]$ such that one of $\{c_t, c_{t,j} : j > t\}$ is not divisible by 2. Separating out the terms involving x_t, we rewrite f as follows:

$$f = c_{t,t} \cdot x_t^2 + x_t \cdot f_1(x_1, \ldots, \widehat{x_t}, \ldots, x_n) + f_2(x_1, \ldots, \widehat{x_t}, \ldots, x_n), \tag{12.22}$$

where f_1 is an affine linear function and f_2 is a quadratic polynomial. Both f_1 and f_2 are over variables $\{x_1, \ldots, x_n\} - \{x_t\}$. Here the notation $\widehat{x_t}$ means that x_t does not appear in the polynomial. Moreover,

$$f_1(x_1, \ldots, \widehat{x_t}, \ldots, x_n) = \sum_{i<t} c_{i,t}x_i + \sum_{j>t} c_{t,j}x_j + c_t. \tag{12.23}$$

From the minimality of t, $c_{i,t}$ is even for all $i < t$, and at least one of $\{c_{t,j}, c_t : j > t\}$ is odd. We claim that

$$Z_q(f) = \sum_{x_1,\ldots,x_n \in \mathbb{Z}_q} \omega_q^{f(x_1,\ldots,x_n)} = \sum_{\substack{x_1,\ldots,x_n \in \mathbb{Z}_q \\ f_1(x_1,\ldots,\widehat{x_t},\ldots,x_n) \equiv 0 \bmod 2}} \omega_q^{f(x_1,\ldots,x_n)}. \tag{12.24}$$

This is because

$$\sum_{\substack{x_1,\ldots,x_n \in \mathbb{Z}_q \\ f_1 \equiv 1 \bmod 2}} \omega_q^{f(x_1,\ldots,x_n)} = \sum_{\substack{x_1,\ldots,\widehat{x_t},\ldots,x_n \in \mathbb{Z}_q \\ f_1 \equiv 1 \bmod 2}} \sum_{x_t \in \mathbb{Z}_q} \omega_{2^k}^{c_{t,t}x_t^2 + x_t f_1 + f_2}.$$

However, for any fixed $x_1, \ldots, \widehat{x_t}, \ldots, x_n$, we have

$$\sum_{x_t \in \mathbb{Z}_q} \omega_{2^k}^{c_{t,t}x_t^2 + x_t f_1 + f_2} = \omega_{2^k}^{f_2} \sum_{x_t \in [0:2^{k-1}-1]} \omega_{2^k}^{c_{t,t}x_t^2 + x_t f_1} + \omega_{2^k}^{c_{t,t}(x_t + 2^{k-1})^2 + (x_t + 2^{k-1}) f_1}$$
$$= \omega_{2^k}^{f_2} \left(1 + (-1)^{f_1}\right) \sum_{x_t \in [0:2^{k-1}-1]} \omega_{2^k}^{c_{t,t}x_t^2 + x_t f_1} = 0,$$

since $f_1 \equiv 1 \bmod 2$. We used $(x + 2^{k-1})^2 \equiv x^2 \pmod{2^k}$ in the first equality.

Recall that f_1 (see (12.23)) is an affine form of $\{x_1, \ldots, \widehat{x_t}, \ldots, x_n\}$, that $c_{i,t}$ is even for all $i < t$, and that one of $\{c_{t,j}, c_t : j > t\}$ is odd. We consider two cases.

In the first case, $c_{t,j}$ is even for all $j > t$ and c_t is odd. Then for any assignment $(x_1, \ldots, \widehat{x_t}, \ldots, x_n)$ in $\mathbb{Z}_q^{n-1}$, f_1 is odd. As a result, by (12.24), $Z_q(f)$ is trivially zero.

In the second case, there exists at least one $j > t$ such that $c_{t,j}$ is odd. We let $\ell > t$ be the smallest of such j. Then we substitute the variable x_ℓ in f with a new variable x'_ℓ over $\mathbb{Z}_q$, where (since $c_{t,\ell}$ is odd, $c_{t,\ell}$ is invertible in $\mathbb{Z}_q$)

$$x_\ell = c_{t,\ell}^{-1} \left(2x'_\ell - \left(\sum_{i<t} c_{i,t} x_i + \sum_{j>t, j \neq \ell} c_{t,j} x_j + c_t \right) \right). \tag{12.25}$$

Let f' denote the new quadratic polynomial in $\mathbb{Z}_q[x_1, \ldots, x'_\ell, \ldots, x_n]$. We claim that

$$Z_q(f') = 2 \cdot Z_q(f) = 2 \cdot \sum_{\substack{x_1, \ldots, x_n \in \mathbb{Z}_q \\ f_1 \equiv 0 \bmod 2}} \omega_q^{f(x_1, \ldots, x_n)}.$$

To see it, we define a map from $\mathbb{Z}_q^n$ to $\mathbb{Z}_q^n$: $(x_1, \ldots, x'_\ell, \ldots, x_n) \mapsto (x_1, \ldots, x_\ell, \ldots, x_n)$, where x_ℓ satisfies (12.25). The range of the map is the set of $(x_1, \ldots, x_\ell, \ldots, x_n) \in \mathbb{Z}_q^n$ such that f_1 is even and every such tuple has two preimages in $\mathbb{Z}_q^n$. The claim follows.

So to compute $Z_q(f)$, we only need to compute $Z_q(f')$, and the advantage of $f' \in \mathbb{Z}_q[x_1, \ldots, x'_\ell, \ldots, x_n]$ over f is the following property that we are going to prove:

(*Even*) Every cross term and linear term that involves $x_1, \ldots, x_t$ has an even coefficient in f'.

To show this, we partition the terms of f' that we are interested in into three groups: cross and linear terms that involve x_t; linear terms x_s, $s < t$; and cross terms of the form $x_s x_{s'}$, where $s < s', s < t$.

First, we consider the expression (12.22) of f after the substitution. The first term $c_{t,t}x_t^2$ remains the same; the second term $x_t f_1$ becomes $2x_t x'_\ell$ by (12.25); x_t does not appear in the third term, even after the substitution. So (*Even*) holds for x_t.

Second, we consider the coefficient c'_s of the linear term x_s in f', where $s < t$. Only the following terms in f can possibly contribute to c'_s:

$$c_s x_s, \ c_{\ell,\ell} x_\ell^2, \ c_{s,\ell} x_s x_\ell, \text{ and } \ c_\ell x_\ell.$$

By the minimality of t, both c_s and $c_{s,\ell}$ are even. For $c_{\ell,\ell}x_\ell^2$ and $c_\ell x_\ell$, although we do not know whether $c_{\ell,\ell}$ and c_ℓ are even or odd, we know that the coefficient $-c_{t,\ell}^{-1} c_{s,t}$ of x_s in (12.25) is even since $c_{s,t}$ is even. So, every term in the list above makes an even contribution to c'_s and thus c'_s is even.

Finally, we consider the coefficient $c'_{s,s'}$ of the term $x_s x_{s'}$ in f', where $s < s'$ and $s < t$. Similarly, only the following terms in f can possibly contribute to $c'_{s,s'}$ (here we consider the general case when $s' \neq \ell$; the special case when $s' = \ell$ is easier):

$$c_{s,s'} x_s x_{s'}, \ c_{\ell,\ell} x_\ell^2, \ c_{s,\ell} x_s x_\ell, \text{ and } \ c_{\ell,s'} x_\ell x_{s'} \ (\text{or } \ c_{s',\ell} x_{s'} x_\ell).$$

By the minimality of t, $c_{s,s'}$ and $c_{s,\ell}$ are even. Moreover, the coefficient $-c_{t,\ell}^{-1}c_{s,t}$ of x_s in (12.25) is even. As a result, every term in the list above makes an even contribution to $c'_{s,s'}$ and thus $c'_{s,s'}$ is even.

To summarize, after substituting x_ℓ with x'_ℓ using (12.25) we get a new quadratic polynomial f' such that $Z_q(f') = 2Z_q(f)$, and every cross term and linear term that involves $x_1, \ldots, x_t$ has an even coefficient in f'. We can then repeat this substitution procedure on f': We either show that $Z_q(f') = 0$ or get a quadratic polynomial f'' such that $Z_q(f'') = 2Z_q(f')$ and the parameter t increases by at least one. So given a quadratic polynomial f, we can, in polynomial time, either show that $Z_q(f) = 0$ or get a new quadratic $g \in \mathbb{Z}_q[x_1, \ldots, x_n]$ such that $Z_q(f) = 2^r \cdot Z_q(g)$ for some known integer $r \in [0 : n]$, and every cross term and linear term has an even coefficient in g.

Now it suffices to compute $Z_q(g)$. We show that given such a polynomial g in n variables, we can reduce it to either $\mathsf{EVAL}(2^{k-1}) = \mathsf{EVAL}(q/2)$ or to the computation of $Z_q(g')$, in which g' is a quadratic polynomial in $n-1$ variables. Let

$$g = \sum_{i \le j \in [n]} a_{i,j} x_i x_j + \sum_{i \in [n]} a_i x_i + a.$$

We consider two cases: $a_{i,i}$ is even for all $i \in [n]$, or at least one of the $a_{i,i}$'s is odd. In the first case, $a_{i,j}$ and a_i are even for all $i \le j \in [n]$. Let $a'_{i,j}$ and a'_i denote integers in $[0 : 2^{k-1} - 1]$ that satisfy $a_{i,j} \equiv 2a'_{i,j},\ a_i \equiv 2a'_i \pmod q$, respectively. Then,

$$Z_q(g) = \omega_q^a \cdot \sum_{x_1,\ldots,x_n \in \mathbb{Z}_q} \omega_q^{2\left(\sum_{i \le j \in [n]} a'_{i,j} x_i x_j + \sum_{i \in [n]} a'_i x_i\right)} = 2^n \cdot \omega_q^a \cdot Z_{2^{k-1}}(g'),$$

where g' is the quadratic polynomial over $\mathbb{Z}_{q/2} = \mathbb{Z}_{2^{k-1}}$ in the exponent. This reduces the computation of $Z_q(g)$ to $Z_{q/2}(g')$.

In the second case, without loss of generality, we assume $a_{1,1}$ is odd. Then

$$f = a_{1,1}(x_1^2 + 2x_1 g_1) + g_2 = a_{1,1}(x_1 + g_1)^2 + g',$$

where g_1 is an affine form and g_2, g' are quadratic polynomials, all of which are over $x_2, \ldots, x_n$. We are able to do this because $a_{1,j}$ and a_1, $j \ge 2$, are even. Now

$$Z_q(g) = \sum_{x_2,\ldots,x_n \in \mathbb{Z}_q} \omega_q^{g'} \cdot \sum_{x_1 \in \mathbb{Z}_q} \omega_q^{a_{1,1}(x_1+g_1)^2} = Z_q(g') \sum_{x \in \mathbb{Z}_q} \omega_q^{a_{1,1}x^2}.$$

The last equation is because the sum over $x_1 \in \mathbb{Z}_q$ is independent of the value of g_1. This reduces $Z_q(g)$ to $Z_q(g')$ in which g' is a quadratic polynomial in $n-1$ variables.

To sum up, given any quadratic polynomial f, we can, in polynomial time, either output the correct value of $Z_q(f)$ or reduce one of the two parameters, k or n, by at least one. This gives us a polynomial time algorithm to evaluate $Z_q(f)$. □

This concludes the proof of Theorem 1.1 for the bipartite case.

Remark 12.10. Back in section 1, we mentioned that Holant(Ω) for $\Omega = (G, \mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3)$ are all tractable, and the tractability boils down to the exponential sum

$$\sum_{x_1,x_2,\ldots,x_n \in \{0,1\}} i^{L_1 + L_2 + \cdots + L_s} \tag{12.26}$$

being computable in polynomial time. This can also be derived from Theorem 12.1.

First, each mod 2 sum L_j in (12.26) can be replaced by its square $(L_j)^2$, because $L_j = 0, 1 \pmod 2$ iff $(L_j)^2 = 0, 1 \pmod 4$, respectively. So, (12.26) can be expressed as a sum of the form $i^{Q(x_1,x_2,\ldots,x_n)}$, where Q is an ordinary sum of squares of affine forms with integer coefficients and, in particular, a quadratic polynomial with integer coefficients. For a sum of squares of affine forms Q, if we evaluate each $x_i \in \{0,1,2,3\}$, we may take $x_i \bmod 2$, and this reduces (12.26) to $\mathsf{EVAL}(4)$:

$$\sum_{x_1,x_2,\ldots,x_n \in \mathbb{Z}_4} i^{Q(x_1,x_2\ldots,x_n)} = 2^n \sum_{x_1,x_2,\ldots,x_n \in \{0,1\}} i^{Q(x_1,x_2,\ldots,x_n)}.$$

13. Proof of Theorem 6.3. Let $\mathbf{A}$ be a symmetric, nonbipartite, and purified matrix. After collecting its entries of equal norm in decreasing order (by permuting the rows and columns of $\mathbf{A}$), there exist a positive integer N and two sequences $\boldsymbol{\kappa}$ and $\mathbf{m}$ such that $(\mathbf{A}, (N, \boldsymbol{\kappa}, \mathbf{m}))$ satisfies the following condition:

$(\mathcal{S}_1')$ $\mathbf{A}$ is an $m \times m$ symmetric matrix. $\boldsymbol{\kappa} = (\kappa_1, \ldots, \kappa_s)$ is a strictly decreasing sequence of positive rational numbers, where $s \geq 1$. $\mathbf{m} = (m_1, \ldots, m_s)$ is a sequence of positive integers such that $m = \sum m_i$. The rows and columns of $\mathbf{A}$ are indexed by $\mathbf{x} = (x_1, x_2)$, where $x_1 \in [s]$ and $x_2 \in [m_{x_1}]$. For all $\mathbf{x}, \mathbf{y}$, $\mathbf{A}$ satisfies

$$A_{\mathbf{x},\mathbf{y}} = A_{(x_1,x_2),(y_1,y_2)} = \kappa_{x_1}\kappa_{y_1} S_{\mathbf{x},\mathbf{y}},$$

where $\mathbf{S} = \{S_{\mathbf{x},\mathbf{y}}\}$ is a symmetric matrix in which every entry is a power of ω_N:

$$\mathbf{A} = \begin{pmatrix} \kappa_1\mathbf{I}_{m_1} & & & \\ & \kappa_2\mathbf{I}_{m_2} & & \\ & & \ddots & \\ & & & \kappa_s\mathbf{I}_{m_s} \end{pmatrix} \begin{pmatrix} \mathbf{S}_{(1,*),(1,*)} & \mathbf{S}_{(1,*),(2,*)} & \cdots & \mathbf{S}_{(1,*),(s,*)} \\ \mathbf{S}_{(2,*),(1,*)} & \mathbf{S}_{(2,*),(2,*)} & \cdots & \mathbf{S}_{(2,*),(s,*)} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{S}_{(s,*),(1,*)} & \mathbf{S}_{(s,*),(2,*)} & \cdots & \mathbf{S}_{(s,*),(s,*)} \end{pmatrix} \begin{pmatrix} \kappa_1\mathbf{I}_{m_1} & & & \\ & \kappa_2\mathbf{I}_{m_2} & & \\ & & \ddots & \\ & & & \kappa_s\mathbf{I}_{m_s} \end{pmatrix},$$

where $\mathbf{I}_{m_i}$ is the $m_i \times m_i$ identity matrix. We let $I = \{(i,j) : i \in [s], j \in [m_i]\}$.

The proof of Theorem 6.3, just like the one of Theorem 5.3, consists of five steps. All the proofs use the following strategy. We construct from the $m \times m$ matrix $\mathbf{A}$ its bipartization $\mathbf{A}'$, a $2m \times 2m$ symmetric matrix. Then we just apply the lemmas for the bipartite case to $\mathbf{A}'$ and show that $\mathbf{A}'$ is either #P-hard or has certain properties. Finally, we use these properties of $\mathbf{A}'$ to derive properties of $\mathbf{A}$.

To this end, we need the following lemma.

LEMMA 13.1. *Let $\mathbf{A}$ be a symmetric matrix, and let $\mathbf{A}'$ be its bipartization. Then* $\mathsf{EVAL}(\mathbf{A}') \leq \mathsf{EVAL}(\mathbf{A})$.

Proof. Suppose $\mathbf{A}$ is an $m \times m$ matrix. Let G be a connected undirected graph. If G is not bipartite, then $Z_{\mathbf{A}'}(G)$ is trivially 0, because $\mathbf{A}'$ is the bipartization of $\mathbf{A}$. Otherwise, assume that $G = (U \cup V, E)$ is bipartite and connected; let $u^* \in U$. Then

$$Z_{\mathbf{A}}(G, u^*, i) = Z_{\mathbf{A}'}(G, u^*, i) = Z_{\mathbf{A}'}(G, u^*, m+i) \quad \text{for any } i \in [m].$$

It then follows that $Z_{\mathbf{A}'}(G) = 2Z_{\mathbf{A}}(G)$ and $\mathsf{EVAL}(\mathbf{A}') \leq \mathsf{EVAL}(\mathbf{A})$. ∎

13.1. Step 2.1.

LEMMA 13.2. *Suppose that $(\mathbf{A}, (N, \boldsymbol{\kappa}, \mathbf{m}))$ satisfies $(\mathcal{S}_1')$. Then either $\mathsf{EVAL}(\mathbf{A})$ is #P-hard or $(\mathbf{A}, (N, \boldsymbol{\kappa}, \mathbf{m}))$ satisfies the following condition:*

$(\mathcal{S}_2')$ *For all $\mathbf{x}, \mathbf{x}' \in I$, either there exists an integer k such that $\mathbf{S}_{\mathbf{x},*} = \omega_N^k \cdot \mathbf{S}_{\mathbf{x}',*}$, or for every $j \in [s]$, $\langle \mathbf{S}_{\mathbf{x},(j,*)}, \mathbf{S}_{\mathbf{x}',(j,*)} \rangle = 0$.*

Proof. Let $\mathbf{A}'$ be the bipartization of $\mathbf{A}$. Suppose that $\mathsf{EVAL}(\mathbf{A})$ is not #P-hard. From Lemma 13.1, $\mathsf{EVAL}(\mathbf{A}') \leq \mathsf{EVAL}(\mathbf{A})$ and thus $\mathsf{EVAL}(\mathbf{A}')$ is not #P-hard. Note that the $\mathbf{S}$ matrix in Lemma 8.5 is exactly the same $\mathbf{S}$ here. Also $(\mathbf{A}', (N, \boldsymbol{\kappa}, \boldsymbol{\kappa}, \mathbf{m}, \mathbf{m}))$ satisfies condition $(\mathcal{S}_1)$, so by Lemma 8.5 together with the assumption that $\mathbf{A}'$ is not #P-hard, $\mathbf{S}$ satisfies $(\mathcal{S}_2)$ which is exactly the same as $(\mathcal{S}_2')$ here. (For Lemma 8.5, $\mathbf{S}$ also needs to satisfy $(\mathcal{S}_3)$, but since $\mathbf{S}$ is symmetric here, $(\mathcal{S}_3)$ is the same as $(\mathcal{S}_2)$.) □

We have the following corollary. The proof is the same as that of Corollary 8.6.

COROLLARY 13.3. *For all $i, j \in [s]$, $\mathbf{S}_{(i,*),(j,*)}$ has the same rank as $\mathbf{S}$.*

Next we build a pair $(\mathbf{F}, \mathfrak{D})$ and apply the cyclotomic reduction lemma on $\mathbf{A}$.

Let $h = \operatorname{rank}(\mathbf{S})$. By Corollary 13.3, there exist $1 \leq i_1 < \cdots < i_h \leq m_1$ such that the $\{(1, i_1), \ldots, (1, i_h)\} \times \{(1, i_1), \ldots, (1, i_h)\}$ submatrix of $\mathbf{S}$ has full rank h (using the fact that $\mathbf{S}$ is symmetric). Without loss of generality (if this is not the case, we can apply an appropriate permutation Π to the rows and columns of $\mathbf{A}$ so that the new $\mathbf{S}$ has this property), assume $i_k = k$ for all $k \in [h]$. Let $\mathbf{H}$ denote this $h \times h$ symmetric matrix: $H_{i,j} = S_{(1,i),(1,j)}$. From Corollary 13.3 and Lemma 13.2, for every index $\mathbf{x} \in I$, there exist two unique integers $j \in [h]$ and $k \in [0 : N-1]$ such that

$$\mathbf{S}_{\mathbf{x},*} = \omega_N^k \cdot \mathbf{S}_{(1,j),*} \quad \text{and} \quad \mathbf{S}_{*,\mathbf{x}} = \omega_N^k \cdot \mathbf{S}_{*,(1,j)}. \tag{13.1}$$

This gives us a partition of the index set I

$$\mathscr{R} = \big\{R_{(i,j),k} : i \in [s], j \in [h], k \in [0 : N-1]\big\}.$$

For every $\mathbf{x} \in I$, $\mathbf{x} \in R_{(i,j),k}$ iff $i = x_1$ and $\mathbf{x}, j, k$ satisfy (13.1). By Corollary 13.3,

$$\bigcup_{k \in [0:N-1]} R_{(i,j),k} \neq \emptyset \quad \text{for all } i \in [s] \text{ and } j \in [h].$$

Now we define $(\mathbf{F}, \mathfrak{D})$ and use the cyclotomic reduction lemma and $\mathscr{R}$ to show that $\mathsf{EVAL}(\mathbf{F}, \mathfrak{D}) \equiv \mathsf{EVAL}(\mathbf{A})$. First, $\mathbf{F}$ is an $sh \times sh$ matrix. We use $I' = [s] \times [h]$ to index the rows and columns of $\mathbf{F}$. Then

$$F_{\mathbf{x},\mathbf{y}} = \kappa_{x_1} \kappa_{y_1} H_{x_2,y_2} = \kappa_{x_1} \kappa_{y_1} S_{(1,x_2),(1,y_2)} \quad \text{for all } \mathbf{x}, \mathbf{y} \in I',$$

or equivalently,

$$\mathbf{F} = \begin{pmatrix} \kappa_1\mathbf{I} & & & \\ & \kappa_2\mathbf{I} & & \\ & & \ddots & \\ & & & \kappa_s\mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{H} & \mathbf{H} & \ldots & \mathbf{H} \\ \mathbf{H} & \mathbf{H} & \ldots & \mathbf{H} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{H} & \mathbf{H} & \ldots & \mathbf{H} \end{pmatrix} \begin{pmatrix} \kappa_1\mathbf{I} & & & \\ & \kappa_2\mathbf{I} & & \\ & & \ddots & \\ & & & \kappa_s\mathbf{I} \end{pmatrix},$$

where $\mathbf{I}$ is the $h \times h$ identity matrix.

Second, $\mathfrak{D} = (\mathbf{D}^{[0]}, \ldots, \mathbf{D}^{[N-1]})$ is a sequence of N diagonal matrices of the same size as $\mathbf{F}$. We use I' to index its diagonal entries. The $\mathbf{x}$th entries are

$$D_{\mathbf{x}}^{[r]} = \sum_{k=0}^{N-1} \big|R_{(x_1,x_2),k}\big| \cdot \omega_N^{kr} \quad \text{for all } r \in [0 : N-1], \mathbf{x} \in I'.$$

We use the cyclotomic reduction lemma (Lemma 8.2) to prove the next lemma.

LEMMA 13.4. $\mathsf{EVAL}(\mathbf{A}) \equiv \mathsf{EVAL}(\mathbf{F}, \mathfrak{D})$.

Proof. Let $\mathbf{x}, \mathbf{y} \in I$, $\mathbf{x} \in R_{(x_1,j),k}$ and $\mathbf{y} \in R_{(y_1,j'),k'}$ for some j, k, j', k'. By (13.1),

$$A_{\mathbf{x},\mathbf{y}} = \kappa_{x_1}\kappa_{y_1} S_{\mathbf{x},\mathbf{y}} = \kappa_{x_1}\kappa_{y_1} S_{(1,j),(1,j')} \cdot \omega_N^{k+k'} = F_{(x_1,j),(y_1,j')} \cdot \omega_N^{k+k'}.$$

So $\mathbf{A}$ can be generated from $\mathbf{F}$ using $\mathscr{R}$. The construction of $\mathfrak{D}$ implies that $\mathfrak{D}$ can be generated from $\mathscr{R}$. The lemma follows from the cyclotomic reduction lemma. ∎

13.2. Steps 2.2 and 2.3. Now we have a pair $(\mathbf{F}, \mathfrak{D})$ that satisfies the following condition (*Shape*′):

(*Shape*′$_1$) $\mathbf{F} \in \mathbb{C}^{m\times m}$ is a symmetric $s \times s$ block matrix. (The m here is different from the m used in Step 2.1.) We use $I = [s] \times [h]$ to index its rows and columns.

(*Shape*′$_2$) There are a strictly decreasing sequence $\boldsymbol{\kappa} = (\kappa_1, \ldots, \kappa_s)$ of positive rational numbers together with an $h \times h$ matrix $\mathbf{H}$ of full rank, whose entries are all powers of ω_N, for some $N \geq 1$. We have

$$F_{\mathbf{x},\mathbf{y}} = \kappa_{x_1}\kappa_{y_1} H_{x_2,y_2} \quad \text{for all } \mathbf{x}, \mathbf{y} \in I.$$

(*Shape*′$_3$) $\mathfrak{D} = (\mathbf{D}^{[0]}, \ldots, \mathbf{D}^{[N-1]})$ is a sequence of N $m \times m$ diagonal matrices. $\mathfrak{D}$ satisfies $(\mathcal{T}_3)$, so for all $r \in [N-1]$ and $\mathbf{x} \in I$, we have

$$D^{[r]}_{\mathbf{x}} = \overline{D^{[N-r]}_{\mathbf{x}}}.$$

Now suppose $\mathsf{EVAL}(\mathbf{F}, \mathfrak{D})$ is not #P-hard.

We define $(\mathbf{C}, \mathfrak{D}')$: $\mathbf{C}$ is the bipartization of $\mathbf{F}$; $\mathfrak{D}'$ is a sequence of N copies of

$$\begin{pmatrix} \mathbf{D}^{[r]} & \\ & \mathbf{D}^{[r]} \end{pmatrix}.$$

The proof of the following lemma is the same as that of Lemma 13.1.

LEMMA 13.5. $\mathsf{EVAL}(\mathbf{C}, \mathfrak{D}') \leq \mathsf{EVAL}(\mathbf{F}, \mathfrak{D})$.

By Lemma 13.5, $\mathsf{EVAL}(\mathbf{C}, \mathfrak{D}') \leq \mathsf{EVAL}(\mathbf{F}, \mathfrak{D})$ and thus $\mathsf{EVAL}(\mathbf{C}, \mathfrak{D}')$ is not #P-hard. By (*Shape*′$_1$)–(*Shape*′$_3$), $(\mathbf{C}, \mathfrak{D}')$ also satisfies (*Shape*$_1$)–(*Shape*$_3$). It then follows from Lemmas 8.8 and 8.11 that $(\mathbf{C}, \mathfrak{D}')$ also satisfies (*Shape*$_4$)–(*Shape*$_6$). Since $(\mathbf{C}, \mathfrak{D}')$ is built from $(\mathbf{F}, \mathfrak{D})$, the latter must satisfy the following conditions:

(*Shape*′$_4$) $\mathbf{H}/\sqrt{h}$ is unitary: $\langle \mathbf{H}_{i,*}, \mathbf{H}_{j,*} \rangle = \langle \mathbf{H}_{*,i}, \mathbf{H}_{*,j} \rangle = 0$ for all $i \neq j \in [h]$.

(*Shape*′$_5$) For all $\mathbf{x} \in I$,

$$D^{[0]}_{\mathbf{x}} = D^{[0]}_{(x_1,1)}.$$

(*Shape*′$_6$) For each $r \in [N-1]$, there are diagonal matrices $\mathbf{K}^{[r]} \in \mathbb{C}^{s\times s}, \mathbf{L}^{[r]} \in \mathbb{C}^{h\times h}$. The norm of every diagonal entry in $\mathbf{L}^{[r]}$ is either 0 or 1. We have

$$\mathbf{D}^{[r]} = \mathbf{K}^{[r]} \otimes \mathbf{L}^{[r]} \quad \text{for all } r \in [N-1].$$

For all $r \in [N-1]$, $\mathbf{K}^{[r]} = \mathbf{0}$ implies $\mathbf{L}^{[r]} = \mathbf{0}$; $\mathbf{L}^{[r]} \neq \mathbf{0}$ implies one of its entries is 1.

In particular, (*Shape*′$_5$) means that by setting

$$K^{[0]}_i = D^{[0]}_{(i,1)} \quad \text{and} \quad L^{[0]}_j = 1 \quad \text{for all } i \in [s] \text{ and } j \in [h],$$

we have $\mathbf{D}^{[0]} = \mathbf{K}^{[0]} \otimes \mathbf{L}^{[0]}$. By $(\mathcal{T}_3)$ in (*Shape*′$_3$), entries of $\mathbf{K}^{[0]}$ are positive integers.

13.3. Step 2.4. Suppose $(\mathbf{F}, \mathfrak{D})$ satisfies $(Shape'_1)$–$(Shape'_6)$. From $(Shape'_2)$ we have $\mathbf{F} = \mathbf{M} \otimes \mathbf{H}$, where $\mathbf{M}$ is an $s \times s$ matrix of rank 1: $M_{i,j} = \kappa_i \kappa_j$ for all $i, j \in [s]$.

We reduce $\mathsf{EVAL}(\mathbf{F}, \mathfrak{D})$ to two problems $\mathsf{EVAL}(\mathbf{M}, \mathfrak{K})$ and $\mathsf{EVAL}(\mathbf{H}, \mathfrak{L})$, where

$$\mathfrak{K} = \left(\mathbf{K}^{[0]}, \ldots, \mathbf{K}^{[N-1]}\right) \quad \text{and} \quad \mathfrak{L} = \left(\mathbf{L}^{[0]}, \ldots, \mathbf{L}^{[N-1]}\right).$$

The proof of the following lemma is essentially the same as that of Lemma 8.24.

LEMMA 13.6. $\mathsf{EVAL}(\mathbf{F}, \mathfrak{D}) \equiv \mathsf{EVAL}(\mathbf{H}, \mathcal{L})$.

13.4. Step 2.5. Finally we normalize the matrix $\mathbf{H}$ in the same way we did for the bipartite case and obtain a new pair that (1) satisfies conditions $(\mathcal{U}'_1)$–$(\mathcal{U}'_4)$ and (2) is polynomial-time equivalent to $\mathsf{EVAL}(\mathbf{H}, \mathcal{L})$.

14. Proofs of Theorems 6.4 and 6.7. Suppose $((M, N), \mathbf{F}, \mathfrak{D})$ satisfies $(\mathcal{U}'_1)$–$(\mathcal{U}'_4)$. We prove Theorems 6.4 and 6.7 in this section. We first prove that if $\mathbf{F}$ does not satisfy the group condition $(\mathcal{GC})$, then $\mathsf{EVAL}(\mathbf{F}, \mathfrak{D})$ is #P-hard. This is done by applying Lemma 9.1 (for the bipartite case) to the bipartization $\mathbf{C}$ of $\mathbf{F}$.

LEMMA 14.1. *Suppose $((M, N), \mathbf{F}, \mathfrak{D})$ satisfies conditions $(\mathcal{U}'_1)$–$(\mathcal{U}'_4)$. Then either the matrix $\mathbf{F}$ satisfies the group condition $(\mathcal{GC})$ or $\mathsf{EVAL}(\mathbf{F}, \mathfrak{D})$ is #P-hard.*

Proof. Assume $\mathsf{EVAL}(\mathbf{F}, \mathfrak{D})$ is not #P-hard. Let $\mathbf{C}$ and $\mathfrak{E} = (\mathbf{E}^{[0]}, \ldots, \mathbf{E}^{[N-1]})$ be

$$\mathbf{C} = \begin{pmatrix} \mathbf{0} & \mathbf{F} \\ \mathbf{F} & \mathbf{0} \end{pmatrix} \quad \text{and} \quad \mathbf{E}^{[r]} = \begin{pmatrix} \mathbf{D}^{[r]} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}^{[r]} \end{pmatrix} \quad \text{for all } r \in [0 : N-1].$$

By $(\mathcal{U}'_1)$–$(\mathcal{U}'_4)$, $((M, N), \mathbf{C}, \mathfrak{E})$ satisfies $(\mathcal{U}_1)$–$(\mathcal{U}_4)$. Furthermore, using Lemma 13.5, we have $\mathsf{EVAL}(\mathbf{C}, \mathfrak{E}) \leq \mathsf{EVAL}(\mathbf{F}, \mathfrak{D})$ and thus $\mathsf{EVAL}(\mathbf{C}, \mathfrak{E})$ is also not #P-hard. It follows from Lemma 9.1 that $\mathbf{F}$ satisfies the group condition $(\mathcal{GC})$. ∎

14.1. Proof of Theorem 6.4. We prove Theorem 6.4 again, using $\mathbf{C}$ and $\mathfrak{E}$ again.

Suppose $\mathsf{EVAL}(\mathbf{F}, \mathfrak{D})$ is not #P-hard. On the one hand, $\mathsf{EVAL}(\mathbf{C}, \mathfrak{E}) \leq \mathsf{EVAL}(\mathbf{F}, \mathfrak{D})$ and $\mathsf{EVAL}(\mathbf{C}, \mathfrak{E})$ is also not #P-hard. On the other hand, $((M, N), \mathbf{C}, \mathfrak{E})$ satisfies conditions $(\mathcal{U}_1)$–$(\mathcal{U}_4)$. Thus, using Theorem 5.4, $\mathfrak{E}$ must satisfy $(\mathcal{U}_5)$: Every entry of $\mathbf{E}^{[r]}$, $r \in [N-1]$, is either 0 or a power of ω_N. It then follows directly that every entry of $\mathbf{D}^{[r]}$, $r \in [N-1]$, is either 0 or a power of ω_N.

14.2. Proof of Theorem 6.7. In this section we prove Theorem 6.7. However, we cannot simply reduce it, using $(\mathbf{C}, \mathfrak{E})$, to the bipartite case (Theorem 5.6), because in Theorem 6.7, we are only allowed to permute the rows and columns symmetrically, while in Theorem 5.6, one can use two different permutations to permute the rows and columns. But as we will see below, for most of the lemmas we need here, their proofs are exactly the same as those for the bipartite case. The only exception is the counterpart of Lemma 9.7, in which we have to bring in the generalized Fourier matrices (see Definitions 5.5 and 6.6).

Suppose $\mathbf{F}$ satisfies $(\mathcal{GC})$. Let F^{R} denote the set of row vectors $\{\mathbf{F}_{i,*}\}$ of $\mathbf{F}$ and F^{C} denote the set of column vectors $\{\mathbf{F}_{*,j}\}$ of $\mathbf{F}$. Since $\mathbf{F}$ satisfies $(\mathcal{GC})$, by Property 9.2, both F^{R} and F^{C} are finite Abelian groups of order m, under the Hadamard product.

We start by proving a symmetric version of Lemma 9.5, stating that when $M = pq$ and $\gcd(p, q) = 1$ (note that p and q are not necessarily primes), a permutation of $\mathbf{F}$ is the tensor product of two smaller discrete unitary matrices, both of which satisfy the group condition.

LEMMA 14.2. *Suppose $\mathbf{F} \in \mathbb{C}^{m \times m}$ is symmetric and M-discrete unitary and satisfies $(\mathcal{GC})$. Moreover, $M = pq$, $p, q > 1$, and $\gcd(p, q) = 1$. Then there is a*

permutation Π *of* $[0:m-1]$ *such that* $\mathbf{F}_{\Pi,\Pi} = \mathbf{F}' \otimes \mathbf{F}''$, *where* $\mathbf{F}'$ *is a symmetric* p*-discrete unitary matrix,* $\mathbf{F}''$ *is a symmetric* q*-discrete unitary matrix, and both of them satisfy* $(\mathcal{GC})$.

Proof. The proof is almost the same as that of Lemma 9.5. Since $\mathbf{F}$ is symmetric the two bijections f, g that we defined in the proof of Lemma 9.5, from $[0:m-1]$ to $[0:m'-1] \times [0:m''-1]$, are exactly the same. ∎

As a result, we only need to deal with the case when $M = p^\beta$ is a prime power.

LEMMA 14.3. *Suppose* $\mathbf{F} \in \mathbb{C}^{m\times m}$ *is symmetric and* M*-discrete unitary and satisfies* $(\mathcal{GC})$. *Moreover,* $M = p^\beta$ *is a prime power,* $p \neq 2$, *and* $\beta \geq 1$. *Then there must exist an integer* $k \in [0:m-1]$ *such that* $p \nmid \alpha_{k,k}$, *where* $F_{k,k} = \omega_M^{\alpha_{k,k}}$.

Proof. For $i, j \in [0:m-1]$, we let $\alpha_{i,j}$ denote the integer in $[0:M-1]$ such that $F_{i,j} = \omega_M^{\alpha_{i,j}}$. Assume the lemma is not true, that is, $p \mid \alpha_{k,k}$ for all k. Then because $\mathbf{F}$ is M-discrete unitary, there must exist $i \neq j \in [0:m-1]$ such that $p \nmid \alpha_{i,j}$. Without loss of generality, we assume $p \nmid \alpha_{2,1} = \alpha_{1,2}$.

By $(\mathcal{GC})$, there exists a $k \in [0:m-1]$ such that $\mathbf{F}_{k,*} = \mathbf{F}_{1,*} \circ \mathbf{F}_{2,*}$. However,

$$\omega_M^{\alpha_{k,k}} = F_{k,k} = F_{1,k}F_{2,k} = F_{k,1}F_{k,2} = F_{1,1}F_{2,1}F_{1,2}F_{2,2} = \omega_M^{\alpha_{1,1}+\alpha_{2,2}+2\alpha_{1,2}},$$

and $\alpha_{k,k} \equiv \alpha_{1,1} + \alpha_{2,2} + 2\alpha_{1,2} \pmod{M}$ implies that $0 \equiv 0 + 0 + 2\alpha_{1,2} \pmod{p}$. Since $p \neq 2$ and $p \nmid \alpha_{1,2}$, we get a contradiction. ∎

The next lemma is the symmetric version of Lemma 9.7 showing that when there exists a diagonal entry $F_{k,k}$ such that $p \nmid \alpha_{k,k}$, $\mathbf{F}$ is the tensor product of a Fourier matrix and a discrete unitary matrix. Note that this lemma also applies to the case when $p = 2$. So the only case left is when $p = 2$ but $2 \mid \alpha_{i,i}$ for all $i \in [0:m-1]$.

LEMMA 14.4. *Suppose* $\mathbf{F} \in \mathbb{C}^{m\times m}$ *is symmetric and* M*-discrete unitary and satisfies* $(\mathcal{GC})$. *Moreover,* $M = p^\beta$ *is a prime power. If there exists a* $k \in [0:m-1]$ *such that* $F_{k,k} = \omega_M^\alpha$ *and* $p \nmid \alpha$, *then there exists a permutation* Π *such that* $\mathbf{F}_{\Pi,\Pi} = \boldsymbol{\mathcal{F}}_{M,\alpha} \otimes \mathbf{F}'$, *where* $\mathbf{F}'$ *is a symmetric and* M'*-discrete unitary matrix that satisfies condition* $(\mathcal{GC})$ *with* $M' \mid M$.

Proof. The proof is the same as the one of Lemma 9.7 by setting $a = b = k$. The only thing to notice is that since $\mathbf{F}$ is symmetric, the two bijections f and g that we defined in the proof of Lemma 9.7 are the same. Thus, the row permutation and the column permutation applied on $\mathbf{F}$ are the same. Since $F_{k,k} = \omega_M^\alpha$, (9.12) becomes

$$G_{(x_1,x_2),(y_1,y_2)} = \omega_M^{\alpha x_1 y_1} \cdot G_{(0,x_2),(0,y_2)}.$$

This explains why we need to use the Fourier matrix $\boldsymbol{\mathcal{F}}_{M,\alpha}$ here. ∎

Finally, we deal with the case when $p = 2$ and $2 \mid \alpha_{i,i}$ for all $i \in [0:m-1]$.

LEMMA 14.5. *Suppose* $\mathbf{F} \in \mathbb{C}^{m\times m}$ *is symmetric and* M*-discrete unitary and satisfies* $(\mathcal{GC})$ *with* $M = 2^\beta$ *and* $2 \mid \alpha_{i,i}$ *for all* $i \in [0:m-1]$. *Then there exist a permutation* Π *and a* 2×2 *symmetric nondegenerate matrix* $\mathbf{W}$ *over* $\mathbb{Z}_M$ (see section 6.3.2 and Definition 6.6), *such that* $\mathbf{F}_{\Pi,\Pi} = \boldsymbol{\mathcal{F}}_{M,\mathbf{W}} \otimes \mathbf{F}'$, *where* $\mathbf{F}'$ *is a symmetric,* M'*-discrete unitary matrix that satisfies* $(\mathcal{GC})$ *with* $M' \mid M$.

Proof. By Property 9.6, there are two integers $a \neq b$ such that $F_{a,b} = F_{b,a} = \omega_M$. Let $F_{a,a} = \omega^{\alpha_a}$ and $F_{b,b} = \omega^{\alpha_b}$. The assumption of the lemma implies that $2 \mid \alpha_a, \alpha_b$.

We let $S^{a,b}$ denote the following subset of F^{R}:

$$S^{a,b} = \{\mathbf{u} \in F^{\mathrm{R}} : u_a = u_b = 1\}.$$

Clearly $S^{a,b}$ is a subgroup of F^{R}. On the other hand, let S^a denote the subgroup of F^{R} that is generated by $\mathbf{F}_{a,*}$, and let S^b denote the subgroup generated by $\mathbf{F}_{b,*}$:

$$S^a = \{(\mathbf{F}_{a,*})^0, (\mathbf{F}_{a,*})^1, \ldots, (\mathbf{F}_{a,*})^{M-1}\} \quad \text{and} \quad S^b = \{(\mathbf{F}_{b,*})^0, (\mathbf{F}_{b,*})^1, \ldots, (\mathbf{F}_{b,*})^{M-1}\}.$$

We have $|S^a| = |S^b| = M$ since $F_{a,b} = \omega_M$. It is clear that $(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3) \mapsto \mathbf{u}_1 \circ \mathbf{u}_2 \circ \mathbf{u}_3$ is a group homomorphism from $S^a \times S^b \times S^{a,b}$ to F^{R}. We show that it is surjective.

Toward this end, we first note that

$$\mathbf{W} = \begin{pmatrix} \alpha_a & 1 \\ 1 & \alpha_b \end{pmatrix}$$

is nondegenerate. This follows from Lemma 6.5, since $\det(\mathbf{W}) = \alpha_a\alpha_b - 1$ is odd.

First, we show that $(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3) \mapsto \mathbf{u}_1 \circ \mathbf{u}_2 \circ \mathbf{u}_3$ is surjective. This is because for any $\mathbf{u} \in F^{\mathrm{R}}$, there exist integers k_1 and k_2 such that (since $\mathbf{W}$ is nondegenerate, by Lemma 6.5, $\mathbf{x} \mapsto \mathbf{W}\mathbf{x}$ is a bijection)

$$u_a = F_{a,a}^{k_1} \cdot F_{b,a}^{k_2} = \omega_M^{\alpha_a k_1 + k_2} \quad \text{and} \quad u_b = F_{a,b}^{k_1} \cdot F_{b,b}^{k_2} = \omega_M^{k_1 + \alpha_b k_2}.$$

Thus, $\mathbf{u} \circ \overline{\mathbf{F}_{a,*}^{k_1}} \circ \overline{\mathbf{F}_{b,*}^{k_2}} \in S^{a,b}$. It follows that $\mathbf{u} = \mathbf{F}_{a,*}^{k_1} \circ \mathbf{F}_{b,*}^{k_2} \circ \mathbf{u}_3$ for some $\mathbf{u}_3 \in S^{a,b}$.

Second, we show that it is also injective. Assume this is not the case. Then there exist $k_1, k_2, k_1', k_2' \in \mathbb{Z}_M$, and $\mathbf{u}, \mathbf{u}' \in S^{a,b}$ such that $(k_1, k_2, \mathbf{u}) \neq (k_1', k_2', \mathbf{u}')$ but

$$(\mathbf{F}_{a,*})^{k_1} \circ (\mathbf{F}_{b,*})^{k_2} \circ \mathbf{u} = (\mathbf{F}_{a,*})^{k_1'} \circ (\mathbf{F}_{b,*})^{k_2'} \circ \mathbf{u}'.$$

If $k_1 = k_1'$ and $k_2 = k_2'$, then $\mathbf{u} = \mathbf{u}'$, contradiction. Therefore, we may assume that

$$\boldsymbol{\ell} = (\ell_1, \ell_2)^{\mathrm{T}} = (k_1 - k_1', k_2 - k_2')^{\mathrm{T}} \neq \mathbf{0}.$$

By restricting on the ath and bth entries, we get $\mathbf{W}\boldsymbol{\ell} = \mathbf{0}$. This contradicts the fact that $\mathbf{W}$ is nondegenerate.

Since $(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3) \mapsto \mathbf{u}_1 \circ \mathbf{u}_2 \circ \mathbf{u}_3$ is a group isomorphism, we have $|S^{a,b}| = m/M^2$, which we denote by n. Let $S^{a,b} = \{\mathbf{v}_0 = \mathbf{1}, \mathbf{v}_1, \ldots, \mathbf{v}_{n-1}\}$. There is a bijection f from $[0 : m-1]$ to $[0 : M-1] \times [0 : M-1] \times [0 : n-1]$, $f(i) = (f_1(i), f_2(i), f_3(i))$, with

$$\mathbf{F}_{i,*} = (\mathbf{F}_{a,*})^{f_1(i)} \circ (\mathbf{F}_{b,*})^{f_2(i)} \circ \mathbf{v}_{f_3(i)} \quad \text{for all } i \in [0 : m-1]. \tag{14.1}$$

Since $\mathbf{F}$ is symmetric, this also implies that

$$\mathbf{F}_{*,j} = (\mathbf{F}_{*,a})^{f_1(j)} \circ (\mathbf{F}_{*,b})^{f_2(j)} \circ \mathbf{v}_{f_3(j)} \quad \text{for all } j \in [0 : m-1]. \tag{14.2}$$

Note that $f(a) = (1, 0, 0)$ and $f(b) = (0, 1, 0)$.

Next we permute $\mathbf{F}$ to get a new matrix $\mathbf{G}$. For convenience, we use (x_1, x_2, x_3), where $x_1, x_2 \in [0 : M-1]$ and $x_3 \in [0 : n-1]$, to index the rows and columns of $\mathbf{G}$. We permute $\mathbf{F}$ using $\Pi(x_1, x_2, x_3) = f^{-1}(x_1, x_2, x_3)$:

$$G_{(x_1,x_2,x_3),(y_1,y_2,y_3)} = F_{\Pi(x_1,x_2,x_3),\Pi(y_1,y_2,y_3)}. \tag{14.3}$$

Then by (14.1) and (14.2),

$$\begin{aligned} \mathbf{G}_{(x_1,x_2,x_3),*} &= (\mathbf{G}_{(1,0,0),*})^{x_1} \circ (\mathbf{G}_{(0,1,0),*})^{x_2} \circ \mathbf{G}_{(0,0,x_3),*} \quad \text{and} \\ \mathbf{G}_{*,(y_1,y_2,y_3)} &= (\mathbf{G}_{*,(1,0,0)})^{y_1} \circ (\mathbf{G}_{*,(0,1,0)})^{y_2} \circ \mathbf{G}_{*,(0,0,y_3)}. \end{aligned}$$

As a result,

$$G_{(x_1,x_2,x_3),(y_1,y_2,y_3)} = (G_{(1,0,0),(y_1,y_2,y_3)})^{x_1} (G_{(0,1,0),(y_1,y_2,y_3)})^{x_2}\, G_{(0,0,x_3),(y_1,y_2,y_3)}.$$

We analyze the three factors. First, we have

$$G_{(1,0,0),(y_1,y_2,y_3)} = F_{a,a}^{y_1} \cdot F_{a,b}^{y_2} \cdot v_{y_3,a} = \omega_M^{\alpha_a y_1 + y_2},$$

where $v_{y_3,a}$ is the ath entry of $\mathbf{v}_{y_3}$. Similarly, $G_{(0,1,0),(y_1,y_2,y_3)} = \omega_M^{y_1+\alpha_b y_2}$. Second,

$$G_{(0,0,x_3),(y_1,y_2,y_3)} = (G_{(0,0,x_3),(1,0,0)})^{y_1} (G_{(0,0,x_3),(0,1,0)})^{y_2} G_{(0,0,x_3),(0,0,y_3)}.$$

By (14.3) and (14.2) we have

$$G_{(0,0,x),(1,0,0)} = F_{\Pi(0,0,x),\Pi(1,0,0)} = F_{\Pi(0,0,x),a}.$$

Then by (14.1), $F_{\Pi(0,0,x),a} = v_{x,a} = 1$. Similarly, $G_{(0,0,x),(0,1,0)} = v_{x,b} = 1$. Therefore,

$$G_{(x_1,x_2,x_3),(y_1,y_2,y_3)} = \omega_M^{\alpha_a x_1 y_1 + x_1 y_2 + x_2 y_1 + \alpha_b x_2 y_2} \cdot G_{(0,0,x_3),(0,0,y_3)}.$$

So $\mathbf{G} = \boldsymbol{\mathcal{F}}_{M,\mathbf{W}} \otimes \mathbf{F}'$; $\mathbf{F}' \equiv (F'_{i,j} = G_{(0,0,i),(0,0,j)})$ is symmetric; $\mathbf{W}$ is nondegenerate.

The only thing left is to show $\mathbf{F}'$ is discrete unitary and satisfies $(\mathcal{GC})$. $\mathbf{F}'$ satisfies $(\mathcal{GC})$ because $S^{a,b}$ is a group and thus is closed under the Hadamard product. To see that $\mathbf{F}'$ is discrete unitary, we have

$$0 = \langle \mathbf{G}_{(0,0,i),*}, \mathbf{G}_{(0,0,j),*} \rangle = M^2 \cdot \langle \mathbf{F}'_{i,*}, \mathbf{F}'_{j,*} \rangle \quad \text{for any } i \neq j \in [0 : n-1].$$

Since $\mathbf{F}'$ is symmetric, columns $\mathbf{F}'_{*,i}$ and $\mathbf{F}'_{*,j}$ are also orthogonal. ∎

Theorem 6.7 then follows from Lemmas 14.3, 14.4, and 14.5.

15. Proofs of Theorems 6.8 and 6.9. Suppose $((M,N),\mathbf{F},\mathfrak{D},(\mathbf{d},\mathcal{W},\mathbf{p},\mathbf{t},\mathcal{Q},\mathcal{K}))$ satisfies condition $(\mathcal{R}')$. We prove Theorem 6.8: either $\mathsf{EVAL}(\mathbf{F},\mathfrak{D})$ is #P-hard or $\mathfrak{D}$ satisfies conditions $(\mathcal{L}'_1)$ and $(\mathcal{L}'_2)$.

Suppose $\mathsf{EVAL}(\mathbf{F},\mathfrak{D})$ is not #P-hard. We use $(\mathbf{C},\mathfrak{E})$ to denote the bipartization of $(\mathbf{F},\mathfrak{D})$. The plan is to show that $(\mathbf{C},\mathfrak{E})$ with appropriate $\mathbf{p}'$, $\mathbf{t}'$, and $\mathcal{Q}'$ satisfies $(\mathcal{R})$.

To see this, we permute $\mathbf{C}$ and $\mathfrak{E}$ using the following permutation Σ. We index the rows and columns of $\mathbf{C}$ and $\mathbf{E}^{[r]}$ using $\{0,1\} \times \mathbb{Z}_{\mathbf{d}}^2 \times \mathbb{Z}_{\mathcal{Q}}$. We set $\Sigma(1,\mathbf{y}) = (1,\mathbf{y})$ for all $\mathbf{y} \in \mathbb{Z}_{\mathbf{d}}^2 \times \mathbb{Z}_{\mathcal{Q}}$, that is, Σ fixes pointwise the second half of the rows and columns, and $\Sigma(0,\mathbf{x}) = (0,\mathbf{x}')$, where $\mathbf{x}'$ satisfies

$$x_{0,i,1} = W_{1,1}^{[i]} x'_{0,i,1} + W_{2,1}^{[i]} x'_{0,i,2}, \quad x_{0,i,2} = W_{1,2}^{[i]} x'_{0,i,1} + W_{2,2}^{[i]} x'_{0,i,2} \quad \text{for all } i \in [g],$$

and $x_{1,i,j} = k_{i,j} \cdot x'_{1,i,j}$ for all $i \in [s], j \in [t_i]$. See $(\mathcal{R}')$ for the definitions of these symbols.

Before proving properties of $\mathbf{C}_{\Sigma,\Sigma}$ and $\mathfrak{E}_\Sigma$, we need to verify that Σ is indeed a permutation. This follows from the fact that $\mathbf{W}^{[i]}$, for every $i \in [g]$, is nondegenerate over $\mathbb{Z}_{d_i}$, and $k_{i,j}$ for all $i \in [s]$ and $j \in [t_i]$ satisfies $\gcd(k_{i,j}, q_{i,j}) = 1$ (so $\mathbf{x}'$ above is unique). We use Σ_0 to denote the $(0,*)$-part of Σ and I to denote the identity map:

$$\Sigma(0,\mathbf{x}) = (0,\Sigma_0(\mathbf{x})) = (0,\mathbf{x}') \quad \text{for all } \mathbf{x} \in \mathbb{Z}_{\mathbf{d}}^2 \times \mathbb{Z}_{\mathcal{Q}}.$$

Now we can write $\mathbf{C}_{\Sigma,\Sigma}$ and $\mathfrak{E}_\Sigma = (\mathbf{E}_\Sigma^{[0]}, \ldots, \mathbf{E}_\Sigma^{[N-1]})$ as

$$\mathbf{C}_{\Sigma,\Sigma} = \begin{pmatrix} \mathbf{0} & \mathbf{F}_{\Sigma_0,I} \\ \mathbf{F}_{I,\Sigma_0} & \mathbf{0} \end{pmatrix} \quad \text{and} \quad \mathbf{E}_\Sigma^{[r]} = \begin{pmatrix} \mathbf{D}_{\Sigma_0}^{[r]} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}^{[r]} \end{pmatrix} \tag{15.1}$$

for all $r \in [0 : N-1]$. We make the following two observations: Observation 1: $\mathsf{EVAL}(\mathbf{C}_{\Sigma,\Sigma}, \mathfrak{E}_\Sigma) \equiv \mathsf{EVAL}(\mathbf{C}, \mathfrak{E}) \le \mathsf{EVAL}(\mathbf{F}, \mathfrak{D})$ and thus $\mathsf{EVAL}(\mathbf{C}_{\Sigma,\Sigma}, \mathfrak{E}_\Sigma)$ is not #P-hard. Observation 2: $\mathbf{F}_{\Sigma_0,I}$ satisfies

$$\left(\mathbf{F}_{\Sigma_0,I}\right)_{\mathbf{x},\mathbf{y}} = F_{\mathbf{x}',\mathbf{y}} = \prod_{i\in[g]} \omega_{d_i}^{(x'_{0,i,1}\ x'_{0,i,2})\cdot\mathbf{W}^{[i]}\cdot(y_{0,i,1}\ y_{0,i,2})^{\mathrm{T}}} \prod_{i\in[s],j\in[t_i]} \omega_{q_{i,j}}^{k_{i,j}\cdot x'_{1,i,j}y_{1,i,j}}$$
$$= \prod_{i\in[g]} \omega_{d_i}^{x_{0,i,1}y_{0,i,1}+x_{0,i,2}y_{0,i,2}} \prod_{i\in[s],j\in[t_i]} \omega_{q_{i,j}}^{x_{1,i,j}y_{1,i,j}}.$$

By Observation 2, it is easy to show that $\mathbf{C}_{\Sigma,\Sigma}$ and $\mathfrak{E}_\Sigma$ (together with appropriate $\mathbf{q}', \mathbf{t}', \mathcal{Q}'$) satisfy condition $(\mathcal{R})$. Since $\mathsf{EVAL}(\mathbf{C}_{\Sigma,\Sigma}, \mathfrak{E}_\Sigma)$ by Observation 1 is not #P-hard, it follows from Theorem 5.8 and (15.1) that $\mathbf{D}^{[r]}$ satisfy $(\mathcal{L}_2)$ and $(\mathcal{L}_3)$. This proves Theorem 6.8 since $(\mathcal{L}'_1)$ and $(\mathcal{L}'_2)$ follow from $(\mathcal{L}_2)$ and $(\mathcal{L}_3)$, respectively.

We continue to prove Theorem 6.9. Suppose $\mathsf{EVAL}(\mathbf{F}, \mathfrak{D})$ is not #P-hard. Then the argument above shows that $(\mathbf{C}_{\Sigma,\Sigma}, \mathfrak{E}_\Sigma)$ (with appropriate $\mathbf{p}', \mathbf{t}', \mathcal{Q}'$) satisfies both $(\mathcal{R})$ and $(\mathcal{L})$. Since by Observation 1, $\mathsf{EVAL}(\mathbf{C}_{\Sigma,\Sigma}, \mathfrak{E}_\Sigma)$ is not #P-hard, by Theorem 5.9 and (15.1), $\mathbf{D}^{[r]}$ satisfies $(\mathcal{D}_2)$ and $(\mathcal{D}_4)$ for all $r \in \mathcal{Z}$. $(\mathcal{D}'_1)$ follows from $(\mathcal{D}_2)$.

To prove $(\mathcal{D}'_2)$, let $\mathbf{F}' = \mathbf{F}_{\Sigma_0,I}$. By $(\mathcal{D}_4)$, for any $r \in \mathcal{Z}$, $k \in [s]$ and $\mathbf{a} \in \Gamma^{\mathrm{lin}}_{r,k}$, there exist $\mathbf{b} \in \tilde{\mathbb{Z}}_{\mathbf{q}_k}$ and $\alpha \in \mathbb{Z}_N$ such that

$$\omega_N^\alpha \cdot F'_{\widetilde{\mathbf{b}},\mathbf{x}} = D^{[r]}_{\mathbf{x}+\widetilde{\mathbf{a}}} \cdot \overline{D^{[r]}_{\mathbf{x}}} \quad \text{for all } \mathbf{x} \in \Gamma_r, \text{ where } \mathbf{F}'_{\widetilde{\mathbf{b}},*} = \mathbf{F}_{\Sigma_0(\widetilde{\mathbf{b}}),*}.$$

Since Σ_0 works within each prime factor, there exists a $\mathbf{b}' \in \tilde{\mathbb{Z}}_{\mathbf{q}_k}$ such that $\Sigma_0(\widetilde{\mathbf{b}}) = \widetilde{\mathbf{b}'}$ and $(\mathcal{D}'_2)$ follows.

16. Tractability: Proof of Theorem 6.10. The proof of Theorem 6.10 is similar to that of Theorem 5.10 for the bipartite case presented in section 12.

Let $((M,N), \mathbf{F}, \mathfrak{D}, (\mathbf{d}, \mathcal{W}, \mathbf{p}, \mathbf{t}, \mathcal{Q}, \mathcal{K}))$ be a tuple that satisfies $(\mathcal{R}')$, $(\mathcal{L}')$, and $(\mathcal{D}')$. The proof has two steps. First we use $(\mathcal{R}'), (\mathcal{L}'), (\mathcal{D}')$ to decompose $\mathsf{EVAL}(\mathbf{F}, \mathfrak{D})$ into s subproblems (recall s is the length of the sequence $\mathbf{p}$), denoted by $\mathsf{EVAL}(\mathbf{F}^{[i]}, \mathfrak{D}^{[i]})$, $i \in [s]$, such that if every $\mathsf{EVAL}(\mathbf{F}^{[i]}, \mathfrak{D}^{[i]})$ is tractable, then so is $\mathsf{EVAL}(\mathbf{F}, \mathfrak{D})$. Second, we reduce each $\mathsf{EVAL}(\mathbf{F}^{[i]}, \mathfrak{D}^{[i]})$ to $\mathsf{EVAL}(\pi)$ for some prime power π.

By Theorem 12.1, $\mathsf{EVAL}(\pi)$ can be solved in polynomial time for any fixed prime power π. Thus, $\mathsf{EVAL}(\mathbf{F}^{[i]}, \mathfrak{D}^{[i]})$ is tractable for all $i \in [s]$, and so is $\mathsf{EVAL}(\mathbf{F}, \mathfrak{D})$.

16.1. Step 1. Fix i to be any index in $[s]$. We start by defining $\mathbf{F}^{[i]}$ and $\mathfrak{D}^{[i]}$. Recall the definition of $\tilde{\mathbb{Z}}_{\mathbf{q}_i}$ from section 6.3.3. For any $\mathbf{x} \in \tilde{\mathbb{Z}}_{\mathbf{q}_i}$, we use $\widetilde{\mathbf{x}} \in \prod_{j=1}^s \tilde{\mathbb{Z}}_{\mathbf{q}_j}$ to denote the vector such that $(\widetilde{\mathbf{x}})_i = \mathbf{x}$ and $(\widetilde{\mathbf{x}})_j = \mathbf{0}$ for all $j \ne i$.

$\mathbf{F}^{[i]}$ is an $m_i \times m_i$ symmetric matrix, where $m_i = |\tilde{\mathbb{Z}}_{\mathbf{q}_i}|$. We use $\tilde{\mathbb{Z}}_{\mathbf{q}_i}$ to index the rows and columns of $\mathbf{F}^{[i]}$. Then

$$F^{[i]}_{\mathbf{x},\mathbf{y}} = F_{\widetilde{\mathbf{x}},\widetilde{\mathbf{y}}} \quad \text{for all } \mathbf{x}, \mathbf{y} \in \tilde{\mathbb{Z}}_{\mathbf{q}_i}.$$

By condition $(\mathcal{R}'_3)$, it is easy to see that

$$\mathbf{F} = \mathbf{F}^{[1]} \otimes \ldots \otimes \mathbf{F}^{[s]}. \tag{16.1}$$

$\mathfrak{D}^{[i]} = (\mathbf{D}^{[i,0]}, \ldots, \mathbf{D}^{[i,N-1]})$ is a sequence of $m_i \times m_i$ diagonal matrices: $\mathbf{D}^{[i,0]}$ is the $m_i \times m_i$ identity matrix; for every $r \in [N-1]$, the $\mathbf{x}$th entry of $\mathbf{D}^{[i,r]}$ is

$$D^{[i,r]}_{\mathbf{x}} = D^{[r]}_{\mathbf{ext}_r(\mathbf{x})} \quad \text{for all } \mathbf{x} \in \tilde{\mathbb{Z}}_{\mathbf{q}_i}.$$

By condition $(\mathcal{D}_1')$, we have

$$\mathbf{D}^{[r]} = \mathbf{D}^{[1,r]} \otimes \ldots \otimes \mathbf{D}^{[s,r]} \quad \text{for all } r \in [0 : N-1]. \tag{16.2}$$

It then follows from (16.1) and (16.2) that

$$Z_{\mathbf{F},\mathfrak{D}}(G) = Z_{\mathbf{F}^{[1]},\mathfrak{D}^{[1]}}(G) \times \ldots \times Z_{\mathbf{F}^{[s]},\mathfrak{D}^{[s]}}(G)$$

for all graphs G. As a result, we have the following lemma.

LEMMA 16.1. *If* $\mathsf{EVAL}(\mathbf{F}^{[i]}, \mathfrak{D}^{[i]})$ *is tractable for all* $i \in [s]$*, then* $\mathsf{EVAL}(\mathbf{F}, \mathfrak{D})$ *is also tractable.*

Recall that $\mathcal{Z}$ is the set of $r \in [N-1]$ such that $\mathbf{D}^{[r]} \neq \mathbf{0}$; $\Gamma_{r,i}$ is a coset in $\tilde{\mathbb{Z}}_{\mathbf{q}_i}$ for each $i \in [s]$ such that $\Gamma_r = \Gamma_{r,1} \times \cdots \times \Gamma_{r,s}$. We use $(\mathcal{D}_2')$ to prove the next lemma.

LEMMA 16.2. *Given* $r \in \mathcal{Z}, i \in [s], \mathbf{a} \in \Gamma_{r,i}^{\mathrm{lin}}$*, there are* $\mathbf{b} \in \tilde{\mathbb{Z}}_{\mathbf{q}_i}, \alpha \in \mathbb{Z}_N$ *such that*

$$D^{[i,r]}_{\mathbf{x}+\mathbf{a}} \cdot \overline{D^{[i,r]}_{\mathbf{x}}} = \omega_N^{\alpha} \cdot F^{[i]}_{\mathbf{b},\mathbf{x}} \quad \textit{for all } \mathbf{x} \in \Gamma_{r,i}.$$

Proof. By the definition of $\mathbf{D}^{[i,r]}$, we have

$$D^{[i,r]}_{\mathbf{x}+\mathbf{a}} \cdot \overline{D^{[i,r]}_{\mathbf{x}}} = D^{[r]}_{\mathbf{ext}_r(\mathbf{x}+\mathbf{a})} \cdot \overline{D^{[r]}_{\mathbf{ext}_r(\mathbf{x})}} = D^{[r]}_{\mathbf{ext}_r(\mathbf{x})+\tilde{\mathbf{a}}} \cdot \overline{D^{[r]}_{\mathbf{ext}_r(\mathbf{x})}}.$$

Then by condition $(\mathcal{D}_2')$, we know there exist $\mathbf{b} \in \tilde{\mathbb{Z}}_{\mathbf{q}_i}$ and $\alpha \in \mathbb{Z}_N$ such that

$$D^{[i,r]}_{\mathbf{x}+\mathbf{a}} \cdot \overline{D^{[i,r]}_{\mathbf{x}}} = \omega_N^{\alpha} \cdot F_{\tilde{\mathbf{b}},\mathbf{ext}_r(\mathbf{x})} = \omega_N^{\alpha} \cdot F^{[i]}_{\mathbf{b},\mathbf{x}} \quad \text{for all } \mathbf{x} \in \Gamma_{r,i},$$

and the lemma is proved. ∎

16.2. Step 2. For convenience, we let $\mathsf{EVAL}(\mathbf{F}, \mathfrak{D})$ denote one of the problems $\mathsf{EVAL}(\mathbf{F}^{[i]}, \mathfrak{D}^{[i]})$ we defined in the last step. By conditions $(\mathcal{R}')$, $(\mathcal{L}')$, $(\mathcal{D}')$ and Lemma 16.2, we summarize the properties of $(\mathbf{F}, \mathfrak{D})$ as follows. We will use these properties to show that $\mathsf{EVAL}(\mathbf{F}, \mathfrak{D})$ is tractable.

$(\mathcal{F}_1')$ There is a prime p and a nonincreasing sequence $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_h)$ of powers of p. $\mathbf{F}$ is an $m \times m$ symmetric matrix, where $m = \pi_1 \ldots \pi_h$. We let π denote π_1 and use $\mathbb{Z}_{\boldsymbol{\pi}} \equiv \mathbb{Z}_{\pi_1} \times \cdots \times \mathbb{Z}_{\pi_h}$ to index the rows and columns of $\mathbf{F}$. We also let $\mathcal{T}$ denote the set of pairs $(i,j) \in [h] \times [h]$ such that $\pi_i = \pi_j$. Then there exist $c_{i,j} \in \mathbb{Z}_{\pi_i} = \mathbb{Z}_{\pi_j}$, for all $(i,j) \in \mathcal{T}$, such that $c_{i,j} = c_{j,i}$ and

$$F_{\mathbf{x},\mathbf{y}} = \prod_{(i,j)\in\mathcal{T}} \omega_{\pi_i}^{c_{i,j}x_iy_j} \quad \text{for all } \mathbf{x} = (x_1, \ldots, x_h), \quad \mathbf{y} = (y_1, \ldots, y_h) \in \mathbb{Z}_{\boldsymbol{\pi}},$$

where $x_i \in \mathbb{Z}_{\pi_i}$ denotes the ith entry of $\mathbf{x}$. We express $\mathbf{F}$ in this very general form to unify the proofs for the two slightly different cases: $(\mathbf{F}^{[1]}, \mathfrak{D}^{[1]})$ and $(\mathbf{F}^{[i]}, \mathfrak{D}^{[i]})$, $i \geq 2$.

$(\mathcal{F}_2')$ $\mathfrak{D} = (\mathbf{D}^{[0]}, \ldots, \mathbf{D}^{[N-1]})$ is a sequence of N $m \times m$ diagonal matrices, where $N \geq 1$ and $\pi \mid N$. $\mathbf{D}^{[0]}$ is the identity matrix; every diagonal entry of $\mathbf{D}^{[r]}$, $r \in [N-1]$ is either 0 or a power of ω_N. We also use $\mathbb{Z}_{\boldsymbol{\pi}}$ to index the diagonal entries of $\mathbf{D}^{[r]}$.

$(\mathcal{F}_3')$ For every $r \in [0 : N-1]$, let Γ_r denote the set of $\mathbf{x} \in \mathbb{Z}_{\boldsymbol{\pi}}$ such that the $\mathbf{x}$th entry of $\mathbf{D}^{[r]}$ is nonzero, and let $\mathcal{Z}$ denote the set of r such that $\Gamma_r \neq \emptyset$. For every $r \in \mathcal{Z}$, Γ_r is a coset in $\mathbb{Z}_{\boldsymbol{\pi}}$. Moreover, for every $r \in \mathcal{Z}$, there is a vector $\mathbf{a}^{[r]} \in \Gamma_r$ such that the $(\mathbf{a}^{[r]})$th entry of $\mathbf{D}^{[r]}$ is 1.

$(\mathcal{F}_4')$ For all $r \in \mathcal{Z}$ and $\mathbf{a} \in \Gamma_r^{\mathrm{lin}}$, there exist $\mathbf{b} \in \mathbb{Z}_{\boldsymbol{\pi}}$ and $\alpha \in \mathbb{Z}_N$ such that

$$D^{[r]}_{\mathbf{x}+\mathbf{a}} \cdot \overline{D^{[r]}_{\mathbf{x}}} = \omega_N^{\alpha} \cdot F_{\mathbf{b},\mathbf{x}} \quad \text{for all } \mathbf{x} \in \Gamma_r.$$

Let G be an undirected graph. Below we reduce the computation of $Z_{\mathbf{F},\mathfrak{D}}(G)$ to $\mathsf{EVAL}(\widehat{\pi})$, where $\widehat{\pi} = \pi$ if $p \neq 2$ and $\widehat{\pi} = 2\pi$ if $p = 2$. Given $a \in \mathbb{Z}_{\pi_i}$ for some $i \in [h]$, we use $\widehat{a}$ to denote an element in $\mathbb{Z}_{\widehat{\pi}}$ such that $\widehat{a} \equiv a \pmod{\pi_i}$. For definiteness we can choose a itself if we consider a to be an integer between 0 and $\pi_i - 1$.

Let $G = (V, E)$. We let V_r, $r \in [0 : N-1]$, denote the set of vertices in V whose degree is $r \bmod N$. We decompose E into $E_{i,j}$, $i \leq j \in [0 : N-1]$, where $E_{i,j}$ contains the set of edges between V_i and V_j. Clearly, if $V_r \neq \emptyset$ for some $r \notin \mathcal{Z}$, then $Z_{\mathbf{F},\mathfrak{D}}(G)$ is trivially 0. Thus, we assume $V_r = \emptyset$ for all $r \notin \mathcal{Z}$. In this case, we have

$$Z_{\mathbf{F},\mathfrak{D}}(G) = \sum_{\xi} \left[\prod_{r \in \mathcal{Z}} \prod_{v \in V_r} D^{[r]}_{\mathbf{x}_v} \right] \left[\prod_{r \leq r' \in \mathcal{Z}} \prod_{uv \in E_{r,r'}} F_{\mathbf{x}_u, \mathbf{x}_v} \right],$$

where the sum ranges over all assignments $\xi = (\xi_r \colon V_r \to \Gamma_r \,|\, r \in \mathcal{Z})$ with $\xi(v) = \mathbf{x}_v$.

By Lemma 12.4, we know that for every $r \in \mathcal{Z}$, there exist a positive integer s_r and an $s_r \times h$ matrix $\mathbf{A}^{[r]}$ over $\mathbb{Z}_{\widehat{\pi}}$ that give us a *uniform* map $\gamma^{[r]}$ (see Lemma 12.4 for the definition) from $\mathbb{Z}_{\widehat{\pi}}^{s_r}$ to Γ_r:

$$\gamma^{[r]}_i(\mathbf{x}) = \left(\mathbf{x}\mathbf{A}^{[r]}_{*,i} + \widehat{\mathfrak{a}}^{[r]}_i \pmod{\pi_i} \right) \quad \text{for all } i \in [h].$$

For every $r \in \mathcal{Z}$, we have $\gamma^{[r]}(\mathbf{0}) = \mathfrak{a}^{[r]} \in \Gamma_r$. Since $\gamma^{[r]}$ is uniform and we know the multiplicity of this map, in order to compute $Z_{\mathbf{F},\mathfrak{D}}(G)$ it suffices to compute

$$\sum_{(\mathbf{x}_v)} \left[\prod_{r \in \mathcal{Z}} \prod_{v \in V_r} D^{[r]}_{\gamma^{[r]}(\mathbf{x}_v)} \right] \left[\prod_{r \leq r' \in \mathcal{Z}} \prod_{uv \in E_{r,r'}} F_{\gamma^{[r]}(\mathbf{x}_u), \gamma^{[r']}(\mathbf{x}_v)} \right],$$

where the sum is over

$$\left(\mathbf{x}_v \in \mathbb{Z}_{\widehat{\pi}}^{s_r} : v \in V_r, r \in \mathcal{Z} \right) = \prod_{r \in \mathcal{Z}} (\mathbb{Z}_{\widehat{\pi}}^{s_r})^{|V_r|}.$$

If for every $r \in \mathcal{Z}$, there is a quadratic polynomial $f^{[r]}$ over $\mathbb{Z}_{\widehat{\pi}}$ such that

$$D^{[r]}_{\gamma^{[r]}(\mathbf{x})} = \omega_{\widehat{\pi}}^{f^{[r]}(\mathbf{x})} \quad \text{for all } \mathbf{x} \in \mathbb{Z}_{\widehat{\pi}}^{s_r}, \tag{16.3}$$

and for all $r, r' : r \leq r' \in \mathcal{Z}$, there is a quadratic polynomial $f^{[r,r']}$ over $\mathbb{Z}_{\widehat{\pi}}$ such that

$$F_{\gamma^{[r]}(\mathbf{x}), \gamma^{[r']}(\mathbf{y})} = \omega_{\widehat{\pi}}^{f^{[r,r']}(\mathbf{x},\mathbf{y})} \quad \text{for all } \mathbf{x} \in \mathbb{Z}_{\widehat{\pi}}^{s_r} \text{ and } \mathbf{y} \in \mathbb{Z}_{\widehat{\pi}}^{s_{r'}}, \tag{16.4}$$

then we can reduce the computation of $Z_{\mathbf{F},\mathfrak{D}}(G)$ to $\mathsf{EVAL}(\widehat{\pi})$ and finish the proof.

First, we deal with (16.4). By $(\mathcal{F}'_1)$, the following function satisfies (16.4):

$$f^{[r,r']}(\mathbf{x},\mathbf{y}) = \sum_{(i,j) \in \mathcal{T}} c_{i,j} \frac{\widehat{\pi}}{\pi_i} \gamma^{[r]}_i(\mathbf{x}) \gamma^{[r']}_j(\mathbf{y}) = \sum_{(i,j) \in \mathcal{T}} \widehat{c}_{i,j} \frac{\widehat{\pi}}{\pi_i} \left(\mathbf{x}\mathbf{A}^{[r]}_{*,i} + \widehat{\mathfrak{a}}^{[r]}_i \right) \left(\mathbf{y}\mathbf{A}^{[r']}_{*,j} + \widehat{\mathfrak{a}}^{[r']}_j \right).$$

Note that $(i,j) \in \mathcal{T}$ implies that $\pi_i = \pi_j$ and thus

$$\gamma^{[r]}_i(\mathbf{x}),\ \gamma^{[r']}_j(\mathbf{y}) \in \mathbb{Z}_{\pi_i} = \mathbb{Z}_{\pi_j}.$$

To be able to substitute the $(\bmod\ \pi_i)$ expressions for $\gamma^{[r]}_i(\mathbf{x})$ and $\gamma^{[r']}_j(\mathbf{y})$, the presence of $\widehat{\pi}/\pi_i$ is crucial. It is also clear that this is a quadratic polynomial over $\mathbb{Z}_{\widehat{\pi}}$.

Next we prove the existence of the quadratic polynomial $f^{[r]}$. Let us fix r to be an index in $\mathcal{Z}$. We use $\mathbf{e}_i$ for each $i \in [s_r]$ to denote the unit vector in $\mathbb{Z}_{\widehat{\pi}}^{s_r}$ whose ith entry is 1 and whose other entries are 0. Using $(\mathcal{F}_4')$, we know that for every $i \in [s_r]$, there exist $\alpha_i \in \mathbb{Z}_N$ and $\mathbf{b}_i = (b_{i,1}, \ldots, b_{i,h}) \in \mathbb{Z}_{\boldsymbol{\pi}}$, where $b_{i,j} \in \mathbb{Z}_{\pi_j}$, such that

$$D^{[r]}_{\gamma^{[r]}(\mathbf{x}+\mathbf{e}_i)} \cdot \overline{D^{[r]}_{\gamma^{[r]}(\mathbf{x})}} = \omega_N^{\alpha_i} \cdot \prod_{j\in[h]} \omega_{\pi_j}^{b_{i,j}\cdot\gamma_j^{[r]}(\mathbf{x})} \quad \text{for all } \mathbf{x} \in \mathbb{Z}_{\widehat{\pi}}^{s_r},$$

because $\gamma^{[r]}(\mathbf{x}+\mathbf{e}_i) - \gamma^{[r]}(\mathbf{x})$ is a vector in $\mathbb{Z}_{\boldsymbol{\pi}}$ that is independent of $\mathbf{x}$.

With the same argument used in the proof of Theorem 5.10 ((12.14) and (12.15)), $\omega_N^{\alpha_i}$ must be a power of $\omega_{\widehat{\pi}}$ for all $i \in [s_r]$. As a result, there exists $\beta_i \in \mathbb{Z}_{\widehat{\pi}}$ such that

$$D^{[r]}_{\gamma^{[r]}(\mathbf{x}+\mathbf{e}_i)} \cdot \overline{D^{[r]}_{\gamma^{[r]}(\mathbf{x})}} = \omega_{\widehat{\pi}}^{\beta_i} \cdot \prod_{j\in[h]} \omega_{\pi_j}^{b_{i,j}\cdot\gamma_j^{[r]}(\mathbf{x})} \quad \text{for all } \mathbf{x} \in \mathbb{Z}_{\widehat{\pi}}^{s_r}. \tag{16.5}$$

By the argument used in the proof of Theorem 5.10, every nonzero entry of $\mathbf{D}^{[r]}$ is a power of $\omega_{\widehat{\pi}}$. As a result, there exists a function $f^{[r]}$ from $\mathbb{Z}_{\widehat{\pi}}^{s_r}$ to $\mathbb{Z}_{\widehat{\pi}}$ that satisfies (16.3). To see that $f^{[r]}$ is indeed a quadratic polynomial, by (16.5), we have

$$f^{[r]}(\mathbf{x}+\mathbf{e}_i) - f^{[r]}(\mathbf{x}) = \beta_i + \sum_{j\in[h]} \left(\widehat{b}_{i,j} \frac{\widehat{\pi}}{\pi_j} \left(\mathbf{x}\mathbf{A}^{[r]}_{*,j} + \widehat{\mathfrak{a}}_j^{[r]} \right) \right) \quad \text{for all } i \in [s_r], \mathbf{x} \in \mathbb{Z}_{\widehat{\pi}}^{s_r},$$

which is an affine linear form of $\mathbf{x}$ with all coefficients from $\mathbb{Z}_{\widehat{\pi}}$.

By using Lemmas 12.5 and 12.6, we know that $f^{[r]}$ is a quadratic polynomial over $\mathbb{Z}_{\widehat{\pi}}$, and this finishes the reduction from $\mathsf{EVAL}(\mathbf{F}, \mathfrak{D})$ to $\mathsf{EVAL}(\widehat{\pi})$.

17. Decidability in polynomial time: Proof of Theorem 1.2. Finally, we prove Theorem 1.2, i.e., the following decision problem is computable in polynomial time: Given a symmetric $\mathbf{A} \in \mathbb{C}^{m\times m}$ in which every entry $A_{i,j}$ is algebraic, decide if $\mathsf{EVAL}(\mathbf{A})$ is tractable or is #P-hard.

We follow the model of computation discussed in section 2.2. Let

$$\mathscr{A} = \{A_{i,j} : i, j \in [m]\} = \{a_j : j \in [n]\}$$

for some $n \geq 1$ and let α be a primitive element of $\mathbb{Q}(\mathscr{A})$. Thus, $\mathbb{Q}(\mathscr{A}) = \mathbb{Q}(\alpha)$.

The input of the problem consists of the following three parts:

1. a minimal polynomial $F(x) \in \mathbb{Q}[x]$ of α;
2. a rational approximation $\widehat{\alpha}$ that uniquely determines α as a root of $F(x)$;
3. the standard representation of $A_{i,j}$ with respect to α and $F(x)$, $i, j \in [m]$.

The input size of the decision problem is then the length of the binary string needed to describe all these three parts.

Given $\mathbf{A}$, we follow the proof of Theorem 1.1 as follows. First by Lemma 4.6, we can assume without loss of generality that $\mathbf{A}$ is connected. Then we follow the proof sketch described in sections 5 and 6, depending on whether the matrix $\mathbf{A}$ is bipartite or nonbipartite. We assume that $\mathbf{A}$ is connected and bipartite below. The proof for the nonbipartite case is similar.

17.1. Step 1. We show that either $\mathsf{EVAL}(\mathbf{A})$ is #P-hard or we can construct a purified matrix $\mathbf{A}'$ such that $\mathsf{EVAL}(\mathbf{A}) \equiv \mathsf{EVAL}(\mathbf{A}')$ and then pass $\mathbf{A}'$ down to Step 2. We follow the proof of Theorem 5.2. First, we prove that given $\mathscr{A}$, a generating

set $\mathscr{G} \subset \mathbb{Q}(\mathscr{A})$ of $\mathscr{A}$ can be computed in polynomial time. Recall the definition of a generating set from Definition 7.2. We denote the input size as $\widehat{m}$. Thus, $\widehat{m} \geq m$.

THEOREM 17.1. *Given a finite set of nonzero algebraic numbers* $\mathscr{A}$ (under the model of computation described in section 2.2)*, one can in polynomial time (in* $\widehat{m}$*) find* (1) *a generating set* $\mathscr{G} = \{g_1, \ldots, g_d\}$ *of* $\mathscr{A}$ *and* (2) *for every number* $a \in \mathscr{A}$ *the unique tuple* $(k_1, \ldots, k_d) \in \mathbb{Z}^d$ *such that* $a/(g_1^{k_1} \cdots g_d^{k_d})$ *is a root of unity.*

We start the proof with the following lemma.

LEMMA 17.2. *Let*

$$L = \Big\{ \big(x_1, \ldots, x_n\big) \in \mathbb{Z}^n \;\Big|\; a_1^{x_1} \cdots a_n^{x_n} = 1 \Big\}.$$

Let S *be the* $\mathbb{Q}$*-span of* L*, and let* $L' = \mathbb{Z}^n \cap S$*. Then*

$$L' = \Big\{ \big(x_1, \ldots, x_n\big) \in \mathbb{Z}^n \;\Big|\; a_1^{x_1} \cdots a_n^{x_n} \text{ is a root of unity} \Big\}. \tag{17.1}$$

Proof. Clearly L is a lattice, being a discrete subgroup of $\mathbb{Z}^n$. Also L' is a lattice, and $L \subseteq L'$. Suppose $(x_1, \ldots, x_n) \in \mathbb{Z}^n$ is in the lattice in (17.1). Then there exists a nonzero integer ℓ such that $(a_1^{x_1} \cdots a_n^{x_n})^\ell = 1$. As a result, $\ell(x_1, \ldots, x_n) \in L$ and thus $(x_1, \ldots, x_n) \in S$, the $\mathbb{Q}$-span of L.

Conversely, if $\dim(L) = 0$, then $L = \{(0, \ldots, 0)\} = S = L'$. Suppose $\dim(L) > 0$, and we let $\mathbf{b}_1, \ldots, \mathbf{b}_t$ be a basis for L, where $t \in [n]$. Let $(x_1, \ldots, x_n) \in \mathbb{Z}^n \cap S$; then there exist rational numbers $r_1, \ldots, r_t$ such that $(x_1, \ldots, x_n) = \sum_{i=1}^t r_i \mathbf{b}_i$. We have

$$a_1^{x_1} \cdots a_n^{x_n} = \prod_{j=1}^n a_j^{\sum_{i=1}^t r_i b_{i,j}}.$$

Let N be a positive integer such that Nr_i is an integer for $i \in [t]$. Then

$$\big(a_1^{x_1} \cdots a_n^{x_n}\big)^N = \prod_{i=1}^t \left(\prod_{j=1}^n a_j^{b_{i,j}} \right)^{Nr_i} = 1.$$

Thus $a_1^{x_1} \cdots a_n^{x_n}$ is a root of unity and $(x_1, \ldots, x_n)$ is in the lattice in (17.1). ∎

To prove Theorem 17.1, we will also need the following theorem by Ge [19, 20].

THEOREM 17.3 (see [19, 20]). *Given a finite set of nonzero algebraic numbers* $\mathscr{A} = \{a_1, \ldots, a_n\}$ (under the model of computation described in section 2.2)*, one can in polynomial time find a lattice basis for the lattice* L *given by*

$$L = \Big\{ \mathbf{x} = \big(x_1, \ldots, x_n\big) \in \mathbb{Z}^n \;\Big|\; a_1^{x_1} \cdots a_n^{x_n} = 1 \Big\}.$$

Proof of Theorem 17.1. Conceptually this is what we will do: We first use Ge's algorithm to compute a basis for L. Then we show how to compute a basis for L' efficiently. Finally, we compute a basis for $\mathbb{Z}^n/L'$. This basis for $\mathbb{Z}^n/L'$ will define our generating set for $\mathscr{A}$.

More precisely, given the set $\mathscr{A} = \{a_1, \ldots, a_n\}$, we let $\kappa = \{\mathbf{k}_1, \ldots, \mathbf{k}_t\}$ denote the lattice basis for L found by Ge's algorithm [19, 20], where $0 \leq t \leq n$. This basis has polynomially many bits in each integer entry $k_{i,j}$. Here are two easy cases:

1. If $t = 0$, then we can take $g_i = a_i$ as the generators, $1 \leq i \leq n$. There is no nontrivial relation $a_1^{k_1} \cdots a_n^{k_n} =$ a root of unity for any $(k_1, \ldots, k_n) \in \mathbb{Z}^n$ other than $\mathbf{0}$; otherwise a suitable nonzero integer power gives a nontrivial lattice point in L.

2. If $t = n$, then $S = \mathbb{Q}^n$ and $L' = \mathbb{Z}^n$; hence every a_i is a root of unity. In this case, the empty set is a generating set for $\mathscr{A}$.

Assume $0 < t < n$. We will compute from the basis κ a basis β for $L' = \mathbb{Z}^n \cap S$, where S is the $\mathbb{Q}$-span of L; then we compute a basis γ for the quotient lattice $\mathbb{Z}^n/L'$. Both lattice bases γ and β will have polynomially many bits in each integer entry.

Before showing how to compute β and γ, it is clear that $\dim L' = \dim L = t$ and $\dim(\mathbb{Z}^n/L') = n - t$. Let

$$\gamma = \{\mathbf{x}_1, \ldots, \mathbf{x}_{n-t}\} \quad \text{and} \quad \beta = \{\mathbf{y}_1, \ldots, \mathbf{y}_t\}.$$

We define the following set $\{g_1, \ldots, g_{n-t}\}$ from γ as follows:

$$g_j = a_1^{x_{j,1}} a_2^{x_{j,2}} \cdots a_n^{x_{j,n}}, \quad \text{where } \mathbf{x}_j = (x_{j,1}, x_{j,2}, \ldots, x_{j,n}).$$

We check that $\{g_1, \ldots, g_{n-t}\}$ is a generating set of $\mathscr{A}$. Clearly, being exponentials, all $g_j \neq 0$. Suppose for some $(c_1, \ldots, c_{n-t}) \in \mathbb{Z}^{n-t}$, $g_1^{c_1} \cdots g_{n-t}^{c_{n-t}}$ is a root of unity. Since

$$g_1^{c_1} g_2^{c_2} \cdots g_{n-t}^{c_{n-t}} = a_1^{\sum_{j=1}^{n-t} c_j x_{j,1}} a_2^{\sum_{j=1}^{n-t} c_j x_{j,2}} \cdots a_n^{\sum_{j=1}^{n-t} c_j x_{j,n}},$$

we have

$$\left(\sum_{j=1}^{n-t} c_j x_{j,1}, \sum_{j=1}^{n-t} c_j x_{j,2}, \ldots, \sum_{j=1}^{n-t} c_j x_{j,n} \right) = \sum_{j=1}^{n-t} c_j \mathbf{x}_j \in L'.$$

It follows that $c_j = 0$ for all $j \in [n-t]$.

On the other hand, by the definition of $\mathbb{Z}^n/L'$, for every $(k_1, \ldots, k_n) \in \mathbb{Z}^n$, there exists a unique sequence of integers $c_1, \ldots, c_{n-t} \in \mathbb{Z}$ such that

$$(k_1, \ldots, k_n) - \sum_{j=1}^{n-t} c_j \mathbf{x}_j \in L'.$$

In particular, for $\mathbf{e}_i = (0, \ldots, 1, \ldots, 0)$, where there is a single 1 in the ith position, there exist integers $c_{i,j}$, $i \in [n]$ and $j \in [n-t]$, such that

$$\mathbf{e}_i - \sum_{j=1}^{n-t} c_{i,j} \mathbf{x}_j \in L'.$$

As a result, we have

$$\frac{a_i}{a_1^{\sum_{j=1}^{n-t} c_{i,j} x_{j,1}} a_2^{\sum_{j=1}^{n-t} c_{i,j} x_{j,2}} \cdots a_n^{\sum_{j=1}^{n-t} c_{i,j} x_{j,n}}} = \frac{a_i}{g_1^{c_{i,1}} \cdots g_{n-t}^{c_{i,n-t}}}$$

is a root of unity. This completes the construction of a generating set $\mathscr{G}$ for $\mathscr{A}$. In the following, we compute the bases γ and β in polynomial time, given κ.

First, we may change the first vector $\mathbf{k}_1 = (k_{1,1}, \ldots, k_{1,n})$ in κ to be a *primitive* vector, meaning that $\gcd(k_{1,1}, \ldots, k_{1,n}) = 1$, by factoring out the gcd. If the gcd is greater than 1, then this changes the lattice L, but it does not change the $\mathbb{Q}$-span S and thus there is no change to L'.

In addition, there exists a unimodular matrix $\mathbf{M}_1$ such that

$$(k_{1,1}, \ldots, k_{1,n})\, \mathbf{M}_1 = (1, 0, \ldots, 0) \in \mathbb{Z}^n.$$

This is just the extended Euclidean algorithm. (An integer matrix $\mathbf{M}_1$ is *unimodular* iff its determinant is ± 1 or, equivalently, it has an integral inverse matrix.)

Now consider the $t \times n$ matrix

$$\begin{pmatrix} u_{1,1} & \dots & u_{1,n} \\ \vdots & \ddots & \vdots \\ u_{t,1} & \dots & u_{t,n} \end{pmatrix} = \begin{pmatrix} k_{1,1} & \dots & k_{1,n} \\ \vdots & \ddots & \vdots \\ k_{t,1} & \dots & k_{t,n} \end{pmatrix} \mathbf{M}_1.$$

This is also an integral matrix as $\mathbf{M}_1$ is integral. Moreover its first row is $(1, 0, \dots, 0)$. We may perform row transformations to make $u_{2,1} = 0, \dots, u_{t,1} = 0$. Performing the same transformations on the right-hand side replaces the basis κ by another basis for the same lattice, and L' is unchanged. We still use $\kappa = \{\mathbf{k}_1, \dots, \mathbf{k}_t\}$ to denote this new basis.

Next, consider the entries $u_{2,2}, \dots, u_{2,n}$. If $\gcd(u_{2,2}, \dots, u_{2,n}) > 1$ we may divide out this gcd. Since the second row satisfies

$$\big(k_{2,1}, k_{2,2}, \dots, k_{2,n}\big) = \big(0, u_{2,2}, \dots, u_{2,n}\big)\, \mathbf{M}_1^{-1},$$

this gcd must also divide $k_{2,1}, k_{2,2}, \dots, k_{2,n}$. (In fact, this is also the gcd of $(k_{2,1}, k_{2,2}, \dots, k_{2,n})$.) This division updates the basis κ by another basis, which changes the lattice L, but still it does not change the $\mathbb{Q}$-span S and thus the lattice L' remains unchanged. We continue to use the same κ to denote this updated basis.

For the same reason, there exists an $(n-1) \times (n-1)$ unimodular $\mathbf{M}'$ such that

$$\big(u_{2,2}, \dots, u_{2,n}\big)\, \mathbf{M}' = \big(1, 0, \dots, 0\big) \in \mathbb{Z}^{n-1}.$$

Append a 1 at the $(1,1)$ position. This defines a second $n \times n$ unimodular matrix $\mathbf{M}_2$ such that we may update the matrix equation as follows:

$$\begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & u_{3,2} & u_{3,3} & \dots & u_{3,n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & u_{t,2} & u_{t,3} & \dots & u_{t,n} \end{pmatrix} = \begin{pmatrix} k_{1,1} & \dots & k_{1,n} \\ \vdots & \ddots & \vdots \\ k_{t,1} & \dots & k_{t,n} \end{pmatrix} \mathbf{M}_1 \mathbf{M}_2.$$

Now we may kill off the entries $u_{3,2}, \dots, u_{t,2}$, accomplished by row transformations which do not change L or L'. It follows that we can finally find a unimodular matrix $\mathbf{M}^*$ such that the updated κ satisfies

$$\begin{pmatrix} k_{1,1} & \dots & k_{1,n} \\ \vdots & \ddots & \vdots \\ k_{t,1} & \dots & k_{t,n} \end{pmatrix} \mathbf{M}^* = \begin{pmatrix} 1 & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 & 0 & \dots & 0 \end{pmatrix}. \tag{17.2}$$

The right-hand side is the $t \times t$ identity matrix $\mathbf{I}_t$ with an all-zero $t \times (n-t)$ matrix appended. The updated κ here is a lattice basis for a lattice $\widehat{L}$ which has the same $\mathbb{Q}$-span S as L. It is also a full-dimensional sublattice of (the unchanged) L'.

We claim this updated $\kappa = \{\mathbf{k}_1, \dots, \mathbf{k}_t\}$ is actually a lattice basis for L' and thus $\widehat{L} = L'$. Assume for some rational numbers $r_1, \dots, r_t$ the vector $\sum_{i=1}^{t} r_i \mathbf{k}_i \in \mathbb{Z}^n$. Then multiplying $(r_1, \dots, r_t)$ to the left in (17.2) implies that $r_1, \dots, r_t$ are integers.

This completes the computation of a basis for L'. As the only operations we perform are Gaussian eliminations and gcd computations, this is in polynomial time, and the number of bits in every entry is always polynomially bounded.

Finally we describe the computation of a basis for the quotient lattice $\mathbb{Z}^n/L'$.

We start with a basis κ for L' as computed above and extend it to a basis for $\mathbb{Z}^n$. The extended part will then be a basis for $\mathbb{Z}^n/L'$. Suppose that we are given the basis κ for L' together with a unimodular matrix $\mathbf{M}^*$ satisfying (17.2). Then consider the $n \times n$ matrix $(\mathbf{M}^*)^{-1}$. Since $(\mathbf{M}^*)^{-1} = \mathbf{I}_n(\mathbf{M}^*)^{-1}$, the first t rows of $(\mathbf{M}^*)^{-1}$ are precisely the κ matrix. We define the basis for $\mathbb{Z}^n/L'$ to be the last $n-t$ row vectors of $(\mathbf{M}^*)^{-1}$. It can be easily verified that this is a lattice basis for $\mathbb{Z}^n/L'$. □

With Theorem 17.1, we can now follow the proof of Theorem 5.2. By using the generating set, we construct the matrix $\mathbf{B}$ as in section 7.2. Every entry of $\mathbf{B}$ is the product of a nonnegative integer and a root of unity with $\mathsf{EVAL}(\mathbf{A}) \equiv \mathsf{EVAL}(\mathbf{B})$.

We then check whether $\mathbf{B}'$, where $B'_{i,j} = |B_{i,j}|$ for all i, j, satisfies the conditions imposed by the dichotomy theorem of Bulatov and Grohe. (Note that every entry of $\mathbf{B}'$ is a nonnegative integer.) If $\mathbf{B}'$ does not satisfy, then $\mathsf{EVAL}(\mathbf{B}')$ is #P-hard, and so is $\mathsf{EVAL}(\mathbf{A})$ by Lemma 7.5. Otherwise, $\mathbf{B}$ must be a purified matrix, and we pass it down to the next step.

17.2. Step 2. We follow the proof of Theorem 5.3. After rearranging the rows and columns of the purified matrix $\mathbf{B}$, we check the orthogonality condition imposed by Lemma 8.5. If $\mathbf{B}$ satisfies the orthogonality condition, we can use the cyclotomic reduction to construct efficiently a pair $(\mathbf{C}, \mathfrak{D})$ from $\mathbf{B}$, which satisfies the conditions (Shape_1), (Shape_2), (Shape_3) and satisfies $\mathsf{EVAL}(\mathbf{B}) \equiv \mathsf{EVAL}(\mathbf{C}, \mathfrak{D})$.

Next, we check whether the pair $(\mathbf{C}, \mathfrak{D})$ satisfies (Shape_4) and (Shape_5). If either of these two conditions is not satisfied, $\mathsf{EVAL}(\mathbf{C}, \mathfrak{D})$ is #P-hard, and so is $\mathsf{EVAL}(\mathbf{B})$. Finally we check the rank-1 condition, which implies (Shape_6), as imposed by Lemma 8.12 on $(\mathbf{C}, \mathfrak{D})$. With (Shape_1)–(Shape_6), we follow section 8.6 to construct a tuple $((M, 2N), \mathbf{X}, \mathfrak{Y}')$ that satisfies $(\mathcal{U}_1)$–$(\mathcal{U}_4)$, and $\mathsf{EVAL}(\mathbf{C}, \mathfrak{D}) \equiv \mathsf{EVAL}(\mathbf{X}, \mathfrak{Y}')$. We then pass the tuple $((M, 2N), \mathbf{X}, \mathfrak{Y}')$ down to Step 3.

17.3. Step 3. We follow Theorems 5.4, 5.6, 5.8, and 5.9. First, $(\mathcal{U}_5)$ in Theorem 5.4 can be verified efficiently. In Theorem 5.6, we need to check if the matrix $\mathbf{F}$ has a Fourier decomposition, after an appropriate permutation of its rows and columns. This decomposition, if $\mathbf{F}$ has one, can be computed efficiently by first checking the group condition in Lemma 9.1 and then following the proofs of both Lemma 9.5 and Lemma 9.7. Finally, it is easy to see that all the conditions imposed by Theorems 5.8 and 5.9 can be checked in polynomial time.

If $\mathbf{A}$ and other matrices, pairs, or tuples derived from $\mathbf{A}$ satisfy all the conditions in these three steps, then by the tractability part of the dichotomy theorem, $\mathsf{EVAL}(\mathbf{A})$ is solvable in polynomial time. From this, we obtain the polynomial-time decidability of the complexity dichotomy, and Theorem 1.2 is proved.

18. Acknowledgments. We would like to thank Al Aho, Miki Ajtai, Sanjeev Arora, Dick Askey, Paul Beame, Richard Brualdi, Andrei Bulatov, Xiaotie Deng, Alan Frieze, Martin Grohe, Pavol Hell, Lane Hemaspaandra, Kazuo Iwama, Gabor Kun, Dick Lipton, Tal Malkin, Christos Papadimitriou, Mike Paterson, Rocco Servedio, Endre Szemerédi, Shang-Hua Teng, Joe Traub, Osamu Watanabe, Avi Wigderson, and Mihalis Yannakakis for their interest and many comments. We thank especially Martin Dyer, Leslie Goldberg, Mark Jerrum, Marc Thurley, Leslie Valiant, and Mingji Xia for in-depth discussions. We are truly grateful to the reviewers for their dedication

in carefully reading through this long paper; they offered many valuable critiques and suggestions for improvements. We have greatly benefited from their comments.

REFERENCES

[1] A. Bulatov, *The complexity of the counting constraint satisfaction problem*, in Proceedings of the 35th International Colloquium on Automata, Languages and Programming, 2008, pp. 646–661.

[2] A. Bulatov, *The Complexity of the Counting Constraint Satisfaction Problem*, Electronic Colloquium on Computational Complexity, 2009.

[3] A. Bulatov, M. E. Dyer, L. A. Goldberg, M. Jalsenius, M. R Jerrum, and D. M. Richerby, *The complexity of weighted and unweighted #CSP*, J. Comput. System Sci., 78 (2012), pp. 681–688.

[4] A. Bulatov and M. Grohe, *The complexity of partition functions*, Theoret. Comput. Sci., 348 (2005), pp. 148–186.

[5] J.-Y. Cai and X. Chen, *A decidable dichotomy theorem on directed graph homomorphisms with nonnegative weights*, in Proceedings of the 51st Annual IEEE Symposium on Foundations of Computer Science, 2010, pp. 437–446.

[6] J.-Y. Cai and X. Chen, *Complexity of counting CSP with complex weights*, in Proceedings of the 44th Symposium on Theory of Computing, 2012, pp. 909–920.

[7] J.-Y. Cai, X. Chen, and P. Lu, *Non-negatively weighted #CSPs: An effective complexity dichotomy*, in Proceedings of the 26th Annual IEEE Conference on Computational Complexity, 2011.

[8] J.-Y. Cai and P. Lu, *Holographic algorithms: From art to science*, in Proceedings of the 39th Annual ACM Symposium on Theory of Computing, 2007, pp. 401–410.

[9] J.-Y. Cai, P. Lu, and M. Xia, *Holant problems and counting CSP*, in Proceedings of the 41st ACM Symposium on Theory of Computing, 2009, pp. 715–724.

[10] L. Carlitz, *Kloosterman sums and finite field extensions*, Acta Arithmetica, 16 (1969), pp. 179–193.

[11] N. Creignou, S. Khanna, and M. Sudan, *Complexity Classifications of Boolean Constraint Satisfaction Problems*, SIAM Monogr. Discrete Math. Appl., SIAM, Philadelphia, 2001.

[12] M. E. Dyer, L. A. Goldberg, and M. Paterson, *On counting homomorphisms to directed acyclic graphs*, J. ACM, 54 (2007).

[13] M. E. Dyer and C. Greenhill, *The complexity of counting graph homomorphisms*, Random Structures Algorithms, 17 (2000), pp. 260–289.

[14] M. E. Dyer and D. M. Richerby, *On the complexity of #CSP*, in Proceedings of the 42nd ACM Symposium on Theory of Computing, 2010, pp. 725–734.

[15] A. Ehrenfeucht and M. Karpinski, *The Computational Complexity of (XOR, AND)-Counting Problems*, Tech. report TR-8543, Universität Bonn, 1990.

[16] T. Feder and M. Vardi, *The computational structure of monotone monadic SNP and constraint satisfaction: A study through Datalog and group theory*, SIAM J. Comput., 28 (1999), pp. 57–104.

[17] R. Feynman, R. Leighton, and M. Sands, *The Feynman Lectures on Physics*, Addison-Wesley, Reading, MA, 1970.

[18] M. Freedman, L. Lovász, and A. Schrijver, *Reflection positivity, rank connectivity, and homomorphism of graphs*, J. AMS, 20 (2007), pp. 37–51.

[19] G. Ge, *Algorithms Related to Multiplicative Representations of Algebraic Numbers*, Ph.D. thesis, University of California–Berkeley, 1993.

[20] G. Ge, *Testing equalities of multiplicative representations in polynomial time*, in Proceedings of the 34th Annual IEEE Symposium on Foundations of Computer Science, 1993, pp. 422–426.

[21] L. A. Goldberg, M. Grohe, M. Jerrum, and M. Thurley, *A complexity dichotomy for partition functions with mixed signs*, SIAM J. Comput., 39 (2010), pp. 3336–3402.

[22] P. Hell and J. Nešetřil, *On the complexity of H-coloring*, J. Combin. Theory Ser. B, 48 (1990), pp. 92–110.

[23] P. Hell and J. Nešetřil, *Graphs and Homomorphisms*, Oxford University Press, New York, 2004.

[24] N. Jacobson, *Basic Algebra* I, W.H. Freeman, New York, 1985.

[25] S. Lang, *Algebra*, 3rd ed., Springer-Verlag, New York, 2002.

[26] H. W. Lenstra, *Algorithms in algebraic number theory*, Bull. Amer. Math. Soc., 26 (1992), pp. 211–244.

[27] R. Lidl and H. Niederreiter, *Finite Fields*, Encyclopedia Math. Appl., Cambridge University Press, Cambridge, UK, 1997.
[28] L. Lovász, *Operations with structures*, Acta Math. Hungar., 18 (1967), pp. 321–328.
[29] L. Lovász, *The rank of connection matrices and the dimension of graph algebras*, European J. Combin., 27 (2006), pp. 962–970.
[30] P. Morandi, *Field and Galois Theory*, Grad. Texts in Math. 167, Springer, New York, 1996.
[31] T.J. Schaefer, *The complexity of satisfiability problems*, in Proceedings of the 10th Annual ACM Symposium on Theory of Computing, 1978, pp. 216–226.
[32] A. Schrijver, *Graph invariants in the spin model*, J. Combin. Theory Ser. B, 99 (2009), pp. 502–511.
[33] M. Thurley, *The Complexity of Partition Functions*, Ph.D. thesis, Humboldt Universität zu Berlin, 2009.
[34] M. Thurley, *The Complexity of Partition Functions on Hermitian Matrices*, arXiv:1004.0992, 2010.
[35] L. G. Valiant, *Holographic algorithms (extended abstract)*, in Proceedings of the 45th Annual IEEE Symposium on Foundations of Computer Science, 2004, pp. 306–315.
[36] L. G. Valiant, *Accidental algorthims*, in Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science, 2006, pp. 509–517.
[37] L. G. Valiant, *Holographic algorithms*, SIAM J. Comput., 37 (2008), pp. 1565–1594.

Paper 7

comput. complex. **19** (2010), 37–56
1016-3328/10/010037-20, *published online* 24 February 2010
DOI 10.1007/s00037-009-0284-2

computational complexity

QUADRATIC LOWER BOUND FOR PERMANENT VS. DETERMINANT IN ANY CHARACTERISTIC

Jin-Yi Cai, Xi Chen, and Dong Li

Abstract. In Valiant's theory of arithmetic complexity, the classes VP and VNP are analogs of P and NP. A fundamental problem concerning these classes is the Permanent and Determinant Problem: Given a field $\mathbb{F}$ of characteristic $\neq 2$, and an integer n, what is the minimum m such that the permanent of an $n \times n$ matrix $\mathbf{X} = (x_{ij})$ can be expressed as a determinant of an $m \times m$ matrix, where the entries of the determinant matrix are affine linear functions of x_{ij}'s, and the equality is in $\mathbb{F}[\mathbf{X}]$. Mignon and Ressayre (2004) proved a quadratic lower bound $m = \Omega(n^2)$ for fields of characteristic 0. We extend the Mignon–Ressayre quadratic lower bound to all fields of characteristic $\neq 2$.

Keywords. Arithmetic complexity; determinant; permanent; finite field.

Subject classification. 68Q17.

1. Introduction

Given a set of n^2 indeterminates $\mathbf{X} = (x_{i,j})_{i,j=1,\ldots,n}$ over a field $\mathbb{F}$, we can define

$$\det(\mathbf{X}) = \sum_{\pi \in S_n} \operatorname{sign}(\pi) \prod_{i=1}^{n} x_{i,\pi(i)} \quad \text{and} \quad \operatorname{per}(\mathbf{X}) = \sum_{\pi \in S_n} \prod_{i=1}^{n} x_{i,\pi(i)}\,.$$

The determinant function (det) is certainly one of the most well-studied functions in mathematics. The permanent function (per) is also well-studied, especially in combinatorics (Minc 1978). For example, if $\mathbf{A}$ is a 0-1 matrix then $\operatorname{per}(\mathbf{A})$ counts the number of perfect matchings in a bipartite graph with adjacency matrix $\mathbf{A}$.

These well-known functions took on important new meanings when viewed from the computational complexity perspective. It is well known that the deter-

minant can be computed in polynomial time. In fact it can be computed in the complexity class NC^2. By contrast, Valiant (1979a,b) showed that computing the permanent is #P-complete.

In fact Valiant has developed a substantial theory (see also Bürgisser (2000) and Bürgisser, Clausen & Shokrollahi (1997)). The two complexity classes $\mathrm{VP}_{\mathbb{F}}$ and $\mathrm{VNP}_{\mathbb{F}}$ are the analogs of P and NP in this theory of arithmetic complexity, and the two functions, det and per, are the central objects in the two classes, respectively. It was shown that the complexity of computing the permanent characterizes the class $\mathrm{VNP}_{\mathbb{F}}$ and the complexity of computing the determinant (almost) characterizes the class $\mathrm{VP}_{\mathbb{F}}$.

More precisely, a family of polynomials $\{f_n\}$ is in $\mathrm{VP}_{\mathbb{F}}$ if $\deg(f_n) = n^{O(1)}$ and there is a family of arithmetic circuits of size $n^{O(1)}$ computing $\{f_n\}$. A family of polynomials $\{g_n\}$ is in $\mathrm{VNP}_{\mathbb{F}}$ if $\deg(g_n) = n^{O(1)}$, and there exists a family of polynomials $\{f_n\} \in \mathrm{VP}_{\mathbb{F}}$ such that

$$g_n(x_1, \ldots, x_n) = \sum_{y_1,\ldots,y_m \in \{0,1\}} f_{n+m}(x_1, \ldots, x_n, y_1, \ldots, y_m)\,,$$

where $m = n^{O(1)}$. We say that $\{f_n\}$ is a projection of $\{g_m\}$ if there are some $\alpha_1, \ldots, \alpha_m \in \mathbb{F} \cup \{x_1, \ldots, x_n\}$, such that $f_n(x_1, \ldots, x_n) = g_m(\alpha_1, \ldots, \alpha_m)$. It is a p-projection if $m = n^{O(1)}$. A projection is a particularly simple reduction. It is a special case of an affine linear reduction, where each α_i is an affine linear function of x_i's. Valiant proved that

THEOREM 1.1 (Valiant). *For any field* $\mathbb{F}$, $\mathrm{per} \in \mathrm{VNP}_{\mathbb{F}}$. *Moreover, for any* $\mathbb{F}$ *with* $\mathrm{char}\ \mathbb{F} \neq 2$, *any* $\{f_n\} \in \mathrm{VNP}_{\mathbb{F}}$ *is a p-projection of* per.

It is also known that det is in $\mathrm{VP}_{\mathbb{F}}$ (e.g., see Borodin, von zur Gathen & Hopcroft (1982)). More exact characterizations of det were given in terms of polynomial-sized arithmetic branching programs (Damm 1991; Toda 1991; Vinay 1991).

THEOREM 1.2 (Valiant). *Any polynomial* f_n *is a projection of* $\det_m$ *of an* $m \times m$ *matrix, where* m *is linear in the* formula size *of* f_n. *In particular, if* $\{f_n\}$ *has polynomial formula size, then* $\{f_n\}$ *is a p-projection of* det. *Also if* $\{f_n\} \in \mathrm{VP}_{\mathbb{F}}$, *then* f_n *is the projection of* $\det_m$ *for some* $m = n^{O(\log n)}$.

By Ryser's formula (Minc 1978), per_n has formula size $O(n^2 2^n)$. Thus by Valiant's theorem it is the projection of $\det_m$, where $m = O(n^2 2^n)$. Furthermore, if we view Ryser's formula as on the truncated linear row sums directly (instead of on the variables), then Valiant's theorem implies that

THEOREM 1.3. *For any n, there exists a collection $\mathbf{A}$ of affine linear functions $A_{k,l}(\mathbf{X})$ over n^2 variables, where $1 \le k, l \le m = O(2^n)$, such that*

$$\mathrm{per}_n(\mathbf{X}) = \det_m\big(\mathbf{A}(\mathbf{X})\big)\,.$$

It is remarkable that this is the best general upper bound known for this.

DEFINITION 1.4. *The* determinantal complexity dc *of f is the minimum integer m such that there exist affine linear functions $A_{k,l}(\mathbf{X})$, $1 \le k, l \le m$, which satisfy $f(\mathbf{X}) = \det_m(\mathbf{A}(\mathbf{X}))$.*

The question addressed in this paper is about $\mathsf{dc}(\mathrm{per})$. Valiant's analog of P $\neq$ NP will follow if one can show a lower bound

$$\mathsf{dc}(\mathrm{per}_n) = n^{\omega(\log n)}\,.$$

Actually in some sense, this question has a longer history. Pólya (1913) was the first to ask a question on when one can express a permanent as a modified determinant. He noticed that

$$\mathrm{per}\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \det\begin{pmatrix} a & -b \\ c & d \end{pmatrix},$$

and asked if there are any similar equations, by affixing ± 1 to the n^2 variables for $n \ge 3$. This was answered in the negative by Szegö (1913). This line of inquiry culminated in

THEOREM 1.5 (Marcus & Minc 1961). *If* $\mathrm{char}\,\mathbb{F} = 0$ *and $n \ge 3$, then there are no homogeneous linear functions $f_{k,\ell}$ in the indeterminates $x_{i,j}$, $1 \le i, j, k, \ell \le n$, such that*

$$\mathrm{per}_n(x_{i,j}) = \det_n(f_{k,\ell})\,.$$

In terms of $\mathsf{dc}(\mathrm{per}_n)$, this celebrated theorem is equivalent to

$$\mathsf{dc}(\mathrm{per}_n) \ge n + 1\,,$$

over any field of $\mathrm{char}\,\mathbb{F} = 0$ (Note that if the permanental matrix is also $n \times n$ then clearly constant terms in affine linear equations do not help, as seen by the homogeneous part.).

The first non-trivial lower bound for $\mathsf{dc}(\mathrm{per}_n)$ is by von zur Gathen (1985), who showed that $\mathsf{dc}(\mathrm{per}_n) \ge \sqrt{8/7}n$. This was proved for p-projections. Von zur Gathen's result was then improved independently by Babai and Seress as

reported in (von zur Gathen 1987), by Cai (1990), and by Meshulam (1989). Their results were (ignoring lower order terms)

$$\mathsf{dc}(\mathrm{per}_n) \geq \sqrt{2}n\,.$$

This rather weak lower bound stood as the best bound until 2004, when Mignon and Ressayre proved that

$$\mathsf{dc}(\mathrm{per}_n) \geq n^2/2,$$

over any field of char 0. Over a field of char $\mathbb{F} \neq 2$, the best bound is a recent unpublished result by Valiant (2007), which is $\Omega(n^{5/4})$ for projections.

More important than the lower bound $\sqrt{8/7}n$, von zur Gathen (1985) introduced a method of taking derivatives and then comparing appropriate dimensions/ranks. The follow-up improvements to $\sqrt{2}n$ all use this approach.

The Mignon–Ressayre breakthrough (2004) uses a new idea: Take second-order derivatives.

The key step in their proof is to lower bound the rank of the second-order derivative matrix $\mathbf{H}$ of the permanent at a certain matrix $\mathbf{X}_0$. However, their proof encounters a major difficulty when char $\mathbb{F} \neq 0$. The matrix $\mathbf{H}$ at $\mathbf{X}_0$ has various non-zero entries, which is a necessary condition to being of high rank. However, these non-zero entries are all divisible by large factorials. Thus when char $\mathbb{F} = p \neq 0$, a constant, these entries are all zero, and the matrix $\mathbf{H}$ becomes $\mathbf{0}$. In this paper, we overcome this difficulty by considering another explicit construction of $\mathbf{X}_0$.

We mention some other related results. Jerrum & Snir (1982) showed that any monotone arithmetic circuit family that computes permanent must have exponential size. For depth-three arithmetic circuits over fields of char $\mathbb{F} = 0$, Shpilka & Wigderson (2001) proved that the permanent and determinant require circuit size $\Omega(n^2)$. For depth-three arithmetic circuits over finite characteristic, Grigoriev & Razborov (2000) showed an exponential lower bound for both determinant and permanent. Raz (2004; 2009) proved a lower bound of $n^{\Omega(\log n)}$ on the size of families of multilinear formulas computing permanent and determinant. For syntactically multilinear arithmetic circuits, Raz, Shpilka & Yehudayoff (2007) proved a $\Omega(n^{4/3}/\log^2 n)$ lower bound for an explicit multilinear function. A survey of some work on this Permanent and Determinant Problem can be found in Agrawal (2006), where it also discusses an algebraic geometry approach by Mulmulay & Sohoni (2002) and connections to the pseudorandom generator used in the AKS proof for primality (Agrawal 2005; Agrawal, Kayal & Saxena 2004).

The paper is organized as follows. In Section 2, we discuss the general approach by Mignon and Ressayre (2004) and state our main result. In Section 3, we prove an $\Omega(n^2)$ lower bound that is valid for all fields of characteristic $\neq 2$. Finally, in Section 4, we indicate how to improve the leading constant in our $\Omega(n^2)$ lower bound to match the Mignon–Ressayre bound.

2. The approach and the theorem

2.1. The proof by Mignon and Ressayre. Given an $n \times n$ matrix $\mathbf{X} = (x_{i,j})_{i,j=1,2,\ldots,n}$ over a field $\mathbb{F}$, it is clear that both $\det(\mathbf{X})$ and $\mathrm{per}(\mathbf{X})$ are polynomials of degree n over n^2 variables. Their partial derivatives of all orders are defined formally.

We let $\mathbf{H}(\mathbf{X}) = (H_{ij,kl})_{i,j,k,l=1,2,\ldots,n}$ denote the *Hessian* matrix of $\mathrm{per}(\mathbf{X})$:

$$H_{ij,kl} = \frac{\partial^2 \mathrm{per}(\mathbf{X})}{\partial x_{i,j}\, \partial x_{k,l}} \in \mathbb{F}[\mathbf{X}]\,, \quad \text{for all} \quad 1 \le i, j, k, l \le n\,.$$

Similarly, we can define the Hessian of $\det(\mathbf{X})$, and denote it by $\mathbf{H}_{\det}(\mathbf{X})$.

Now suppose there exists a collection $\mathbf{A}$ of m^2 affine linear functions, where

$$\mathbf{A} = \big\{ A_{k,l}(x_{1,1}, x_{1,2}, \ldots, x_{n,n}), \text{ where } k, l : 1 \le k, l \le m \big\}\,,$$

such that in the polynomial ring $\mathbb{F}[\mathbf{X}]$,

$$\mathrm{per}_n(\mathbf{X}) = \det_m\Big(\big(A_{k,l}(\mathbf{X})\big)_{1 \le k,l \le m} \Big)\,. \tag{2.1}$$

The first step in the proof by Mignon & Ressayre (2004) is to transform $\mathbf{A}$ to a normal form. Consider a fixed matrix $\mathbf{X}_0 \in \mathbb{F}^{n \times n}$ such that $\mathrm{per}(\mathbf{X}_0) = 0$. We expand the affine linear functions $A_{k,l}(\mathbf{X})$ at $\mathbf{X}_0$, and write

$$\big(A_{k,l}(\mathbf{X})\big) = \big(L_{k,l}(\mathbf{X} - \mathbf{X}_0)\big) + \mathbf{Y}_0$$

for some homogeneous linear functions $L_{k,l}$ and some matrix $\mathbf{Y}_0 \in \mathbb{F}^{m \times m}$. It then follows from (2.1) that $\det(\mathbf{Y}_0) = \mathrm{per}(\mathbf{X}_0) = 0$. Now let $\mathbf{C}$ and $\mathbf{D}$ be two non-singular matrices such that $\mathbf{C}\mathbf{Y}_0\mathbf{D}$ is a diagonal matrix

$$\begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_s \end{pmatrix}, \quad \text{where} \quad s < m\,.$$

It follows from previous work (Cai 1990; von zur Gathen 1987; Meshulam 1989) that if (2.1) holds, then this s must be $m - 1$. (But it will also follow easily

from the Mignon–Ressayre proof.) Since the first row and column of $\mathbf{C}\mathbf{Y}_0\mathbf{D}$ are both zero, we may multiply diagonal matrices $\mathrm{diag}(\det(\mathbf{C})^{-1}, 1, \ldots, 1)$ and $\mathrm{diag}(\det(\mathbf{D})^{-1}, 1, \ldots, 1)$ to the left and right, so we may just assume $\det(\mathbf{C}) = \det(\mathbf{D}) = 1$. It follows that, by (multiplying matrices $\mathbf{C}$ and $\mathbf{D}$ to the left and right, and) renaming $L_{k,l}$ and $\mathbf{Y}_0$, we may assume (2.1) takes the form

$$\mathrm{per}(\mathbf{X}) = \det\Big(\big(L_{k,l}\,(\mathbf{X} - \mathbf{X}_0)\big) + \mathbf{Y}_0\Big)\,,$$

where $\mathbf{Y}_0 = \mathrm{diag}(0, 1, \ldots, 1)$.

Now we can take second-order derivatives, and evaluate them at $\mathbf{X}_0$. By the chain rule, we have

$$\mathbf{H}(\mathbf{X}_0) = \mathbf{L} \cdot \mathbf{H}_{\det}(\mathbf{Y}_0) \cdot \mathbf{L}^T\,,$$

where $\mathbf{L}$ is an $n^2 \times m^2$ matrix over $\mathbb{F}$. It immediately follows that

$$\mathrm{rank}\big(\mathbf{H}(\mathbf{X}_0)\big) \le \mathrm{rank}\big(\mathbf{H}_{\det}(\mathbf{Y}_0)\big)\,.$$

It is relatively easy to derive a $O(m)$ upper bound for the rank of $\mathbf{H}_{\det}(\mathbf{Y}_0)$. Notice that when one takes a partial derivative $\partial/\partial x_{ij}$ on the determinant (as well as on the permanent), one simply gets the minor after striking out row i and column j. Second order derivative $\partial^2/\partial x_{ij}\partial x_{kl}$ simply strikes out rows $\{i, k\}$ and columns $\{j, l\}$. By the form of $\mathbf{Y}_0$, to get a non-zero value for an entry (ij, kl) in $\mathbf{H}_{\det}(\mathbf{Y}_0)$, it must be that $1 \in \{i, k\}$ and $1 \in \{j, l\}$. In fact the only non-zero entries are

$$(ij, kl) = (11, tt), (tt, 11), (1t, t1) \;\text{ or }\; (t1, 1t)\,,$$

for all $t > 1$. This immediately gives a $2m$ upper bound for $\mathrm{rank}(\mathbf{H}_{\det}(\mathbf{Y}_0))$. (If we did not assume $s = m - 1$, then it would have been even more difficult to get a non-zero entry in $\mathbf{H}_{\det}(\mathbf{Y}_0)$. If $s = m - 2$, there could be at most $O(1)$ many non-zero entries. If $s < m - 2$, there are no non-zero entries.)

The real work of their proof is to find an explicit $\mathbf{X}_0$ such that $\mathrm{per}(\mathbf{X}_0) = 0$ and yet $\mathrm{rank}(\mathbf{H}(\mathbf{X}_0))$ is high. For the case when $\mathrm{char}\,\mathbb{F} = 0$, they constructed an infinite sequence of $n \times n$ matrices $\mathbf{X}_0$ such that $\mathrm{per}_n(\mathbf{X}_0) = 0$ and the rank of the $n^2 \times n^2$ matrix $\mathbf{H}(\mathbf{X}_0)$ is full. This gives their quadratic lower bound $m = \Omega(n^2)$.

THEOREM 2.2 (Mignon and Ressayre). *For any field of characteristic* 0,

$$\mathsf{dc}(\mathrm{per}_n) \ge n^2/2\,.$$

However, their matrices $\mathbf{X}_0$ do not work for fields $\mathbb{F}$ with small characteristics, e.g., 3. All entries of $\mathbf{H}(\mathbf{X}_0)$ are divisible by large factorials, and thus, divisible by char $\mathbb{F}$. As a result, $\mathbf{H}(\mathbf{X}_0)$ becomes the zero matrix of rank 0. In a way, to get non-zero values for entries in $\mathbf{H}(\mathbf{X}_0)$, which are permanental minors of $\mathbf{X}_0$, and yet to be able to analyze the rank, the most natural approach is to assign pretty uniform values for $\mathbf{X}_0$. This is what was done. But these entries are non-zero by virtue of the fact that they are sums of constant terms with a large factorial number of terms. Thus the appearance of large factorials in $\mathbf{H}(\mathbf{X}_0)$ is not surprising. To avoid these factorials, we have to be more judicious in our choice of $\mathbf{X}_0$. We need it to be not terribly uniform, and yet sufficiently structured so that we can still calculate the rank for $\mathbf{H}(\mathbf{X}_0)$.

2.2. Our main result. Our main result is a new construction of matrices $\mathbf{X}_0$ such that $\mathbf{H}(\mathbf{X}_0)$ has almost full rank over any field of char $\mathbb{F} \neq 2$. More exactly, we will prove the following theorem in Section 4:

THEOREM 2.3. *Let $p > 2$ be a prime, then*

(i) If $p \neq 23$, then for every $n > 2$ that satisfies $p \big| (n+1)$, there exists an $(n+1) \times (n+1)$ matrix $\mathbf{X}_0$ over finite field $\mathbb{F}_p$ such that

$$\operatorname{per}(\mathbf{X}_0) \equiv 0 \pmod{p} \quad \textit{and} \quad \operatorname{rank}\big(\mathbf{H}(\mathbf{X}_0)\big) \geq (n-2)(n-3)\,;$$

(ii) If $p \neq 3, 5$, then for every $n > 1$ that satisfies $p \big| (n+2)$, there exists an $(n+1) \times (n+1)$ matrix $\mathbf{X}_0$ over finite field $\mathbb{F}_p$ such that

$$\operatorname{per}(\mathbf{X}_0) \equiv 0 \pmod{p} \quad \textit{and} \quad \operatorname{rank}\big(\mathbf{H}(\mathbf{X}_0)\big) \geq (n-2)(n-3)\,.$$

This implies a quadratic lower bound for $\mathsf{dc}(\mathrm{per})$ over field $\mathbb{F}_p$. We remark that a lower bound for $\mathbb{F}_p$ is also valid over $\mathbb{Q}$.

COROLLARY 2.4. *For every prime $p \neq 2$, there exist infinitely many positive integers n such that $\mathsf{dc}(\mathrm{per}_n) \geq (n-2)(n-3)/2$ over a field of* char $\mathbb{F} = p$.

To prove the theorem, we introduce, for all $v \in \mathbb{F}_p$ and $n \geq 1$, the following $(n+1) \times (n+1)$ matrix $\mathbf{M}_v^n = (M_{i,j})$: $M_{(n+1),(n+1)} = v$ and $M_{i,i} = M_{(n+1),i} = M_{i,(n+1)} = 1$ for all $i : 1 \leq i \leq n$, and $M_{i,j} = 0$ otherwise. For example,

$$\mathbf{M}_2^3 = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 2 \end{pmatrix}.$$

In Section 4, we will prove the two cases of Theorem 2.3 using $\mathbf{M}_1^n$ and $\mathbf{M}_2^n$, respectively. Given $v \in \mathbb{F}_p$ and $n \geq 1$, the following lemma essentially defines the Hessian matrix $\mathbf{H}(\mathbf{M}_v^n)$ of $\mathbf{M}_v^n$.

LEMMA 2.5. *Let* $\mathbf{H}(\mathbf{M}_v^n) = (H_{ij,kl})$. *Then for all* $i, j : 1 \leq i \neq j \leq n$ *and* $k, l : 1 \leq k \neq l \leq n$, *we have*

$$H_{ij,kl} \equiv \begin{cases} v+n-2 & \text{if } k = j \text{ and } l = i\,; \\ 1 & \text{if } k = j \text{ and } l \neq i, j\,; \\ 1 & \text{if } l = i \text{ and } k \neq i, j\,; \\ 0 & \text{otherwise}\,. \end{cases}$$

For $i, j : 1 \leq i \neq j \leq n$, we let $\mathbf{H}_{ij}$ denote the $(n^2 - n)$-dimensional vector truncated from the $(ij)^{th}$ row of $\mathbf{H}(\mathbf{M}_v^n)$, where we only keep its $(kl)^{th}$ entry if $1 \leq k \neq l \leq n$. For all i, j, k, l satisfying $1 \leq i \neq j \leq n$ and $1 \leq k \neq l \leq n$, the following lemma shows the possible values of the inner product $\mathbf{H}_{ij} \cdot \mathbf{H}_{kl}$.

LEMMA 2.6. *Assume* i *and* j *satisfy* $1 \leq i \neq j \leq n$, *then we have*

1. $\mathbf{H}_{ij} \cdot \mathbf{H}_{ij} = (v+n-2)^2 + 2(n-2)$;
2. $\mathbf{H}_{ij} \cdot \mathbf{H}_{ji} = 0$;
3. *for* $1 \leq k \leq n$ *and* $k \neq i, j$, $\mathbf{H}_{ij} \cdot \mathbf{H}_{ik} = \mathbf{H}_{ij} \cdot \mathbf{H}_{kj} = 2(v+n-2) + n - 3$;
4. *for* $1 \leq k \leq n$ *and* $k \neq i, j$, $\mathbf{H}_{ij} \cdot \mathbf{H}_{ki} = \mathbf{H}_{ij} \cdot \mathbf{H}_{jk} = 1$; *and*
5. *for* $1 \leq k \neq l \leq n$ *and* $\{k, l\} \cap \{i, j\} = \emptyset$, $\mathbf{H}_{ij} \cdot \mathbf{H}_{kl} = 2$.

PROOF. We only prove the first and third cases here. The other cases can be proved similarly.

For the first case, we run all possibilities (kl), where $k, l : 1 \leq k \neq l \leq n$, and the only non-zero entries in $\mathbf{H}_{ij} \cdot \mathbf{H}_{ij}$ are $(v+n-2)^2$ for the index (ji), and 1 for indices (jt) and (ti), where $1 \leq t \leq n$ and $t \neq i, j$. As a result, $\mathbf{H}_{ij} \cdot \mathbf{H}_{ij} = (v+n-2)^2 + 2(n-2)$. For the third case, the only non-zero entries in $\mathbf{H}_{ij} \cdot \mathbf{H}_{ik}$ are $(v+n-2)$ for indices (ji) and (ki), and 1 for indices (ti) where $1 \leq t \leq n$ and $t \neq i, j, k$. As a result, $\mathbf{H}_{ij} \cdot \mathbf{H}_{ik} = 2(v+n-2) + n - 3$. □

We also need the following lemma concerning the determinant of matrices of a specific form.

LEMMA 2.7. *Let $\mathbf{A} = (A_{i,j})_{i,j=1,\ldots,n}$ be an $n \times n$ matrix over $\mathbb{F}_p$, which satisfies $A_{i,i} = \alpha$ for all $1 \le i \le n$ and $A_{i,j} = \beta$ otherwise. Then we have*

$$\det(\mathbf{A}) = \big(\alpha + (n-1)\beta\big)(\alpha - \beta)^{n-1}\,.$$

PROOF. First, we add the ith row of $\mathbf{A}$ to the first row for all $i : 1 < i \le n$. As a result, we have

$$\det(\mathbf{A}) = \det\begin{pmatrix} \gamma & \gamma & \cdots & \gamma \\ \beta & \alpha & \cdots & \beta \\ \vdots & \vdots & \ddots & \vdots \\ \beta & \beta & \cdots & \alpha \end{pmatrix} = \gamma \cdot \det(\mathbf{B})\,, \quad \text{where} \quad \mathbf{B} = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ \beta & \alpha & \cdots & \beta \\ \vdots & \vdots & \ddots & \vdots \\ \beta & \beta & \cdots & \alpha \end{pmatrix}$$

and $\gamma = \alpha + (n-1)\beta$. Second, for each $i : 1 < i \le n$, we subtract $(\beta, \beta, \ldots, \beta)$ from the ith row of $\mathbf{B}$:

$$\det(\mathbf{B}) = \det\begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & \alpha - \beta & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \alpha - \beta \end{pmatrix} = (\alpha - \beta)^{n-1}\,.$$

The lemma then follows. □

3. A weaker theorem

In this section, we prove the following weaker version of Theorem 2.3.

LEMMA 3.1. *Let $p > 2$ be a prime, then for any sufficiently large n satisfying $p\,|(n+1)$, we have $\operatorname{per}(\mathbf{M}_1^n) \equiv 0 \pmod p$ and $\operatorname{rank}(\mathbf{H}(\mathbf{M}_1^n)) = \Omega(n^2)$.*

PROOF. In the proof, we denote matrix $\mathbf{M}_1^n$ by $\mathbf{M}$. Clearly,

$$\operatorname{per}(\mathbf{M}) = n + 1 \equiv 0 \pmod p\,,$$

so we only need to prove the second part.

Let S be a *maximal* subset of integers $\{i : 1 \le i < n/2\}$ with $|S| \equiv 2 \pmod p$, and T be a *maximal* subset of $\{j : n/2 \le j \le n\}$ with $|T| \equiv 2 \pmod p$. Both $|S|$ and $|T|$ are $\Omega(n)$.

Next, we will show that there exists a sub-matrix $\mathbf{R}$ of $\mathbf{H}(\mathbf{M})$ with $|S| \cdot |T|$ rows, such that, $\det(\mathbf{R}\mathbf{R}^T)$ is non-zero. As a result, we have

$$\operatorname{rank}\big(\mathbf{H}(\mathbf{M})\big) \ge \operatorname{rank}(\mathbf{R}) \ge \operatorname{rank}(\mathbf{R}\mathbf{R}^T) = |S| \cdot |T| = \Omega(n^2)\,,$$

and the lemma follows.

To get the matrix $\mathbf{R}$, we choose the following subset of rows and columns of $\mathbf{H}(\mathbf{M})$: rows (ij), where $i \in S$ and $j \in T$; and columns (kl), where $1 \le k \ne l \le n$. So $\mathbf{R}$ is an $(|S| \cdot |T|) \times (n^2 - n)$ matrix. Let $S = \{i_1, i_2, \ldots, i_{|S|}\}$ and $T = \{j_1, j_2, \ldots, j_{|T|}\}$, then we can write $\mathbf{R}$ as

$$\mathbf{R} = \begin{pmatrix} \mathbf{H}_{i_1 j_1} \\ \mathbf{H}_{i_1 j_2} \\ \vdots \\ \mathbf{H}_{i_1 j_{|T|}} \\ \mathbf{H}_{i_2 j_1} \\ \vdots \\ \mathbf{H}_{i_{|S|} j_{|T|}} \end{pmatrix},$$

where $\mathbf{H}_{ij}$ is the $(n^2 - n)$-dimensional vector truncated from the $(ij)^{th}$ row of $\mathbf{H}(\mathbf{M})$. Consider the inner products of arbitrary two rows of $\mathbf{R}$. By Lemma 2.6 we have for $i \in S$ and $j \in T$,

1. $\mathbf{H}_{ij} \cdot \mathbf{H}_{ij} = (v + n - 2)^2 + 2(n - 2) \equiv -2 \pmod p$, since $v = 1$, $n \equiv -1 \pmod p$ by the assumption;

2. when $j' \ne j$ and $j' \in T$, $\mathbf{H}_{ij} \cdot \mathbf{H}_{ij'} = 2(v + n - 2) + (n - 3) \equiv -8 \not\equiv 0 \pmod p$;

3. when $i' \ne i$ and $i' \in S$, $\mathbf{H}_{ij} \cdot \mathbf{H}_{i'j} \equiv -8 \not\equiv 0 \pmod p$;

4. when $i' \ne i$, $j' \ne j$, $i' \in S$ and $j' \in T$, $\mathbf{H}_{ij} \cdot \mathbf{H}_{i'j'} \equiv 2 \pmod p$.

Now we can write $\mathbf{R}\mathbf{R}^T$ as an $|S| \times |S|$ block matrix:

$$\mathbf{R}\mathbf{R}^T = \begin{pmatrix} \mathbf{A} & \mathbf{B} & \mathbf{B} & \cdots & \mathbf{B} \\ \mathbf{B} & \mathbf{A} & \mathbf{B} & \cdots & \mathbf{B} \\ \mathbf{B} & \mathbf{B} & \mathbf{A} & \cdots & \mathbf{B} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{B} & \mathbf{B} & \mathbf{B} & \cdots & \mathbf{A} \end{pmatrix}, \quad \text{where}$$

$$\mathbf{A} = \begin{pmatrix} a & b & b & \cdots & b \\ b & a & b & \cdots & b \\ b & b & a & \cdots & b \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ b & b & b & \cdots & a \end{pmatrix} \quad \text{and} \quad \mathbf{B} = \begin{pmatrix} b & c & c & \cdots & c \\ c & b & c & \cdots & c \\ c & c & b & \cdots & c \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c & c & c & \cdots & b \end{pmatrix},$$

are both $|T| \times |T|$ matrices with $a = -2$, $b = -8$, and $c = 2$.

We apply the following operations to $\mathbf{R}\mathbf{R}^T$: subtract the second last column from the last column of $\mathbf{R}\mathbf{R}^T$ (Here what we mean by "a column" is a whole block column of $\mathbf{R}\mathbf{R}^T$). Then subtract the third last column from the second last column ... till subtract the first column from the second column. We end up with

$$\begin{pmatrix} \mathbf{A} & \mathbf{B}-\mathbf{A} & 0 & \cdots & 0 \\ \mathbf{B} & \mathbf{A}-\mathbf{B} & \mathbf{B}-\mathbf{A} & \cdots & 0 \\ \mathbf{B} & 0 & \mathbf{A}-\mathbf{B} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{B} & 0 & 0 & \cdots & \mathbf{B}-\mathbf{A} \\ \mathbf{B} & 0 & 0 & \cdots & \mathbf{A}-\mathbf{B} \end{pmatrix}.$$

Then we add the first row to the second row. Add the second row to the third row, etc. Finally, we get

$$\begin{pmatrix} \mathbf{A} & \mathbf{B}-\mathbf{A} & 0 & 0 & \cdots & 0 \\ \mathbf{A}+\mathbf{B} & 0 & \mathbf{B}-\mathbf{A} & 0 & \cdots & 0 \\ \mathbf{A}+2\mathbf{B} & 0 & 0 & \mathbf{B}-\mathbf{A} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{A}+(|S|-2)\mathbf{B} & 0 & 0 & 0 & \cdots & \mathbf{B}-\mathbf{A} \\ \mathbf{A}+(|S|-1)\mathbf{B} & 0 & 0 & 0 & \cdots & 0 \end{pmatrix}.$$

Clearly all these operations do not change its determinant. By Lemma 2.7, we have (Here we use s and t to denote $|S|-1$ and $|T|-1$, respectively)

$$\begin{aligned}\det(\mathbf{R}\mathbf{R}^T) &= \pm\det(\mathbf{A}+s\mathbf{B})\cdot\big(\det(\mathbf{B}-\mathbf{A})\big)^s\\ &= \pm\big(a+sb+t(b+sc)\big)\big(a+sb-(b+sc)\big)^t\\ &\qquad\Big(\big(b-a+t(c-b)\big)\big(b-a-(c-b)\big)^t\Big)^s\\ &\equiv \pm(-16)(-4)^t\big((4)(-16)^t\big)^s \not\equiv 0 \pmod p\,,\end{aligned}$$

since $p>2$ is a prime. As a result, we have $\operatorname{rank}(\mathbf{R}\mathbf{R}^T)=|S|\cdot|T|$, and the lemma is proven. □

4. Proof of the main theorem

In this section, we prove Theorem 2.3. As already mentioned in Section 2.2, we will use $\mathbf{M}_1^n$ and $\mathbf{M}_2^n$ to prove the two cases, respectively. The idea behind the proof is similar to the previous one. However, the sub-matrix $\mathbf{R}$ we pick this time is a square matrix with n^2-n rows. By showing that the rank of $\mathbf{R}\mathbf{R}^T$ is almost full, the theorem follows.

Proof of Theorem 2.3. Let $v=1$ in the first case and $v=2$ in the second case. Note that in both cases, we have $n\equiv -v \pmod p$.

Let $S=\{(i,j):1\le i\ne j\le n\}$. Then we use $\mathbf{R}_v$ to denote the following sub-matrix of $\mathbf{H}(\mathbf{M}_v^n)$: Row (or column) (ij) of $\mathbf{H}(\mathbf{M}_v^n)$ is selected if and only if $(i,j)\in S$. Thus, $\mathbf{R}_v$ is an $(n^2-n)\times(n^2-n)$ matrix. Again, we write $\mathbf{R}_v$ as

$$\mathbf{R}_v=\begin{pmatrix}\mathbf{H}_{12}\\ \mathbf{H}_{13}\\ \vdots\\ \mathbf{H}_{1n}\\ \mathbf{H}_{21}\\ \mathbf{H}_{23}\\ \vdots\\ \mathbf{H}_{n(n-1)}\end{pmatrix},$$

where $\mathbf{H}_{ij}$ is the (n^2-n)-dimensional vector truncated from the $(ij)^{th}$ row of the original matrix $\mathbf{H}(\mathbf{M}_v^n)$. Again, by using Lemma 2.6 we have the following cases (under the assumption that $n\equiv -v \pmod p$): For $(i,j)\in S$,

1. $\mathbf{H}_{ij} \cdot \mathbf{H}_{ij} = (v+n-2)^2 + 2(n-2) \equiv -2v \pmod{p}$. We denote $-2v$ by a.
2. $\mathbf{H}_{ij} \cdot \mathbf{H}_{ji} = 0$.
3. when $1 \leq k \leq n$ and $k \neq i, j$, $\mathbf{H}_{ij} \cdot \mathbf{H}_{ik} = \mathbf{H}_{ij} \cdot \mathbf{H}_{kj} = 2(v+n-2)+(n-3) \equiv -(v+7) \pmod{p}$. We denote $-(v+7)$ by b.
4. when $1 \leq k \leq n$ and $k \neq i, j$, $\mathbf{H}_{ij} \cdot \mathbf{H}_{ki} = \mathbf{H}_{ij} \cdot \mathbf{H}_{jk} = 1$.
5. when $1 \leq k \neq l \leq n$ and $\{k, l\} \cap \{i, j\} = \emptyset$, $\mathbf{H}_{ij} \cdot \mathbf{H}_{kl} = 2$.

Therefore, $\mathbf{R}_v \mathbf{R}_v^T$ is an $n \times n$-block matrix in which each block is an $(n-1) \times (n-1)$ matrix. An example, when $n = 6$, is shown in Figure 4.1.

In Figure 4.1, notice that the $(1,2)$th block can be transformed into the $(1,6)$th block with the following operations: Move the 1st row to the 5th row and then move the 2nd-5th rows up by one row. One can also transform the $(1,6)$th block into the $(5,6)$th block by simply moving the 1st column to the 5th column and moving the 2nd-5th columns one column left. Let $\mathbf{A}$ and $\mathbf{B}$ be the following $(n-1) \times (n-1)$ matrices,

$$\mathbf{A} = \begin{pmatrix} a & b & b & b & \cdots & b \\ b & a & b & b & \cdots & b \\ b & b & a & b & \cdots & b \\ b & b & b & a & \cdots & b \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ b & b & b & b & \cdots & a \end{pmatrix} \quad \text{and} \quad \mathbf{B} = \begin{pmatrix} 0 & 1 & 1 & 1 & \cdots & 1 \\ 1 & b & 2 & 2 & \cdots & 2 \\ 1 & 2 & b & 2 & \cdots & 2 \\ 1 & 2 & 2 & b & \cdots & 2 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 2 & 2 & 2 & \cdots & b \end{pmatrix},$$

then we formally state the property observed above in the following lemma.

LEMMA 4.1. *The $(1,2)$th block of $\mathbf{R}_v \mathbf{R}_v^T$ is $\mathbf{B}$. For any $i : 1 \leq i \leq n-1$, let $\mathbf{C}_i$ denote the following $(n-1) \times (n-1)$ matrix:*

$$\mathbf{C}_i = \begin{pmatrix} \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & 1 \\ 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \end{pmatrix}_{i \times i} & \\ & \mathbf{I}_{n-1-i} \end{pmatrix}.$$

Then for all $i, j : 1 \leq i < j \leq n$, the (i,j)th block of $\mathbf{R}_v \mathbf{R}_v^T$ is $\mathbf{C}_{j-1}^T \mathbf{B} \mathbf{C}_i$.

	12	13	14	15	16	21	23	24	25	26	31	32	34	35	36	41	42	43	45	46	51	52	53	54	56	61	62	63	64	65
12	a	b	b	b	b	0	1	1	1	1	1	b	2	2	2	1	b	2	2	2	1	b	2	2	2	1	b	2	2	2
13	b	a	b	b	b	1	b	2	2	2	0	1	1	1	1	1	2	b	2	2	1	2	b	2	2	1	2	b	2	2
14	b	b	a	b	b	1	2	b	2	2	1	2	b	2	2	0	1	1	1	1	1	2	2	b	2	1	2	2	b	2
15	b	b	b	a	b	1	2	2	b	2	1	2	2	b	2	1	2	2	b	2	0	1	1	1	1	1	2	2	2	b
16	b	b	b	b	a	1	2	2	2	b	1	2	2	2	b	1	2	2	2	b	1	2	2	2	b	0	1	1	1	1
21	0	1	1	1	1	a	b	b	b	b	b	1	2	2	2	b	1	2	2	2	b	1	2	2	2	b	1	2	2	2
23	1	b	2	2	2	b	a	b	b	b	1	0	1	1	1	2	1	b	2	2	2	1	b	2	2	2	1	b	2	2
24	1	2	b	2	2	b	b	a	b	b	2	1	b	2	2	1	0	1	1	1	2	1	2	b	2	2	1	2	b	2
25	1	2	2	b	2	b	b	b	a	b	2	1	2	b	2	2	1	2	b	2	1	0	1	1	1	2	1	2	2	b
26	1	2	2	2	b	b	b	b	b	a	2	1	2	2	b	2	1	2	2	b	2	1	2	2	b	1	0	1	1	1
31	1	0	1	1	1	b	1	2	2	2	a	b	b	b	b	b	2	1	2	2	b	2	1	2	2	b	2	1	2	2
32	b	1	2	2	2	1	0	1	1	1	b	a	b	b	b	2	b	1	2	2	2	b	1	2	2	2	b	1	2	2
34	2	1	b	2	2	2	1	b	2	2	b	b	a	b	b	1	1	0	1	1	2	2	1	b	2	2	2	1	b	2
35	2	1	2	b	2	2	1	2	b	2	b	b	b	a	b	2	2	1	b	2	1	1	0	1	1	2	2	1	2	b
36	2	1	2	2	b	2	1	2	2	b	b	b	b	b	a	2	2	1	2	b	2	2	1	2	b	1	1	0	1	1
41	1	1	0	1	1	b	2	1	2	2	b	2	1	2	2	a	b	b	b	b	b	2	2	1	2	b	2	2	1	2
42	b	2	1	2	2	1	1	0	1	1	2	b	1	2	2	b	a	b	b	b	2	b	2	1	2	2	b	2	1	2
43	2	b	1	2	2	2	b	1	2	2	1	1	0	1	1	b	b	a	b	b	2	2	b	1	2	2	2	b	1	2
45	2	2	1	b	2	2	2	1	b	2	2	2	1	b	2	b	b	b	a	b	1	1	1	0	1	2	2	2	1	b
46	2	2	1	2	b	2	2	1	2	b	2	2	1	2	b	b	b	b	b	a	2	2	2	1	b	1	1	1	0	1
51	1	1	1	0	1	b	2	2	1	2	b	2	2	1	2	b	2	2	1	2	a	b	b	b	b	b	2	2	2	1
52	b	2	2	1	2	1	1	1	0	1	2	b	2	1	2	2	b	2	1	2	b	a	b	b	b	2	b	2	2	1
53	2	b	2	1	2	2	b	2	1	2	1	1	1	0	1	2	2	b	1	2	b	b	a	b	b	2	2	b	2	1
54	2	2	b	1	2	2	2	b	1	2	2	2	b	1	2	1	1	1	0	1	b	b	b	a	b	2	2	2	b	1
56	2	2	2	1	b	2	2	2	1	b	2	2	2	1	b	2	2	2	1	b	b	b	b	b	a	1	1	1	1	0
61	1	1	1	1	0	b	2	2	2	1	b	2	2	2	1	b	2	2	2	1	b	2	2	2	1	a	b	b	b	b
62	b	2	2	2	1	1	1	1	1	0	2	b	2	2	1	2	b	2	2	1	2	b	2	2	1	b	a	b	b	b
63	2	b	2	2	1	2	b	2	2	1	1	1	1	1	0	2	2	b	2	1	2	2	b	2	1	b	b	a	b	b
64	2	2	b	2	1	2	2	b	2	1	2	2	b	2	1	1	1	1	1	0	2	2	2	b	1	b	b	b	a	b
65	2	2	2	b	1	2	2	2	b	1	2	2	2	b	1	2	2	2	b	1	1	1	1	1	0	b	b	b	b	a

Figure 4.1: An example of matrix $\mathbf{R}_v\mathbf{R}_v^T$ when $n = 6$.

PROOF. To prove the lemma, it suffices to show that, for all $i, j : 1 \le i < j < n$, the $(i, j+1)$th block of $\mathbf{R}_v\mathbf{R}_v^T$ can be obtained from its (i, j)th block by exchanging the $(j-1)$th and jth rows; and for all $i, j : 1 \le i < j-1 < n$, the $(i+1, j)^{th}$ block of $\mathbf{R}_v\mathbf{R}_v^T$ can be obtained from its $(i, j)^{th}$ block by exchanging the ith and $(i+1)$th columns. We only prove the first statement here. Assume i and j satisfy $1 \le i < j < n$. We define the following mappings:

$$\gamma(l) = \begin{cases} l & l \neq j, j+1\,; \\ j+1 & l = j\,; \\ j & l = j+1\,, \end{cases} \quad \text{and} \quad \sigma_r(l) = \begin{cases} l & l \le r-1\,; \\ l+1 & l \ge r\,, \end{cases}$$

for all $r \in \mathbb{Z}$. One can easily check that for any $l \in \mathbb{Z}$, $\gamma(\sigma_j(l)) = \sigma_{j+1}(l)$.

First, our analysis of $\mathbf{H}_{ij} \cdot \mathbf{H}_{kl}$ implies that

$$\mathbf{H}_{ij} \cdot \mathbf{H}_{kl} = \mathbf{H}_{\gamma(i)\gamma(j)} \cdot \mathbf{H}_{\gamma(k)\gamma(l)}\,.$$

This is because the value of $\mathbf{H}_{ij} \cdot \mathbf{H}_{kl}$ only depends on the equality relations between indices i, j and k, l (e.g., whether i is equal to k or not). As a result, exchanging j and $j+1$ does not change the inner product.

Second, for all $k, k' : 1 \le k, k' \le n-1$, we observe that the (k, k')th entry of the (i, j)th block of $\mathbf{R}_v\mathbf{R}_v^T$ is $\mathbf{H}_{i,\sigma_i(k)} \cdot \mathbf{H}_{j,\sigma_j(k')}$, while the (k, k')th entry of its $(i, j+1)$th block is $\mathbf{H}_{i,\sigma_i(k)} \cdot \mathbf{H}_{j+1,\sigma_{j+1}(k')}$. To compare the two blocks, we need to consider the following cases about k:

1. $k < j-1$. Then $\sigma_i(k) \le k+1 < j$, and $\gamma(\sigma_i(k)) = \sigma_i(k)$. As a result,

$$\mathbf{H}_{i,\sigma_i(k)} \cdot \mathbf{H}_{j,\sigma_j(k')} = \mathbf{H}_{\gamma(i),\gamma(\sigma_i(k))} \cdot \mathbf{H}_{\gamma(j),\gamma(\sigma_j(k'))} = \mathbf{H}_{i,\sigma_i(k)} \cdot \mathbf{H}_{j+1,\sigma_{j+1}(k')}\,.$$

2. $k > j$. Similarly, we have $\mathbf{H}_{i,\sigma_i(k)} \cdot \mathbf{H}_{j,\sigma_j(k')} = \mathbf{H}_{i,\sigma_i(k)} \cdot \mathbf{H}_{j+1,\sigma_{j+1}(k')}$.

3. $k = j-1$, then $\gamma(\sigma_i(k)) = j+1 = \sigma_i(j)$. Therefore,

$$\mathbf{H}_{i,\sigma_i(k)} \cdot \mathbf{H}_{j,\sigma_j(k')} = \mathbf{H}_{\gamma(i),\gamma(\sigma_i(k))} \cdot \mathbf{H}_{\gamma(j),\gamma(\sigma_j(k'))} = \mathbf{H}_{i,\sigma_i(j)} \cdot \mathbf{H}_{j+1,\sigma_{j+1}(k')}\,.$$

4. $k = j$, then $\gamma(\sigma_i(k)) = j = \sigma_i(j-1)$. Similarly,

$$\mathbf{H}_{i,\sigma_i(k)} \cdot \mathbf{H}_{j,\sigma_j(k')} = \mathbf{H}_{\gamma(i),\gamma(\sigma_i(k))} \cdot \mathbf{H}_{\gamma(j),\gamma(\sigma_j(k'))} = \mathbf{H}_{i,\sigma_i(j-1)} \cdot \mathbf{H}_{j+1,\sigma_{j+1}(k')}\,.$$

The lemma then follows. □

Now we know $\mathbf{R}_v\mathbf{R}_v^T$ has the following form (We let $*$ denote the blocks we don't care, although we know exactly what they are since $\mathbf{R}_v\mathbf{R}_v^T$ is symmetric):

$$\begin{pmatrix} \mathbf{A} & \mathbf{C}_1^T\mathbf{B}\mathbf{C}_1 & \mathbf{C}_2^T\mathbf{B}\mathbf{C}_1 & \cdots & \mathbf{C}_{n-2}^T\mathbf{B}\mathbf{C}_1 & \mathbf{C}_{n-1}^T\mathbf{B}\mathbf{C}_1 \\ * & \mathbf{A} & \mathbf{C}_2^T\mathbf{B}\mathbf{C}_2 & \cdots & \mathbf{C}_{n-2}^T\mathbf{B}\mathbf{C}_2 & \mathbf{C}_{n-1}^T\mathbf{B}\mathbf{C}_2 \\ * & * & \mathbf{A} & \cdots & \mathbf{C}_{n-2}^T\mathbf{B}\mathbf{C}_3 & \mathbf{C}_{n-1}^T\mathbf{B}\mathbf{C}_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ * & * & * & \cdots & \mathbf{A} & \mathbf{C}_{n-1}^T\mathbf{B}\mathbf{C}_{n-1} \\ * & * & * & \cdots & * & \mathbf{A} \end{pmatrix}.$$

Again, we will apply matrix operations to $\mathbf{R}_v\mathbf{R}_v^T$. But before that, we need to prove the following key property about the block matrices in $\mathbf{R}_v\mathbf{R}_v^T$: The difference between the $(i+1,j+1)$th and $(i+1,j)$th blocks of $\mathbf{R}_v\mathbf{R}_v^T$ is exactly the same as the difference between the $(i,j+1)$th and (i,j)th blocks.

LEMMA 4.2. *For all $1 \le i < j \le n$ such that $i+1 < j$ and $j+1 \le n$, we have*

$$(\mathbf{C}_j^T - \mathbf{C}_{j-1}^T)\mathbf{B}\mathbf{C}_{i+1} = (\mathbf{C}_j^T - \mathbf{C}_{j-1}^T)\mathbf{B}\mathbf{C}_i\,.$$

PROOF. For $k : 1 \le k \le n-1$, we use $\mathbf{B}_k$ to denote the k^{th} row vector of $\mathbf{B}$. We also use $\mathbf{B}'$ to denote $(\mathbf{C}_j^T - \mathbf{C}_{j-1}^T)\mathbf{B}$, and $\mathbf{B}'_k$ to denote the k^{th} row of $\mathbf{B}'$. It is not hard to check that $\mathbf{B}'_{j-1} = \mathbf{B}_j - \mathbf{B}_1$, $\mathbf{B}'_j = \mathbf{B}_1 - \mathbf{B}_j$, and $\mathbf{B}'_k = \mathbf{0}$ for all $k \ne j-1, j$.

On the other hand, all the entries of vector $\mathbf{B}_j - \mathbf{B}_1$ are equal to 1 except the jth entry which is equal to $b-1$. As we assumed that $i+1 < j$, we have $\mathbf{B}'\mathbf{C}_{i+1} = \mathbf{B}'\mathbf{C}_i = \mathbf{B}'$, and the lemma is proven. □

We apply the following operations to $\mathbf{R}_v\mathbf{R}_v^T$: subtract the second last column from the last column of $\mathbf{R}_v\mathbf{R}_v^T$, then subtract the third last column from the second last column ... till subtract the first column from the second column. Let $\mathbf{P}$ denote the upper right sub-matrix, after the operations, of $\mathbf{R}_v\mathbf{R}_v^T$ containing $(n-1)\times(n-1)$ blocks:

$$\begin{pmatrix} \mathbf{C}_1^T(\mathbf{B}-\mathbf{A})\mathbf{C}_1 & (\mathbf{C}_2^T - \mathbf{C}_1^T)\mathbf{B}\mathbf{C}_1 & (\mathbf{C}_3^T - \mathbf{C}_2^T)\mathbf{B}\mathbf{C}_1 & \cdots & (\mathbf{C}_{n-1}^T - \mathbf{C}_{n-2}^T)\mathbf{B}\mathbf{C}_1 \\ * & \mathbf{C}_2^T(\mathbf{B}-\mathbf{A})\mathbf{C}_2 & (\mathbf{C}_3^T - \mathbf{C}_2^T)\mathbf{B}\mathbf{C}_2 & \cdots & (\mathbf{C}_{n-1}^T - \mathbf{C}_{n-2}^T)\mathbf{B}\mathbf{C}_2 \\ * & * & \mathbf{C}_3^T(\mathbf{B}-\mathbf{A})\mathbf{C}_3 & \cdots & (\mathbf{C}_{n-1}^T - \mathbf{C}_{n-2}^T)\mathbf{B}\mathbf{C}_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ * & * & * & \cdots & \mathbf{C}_{n-1}^T(\mathbf{B}-\mathbf{A})\mathbf{C}_{n-1} \end{pmatrix}.$$

Next, we transform $\mathbf{P}$ as follows: Subtract the second last row from the last row, then subtract the third last row from the second last row ... till subtract the first row from the second row. We only need to focus on the lower right part of $\mathbf{P}$ containing $(n-2)\times(n-2)$ blocks, which we denote by $\mathbf{P}^*$. It directly follows from Lemma 4.2 that $\mathbf{P}^*$ is a lower triangular block matrix, and the block matrices along the diagonal are:

$$\big(\mathbf{C}_i^T(\mathbf{B}-\mathbf{A})\mathbf{C}_i - (\mathbf{C}_i^T - \mathbf{C}_{i-1}^T)\mathbf{B}\mathbf{C}_{i-1}\big)\,, \quad i = 2,3,\ldots,n-1\,.$$

On the other hand, as implied by the proof of Lemma 4.2, the rank of matrix $(\mathbf{C}_i^T - \mathbf{C}_{i-1}^T)\mathbf{B}\mathbf{C}_{i-1}$ is exactly 1, so

$$\operatorname{rank}(\mathbf{R}_v\mathbf{R}_v^T) \geq \sum_{i=2}^{n-1}\Big(\operatorname{rank}\big(\mathbf{C}_i^T(\mathbf{B}-\mathbf{A})\mathbf{C}_i\big) - 1\Big) = (n-2)\big(\operatorname{rank}(\mathbf{B}-\mathbf{A}) - 1\big)\,.$$

Finally, by Lemma 2.7, the determinant of the lower right $(n-2)\times(n-2)$ sub-matrix of $\mathbf{B}-\mathbf{A}$ is (by setting $\alpha = b-a$ and $\beta = 2-b$)

$$\begin{aligned}\big((b-a)+(n-3)(2-b)\big)\big((b-a)-(2-b)\big)^{n-3} \\ \equiv \begin{cases}(-46)(-16)^{n-3} \pmod p & \text{when } v=1\,; \quad \text{and}\\ (-60)(-16)^{n-3} \pmod p & \text{when } v=2\,.\end{cases}\end{aligned}$$

As a result, we have

$$\begin{aligned}&\operatorname{rank}\big(\mathbf{H}(\mathbf{M}_1^n)\big) \geq \operatorname{rank}(\mathbf{R}_1\mathbf{R}_1^T) \geq (n-2)(n-3)\,, \quad \text{when} \quad p\neq 23\,; \quad \text{and}\\ &\operatorname{rank}\big(\mathbf{H}(\mathbf{M}_2^n)\big) \geq \operatorname{rank}(\mathbf{R}_2\mathbf{R}_2^T) \geq (n-2)(n-3)\,, \quad \text{when} \quad p\neq 3,5\,. \qquad \square\end{aligned}$$

A natural question is what makes this sequence of matrices work for the proof. We can only offer our take on this. We believe that probably most matrices $\mathbf{X}$, where $\operatorname{per}(\mathbf{X}) = 0$, will work, i.e., its Hessian will have a quadratic rank. The problem is rather how to prove this. Over characteristic 0, Mignon and Ressayre gave a construction which is essentially the all 1 matrix (except the $(1,1)$ entry to make $\operatorname{per}(\mathbf{X}) = 0$). This makes most second derivatives in the Hessian of the permanent a constant (but involving a large factorial). The key to our matrix is to choose it sufficiently uniform so that we can still prove its rank analytically, but not that uniform so as to involve large factorials.

Acknowledgements

We wish to thank Les Valiant for many interesting discussions on the topic. This work would not have been possible for us without the discussions and comments from Les.

Jin-Yi Cai's work was supported by NSF Grant CCR-0511679. Xi Chen's work was supported by NSF Grant DMS-0635607, CCF-0832797, and Princeton Center for Theoretical Computer Science. Dong Li was supported by NSF Grant DMS-0635607 and the start-up funding from the Mathematics Department of University of Iowa.

References

M. Agrawal (2005). Proving lower bounds via pseudo-random generators. In *Proceedings of the 25th Conference on Foundations of Software Technology and Theoretical Computer Science*, 92–105.

M. Agrawal (2006). Determinant versus permanent. In *Proceedings of the International Congress of Mathematicians*, 985–997.

M. Agrawal, N. Kayal & N. Saxena (2004). PRIMES is in P. *Annals of Mathematics* **160**(2), 781–793.

A. Borodin, J. von zur Gathen & J. E. Hopcroft (1982). Fast parallel matrix and GCD computations. In *Proceedings of the 23rd Annual Symposium on Foundations of Computer Science*, 65–71.

P. Bürgisser (2000). *Completeness and Reduction in Algebraic Complexity Theory.* Springer-Verlag.

P. Bürgisser, M. Clausen & M. A. Shokrollahi (1997). *Algebraic Complexity Theory.* Grundlehren der mathematischen Wissenschaften. Springer.

J. Cai (1990). A note on the determinant and permanent problem. *Information and Computation* **84**(1), 119–127.

C. Damm (1991). DET = $\mathrm{L}^{\#\mathrm{L}}$. *Technical Report Informatik-preprint 8,* Fachbereich Informatik der Humboldt Universität zu Berlin.

J. von zur Gathen (1985). Permanent and determinant. In *Proceedings of the 27th Annual IEEE Symposium on Foundations of Computer Science*, 398–401.

J. von zur Gathen (1987). Permanent and determinant. *Linear Algebra and its Applications* **96**, 87–100.

D. Grigoriev & A. Razborov (2000). Exponential lower bounds for depth 3 arithmetic circuits in algebras of functions over finite fields. *Applicable Algebra in Engineering, Communication and Computing* **10**(6), 465–487.

M. Jerrum & M. Snir (1982). Some exact complexity results for straight-line computations over semirings. *Journal of the ACM* **29**(3), 874–897.

M. Marcus & H. Minc (1961). On the relation between the determinant and the permanent. *Illinois Journal of Mathematics* **5**, 376–381.

R. Meshulam (1989). Two extremal matrix problems. *Linear Algebra and its Applications* **114/115**, 261–271.

T. Mignon & N. Ressayre (2004). A quadratic bound for the determinant and permanent problem. *International Mathematics Research Notices* 4241–4253.

H. Minc (1978). *Permanents*. Encyclopedia of Mathematics and its Applications, vol 6. Addison-Wesley.

K. Mulmulay & M. Sohoni (2002). Geometric complexity theory, P vs. NP, and explicit obstructions. *SIAM Journal on Computing* **31**(2), 496–526.

G. Pólya (1913). Aufgabe 424. *Archiv der Mathematik und Physik* **20**, 271.

R. Raz (2004). Multi-linear formulas for permanent and determinant are of super-polynomial size. In *Proceedings of the 36th Annual ACM Symposium on Theory of Computing*, 633–641.

R. Raz (2009). Multi-linear formulas for permanent and determinant are of super-polynomial size. *Journal of the ACM* **56**(2), 1–17.

R. Raz, A. Shpilka & A. Yehudayoff (2007). A lower bound for the size of syntactically multilinear arithmetic circuits. In *Proceedings of the 48th Annual IEEE Symposium on Foundations of Computer Science*, 438–448.

A. Shpilka & A. Wigderson (2001). Depth-3 arithmetic circuits over fields of characteristic zero. *Computational Complexity* **10**(1), 1–27.

G. Szegö (1913). Zu Aufgabe 424. *Archiv der Mathematik und Physik* **21**, 291–292.

S. Toda (1991). Counting problems computationally equivalent to the determinant. *manuscript*.

L. G. Valiant (1979a). The complexity of computing the permanent. *Theoretical Computer Science* **8**(2), 189–201.

L. G. Valiant (1979b). The complexity of enumeration and reliability problems. *SIAM Journal on Computing* **8**(3), 410–421.

L. G. Valiant (2007). Private communication.

V. Vinay (1991). Counting auxiliary pushdown automata and semi-unbounded arithmetic circuits. In *Proceedings of the Structure in Complexity Theory Conference*, 270–284.

Manuscript received 24 December 2008

Jin-Yi Cai
Computer Sciences Department
University of Wisconsin-Madison
Madison, WI 53706-1685, USA
jyc@cs.wisc.edu

Xi Chen
Department of Computer Science
Princeton University
Princeton, NJ 08540-5233, USA
csxichen@gmail.com

Dong Li
School of Mathematics
Institute for Advanced Study
Princeton, NJ 08540, USA
dongli@math.ias.edu

Paper 8

Journal of Computer and System Sciences 77 (2011) 41–61

Contents lists available at ScienceDirect

Journal of Computer and System Sciences

www.elsevier.com/locate/jcss

Holographic algorithms: From art to science ☆

Jin-Yi Cai [a,*,1], Pinyan Lu [b,2]

[a] Computer Sciences Department, University of Wisconsin, Madison, WI 53706, USA
[b] Microsoft Research Asia, Beijing, 100190, PR China

ARTICLE INFO

Article history:
Received 23 June 2009
Received in revised form 22 December 2009
Accepted 7 June 2010
Available online 11 June 2010

Keywords:
Holographic algorithm
Matchgate
Matchgate realizability

ABSTRACT

We develop the theory of holographic algorithms initiated by Leslie Valiant. First we define a basis manifold. Then we characterize algebraic varieties of realizable symmetric generators and recognizers on the basis manifold, and give a polynomial time decision algorithm for the simultaneous realizability problem. These results enable one to decide whether suitable signatures for a holographic algorithm are realizable, and if so, to find a suitable linear basis to realize these signatures by an efficient algorithm. Using the general machinery we are able to give unexpected holographic algorithms for some counting problems, modulo certain Mersenne type integers. These counting problems are #P-complete without the moduli. Going beyond symmetric signatures, we define d-admissibility and d-realizability for general signatures, and give a characterization of 2-admissibility and some general constructions of admissible and realizable families.

1. Introduction

It is a testament to the enormous impact of NP-completeness theory [2,18] that the *conjecture* $P \neq NP$ has become a leading hypothesis in all computer science and mathematics. We consider it a great honor and privilege to dedicate this paper to the 2009 Kyoto Prize Laureate Prof. Richard M. Karp, a founder of this theory.

The NP-completeness theory is so well established that most computer scientists consider it a proof of computational intractability in terms of worst-case complexity if a problem is proved to be NP-complete. Expressed in terms of complexity classes, it has become more or less an article of faith among theoretical computer scientists that the *conjecture* $P \neq NP$ holds. The theory of holographic algorithms, however, provides a cautionary coda, that our understanding of the ultimate capability of polynomial time algorithms is far from well understood.

Certainly there are good reasons to believe the *conjecture* $P \neq NP$, not the least of which is the fact that the usual algorithmic paradigms seem unable to handle any of the NP-hard problems. Such statements are made credible by decades of in-depth study of these methodologies. On the other hand, there are some "surprising" polynomial time algorithms for problems which, on appearance, would seem to require exponential time. One such example is to count the number of perfect matchings in a planar graph (the FKT method) [19,20,25]. In [27,29] L. Valiant introduced an algorithmic design technique of breathtaking originality, called *holographic algorithms*. Computation in these algorithms is expressed and interpreted through a choice of linear basis vectors in an exponential "holographic" mix, and then it is carried out by the FKT method via the Holant Theorem. This methodology has produced polynomial time algorithms for a variety of problems ranging from re-

☆ A preliminary version of this paper appeared in the 39th ACM Symposium on Theory of Computing (STOC 2007) (J.-Y. Cai and P. Lu (2007) [7]).

* Corresponding author.
E-mail addresses: jyc@cs.wisc.edu (J.-Y. Cai), pinyanl@microsoft.com (P. Lu).

[1] Supported by NSF CCR-0208013 and CCR-0511679.

[2] Work performed while the author was a graduate student at Tsinghua University.

0022-0000/$ – see front matter
doi:10.1016/j.jcss.2010.06.005

strictive versions of satisfiability, vertex cover, to other graph problems such as edge orientation and node/edge deletion. No polynomial time algorithms were known for any of these problems, and some minor variations are known to be NP-hard.

These holographic algorithms are quite unusual compared to other kinds of algorithms (except perhaps quantum algorithms). At the heart of the computation is a process of introducing and then canceling exponentially many computational fragments. Invariably the success of this methodology on a particular problem boils down to finding a certain "exotic" object represented by a *signature*.

For example, Valiant showed [30] that the restrictive SAT problem $\#_7$Pl-Rtw-Mon-3CNF (counting the number of satisfying assignments of a planar read-twice monotone 3CNF formula, modulo 7) is solvable in P. The same problem #Pl-Rtw-Mon-3CNF without mod 7 is known to be #P-complete, a result due to Xia et al. [31]; the problem mod 2, $\#_2$Pl-Rtw-Mon-3CNF, is known to be $\oplus$P-complete (thus NP-hard), a result due to Valiant [30]. The surprising tractability mod 7 is due to the unexpected existence of suitable generators and recognizers over $\mathbf{Z}_7$.

These signatures are specified by families of algebraic equations. These families of equations are typically exponential in size. Searching for their solutions is what Valiant called "the enumeration" of "freak objects" in his "Accidental algorithm" paper [30].[3] Dealing with such algebraic equations can be difficult due to the exponential size. So far the successes have been an expression of *artistic* inspiration.

To sustain a belief in $\mathrm{P} \neq \mathrm{NP}$, we must develop a systematic understanding of the capabilities of holographic algorithms. One might take the view that the problems such as $\#_7$Pl-Rtw-Mon-3CNF that have been solved in this framework are a little contrived. But the point is that when we surveyed potential algorithmic approaches with P vs. NP in mind, these algorithms were not part of the repertoire. Presumably the same "intuition" for $\mathrm{P} \neq \mathrm{NP}$ would have applied equally to $\#_7$Pl-Rtw-Mon-3CNF and to $\#_2$Pl-Rtw-Mon-3CNF. Thus, Valiant suggested in [29], "any proof of $\mathrm{P} \neq \mathrm{NP}$ may need to explain, and not only to imply, the unsolvability" of NP-hard problems using this approach.

While finding "exotic" solutions such as the signature for $\#_7$Pl-Rtw-Mon-3CNF is inspired artistry, the situation with ever more complicated algebraic constraints on such signatures (for other problems) can quickly overwhelm such an artistic approach (as well as a computer search). At any rate, failure to find such solutions to a particular algebraic system yields no proof that such solutions do not exist, and it generally does not give us any insight as to why. We need a more *scientific* understanding. The aim of this paper is to build toward such an understanding.

In this paper we have achieved a complete account for all realizable symmetric signatures. Using this we can show why the modulus 7 happens to be *the* modulus that works for $\#_7$Pl-Rtw-Mon-3CNF. Underlying this is the fact that 7 is $2^3 - 1$, and for any odd prime p, any prime factor q of the Mersenne number $2^p - 1$ has $q \equiv \pm 1 \bmod 8$, and therefore 2 is a quadratic residue in $\mathbf{Z}_q$. Generalizing this, we show that $\#_{2^k-1}$Pl-Rtw-Mon-kCNF is in P for all $k \geqslant 3$ (the problem is trivial for $k \leqslant 2$). Furthermore, no suitable signatures exist for any modulus other than factors of $2^k - 1$ for this problem.

When designing a holographic algorithm for any particular problem, the essential step is to decide whether there is a linear basis for which certain signatures of both generators and recognizers can be simultaneously realized (we give a quick review of terminologies in Section 2. See [29,27,4,5] for more details). Frequently these signatures are symmetric signatures. Our understanding of symmetric signatures has advanced to the point where it is possible to give a polynomial time algorithm to decide the simultaneous realizability problem. If a matchgate has arity n, the signature has size 2^n. However for symmetric signatures we have a compact form, and the running time of the decision algorithm is polynomial in n. With this structural understanding we can give (i) a complete account of all the previous successes of holographic algorithms using symmetric signatures [29,5,30]; (ii) generalizations such as $\#_{2^k-1}$Pl-Rtw-Mon-kCNF and a similar problem for vertex cover, when this is possible; and (iii) a proof when this is not possible. We think this is an important step in our understanding of holographic algorithms, from *art* to *science*.

In order to investigate realizability of signatures, we found it useful to introduce a basis manifold $\mathcal{M}$, which is defined to be the set of all possible bases modulo an equivalence relation. This is a useful language for the discussion of symmetric signatures; it becomes essential for the general signatures. We define the notions of d-admissibility and d-realizability. To be d-admissible is to have a d-dimensional solution subvariety in $\mathcal{M}$, satisfying all the parity requirements. This is a part of the requirements in order to be realizable. To be d-realizable is to have a d-dimensional solution subvariety in $\mathcal{M}$ for all realizability requirements, which include the parity requirements as well as the *useful Grassmann–Plücker identities* [5,28], called the matchgate identities. To have 0-realizability is a necessary condition. But to get holographic algorithms one needs simultaneous realizability of both generators and recognizers. This is accomplished by having a non-empty intersection of the respective subvarieties for the realizability of generators and recognizers. And this tends to be accomplished by having d-realizability (which implies d-admissibility), for $d \geqslant 1$, on at least one side. Therefore it is important to investigate d-realizability and d-admissibility for $d \geqslant 1$. We give a complete characterization of 2-admissibility. We also give some non-trivial 1-admissible families, and 1- or 2-realizable families.

This paper is organized as follows. In Section 2 we give a review of terminologies. In Section 3 we define the basis manifold $\mathcal{M}$ which will be used to express our results throughout. In Section 4 we describe our results on simultaneous realizability of recognizers and generators, culminating in the polynomial time decision procedure. In Section 5 we describe

[3] From [30]: "The objects enumerated are sets of polynomial systems such that the solvability of any one member would give a polynomial time algorithm for a specific problem. ... the situation with the P = NP question is not dissimilar to that of other unresolved enumerative conjectures in mathematics. The possibility that accidental or freak objects in the enumeration exist cannot be discounted, if the objects in the enumeration have not been systematically studied previously."

our results on $\#_{2^k-1}$Pl-Rtw-Mon-kCNF and on vertex cover. Further illustrations of the power of the general machinery are given in Section 6. In Section 7 we go beyond symmetric signatures, and give some general results regarding d-admissibility and d-realizability.

2. Some background

In this section, for the convenience of readers, we review some definitions and results. More details can be found in [27,29,28,5,4,3].

Let $G = (V, E, W)$ and $G' = (V', E', W')$ be weighted undirected planar graphs, where V and V' are vertices, E and E' are edges, and W and W' are edge weights. A *generator matchgate* Γ is a tuple (G, X) where $X \subseteq V$ is a set of external *output* nodes. A *recognizer matchgate* Γ' is a tuple (G', Y) where $Y \subseteq V'$ is a set of external *input* nodes. The external nodes are ordered counter-clock wise on the external face. Γ is called an odd (resp. even) matchgate if it has an odd (resp. even) number of nodes.

Each matchgate is assigned a *signature* tensor. A generator Γ with m output nodes is assigned a contravariant tensor $\mathbf{G} \in V_0^m$ of type $\binom{m}{0}$, where V_0^m is the tensor space spanned by the m-fold tensor products of the standard basis $\mathbf{b} = [\mathbf{b}_0, \mathbf{b}_1] = \left[\binom{1}{0}, \binom{0}{1}\right]$. The tensor $\mathbf{G}$ under the standard basis $\mathbf{b}$ has the form

$$\sum G^{i_1 i_2 \ldots i_m} \mathbf{b}_{i_1} \otimes \mathbf{b}_{i_2} \otimes \cdots \otimes \mathbf{b}_{i_m},$$

where

$$G^{i_1 i_2 \ldots i_m} = \text{PerfMatch}(G - Z),$$

where $\text{PerfMatch}(G - Z) = \sum_M \prod_{(i,j)\in M} w_{ij}$, is a sum over all perfect matchings M in $G - Z$, w_{ij} is the weight of the edge (i, j), and where Z is the subset of the output nodes of Γ having the characteristic sequence $\chi_Z = i_1 i_2 \ldots i_m$. Similarly a recognizer Γ' with m input nodes is assigned a covariant tensor $\mathbf{R} \in V_m^0$ of type $\binom{0}{m}$. This tensor under the standard (dual) basis $\mathbf{b}^*$ has the form

$$\sum R_{i_1 i_2 \ldots i_m} \mathbf{b}^{i_1} \otimes \mathbf{b}^{i_2} \otimes \cdots \otimes \mathbf{b}^{i_m},$$

where

$$R_{i_1 i_2 \ldots i_m} = \text{PerfMatch}\big(G' - Z\big),$$

where Z is the subset of the input nodes of Γ' having the characteristic sequence $\chi_Z = i_1 i_2 \ldots i_m$.

In particular, $\mathbf{G}$ transforms as a contravariant tensor under a basis transformation $\boldsymbol{\beta}_j = \sum_i \mathbf{b}_i t_j^i$,

$$\big(G'\big)^{j_1 j_2 \ldots j_m} = \sum G^{i_1 i_2 \ldots i_m} \tilde{t}_{i_1}^{j_1} \tilde{t}_{i_2}^{j_2} \ldots \tilde{t}_{i_m}^{j_m},$$

where $(\tilde{t}_i^j)$ is the inverse matrix of (t_j^i). Similarly, $\mathbf{R}$ transforms as a covariant tensor, namely

$$\big(R'\big)_{j_1 j_2 \ldots j_m} = \sum R_{i_1 i_2 \ldots i_m} t_{j_1}^{i_1} t_{j_2}^{i_2} \ldots t_{j_m}^{i_m}.$$

A signature is *symmetric* if each entry only depends on the Hamming weight of the index $i_1 i_2 \ldots i_m$. This notion is invariant under a basis transformation. A symmetric signature is denoted by $[\sigma_0, \sigma_1, \ldots, \sigma_m]$, where σ_i denotes the value of a signature entry whose Hamming weight of its index is i.

A *matchgrid* $\Omega = (A, B, C)$ is a weighted planar graph consisting of a disjoint union of: a set of g generators $A = (A_1, \ldots, A_g)$, a set of r recognizers $B = (B_1, \ldots, B_r)$, and a set of f connecting edges $C = (C_1, \ldots, C_f)$, where each C_i edge has weight 1 and joins an output node of a generator with an input node of a recognizer, so that every input and output node in every constituent matchgate has exactly one such incident connecting edge.

Let $\mathbf{G} = \bigotimes_{i=1}^{g} \mathbf{G}(A_i)$ be the tensor product of all the generator signatures, and let $\mathbf{R} = \bigotimes_{j=1}^{r} \mathbf{R}(B_j)$ be the tensor product of all the recognizer signatures. Then $\text{Holant}(\Omega)$ is defined to be the contraction of the two product tensors, under some basis $\boldsymbol{\beta}$, where the corresponding indices match up according to the f connecting edges C_k:

$$\text{Holant}(\Omega) = \langle \mathbf{R}, \mathbf{G} \rangle = \sum_{x \in \boldsymbol{\beta}^{\otimes f}} \left\{ \left[\prod_{1 \leqslant i \leqslant g} \mathbf{G}(A_i, x|_{A_i}) \right] \cdot \left[\prod_{1 \leqslant j \leqslant r} \mathbf{R}\big(B_j, x^*\big|_{B_j}\big) \right] \right\}.$$

(If we write the tensor product for the covariant tensor $\mathbf{R}$ as a row vector of dimension 2^f, and write the contravariant tensor $\mathbf{G}$ as a column vector of dimension 2^f, then $\text{Holant}(\Omega)$ is just the inner product of these two vectors.)

Valiant's beautiful Holant Theorem is

Theorem 2.1 *(Valiant). For any matchgrid Ω over any basis $\boldsymbol{\beta}$, let G be its underlying weighted graph, then*

$$\text{Holant}(\Omega) = \text{PerfMatch}(G).$$

The FKT algorithm can compute the perfect matching polynomial $\text{PerfMatch}(G)$ for a planar graph in polynomial time. This algorithm gives an orientation of the edges of the planar graph, which assigns a ± 1 factor to each edge weight. It then evaluates the Pfaffian of the skew-symmetric matrix of the graph.

Pfaffians satisfy the Grassmann–Plücker identities [24].

Theorem 2.2. *For any $n \times n$ skew-symmetric matrix M, and any $I = \{i_1, \ldots, i_K\} \subseteq [n]$ and $J = \{j_1, \ldots, j_L\} \subseteq [n]$,*

$$\sum_{l=1}^{L} (-1)^l \text{Pf}(j_l, i_1, \ldots, i_K)\text{Pf}(j_1, \ldots, \widehat{j_l}, \ldots, j_L) + \sum_{k=1}^{K} (-1)^k \text{Pf}(i_1, \ldots, \widehat{i_k}, \ldots, i_K)\text{Pf}(i_k, j_1, \ldots, j_L) = 0,$$

where the notation $\hat{i}$ indicates that the entry i is omitted.

A set of the so-called *useful* Grassmann–Plücker identities have been proved to characterize planar matchgate signatures [28,3,5]. These are called matchgate identities.

We state some theorems from [6], which will be used.

Theorem 2.3. *A symmetric signature $[x_0, x_1, \ldots, x_n]$ for a recognizer is realizable under the basis $\boldsymbol{\beta} = [n, p] = \left[\binom{n_0}{n_1}, \binom{p_0}{p_1}\right]$* iff *it takes one of the following forms*:

- *Form* 1: *there exist* (*arbitrary*) *constants λ, s, t and ϵ where $\epsilon = \pm 1$, such that for all i, $0 \leqslant i \leqslant n$,*

$$x_i = \lambda\left[(sn_0 + tn_1)^{n-i}(sp_0 + tp_1)^i + \epsilon(sn_0 - tn_1)^{n-i}(sp_0 - tp_1)^i\right]. \tag{1}$$

- *Form* 2: *there exists an* (*arbitrary*) *constant λ, such that for all i, $0 \leqslant i \leqslant n$,*

$$x_i = \lambda\left[(n-i)n_0(p_1)^i(n_1)^{n-1-i} + ip_0(p_1)^{i-1}(n_1)^{n-i}\right]. \tag{2}$$

- *Form* 3: *there exists an* (*arbitrary*) *constant λ, such that for all i, $0 \leqslant i \leqslant n$,*

$$x_i = \lambda\left[(n-i)n_1(p_0)^i(n_0)^{n-1-i} + ip_1(p_0)^{i-1}(n_0)^{n-i}\right]. \tag{3}$$

We take the convention that $\alpha^0 = 1$ and $0 \cdot \alpha^{0-1} = 0$.

Theorem 2.4. *A symmetric signature $[x_0, x_1, \ldots, x_n]$ for a generator is realizable under the basis $\boldsymbol{\beta} = [n, p] = \left[\binom{n_0}{n_1}, \binom{p_0}{p_1}\right]$ (more precisely in the dual basis)* iff *it takes one of the following forms*:

- *Form* 1: *there exist* (*arbitrary*) *constants λ, s, t and ϵ where $\epsilon = \pm 1$, such that for all i, $0 \leqslant i \leqslant n$,*

$$x_i = \lambda\left[(sp_1 - tp_0)^{n-i}(-sn_1 + tn_0)^i + \epsilon(sp_1 + tp_0)^{n-i}(-sn_1 - tn_0)^i\right]. \tag{4}$$

- *Form* 2: *there exists an* (*arbitrary*) *constant λ, such that for all i, $0 \leqslant i \leqslant n$,*

$$x_i = \lambda\left[(n-i)p_1(n_0)^i(-p_0)^{n-1-i} - in_1(n_0)^{i-1}(-p_0)^{n-i}\right]. \tag{5}$$

- *Form* 3: *there exists an* (*arbitrary*) *constant λ, such that for all i, $0 \leqslant i \leqslant n$,*

$$x_i = \lambda\left[-(n-i)p_0(-n_1)^i(p_1)^{n-1-i} + in_0(-n_1)^{i-1}(p_1)^{n-i}\right]. \tag{6}$$

Theorem 2.5. *A symmetric signature $[x_0, x_1, \ldots, x_n]$ is realizable on some basis iff there exist three constants a, b, c (not all zero), such that for all k, $0 \leqslant k \leqslant n-2$,*

$$ax_k + bx_{k+1} + cx_{k+2} = 0. \tag{7}$$

The following two simple lemmas are used in the proof of Lemmas 4.5 and 4.6.

Lemma 2.1. *Suppose a sequence* $(x_i)_{i=0,1,\ldots,n}$, *where* $n \geqslant 3$, *has the following form*: $x_i = A\alpha^i + B\beta^i$ $(AB \neq 0,\ \alpha \neq \beta)$, *then the representation is unique. That is, if* $x_i = A'(\alpha')^i + B'(\beta')^i$ $(i = 0, 1, \ldots, n,\ n \geqslant 3)$, *then* $A' = A$, $B' = B$, $\alpha' = \alpha$, $\beta' = \beta$ *or* $A' = B$, $B' = A$, $\alpha' = \beta$, $\beta' = \alpha$.

Lemma 2.2. *Suppose a sequence* $(x_i)_{i=0,1,\ldots,n}$, *where* $n \geqslant 3$, *has the following form*: $x_i = Ai\alpha^{i-1} + B\alpha^i$ $(A \neq 0)$, *then the representation is unique. That is, if* $x_i = A'i(\alpha')^{i-1} + B'(\alpha')^i$ $(i = 0, 1, \ldots, n,\ n \geqslant 3)$, *then* $A' = A$, $B' = B$, $\alpha' = \alpha$.

These follow from the fact that second order homogeneous linear recurrence sequence has a unique representation.

3. The basis manifold $\mathcal{M}$

In holographic algorithms, computations are expressed in terms of a set of linear basis vectors of dimension 2^k, where k is called the size of the basis. In almost all cases [29,3], the successful design of a holographic algorithm was accomplished by a basis of size 1. In [30], initially Valiant used a basis of size 2 to show $\#_7$Pl-Rtw-Mon-3CNF $\in$ P. Then it was pointed out in [6] that even in that case the same can be done with a basis of size 1. In [8] and [9], we show that this is generally true, i.e., higher dimensional bases *do not* extend the reach of holographic algorithms. Therefore, in this paper we will develop our theory exclusively with bases of size 1; but our results are universally applicable.

We will identify the set of 2-dimensional bases $\left[\binom{n_0}{n_1}, \binom{p_0}{p_1}\right]$ with $\mathrm{GL}_2(\mathbf{F})$. Over the complex field $\mathbf{F} = \mathbf{C}$, it has dimension 4. However, the following simple proposition (Proposition 4.3 of [29]) shows that the essential underlying structure has only dimension 2.

Proposition 3.1 *(Valiant). (See [29].) If there is a generator* (*recognizer*) *with certain signature for size one basis* $\{(n_0, n_1), (p_0, p_1)\}$ *then there is a generator* (*recognizer*) *with the same signature for size one basis* $\{(xn_0, yn_1), (xp_0, yp_1)\}$ *or* $\{(xn_1, yn_0), (xp_1, yp_0)\}$ *for any* $x, y \in \mathbf{F}$ *and* $xy \neq 0$.

This leads to the following definition of an equivalence relation:

Definition 3.1. Two bases $\boldsymbol{\beta} = [n, p] = \left[\binom{n_0}{n_1}, \binom{p_0}{p_1}\right]$ and $\boldsymbol{\beta'} = [n', p'] = \left[\binom{n'_0}{n'_1}, \binom{p'_0}{p'_1}\right]$ are equivalent, denoted by $\boldsymbol{\beta} \sim \boldsymbol{\beta'}$, iff there exist $x, y \in \mathbf{F}^*$, the non-zero elements in $\mathbf{F}$, such that $n'_0 = xn_0$, $p'_0 = xp_0$, $n'_1 = yn_1$, $p'_1 = yp_1$ or $n'_0 = xn_1$, $p'_0 = xp_1$, $n'_1 = yn_0$, $p'_1 = yp_0$.

In other words, to obtain $\boldsymbol{\beta'}$ from $\boldsymbol{\beta}$, viewed as a 2×2 matrix, we can multiply each row by a non-zero constant, or exchange the two rows.

Theorem 3.1. $\mathrm{GL}_2(\mathbf{F})/\sim$ *is a* 2*-dimensional manifold* (*for* $\mathbf{F} = \mathbf{C}$ *or* $\mathbf{R}$).

We call this the *basis manifold* $\mathcal{M}$. For $\mathbf{F} = \mathbf{R}$, it can be shown that topologically $\mathcal{M}$ is a Möbius strip. From now on we identify a basis $\boldsymbol{\beta}$ with its equivalence class containing it. When it is permissible, we use the dehomogenized coordinates $\binom{1\ x}{1\ y}$ to represent a point (i.e., a basis class) in $\mathcal{M}$. We will assume char.$\mathbf{F} \neq 2$.

4. Simultaneous realizability of symmetric signatures

In [6], we gave a complete characterization of all the realizable symmetric signatures (Theorems 2.3–2.5). These tell us exactly what signatures can be realized over *some* bases. However, to construct a holographic algorithm, one needs to realize some generators and recognizers simultaneously. In terms of $\mathcal{M}$, a given generator (recognizer) defines a (possibly empty) subvariety which consists of all the bases over which it is realizable. The simultaneous realizability is equivalent to a non-empty intersection of these subvarieties. Thus we have to go beyond Theorems 2.3–2.5. For every signature which is realizable according to Theorem 2.5, we need to determine the subvariety where it is realizable.

Definition 4.1. Let $B_{rec}([x_0, x_1, \ldots, x_n])$ (resp. $B_{gen}([x_0, x_1, \ldots, x_n])$) be the set of all possible bases in $\mathcal{M}$ for which a symmetric signature $[x_0, x_1, \ldots, x_n]$ for a recognizer (resp. a generator) is realizable. We also use $B_{rec}(R)$ and $B_{gen}(G)$ to denote the realizability subvarieties for general (unsymmetric) signatures R and G.

Since the identically zero signature is realizable in every basis, we will assume the signature is not identically zero in the following discussion.

4.1. Realizability of recognizers

The following lemmas give a complete and mutually exclusive list of realizable symmetric signatures for recognizers.

Lemma 4.1.

$$B_{rec}\big(\lambda[a^n, a^{n-1}b, \ldots, b^n]\big) = \left\{ \left[\binom{a}{n_1}, \binom{b}{p_1} \right] \in \mathcal{M} \;\middle|\; n_1, p_1 \in \mathbf{F} \right\}.$$

Remark. Every signature with arity 1 is trivially of this form. We will omit the scalar factor λ below as it is trivial. Since we will exclude the identically 0 signature, a and b are not both 0.

Proof of Lemma 4.1. If $n = 1$, the standard signature can and can only be $(\lambda, 0)$ or $(0, \lambda)$ (where λ is arbitrary). One entry of the signature must be zero due to the parity requirement, as matchgates are defined in terms of perfect matchings. So the signature over the basis $\left[\binom{n_0}{n_1}, \binom{p_0}{p_1}\right]$ is $(\lambda n_0, \lambda p_0)$ or $(\lambda n_1, \lambda p_1)$. Since we require the signature to be (a, b), all possible bases as expressed in $\mathcal{M}$ are $\left[\binom{a}{n_1}, \binom{b}{p_1}\right]$, taking into account the equivalence relation $\sim$, where n_1, p_1 are arbitrary, except $ap_1 - bn_1 \neq 0$.

Now we assume $n > 1$. First suppose this signature is expressed as Form 1 of Theorem 2.3.

In Form 1, denote by $u_0 = sn_0 + tn_1$, $u_1 = sp_0 + tp_1$, $v_0 = sn_0 - tn_1$, and $v_1 = sp_0 - tp_1$. Then up to a constant factor λ, for each $0 \leqslant i \leqslant n$, we have $u_0^{n-i}u_1^i + \epsilon v_0^{n-i}v_1^i = a^{n-i}b^i$.

We first assume $u_0v_0 \neq 0$. Then by multiplying the equations for $i = 0$ and $i = 2$, and multiplying the equation for $i = 1$ by itself, we get an equation on u_0, u_1, v_0 and v_1. After some simplifications we get $u_1/u_0 = v_1/v_0$. Denote this common ratio by ρ.

We claim in this case $a \neq 0$, and $\rho = b/a$. Assume for a contradiction that $a = 0$, then all entries of the signature are 0 for $i = 0, \ldots, n-1$. However the entry at $i = n$ is obtained from the entry at $i = n-1$ by multiplying with the ratio ρ, and thus it is also 0. Then it follows that $b = 0$ as well, contrary to assumption. Therefore $a \neq 0$ and b/a is defined. Now consider the signature entry at $i = 0$ and $i = 1$. The entry at $i = 0$ is $a^n \neq 0$, and the entry at $i = 1$ is obtained by multiplying the non-zero entry at $i = 0$ by the ratio b/a, as well as by the common ratio ρ. It follows that $\rho = b/a$.

Hence $bu_0 = au_1$ and $bv_0 = av_1$. Then by the definitions of u_i and v_j, it follows that $bsn_0 = asp_0$ and $btn_1 = atp_1$. Because at least one of a, b is non-zero, we claim that this implies either $s = 0$ or $t = 0$. Otherwise, $st \neq 0$, we have $n_0p_1 - n_1p_0 = 0$. This is impossible. So we must have $s = 0$ or $t = 0$ (and not both zero since otherwise the signature is identically zero). Now in either cases, it is easy to verify that all the possible bases are $\left[\binom{a}{n_1}, \binom{b}{p_1}\right] \in \mathcal{M}$, taking into account the equivalence relation $\sim$, where n_1, p_1 are arbitrary, except $ap_1 - bn_1 \neq 0$.

The same conclusion holds if we assume $u_1v_1 \neq 0$. To complete the proof, assume both $u_0v_0 = 0$ and $u_1v_1 = 0$. By symmetry, suppose $u_0 = 0$ (the other cases are symmetric). In this case if $u_1 = 0$ as well, then $s = t = 0$ since the determinant $n_0p_1 - n_1p_0 \neq 0$. Then $v_0 = v_1 = 0$ and the signature is identically zero. Hence $u_1 \neq 0$. Then $v_1 = 0$. It follows that the signature has the form $\lambda[\epsilon v_0^n, 0, \ldots, 0, u_1^n]$, where there are a non-empty segment of zeros corresponding to $0 < i < n$. These are of the form $a^{n-i}b^i$, and thus $ab = 0$. But then the signature entry is zero at either $i = 0$ or at $i = n$. Since $u_1 \neq 0$, we get $v_0 = 0$. The statement of the lemma clearly holds when $u_0 = v_0 = 0$.

Now suppose the signature is expressed as Form 2 of Theorem 2.3. (The case with Form 3 is symmetric, exchanging subscript 0 for 1 in the basis.)

In that expression, if $n_1 = 0$, then $a = 0$ since $x_0 = a^n = nn_0n_1^{n-1} = 0$. At $i = n-1$, $x_{n-1} = n_0p_1^{n-1} = a^{n-1}b = 0$. This gives $n_0 = 0$ or $p_1 = 0$, together with $n_1 = 0$, we get a singular basis.

So we have $n_1 \neq 0$. Then we claim $a \neq 0$. Otherwise at $i = 0$, $x_0 = nn_0n_1^{n-1} = a^n = 0$, which implies that $n_0 = 0$. At $i = 1$, $x_1 = p_0n_1^{n-1} = a^{n-1}b = 0$, we get $p_0 = 0$. This gives a singular basis. So $a \neq 0$, and from the above we also get $n_0 \neq 0$. Then up to a scalar factor $a^n = 1$, $a^{n-1}b = c + \rho$, and $a^{n-2}b^2 = 2c\rho + \rho^2$, for $c = (n_1p_0 - n_0p_1)n_1/nn_0$ and $\rho = p_1/n_1$. It follows that $2c\rho + \rho^2 = (c + \rho)^2$, which implies that $c = 0$. As the determinant $n_1p_0 - n_0p_1 \neq 0$, and $n_1 \neq 0$, we get a contradiction.

This completes the proof. □

Definition 4.2. A symmetric signature $[x_0, x_1, \ldots, x_n]$, where $n \geqslant 2$, is called non-degenerate iff $\operatorname{rank}\left[\begin{smallmatrix} x_0 & \cdots & x_{n-1} \\ x_1 & \cdots & x_n \end{smallmatrix}\right] = 2$. Otherwise it is degenerate.

The signature is identically 0 iff $\operatorname{rank}\left[\begin{smallmatrix} x_0 & \cdots & x_{n-1} \\ x_1 & \cdots & x_n \end{smallmatrix}\right] = 0$. It has rank 1 iff it can be expressed as $\lambda[a^n, a^{n-1}b, \ldots, b^n]$, for $\lambda \neq 0$, and a, b not both 0. In the following we assume the signature is non-degenerate. We directly handle the case for arity $n = 2$ next.

Lemma 4.2.

$$B_{rec}\big([x_0, x_1, x_2]\big) = \left\{ \left[\binom{n_0}{n_1}, \binom{p_0}{p_1} \right] \in \mathcal{M} \;\middle|\; \begin{array}{c} x_0p_1^2 - 2x_1p_1n_1 + x_2n_1^2 = 0,\ x_0p_0^2 - 2x_1p_0n_0 + x_2n_0^2 = 0 \\ \text{or } x_0p_0p_1 - x_1(n_0p_1 + n_1p_0) + x_2n_0n_1 = 0 \end{array} \right\}.$$

Proof. Under the equivalence relation, we can assume $n_0p_1 - n_1p_0 = 1$.

Then $\left[\binom{n_0}{n_1}, \binom{p_0}{p_1}\right]^{-1} = \left[\binom{p_1}{-n_1}, \binom{-p_0}{n_0}\right]$. So the standard signature of $[x_0, x_1, x_2]$ is

$$\left[x_0p_1^2 - 2x_1p_1n_1 + x_2n_1^2, x_0p_0p_1 - x_1(n_0p_1 + n_1p_0) + x_2n_0n_1, x_0p_0^2 - 2x_1p_0n_0 + x_2n_0^2\right].$$

The fact that the only constraint of a standard signature of arity 2 is the parity constraint completes the proof. □

In the following we assume the signature has arity $n \geqslant 3$, and non-degenerate. In this case, we note that the constants a, b, c in Theorem 2.5 are unique up to a scalar factor. In fact if there are two linearly independent triples (a, b, c), then the following matrix

$$\begin{bmatrix} x_0 & x_1 & \dots & x_{n-2} \\ x_1 & x_2 & \dots & x_{n-1} \\ x_2 & x_3 & \dots & x_n \end{bmatrix}$$

has rank $\leqslant 1$. The first row and the last row are not both zero, otherwise the signature is identically zero (by $n \geqslant 3$). It follows that the matrix

$$\begin{bmatrix} x_0 & x_1 & \dots & x_{n-1} \\ x_1 & x_2 & \dots & x_n \end{bmatrix}$$

also has rank 1, hence the signature is degenerate.

Lemma 4.3. *Let $\lambda_1 \neq 0$. Let $p = \mathrm{char}.\mathbf{F}$. Suppose $p = 0$, or $p \nmid n$,*

$$B_{rec}([0, 0, \dots, 0, \lambda_1, \lambda_2]) = \left\{\left[\begin{pmatrix} 0 \\ n\lambda_1 \end{pmatrix}, \begin{pmatrix} 1 \\ \lambda_2 \end{pmatrix}\right]\right\}.$$

For $p \mid n$ and $\lambda_2 = 0$, $B_{rec}([0, 0, \dots, 0, \lambda_1, 0]) = \left\{\left[\binom{0}{n_1}, \binom{1}{p_1}\right] \in \mathcal{M} \mid n_1, p_1 \in \mathbf{F}\right\}$. For $p \mid n$ and $\lambda_2 \neq 0$, the signature $[0, 0, \dots, 0, \lambda_1, \lambda_2]$ is not realizable.

Proof. Its reversal signature $[\lambda_2, \lambda_1, 0, \dots, 0]$ is a special case of Lemma 4.6 (with $\alpha = 0$). □

Lemma 4.4. *For $AB \neq 0$,*

$$B_{rec}([A, A\alpha, A\alpha^2, \dots, A\alpha^n + B]) = \left\{\left[\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} \alpha + \omega \\ \alpha - \omega \end{pmatrix}\right] \,\middle|\, \omega^n = \pm\frac{B}{A}\right\}.$$

Proof. Its reversal signature $[A\alpha^n + B, A\alpha^{n-1}, \dots, A\alpha, A]$ is a spacial case of Lemma 4.5. (This proof assumes $\alpha \neq 0$. For $\alpha = 0$, it can be directly verified.) □

In the following we use the fact that the triple (a, b, c) in the statement of Theorem 2.5 is unique up to a scalar factor. Also in the remaining cases we may assume $c \neq 0$. So we have a unique characteristic equation $cx^2 + bx + a = 0$, which has two roots α and β. In particular Forms 1, 2 and 3 from Theorem 2.3 are mutually exclusive. If $\alpha \neq \beta$, we have the following lemma:

Lemma 4.5. *For $AB \neq 0$ and $\alpha \neq \beta$,*

$$B_{rec}([A\alpha^i + B\beta^i \mid i = 0, 1, \dots, n]) = \left\{\left[\begin{pmatrix} 1 + \omega \\ 1 - \omega \end{pmatrix}, \begin{pmatrix} \alpha + \beta\omega \\ \alpha - \beta\omega \end{pmatrix}\right] \,\middle|\, \omega^n = \pm\frac{B}{A}\right\}.$$

Remark. We denote $0^0 = 1$.

Proof of Lemma 4.5. From $A + B = x_0$, $A\alpha + B\beta = x_1$, we can solve uniquely for A, B. We have $AB \neq 0$; otherwise $\{x_i\}$ has the form $\{a^{n-i}b^i\}$, which has been dealt with in Lemma 4.1. Having two distinct eigenvalues $\alpha \neq \beta$, this signature must be expressed as Form 1 of Theorem 2.3. Let $u_0 = sn_0 + tn_1$, $u_1 = sp_0 + tp_1$, $v_0 = sn_0 - tn_1$, and $v_1 = sp_0 - tp_1$. Then $A\alpha^i + B\beta^i = u_0^{n-i}u_1^i + \epsilon v_0^{n-i}v_1^i$.

We claim $u_0 \neq 0$. Otherwise, for $i = 0, 1, \dots, n-1$, the signature entry at i is $\epsilon v_0^{n-i}v_1^i$. It follows that $(A + B)(A\alpha^2 + B\beta^2) = (A\alpha + B\beta)^2$, and since $AB \neq 0$, we get $\alpha = \beta$, a contradiction. Similarly we have $v_0 \neq 0$.

Hence we have two expressions

$$x_i = A\alpha^i + B\beta^i = u_0^n\left(\frac{u_1}{u_0}\right)^i + \epsilon v_0^n\left(\frac{v_1}{v_0}\right)^i.$$

From Lemma 2.1, we know that the representation is unique. So $u_0^n = A$, $\epsilon v_0^n = B$, $\frac{u_1}{u_0} = \alpha$ and $\frac{v_1}{v_0} = \beta$ (exchanging notations A with B, and α with β if necessary). It follows that $\left[\binom{2sn_0}{2tn_1}, \binom{2sp_0}{2tp_1}\right] = \left[\binom{u_0+v_0}{u_0-v_0}, \binom{u_1+v_1}{u_1-v_1}\right]$. Since $\alpha \neq \beta$, we can show $st \neq 0$, by the same proof showing $u_0 \neq 0$ and $v_0 \neq 0$. Now let $\omega = v_0/u_0$, then $\omega^n = \pm B/A$, and

$$\left[\binom{n_0}{n_1}, \binom{p_0}{p_1}\right] \sim \left[\binom{2sn_0}{2tn_1}, \binom{2sp_0}{2tp_1}\right] \sim \left[\binom{1+\omega}{1-\omega}, \binom{\alpha+\beta\omega}{\alpha-\beta\omega}\right].$$

This completes the proof. □

If the characteristic roots $\alpha = \beta$, we have the following lemma:

Lemma 4.6. *Let $p = \mathrm{char}.\mathbf{F}$ and let $A \neq 0$.*

Case 1: *$p = 0$ or $p \nmid n$.*

$$B_{rec}\left(\left[Ai\alpha^{i-1} + B\alpha^i \mid i = 0, 1, \ldots, n\right]\right) = \left\{\left[\binom{1}{B}, \binom{\alpha}{nA + B\alpha}\right]\right\}.$$

Case 2: *$p \mid n$ and $x_0 = 0$. In this case, the signature has entries $x_i = Ai\alpha^{i-1}$, with $B = 0$ in the above form.*

$$B_{rec}\left(\left[Ai\alpha^{i-1} \mid i = 0, 1, \ldots, n\right]\right) = \left\{\left[\binom{1}{n_1}, \binom{\alpha}{p_1}\right] \in \mathcal{M} \;\middle|\; n_1, p_1 \in \mathbf{F}\right\}.$$

Case 3: *$p \mid n$ and $x_0 \neq 0$. In this case the signature $[Ai\alpha^{i-1} + B\alpha^i \mid i = 0, 1, \ldots, n]$ is not realizable.*

Remark. If $\alpha = 0$, and $i = 0$, we take the convention that $i\alpha^{i-1} = 0$, and also $\alpha^i = 1$.

Proof of Lemma 4.6. In Case 1, from $B = x_0$, $A + B\alpha = x_1$, we can solve uniquely for A, B. We have $A \neq 0$, so Lemma 2.2 applies. From Lemma 2.2, we know that the representation is unique. From Form 2 of Theorem 2.3 we claim $n_1 \neq 0$. Otherwise, all signature entries $x_i = 0$ for $i = 0, \ldots, n-2$. Since $n \geqslant 3$, we have $x_0 = x_1 = 0$, which implies that $A = 0$, contrary to assumption. In the following we assume Form 2 of Theorem 2.3, Form 3 will give an equivalent basis. Then we have $x_i = i(n_1p_0 - n_0p_1)n_1^n(\frac{p_1}{n_1})^{i-1} + nn_0n_1^{n-1}(\frac{p_1}{n_1})^i$. So by uniqueness $(n_1p_0 - n_0p_1)n_1^n = A$, $\frac{p_1}{n_1} = \alpha$, $nn_0n_1^{n-1} = B$. Since $n_1 \neq 0$, under the equivalence relation, we can let $n_1 = 1$, then we have the unique solution $n_0 = B/n$, $p_1 = \alpha$, $p_0 = A + \frac{B\alpha}{n}$. We omit the proofs for Cases 2 and 3. □

The above list of realizable symmetric signatures for recognizers is complete and mutually exclusive. To see that, by Theorem 2.5, we have a recurrence relation for any realizable signature. The case for any degenerate signature, including the case $n = 1$, is handled in Lemma 4.1. Now assume the signature is non-degenerate. The case $n = 2$ is handled in Lemma 4.2. Next we assume the signature is non-degenerate and arity $n \geqslant 3$. Then Theorem 2.5 provides a tuple $(a, b, c) \neq 0$, unique up to a non-zero constant multiple. If $c \neq 0$ this defines a unique second order recurrence relation. If $a \neq 0$ this defines a unique second order recurrence relation for the reversal. (If both $a = 0$ and $c = 0$, this defines the signature $[A, 0, \ldots, 0, B]$ where $AB \neq 0$, due to non-degeneracy. This is included in Lemma 4.4, with $\alpha = 0$.) Assume $c \neq 0$ then the recurrence relation is second order and has eigenvalues α and β. Depending on whether it has a pair of distinct eigenvalues or a double eigenvalue, we have Lemmas 4.5 and 4.6. The case when the recurrence relation is for the reversal signature results in the same expression, except in the case when one of the eigenvalue is 0. And these special cases are handled in Lemmas 4.4 and 4.3 respectively.

4.2. Realizability of generators

The following lemmas give a complete and mutually exclusive list of realizable symmetric signatures for generators. They can be proved similarly.

Lemma 4.7.

$$B_{gen}\left(\lambda\left[a^n, a^{n-1}b, \ldots, b^n\right]\right) = \left\{\left[\binom{n_0}{-b}, \binom{p_0}{a}\right] \;\middle|\; n_0, p_0 \in \mathbf{F}\right\}.$$

Lemma 4.8.

$$B_{gen}\left([x_0, x_1, x_2]\right) = \left\{\left[\binom{n_0}{n_1}, \binom{p_0}{p_1}\right] \in \mathcal{M} \;\middle|\; \begin{array}{c} x_0n_0^2 + 2x_1n_0p_0 + x_2p_0^2 = 0, x_0n_1^2 + 2x_1n_1p_1 + x_2p_1^2 = 0 \\ \text{or } x_0n_0n_1 + x_1(n_0p_1 + n_1p_0) + x_2p_0p_1 = 0 \end{array}\right\}.$$

Lemma 4.9. *Let $\lambda_1 \neq 0$. Let $p = \mathrm{char}.\mathbf{F}$. Suppose $p = 0$, or $p \nmid n$,*

$$B_{gen}([0,0,\ldots,0,\lambda_1,\lambda_2]) = \left\{\left[\begin{pmatrix} -\lambda_2 \\ 1 \end{pmatrix}, \begin{pmatrix} n\lambda_1 \\ 0 \end{pmatrix}\right]\right\}.$$

For $p \mid n$ and $\lambda_2 = 0$, $B_{gen}([0,0,\ldots,0,\lambda_1,0]) = \left\{\left[\binom{1}{n_1}, \binom{0}{p_1}\right] \in \mathcal{M} \mid n_1, p_1 \in \mathbf{F}\right\}$. For $p \mid n$ and $\lambda_2 \neq 0$, then $[0,0,\ldots,0,\lambda_1,\lambda_2]$ is not realizable.

Lemma 4.10. *For $AB \neq 0$,*

$$B_{gen}([A, A\alpha, A\alpha^2, \ldots, A\alpha^n + B]) = \left\{\left[\begin{pmatrix} \omega - \alpha \\ -\alpha - \omega \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix}\right] \,\middle|\, \omega^n = \pm\frac{B}{A}\right\}.$$

Lemma 4.11. *For $AB \neq 0$ and $\alpha \neq \beta$,*

$$B_{gen}(\{A\alpha^i + B\beta^i \mid i = 0,1,\ldots,n\}) = \left\{\left[\begin{pmatrix} \beta\omega - \alpha \\ -\alpha - \beta\omega \end{pmatrix}, \begin{pmatrix} 1-\omega \\ 1+\omega \end{pmatrix}\right] \,\middle|\, \omega^n = \pm\frac{B}{A}\right\}.$$

Lemma 4.12. *Let $p = \mathrm{char}.\mathbf{F}$ and let $A \neq 0$.*

Case 1: *$p = 0$ or $p \nmid n$.*

$$B_{gen}(\{Ai\alpha^{i-1} + B\alpha^i \mid i = 0,1,\ldots,n\}) = \left\{\left[\begin{pmatrix} nA + B\alpha \\ -\alpha \end{pmatrix}, \begin{pmatrix} -B \\ 1 \end{pmatrix}\right]\right\}.$$

Case 2: *$p \mid n$ and $x_0 = 0$. In this case, the signature has entries $x_i = Ai\alpha^{i-1}$, with $B = 0$ in the above form*:

$$B_{gen}([Ai\alpha^{i-1} \mid i = 0,1,\ldots,n]) = \left\{\left[\begin{pmatrix} -\alpha \\ n_1 \end{pmatrix}, \begin{pmatrix} 1 \\ p_1 \end{pmatrix}\right] \in \mathcal{M} \,\middle|\, n_1, p_1 \in \mathbf{F}\right\}.$$

Case 3: *$p \mid n$ and $x_0 \neq 0$. In this case the signature $[Ai\alpha^{i-1} + B\alpha^i \mid i = 0,1,\ldots,n]$ is not realizable.*

4.3. Simultaneous realizability

Definition 4.3. The Simultaneous Realizability Problem (SRP):
Input: A set of symmetric signatures for generators and/or recognizers.
Output: A common basis of these signatures if any exists; "NO" if they are not simultaneously realizable.

Algorithm.

For every signature $[x_0, x_1, \ldots x_n]$, check if it satisfies Theorem 2.5.
If not, output "NO" and halt.
Otherwise find $B_{gen}([x_0, x_1, \ldots x_n])$ or $B_{rec}([x_0, x_1, \ldots x_n])$ according to one of the lemmas.
Check if these subvarieties have a non-empty intersection.

Theorem 4.1. *This is a polynomial time algorithm for SRP. (If $p = \mathrm{char}.\mathbf{F}$ is a large prime and is considered part of the input, i.e., input size includes $\log p$, then the problem is in RP.)*

Proof. Checking whether every input signature satisfies Theorem 2.5 can obviously be done in polynomial time. To find the right form and then the right lemma for a signature which satisfies Theorem 2.5 can also be done in polynomial time as they are mutually exclusive.

Every subvariety of bases from Lemmas 4.1 to 4.6 and from Lemmas 4.7 to 4.12 is of one of three kinds: a finite set of points (of linear size), a line or a quadratic curve. More precisely, consider recognizers; the situation for generators is similar. Expressing things in terms of the manifold $\mathcal{M}$ shows that: For Lemma 4.1 we get a line with $x = \mathrm{const.}$ (in the notation defining $\mathcal{M}$). For Lemma 4.2 we get a union of two sets. The first is finite, where both x and y satisfy a quadratic polynomial (and by projective closure). Therefore there are at most 4 points in $\mathcal{M}$. The second set is defined by an equation of the form $Axy + B(x + y) + C = 0$ (and by projective closure), where A, B, C are known constants. Note that if we had two sets of this type (from Lemma 4.2 and/or Lemma 4.8) we can eliminate A and get a linear equation. (Solving quadratic equations over large finite field may require randomized polynomial time.)

For Lemma 4.3 we have either a single point for $p \mid n$ or a line "at infinity". Lemma 4.6 is similar, where we have either a point or a line $x = \mathrm{const.}$ For Lemma 4.4, we get at most n points from the equation $\omega^n = \mathrm{const.}$ If we are in **C** (more

precisely in **Q** or an algebraic extension field of **Q**) then the computation is clearly in P. For fields of finite characteristic, since n is given in unary, the computation is in P, provided p is fixed (or at most $O(\log n)$). For large p (the field size is exponential in n), this can be done in RP (i.e., in randomized polynomial time). We need to be able to solve equations such as $X^n = \text{const}$. These can be done in randomized polynomial time; see [1] for more details. □

5. Some not-so-accidental algorithms

In [30], Valiant gave polynomial time algorithms for $\#_7$Pl-Rtw-Mon-3CNF and $\#_7$Pl-3/2Bip-VC, and he called them "accidental algorithms". In this section, we show how such algorithms can be developed almost "mechanically". This approach has the advantage that one gains more understanding of what can or cannot be accomplished. With this machinery we are able to generalize his result to Pl-Rtw-Mon-kCNF and Pl-k/2Bip-VC, for a general k. We show that there is a unique modulus $2^k - 1$ for which we can design such a holographic algorithm which counts the number of solutions. In the case of $k = 3$, this shows why 7 is special.

5.1. $\#_{2^k-1}$Pl-Rtw-Mon-kCNF

For #Pl-Rtw-Mon-kCNF, we are given a planar formula [16] in kCNF form, where each variable appears positively, and each appears in exactly 2 clauses. The problem is to count the number of satisfying assignments. As noted earlier, this counting problem is #P-complete already for $k = 3$.

To solve the problem by a holographic algorithm, we wish to replace each variable by a generator with the signature $[1, 0, 1]$, and each clause by a recognizer with the signature $[0, 1, 1, \ldots, 1]$ (with k 1's). The symmetric signature $[1, 0, 1]$ corresponds to a consistent truth assignment on two edges leading to clauses (i.e. the equality function $=_2$ on two Boolean inputs), and $[0, 1, 1, \ldots, 1]$ corresponds to a Boolean OR function for the clause. If we connect the generators and recognizers in a natural way, by the *Holant Theorem* [29] this would solve #Pl-Rtw-Mon-kCNF in polynomial time (if the signatures are realizable over **Q**).

Then the question boils down to whether there is a basis in $\mathcal{M}$ where $[1, 0, 1]$ for a generator and $[0, 1, 1, \ldots, 1]$ (with k 1's) for a recognizer can be simultaneously realized. For this, we use our machinery.

From Lemma 4.5, with $A = 1$, $B = -1$, $\alpha = 1$, $\beta = 0$, we have

$$B_{rec}\big([0, 1, 1, \ldots, 1]\big) = \left\{ \left[\begin{pmatrix} 1+\omega \\ 1-\omega \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right] \,\middle|\, \omega^k = \pm 1 \right\}.$$

We look for some $\omega^k = \pm 1$, such that $\left[\binom{1+\omega}{1-\omega}, \binom{1}{1}\right] \in B_{gen}([1, 0, 1])$.

According to Lemma 4.8, we want $(1+\omega)^2 + 1 = (1-\omega)^2 + 1 = 0$ or $(1+\omega)(1-\omega) + 1 = 0$.

The first case is impossible, and in the second case we require $\omega^2 = 2$. Together with the condition $\omega^k = \pm 1$, we have $2^k - 1 = 0$. From this we can already see that for every prime $p \mid 2^k - 1$, $\#_p$Pl-Rtw-Mon-kCNF is computable in polynomial time. In particular this is true for every Mersenne prime $2^q - 1$. (Note that $\omega^2 = 2$ means that 2 is a quadratic residue.) More generally we have:

Theorem 5.1. *There is a polynomial time algorithm for $\#_{2^k-1}$Pl-Rtw-Mon-kCNF. Furthermore, any modulus m for which the appropriate signatures exist must be a divisor of $2^k - 1$.*

Proof. Our discussion above already shows that the modulus $2^k - 1$ is the best we can do. (Formally speaking we should present a generalization of the Holant Theorem [29] over a ring such as $\mathbf{Z}_{2^k-1}$, which we will omit here.) We now give the polynomial algorithms in two cases:

Case 1. k is even.

Over the complex numbers **C**, from Lemmas 4.8 and 4.4, we can see that a generator for $[1, 0, 1]$ and a recognizer for $[1 + \epsilon 2^{k/2}, 1, 1, \ldots, 1]$ (where there are k 1's, and $\epsilon = \pm 1$) are simultaneously realizable in the basis $\boldsymbol{\beta} = \left[\binom{1+\sqrt{2}}{1-\sqrt{2}}, \binom{1}{1}\right]$.

Setting $\epsilon = 1$ and replacing each variable by a generator and each clause by a recognizer with the corresponding signatures, we obtain a matchgrid Ω with the underlying weighted planar graph G. Then the Holant Theorem [29] tells us

$$\text{Holant}(\Omega) = \text{PerfMatch}(G). \tag{8}$$

We will denote this value by X.

From the left-hand side of (8) we know that X is an integer because every entry in the signatures of generators and recognizers is an integer. Furthermore we have

$$X \equiv \#\text{Pl-Rtw-Mon-}k\text{CNF} \pmod{1 + 2^{k/2}}.$$

From the right-hand side of (8) we know that X can be computed in polynomial time using the FKT algorithm for perfect matchings of a planar graph. The planar graph has weights from the subfield $\mathbf{Q}(\sqrt{2}) \subset \mathbf{C}$, which poses no problem to the Pfaffian evaluation of FKT in polynomial time.

Therefore $\#_{2^{k/2}+1}$Pl-Rtw-Mon-kCNF can be computed in polynomial time. Similarly, setting $\epsilon = -1$, we can compute $\#_{2^{k/2}-1}$Pl-Rtw-Mon-kCNF in polynomial time.

Since $(2^{k/2}+1, 2^{k/2}-1) = 1$ and $2^k - 1 = (2^{k/2}+1)(2^{k/2}-1)$, we can apply Chinese remaindering to get a polynomial time algorithm for $\#_{2^k-1}$Pl-Rtw-Mon-kCNF.

Case 2. k is odd.

Consider the ring $\mathbf{Z}_{2^k-1}$. (Formally we could develop the theory over such a ring, and consider invertible elements and matrices for the basis manifold. But we will omit this formality; everything we need can be easily done by a slight modification of the proofs given before.) Let $r = 2^{(k+1)/2} \in \mathbf{Z}_{2^k-1}$. Then r satisfies $r^2 = 2$ in $\mathbf{Z}_{2^k-1}$. We denote this r by $\sqrt{2}$. Then $1 - (\sqrt{2})^k = 1 - (2^k)^{(k+1)/2} = 0$ in $\mathbf{Z}_{2^k-1}$.

Therefore over this ring $\mathbf{Z}_{2^k-1}$ and with the basis $\boldsymbol{\beta} = \left[\binom{1+\sqrt{2}}{1-\sqrt{2}}, \binom{1}{1}\right] = \left[\binom{1+2^{(k+1)/2}}{1-2^{(k+1)/2}}, \binom{1}{1}\right]$, we have a generator for $[1, 0, 1]$ and a recognizer for $[0, 1, 1, \ldots, 1]$ (with k 1's) according to Lemmas 4.8 and 4.4. As a result, we have a polynomial time algorithm for $\#_{2^k-1}$Pl-Rtw-Mon-kCNF. (It is in this case where k is odd, we need 2 as a quadratic residue in $\mathbf{Z}_p$ for primes $p \mid 2^k - 1$, as discussed in Section 1.) □

5.2. $\#_{2^k-1}$Pl-k/2Bip-VC

In this problem, we are given a planar bipartite graph with left degree k and right degree 2. These are called regular $(k, 2)$-bipartite graphs. We wish to count the number of vertex covers mod $2^k - 1$. The counting problem for this class of graphs mod 2 is $\oplus$P-complete and thus NP-hard [30]. Consider an arbitrary subset S of vertices from the right. Every vertex v on the left *either* has all its k adjacent vertices in S, in which case there are exactly two choices to extend at v to a vertex cover, *or* has some of its k adjacent vertices not in S, in which case there is exactly one choice to extend at v to a vertex cover. Thus, following the general recipe for holographic algorithms, we want to construct a generator with signature $[1, 0, 1]$ and a recognizer with signature $[2, 1, 1, \ldots, 1]$ (with k 1's), to be simultaneously realized over some basis.

From Lemma 4.5, where $A = 1$, $B = 1$, $\alpha = 1$, $\beta = 0$, we have:

$$B_{rec}\big([2, 1, 1, \ldots, 1]\big) = \left\{ \left[\begin{pmatrix} 1+\omega \\ 1-\omega \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right] \,\middle|\, \omega^k = \pm 1 \right\}.$$

We realize that this set is exactly the same as $B_{rec}([0, 1, 1, \ldots, 1])$. Then the proof in Section 5.1 gives us:

Theorem 5.2. *There is a polynomial time algorithm for $\#_{2^k-1}$Pl-k/2Bip-VC. Furthermore, any modulus m for which the appropriate signatures exist must be a divisor of $2^k - 1$.*

Our general machinery not only can find the required signatures when they exist, but also can prove certain desired signatures do not exist or cannot be simultaneously realized. As an example, one may wish to extend the previous two problems to allow more than Read-twice as in #Pl-R_l-Mon-kCNF, where $l > 2$. This calls for a simultaneous realizability of $[1, 0, 0, \ldots, 0, 1]$ (where there are $(l-1)$ 0's) and $[0, 1, 1, \ldots, 1]$ (where there are k 1's). This can be shown to result in an empty intersection on $\mathcal{M}$.

5.3. An edge-vertex cover problem

Another way to think of a regular $(k, 2)$-bipartite graph is to identify every degree 2 vertex on the right together with its two incident edges as a new edge. Then we obtain precisely the class of k-regular graphs. We say a subset of edges *and* vertices is an *edge-vertex cover* if every vertex is either in the subset or all of its k incident edges are in the subset. We consider the following edge-vertex cover problem $\#_{2^k-1}$Pl-k-Reg-EVC: Given a planar k-regular graph G, count the number of edge-vertex covers of G mod $2^k - 1$.

It is clear that this problem is really the same problem as the one in Section 5.2 and thus the same algorithm also gives a polynomial time algorithm for this problem.

Theorem 5.3. *There is a polynomial time algorithm for $\#_{2^k-1}$Pl-k-Reg-EVC. Furthermore, any modulus m for which the appropriate signatures exist must be a divisor of $2^k - 1$.*

5.4. A problem from neural networks

Consider the following planar two-level neural network N: The input nodes are Boolean variables $x_1, \ldots, x_n$. Each x_i has fan-out 2. The intermediate level nodes v all have fan-in k from the x_i's. The output of v feeds into the top node and can have $c+1$ different values $0, 1, \ldots, c$. If all k inputs of v are 0 then the output of v is 0 (unexcited state). Otherwise, the output of v can be any of the $c+1$ values (excited state). The problem is to count the total number of output (firing) patterns as received at the top node. (In the following, for simplicity we state the result for an odd c. We have a parallel set of results for c even, but the statement has some number theoretic complications.)

$\#_{2^k-c^2}$NN$k/c$-Firing-Pattern.
Input: A two-level neural network with parameters k and c as above.
Output: The number mod $(2^k - c^2)$ of all possible firing patterns.

First we suppose k is even. Then we do it over $\mathbf{C}$ by taking $\omega = \sqrt{2}$. We can use the same basis in Section 5.1 to realize the signature $[1 + 2^{k/2}, 1, 1, \ldots, 1]$ (with k 1's) for a recognizer and the signature $[1, 0, 1]$ for a generator simultaneously. This is verified by $\omega^2 = 2$ and $\omega^k = 2^{k/2}$.

Let X be the value of the Holant. With mod $2^{k/2} - c$, the recognizer signature is the same as $[1 + c, 1, 1, \ldots, 1]$. Thus

$$X \equiv \#\text{NN}k/c\text{-Firing-Pattern} \pmod{2^{k/2} - c}.$$

Similarly we can also achieve the signature $[1 - 2^{k/2}, 1, 1, \ldots, 1]$ (with k 1's) for a recognizer and the signature $[1, 0, 1]$ for a generator simultaneously. This is verified by $\omega^2 = 2$ and $\omega^k = -(-2^{k/2})$. This recognizer signature is congruent to $[1 + c, 1, 1, \ldots, 1]$ mod $2^{k/2} + c$. Thus we can compute in polynomial time some value X' for a Holant, where

$$X' \equiv \#\text{NN}k/c\text{-Firing-Pattern} \pmod{2^{k/2} + c}.$$

Then by Chinese remaindering, we can compute the value #NN$k/c$-Firing-Pattern modulo the l.c.m. of $2^{k/2} - c$ and $2^{k/2} + c$. Since c is odd, this is $2^k - c^2$.

Now we suppose k is odd. As c is relatively prime to $N = 2^k - c^2$, there exists a c' such that $cc' \equiv 1 \bmod N$. Take $\omega = 2^{(k+1)/2}c'$. Then $\omega^2 = 2^{k+1}c'^2 \equiv 2 \bmod N$. Also $\omega^k = (2^k)^{(k+1)/2}c'^k \equiv c^{k+1}c'^k \equiv c \bmod N$. Thus we can construct $[1 + c, 1, 1, \ldots, 1]$ (with k 1's) for a recognizer and the signature $[1, 0, 1]$ for a generator simultaneously in the ring $\mathbf{Z}_N$ directly.

6. Some more examples

In [29] Valiant gave a list of combinatorial problems all of which can be solved by holographic algorithms. In each case, a "magic" design of matchgates and signatures were presented to derive the algorithm. With our machinery, we can show all these problems can be systematically derived. In particular, we will see how the two mysterious bases **b1** and **b2** show up naturally. The framework here can handle all the problems from [29]. (But for PL-FO-2-COLOR, which uses a basis of three vectors, it is more naturally dealt with in the context of more general bases.)

6.1. Not-All-Equal gate

In [29], four problems employ the NAE (Not-All-Equal) gate $[0, 1, 1, 0]$. They are #PL-3-NAE-SAT, #PL-3-NAE-ICE, #PL-3-$(1, 1)$-CYCLECHAIN and PL-NODE-BIPARTITION (this last one uses a generator with signature $[x, 1, 1, x]$).

Notice that they have a common restriction of "maximum degree 3". This is necessary because if $k > 3$, then $[0, 1, 1, \ldots, 1, 0]$ ($(k-1)$ 1's) is not realizable. This is a result of [5], but it's easy to see now.

For the case of degree 3, by Lemma 4.5, taking α, β to be the two roots of $x^2 - x + 1 = 0$ and $A/B = -1$, we have $B_{rec}([0, 1, 1, 0]) = \left\{\left[\binom{1+\omega}{1-\omega}, \binom{\alpha+\beta\omega}{\alpha-\beta\omega}\right] \mid \omega^3 = \pm 1\right\}$.

Noticing that $\alpha^3 = -1$ and $\alpha\beta = 1$, letting $\omega = \alpha$, we have (using $\sim$ on $\mathcal{M}$)

$$\left[\binom{1+\omega}{1-\omega}, \binom{\alpha+\beta\omega}{\alpha-\beta\omega}\right] = \left[\binom{1+\alpha}{1-\alpha}, \binom{\alpha+\beta\alpha}{\alpha-\beta\alpha}\right] = \left[\binom{1}{1}, \binom{1}{-1}\right].$$

This is **b2** in [29]. Actually for each of the four problems, in order to intersect with the subvarieties of other generators and recognizers, this is the only choice. We omit the details.

6.2. $\#_{k+1}2/k$-X-Matchings

Input: A planar bipartite graph $G = (V_1, V_2, E)$. Nodes in V_1 and V_2 have degrees 2 and k respectively.
Output: The number mod $(k+1)$ of all (not necessarily perfect) matchings.

This problem is a slight variation on #X-Matchings from [29], which has general weights on edges and uses an *unsymmetric* signature. (We will discuss unsymmetric signatures in Section 7.) The case $k = 4$ was explicitly stated in [29], but the proof there clearly also handles general k. Jerrum [17] showed that counting matchings for planar graphs is #P-complete. Vadhan [26] showed that this remains #P-complete for planar bipartite graphs of degree 6.

For this problem we are looking for a generator with signature $[1, 1, 0]$ and a recognizer with signature $[1, 1, 0, \ldots, 0]$ ($(k-1)$ 0's) simultaneously. From Lemma 4.6, with $A = B = 1$, $\alpha = 0$, we have: $B_{rec}([1, 1, 0, \ldots, 0]) = \{[\binom{1}{1}, \binom{0}{k}]\}$. We hope that $[\binom{1}{1}, \binom{0}{k}] \in B_{gen}([1, 1, 0])$.

From Lemma 4.8, we must have $k + 1 = 0$. So we can only work inside the ring $\mathbf{Z}_{k+1}$.

Remark. In $\mathbf{Z}_{k+1}$, this basis $[\binom{1}{1}, \binom{0}{k}]$ in $\mathcal{M}$ under the equivalence relation $\sim$ is exactly **b1** in [29].

Theorem 6.1. *There is a polynomial time algorithms for $\#_{k+1}2/k$-X-Matchings. Any modulus m for which the appropriate signatures exist must be a divisor of $k + 1$.*

6.3. $\oplus$PL-EVEN-LIN2

In this problem, we wish to construct generators for $[1, x, 1]$, $[x, 1, x]$, $[1, 0, 1]$, $[0, 1, 0]$, $[1, 0, 0, \ldots, 0, 1]$ and recognizers for $[1, 0, -1, 0, 1]$, $[0, 1, 0, -1, 0]$, $[1, 0, 1]$, $[0, 1, 0]$.

By Lemma 4.5, for $A = B = 1/2$, $\alpha = i$, $\beta = -i$ (here $i = \sqrt{-1}$), we have

$$B_{rec}\big([1, 0, -1, 0, 1]\big) = \left\{ \left[\begin{pmatrix} 1+\omega \\ 1-\omega \end{pmatrix}, \begin{pmatrix} i-i\omega \\ i+i\omega \end{pmatrix} \right] \,\middle|\, \omega^4 = \pm 1 \right\}.$$

We hope that $[\binom{1+\omega}{1-\omega}, \binom{i-i\omega}{i+i\omega}]$ is also a basis for the recognizer $[0, 1, 0]$.

By Lemma 4.2, we require that $(1+\omega)(i+i\omega) + (1-\omega)(i-i\omega) = 0$. That is, $\omega = i$, and

$$\left[\begin{pmatrix} 1+\omega \\ 1-\omega \end{pmatrix}, \begin{pmatrix} i-i\omega \\ i+i\omega \end{pmatrix} \right] = \left[\begin{pmatrix} 1+i \\ 1-i \end{pmatrix}, \begin{pmatrix} i+1 \\ i-1 \end{pmatrix} \right] = \left[\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \end{pmatrix} \right].$$

We can easily verify that this is also a basis for the other recognizers and generators and we remark that this basis is precisely **b2** in [29]. One can also prove 2 is the only modulus for this problem.

7. Beyond symmetric signatures

The theory of symmetric signatures has been satisfactorily developed. Symmetric signatures are particularly useful because they have clear combinatorial meanings. However general (i.e. unsymmetric) signatures have also been used before. To understand completely the power of holographic algorithms, we must study unsymmetric signatures as well. (In the following, we discuss generators only; the situation for recognizers is similar.)

Following the framework in [4], a generator is a contravariant tensor of the form $G = (g^{i_1 i_2 \cdots i_n})$ where the index $i_1 i_2 \ldots i_n \in \{0, 1\}^n$. We also denote $G = (g^S)$ where $S \subseteq [n]$, and $g^S = g^{\chi_S(1)\chi_S(2)\cdots\chi_S(n)}$. A generator signature G is realizable on a basis $\boldsymbol{\beta}$ iff the standard signature $G' = \boldsymbol{\beta}^{\otimes n} G$ can be realized by some planar matchgate. There are two conditions for a standard signature (g'^S) to be realizable:

Parity constraints: Either $g'^S = 0$ for all $|S|$ even, or $g'^S = 0$ for all $|S|$ odd.
Matchgate identities: G' satisfies all the *useful Grassmann–Plücker identities*.

Definition 7.1. A tensor G is *admissible* as a generator on a basis $\boldsymbol{\beta}$ iff $G' = \boldsymbol{\beta}^{\otimes n} G$ satisfies the *parity constraints*. Let $B^p_{gen}(G)$ denote the subset of $\mathcal{M}$ for which G is admissible as a generator.

By definition we have $B_{gen}(G) \subseteq B^p_{gen}(G)$ for all G.

For symmetric signatures, we already observed that there are some different levels of realizability. Some signatures are realizable on isolated points, while others are realizable on lines or curves. Any success of getting a holographic algorithm typically results from either a generator or a recognizer having more than isolated points of realizability. In terms of $\mathcal{M}$, this refers to the dimension of the subvariety $B_{gen}(G)$ (and the corresponding subvarieties for recognizers). More precisely,

Definition 7.2. A generator G is called d-realizable (resp. d-admissible) for an integer $d \geqslant 0$ iff $B_{gen}(G) \subseteq \mathcal{M}$ (resp. $B^p_{gen}(G) \subseteq \mathcal{M}$) is a (non-empty) algebraic subset of dimension at least d.

By definition, if a generator G is d-realizable, then it is d-admissible.

Remark. Since $\mathcal{M}$ has dimension two, 2-realizability is universal realizability which means that G is realizable on any basis. This is because the conditions defining realizability are polynomial equations (with coefficients from (g^S), and variables on $\mathcal{M}$). If there is at least one polynomial which is not identically 0, the algebraic set has dimension $\leqslant 1$. Using any 2-realizable signature is a freebie in the design of holographic algorithms; it places no restriction on the rest of the design. Therefore they are particularly desirable.

7.1. Characterization of 2-admissibility

The following theorem is a complete characterization of 2-admissibility over fields of characteristic 0. It uses rank estimates related to the *Kneser Graph* $\mathrm{KG}_{2k+1,k}$ [21–23,12–15].

Theorem 7.1. *G is 2-admissible iff* (1) *$n = 2k$ is even*; (2) *all $g^S = 0$ except for $|S| = k$*; *and* (3) *for all $T \subseteq [n]$ with $|T| = k+1$,*

$$\sum_{S\subset T,\, |S|=k} g^S = 0. \tag{9}$$

The solution space is a linear subspace of dimension $\frac{1}{k+1}\binom{2k}{k}$ (the Catalan number).

Consider all subsets of $[n]$ of a certain cardinality. Let $0 \leqslant k \leqslant \ell \leqslant n$, and let $A_{k,\ell,n}$ denote the $\binom{n}{k} \times \binom{n}{\ell}$ Boolean matrix indexed by (A, B), where $A, B \subseteq [n]$ and $|A| = k, |B| = \ell$, and the entry at (A, B) is $\chi_{[A\subseteq B]}$, i.e., it is 1 if $A \subseteq B$ and 0 otherwise. It is known that over the rationals $\mathbf{Q}$, the rank $\mathrm{rk}(A_{k,\ell,n}) = \min\{\binom{n}{k}, \binom{n}{\ell}\}$ [12–15]. We will not deal with finite characteristics here. The situation with finite characteristic p is interesting and is more involved. For example, Linial and Rothschild [15] proved exact rank formula for characteristic 2 and 3. The rank "defect" compared to the characteristic 0 case provides more admissible signatures. This will be discussed in future work.

We restate the definition of d-admissibility in more detail.

Definition 7.3. $G = (g^S)_{S\subseteq[n]}$ is called d-admissible if the following algebraic variety V has dimension at least d, where $V = V_0 \cup V_1 \subseteq \mathcal{M}$, and V_0 (resp. V_1) is defined by the set of all parity requirements for the generator signature of an odd (resp. even) matchgate.

More precisely, consider V_0. We take a point (in dehomogenized coordinates) $\left(\begin{smallmatrix} 1 & x \\ 1 & y \end{smallmatrix}\right) \in \mathcal{M}$. We also denote $x_0 = x$, $x_1 = y$. Let $T \subseteq [n]$ with $|T|$ even. Then we require

$$\left\langle \bigotimes_{\sigma=1}^{n} [1, x_{[\sigma\in T]}], G \right\rangle = 0.$$

Note that the left-hand side is precise the entry of the standard signature indexed at T, under the (contravariant) basis transformation. Similarly we define V_1, where the equations are over all T with an odd cardinality.

We note that

$$\left\langle \bigotimes_{\sigma=1}^{n} [1, x_{[\sigma\in T]}], G \right\rangle = \sum_{\substack{0\leqslant i\leqslant n-|T| \\ 0\leqslant j\leqslant |T|}} x^i y^j \sum_{\substack{A\subseteq T^c,\, |A|=i \\ B\subseteq T,\, |B|=j}} g^{A\cup B}. \tag{10}$$

If $\dim(V) = 2$, then either $\dim(V_0) = 2$ or $\dim(V_1) = 2$. For $\dim(V_0) = 2$, we have the following: For all $T \subseteq [n]$ with $|T|$ even, and for all $0 \leqslant i \leqslant n - |T|$ and $0 \leqslant j \leqslant |T|$,

$$\sum_{A\subseteq T^c,\, B\subseteq T,\, |A|=i,\, |B|=j} g^{A\cup B} = 0. \tag{11}$$

(If there is one equation not satisfied, then there is at least one non-trivial polynomial among the parity requirements, which implies $\dim(V_0) \leqslant 1$.) For $\dim(V_1) = 2$, the above holds for all $|T|$ odd. Continuing with $\dim(V_0) = 2$, by taking $i = 0$, we get for all $T \subseteq [n]$ with $|T|$ even, and $j \leqslant |T|$,

$$\sum_{S\subseteq T, |S|=j} g^S = 0. \tag{12}$$

Also by taking $j = 0$, we get for all $i \leqslant n - |T|$,

$$\sum_{S\subseteq T^c,\, |S|=i} g^S = 0.$$

If $S \subseteq [n]$ with $|S|$ even, then we may take $T = S$ and $j = |T|$, and it follows that

$$g^S = 0.$$

If n is odd, then T is even and T^c is odd, and together they range over all possible subsets of $[n]$. It follows that

$$g^S = 0,$$

for all $S \subseteq [n]$. That is, G is trivial.

An identical argument also shows that for n odd and $\dim(V_1) = 2$, the trivial $G \equiv 0$ is the only possibility.

Now we assume $n = 2k$ is even, and continuing with $\dim(V_0) = 2$. Both T and T^c are even. Pick any T even and $i = n - |T|$, we get

$$\sum_{A \subseteq T^c,\, B \subseteq T,\, |A|=i,\, |B|=j} g^{A \cup B} = \sum_{S \supseteq T^c,\, |S|=i+j} g^S = 0,$$

i.e. for all even $T' \subseteq [n]$ and all $i \geqslant |T'|$,

$$\sum_{S \supseteq T',\, |S|=i} g^S = 0. \tag{13}$$

If $|S| = i < k$, we form the following system of equations from (12),

$$\sum_{S \subseteq T,\, |S|=i} g^S = 0,$$

where T ranges over all subsets of $[n]$ with $|T| = t$, and $t = i$ or $i + 1$, whichever is even. This linear system has rank $\binom{n}{i}$. It follows that $g^S = 0$ for all $|S| < k$.

Similarly if $|S| = i > k$, we can use (13) with $|T| = i$ or $i - 1$, whichever is even, and summing over all subsets S containing T. This linear system also has rank $\binom{n}{i}$. It follows that $g^S = 0$ for all $|S| > k$.

Therefore the only non-zero entries of G are among g^S with half weight $|S| = k$. Also with $\dim(V_0) = 2$, we may assume k is odd. Otherwise, we already know $g^S = 0$ for all $|S|$ even.

A similar argument for V_1 shows that, in order for $\dim(V_1) = 2$, we must have $n = 2k$ even, all $g^S = 0$ except for $|S| = k$ and k is even.

Summarizing, we have

Lemma 7.1. *If G is 2-admissible, then $n = 2k$ is even, all $g^S = 0$ except for $|S| = k$. If k is odd (resp. even) then the only possibility is $\dim(V_0) = 2$ (resp. $\dim(V_1) = 2$). Moreover, for all $T \subseteq [n]$ with $|T| = k + 1$,*

$$\sum_{S \subseteq T,\, |S|=k} g^S = 0. \tag{14}$$

Next we prove that the conditions in Lemma 7.1 are also sufficient for G being 2-admissible, i.e., we prove (11), thus all the polynomials in (10) are identically zero.

Suppose k odd. We prove $\dim(V_0) = 2$. A similar argument does for k even and $\dim(V_1) = 2$. We only need to verify (11) for all $i + j = k$, namely for all $T \subseteq [n]$ with $|T|$ even, and for all $0 \leqslant i \leqslant n - |T|$, and $0 \leqslant j = k - i \leqslant |T|$,

$$\sum_{A \subseteq T^c,\, B \subseteq T,\, |A|=i,\, |B|=k-i} g^{A \cup B} = 0. \tag{15}$$

Denote by $t = |T|$ and $s = n - |T|$. By exchanging T and T^c (both being even subsets of $[n]$) we may assume $s \leqslant t$. Since k is odd, we have the strict $s < t$, for otherwise $s = t = k$ would be odd.

We prove (15) by induction on $i \geqslant 0$. The base case is $i = 0$ and $j = k$. Let's consider all $U \subseteq T$ with $|U| = k + 1$. Note that as $t \geqslant k + 1$, this is not vacuous. By (14) we have

$$\sum_{S \subseteq U,\, |S|=k} g^S = 0.$$

Summing over all such U, and consider how many times each $S \subseteq [n]$ with $|S| = k$ appears in the sum, we get

$$\sum_{\substack{A \subseteq T^c,\, |A|=0 \\ B \subseteq T,\, |B|=k}} g^{A \cup B} = \sum_{S \subseteq T,\, |S|=k} g^S = \frac{1}{\binom{t-k}{1}} \sum_{\substack{U \subseteq T \\ |U|=k+1}} \sum_{S \subseteq U,\, |S|=k} g^S = 0. \tag{16}$$

Inductively we assume (15) has been proved for $i-1$, for some $i \geqslant 1$. Consider i and $j = k-i$. We may assume $i \leqslant s$; otherwise we are done. Also $k-i+1 \leqslant k+1 \leqslant t$. Consider all subsets $U = U_1 \cup U_2 \subseteq [n]$, where $U_1 \subseteq T^c$, $U_2 \subseteq T$, with $|U_1| = i$ and $|U_2| = k-i+1$. Note that $|U| = k+1$. We have

$$0 = \sum_{S \subseteq U,\, |S|=k} g^S = \sum_{A \subseteq U_1,\, |A|=i-1} g^{A \cup U_2} + \sum_{B \subseteq U_2,\, |B|=k-i} g^{U_1 \cup B},$$

as all sets $S \subseteq U$ with $|S| = k$ are classified into two classes according to whether $|S \cap U_1| = i-1$ or i. Then summing over all such U,

$$0 = \sum_{U} \sum_{S \subseteq U, |S|=k} g^S = \binom{s-(i-1)}{1} \sum_{\substack{A \subseteq T^c,\, |A|=i-1 \\ B \subseteq T,\, |B|=k-i+1}} g^{A \cup B} + \binom{t-(k-i)}{1} \sum_{\substack{A \subseteq T^c,\, |A|=i \\ B \subseteq T,\, |B|=k-i}} g^{A \cup B},$$

by considering how many times each S of the two classes appears in the sum $\sum_U \sum_S$. Since the first sum is 0 by inductive hypothesis, and $t-k+i \geqslant 1$, the second sum is also zero. Thus

$$\sum_{A \subseteq T^c,\, B \subseteq T,\, |A|=i,\, |B|=k-i} g^{A \cup B} = 0.$$

This proves Theorem 7.1.

We can further prove:

Theorem 7.2. *If G is 2-admissible with arity $2k$, then $\forall \boldsymbol{\beta} = \left(\begin{smallmatrix} n_0 & p_0 \\ n_1 & p_1 \end{smallmatrix}\right) \in \mathcal{M}$, $\boldsymbol{\beta}^{\otimes 2k} G = (n_0 p_1 - n_1 p_0)^k G$.*

In order to prove this theorem, we first prove the following lemma:

Lemma 7.2. *Let G be 2-admissible with arity $2k$, $S \subseteq [2k]$ with $|S| = k$, and $A \subseteq S^c$. Then*

$$\sum_{B \subseteq S \text{ and } |B|=k-|A|} g^{A \cup B} = (-1)^{|A|} g^S.$$

Proof. We prove it by induction on $|A| \geqslant 0$.

The case $|A| = 0$ is obvious.

Inductively we assume the lemma has been proved for all $|A| \leqslant i-1$, for some $i \geqslant 1$. Letting $|A| = i > 0$ and letting G be 2-admissible, it follows from Lemma 7.1 that we have

$$\sum_{C \subseteq A \cup S \text{ and } |C|=k} g^C = 0.$$

Then

$$\begin{aligned} 0 &= \sum_{C \subseteq A \cup S \text{ and } |C|=k} g^C \\ &= \sum_{B \subseteq S \text{ and } |B|=k-|A|} g^{A \cup B} + \sum_{t=0}^{|A|-1} \sum_{A_1 \subseteq A,\, |A_1|=t} \sum_{B \subseteq S,\, |B|=k-|A_1|} g^{A_1 \cup B}, \end{aligned}$$

according to $t = |A \cap C| = 0, 1, \ldots, |A|$. Since $|A_1| = t \leqslant |A|-1$, by induction we have:

$$\sum_{B \subseteq S,\, |B|=k-|A_1|} g^{A_1 \cup B} = (-1)^{|A_1|} g^S = (-1)^t g^S.$$

So

$$\begin{aligned} 0 &= \sum_{B \subseteq S \text{ and } |B|=k-|A|} g^{A \cup B} + g^S \sum_{t=0}^{|A|-1} (-1)^t \binom{|A|}{t} \\ &= \sum_{B \subseteq S \text{ and } |B|=k-|A|} g^{A \cup B} - (-1)^{|A|} g^S. \end{aligned}$$

From the last equation, we have

$$\sum_{B \subseteq S \text{ and } |B|=k-|A|} g^{A \cup B} = (-1)^{|A|} g^S.$$

This completes the proof. □

Corollary 7.1. *If G is any 2-admissible signature, then $\forall S \subseteq [2k]$, $g^S = (-1)^k g^{S^c}$.*

Now we can prove Theorem 7.2.

Proof. To simplify notations, we use the dehomogenized coordinates $\boldsymbol{\beta} = \binom{1\ x}{1\ y} = \binom{1\ x_0}{1\ x_1}$. Some exceptional cases can be proved directly.

First it is obvious that $\boldsymbol{\beta}^{\otimes 2k} G$ is also 2-admissible. So for any $S \subseteq [2k]$ and $|S| \neq k$,

$$\left\langle \bigotimes_{\sigma=1}^{n} [1, x_{[\sigma \in S]}], G \right\rangle \equiv 0.$$

Now let $S \subseteq [2k]$ and $|S| = k$,

$$\left\langle \bigotimes_{\sigma=1}^{n} [1, x_{[\sigma \in S]}], G \right\rangle = \sum_{0 \leqslant i \leqslant k} x^i y^{k-i} \sum_{A \subseteq S^c,\, |A|=i} \sum_{B \subseteq S,\, |B|=k-i} g^{A \cup B}.$$

By Lemma 7.2 and for $A \subseteq S^c$, $|A| = i$, we have

$$\sum_{B \subseteq S,\, |B|=k-i} g^{A \cup B} = (-1)^i g^S.$$

So

$$\left\langle \bigotimes_{\sigma=1}^{n} [1, x_{[\sigma \in S]}], G \right\rangle = \sum_{0 \leqslant i \leqslant k} x^i y^{k-i} \sum_{A \subseteq S^c,\, |A|=i} (-1)^i g^S = g^S \sum_{0 \leqslant i \leqslant k} x^i y^{k-i} (-1)^i \binom{k}{i} = (y-x)^k g^S.$$

This completes the proof. □

Since a scaling preserves realizability, the theorem gives:

Corollary 7.2. *If a 2-admissible G is realizable on some basis (e.g., on the standard basis), then it is realizable on any basis, which means it is 2-realizable.*

For $n = 6$, all 2-admissible G's form a 5-dimensional linear space. Applying the matchgate identities, we find that there are 5 different 2-realizable signatures (up to scaling). Let G_1 and G_2 be the following

$$g_1^\alpha = \begin{cases} 1, & \alpha \in \{000111, 011001, 101010, 110100\}, \\ -1, & \alpha \in \{111000, 100110, 010101, 001011\}, \\ 0, & \text{otherwise}, \end{cases}$$

$$g_2^\alpha = \begin{cases} 1, & \alpha \in \{010101, 011010, 100110, 101001\}, \\ -1, & \alpha \in \{101010, 100101, 011001, 010110\}, \\ 0, & \text{otherwise}. \end{cases}$$

Then all the 2-realizable signatures are obtained by cyclically rotating the indices of G_1 or G_2. (Rotating 3 bits on G_1 is G_1 itself up to a scaling factor -1; rotating 2 bits on G_2 gives G_2 back. So there are 3 different 2-realizable signatures from rotating G_1 and 2 different ones from rotating G_2. See Figs. 1 and 2.)

It turns out that all of these can be obtained from the *planar tensor product* operation which we define next.

Definition 7.4. Let $\mathrm{Rot}_r(G)$ be the tensor obtained by circularly rotating clockwise the coordinates of G by r bits. Let $G \otimes G'$ be the tensor product with all indices of G before all indices of G'. A planar tensor product is a finite sequence of operations of $\mathrm{Rot}_r(G)$ and $G \otimes G'$.

By direct constructions and matchgate identities, we can prove the following theorem.

Theorem 7.3. *$B_{gen}(\mathrm{Rot}_r(G)) = B_{gen}(G)$ and $B_{gen}(G_1 \otimes G_2) = B_{gen}(G_1) \cap B_{gen}(G_2)$. Thus a planar tensor product preserves B_{gen}.*

Theorem 7.4. *Each of the five 2-realizable signatures for $n = 6$ is obtainable as a planar tensor product from $(0, 1, -1, 0)$.*

Valiant [30] already noted that $(0, 1, -1, 0)$ is realizable under all bases, i.e., 2-realizable in our terminology. From $(0, 1, -1, 0)$, we can construct a family of 2-realizable signatures for any arity $2k$ by planar tensor product. It is an open question if this family (up to scaling) captures all the 2-realizable signatures. This is true for $n \leqslant 6$.

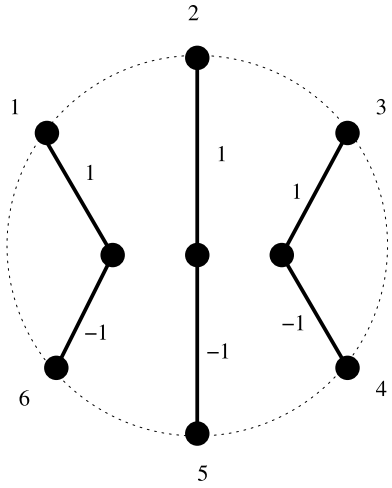


Fig. 1. One planar tensor product for arity 6.

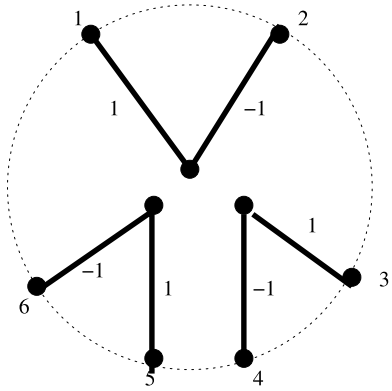


Fig. 2. Another planar tensor product for arity 6.

Definition 7.5. A signature G is called prime iff it cannot be decomposed as a planar tensor product of two signatures of positive arity.

In particular $(0, 1, -1, 0)$ is a prime 2-realizable signature. The above open problem is essentially whether $(0, 1, -1, 0)$ is the unique prime 2-realizable signature (up to scaling).

7.2. 1*-admissibility and* 1*-realizability*

1-admissibility (resp. 1-realizability) is strictly weaker than 2-admissibility (resp. 2-realizability). In this section, we give some constructions of 1-admissible and 1-realizable families which are not in general 2-admissible or 2-realizable. These are in fact prime signatures. Planar tensor product can be applied to construct more 1-realizable families.

First we give a family of 1-admissible generators.

Theorem 7.5. *Letting $n = 2k$ be even, we have all $g^S = 0$ except for those $|S| = k$. Finally for all $S \subset [n]$ with $|S| = k$, $g^S = g^{S^c}$. Then G is* 1*-admissible.*

Proof. We prove this by showing that $\forall x$, $\left(\begin{smallmatrix} 1 & x \\ 1 & -x \end{smallmatrix}\right) \in V_1$, where V_1 is defined in Definition 7.3. Let $T \subset [n]$ with $|T|$ odd. Then we require the following polynomial to be identically zero:

$$\left\langle \bigotimes_{\sigma=1}^{n} [1, x_{[\sigma \in T]}], G \right\rangle \equiv 0,$$

where $x_0 = x$ and $x_1 = -x$. In the above setting, we have

$$\left\langle \bigotimes_{\sigma=1}^{n} [1, x_{[\sigma\in T]}], G \right\rangle = x^k \sum_{\max\{0,|T|-k\}\leqslant i\leqslant \min\{k,|T|\}} (-1)^i \sum_{\substack{A\subseteq T^c,\,|A|=k-i \\ B\subseteq T,\,|B|=i}} g^{A\cup B}.$$

We assume that $k \geqslant |T|$ (the case $k < |T|$ is similar). Then the outer sum is $\sum_{i=0}^{|T|}$. Since $|T|$ is odd, the first and the last term of the outer sum cancel out. Similarly the second and the second last term cancel out, and so on. There are altogether an even number $|T|+1$ of terms of this outer sum over i, and the term indexed by i and by $|T|-i$ cancel out. It follows that this summation is identically 0. This completes the proof. □

For $n = 4$, in order to be 1-realizable, the matchgate identities further require $g^{0011}g^{1001} = 0$. This gives the following two 1-realizable signatures (they are prime for $a^2 \neq b^2$):

$$g^\alpha = \begin{cases} a, & \alpha \in \{0101, 1010\}, \\ b, & \alpha \in \{0011, 1100\}, \\ 0, & \text{otherwise} \end{cases}$$

and

$$g^\alpha = \begin{cases} a, & \alpha \in \{0101, 1010\}, \\ b, & \alpha \in \{1001, 0110\}, \\ 0, & \text{otherwise.} \end{cases}$$

This family of 1-realizable signatures has been used in a subsequent paper [10] to obtain some surprising holographic algorithms.

Next, we present another family of 1-realizable signatures, which are not subsumed by any of the above. It also has some generalized symmetry. It can be viewed as a generalization of Case 2 in Lemma 4.12.

Theorem 7.6. *For any $g_1, g_2, \ldots, g_n, \alpha \in \mathbf{F}$, where $g_1 + g_2 + \cdots + g_n = 0$, let $G = (g^S)_{S\subseteq[n]}$ be defined as follows*

$$g^S = \alpha^{|S|-1} \sum_{i\in S} g_i.$$

Then G is 1-realizable and

$$B_{gen}(G) = \left\{ \left[\begin{pmatrix} -\alpha \\ n_1 \end{pmatrix}, \begin{pmatrix} 1 \\ p_1 \end{pmatrix} \right] \in \mathcal{M} \;\middle|\; n_1, p_1 \in \mathbf{F} \right\}.$$

Proof. For simplicity, we use the dehomogenized coordinates $\left(\begin{smallmatrix} 1 & x \\ 1 & y \end{smallmatrix}\right)$ where $x = -1/\alpha$. Some exceptional cases such as $\alpha = 0$ can be proved directly (we use the convention that $\alpha^0 = 1$ and $0 \cdot \alpha^{0-1} = 0$ even when $\alpha = 0$).

Let $T \subseteq [n]$. If $|T| = 0$ or $|T| = n$, by (10) and the definition of G, it follows easily that

$$\left\langle \bigotimes_{\sigma=1}^{n} [1, x_{[\sigma\in T]}], G \right\rangle = 0.$$

Otherwise we have

$$\begin{aligned}
&\left\langle \bigotimes_{\sigma=1}^{n} [1, x_{[\sigma\in T]}], G \right\rangle \\
&= \sum_{\substack{0\leqslant i\leqslant n-|T| \\ 0\leqslant j\leqslant |T|}} x^i y^j \sum_{\substack{A\subseteq T^c,\,|A|=i \\ B\subseteq T,\,|B|=j}} g^{A\cup B} \\
&= \sum_{\substack{0\leqslant i\leqslant n-|T| \\ 0\leqslant j\leqslant |T|}} x^i y^j \sum_{\substack{A\subseteq T^c,\,|A|=i \\ B\subseteq T,\,|B|=j}} \alpha^{|A\cup B|-1} \sum_{k\in A\cup B} g_k \\
&= \sum_{\substack{0\leqslant i\leqslant n-|T| \\ 0\leqslant j\leqslant |T|}} x^i y^j \alpha^{i+j-1} \sum_{\substack{A\subseteq T^c,\,|A|=i \\ B\subseteq T,\,|B|=j}} \left(\sum_{k\in A} g_k + \sum_{l\in B} g_l \right) \\
&= \sum_{\substack{0\leqslant i\leqslant n-|T| \\ 0\leqslant j\leqslant |T|}} x^i y^j \alpha^{i+j-1} \left(\binom{|T|}{j} \binom{|T^c|-1}{i-1} \sum_{k\in T^c} g_k + \binom{|T^c|}{i} \binom{|T|-1}{j-1} \sum_{l\in T} g_l \right)
\end{aligned}$$

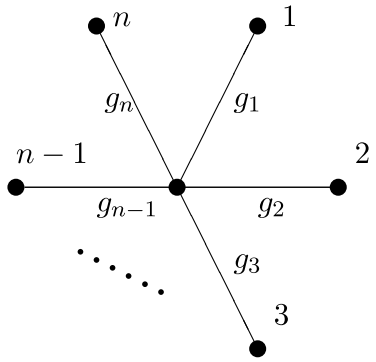


Fig. 3. 1-realizability.

$$= \sum_{k\in T^c} g_k \Bigg(\sum_{\substack{0\leqslant i\leqslant n-|T| \\ 0\leqslant j\leqslant |T|}} x^i y^j \alpha^{i+j-1} \binom{|T|}{j} \binom{n-|T|-1}{i-1} - \sum_{\substack{0\leqslant i\leqslant n-|T| \\ 0\leqslant j\leqslant |T|}} x^i y^j \alpha^{i+j-1} \binom{n-|T|}{i} \binom{|T|-1}{j-1} \Bigg)$$

$$= \sum_{k\in T^c} g_k \big(x(1+\alpha x)^{n-|T|-1}(1+\alpha y)^{|T|} - y(1+\alpha x)^{n-|T|}(1+\alpha y)^{|T|-1} \big).$$

If $|T| < n-1$, the above equation is identically 0 when $x = -1/\alpha$.

For $|T| = n-1$, suppose $T = [n] - \{t\}$, then at $x = -1/\alpha$, the value of the above equation is λg_t where $\lambda = -(1+\alpha y)^{n-1}/\alpha$. This standard signature is realizable by the star (see Fig. 3). □

Remark. When $n = 2$, this generator is the 2-realizable signature $(0, 1, -1, 0)$.

Addendum. In this paper we could only prove a characterization of 2-admissibility, some results on 2-realizability and constructed some families of 1-admissible and 1-realizable signatures. In a subsequent paper [11], we have proved a complete characterization of 2-realizability, which confirms the conjecture here. And the characterization of 2-admissibility in this paper serves as a good start point of that result. In [11], we also give some characterizations of 1-admissibility and 1-realizability.

Acknowledgments

We would like to thank Leslie Valiant for many comments and discussions. We also thank Eric Bach, Xi Chen, Steve Cook, Jon Kleinberg, Edith Hemaspaandra, Lane Hemaspaandra, Joseph Landsberg, Jason Morton, Salil Vadhan, Avi Wigderson and Mingji Xia for their comments and interests. We especially thank the anonymous referees for this paper, both for the conference version and for the journal version.

References

[1] E. Bach, J. Shallit, Algorithmic Number Theory, vol. 1: Efficient Algorithms, MIT Press, 1996.
[2] S. Cook, The complexity of theorem proving procedures, in: Proceedings of the Third Annual ACM Symposium on Theory of Computing, 1971, pp. 151–158.
[3] J.-Y. Cai, Vinay Choudhary, Some results on matchgates and holographic algorithms, in: Proceedings of ICALP 2006, Part I, in: Lecture Notes in Comput. Sci., vol. 4051, 2006, pp. 703–714; Int. J. Software Informatics 1 (1) (2007) 3–36; Also available at Electronic Colloquium on Computational Complexity TR06-048, 2006.
[4] J.-Y. Cai, Vinay Choudhary, Valiant's Holant Theorem and matchgate tensors, in: Proceedings of TAMC 2006, in: Lecture Notes in Comput. Sci., vol. 3959, 2006, pp. 248–261; Theoret. Comput. Sci. 384 (1) (2007) 22–32; Also available at Electronic Colloquium on Computational Complexity Report TR05-118.
[5] J.-Y. Cai, Vinay Choudhary, Pinyan Lu, On the theory of matchgate computations, in: IEEE Conference on Computational Complexity, 2007, pp. 305–318.
[6] J.-Y. Cai, Pinyan Lu, On symmetric signatures in holographic algorithms, in: Proceedings of STACS 2007, in: Lecture Notes in Comput. Sci., vol. 4393, 2007, pp. 429–440; Theory Comput. Syst. 46 (3) (2010) 398–415.
[7] J.-Y. Cai, Pinyan Lu, Holographic algorithms: From art to science, in: Proceedings of STOC, 2007, pp. 401–410.
[8] J.-Y. Cai, Pinyan Lu, Bases collapse in holographic algorithms, in: IEEE Conference on Computational Complexity, 2007, pp. 292–304, Comput. Complexity 17 (2) (2008) 254–281.
[9] J.-Y. Cai, Pinyan Lu, Holographic algorithms: The power of dimensionality resolved, in: Proceedings of ICALP, 2007, pp. 631–642, Theoret. Comput. Sci. 410 (18) (2009) 1618–1628.
[10] J.-Y. Cai, Pinyan Lu, Holographic algorithms with unsymmetric signatures, in: ACM–SIAM Symposium on Discrete Algorithms (SODA), 2008, pp. 54–63.
[11] J.-Y. Cai, Pinyan Lu, Signature theory in holographic algorithms, in: S.H. Hong, H. Nagamochi, T. Fukunaga (Eds.), Proceedings of ISAAC, in: Lecture Notes in Comput. Sci., vol. 5369, Springer, 2008, pp. 568–579.
[12] W. Foody, A. Hedayat, On theory and applications of BIB designs with repeated blocks, Ann. Statist. 5 (1977) 932–945.
[13] W. Foody, A. Hedayat, Note: Correction to "On theory and application of BIB designs with repeated blocks", Ann. Statist. 7 (4) (1979) 925.
[14] R.L. Graham, S.-Y.R. Li, W.-C.W. Li, On the structure of t-designs, SIAM. J. Algebraic Discrete Methods 1 (1980) 8.
[15] N. Linial, B. Rothschild, Incidence matrices of subsets—A rank formula, SIAM. J. Algebraic Discrete Methods 2 (1981) 333.

[16] D. Lichtenstein, Planar formulae and their uses, SIAM J. Comput. 11 (2) (1982) 329–343.
[17] M. Jerrum, Two-dimensional monomer-dimer systems are computationally intractable, J. Stat. Phys. 48 (1987) 121–134; J. Stat. Phys. 59 (1990) 1087–1088, Erratum.
[18] R.M. Karp, Reducibility among combinatorial problems, in: Raymond E. Miller, James W. Thatcher (Eds.), Complexity of Computer Computations, Plenum, New York, 1972, pp. 85–103.
[19] P.W. Kasteleyn, The statistics of dimers on a lattice, Physica 27 (1961) 1209–1225.
[20] P.W. Kasteleyn, Graph theory and crystal physics, in: F. Harary (Ed.), Graph Theory and Theoretical Physics, Academic Press, London, 1967, pp. 43–110.
[21] M. Kneser, Aufgabe 360, Jahresber. Deutsch. Math.-Verein. 2 (58) (1955) 27.
[22] L. Lovász, Kneser's conjecture, chromatic number, and homotopy, J. Combin. Theory Ser. A 25 (1978) 319–324.
[23] J. Matoušek, A combinatorial proof of Kneser's conjecture, Combinatorica 24 (1) (2004) 163–170.
[24] K. Murota, Matrices and Matroids for Systems Analysis, Springer, Berlin, 2000.
[25] H.N.V. Temperley, M.E. Fisher, Dimer problem in statistical mechanics – an exact result, Philos. Magazine 6 (1961) 1061–1063.
[26] S.P. Vadhan, The complexity of counting in sparse, regular, and planar graphs, SIAM J. Comput. 31 (2001) 398–427.
[27] L.G. Valiant, Quantum circuits that can be simulated classically in polynomial time, SIAM J. Comput. 31 (4) (2002) 1229–1254.
[28] L.G. Valiant, Expressiveness of matchgates, Theoret. Comput. Sci. 281 (1) (2002) 457–471.
[29] L.G. Valiant, Holographic algorithms (extended abstract), in: Proc. 45th IEEE Symposium on Foundations of Computer Science, 2004, pp. 306–315; A more detailed version appeared in Electronic Colloquium on Computational Complexity Report TR05-099.
[30] L.G. Valiant, Accidental algorithms, in: Proc. 47th Annual IEEE Symposium on Foundations of Computer Science, 2006, pp. 509–517.
[31] Mingji Xia, Peng Zhang, Wenbo Zhao, Computational complexity of counting problems on 3-regular planar graphs, Theoret. Comput. Sci. 384 (2007) 111–125.

Paper 9

Available online at www.sciencedirect.com
ScienceDirect
Journal of Computer and System Sciences 73 (2007) 25–35
JOURNAL OF COMPUTER AND SYSTEM SCIENCES
www.elsevier.com/locate/jcss

$S_2^p \subseteq ZPP^{NP}$ ☆

Jin-Yi Cai

Computer Sciences Department, University of Wisconsin, Madison, WI 53706, USA

Received 17 October 2002; received in revised form 11 July 2003

Abstract

We show that the class S_2^p is contained in ZPP^{NP}. The proof uses universal hashing, approximate counting and witness sampling. As a consequence, a collapse first noticed by Samik Sengupta that the assumption NP has small circuits collapses PH to S_2^p becomes the strongest version to date of the Karp–Lipton Theorem. We also discuss the problem of finding irrefutable proofs for S_2^p in ZPP^{NP}.

Keywords: Complexity Theory; Complexity classes; Symmetric alternation; Karp–Lipton Theorem; Approximate counting; Witness sampling; Irrefutable proof

1. Introduction

The class S_2^p was introduced independently by Canetti [10] and Russell and Sundaram [24] in the mid 1990's. Suppose there are two competing all powerful provers Y and Z. A string x is given, Y wishes to convince us that $x \in L$, and Z wishes to convince us the opposite $x \notin L$. We—the verifier—have only deterministic polynomial time computing power. A language L is in S_2^p iff there is a P-time predicate P such that the following holds:

If $x \in L$ then there exists a y, such that for all z, $P(x, y, z)$ holds;
If $x \notin L$ then there exists a z, such that for all y, $\neg P(x, y, z)$ holds, where both y and z are polynomially bounded in the length of x.

In other words, if $x \in L$ then Y has irrefutable proof y which can withstand any challenge z from Z; and if $x \notin L$ then Z has irrefutable proof z which can withstand any challenge y from Y.

The motivation by both Canetti [10] and Russell and Sundaram [24] was to provide a refinement of the Sipser–Lautemann Theorem (with contribution by Gacs) that $BPP \subseteq \Sigma_2^p \cap \Pi_2^p$ [21,26,27]. Indeed, Canetti [10] extended Lautemann's proof to show that $BPP \subseteq S_2^p$, whereas Russell and Sundaram [24] showed further that $MA \subseteq S_2^p$. Note

☆ Research supported in part by NSF CCR-0196197, CCR-0208013 and a Guggenheim Fellowship. A preliminary version [9] appeared in FOCS 2001 [Jin-Yi Cai, $S_2^p \subseteq ZPP^{NP}$, in: Proc. 42nd IEEE Symposium on Foundations of Computer Science (FOCS), 2001, pp. 620–628].
E-mail address: jyc@cs.wisc.edu.

0022-0000/$ – see front matter
doi:10.1016/j.jcss.2003.07.015

that BPP $\subseteq$ MA is direct from definition (the two-sided error version) of MA [4,5,11,12], thus BPP $\subseteq$ MA $\subseteq \mathrm{S}_2^p$. Also it is known that $\mathrm{P}^{\mathrm{NP}} \subseteq \mathrm{S}_2^p$ [24].

As to upper bound of S_2^p, the only known containment is by definition $\mathrm{S}_2^p \subseteq \Sigma_2^p \cap \Pi_2^p$ (see Section 2). Goldreich and Zuckerman [13] surveyed a number of interesting classes between P and the second level of the Polynomial-time Hierarchy Σ_2^p and Π_2^p. These classes include ZPP, RP, BPP, NP, P^{NP}, MA, AM, $\mathrm{ZPP}^{\mathrm{NP}}$ and S_2^p. They called the classes listed here up to P^{NP} "Traditional classes—classes of the 1970's," the class Arthur–Merlin "a class of the 1980's," and the class S_2^p "a class of the 1990's," underscoring that not much is yet known about this class S_2^p. In their paper [13] Goldreich and Zuckerman gave a number of elegant proofs of known results with the strikingly sharp amplification technique due to Zuckerman [30]. They also prove an interesting result MA $\subseteq \mathrm{ZPP}^{\mathrm{NP}}$. This last result was new in 1997 when [13] appeared; it was independently obtained by Arvind and Köbler [1–3]. In summarizing the known facts about all these classes between P and Σ_2^p and Π_2^p it was observed that both S_2^p and $\mathrm{ZPP}^{\mathrm{NP}}$ appear to share all the known containment properties both below and above [13]. How these two classes are related was unknown.

The main result of this paper is

Theorem 1. $\mathrm{S}_2^p \subseteq \mathrm{ZPP}^{\mathrm{NP}}$.

The proof uses universal hashing, approximate counting and witness sampling. We also discuss the problem of finding irrefutable proofs in $\mathrm{ZPP}^{\mathrm{NP}}$.

There is an interesting consequence of this result with respect to the well-known Karp–Lipton Theorem concerning sparse sets (with contribution by Sipser) [18]. This theorem says, if NP is Cook-reducible ($\leqslant_T^p$) to sparse sets, or equivalently, if **SAT** has polynomial size circuits, then the Polynomial-time Hierarchy collapses to its second level: $\mathrm{PH} = \Sigma_2^p \cap \Pi_2^p$. Many researchers have since tried to improve on this signature theorem—to simplify the proof and to strengthen the collapse. On the one hand, there emerged what I consider to be the "book" proof (as Erdös would say) of the theorem (as far as I know John Hopcroft [16] was the first to give essentially this proof):

> To simulate Π_2^p by Σ_2^p, guess a poly-size circuit C for **SAT**, modify C via self-reducibility so that whenever $C(\phi) = 1$ it also produces a satisfying assignment to ϕ, then check all universal paths of the Π_2^p computation lead to a satisfiable formula.

Samik Sengupta [25] first noticed that this "book" proof actually gave the collapse to S_2^p. (See Section 6.)

While the proof of Karp–Lipton Theorem becomes extremely transparent, more research effort went into trying to extend this beautiful result. Much work was done on the general theme (we mention some in Section 6). Over the years there have been steady improvements on the exact level of collapse of PH, assuming **SAT** has small circuits. In this regard, the best result so far is due to Bshouty et al. [7] and Köbler and Watanabe [20]. Their result states that if NP has polynomial size circuits, then the Polynomial-time Hierarchy collapses to $\mathrm{ZPP}^{\mathrm{NP}}$. Admittedly the proofs of the theorem of Bshouty et al. and Köbler–Watanabe are more involved than the "book" proof of the basic version of the Karp–Lipton Theorem and depend on previous interesting results by Jerrum, Valiant and V. Vazirani [17] and others [8].

By the new theorem $\mathrm{S}_2^p \subseteq \mathrm{ZPP}^{\mathrm{NP}}$ (unconditionally), the (currently) strongest Karp–Lipton Theorem becomes the following Theorem 2. (See Section 6.)

Theorem 2 *(Sengupta). If* **SAT** *has polynomial size circuits, then the Polynomial-time Hierarchy collapses to* S_2^p.

Theorem 1 also subsumes the result MA $\subseteq \mathrm{ZPP}^{\mathrm{NP}}$ by Goldreich–Zuckerman [13] and Arvind–Köbler [1], as we know from Russell and Sundaram [24] that MA $\subseteq \mathrm{S}_2^p$.

2. Preliminaries

The class S_2^p was defined by Russell and Sundaram [24] as follows: $L \in \mathrm{S}_2^p$ iff there is a P-time computable 0-1 function P on three arguments, such that

$$x \in L \implies \left(\exists^p y\right)\left(\forall^p z\right) \quad \left[P(x, y, z) = 1\right], \tag{1}$$

$$x \notin L \implies \left(\exists^p z\right)\left(\forall^p y\right) \quad \left[P(x, y, z) = 0\right], \tag{2}$$

where as usual "$\exists^p y$" stands for "$\exists y \in \{0,1\}^{p_1(|x|)}$" for some polynomial $p_1(\cdot)$. Similarly "$\forall^p z$" stands for "$\forall z \in \{0,1\}^{p_2(|x|)}$" for some polynomial $p_2(\cdot)$. By padding we can suitably extend the length of both y and z, and henceforth we can assume they both vary over the same length n which is a power of 2, and n is polynomially bounded in the length of x.

Given x, for convenience, for a pair (y,z) we say y beats z if $P(x,y,z)=1$, and z beats y if $P(x,y,z)=0$.

It is immediately clear that both implications "$\Rightarrow$" can be replaced by the if and only if relation "$\Leftrightarrow$" without changing the class S_2^p. For instance, suppose $(\exists^p y)\ (\forall^p z)\ [P(x,y,z)=1]$, let y_0 be such a y. Then certainly $x \in L$, else we would have a z_0 such that $(\forall^p y)\ [P(x,y,z_0)=0]$, which is clearly a contradiction to $P(x,y_0,z_0)=1$. Similarly $(\exists^p z)\ (\forall^p y)\ [P(x,y,z)=0]$ implies $x \notin L$. Thus

$$\begin{aligned} x \in L \quad &\Longleftrightarrow \quad \left(\exists^p y\right)\left(\forall^p z\right) \quad \left[P(x,y,z)=1\right], \\ x \notin L \quad &\Longleftrightarrow \quad \left(\exists^p z\right)\left(\forall^p y\right) \quad \left[P(x,y,z)=0\right]. \end{aligned}$$

It follows from this if and only if condition that $\mathrm{S}_2^p \subseteq \Sigma_2^p \cap \Pi_2^p$. In fact S_2^p consists of precisely those languages in $\Sigma_2^p \cap \Pi_2^p$ where membership in both Σ_2^p and Π_2^p is demonstrated by the same predicate P.

Canetti [10] defined the class S_2^p as follows: $L \in \mathrm{S}_2^p$ iff there is a P-time computable 0-1 function P on three arguments, such that for all x,

$$\left(\exists^p y\right)\left(\forall^p z\right) \quad \left[P(x,y,z)=\chi_L(x)\right]$$

and

$$\left(\exists^p z\right)\left(\forall^p y\right) \quad \left[P(x,y,z)=\chi_L(x)\right],$$

where χ_L is the characteristic function of L.

Clearly the Canetti definition implies the Russell–Sundaram definition. The reverse implication also holds. For completeness we sketch a simple proof (see [10,24] for more details). Suppose a predicate P is given in the Russell–Sundaram definition. We define an extended predicate $\hat{P}$ to satisfy the Canetti definition. For x, suppose y and z vary over $\{0,1\}^n$. Then $\hat{P}$ is defined over $\{0,1\}^{|x|} \times \{0,1\}^{n+1} \times \{0,1\}^{n+1}$:

$$\begin{aligned} &\hat{P}(x,1y,1z)=1, \\ &\hat{P}(x,1y,0z)=P(x,y,z), \\ &\hat{P}(x,0y,1z)=P(x,z,y), \\ &\hat{P}(x,0y,0z)=0. \end{aligned}$$

This can be rephrased in the language of boolean matrices. Thus, for the Russell–Sundaram definition, the predicate P, for a given x, corresponds to a boolean matrix M whose rows and columns are indexed by y and $z \in \{0,1\}^n$, respectively. When $x \in L$, there exists an all-1 row; and when $x \notin L$, there exists an all-0 column. In this language, the Canetti definition requires that, when $x \in L$, there exist both an all-1 row as well as an all-1 column; and when $x \notin L$, there exist both an all-0 row as well as an all-0 column.

To go from the Russell–Sundaram definition to the Canetti definition, we simply take the matrix M from the Russell–Sundaram definition, and form the new matrix

$$\begin{pmatrix} 0 & M^T \\ M & J \end{pmatrix},$$

where J denotes the all-1 matrix, and M^T denotes the transpose of M.

ZPP denotes zero-error probabilistic polynomial time. $\mathrm{ZPP}^{\mathrm{NP}}$ is the class accepted by zero-error probabilistic polynomial time oracle Turing machines using an NP oracle. By Cook's Theorem, we can assume without loss of generality that this oracle is the set of satisfiable boolean formulae **SAT**.

3. Main theorem

To prove the main Theorem 1, we proceed as follows. Let x be given. Let $\{0,1\}^n$ be the witness sets for both provers Y and Z. Here n is polynomially bounded by $|x|$, and is a power of 2.

We will grow a list $Y_k \subset \{0,1\}^n$ of y's, where $|Y_k| = k$, and $k = 1, 2, \ldots, n^{O(1)}$; initially Y_1 can be arbitrarily given, for example $Y_1 = \{0^n\}$. In the kth stage, with Y_k in hand, we ask the **SAT** oracle whether there exists a $z \in \{0,1\}^n$ such that $P(x, y, z) = 0$ for every $y \in Y_k$, i.e., a z that beats every $y \in Y_k$. Let

$$Z(Y_k) = \big\{z \in \{0,1\}^n \;\big|\; (\forall y \in Y_k)\,\big[P(x, y, z) = 0\big]\big\}.$$

Then the question we ask the **SAT** oracle is whether $Z(Y_k) \neq \emptyset$.

Since $|Y_k| = k$ is polynomially bounded, this is clearly a **SAT** query by Cook's Theorem. If the answer is No, i.e., $Z(Y_k) = \emptyset$, then we can already conclude that $x \in L$ and halt. This is because if it were the case that $x \notin L$, by definition it is guaranteed that some z_0 exists beats *all* y, which certainly include all $y \in Y_k$. Note that in this case we concluded $x \in L$, even though we may not have found a witness y_0 which beats every z as promised in the definition.

Hence let us assume the answer to the **SAT** query is Yes, i.e., $Z(Y_k) \neq \emptyset$.

Next we would like to append Y_k to Y_{k+1}. Our goal is, either to find conclusively that $x \notin L$, or to find a new y^* to be appended to the list Y_k so that the corresponding $Z(Y_{k+1})$ is shrunk significantly.

More precisely, we would like either to find conclusively $x \notin L$, or to find with high probability a new y^* such that $|Z(Y_{k+1})| \leqslant |Z(Y_k)|/2$, where $Y_{k+1} = Y_k \cup \{y^*\}$. If so, we would guarantee that the size $|Z(Y_k)|$ shrinks geometrically every step by a constant fraction with high probability, and thus in polynomial time with high probability we either find out $x \notin L$, or we end up in the case with $Z(Y_k) = \emptyset$, in which case we can conclude that $x \in L$ as discussed earlier.

Lemma 1. *For every set S in* P*, there is a probabilistic sampling procedure A using a* **SAT** *oracle, such that for every n, and for every $0 < \varepsilon < 1$, $A(n, \varepsilon)$ samples at most $O(n/\varepsilon)$ elements $S' \subseteq S^{=n} = S \cap \{0,1\}^n$ in such a way that, for every subset $T \subseteq S^{=n}$, with $|T| > \varepsilon |S^{=n}|$,*

$$\Pr[S' \cap T = \emptyset] \leqslant \frac{1}{2^{2n}}.$$

The algorithm runs in time $(n/\varepsilon)^{O(1)}$.

We will discuss Lemma 1 in the next section. For now we assume Lemma 1.

For any witness $y' \in \{0,1\}^n$, consider the set

$$T_{y'} := Z\big(Y_k \cup \{y'\}\big) = \big\{z \in Z(Y_k) \;\big|\; P(x, y', z) = 0\big\}.$$

We say that a $y' \in \{0,1\}^n$ is a "bad witness" with respect to $Z(Y_k)$ if

$$\big|T_{y'}\big| = \big|\big\{z \in Z(Y_k) \;\big|\; P(x, y', z) = 0\big\}\big| > \frac{|Z(Y_k)|}{2}.$$

That is, y' is a "bad witness" iff more than $1/2$ of $Z(Y_k)$ beat this y'. Thus for any fixed bad witness y', by Lemma 1 with $\varepsilon = 1/2$, we can sample a polynomial number of $z \in Z(Y_k)$, call the set Z', such that the probability

$$\Pr[Z' \cap T_{y'} = \emptyset] \leqslant \frac{1}{2^{2n}}.$$

Since there are at most 2^n bad witnesses,

$$\Pr\big[\big(\exists \text{ a bad witness } y' \in \{0,1\}^n\big)\; [Z' \cap T_{y'} = \emptyset]\big] \leqslant \frac{1}{2^n}.$$

Suppose now for every bad witness $y' \in \{0,1\}^n$, the sample set Z' has a non-empty intersection with $T_{y'} = Z(Y_k \cup \{y'\})$. That means that for every bad witness y', y' cannot beat all of Z'. With the polynomial sized set Z' in hand, we ask the **SAT** oracle once again whether there is a y which beats all these $z \in Z'$. Again this is a **SAT** query by Cook's Theorem. If the answer is No, then we know $x \notin L$ since otherwise there is a y which beats all $z \in \{0,1\}^n$, and certainly y beats all these $z \in Z'$. So we reject x and halt.

If the answer is Yes, we use self-reducibility of the **SAT** oracle to obtain one such y^*. Notice that by now there is no bad witness y' which can beat all of Z'. Thus this y^* is not a bad witness. This is true with probability $\geqslant 1 - 1/2^n$. We then define $Y_{k+1} = Y_k \cup \{y^*\}$. Then with high probability we have

$$\big|Z(Y_{k+1})\big| \leqslant \frac{|Z(Y_k)|}{2}.$$

As remarked earlier this gives our $\mathrm{ZPP}^{\mathrm{NP}}$ algorithm.

4. A sampling lemma

The Sampling Lemma 1 follows from the work of Jerrum, Valiant and V. Vazirani [17]. However Lemma 1 has a relatively simple proof based on universal hashing. We give a self-contained account in this section using the notion of *isolation* of Sipser [26] (see also [28]).

Consider a family of hash functions:

$$\left\{h_s\colon \{0,1\}^n \to \{0,1\}^k\right\}_{s\in\mathcal{S}}.$$

Recall that a family of hash functions is 2-universal if for every pair of distinct $x \neq y$ in $\{0,1\}^n$, and for every $\alpha, \beta \in \{0,1\}^k$, $\Pr_{s\in\mathcal{S}}[h_s(x) = \alpha \wedge h_s(y) = \beta] = 1/2^{2k}$, i.e., $h_s(x)$ and $h_s(y)$ are pair-wise independent and uniformly distributed when $s \in_R \mathcal{S}$. It is well known such a family of 2-universal hash functions exists and can be easily constructed with small sample space, e.g., $h_{a,b}(x) = ax + b$ and then truncate to k bits, where a, b and x range over a finite field $\mathbf{GF}[2^n]$.

Here is an outline of the proof of Lemma 1. First we will use hash functions and the **SAT** oracle to get an approximate count of the subset $S^{=n}$. We will use the notion of *isolation* of Sipser [26] for this. Using the **SAT** oracle we can decide if $S^{=n} = \emptyset$. If so then Lemma 1 is vacuously true (no subset T exists with $|T| > \varepsilon|S^{=n}|$). Suppose $S^{=n} \neq \emptyset$. Then we will devise a simple sampling strategy based on an estimate of the number of points with unique inverse images from $S^{=n}$ under a random hash function. The details follow.

Given $x \neq y$, we say x **collides** with y under h_s if $h_s(x) = h_s(y)$. For a subset $E \subseteq \{0,1\}^n$, we say that h_s **isolates** $x \in E$ iff x does not collide under h_s with any other element of E. The following lemma of Sipser is well known and follows from a simple probability estimate [26].

Lemma 2. *Let $E \subseteq \{0,1\}^n$, and let $\{h_s\colon \{0,1\}^n \to \{0,1\}^k\}_{s\in\mathcal{S}}$ be a family of 2-universal hash functions of cardinality 2^{2n} with $1 \leqslant k \leqslant n$. Then for all $m \geqslant k$,*

(1) *if $|E| \leqslant 2^{k-1}$ then*

$$\Pr_{s_1,\ldots,s_m\in_R\mathcal{S}}[\forall x \in E \text{ some } h_{s_i} \text{ isolates } x] \geqslant 1 - \frac{1}{2^{m-k+1}};$$

(2) *if $|E| > m2^k$ then*

$$\Pr_{s_1,\ldots,s_m\in_R\mathcal{S}}[\forall x \in E \text{ some } h_{s_i} \text{ isolates } x] = 0.$$

For our set $E = S^{=n}$, there is some k_e, where $1 \leqslant k_e \leqslant n$, such that $2^{k_e-1} \leqslant |E| \leqslant 2^{k_e}$. If we take every k in the range $1 \leqslant k \leqslant n+1$, and randomly pick $m = 4n$ hash functions $h_{s_1}, \ldots, h_{s_m} : \{0,1\}^n \to \{0,1\}^k$, with probability $\geqslant 1 - \frac{1}{2^{3n}}$, at least for $k = k_e + 1$, we would get *isolation*. For each k we ask the **SAT** oracle, whether the chosen set of $h_{s_1}, \ldots, h_{s_m}$ has the property that "$\forall x \in E$, one of h_i isolates x." Since there are only $m = 4n$ hash functions this is a **SAT** query. We pick the least k_0 such that the oracle confirms *isolation*. We abort if for no k the chosen hash functions achieve *isolation*. With probability $\geqslant 1 - \frac{1}{2^{3n}}$ we do not abort, and we get $k_0 \leqslant k_e + 1$. Also by the second part of Lemma 2, we know definitely $|E| \leqslant 4n2^{k_0}$.

Denote by $U = 4n2^{k_0}$. This is defined with probability $\geqslant 1 - \frac{1}{2^{3n}}$. Whenever k_0 is defined, U is an upper bound of $|E|$. Also, with probability $\geqslant 1 - \frac{1}{2^{3n}}$, U is defined and it is not too far from a lower bound of $|E|$,

$$\frac{U}{16n} \leqslant |E| \leqslant U.$$

Let $r = 2^{\lceil \log_2 1/\varepsilon \rceil}$, so that $1/\varepsilon \leqslant r < 2/\varepsilon$. Also $r \geqslant 2$ as $\varepsilon < 1$. Let $R = \{0,1\}^{k_0+\log_2 n+\lceil \log_2 1/\varepsilon \rceil+4}$. Then $|R| = 4rU$.

The sampling procedure can be summarized as follows: First we get an estimate U as described above. Then, for each $1 \leqslant i \leqslant 3n$, uniformly and independently choose a hash function $h_i : \{0,1\}^n \to R$. Now repeat the following $2^{10}r^2n^2$ times for each h_i: Uniformly and independently pick a target $\alpha \in R$. Ask the **SAT** oracle whether it has an inverse image from the set $E = S^{=n}$. Since S is in P, this is a **SAT** query. If $\alpha \in h_i(E)$, we use self-reducibility to get one inverse image. This inverse image is a sample point. We exit the "repeat" loop as soon as we obtain $4rn$ samples.

```
1. Get estimate U = 4n2^{k_0}
2. For i = 1, ..., 3n
3.     Randomly pick h_{s_i} : {0,1}^n → R with |R| = 4rU
4.     Repeat 2^{10} r^2 n^2 times steps 5 and 6
5.         Randomly pick α ∈ R
6.         Try to find an x ∈ E s.t. h_{s_i}(x) = α using SAT
7.     if found 4rn points, Goto 3 with i := i + 1.
```

Consider $3n$ hash functions $h_1, h_2, \ldots, h_{3n}$ uniformly and independently chosen. For any such h, define the random variable C to be the number of colliding pairs,

$$C = \sum_{\{x,y\}\subseteq E, x\neq y} \chi_{[h(x)=h(y)]}.$$

The expectation of C is

$$\mathbf{E}[C] = \sum_{\{x,y\}\subseteq E, x\neq y} \Pr\big[h(x)=h(y)\big] = \binom{|E|}{2}\frac{1}{|R|} < \frac{|E|}{8r}.$$

Hence by Markov's inequality

$$\Pr\big[C \geqslant \varepsilon|E|/4\big] \leqslant \frac{1}{2}. \tag{3}$$

We say a point $\alpha \in R$ is a *unique image* if there is a unique $x \in E$ such that $h(x) = \alpha$. Suppose $C \leqslant \varepsilon|E|/4$, then there can be at most $\varepsilon|E|/2$ many $x \in E$ involved in a collision, i.e., such that there exists some $y \neq x$, $y \in E$, $h(x) = h(y)$. At least $(1 - \varepsilon/2)|E| \geqslant |E|/2$ elements of E are mapped to a unique image. Also by assumption $|T| > \varepsilon|E|$, at least $\varepsilon|E|/2$ many elements from T are mapped to a unique image.

For each h_i, the sampling procedure will produce $O(n/\varepsilon)$ points in time $(n/\varepsilon)^{O(1)}$. The probability that the procedure fails to produce any point from T is bounded by the sum of probabilities of the following events:

(E1) One did not get a good estimate U; or else,
(E2) $\forall 1 \leqslant i \leqslant 3n$, the collision set for h_i is large: $|C_i| \geqslant \varepsilon|E|/4$; or else,
(E3) the first i for which the C_i is small, yet less than $4rn$ points from $h_i(E)$ are picked; or else,
(E4) for this i the first $4rn$ points from $h_i(E)$ all do not produce points from T.

We have seen $\Pr[E1] \leqslant 2^{-3n}$. Also, $\Pr[E2] \leqslant 2^{-3n}$ by (3).

For (E3), we use the following version of Chernoff Bound:

Chernoff Bound. For any $0 < p < 1$ and $0 < \delta \leqslant p(1-p)$, if $X_i, i = 1, \ldots, \ell$ are i.i.d Bernoulli 0-1 variables with $\Pr[X_i = 1] = p$, then

$$\Pr\left[\left|\sum_{i=1}^{\ell} X_i - p\ell\right| \geqslant \delta\ell\right] \leqslant 2e^{-\frac{\delta^2\ell}{2p(1-p)}}. \tag{4}$$

If $|C_i| \leqslant \varepsilon|E|/4$, then $|h_i(E)| \geqslant |E|/2 \geqslant U/32n$, thus a target α belongs to $h_i(E)$ has probability at least $|h_i(E)|/|R| \geqslant \frac{1}{2^7 rn}$. Thus in our case, $p \geqslant \frac{1}{2^7 rn}$, $\ell = 2^{10}r^2n^2$, and let $\delta = p/2$. Then a simple calculation gives

$$\Pr[E3] \leqslant 2e^{-rn} \leqslant 2e^{-2n}.$$

Finally for (E4), for this h_i, $\varepsilon|E|/2$ many elements from T are mapped to unique images, thus each time a random $\alpha \in h_i(E)$ is picked, it has probability at least $\varepsilon|E|/(2|h_i(E)|) \geqslant \varepsilon/2 \geqslant 1/2r$ to give a sample point from T. (If $\alpha \in h_i(T)$ is a unique image, then the self-reducibility procedure with **SAT** *will* produce a pre-image from T.) It follows that $\Pr[E4] \leqslant (1 - 1/2r)^{4rn} < e^{-2n}$.

Adding up all the error probabilities, we get

$$\Pr[S' \cap T = \emptyset] \leqslant \frac{1}{2^{2n}}.$$

The procedure as stated will produce $O(n^2/\varepsilon)$ points. (This is sufficient for our Theorem 1.) However, for each hash function h_i one can check whether the collision set C_i is approximately small probabilistically using **SAT**, and proceed to produce $4rn$ samples only for the first h_i for which the C_i is found small. The modified procedure produces only $O(n/\varepsilon)$ points in $(n/\varepsilon)^{O(1)}$ time. This completes the proof of Lemma 1.

5. In search of irrefutable proofs

Let $L \in \mathrm{S}_2^p$ be defined as in (1), (2). If $x \in L$, then there exists y that beats all z. We call such a y an *irrefutable proof* w.r.t. P. Similarly if $x \notin L$ there are *irrefutable proofs* w.r.t. P, namely any z which beats all y. We have shown that membership $x \in L$ is decidable in $\mathrm{ZPP}^{\mathrm{NP}}$. However in neither case have we produced, in general, an irrefutable proof.

Say $x \in L$, then one simple case is already problematic when we have a polynomial number of y_i's and according to **SAT** there are no z that beat all these y_i's. While this is sufficient to conclude that $x \in L$ (and hence an irrefutable proof y exists), it does not help in locating one such. Moreover, suppose it happens to be that most $y \in \{0,1\}^n$ beats most but not all $z \in \{0,1\}^n$ w.r.t. P, then our proof of Theorem 1 in fact will not find an irrefutable proof with high probability.

However, for any $L \in \mathrm{S}_2^p$, we *can* find an irrefutable proof w.r.t. *some* predicate also defining L.

Theorem 3. *For every $L \in \mathrm{S}_2^p$, there is a P-time predicate Q defining L, such that irrefutable proof w.r.t. Q can be found in* $\mathrm{ZPP}^{\mathrm{NP}}$.[1]

Given L defined via P, define Q as follows:

$$Q(x; y_1, \ldots, y_m; z_1, \ldots, z_m) = 1 \quad \Longleftrightarrow \quad \left|\left\{(i,j) \;\middle|\; 1 \leqslant i,\ j \leqslant m,\ P(x, y_i, z_j) = 1\right\}\right| > \frac{m^2}{2},$$

where x is the input to L, $y_i, z_j \in \{0,1\}^n$, the length $n = |x|^{O(1)}$ is determined by P, and $m = 7n$ or $7n+1$, whichever is odd.

It is clear that Q is defined symmetrically. Also Q defines L: if $x \in L$, one can take all y_i to be an irrefutable proof y w.r.t. P. The case $x \notin L$ is symmetric.

We claim that in $\mathrm{ZPP}^{\mathrm{NP}}$ we can find an irrefutable proof w.r.t. Q in the following strong sense: Suppose $x \in L$, it will find a sequence $y_1, \ldots, y_m$ such that $\forall z \in \{0,1\}^n$,

$$\left|\left\{i \;\middle|\; P(x, y_i, z) = 1,\ 1 \leqslant i \leqslant m\right\}\right| > m/2, \tag{5}$$

and symmetrically if $x \notin L$.

By symmetry, we assume $x \in L$, and have found out this is so in $\mathrm{ZPP}^{\mathrm{NP}}$. The sequence $y_1, \ldots, y_m$ is defined inductively. $y_1, \ldots, y_k$ defines $\{\mathcal{Z}_k\}_{k \geqslant 0}$, a sequence of partitions of $Z = \{0,1\}^n$. $\mathcal{Z}_k = \{Z_{k0}, Z_{k1}, \ldots, Z_{kk}\}$ consists of $k+1$ disjoint subsets of Z, where $Z_{k,i}$ consists of those z for which exactly i of $y_1, \ldots, y_k$ beat it. Formally, for $\mathcal{Z}_0$, let $Z_{00} = Z$. For $k \geqslant 1$, $\mathcal{Z}_k$ is defined as: $\forall z \in Z$, let

$$c_k(z) = c_{y_1,\ldots,y_k}(z) = \left|\left\{j \;\middle|\; P(x, y_j, z) = 1,\ 1 \leqslant j \leqslant k\right\}\right|;$$

then for $0 \leqslant i \leqslant k$,

$$Z_{k,i} = \left\{z \in Z \;\middle|\; c_k(z) = i\right\}.$$

Suppose $\mathcal{Z}_k$ and $y_1, \ldots, y_k$ have been defined. For any y, it divides $Z_{k,i}$ into two parts, $Z_{k,i}^{\epsilon} = \{z \in Z_{k,i} \mid P(x, y, z) = \epsilon\}$, for $\epsilon = 0, 1$. We want to choose $y = y_{k+1}$, so that $|Z_{k,i}^1| \geqslant \frac{3}{4}|Z_{k,i}|$, for all $0 \leqslant i \leqslant k$. Our y_{k+1}

[1] Technically $\mathrm{ZPP}^{\mathrm{NP}}$ is a language class, and thus not for search problems. However the slight abuse of notation is harmless here. The theorem says that a probabilistic P-time algorithm using **SAT** can find *some* irrefutable proof w.h.p. and it never produces a non-irrefutable proof.

will be chosen probabilistically, and we will argue that it satisfies this condition w.h.p. In other words, let $p_{k,i} = \frac{|Z^1_{k,i}|}{|Z_{k,i}|}$ (if $|Z_{k,i}| = 0$, we let $p_{k,i} = 1$), then we require that

$$p_{k,i} \geqslant 3/4 \tag{6}$$

for all $k \geqslant 0$ and $0 \leqslant i \leqslant k$. Note that $Z_{k+1,i+1} = Z^1_{k,i} \cup Z^0_{k,i+1}$, if $y = y_{k+1}$.

Lemma 3. *Let $\{\mathcal{Z}_k\}_{k\geqslant 0}$ be any sequence of partitions of Z, where each $Z_{k,i}$ is divided into a disjoint union $Z_{k,i} = Z^0_{k,i} \cup Z^1_{k,i}$ and $Z_{k+1,i+1} = Z^1_{k,i} \cup Z^0_{k,i+1}$. Suppose $p_{k,i}$ as defined above satisfy* (6)*, then*

$$Z_{m,0} = Z_{m,1} = \cdots = Z_{m,\lfloor \frac{m}{2} \rfloor} = \emptyset,$$

where $m = 7n$ or $7n + 1$, whichever is odd.

We will prove Lemma 3 after we complete the proof of Theorem 3 assuming the lemma.

With $\mathcal{Z}_k$ defined and $y_1, \ldots, y_k \in \{0,1\}^n$ in hand, we can apply Lemma 1 (with $\varepsilon = 3/4$) to each $Z_{k,i}$, $0 \leqslant i \leqslant k$, and probabilistically produce samples $Z'_{k,i} \subseteq Z_{k,i}$, where each $|Z'_{k,i}|$ is polynomially bounded, and such that

$$\Pr\left[\left(\exists y \in \{0,1\}^n\right) y \text{ beats all } Z'_{k,i},\ 0 \leqslant i \leqslant k, \text{ yet } \exists i,\ y \text{ beats at most } \frac{3}{4} \text{ of } Z_{k,i}\right] \leqslant 2^n \cdot (k+1) \cdot \frac{1}{2^{2n}}.$$

For polynomially bounded k, this is exponentially small.

Assume such y does not exist, then we can ask our **SAT** oracle to find a y_{k+1}, via self-reducibility, that beats all $Z'_{k,i}$, $0 \leqslant i \leqslant k$. Such y_{k+1} certainly exists since $x \in L$, and, since all such y beat at least $3/4$ of $Z_{k,i}$, (6) is satisfied with this y_{k+1} for all $0 \leqslant i \leqslant k$. Now it follows from Lemma 3 that the sequence $y_1, \ldots, y_m$ *is* an irrefutable proof w.r.t. Q in the strong sense of (5). Thus except with exponentially small probability $O(n^2/2^n)$ we find an irrefutable proof w.r.t. Q. One more query to **SAT** confirms this.

This completes the proof of Theorem 3 modulo Lemma 3, to which we turn next. Our proof of Lemma 3 will be probabilistic in nature. It should be pointed out that this use of probability has nothing to do with the probabilistic construction of $\mathcal{Z}_k$ in the proof of Theorem 3. The statement of Lemma 3 is completely deterministic.

We define an ensemble of r.v. $\{\tilde{c}_k(z)\colon\ z \in Z\}_{k\geqslant 0}$ where for each $k \geqslant 0$, the family $\{\tilde{c}_k(z)\colon\ z \in Z\}$ is i.i.d. and defined as follows: $\forall z \in Z$, $\tilde{c}_0(z) = 0$, and if $\tilde{c}_k(z) = i$ then $\tilde{c}_{k+1}(z) = i + 1$ or i with probability $p_{k,i}$ and $1 - p_{k,i}$, respectively. Let $\tilde{\mathcal{Z}}_k = \{\tilde{Z}_{k0}, \tilde{Z}_{k1}, \ldots, \tilde{Z}_{kk}\}$ be defined as follows: For $0 \leqslant i \leqslant k$,

$$\tilde{Z}_{k,i} = \left\{z \in Z \;\middle|\; \tilde{c}_k(z) = i\right\}.$$

We can show that

Claim. *The expectation $\mathbf{E}|\tilde{Z}_{k,i}| = |Z_{k,i}|$, for all $k \geqslant 0$ and $0 \leqslant i \leqslant k$.*

To prove this claim, we induct on k, the case $k = 0$ being trivial. Suppose the claim holds for k and for all $0 \leqslant i \leqslant k$. Consider $k + 1$ and $1 \leqslant i \leqslant k + 1$. The case $\mathbf{E}|\tilde{Z}_{k+1,0}| = |Z_{k+1,0}|$ follows from the rest, and the fact that the total cardinality is 2^n.

Denote by $\mathbf{E}_{\leqslant k}$ the expectation taken w.r.t. stages up to k. Since $|\tilde{Z}_{k+1,i}| = \sum_{z\in Z} \chi_{[z\in\tilde{Z}_{k+1,i}]}$, it follows that, for $1 \leqslant i \leqslant k + 1$,

$$\begin{aligned}
\mathbf{E}|\tilde{Z}_{k+1,i}| &= \sum_{z\in Z} \mathbf{E}[\chi_{[z\in\tilde{Z}_{k+1,i}]}] = \sum_{z\in Z} \mathbf{E}[\chi_{[z\in\tilde{Z}_{k,i}]} \cdot \chi_{[z\in\tilde{Z}_{k+1,i}]} + \chi_{[z\in\tilde{Z}_{k,i-1}]} \cdot \chi_{[z\in\tilde{Z}_{k+1,i}]}] \\
&= \sum_{z\in Z} \left\{\mathbf{E}_{\leqslant k}[\chi_{[z\in\tilde{Z}_{k,i}]}] \cdot (1 - p_{k,i}) + \mathbf{E}_{\leqslant k}[\chi_{[z\in\tilde{Z}_{k,i-1}]}] \cdot (p_{k,i-1})\right\} \\
&= (1 - p_{k,i})\mathbf{E}_{\leqslant k}|\tilde{Z}_{k,i}| + p_{k,i-1}\mathbf{E}_{\leqslant k}|\tilde{Z}_{k,i-1}| = (1 - p_{k,i})|Z_{k,i}| + p_{k,i-1}|Z_{k,i-1}| = |Z_{k+1,i}|.
\end{aligned}$$

We next define a second ensemble of r.v. $\{\underline{c}_k(z)\colon z \in Z\}_{k\geqslant 0}$, where again, for fixed $k \geqslant 0$, the family $\{\underline{c}_k(z)\colon z \in Z\}$ is i.i.d. and defined simply as the sum of k Bernoulli independent 0-1 variables with $p = 3/4$. More formally, $\underline{c}_k(z) = \sum_{j=1}^{k} I_j(z)$, where $I_j(z)$ are i.i.d. 0-1 variables with $\Pr[I_j(z) = 1] = 3/4$. Then $\underline{Z}_k = \{\underline{Z}_{k0}, \ldots, \underline{Z}_{kk}\}$ is defined:

$$\underline{Z}_{k,i} = \big\{z \in Z \bigm| \underline{c}_k(z) = i\big\}.$$

We can "realize" $\tilde{\mathcal{Z}}_k$ via $\underline{\mathcal{Z}}_k$ by a "nibbling" process. Note that $\underline{c}_0(z) = 0$, and $\underline{c}_{k+1}(z) = \underline{c}_k(z) + I_k(z)$. Define a third ensemble $\{c_k^*(z)\colon z \in Z\}_{k\geqslant 0}$ via $\underline{c}_k(z)$ as follows: $c_0^*(z) = 0$, and $c_{k+1}^*(z) = c_k^*(z) + I_k(z) + \Delta$, where the "nibble" Δ is a 0-1 r.v. dependent on $c_k^*(z)$ and $I_k(z)$: If $I_k(z) = 1$ then $\Delta = 0$, if $I_k(z) = 0$, and $i = c_k^*(z)$, then $\Delta = 1$ with probability $4p_{k,i} - 3$, and $\Delta = 0$ with probability $4(1 - p_{k,i})$. Note that $0 \leqslant 4p_{k,i} - 3 \leqslant 1$. Given $c_k^*(z)$, the combined effect of $I_k(z) + \Delta$ is a Bernoulli 0-1 variable taking value 1 with probability exactly $p_{k,i}$, independent for every z.

Thus $c_k^*(z)$ has exactly the same distribution as $\tilde{c}_k(z)$. While $\tilde{c}_k(z)$ is independent from $\underline{c}_k(z)$, $c_k^*(z)$ is highly correlated with $\underline{c}_k(z)$: $\forall z, \forall k$,

$$\underline{c}_k(z) \leqslant c_k^*(z).$$

Thus, $\forall z, k, \ell$,

$$\Pr\big[\tilde{c}_k(z) \leqslant \ell\big] = \Pr\big[c_k^*(z) \leqslant \ell\big] \leqslant \Pr\big[\underline{c}_k(z) \leqslant \ell\big].$$

For $\underline{c}_k(z)$, the Chernoff Bound (4) applies directly. Thus if $m \geqslant 7n$ and odd, we take $p = 3/4$ and $\delta = 1/4$ then a short calculation gives,

$$(\forall z) \quad \Pr\left[\underline{c}_m(z) \leqslant \left\lfloor \frac{m}{2} \right\rfloor\right] \leqslant 2e^{-\frac{7}{6}n}.$$

Thus,

$$\sum_{i=0}^{\lfloor \frac{m}{2} \rfloor} |Z_{m,i}| = \sum_{i=0}^{\lfloor \frac{m}{2} \rfloor} \mathbf{E}|\tilde{Z}_{m,i}| = \sum_{z\in Z}\sum_{i=0}^{\lfloor \frac{m}{2} \rfloor} \Pr[z \in \tilde{Z}_{m,i}] = \sum_{z\in Z} \Pr\left[\tilde{c}_m(z) \leqslant \left\lfloor \frac{m}{2} \right\rfloor\right] \leqslant 2^{n+1}e^{-\frac{7}{6}n} < 1.$$

But the cardinalities of the sets $Z_{m,i}$ are all non-negative integers, we must conclude that

$$Z_{m,0} = Z_{m,1} = \cdots = Z_{m,\lfloor \frac{m}{2} \rfloor} = \emptyset.$$

6. An implication for Karp–Lipton

There has been a lot of work on the general theme inspired by the Karp–Lipton Theorem. For example, Mahaney [23] showed that if the sparse oracle is itself in NP (i.e., NP has $\leqslant_T^p$-complete, not just $\leqslant_T^p$-hard sparse set) then PH collapses to Δ_2^p. Long [22] extended this to co-sparse oracles. Arvind et al. [6] showed that under the same assumption as in Karp–Lipton that **SAT** has small circuits then MA = AM. (See [15] for a survey.)

Suppose NP has polynomial size circuits. The Karp–Lipton Theorem says that the Polynomial-time Hierarchy collapses to $\Sigma_2^p \cap \Pi_2^p$. Sengupta [25] pointed out that the same proof collapses the Polynomial-time Hierarchy to S_2^p. To see this we recount the "book" proof, but this time phrase it in terms of provers Y and Z. We only need to show that $\Pi_2^p \subseteq \mathrm{S}_2^p$, then it follows that $\Pi_2^p \subseteq \mathrm{S}_2^p \subseteq \Sigma_2^p$ and hence they are all equal.

Let L be any language in Π_2^p. There is a normal form $L = \{x \mid (\forall^p y)\ (\exists^p z)\ [P(x, y, z)]\}$, where P is a P-time predicate. By Cook's Theorem, without loss of generality we can assume that it takes the form

$$L = \big\{x \bigm| \big(\forall^p s\big)\ [\phi_{x,s} \in \mathbf{SAT}]\big\},$$

where $\phi_{x,s}$ is a boolean formula computable in P-time from x and s. Let the size of $\phi_{x,s}$ be bounded by $p(|x|)$ for some polynomial $p(\cdot)$.

Now to show membership in S_2^p we receive two strings y and z, from provers Y and Z, respectively. We expect the string y to be a poly-size circuit for formulae of size up to $p(|x|)$. For a pair (y, z) we accept if and only if the circuit y says the boolean formula $\phi_{x,z}$ is satisfiable and by self-reducibility produced a satisfying assignment which satisfied it.

We note that there exists a relativized world where the Karp–Lipton Theorem cannot be improved to P^{NP} [14,29].

If one substitutes the predicate P in the definition (1), (2) of S_2^p by a predicate computable in NP∩co-NP, we get the class $S_2[NP \cap \text{co-NP}]$, and we can still prove the inclusion $S_2[NP \cap \text{co-NP}] \subseteq ZPP^{NP}$. Clearly $S_2^p \subseteq S_2[NP \cap \text{co-NP}]$. It is open whether any of the following containments

$$S_2^p \subseteq S_2[NP \cap \text{co-NP}] \subseteq ZPP^{NP}$$

is a proper containment. We note that under suitable hardness assumptions one can prove $P^{NP} = BPP^{NP}$ (see [19]) and thus under these assumptions the above classes all collapse to P^{NP}.

Acknowledgments

I thank Venkat Chakaravarthy, Oded Goldreich, Lane Hemaspaandra, Alex Russell, Uwe Schöning, Alan Selman and Samik Sengupta for interesting discussions and comments. I also thank the anonymous referee for very helpful comments.

References

[1] V. Arvind, J. Köbler, On pseudorandomness and Resource–Bounded measure, in: Proceedings of Conference on the Foundations of Software Technology and Theoretical Computer Science, in: Lecture Notes in Comput. Sci., vol. 1346, Springer-Verlag, 1997, pp. 235–249.
[2] V. Arvind, J. Köbler, Graph isomorphism is low for ZPP^{NP} and other lowness results, in: Proceedings of Annual Symposium on Theoretical Aspects of Computer Science (STACS), 2000, pp. 431–442.
[3] V. Arvind, J. Köbler, New lowness results for ZPP^{NP} and other complexity classes, J. Comput. System Sci. 65 (2) (2002) 257–277.
[4] L. Bábai, Trading group theory for randomness, in: Proceedings of the 17th Annual ACM Symposium on Theory of Computing (STOC), ACM, 1985, pp. 421–429.
[5] L. Bábai, S. Moran, Arthur–Merlin games: A randomized proof system, and a hierarchy of complexity classes, J. Comput. System Sci. 36 (2) (1988) 254–276.
[6] V. Arvind, J. Köbler, U. Schöning, R. Schuler, If NP has polynomial size circuits then MA = AM, Theoret. Comput. Sci. 137 (1995) 279–282.
[7] N. Bshouty, R. Cleve, S. Kannan, R. Gavaldà, C. Tamon, Oracles and queries that are sufficient for exact learning, in: Proceedings of the 17th Annual ACM Conference on Computational Learning Theory, ACM, 1994, pp. 130–139, J. Comput. System Sci. 52 (3) (1996) 421–433.
[8] M. Bellare, O. Goldreich, E. Petrank, Uniform generation of NP-witnesses using an NP-oracle, Inform. and Comput. 163 (2000) 510–526.
[9] Jin-Yi Cai, $S_2^p \subseteq ZPP^{NP}$, in: Proc. 42nd IEEE Symposium on Foundations of Computer Science (FOCS), 2001, pp. 620–628.
[10] R. Canetti, On BPP and the polynomial-time hierarchy, Inform. Process. Lett. 57 (1996) 237–241.
[11] S. Goldwasser, S. Micali, C. Rackoff, The knowledge complexity of interactive proofs, in: Proc. 17th ACM Symp. of Computing, ACM, Providence, RI, 1985, pp. 291–304, SIAM J. Comput. 18 (1) (1989) 186–208.
[12] S. Goldwasser, M. Sipser, Private coins versus public coins in interactive proof systems, in: Proceedings of the 18th Annual ACM Symposium on Theory of Computing (STOC), ACM, 1986, pp. 59–68; in: S. Micali (Ed.), Randomness and Computation; in: Adv. Comput. Res., vol. 5, JAI Press, Greenwich, 1989, pp. 73–90.
[13] O. Goldreich, D. Zuckerman, Another proof that BPP ⊆ PH (and more), in: Electronic Colloquium on Computational Complexity (ECCC), TR97-045, 1997.
[14] H. Heller, On relativized exponential and probabilistic complexity classes, Inform. Control 71 (3) (1986) 231–243.
[15] L.A. Hemachandra, M. Ogiwara, O. Watanabe, How hard are sparse sets?, in: Proceedings of Structure in Complexity Theory Conference, IEEE Comput. Soc. Press, 1992, pp. 222–238.
[16] J. Hopcroft, Recent directions in algorithmic research, in: Proceedings 5th GI Conference on Theoretical Computer Science, in: Lecture Notes in Comput. Sci., vol. 104, Springer-Verlag, 1981, pp. 123–134.
[17] M. Jerrum, L.G. Valiant, V.V. Vazirani, Random generation of combinatorial structures from a uniform distribution, Theoret. Comput. Sci. 43 (1986) 169–188.
[18] R. Karp, R. Lipton, Some connections between nonuniform and uniform complexity classes, in: Proceedings of the 12th ACM Symposium on Theory of Computing, ACM Press, April 1980, pp. 302–309;
An extended version has also appeared as: Turing machines that take advice, Enseign. Math. 28 (1982) 191–209.
[19] A. Klivans, D. van Melkebeek, Graph nonisomorphism has subexponential size proofs unless the polynomial-time hierarchy collapses, in: Proceedings of ACM Symposium on Theory of Computing (STOC), ACM, 1999, pp. 659–667.
[20] J. Köbler, O. Watanabe, New collapse consequences of NP having small circuits, in: ICALP, in: Lecture Notes in Comput. Sci., vol. 944, 1995, pp. 196–207, Journal version: SIAM J. Comput. 28 (1) (1998) 311–324.
[21] C. Lautemann, BPP and the polynomial hierarchy, Inform. Process. Lett. 17 (4) (1983) 215–217.
[22] T. Long, A note on sparse oracles for NP, J. Comput. System Sci. 24 (2) (1982) 224–232.
[23] S. Mahaney, Sparse complete sets for NP: Solution of a conjecture of Berman and Hartmanis, in: Proceedings of 21st IEEE Symposium of Foundations of Computer Science, 1980, pp. 54–60, J. Comput. System Sci. 25 (2) (1982) 130–143.
[24] A. Russell, R. Sundaram, Symmetric alternation captures BPP, Comput. Complexity 7 (2) (1998) 152–162, A preliminary version appeared in Technical Report MIT-LCS-TM-541, 1995.
[25] S. Sengupta, personal communications, 2000.

[26] M. Sipser, A complexity theoretic approach to randomness, in: Proceedings of ACM Symposium On Theory of Computing (STOC), ACM, 1983, pp. 330–335.
[27] L. Stockmeyer, The polynomial-time hierarchy, Theoret. Comput. Sci. 3 (1977) 1–22.
[28] L. Stockmeyer, The complexity of approximate counting, in: Proceedings of ACM Symposium On Theory of Computing (STOC), ACM, 1983, pp. 118–126 (preliminary version);
Journal version: On approximation algorithms for #P, SIAM J. Comput. 14 (4) (1985) 849–861.
[29] C.B. Wilson, Relativized circuit complexity, J. Comput. System Sci. 31 (1985) 169–181.
[30] D. Zuckerman, Simulating BPP using a general weak random source, Algorithmica 16 (4/5) (1996) 367–391.

Paper 10

Journal of Computer an System Sciences **58**, 280–296 (1999)
Article ID jcss.1998.1615, available online at http://www.idealibrary.com on IDEAL®

Sparse Hard Sets for P: Resolution of a Conjecture of Hartmanis

Jin-Yi Cai[1] and D. Sivakumar[2]

Department of Computer Science, State University of New York at Buffalo, 226 Bell Hall, Buffalo, New York 14260
E-mail: {cai, sivak-d}@cs.buffalo.edu

Received November 30, 1995; revised March 26, 1996

Building on a recent breakthrough by Ogihara, we resolve a conjecture made by Hartmanis in 1978 regarding the (non)existence of sparse sets complete for P under logspace many–one reductions. We show that if there exists a sparse hard set for P under logspace many–one reductions, then P = LOGSPACE. We further prove that if P has a sparse hard set under many–one reductions computable in NC^1, then P collapses to NC^1. © 1999 Academic Press

1. INTRODUCTION

A set S is called *sparse* if there are at most a polynomial number of strings in S up to each length n. Sparse sets have been the subject of study in complexity theory for the past 20 years, as they reveal the inherent structure and limitations of computation. Intuitively, a sparse set can be thought of as an encoding of a small amount of information. With this view in mind, the most central questions in the study of sparse sets have been the following:

> What does it mean computationally for a complexity class $\mathscr{C}$ to have a sparse hard set? Can sparse sets be hard or complete for interesting complexity classes such as P, NP, etc?

There are two primary motivations for studying the existence of sparse hard (or sparse complete) sets. The first motivation stems from the connection to nonuniform and Boolean circuit complexity. By a result attributed to A. Meyer (cf. [BH77]), the class of languages that are polynomial time Turing reducible (i.e., by Cook reductions) to a sparse set is precisely the class of languages with polynomial size circuits. Pippenger [Pip79] showed that this is the same as the class P/poly of languages that can be accepted with a polynomial amount of "nonuniform advice."

[1] Research supported in part by NSF Grants CCR-9057486 and CCR-9319093, and by an Alfred P. Sloan Fellowship.

[2] Research supported in part by NSF Grant CCR-9409104. Author's present address: Department of Computer Science, University of Houston, Houston, TX 77204-3475.

0022-0000/99 $30.00

Thus sparse sets serve as a link between uniform complexity theory, which is based on the Turing machine model, and nonuniform complexity theory, which is based on the Boolean circuit model.

Another major motivation for the study of sparse sets, and various reducibilities to them, is concerned with the isomorphism conjectures by Berman and Hartmanis. In 1976, they proved that all the natural NP-complete problems (such as those found in [GJ79]) are isomorphic under polynomial time computable functions [BH77]. Based on this evidence they conjectured that all NP-complete problems under polynomial time many–one reducibility (i.e., Karp reductions) are isomorphic under polynomial time computable bijections. Noting that the densities of any two polynomial time isomorphic sets are polynomially related and that all known NP-complete sets are exponentially dense, they also conjectured that there are no sparse complete sets for NP.

The Berman–Hartmanis isomorphism conjecture has generated a lot of research in this field. Building on earlier work by Fortune [For79], Mahaney [Mah82] showed that if NP has a sparse hard set under polynomial time many–one reducibility, then $P = NP$. This is the definitive result concerning the nonexistence of sparse complete sets for NP under Karp reductions. Note that if $P = NP$, then both conjectures concerning isomorphism and the nonexistence of sparse complete sets for NP are false. Regarding Cook reductions and the connection to circuit complexity, the famous result by Karp and Lipton [KL82], with a contribution by Sipser, showed that if NP has a sparse hard set under Cook reductions, or equivalently, if NP has polynomial size circuits then the polynomial hierarchy collapses to its second level $\Sigma_2^p = \Pi_2^p$. In the subsequent years, considerable research effort has been devoted to studying variations of this problem; we especially mention the results by Ogihara and Watanabe concerning bounded truth table reductions of NP to sparse sets [OW91]; see [HOW92] or [You92a, You92b] for a survey.

The role of sparse hard sets in complexity theory goes further than the connection to NP. In 1978, Hartmanis [Har78] studied the isomorphism question for sets complete for P under logspace many–one reducibility. He observed that all the known P-complete problems were isomorphic under logspace computable bijections and conjectured that all P-complete problems were isomorphic under logspace computable bijections. Similarly, he conjectured that there are no sparse complete sets for P under logspace many–one reductions [Har78]. It is this conjecture that we address in this paper.

The connection between reducibility to sparse sets and polynomial circuit complexity also carries over in an interesting way to the low-level setting. Cook [Coo85] has defined a notion of uniform NC^1 reducibility as a useful notion in studying completeness for complexity classes such as NL and L. This is the analogue of Turing reducibility in low-level complexity. Using Buss' theorem [Bus87, BCG+92] that the boolean formula value problem is in uniform NC^1, it can be shown that a language has a (nonuniform) circuit family of polynomial size and logarithmic depth, that is, the language is in nonuniform NC^1, if and only if it is reducible to a sparse set under uniform NC^1 reductions. This provides another incentive to the study of sparse hard sets for low-level complexity classes. For example, $P \subseteq$ nonuniform NC^1 if and only if there is a sparse hard set for P under

uniform NC^1 reductions. Notice that it is not even known if PSPACE is contained in nonuniform NC^1.

1.1. *Main Result*

The current paper resolves the 1978 conjecture of Hartmanis in the sense of Mahaney; namely we show that there are no sparse complete sets for P under logspace many–one reductions if $P \neq LOGSPACE$. Unlike the NP case, very little progress had been made on this conjecture until very recently. The only known related result until last year is due to Hemachandra, Ogihara, and Toda [HOT94]. They showed that if P has *polylogarithmically* sparse hard sets, then P = SC, the class of languages recognizable in simultaneous polynomial time and polylogarithmic space. Because of the assumption of *polylogarithmic* sparsity, the result leaves an exponential gap. Very recently, Ogihara [Ogi95] made substantial progress toward resolving the conjecture of Hartmanis. He showed that if there is a sparse set S that is hard for P under logspace many–one reductions, then $P \subseteq DSPACE[\log^2 n]$. Our work builds on the work of Ogihara.

The main result of this paper is the following: if there is a sparse set S that is hard for P under logspace many–one reductions, then $P = LOGSPACE$. In fact, we prove the stronger statement: if there is a sparse set S that is hard for P under many–one reductions, then the P-complete circuit-value problem can be solved by a logspace-uniform family of polynomial size, logarithmic depth circuits that make polynomially many parallel calls to the reduction. Consequently, if P has a sparse hard set under many–one reductions computable in logspace-uniform NC^1, then P equals logspace-uniform NC^1.

An interesting aspect of our work is that the techniques we employ are probabilistic and algebraic in nature and are influenced by the recent developments in derandomization techniques, especially constructions of small sample spaces, and the theory of finite fields. The proof of our first theorem begins with a crucial observation due to Ogihara. The main ingredient in the resulting simulation is the solution of a system of linear equations over a finite field. We first prove a probabilistic lemma of general interest. Under the assumption of the existence of a sparse set hard for P, we obtain an RNC^2 simulation of P. Using a "small-bias sample space" construction ([NN90, AGHP90]), we derandomize this algorithm to obtain an NC^2 simulation. Finally, exploiting additional algebraic properties of a closely related construction, we arrive at a Vandermonde system. We then solve the system using closed formulae involving the elementary symmetric polynomials over a certain field and discrete Fourier transforms. The final result is a collapse of P to logspace uniform NC^1. In fact, modulo the complexity of the reduction, the resulting simulation can be done in TC^0.

1.2. *Further Extensions*

The basic techniques involving derandomization and algebraic computation are rather powerful. There are already a number of extensions and many additional results. Those results are primarily concerned with various other reducibilities and

complexity classes. In [CS95], we combine techniques from this paper with the famous result of Immerman and Szelepcsényi [Imm88, Sze87], to resolve a similar conjecture made by Hartmanis concerning sparse hard sets for nondeterministic logspace. In joint work with A. Naik [CNS95], we use a number of additional techniques and extend the results to the case of truth-table and randomized reductions. For truth-table reductions, we exploit error-correcting properties of the small sample space construction and show, e.g., that if there exists a sparse hard set for P under bounded truth-table reductions, then $P = NC^2$. For randomized reductions, we use an algorithm of Goldreich and Levin [GL89] that recovers a linear function over $GF(2)$ by querying an erroneous oracle and show that if there exists a sparse hard set for P under randomized reductions with a two-sided error, then $P \subseteq RNC^1$. To handle superpolynomially sparse sets, we generalize Mulmuley's NC^2 algorithm [Mul87], combined with an idea of Chistov [Chi85], to compute the rank of a matrix over a finite field. As an indication of the effectiveness of our derandomization and algebraic techniques, we note that it took the research community 10 years to take the similar step from many–one reducibility in Mathaney's result for NP to bounded truth-table reducibility in Ogihara and Watanabe's theorem. Very recently, Van Melkebeek [Mel96] has extended the ideas of Section 5 to the case of truth-table reductions, and has shown that if there exists a sparse set hard for P under logspace bounded truth-table reductions, then P = LOGSPACE.

2. PRELIMINARIES

All our notations and definitions are standard. We denote by P the class of all languages recognizable in polynomial time by deterministic Turing machines; NP denotes the class of nondeterministic polynomial time languages. The class of all languages recognizable by deterministic Turing machines that use space no more than $O(\log n)$ is denoted LOGSPACE or L; the corresponding nondeterministic class is denoted by NL. In general, DSPACE$[s(n)]$ denotes the class of languages accepted by deterministic Turing machines, which, on inputs of length n, use space no more than $O(s(n))$.

For circuit and parallel complexity, we use the notation SIZE-DEPTH$[s(n), d(n)]$ to denote the class of languages accepted by a uniform family $\{C_n\}_{n=0}^{\infty}$ of bounded fan-in circuits of size $s(n)$ and depth $d(n)$ for inputs of length n. The criterion for uniformity of the circuit family is usually taken to mean that there is a deterministic space $(\log s(n))$-bounded transducer that, on input 0^n, outputs an encoding of the circuit C_n. The class NC^k is defined as SIZE-DEPTH$[n^{O(1)}, \log^k n]$, and $NC = \bigcup_k NC^k$. (Our NC^1 is logspace-uniform NC^1.) The randomized version of NC^k is denoted by RNC^k. The class AC^0 is defined to be the class of languages that are accepted by families of unbounded fan-in Boolean circuits of polynomial size and constant depth. The class $AC^0[\oplus]$ is defined to be the class of languages that are accepted by families of unbounded fan-in Boolean circuits of polynomial size and constant depth that are allowed to use PARITY gates in addition to AND, OR, and NOT gates. A PARITY gate, on inputs $y_1, y_2, ..., y_k$,

outputs 1 iff the number of y_i's that are 1 is odd. The class TC^0 is defined to be the class of languages that are accepted by families of unbounded fan-in Boolean circuits of polynomial size and constant depth that are allowed to use MAJORITY gates in addition to AND, OR, and NOT gates. A MAJORITY gate, on inputs $y_1, y_2, ..., y_k$, outputs 1 iff the number of y_i's that are 1 is at least $\lceil k/2 \rceil$. The same notion of logspace uniformity applies to the classes AC^0, $AC^0[\oplus]$, and TC^0.

For any language A, let $c_A(n) \doteq \|\{x \in A \mid |x| \leqslant n\}\|$ denote the *census function* for A. A is called (polynomially) *sparse* if $c_A(n)$ is bounded by a polynomial in n.

A Boolean circuit C is a directed acyclic graph with ℓ input nodes labeled 1, ..., ℓ, and one output node. The interior nodes, called *gates*, are labeled from the set $\{\neg, \wedge, \vee\}$, and are respectively called NOT, AND, and OR gates. On any input $x \in \{0, 1\}^n$, the output of each gate is defined in the natural way, including the gate that is the output of the circuit. The *circuit-value problem*, abbreviated *CVP*, of determining whether a Boolean circuit C outputs 1 on input x was shown by Ladner [Lad75] to be complete for P under logspace computable many–one reductions. Cook [Coo85] defined the notion of NC^1 reducibility, and notes that this problem is complete for P under NC^1 reductions. This reducibility is somewhat subtle technically, so we refer the reader to [Coo85] for details. However, we remark that a consequence of the completeness of *CVP* is that if $CVP \in NC^1$, then $P = NC^1$.

All logarithms in this paper are to the base 2.

3. AN RNC^2 SIMULATION

In this section, we consider the hypothesis that there is a polynomially sparse set S hard for P under logspace (or even NC^2) many–one reductions. Note that the sparse set S need not belong to P itself. (Thus our assumption is even weaker than P-*completeness* as stated in Hartmanis' conjecture.) The framework and basic ideas introduced here are used in all our results.

Following Ogihara [Ogi95], we define the set A of tuples $\langle C, x, I, b \rangle$, where C is a boolean circuit, x is an input to C, I is a subset of the gates, and b is a bit (0 or 1), such that the sum mod 2 of the values of the gates chosen in I from C on input x equals b, i.e.,

$$\bigoplus_{i \in I} g_i(x) = b.$$

Clearly, $A \in P$ and hence $A \leqslant_m^L S$. Let f be a logspace computable function such that for all x, $x \in A \Leftrightarrow f(x) \in S$. It is clear that $CVP \leqslant_m^L A$; we will show how to solve *CVP* in RNC^2 using the reduction f from A to S.

We note that for any C, x, I, exactly one of the bits $b = 0, 1$ satisfies the equation, and thus exactly one of $f(\langle C, x, I, 0 \rangle)$ and $f(\langle C, x, I, 1 \rangle)$ is a string in S. Moreover, suppose for two *distinct* subsets I and J and some pair of bits b, b', $f(\langle C, x, I, b \rangle) = f(\langle C, x, J, b' \rangle)$ (we are not assuming that the image is in S). In this case, regardless of whether $\bigoplus_{i \in I} g_i(x) = b$ and $\bigoplus_{i \in J} g_i(x) = b'$ are true

or not, they hold or fail simultaneously. Thus we have an equation mod 2 on the values of the gates of C on input x, namely

$$\bigoplus_{i \in I \triangle J} g_i(x) = b \oplus b',$$

and $I \triangle J \neq \varnothing$.

Fix any C and x, let n denote the number of nodes in C (including the inputs, output, and the interior gates). Let N denote the largest value of $|f(\langle C, x, I, b \rangle)|$ (over all I and b). Clearly N is polynomially bounded in n. Let $p(n)$ be a polynomial function that bounds $c_S(N)$. For notational simplicity we assume $p(n)$ is a power of 2; in particular, we will assume that $\log_2 p(n)$ is always an integer. Since there are only polynomially many strings in S, some string $w \in S$ must be mapped on by at least $2^n/p(n)$ many subsets I. More precisely: for $I \in \{0,1\}^n$, let $b_I = \bigoplus_{i \in I} g_i(x)$ denote the "correct value" of the parity of the gate values chosen by I, and for $w \in S$, define $T_w = \{I \in \{0,1\}^n \mid f(\langle C, x, I, b_I \rangle) = w\}$. Then there is at least one $w \in S$ such that $|T_w| \geqslant 2^n/p(n)$. We will call such w's *popular*.

As described above, any two I and J that that map to the same w give rise to an equation mod 2 on the values of the gates of C on input x. The idea now is to choose polynomially many *random* subsets $I \in \{0,1\}^n$ and compute $f(\langle C, x, I, 0 \rangle)$ and $f(\langle C, x, I, 1 \rangle)$, collecting an equation whenever a "collision" takes place. We remind the reader once again that whenever $f(\langle C, x, I, b \rangle)$ and $f(\langle C, x, J, b' \rangle)$ collide for $I \neq J$, irrespective of whether or not $b = b_I$ and $b' = b_J$ are true, equivalently, irrespective of whether the image is a member of S or not, the equation produced is valid.

The next question is: does the system of equations thus produced have sufficiently high rank, so that we may solve them to infer the g_i's? The following lemma ensures that this process gives us a system of linear equations of rank $n - O(\log n)$, even if we restrict attention to collisions that take place on a single popular w. Of course, we don't know which strings produced by the reduction are in S, but we do know that there *must* be at least one popular $w \in S$. Similarly, we don't know exactly what T_w is, but we know that it is large (by the popularity of w). Thus when the I's are picked at random, we can expect a nontrivial fraction of them to belong to T_w, and therefore, produce a significant number of collisions. The next lemma shows that as a consequence of the existence of a large T_w, the system of equations produced will have sufficiently large rank, with high probability.

3.1. *A Probabilistic Lemma*

Let $B = \{0,1\}^n$ denote the n-dimensional binary cube. With respect to the finite field of two elements $GF(2) = \mathbb{Z}_2$, B is a vector space of dimension n. Let $T \subseteq B$ be an arbitrary subset of the cube. We ask the following question: If we uniformly and independently pick a sequence of m points in B, what can we say about the probability distribution of the dimension of the affine span of those points picked from T as a function of m, n, and $|T|$?

LEMMA 1. *Let k be a power of 2. Suppose $|T| \geqslant 2^n/k$, where $k = n^{O(1)}$, then for $m = 2kn^2 + n + 1 = n^{O(1)}$, if we uniformly and independently pick a sequence of m points in B, the probability that the dimension of the affine span of the points from T is less than $n - \log_2 k$ is at most $e^{-n^2 + O(n \log n)}$.*

Proof. Consider any sequence of points of B being picked by the above process. Let us mark any such sequence $I_1, I_2, ..., I_m$ by a 0–1 sequence of the same length m, according to the following rule: Suppose the subsequence $I_{i_1}, I_{i_2}, ..., I_{i_\ell}$ is the intersection of the sequence $\{I_i\}$ with the set T. I_{i_1} is marked 0. For $j > 1$, precisely those points I_{i_j} are marked 1 if the dimension of the affine span of $I_{i_1}, I_{i_2}, ..., I_{i_j}$ is greater than that of $I_{i_1}, I_{i_2}, ..., I_{i_{j-1}}$. All other points in $\{I_i\}$ are marked 0. This defines a 0–1 sequence σ of length m. We wish to estimate the probability that the number of 1's in σ is small.

The process of uniformly and independently picking a sequence of m points in B induces a probability distribution over the set of 0–1 sequences σ of length m defined as above. Suppose we have picked a sequence $I_1, I_2, ..., I_{i-1}$ which intersects with T in a set whose affine span has dimension $< n - \log_2 k$. Then there are at least $|T| - 2^{n - \log_2 k - 1}$ points of T, which, if picked next, would increase the dimension of the affine span of the intersection. This cardinality is $\geqslant 2^n/k - 2^n/(2k) = 2^n/(2k)$. Hence, for $i > 1$ the conditional probability

$$\Pr[\sigma_i = 1 \mid \text{the number of 1's in } \sigma_1, ..., \sigma_{i-1} < n - \log_2 k] \geqslant 1/(2k).$$

For any sequence σ with strictly fewer than $n - \log_2 k$ many 1's,

$$\Pr[\sigma] \leqslant \left(1 - \frac{1}{2k}\right)^{m - (n - \log_2 k) - 1},$$

which is bounded above by e^{-n^2} if $m = 2kn^2 + n + 1$. Therefore,

$$\Pr[\dim(\text{affine span of } \{I_i\}_{i=1}^m \cap T) < n - \log_2 k] \leqslant \sum_{j < n - \log_2 k} \binom{m}{j} e^{-n^2} < e^{-n^2 + O(n \log n)}.$$

∎

Now we apply the above lemma with $T = T_w$ for some popular $w \in S$. It is clear that we obtain one new equation for each I that gives rise to a "1" in the sequence σ defined in the proof of the lemma. The lemma guarantees that if we try (in parallel) polynomially many uniformly and independently chosen I, with high probability we will obtain a system of linear equations with rank deficiency at most $\log_2 p(n)$. We now describe how we can use these to determine in NC^2 the outputs of all the gates of C on input x.

Without loss of generality, let the rank of the system be $n - \log_2 p(n)$, and let $m (= n^{O(1)})$ denote the number of equations we have. Denote the equations by $E_1, ..., E_m$, and for $i \geqslant 1$ call an equation E_i *useful* if the rank $\mathrm{rk}(E_1, ..., E_i) > \mathrm{rk}(E_1, ..., E_{i-1})$. Clearly the number of useful equations is at least $n - \log_2 p(n)$; without loss of generality, we will assume that we have exactly $n - \log_2 p(n)$ useful equations. Mulmuley [Mul87] gives an algorithm to compute the rank of an $\ell \times n$

matrix, which, for $\ell = n^{O(1)}$, can be implemented by a circuit of depth $O(\log^2 n)$ and size $n^{O(1)}$. For $1 \leqslant i \leqslant m$ we compute in parallel $\mathrm{rk}(E_1, ..., E_i)$ and identify all the useful equations. Now we have $n - \log_2 p(n)$ equations in n variables with rank $n - \log_2 p(n)$. We apply the same process to the columns and identify the $(n - \log_2 p(n))$-many useful columns. We rename the variables so that the first $n - \log_2 p(n)$ columns are all useful. For each of the $p(n)$ possible assignments to the last $\log_2 p(n)$ variables, we create in parallel a system of $n - \log_2 p(n)$ equations as an $(n - \log_2 p(n)) \times (n - \log_2 p(n))$ matrix. Each one of these can be solved in $\log^2 n$ depth and $n^{O(1)}$ size using the algorithm due to Borodin *et al.* [BvzGH82]. For each potential solution we get for the gates of the circuit C on input x, we can check its validity using the local information about the circuit C and input x, such as $x_i = 0$, or $x_i = 1$, or $g_j(x) = g_k(x) \wedge g_\ell(x)$. There will be a unique solution that passes all such tests and we will find the output of $C(x)$ in particular. We have proved

THEOREM 2. *If there is a sparse set that is hard for* P *under logspace or* NC^2 *many–one reductions, then* $\mathrm{P} \subseteq \mathrm{RNC}^2$.

4. DETERMINISTIC CONSTRUCTION

In this section, we use a small sample space construction due to Alon *et al.* [AGHP90], and generalize their result concerning the construction. We apply the generalization to derandomize the probabilistic simulation of Section 3. Under the hypothesis about sparse hard sets, this yields a collapse of P to NC^2.

As before we have $B = \{0, 1\}^n = \mathbb{Z}_2^n$ considered as an n-dimensional vector space over the finite field $\mathbb{Z}_2$. For each $I \in B$, let $b_I = \bigoplus_{i \in I} g_i(x)$ be the "right value." Then the string $w = f(\langle C, x, I, b_I \rangle) \in S$ and this w is called the color of I. The presumed reduction to the sparse set S gives a coloring of B with at most $p(n)$ colors. Let $D \subseteq B$ be a subset of B of cardinality bounded by a certain polynomial in n. The coloring of B induces a coloring of D, thus D is the union of at most $p(n)$ many color classes:

$$D = C_1 \cup C_2 \cup \cdots \cup C_{p(n)}.$$

Let the affine span of C_i be denoted by $L_i + d_i$, where L_i is a linear subspace, and d_i is a displacement vector. Let $L = L_1 + L_2 + \cdots + L_{p(n)}$ be the sum of the linear subspaces. We call L the span of the color classes. L_i is spanned by differences of vectors in C_i. For some spanning set of vectors of L_i, each vector in the set gives us an equation mod 2 of the values of the gates of C with the given input. If we collect a generating set of vectors for each L_i, together they span L. Thus, if we can construct a set D with polynomial size and with $\dim L \geqslant n - O(\log n)$ (irrespective of the coloring), we would have succeeded in derandomizing the construction of the last section. That is, by sampling exhaustively in D, we would have obtained a system of linear equations of rank $\geqslant n - O(\log n)$.

We claim that the above task can be accomplished as follows: given $p(n)$, construct a polynomial sized set D such that for any linear subspace M of B with

dim $M < n - \log_2 p(n)$, and any $p(n)$ displacement vectors $b_1, ..., b_{p(n)} \in B$, the union of the $p(n)$ affine subspaces $\bigcup_{i=1}^{p(n)} (M + b_i)$ does not cover the set D. For if so, then no matter what the induced coloring on D is, the span of the color classes L must be of dimension $\geqslant n - \log_2 p(n)$, simply because the union of at most $p(n)$ affine subspaces $\bigcup_{i=1}^{p(n)} (L + d_i)$ does cover D:

$$\bigcup_{i=1}^{p(n)} (L + d_i) \supseteq \bigcup_{i=1}^{p(n)} (L_i + d_i) \supseteq D.$$

Let $k = 1 + \log_2 p(n) = O(\log n)$. Without loss of generality, we may assume such a linear subspace M has dimension exactly $= n - k$. Any such M can be specified as the null space of a system of linear equations

$$a_{i1} x_1 + a_{i2} x_2 + \cdots + a_{in} x_n = 0,$$

where $i = 1, ..., k$, and the k vectors $\{(a_{i1}, a_{i2}, ..., a_{in}) \mid i = 1, ..., k\}$ are linearly independent vectors in B over $\mathbb{Z}_2$.

Let $m = 2k + \log_2 n + 1 = 2 \log_2 p(n) + \log_2 n + 3 = O(\log n)$. The Galois field $\mathbb{F} = GF(2^m)$ has a vector space structure over $GF(2)$ of dimension m. Choose any basis $\{e_1, ..., e_m\}$; then for $u = \sum_{i=1}^m u_i e_i$ and $v = \sum_{i=1}^m v_i e_i$ in $\mathbb{F}$, we can define an inner product by letting

$$\langle u, v \rangle = \sum_{i=1}^{m} u_i v_i$$

and doing all arithmetic over $\mathbb{Z}_2$.

The (multi)set D is a "small-bias sample space" [AGHP90], defined as

$$D = \{(\langle 1, v \rangle, \langle u, v \rangle, ..., \langle u^{n-1}, v \rangle) \mid u, v \in \mathbb{F}\}.$$

Note that $|D| = 2^{2m} = n^{O(1)}$. Now consider any nonzero vector $a = (a_0, a_1, ..., a_{n-1}) \in B$ and any $b \in \mathbb{Z}_2$. We wish to estimate the size of the intersection of D with the affine hyperplane $\sum_{i=0}^{n-1} a_i x_i = b$. Since the inner product $\langle \cdot, \cdot \rangle$ is bilinear over $\mathbb{Z}_2$ we have

$$\sum_{i=0}^{n-1} a_i \langle u^i, v \rangle = \left\langle \sum_{i=0}^{n-1} a_i u^i, v \right\rangle.$$

Let $q_a(X)$ denote the polynomial $\sum_{i=0}^{n-1} a_i X^i \in \mathbb{F}[X]$. If u is a root of the polynomial $q_a(X)$, then clearly the inner product $\langle \sum_{i=0}^{n-1} a_i u^i, v \rangle = 0$. Now suppose $u \in \mathbb{F}$ is not a root of $q_a(X)$; then $\sum_{i=0}^{n-1} a_i u^i = q_a(u)$ is a nonzero element in $\mathbb{F}$. It is easy to see that for any nonzero $w \in \mathbb{F}$,

$$\Pr_{v \in \mathbb{F}} [\langle w, v \rangle = 0] = 1/2.$$

Thus,

$$\Pr_{u,\, v \in \mathbb{F}}\left[\sum_{i=0}^{n-1} a_i \langle u^i, v\rangle = 0\right]$$
$$= \Pr_{u \in \mathbb{F}}[u \text{ is a root of } q_a(X)] + \Pr_{u \in \mathbb{F}}[u \text{ is not a root of } q_a(X)] \cdot 1/2.$$

But $q_a(X)$ is a nonzero polynomial of degree at most $n-1$; thus,

$$\Pr_{u \in \mathbb{F}}[u \text{ is a root of } q_a(X)] \leqslant \frac{n-1}{2^m}.$$

Collecting terms, we have

$$\frac{1}{2} \leqslant \Pr_{u,\, v \in \mathbb{F}}\left[\sum_{i=0}^{n-1} a_i \langle u^i, v\rangle = 0\right] \leqslant \frac{1}{2} + \frac{n-1}{2^{m+1}}.$$

In particular, if $m > \log_2 n$, both affine hyperplanes $\sum_{i=0}^{n-1} a_i x_i = 0, 1$ must intersect our set D. The bound above was shown in [AGHP90]; we strengthen it to handle $O(\log n)$ linearly independent equations.

In general, consider any k linearly independent equations $\sum_{j=0}^{n-1} a_{ij} x_j = b_i$, where $a_{ij}, b_i \in \mathbb{Z}_2$ and $i = 1, ..., k$. Denote this affine space by Π. Denote the point in D specified by u, v as $D(u, v)$. We wish to estimate the probability $\Pr_{u,\, v \in \mathbb{F}}[D(u, v) \in \Pi]$.

Let Q denote the following set of polynomials: $\{\sum_{i=1}^{k} \beta_i [\sum_{j=0}^{n-1} a_{ij} X^j] \mid \beta_i \in \mathbb{Z}_2,$ but not all $0\}$. We claim that the cardinality of Q is exactly $2^k - 1$, and none of the polynomials in Q is the zero polynomial. This follows from the fact that the vectors $(a_{i0}, ..., a_{i, n-1})$ are independent over $\mathbb{Z}_2$. Let $u \in \mathbb{F}$ be such that no polynomial in Q has u as a root. For such a u,

$$\sum_{j=0}^{n-1} a_{ij} \langle u^i, v\rangle = \left\langle \sum_{i=0}^{n-1} a_{ij} u^i, v \right\rangle = b_i,$$

$i = 1, ..., k$, is a linear equation system on (the m bits of) v with linearly independent coefficient vectors over $\mathbb{Z}_2$. (For otherwise, a nonzero linear combination of the coefficient vectors of v will be zero, which is precisely the same as u being a root of one of the polynomials in Q.) Thus, the conditional probability for v to satisfy this linear equation system is precisely $1/2^k$. However, since $|Q| = 2^k - 1$, and each polynomial in Q is nonzero and of degree at most $n-1$,

$$\Pr_{u \in \mathbb{F}}[u \text{ is a root of some polynomial in } Q] \leqslant (2^k - 1)(n-1)/2^m.$$

Letting $E(u)$ denote the event "u is a root of some polynomial in Q," and letting ρ denote $(2^k - 1)(n-1)/2^m$, we obtain

$$\begin{aligned}
&\Pr_{u,\, v \in \mathbb{F}} [\, D(u, v) \in \Pi\,] \\
&\quad = \Pr_{v \in \mathbb{F}} [\, D(u, v) \in \Pi \mid E(u)\,] \cdot \Pr_{u \in \mathbb{F}} [\, E(u)\,] \\
&\qquad + \Pr_{v \in \mathbb{F}} [\, D(u, v) \in \Pi \mid \neg E(u)\,] \cdot \Pr_{u \in \mathbb{F}} [\, \neg E(u)\,] \\
&\quad \leqslant 1 \cdot \rho + \frac{1}{2^k} \cdot 1 = \frac{1}{2^k} + \rho.
\end{aligned}$$

Similarly, we have

$$\begin{aligned}
&\Pr_{u,\, v \in \mathbb{F}} [\, D(u, v) \in \Pi\,] \\
&\quad = \Pr_{v \in \mathbb{F}} [\, D(u, v) \in \Pi \mid E(u)\,] \cdot \Pr_{u \in \mathbb{F}} [\, E(u)\,] \\
&\qquad + \Pr_{v \in \mathbb{F}} [\, D(u, v) \in \Pi \mid \neg E(u)\,] \cdot \Pr_{u \in \mathbb{F}} [\, \neg E(u)\,] \\
&\quad \geqslant 0 \cdot \rho + \frac{1}{2^k} (1 - \rho) = \frac{1}{2^k} - \frac{\rho}{2^k}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\left| \Pr_{u,\, v \in \mathbb{F}} [\, D(u, v) \in \Pi\,] - \frac{1}{2^k} \right| \\
&\qquad \leqslant \max\{\, \rho/2^k, \rho\} \\
&\qquad = \frac{(2^k - 1)(n - 1)}{2^m} \\
&\qquad < \frac{n}{2^{m-k}},
\end{aligned}$$

which by our choice of m and k is bounded above by $1/2^{k+1}$. Thus, in particular,

$$\Pr_{u,\, v \in \mathbb{F}} [\, D(u, v) \in \Pi\,] > 0.$$

Other than linear independence, the coefficient vectors and the right-hand side vector $b_1, ..., b_k$ in the definition of Π are arbitrary; the total number of the b vectors is $2^k = 2p(n) > p(n)$, and it follows that no linear subspace M of dimension $< n - \log_2 p(n)$ can cover the set D with some $p(n)$ displacements.

THEOREM 3. *If there is a sparse set S which is hard for* P *under* NC^2 *many–one reductions, then* $\mathrm{P} = \mathrm{NC}^2$.

5. THE FINALE: NC^1 SIMULATION

The collapse of $P = NC^2$ under the assumption about sparse sets does not suffice for our ultimate goal of settling the conjecture of Hartmanis. The bottleneck in the randomized and deterministic NC^2 algorithms of the previous sections is the solution of a system of linear equations over $GF(2)$. Whereas solving arbitrary systems of linear equations over finite fields seems to require NC^2, the deterministic construction used in the previous section is highly structured and is suggestive of Vandermonde matrices. In this section, we exploit this structure, together with an appropriate choice of the finite field, to obtain an optimal simulation via *closed formulae*. We show that if there is a sparse set S that is hard for P under many–one reductions computable in logspace, then $P = LOGSPACE$. In fact, we prove the stronger statement

THEOREM 4. *If a sparse set S is hard for* P *under many–one reductions, then the* P*-complete circuit-value problem can be solved by a logspace-uniform family of polynomial size, logarithmic depth circuits that make polynomially many parallel calls to the reduction.*

That is, modulo the complexity of the reduction to the sparse set, the resulting algorithm can be implemented by a uniform circuit of polynomial size and logarithmic depth. It follows that if the reduction is computable in logspace-uniform NC^1, then P equals logspace-uniform NC^1.

Proof. It is known that the polynomial $X^{2\cdot 3^\ell} + X^{3^\ell} + 1 \in \mathbb{Z}_2[X]$ is an irreducible polynomial over $\mathbb{Z}_2$ for all $\ell \geqslant 0$ [vL91]. In the following, by a finite field $GF(2^m)$, where $m = 2 \cdot 3^\ell$, we refer explicitly to the field $\mathbb{Z}_2[X]/(X^{2\cdot 3^\ell} + X^{3^\ell} + 1)$.

Let S be a sparse set hard for P under logspace-computable many–one reductions. As before, we will consider a refinement of the circuit-value problem. Define

$$L = \left\{ \langle C, x, 1^m, u, v \rangle \mid m = 2 \cdot 3^\ell, u, v \in GF(2^m), \sum_{i=0}^{n-1} u^i g_i = v \right\},$$

where C is a boolean circuit and x is an input to C and where $g_0, \ldots, g_{n-1}$ are 0–1 variables that denote the values of the gates of C on input x. Here exponentiation and summation are carried out in the finite field $GF(2^m)$. It is easy to see that $L \in P$, since all the required field arithmetic involved in checking $\sum u^i g_i = v$ can be performed in polynomial time.

Clearly $|\langle C, x, 1^m, u, v \rangle|$ is bounded polynomially in n and m. If f is a logspace-computable function that reduces L to S, the bound on the length of queries made by f on inputs of length $|\langle C, x, 1^m, u, v \rangle|$ is some polynomial $q(n, m)$. Let $p(n, m)$ be a polynomial that bounds the number of strings in S of length at most $q(n, m)$. We will choose the smallest m of the form $2 \cdot 3^\ell$ such that $2^m / p(n, m) \geqslant n$. It is clear that $m = O(\log n)$. Let $\mathbb{F}$ denote the finite extension $GF(2^m)$ of $GF(2)$.

Facts. We first collect some facts about implementing the basic operations of $\mathbb{F}$. The complexity of these operations is important in determining the size, depth, and the uniformity of the circuits that we build.

(1) Adding two elements $\alpha, \beta \in \mathbb{F}$ is just the bitwise exclusive-or of the representations of α and β, and can be done in depth $O(1)$. Adding $n^{O(1)}$-many elements can be done by a circuit of size $n^{O(1)}$ and depth $O(\log n)$, using the obvious recursive doubling strategy. The circuitry to perform these additions are clearly logspace-uniform.

(2) Multiplying two elements $\alpha, \beta \in \mathbb{F}$ can be done using $O(\log m) = O(\log \log n)$ space, or by a circuit of depth $O(\log m) = O(\log \log n)$ and size $m^{O(1)} = (\log n)^{O(1)}$, as follows: For $\gamma \in \mathbb{F}$, let $P_\gamma \in \mathbb{Z}_2[X]$ denote the polynomial whose coefficients are given by the bits of γ. Clearly, $\alpha \cdot \beta = (P_\alpha \cdot P_\beta) \bmod (X^m + X^{m/2} + 1)$. Each of the $2m-1$ coefficients of $P_\alpha \cdot P_\beta$ is the sum (in $\mathbb{Z}_2$) of at most m bits, and can be evaluated in $O(\log m)$ space, or by a circuit of size $O(m^2)$ and depth $O(\log m)$. Finally, implementing the "$\bmod(X^m + x^{m/2} + 1)$" operation on $P_\alpha \cdot P_\beta$ can be done easily in $O(\log m)$ space, or by a circuit of size $O(m)$ and depth $O(\log m)$.

(3) Finding a primitive element ω that generates the multiplicative group $\mathbb{F}^*$ of $\mathbb{F}$ can be done in logspace by exhaustive search. An element $\omega \in \mathbb{F}$ generates $\mathbb{F}^*$ iff the condition "$(\forall \alpha \in \mathbb{F}^*)(\exists i < 2^m)[\omega^i = \alpha]$" holds. The latter condition can be tested using $O(m) = O(\log n)$ space by maintaining two counters, one that runs through all elements α of $\mathbb{F}^*$ and another for the exponent i, and doing the multiplications as described in Fact (2). Note that finding a primitive element is part of the precomputation and does not have to be implemented in NC^1.

(4) Raising the generator ω to any power $i < 2^m$, or computing the discrete logarithm of any element with respect to ω, can be done by table look-up in depth $O(\log n)$. The tables themselves can be precomputed using $O(\log n)$ space.

(5) The following fact is less obvious and will be important: multiplying $k = n^{O(1)}$ elements of $\mathbb{F}$ can be done in $O(\log n)$ depth. Given elements $\alpha_1, \alpha_2, ..., \alpha_k$, first the discrete logarithms $\ell_1, \ell_2, ..., \ell_k$ of the k elements are computed with respect to the generator ω. By Fact (4), this can be done simultaneously in $O(\log n)$ depth and size $n^{O(1)}$. The next task is to add the k $O(\log n)$-bit integers $\ell_1, \ell_2, ..., \ell_k$, and reduce the sum modulo $2^m - 1$. The addition can be done in $O(\log n)$ depth using the folklore 3-to-2 trick, in the following manner. Divide the k integers into $\lceil k/3 \rceil$ groups of three integers each. By computing the "sum" and the "carry" parts of the addition separately, the three integers $\ell_i, \ell_{i+1}, \ell_{i+2}$ in the ith group can be converted into two integers ℓ'_i and ℓ''_i, such that $\ell_i + \ell_{i+1} + \ell_{i+2} = \ell'_i + \ell''_i$. Moreover, this can be accomplished in depth $O(1)$ simultaneously for all groups of three elements, thus producing a list of $2k/3$ elements whose sum equals the sum of the k elements $\ell_1, \ell_2, ..., \ell_k$. By recursively applying this idea, the sum of the k integers can be computed in depth $O(\log_{3/2} k) = O(\log n)$. Since the sum of the k integers is at most $k2^m = n^{O(1)}$, reducing the sum modulo $2^m - 1$ can be easily accomplished by a table look-up in depth $O(\log n)$. It is also clear that the look-up table can be precomputed in space $O(\log n)$. Finally, converting the discrete logarithm into the corresponding field element can be done by table look-up in depth $O(\log n)$.

Remarks. (1) Alternatively, we can take $\mathbb{F}$ to be the finite field $\mathbb{Z}/(a)$ for some prime number a that satisfies $a/p(n, \lceil \log_2 a \rceil) \geqslant n$. Our results are valid with either

choice of $\mathbb{F}$. The important point is that it should be possible to implement the above list of operations in NC^1. We prefer to retain $GF(2^m)$ because it is a natural outgrowth of the ideas from the previous section and because it simplifies exposition of the Boolean complexity of the operations.

(2) The operations listed in Facts (1), (2), and (4) can, in fact, be implemented in $AC^0[\oplus] \subseteq TC^0$. Multiplying $n^{O(1)}$ elements, as described in Fact (5), can also be implemented in TC^0 [CSV84]. Since the proof of Theorem 4 only requires these operations, our proof shows that if there is a sparse P-hard set under many–one reductions computable in logspace-uniform TC^0, then P = logspace-uniform TC^0.

Our parallel algorithm for *CVP* begins by computing $f(\langle C, x, 1^m, u, v\rangle)$ for all $u, v \in \mathbb{F}$. For every $u \in \mathbb{F}$, there is a unique element $v_u \in \mathbb{F}$ such that $\langle C, x, 1^m, u, v_u\rangle \in L$, and therefore, $f(\langle C, x, 1^m, u, v_u\rangle) \in S$. Since $2^m/p(n, m) \geqslant n$, there is at least one string $w \in S$ such that the number of u satisfying $f(\langle C, x, 1^m, u, v_u\rangle) = w$ is at least n. Of course, there could be many such w (not necessarily in S), and we do not know which w is a string in S. To handle this, we will assume that every w that has $\geqslant n$ preimages is a string in S and attempt to solve for the g_i's. As long as there is at least one $w \in S$ that has $\geqslant n$ preimages, one of the assumptions must be correct, and we will have the correct solution. Since we know the details of the circuit C, the solutions can be verified and the incorrect ones weeded out.

Assume, therefore, without loss of generality, that $w \in S$ has $\geqslant n$ preimages. Let $u_1, u_2, ..., u_n$ denote n of them, and let $v_1, v_2, ..., v_n$ denote the corresponding v_u's. The equations

$$1g_0 + u_j g_1 + u_j^2 g_2 + \cdots + u_j^{n-1} g_{n-1} = v_j$$

for $j = 1, 2, ..., n$ form an inhomogeneous system of linear equations, where the coefficients (u_j^i) form a Vandermonde matrix which we will denote by U. Since the u_j's are distinct elements of the field $\mathbb{F}$, the system $Ug = v$ has full rank. It remains to show how to solve this system of equations in logspace-uniform NC^1. Thus the proof of Theorem 4 is complete, modulo the following lemma, which is of general interest. ▮

LEMMA 5. *Let* $\mathbb{F} = GF(2^m)$, *where* $m = O(\log n)$ *and* m *is of the form* $2 \cdot 3^\ell$ *for some integer* $\ell \geqslant 0$. *Solving a system* $Ug = v$ *of* n *equations in* n *unknowns over the field* $\mathbb{F}$ *where* U *is a Vandermonde matrix of full rank over* $\mathbb{F}$, *can be done by an* $O(\log n)$*-space uniform circuit of size* $n^{O(1)}$ *and depth* $O(\log n)$.

Proof. Observe that an equation of the form $\sum_{j=0}^{n-1} g_j u^j = v$ can be viewed as specifying the value of the polynomial $G(u) \doteq \sum_{j=0}^{n-1} g_j u^j$ at the point $u \in \mathbb{F}$. With this viewpoint, our task is to infer the polynomial G, that is, to find the coefficients g_j of G. Clearly if we can evaluate $G(u)$ at n distinct points $u_1, ..., u_n \in \mathbb{F}$, then we can recover the coefficients g_j by Lagrange interpolation as

$$G(u) = \sum_{i=1}^{n} G(u_i)\, Q_i = \sum_{i=1}^{n} v_i Q_i,$$

where

$$Q_i = \frac{(u-u_1)\cdots(u-u_{i-1})(u-u_{i+1})\cdots(u-u_n)}{(u_i-u_1)\cdots(u_i-u_{i-1})(u_i-u_{i+1})\cdots(u_i-u_n)} = \prod_{k\neq i}\frac{(u-u_k)}{(u_i-u_k)}.$$

For $0 \leqslant j < n$, g_j is the coefficient of u^j in $G(u)$. Collecting the terms corresponding to u^j, we have

$$g_j = \sum_{i=1}^{n} (-1)^{i+1} \frac{v_i}{\prod_{k\neq i}(u_k-u_i)} P_{n-j-1}(u_1, ..., \hat{u}_i, ..., u_n).$$

Here $\hat{u}_i$ denotes that u_i is missing from the list $u_1, ..., u_n$, and P_k denotes the kth elementary symmetric polynomial, defined as

$$P_0(y_1, ..., y_\ell) = 1, \quad P_k(u_1, ..., y_\ell) = \sum_{\substack{I\subseteq[\ell]\\ |I|=k}} \prod_{i\in I} y_i, \quad k > 0.$$

By Facts (2), (4), and (5), computing $v_i/(\prod_{k\neq i}(u_k-u_i))$ in NC^1 is fairly straightforward. Hence it suffices to show how to compute the polynomials $P_k(u_1, ..., \hat{u}_i, ..., u_n)$, in logspace-uniform NC^1. A folklore theorem indicates that this can be done in nonuniform NC^1. For our application, however, the uniformity is crucial.

It is easy to see that for $y_1, ..., y_\ell \in \mathbb{F}$, $P_k(y_1, ..., y_\ell)$ equals $P_k(y_1, y_2, ..., y_\ell, 0, 0, ..., 0)$ for any number of extra zeroes. Let $r = |\mathbb{F}^*|$, the number of elements in the multiplicative group of $\mathbb{F}$. We will give an NC^1 algorithm to compute the elementary symmetric polynomial of r elements, not necessarily distinct, from the finite field $\mathbb{F}$. By appending $r-\ell$ zeros, we can then compute $P_k(y_1, y_2, ..., y_\ell)$.

For $0 < k \leqslant r$, the value of the elementary symmetric polynomial $P_k(y_1, y_2, ..., y_r)$ is the coefficient of X^{r-k} in $h(X) \doteq \prod_{i=1}^{r}(X+y_i)$. Note that, given any $\alpha \in \mathbb{F}$, $h(\alpha)$ can be evaluated in NC^1, by Facts (1) and (5).

If we write $h(X)$ as $\sum_{i=0}^{r-1} a_i X^i$, the coefficient $a_i = P_{r-i}(y_1, ..., y_r)$ for $0 \leqslant i < r$. The idea now is to choose α's carefully from $\mathbb{F}$, compute $h(\alpha)$, and compute the coefficients a_i by interpolation. If we choose ω to be a primitive element of order r in $\mathbb{F}^*$, the powers of ω, namely $1 = \omega^0, \omega^1, \omega^2, ..., \omega^{r-1}$, run through the elements of $\mathbb{F}^*$. For $0 \leqslant i < r$, let $b_i = h(\omega^i)$. The relationship between the pointwise values (b_i's) and the coefficients (a_i's) of $h(X)$ can be written as

$$\begin{pmatrix} b_0 \\ b_1 \\ \vdots \\ b_{r-1} \end{pmatrix} = \begin{pmatrix} 1 & \omega^0 & \omega^{0\cdot 2} & \cdots & \omega^{0\cdot(r-1)} \\ 1 & \omega^1 & \omega^{1\cdot 2} & \cdots & \omega^{1\cdot(r-1)} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & \omega^{r-1} & \omega^{(r-1)\cdot 2} & \cdots & \omega^{(r-1)\cdot(r-1)} \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_{r-1} \end{pmatrix}.$$

The above matrix, which we will denote by Ω, is the discrete Fourier transform matrix and a Vandermonde matrix. Since the powers of ω are all distinct, Ω is invertible, and one can compute the coefficients a_i by $(a_0, ..., a_{r-1})^{\mathrm{T}} = \Omega^{-1}(b_0, ..., b_{r-1})^{\mathrm{T}}$. The crucial advantage over the earlier Vandermonde system is

that with this particular choice of Ω, the matrix Ω^{-1} has a simple explicit form: the (i, j)th entry of Ω^{-1} is just $\omega^{-(i-1)(j-1)}$. Computing the coefficients of $h(X)$ is now simply a matrix–vector multiplication. This completes the proof of the lemma. ∎

COROLLARY 6. *If there is a sparse set S that is hard for* P *under logspace-computable many–one reductions, then* P = LOGSPACE.

COROLLARY 7. *If there is a sparse set S that is hard for* P *under many–one reductions computable in logspace-uniform* NC^1, *then* P *equals logspace-uniform* NC^1.

COROLLARY 8. *If there is a set S with census function bounded by* $2^{(\log n)^a}$ *that is hard for* P *under many–one reductions computable in space* $(\log n)^b$, *then* $P \subseteq DSPACE[(\log n)^{ab}]$.

ACKNOWLEDGMENTS

We thank Mitsu Ogihara for showing us his work in a Rochester–Buffalo joint complexity seminar. The resolution of Hartmanis' conjecture would not have been possible without the breakthrough of Ogihara. We also thank Allan Borodin for discussion of the circuit complexity of elementary symmetric polynomials and Dieter van Melkebeek for pointing out an error in the statement of Corollary 8 that appeared in a preliminary draft. We thank Len Adleman, Eric Allender, Steve Cook, Juris Hartmanis, Richard Karp, Ming Li, Steve Mahaney, Ashish Naik, Charlie Rackoff, Ken Regan, Alan Selman, Dieter van Melkebeek, and Umesh Vazirani for interesting discussions and comments. We thank an anonymous referee for helpful comments.

REFERENCES

[AGHP90] N. Alon, O. Goldreich, J. Hastad, and R. Peralta, Simple constructions of almost k-wise independent random variables, *in* "Proc. 31st Annual IEEE Symposium on Foundations of Computer Science, 1990," pp. 544–553.

[BCG+92] S. Buss, S. Cook, A. Gupta, and V. Ramachandran, An optimal parallel algorithm for formula evaluation, *SIAM J. Comput.* **21** (1992), 755–780.

[BH77] L. Berman and J. Hartmanis, On isomorphisms and density of NP and other complete sets, *SIAM J. Comput.* **6** (1977), 305–321. [A preliminary version appeared in STOC 1976]

[Bus87] S. Buss, The boolean formula value problem is in ALOGTIME, *in* "Proc. 19th Annual ACM Symposium on the Theory of Computing, 1987," pp. 123–131.

[BvzGH82] A. Borodin, J. von zur Gathen, and J. Hopcroft, Fast parallel matrix and GCD computations, *Inform. and Control* **52** (1982), 241–256.

[Chi85] A. L. Chistov, Fast parallel calculation of the rank of matrices over a field of arbitrary characteristic, *in* "Proc. 5th International Conference on Fundamental of Computation Theory," Lecture Notes in Computer Science, pp. 63–69, Springer-Verlag, New York/Berlin, 1985.

[CNS95] J. Cai, A. Naik, and D. Sivakumar, On the existence of hard sparse sets under weak reductions, *in* "Proc. Symposium on Theoretical Aspects of Computer Science," Lecture Notes in Computer Science, pp. 307–318, Springer-Verlag, 1996.

[CS95] J. Cai and D. Sivakumar, Resolution of Hartmanis' conjecture for NL-hard sparse sets, *in* "Proc. Third Annual International Computing and Combinatorics Conference (COCOON)," Lecture Notes in Computer Science, Vol. 1276, pp. 62–71, Springer-Verlag, New York/Berlin, 1997.

[CVS84] A. Chandra, L. Stockmeyer, and U. Vishkin, Constant-depth reducibility, *SIAM J. Comput.* **13** (1984), 423–439.

[Coo85] S. Cook, A taxonomy of problems with fast parallel algorithms, *Information and Control* **64** (1985), 2–22.

[For79] S. Fortune, A note on sparse complete sets, *SIAM J. Comput.* **8**, No. 3 (1979), 431–433.

[GJ79] M. R. Garey and D. S. Johnson, "Computers and Intractability: A Guide to the Theory of NP-Completeness," Freeman, New York, 1979.

[GL89] O. Goldreich and L. Levin, A hard-core predicate for all one-way functions, *in* "Proc. 21st Annual ACM Symposium on the Theory of Computing, 1989," pp. 25–32.

[Har78] J. Hartmanis, On log-tape isomorphisms of complete sets, *Theoret. Comput. Sci.* **7**, No. 3 (1978), 273–286.

[HOT94] L. Hemachandra, M. Ogiwara, and S. Toda, Space-efficient recognition of sparse self-reducible languages, *Comput. Complexity* **4** (1994), 262–296.

[HOW92] L. Hemachandra, M. Ogiwara, and O. Watanabe, How hard are sparse sets, *in* "Proc. 7th Annual IEEE Conference on Structure in Complexity Theory, 1992," pp. 222–238.

[Imm88] N. Immerman, Nondeterministic space is closed under complementation, *SIAM J. Comput.* **17** (1988), 935–938.

[KL82] R. Karp and R. Lipton, Turing machines that take advice, *Enseig. Math.* **28**, No. 2 (1982), 191–209. [A prelimary version appeared in STOC 1980].

[Lad75] R. Ladner, The circuit value problem in log space complete for P, *SIGACT News* **7**, No. 1 (1975), 18–20.

[Mah82] S. Mahaney, Sparse complete sets for NP: Solution of a conjecture of Berman and Hartmanis, *J. Comput. System Sci.* **25**, No. 2 (1982), 130–143. [A preliminary version appeared in FOCS 1980.]

[Mel96] D. Van Melkebeek, Reducing P to a sparse set using a constant number of queries collapses P to L, *in* "Proc. 11th IEEE Conference on Computational Complexity Theory, 1996," pp. 88–96.

[Mul87] K. Mulmuley, A fast parallel algorithm to compute the rank of a matrix over an arbitrary field, *Combinatorica* **7**, No. 1 (1987), 101–104.

[NN90] J. Naor and M. Naor, Small-bias probability spaces: efficient constructions and applications, *in* "Proc. 22nd Annual ACM Symposium on the Theory of Computing, 1990," pp. 213–223.

[Ogi95] M. Ogihara, Sparse hard sets for P yield space-efficient algorithms, *in* "Proc. 36th Annual IEEE Symposium on Foundations of Computer Science, 1995," pp. 354–361.

[OW91] M. Ogiwara and O. Watanabe, On polynomial-time bounded truth-table reducibility of NP sets to sparse sets, *SIAM J. Comput.* **20**, No. 3 (1991), 471–483. [A preliminary version appeared in STOC 1990].

[Pip79] N. Pippenger, On simultaneous resource bounds, *in* "Proc. 20th Annual IEEE Symposium on Foundations of Computer Science, 1979," pp. 307–311.

[Sze87] R. Szelepcsényi, The method of forcing for nondeterministic automata, *Bull. EATCS* **33** (1987), 96–100.

[vL91] J. H. van Lint, "Introduction to Coding Theory," Springer-Verlag, 1991.

[You92a] P. Young, How reductions to sparse sets collapse the polynomial-time hierarchy: A primer (Part I), *SIGACT News* **23**, No. 3 (1992), 107–117.

[You92b] P. Young, How reductions to sparse sets collapse the polynomial-time hierarchy: A primer (Part II), *SIGACT News* **23**, No. 4 (1992), 83–94.

Paper 11

COMBINATORICA
Akadémiai Kiadó – Springer-Verlag

COMBINATORICA **12** (4) (1992) 389–410

AN OPTIMAL LOWER BOUND ON THE NUMBER OF VARIABLES FOR GRAPH IDENTIFICATION

JIN-YI CAI*, MARTIN FÜRER† and NEIL IMMERMAN‡

Received November 28, 1988
Revised November 14, 1991

In this paper we show that $\Omega(n)$ variables are needed for first-order logic with counting to identify graphs on n vertices. The k-variable language with counting is equivalent to the $(k-1)$-dimensional Weisfeiler–Lehman method. We thus settle a long-standing open problem. Previously it was an open question whether or not 4 variables suffice. Our lower bound remains true over a set of graphs of color class size 4. This contrasts sharply with the fact that 3 variables suffice to identify all graphs of color class size 3, and 2 variables suffice to identify almost all graphs. Our lower bound is optimal up to multiplication by a constant because n variables obviously suffice to identify graphs on n vertices.

1. Introduction

In this paper we show that $\Omega(n)$ variables are needed for first-order logic with counting to distinguish a sequence of pairs of graphs G_n and H_n. These graphs have $O(n)$ vertices each, have color class size 4, and admit a linear time canonical labeling algorithm. This contrasts sharply with results in [10,27] where it is shown that two variables suffice to identify all trees and almost all graphs, and that three variables suffice to identify all graphs of color class size 3.

Another way to interpret our results is with stable colorings of k-tuples of vertices. The work of Weisfeiler and Lehman [40,39] on combinatorial and group theoretic properties of colored graphs, has inspired the idea of separating the orbits of the automorphism group of a graph by coloring k-tuples of vertices. Sometimes, this approach is called, the k-dimensional Weisfeiler–Lehman method (k-dim W–L). In the late seventies and early eighties, this method was developed by many researchers, including Faradžev, Zemlyachenko, Babai, and Mathon. With $k=1$, this method gives a linear-time graph isomorphism algorithm that works for almost all graphs [10]. Furthermore, the fastest known general graph isomorphism algorithms make use of this method with $k = O(\sqrt{n})$ [11]. It had been conjectured that this method would provide a polynomial time graph isomorphism test at least for graphs of bounded valence. (Valence is a synonym for degree.) Our result disposes of such conjectures.

AMS subject classification code (1991): 03 B 10, 05 C 60, 05 C 85

*Research supported by NSF grant CCR-8709818.

†Research supported by NSF grant CCR-8805978 and Pennsylvania State University Research Initiation grant 428-45.

‡Research supported by NSF grants DCR-8603346 and CCR-8806308.

Up until now, most lower bounds in this area were proved using random graphs. This method does not work when counting is included in the language because as mentioned above, almost all graphs can be identified using only two variables with counting. In our construction we choose graphs T_n, $(n=1,2,\ldots)$ with $O(n)$ vertices and separator size n (Definition 6.3). Then we deterministically modify T_n producing a pair of non-isomorphic graphs G_n, H_n, which agree on all properties expressible with n variables. Our lower bound is linear in the separator size of the graphs T_n. This linear lower bound, combined with a straightforward upper bound (Proposition 7.3) allows us to precisely determine how many variables are needed to identify many classes of graphs in first-order logic, with or without counting.

This paper is organized as follows: In Section 2 we recount some of the history of the Weisfeiler–Lehman method. In Section 3 we give some background in descriptive complexity and explain the significance of this problem from the logical point of view. In Section 4 we introduce some combinatorial games and prove that they characterize logical equivalence in the languages we are considering. In Section 5 we prove the equivalence of the $(k-1)$-dimensional Weisfeiler–Lehman method and the k-variable language with counting. In Section 6 we use the above combinatorial game to prove the linear lower bound. Section 7 describes some corollaries and extensions of this work.

2. History of the Weisfeiler–Lehman Method

An old basic idea in graph isomorphism testing and canonical labeling is the naive vertex classification algorithm as described in Read and Corneil [37]. First, the vertices are labeled or colored with their valences. During the iteration, all labels are extended by the multiset ("set" with possibly multiple elements) of the labels of their neighbors. Between rounds, the labels are replaced by their order numbers in the lexicographic order of all the occurring labels. This always keeps the labels short. The algorithm stops when the set of labels stabilizes, meaning that no new differences between vertices are discovered. A labeling algorithm identifies a class of graphs, if all vertex properties which are invariant under isomorphisms are discovered. In other words, the sets of vertices with the same labels are the orbits of the automorphism group.

The naive vertex classification algorithm, which we want to call the one dimensional Weisfeiler–Lehman method (1-dim W–L), does not solve the worst cases of the graph isomorphism problem. Nevertheless, it is usually a good start, and in fact it succeeds most of the time. Babai, Erdős and Selkow [8] have shown that the 1-dim W–L algorithm already produces normal forms for all but an $n^{-1/7}$ fraction of the n-vertex graphs. This has been improved to a $c^{-n\log n/\log\log n}$ fraction [10] producing an average linear time canonical labeling algorithm by handling the few exceptions with a slow algorithm.

Vertex classification is probably the basis of every practical implementation of a graph isomorphism test. For example, this is the case for the "nauty" package [35], which is said to be the fastest practical graph isomorphism package. It should be mentioned, however, that in addition to vertex refinement, "nauty" makes extensive use of partial automorphism information in its backtrack process; and it is not clear

whether or not our examples may lead to graphs on which "nauty" requires excessive time. In general, it is quite difficult to construct "hard cases" for graph isomorphism.

There is a class of graphs for which the vertex classification algorithm alone is obviously useless, because the algorithm cannot even get started. These are the regular graphs, which have the same degree in each vertex. Here it seems quite natural to go beyond vertex classification to the 2-dim W–L or edge classification algorithm. Initially every ordered pair (u,v) is labeled or colored with one of three possible colors, depending on whether $u = v$ and whether there is an edge $\{u,v\}$. Then information about the multiset of pairs of colors assigned to paths of length 2 from u to v is repeatedly added to the color of (u,v). The algorithm stops when no color class is split any more. A modification of this algorithm has been shown to produce normal forms for all regular graphs in linear average time [29].

The k-tuple coloring algorithm (named k-dim W–L by Babai [13]) classifies k-tuples of vertices. It might color vertices and edges implicitly by using k-tuples with repetition of components. It could start with some encoding of the graph into the labels assigned to the k-tuples. For example, the initial label or color of every k-tuple could be the number of its distinct components except when this number is two. Then two colors could be used to encode the presence or absence of an edge between the two vertices. We prefer to get a quicker start by initially coloring each k-tuple with its isomorphism type. Repeatedly, the color of $(u_1,\ldots,u_k)$ is refined by the n element multiset (containing one element for each vertex v) of k-tuples of colors previously assigned to

$$(v,u_2,\ldots,u_k),(u_1,v,u_3,\ldots,u_k),\ \ldots\ ,(u_1,\ldots,u_{k-1},v)$$

We only need to consider k-tuples of distinct elements, if we finally color the vertices by the multiset of colors of incident k-tuples. A more formal description of the k-dim W–L method is given in Section 5.

Possibly weaker algorithms have been considered. We might call them *special k-dim W–L algorithms*. In Weisfeiler's book [39] only such a method is mentioned and called *deep stabilization*. It consists of individualization followed by a low dimensional (1-dim or 2-dim) W–L algorithm. For every distinguished (i.e., initially colored with unique colors) $(k-1)$-tuple or $(k-2)$-tuple, a 1-dim W–L or a 2-dim W–L respectively is performed to detect invariant properties of vertices or edges. These methods seem to be weaker than the standard k-dim W–L method.

Two of the current authors have independently reinvented the k-dim W–L algorithm in the early eighties and conjectured its capability of identifying the graphs of bounded valence (with k being a suitable function of the valence). Later, we have learned that such conjectures have been around before in the Soviet Union, where the k-dim W–L algorithm (or maybe sometimes a special k-dim W–L algorithm) has been investigated for two decades. Significant results have been obtained by Faradžev's group, which contributed many papers to Weisfeiler's book [39]. The Russians have built a huge algebraic theory with extensive applications around the notion of stable colorings of pairs. The key notion is that of a *cellular algebra* (see [39,30]), which has been discovered in another context and called *coherent configuration* by Higman [21].

Weisfeiler and Lehman have asked whether the special k-dim W–L method with a slowly growing value of k would be sufficient to solve the graph isomorphism

problem. There was actually good reason to conjecture $k=O(\log n)$ or even $O(1)$ to be sufficient.

The second hope was partly based on the following result of Cameron [14], obtained independently by Gol'fand (cf. [19,31]). Let us call a graph *k-regular*, if the number of common neighbors of a k element subset of vertices only depends on the isomorphism type of the subgraph induced by the k vertices. (1-regular and 2-regular graphs are well known as *regular* and *strongly regular* graphs respectively.) Cameron and Gol'fand have shown that apart from the pentagon and the line graph of K(3,3), only the trivial examples of 5-regular graphs exist, namely the disjoint unions of complete graphs of equal size, and their complements (complete multipartite graphs). These graphs are homogeneous, i.e., all isomorphisms of their subgraphs extend to automorphisms. Therefore, they are immune to k-dim W–L refinements for any k: No refinement beyond the isomorphism type of k-tuples will follow. However, for any other graph, the Cameron-Gol'fand result assures us that the 5-dim W–L method will give at least some nontrivial partitioning of the 5-tuples.

Lipton [32] has proved that a special k-dim W–L method with a fixed k is sufficient for canonical labeling of trivalent (degree 3) graphs with arc-transitive automorphism groups. (An arc is an ordered pair of adjacent vertices.)

Support for the $k = O(\log n)$ conjecture has been provided by Gary Miller [36]. He has shown that for certain classes of strongly regular graphs and other combinatorial objects such as Latin squares $k = \log n$ is sufficient. Previously, such graphs have been considered to be difficult examples for isomorphism testing, because of their high degree of regularity and symmetry. The importance of symmetries for graph isomorphism testing has been pointed out by Babai and by Mathon [34] who showed that the graph isomorphism problem is equivalent to computing the order of automorphism groups of graphs.

Individualization followed by a low (1 or 2) dimensional refinement (i.e., the special W–L method) has produced pioneer results in the areas of bounded valence as well as general graph isomorphism and canonical labeling. Babai's technical report [3] started to use group theoretical algorithms to obtain provable upper bounds for isomorphism problems. Not only did he get his well known probabilistic polynomial time isomorphism test for graphs of bounded color class size, he also started the work on bounded valence graphs. Individualization of $k = \sqrt{n}(\log n)^c$ vertices splits a bounded valence graph into color classes of size at most $\sqrt{n}$ resulting in an $\exp(\sqrt{n}(\log n)^c)$ isomorphism test. Subsequently Luks [33] proved, using group theory to greater depth, that isomorphism for graphs of bounded valence is in polynomial time. Finally the canonical labeling problem for graphs of bounded valence has been solved in polynomial time by [11] and [18] independently.

Individualization followed by naive refinement has also been the tool used by Babai to handle strongly regular graphs [4] and primitive coherent configurations [6]. He used individualization of $k=2\sqrt{n}\log n$ vertices. Strongly regular graphs and more generally, coherent configurations are stable under 2-dim W–L. While strongly regular graphs are just undirected graphs, coherent configurations are edge-colored complete directed graphs. A coherent configuration is primitive if the diagonal has one color and all other colors define connected graphs. If a transitive automorphism group is primitive, then 2-dim W–L produces necessarily a primitive coherent configuration. For tournaments, the isomorphism problems of primitive and arbitrary coherent configurations are polynomial time equivalent [11].

The general graph isomorphism problem has been attacked by Zemlyachenko. The method is described in [5] and [41]. By individualization of $O(\sqrt{n})$ vertices and canonical edge-switching, he has been able to reduce the valence to $O(\sqrt{n})$. Combining this with the method of Luks [33], Zemlyachenko obtained the first interesting upper bound for general graph isomorphism [41] (cf. [5]). His bound is $\exp(n^{1-c})$ for some positive constant c. This has subsequently been improved by Babai and Luks [11] to $\exp(n^{1/2+o(1)})$.

Instead of measuring the reduction in the valence, one could ask about the effect of these methods on the color class size. Babai [7] has investigated this splitting power of Zemlyachenko's method combined with 2-dim W–L. The result is that individualizing $k = O(n^{2/3} \log n)$ vertices, and applying Zemlyachenko's method *and* the 2-dim W–L method, he obtains color classes which have $\leq k$ vertices in each connected component of the resulting graph.

3. Logical Background

In [23, 24, 25] one of us has pursued an alternate view of complexity theory in which the complexity of a problem is characterized in terms of the complexity of the simplest first-order sentences expressing the problem. For example, it is shown in [23] that the polynomial-time properties are exactly the properties expressible by first-order sentences iterated[1] polynomially many times:

Fact 3.1. [23]

$$P = \bigcup_{k=1}^{\infty} \mathrm{FO}(\leq)[n^k]$$

The notation $\mathrm{FO}(\leq)[n^k]$ denotes the set of properties describable by a very uniform sequence of sentences $\{\varphi_n\}$ such that each sentence φ_n has length $O(n^k)$ and has a bounded number of variables independent of n.[2] The symbol $\leq$ is included to emphasize the presence of a total ordering on the universe of the input structures. In [24] and in [38] it is also shown that this uniform sequence of formulas can be represented by a least fixed point operator (LFP) applied to a single formula. Thus,

$$P = \mathrm{FO}(\leq) + \mathrm{LFP} = \bigcup_{k=1}^{\infty} \mathrm{FO}(\leq)[n^k].$$

Fact 3.1 gives a natural language expressing exactly the polynomial-time properties of ordered graphs. Let a *graph property* be an order independent property of ordered graphs. One can ask the question,

Question 3.2. Is there a natural language for the polynomial-time graph properties?

Since the notion of "natural" is not well defined, some readers may prefer the more precise question:

---

[1] More precisely, the sentence expressing the property for structures of size n consists of a fixed block of restricted quantifiers written $p(n)$ times, followed by a fixed formula.

[2] In [23] the notation Var&Sz$[O(1), n^k]$ instead of FO$[n^k]$ was used.

Question 3.3. Is there a recursively enumerable listing of a set of Turing machines that accept exactly all the polynomial-time graph properties?

These questions were first asked with respect to database query languages [15]. See [28] for a discussion of the role of ordering in the database context.

We remark that should it be the case that graph canonization (i.e. given a graph return a canonical form such that two graphs are isomorphic iff their canonical forms are equal) is in polynomial time, then the answer to Question 3.3 is, "Yes." Thus a negative answer would imply that P is not equal to NP.

Previous to this paper, the only polynomial-time graph properties known not to be expressible in FO + LFP (without ordering) were "counting problems". For example, that a graph has an even number of edges is not expressible in FO + LFP. In [24] a language which we now call "FO + LFP + COUNT" was proposed as an answer to Question 3.2. This language describes two-sorted structures consisting of an unordered domain of vertices together with an edge predicate, plus an ordered domain of numbers. The domains are joined via counting quantifiers as in Section 3.2. We show in Corollary 7.1 that this language fails badly on certain linear time properties of graphs.

In [27] and [26] the exact number of variables needed to identify various classes of trees with and without counting, respectively, is determined. (Without counting this number increases linearly with the arity of the trees; with counting two variables suffice.) The question of how many variables are needed to identify various classes of graphs is interesting in its own right, and also has applications to temporal logic [26].

In the remainder of this section we explain the logical background we need. Some of this material is described in more detail in [27].

3.1. First-Order Logic

For our purposes, a *graph* will be defined as a finite first-order structure, $G = \langle V_G, E_G \rangle$. V_G is the universe, (the vertices). E_G is a binary relation on V_G, (the edges).

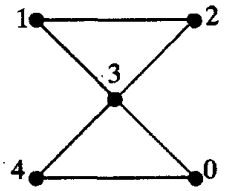


Fig. 1. An Undirected Graph

As an example, the undirected graph, $G_1 = \langle V_1, E_1 \rangle$, pictured in Figure 1 has vertex set $V_1 = \{0,1,2,3,4\}$, and edge relation

$$E_1 = \{\langle 0,3\rangle, \langle 0,4\rangle, \langle 1,2\rangle, \langle 1,3\rangle, \ldots, \langle 4,0\rangle, \langle 4,3\rangle\}$$

consisting of 12 pairs corresponding to the six undirected edges. By convention, we will assume that all structures referred to in this paper have universe $\{0,1,\ldots,n-1\}$ for some natural number n.

The *first-order language* of graph theory is built up in the usual way from the variables $x_1, x_2, \ldots$, the relation symbols E and $=$, the logical connectives $\wedge, \vee, \neg, \rightarrow$, and the quantifiers $\forall$ and $\exists$. The quantifiers range over the vertices of the graph in question. For example consider the following first-order sentence:

$$\varphi \equiv \forall x \forall y [E(x,y) \rightarrow E(y,x) \wedge x \neq y]$$

φ says that G is undirected and loop free. We will only consider graphs that satisfy φ, in symbols: $G \models \varphi$.

It is useful to consider a slightly more general set of structures. The *first-order language of colored graphs* results from the addition of a countable set of unary relations $\{C_1, C_2, \ldots\}$ to the first-order language of graphs.[3] Define a *colored graph* to be a graph that interprets these new unary relations so that all but finitely many of the predicates are false at each vertex. These unary relations may be thought of as colorings of the vertices.

Definition 3.4. For a given language $\mathcal{L}$ we say that the graphs G and H are *$\mathcal{L}$-equivalent* ($G \equiv_{\mathcal{L}} H$) iff for all sentences $\varphi \in \mathcal{L}$,

$$G \models \varphi \Leftrightarrow H \models \varphi.$$

We say that $\mathcal{L}$ *identifies* the graph G iff for all graphs H, if $G \equiv_{\mathcal{L}} H$ then G and H are isomorphic. $\mathcal{L}$ *identifies* a set of graphs S if it identifies every element of S.

Note: For the languages $\mathcal{L}_k, \mathcal{C}_k$ which we consider in this paper, and any graph G, the set of sentences in the language that are true about G has a polynomial size description which may be computed in polynomial-time [27]. Thus any set of graphs identified by $\mathcal{L}_k$ or $\mathcal{C}_k$ has a polynomial-time canonization algorithm.

Of course the First-Order Language of Colored Graphs identifies all colored graphs. From a computational viewpoint it is interesting to consider weaker languages admitting much faster equivalence testing algorithms.

3.2. The Languages $\mathcal{L}_k$ and $\mathcal{C}_k$

Define $\mathcal{L}_k$ to be the set of first-order formulas φ, such that the variables in φ are a subset of $x_1, x_2, \ldots, x_k$. Note that variables in first-order formulas are similar to variables in programs: they can be reused (i.e. requantified).

For example, consider the following sentence in $\mathcal{L}_2$.

$$\psi \equiv \forall x_1 \exists x_2 \Big(E(x_1, x_2) \wedge \exists x_1 [\neg E(x_1, x_2)] \Big)$$

The sentence, ψ, says that every vertex is adjacent to some vertex which is itself not adjacent to every vertex. As an example, the graph from Figure 1 satisfies ψ. Note that the outermost quantifier, $\forall x_1$ refers only to the free occurrence of x_1 within its scope.

Define a *color class* to be the set of vertices which satisfy a particular set of color relations. The *color class size* of a graph is the cardinality of its largest color class. In [27] it is shown that $\mathcal{L}_3$ identifies the set of graphs of color class size 3.

[3] Coloring relations are a clean tool for restricting the automorphisms of graphs. However, all the coloring relations in this paper could be replaced by simple gadgets in the graphs, without changing any of the results.

As noted above, the languages $\mathcal{L}_k$ are too weak to count, or even to express the parity of the number of edges. It is thus natural to strengthen these languages by adding *counting quantifiers* to the languages $\mathcal{L}_k$, thus obtaining the new languages $\mathcal{C}_k$. For each positive integer i, we include the quantifier $(\exists i\, x)$. The meaning of "$(\exists 17\, x_1)\varphi(x_1)$", for example, is that there exist at least 17 vertices such that φ. It is sometimes convenient to use the following abbreviation $(\exists! i\, x)$, meaning that there exists exactly i x's:

$$(\exists! i\, x)\varphi(x) \equiv (\exists i\, x)\varphi(x) \wedge \neg(\exists i+1\, x)\varphi(x)$$

As an example, the following sentence in $\mathcal{C}_2$ says that there exist exactly 17 vertices of degree 5,

$$(\exists! 17\, x_1)(\exists! 5\, x_2)E(x_1, x_2)$$

As an even worse example, the following sentence in $\mathcal{C}_2$ identifies the graph in Figure 1. It says that the whole graph contains exactly 5 vertices and that one vertex is adjacent to four vertices each of which has degree 2.

$$[(\exists! 5\, x_1)(x_1 = x_1)] \wedge [(\exists! 1\, x_1)(\exists 4\, x_2)(E(x_1, x_2) \wedge (\exists! 2\, x_1)E(x_2, x_1))]$$

Note that every sentence in $\mathcal{C}_k$ is equivalent to an ordinary first-order sentence with perhaps many more variables and quantifiers. In Section 5 it is shown that testing $\mathcal{C}_k$ equivalence corresponds to the $(k-1)$-dimensional Weisfeiler–Lehman Method. It thus follows that the language $\mathcal{C}_2$ identifies all trees and almost all graphs. In [27], TIME$(n^k \log n)$ algorithms for testing $\mathcal{L}_k$ or $\mathcal{C}_k$ equivalence of graphs on n vertices are presented.

4. Pebbling Games

We next describe two pebbling games that are equivalent to testing $\mathcal{L}_k$ and $\mathcal{C}_k$ equivalence, respectively. These games are variants of the games of Ehrenfeucht and Fraïssé, [16,17]. The results in this section concerning the $\mathcal{L}_k$ game and the $\mathcal{C}_k$ game originally appeared in [23] and [27], respectively.

Let G and H be two graphs, and let m and k be natural numbers. Define the m-move $\mathcal{L}_k$ *game* on G and H as follows. There are two players, and for each variable x_i, $i=1,\ldots,k$ there is a pair of x_i pebbles.

On each move, Player I picks up the pair of x_i pebbles, for some $i \in \{1,\ldots,k\}$, and he places one of them on a vertex in one of the graphs.[4] Player II must then place the other x_i pebble on a vertex of the other graph.

Define a *k-configuration* on a pair of graphs G, H to be a pair (u, v) of partial functions,

$$u : \{x_1, \ldots, x_k\} \to V_G;\ v : \{x_1, \ldots, x_k\} \to V_H$$

such that the domains of u and v are equal. We will use the notation D_u to denote the domain of the partial function u. Thus a k-configuration on G, H is a valid position of the $\mathcal{L}_k$ game on G, H. Here $u(x_i) = g$ means that an x_i pebble is on

[4] To make the play of the games easier to follow we will use masculine pronouns for Player I and feminine pronouns for Player II.

$g \in V_G$. If $x_i \notin D_u = D_v$ this means that the x_i pebbles are not currently placed on the board.

Let (u_r, v_r) be the configuration of the game after move number r. Then we say *Player I wins the game after move r* if the map that takes $u_r(x_i)$ to $v_r(x_i)$, $i \in D_{u_r}$, is not an isomorphism of the subgraphs induced by these vertices. (Note that if the graphs are colored then an isomorphism must preserve colors as well as edges.) We say that Player I wins the m-move game if for some $r \in \{0, 1, 2, \ldots, m\}$, Player I wins the game after move r. Player II wins iff Player I does not win. Finally, we say that Player II has a winning strategy for the $\mathcal{L}_k$ game on G and H, iff for all m, Player II has a winning strategy for the m-move game on G and H.

Thus Player II has a winning strategy for the $\mathcal{L}_k$ game just if she can always find matching vertices to preserve the isomorphism. [5] Player I is trying to point out a difference between the two graphs and Player II is trying to keep them looking the same.

The number of moves in the $\mathcal{L}_k$ game corresponds to the depth of nesting of quantifiers of the sentences in $\mathcal{L}_k$ needed to distinguish the graphs G and H. Define the language $\mathcal{L}_{k,m}$ to be the restriction of $\mathcal{L}_k$ to formulas of quantifier depth m. The relationship between the $\mathcal{L}_k$ game and the language $\mathcal{L}_k$ is given in Theorem 4.2. Before we state it, we need the following definition.

Definition 4.1. Let G, H be a pair of graphs and let (u, v) be a k-configuration on G, H. We will say that G, u is $\mathcal{L}_{k,m}$-*equivalent* to H, v, in symbols, $G, u \equiv_{\mathcal{L}_{k,m}} H, v$ iff for all formulas $\varphi \in \mathcal{L}_{k,m}$ whose free variables are a subset of D_u,

$$G, u \models \varphi \Leftrightarrow H, v \models \varphi$$

We omit u and v if they denote the nowhere defined partial function. Similarly we will say that G is $\mathcal{L}_k$-*equivalent* to H, in symbols, $G \equiv_{\mathcal{L}_k} H$ iff for all m, $G \equiv_{\mathcal{L}_{k,m}} H$.

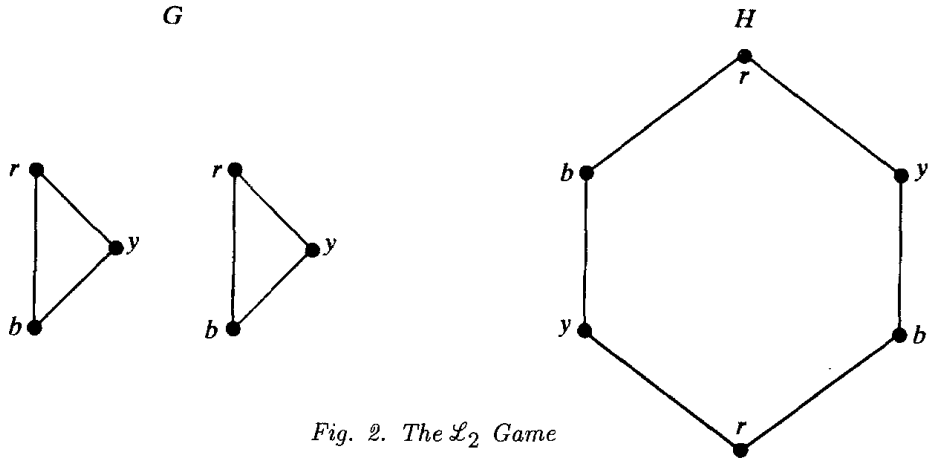


Fig. 2. The $\mathcal{L}_2$ Game

Theorem 4.2. [23] *Player II has a winning strategy for the m-move $\mathcal{L}_k$ game on G, H if and only if $G \equiv_{\mathcal{L}_{k,m}} H$. Thus, Player II has a winning strategy for the $\mathcal{L}_k$ game on G, H iff $G \equiv_{\mathcal{L}_k} H$.*

[5] By definition, the strategy can depend on the given number m of moves, but as G and H are finite, there is actually one strategy winning for all m.

Before we prove Theorem 4.2, we will give a few examples of the game.

Example 4.3. Consider the $\mathscr{L}_2$ game on the graphs G and H shown in Figure 2.

Suppose that Player I's first move is to place an x_1 pebble on a red vertex in G. Player II may answer by putting the other x_1 pebble on either of the red vertices in H. Now suppose Player I puts x_2 on an adjacent yellow vertex in H. Player II has a response because in G, every red vertex has an adjacent yellow vertex. The reader should convince himself or herself that in fact Player II has a winning strategy for the $\mathscr{L}_2$ game on the given G and H. It follows from Theorem 4.2 that G and H agree on all sentences from $\mathscr{L}_2$.

On the other hand, clearly Player I has a win in the 3-move, $\mathscr{L}_3$ game on G and H. He can simply put his pebbles on three points in one of the triangles in G. Since H has no triangle, Player II will lose. Notice that in this case Player I is playing the following sentence from $\mathscr{L}_{3,3}$ which is true of G and false of H:

$$(\exists x_1)(\exists x_2)(\exists x_3)[E(x_1,x_2) \wedge E(x_2,x_3) \wedge E(x_3,x_1)]$$

Finally, a more interesting example of an $\mathscr{L}_3$ game would be with H' consisting of a hexagon like H, but without the colors and G' consisting of a disjoint union of two copies of the hexagon H'. Here Player II has a winning strategy for the 3-move $\mathscr{L}_3$ game, but Player I has a winning strategy for the 4-move $\mathscr{L}_3$ game.[6] His strategy is to play the following sentence true of H' but not of G' saying that every pair of vertices is joined by a path of length at most three:

$$(\forall x)(\forall y)(\exists z)[(E(x,z) \vee x = z) \wedge (\exists x)(E(z,x) \wedge E(x,y))]$$

In order to prove Theorem 4.2 we need the following

Lemma 4.4. *For any relational language with finitely many relation symbols, and any k and m there are only finitely many formulas up to equivalence in $\mathscr{L}_{k,m}$. Furthermore if we include a finite set of additional quantifiers, e.g. counting quantifiers, then the expanded language still only has finitely many formulas up to equivalence.*

Proof. This is easy to see by induction on m. When $m=0$, there are only finitely many variables and only finitely many relation symbols, so only finitely many sets of possible facts about these variables. Assume that there are a total of f different kinds of quantifiers. Inductively, assume there are s_m inequivalent formulas of quantifier depth m. Then there are no more than $2^{2^{fks_m}}$ inequivalent formulas of quantifier depth $m+1$. ∎

Proof of Theorem 4.2. We prove by induction on m, that for all k-configurations (u,v) on G,H the following statements are equivalent:

1. Player II has a winning strategy for the m-move $\mathscr{L}_k$ game on G,H starting from the initial configuration (u,v).
2. $G,u \equiv_{\mathscr{L}_{k,m}} H,v$

[6] The reason we removed the colors is that with the colors there is a sentence from $\mathscr{L}_{3,3}$ distinguishing the two graphs, namely in H every pair of vertices of different color is joined by a path of length at most two.

The base of the induction is immediate because the map from $u(D_u)$ to $v(D_u)$ is an isomorphism of the induced subgraphs iff G,u and H,v agree on all quantifier-free formulas.[7]

Assume the equivalence of (1) and (2) for all m-move games, and let (u,v) be the initial configuration of an $(m+1)$-move game. Assume condition (2) is false and let $\varphi \in \mathcal{L}_{k,m+1}$ be a formula on which G,u and H,v disagree. If φ is a disjunction, conjunction, or negation of smaller formulas then G,u and H,v must disagree on one of these smaller formulas, so we may assume that φ begins with a quantifier. We may assume by symmetry that $\varphi = (\exists x_i)\psi$ and $G,u \models \varphi$, but $H,v \models \neg\varphi$. Player I should then place one of the x_i pebbles on a vertex g such that ψ holds in $G,u(x_i/g)$. No matter what vertex h Player II answers with, we know that $\neg\psi$ will hold in $H,v(x_i/h)$. Letting $(u_1,v_1) = (u(x_i/g),v(x_i/h))$ be the configuration after this move we have that $G,u_1 \not\equiv_{k,m} H,v_1$. Thus by induction Player I has a winning strategy for the remaining m-move game and thus for the original $m+1$-move game.

Conversely, assume that condition (2) is true and let Player I's first move be to place one of the x_i pebbles on some vertex g from G. Let $u_1 = u(x_i/g)$ be the result of this move. Note that there are only finitely many color predicates that any vertex in G or H satisfies. Thus, Lemma 4.4 applies and there is only a finite set $F_{k,m}$ of inequivalent formulas of interest in $\mathcal{L}_{k,m}$. Define S to be the set of formulas in $F_{k,m}$ that are satisfied by G,u_1 and let σ be the conjunction of the finitely many formulas in S. Thus we have that

$$G,u \models (\exists x_i)\sigma$$

It follows that $H,v \models (\exists x_i)\sigma$. Let Player II place the other x_i pebble on a witness h, for σ in H, and let $v_1 = v(x_i/h)$ be the result of this move. By the definition of σ it follows that

$$G,u_1 \equiv_{\mathcal{L}_{k,m}} H,v_1$$

Thus it follows by induction that Player II has a winning strategy for the remaining m-move game and thus also for the original $m+1$-move game. ∎

4.1. The $\mathcal{C}_k$ Game

A modification of the $\mathcal{L}_k$ game provides a combinatorial tool for analyzing the expressive power of $\mathcal{C}_k$. Given a pair of graphs define the $\mathcal{C}_k$ game on G and H as follows: Just as in the $\mathcal{L}_k$ game, we have two players and k pairs of pebbles. The difference is that each move now has two parts.

1. Player I picks up the x_i pebble pair for some i. He then chooses a set A of vertices from one of the graphs. Now Player II answers with a set B of vertices from the other graph. B must have the same cardinality as A.
2. Player I places one of the x_i pebbles on some vertex $b \in B$. Player II answers by placing the other x_i pebble on some $a \in A$.

The definition for winning is as before. What is going on in the two part move is that Player I asserts that there exist $|A|$ vertices in G with a certain property. Player II answers with the same number of such vertices in H. Player I challenges

[7] We are presenting the proof here for the language with no constant or function symbols. The proof goes through when function symbols are present, under the additional assumption that the cardinality of any finitely generated set is finite [26].

one of the vertices in B and Player II replies with an equivalent vertex from A. This game captures expressibility in $\mathcal{C}_k$:

Theorem 4.5. [27] *Player II has a winning strategy for the $\mathcal{C}_k$ game on G, H if and only if $G \equiv_{\mathcal{C}_k} H$.*

Theorem 4.5 follows from Theorem 5.2, which we prove in the next section.

5. $\mathcal{C}_k$-Equivalence Equals $(k-1)$-dim W–L

In this section we describe the k-dimensional Weisfeiler–Lehman method (k-dim W–L). We then prove that a pair of k-tuples of vertices from a graph agree on all formulas in $\mathcal{C}_{k+1}$ iff they are in the same equivalence class arising from the k-dim W–L.

The 1-dim W–L is also called vertex refinement. Let $G = \langle V, E, C_1, \ldots, C_r \rangle$ be a colored graph in which every vertex satisfies exactly one color relation. Let $W^0 : V \to \{1 \ldots n\}$ be given by $W^0(v) = i$ iff $v \in C_i$. We then define W^{r+1}, the refinement of W^r as follows: The new color of each vertex g is defined to be the following tuple:

$$\langle W^r(g), y_1, n_1, \ldots, y_r, n_r \rangle$$

where y_i is the number of vertices of color i that g is adjacent to, and n_i is the number of vertices of color i that g is not adjacent to. In practice, we sort these new colors lexicographically and assign $W^{r+1}(g)$ to be the number of the new color class that g inhabits. However, we retain a table decoding the "meaning" of each of the colors. Thus two vertices are in the same new color class precisely if they were in the same old color class, and they were adjacent to the same number of vertices of each color. We keep refining the coloring until at some level $W^r = W^{r+1}$. We let $\overline{W} = W^r$ and call $\overline{W}$ the *stable refinement* of W^0.

We will see in Theorem 5.2 that stable coloring provides exactly the same information as $\mathcal{C}_2$ equivalence.

Next define the k-dim W–L for $k > 1$ as follows. Let G be a colored graph and let u be a (total) map from $\{x_1, \ldots, x_k\}$ to V_G. Define the initial color $W^0(u)$ according to the isomorphism type of u. That is, $W^0(u) = W^0(v)$ iff the map from $(u(x_1), \ldots, u(x_k))$ to $(v(x_1), \ldots, v(x_k))$ is an isomorphism.

For each $g \in V_G$, define the operation

$$\text{sift}(f, u, g) = \langle f(u(x_1/g)), f(u(x_2/g)), \ldots, f(u(x_k/g)) \rangle$$

Thus $\text{sift}(W^r, u, g)$ is the k-tuple of W^r-colors arising from substituting g in turn for each of the k positions in u.

We define the $r+1$st color of u from the rth color by considering the rth color of u together with the number of vertices g such that $\text{sift}(W^r, u, g) = \bar{t}$ for each possible k-tuple of colors $\bar{t}$. More explicitly, form the new color of u as the tuple:

$$\langle W^r(u),\ \text{SORT}\ \{\text{sift}(W^r, u, g) \mid g \in G\} \rangle$$

As in the one dimensional case, we sort these new colors lexicographically, and assign W^{r+1} according to the ordering that ensues. However, we do retain a table decoding the meaning of each color. Thus for a pair of configurations u, v from

different graphs, $W^{r+1}(u) = W^{r+1}(v)$ iff the numbers of the colors assigned are the same, and the decoding tables for the two graphs are identical. Thus $W^{r+1}(u) = W^{r+1}(v)$ just if $W^r(u) = W^r(v)$ and for each k-tuple of colors $\bar{t}$,

$$|\{g \mid \text{sift}(W^r, u, g) = \bar{t}\}| = |\{g \mid \text{sift}(W^r, v, g) = \bar{t}\}| \tag{5.1}$$

(Note that the difference between the case $k = 1$ and the case $k > 1$ is that in the former case we have to explicitly consider which of the g's are adjacent to $u(x_1)$ in the above definition of new color; whereas, for $k > 1$, this adjacency is part of the information in the initial color of the tuples $u(x_j/g)$ for $j \neq 1$.)

Let $\overline{W}(u)$ denote the stable color of u. Note that there can be at most n^k color classes for a graph with n vertices and thus the algorithm stops after at most n^k iterations.

Theorem 5.2. *Let G, H be a pair of colored graphs and let (u, v) be a k-configuration on G, H, where $k \geq 1$. Then the following are equivalent:*

1. $\overline{W}(u) = \overline{W}(v)$
2. $G, u \equiv_{\mathscr{C}_{k+1}} H, v$
3. *Player II has a winning strategy for the $\mathscr{C}_{k+1}$ game on (G, H), whose initial configuration is (u, v).*

Proof. By induction on r we show that the following are equivalent:

1. $W^r(u) = W^r(v)$
2. $G, u \equiv_{\mathscr{C}_{k+1,r}} H, v$
3. Player II has a winning strategy for the r-move $\mathscr{C}_{k+1}$ game on (G, H) whose initial configuration is (u, v).

The base case is by definition. $W^0(u) = W^0(v)$ iff the map from $u(x_1), \ldots, u(x_k)$ to $v(x_1), \ldots, v(x_k)$ is an isomorphism. This is true iff G, u and H, v satisfy all the same quantifier-free formulas; and it is also the definition of Player II winning the zero move game.

Assume that the equivalence holds for all (u, v) and for all $r < m$.

$(\neg 1 \Rightarrow \neg 2)$: Suppose that $W^m(u) \neq W^m(v)$. There are two cases. If $W^{m-1}(u) \neq W^{m-1}(v)$ then by the inductive assumption there is a formula $\varphi \in \mathscr{C}_{k+1,m-1}$ on which G, u and H, v differ. Otherwise it must be that for some k-tuple of colors, $\bar{t} = (t_1, \ldots, t_k)$, Equation (5.1) fails. Let N be the cardinality of the larger set in Equation (5.1).

By induction, two k-tuples of vertices are in the same $f^{(m-1)}$ color class iff they agree on all formulas from $\mathscr{C}_{k+1,m-1}$. By Lemma 4.4 there are only finitely many inequivalent $\mathscr{C}_{k+1,m-1}$ formulas, when we restrict our attention to graphs with the same finite number of vertices as G. (If G and H have different numbers of vertices, then for $r \geq 1$, all the above conditions are false.) Let ψ_i be the conjunction of the finitely many $\mathscr{C}_{k+1,m-1}$ formulas characterizing the $m-1$ color class i. Thus, for w a k-configuration on $F \in \{G, H\}$

$$W^{m-1}(w) = i \;\Leftrightarrow\; F, w \models (\psi_i)$$

It follows that G,u and H,v differ on the following formula from $\mathcal{C}_{k+1,m}$.[8]

$$(\exists N x_{k+1})(\psi_{t_1}(x_1/x_{k+1}) \wedge \cdots \wedge \psi_{t_k}(x_k/x_{k+1}))$$

$(\neg 2 \Rightarrow \neg 3)$: Suppose that $G,u \models \varphi$ but $H,v \models \neg\varphi$, for some $\varphi \in \mathcal{C}_{k+1,m}$. If φ is a conjunction then G,u and H,v must differ on at least one of the conjuncts, so we may assume that φ is of the form $(\exists N x_i)\psi$. We may assume that x_i is the currently unassigned variable x_{k+1}. On the first move of the game Player I picks up the pair of x_{k+1} pebbles and chooses a set of N vertices g, such that $G,u(x_{k+1}/g) \models \psi$. Whatever Player II chooses as B there will be at least one vertex $h \in B$ such that $H,v(x_{k+1}/h) \models \neg\psi$. Player I puts his pebble on this h. Player II must respond with some $g \in A$. Now $G,u(x_{k+1}/g)$ and $H,v(x_{k+1}/h)$ differ on $\psi \in \mathcal{C}_{k+1,m-1}$. Thus by induction Player II loses the remaining $m-1$ move game.

$(1 \Rightarrow 3)$: Suppose that $W^m(u) = W^m(v)$. It follows that Equation (5.1) holds for each k-tuple of colors $\bar{t}$. Clearly Player I's strongest move involves the presently unused pair of x_{k+1} pebbles. Suppose he picks them up and chooses a set A of N vertices from G. For each $\bar{t}$, let $N_{\bar{t}}$ be the number of vertices $g \in A$ such that $\bar{t} = \text{sift}(W^{m-1},u,g)$. It follows from Equation (5.1) that Player II can put $N_{\bar{t}}$ vertices h into B such $\bar{t} = \text{sift}(W^{m-1},v,h)$.[9]

In the second part of the move Player I will put x_{k+1} on some $h \in B$. Player II should then answer with a $g \in A$ such that

$$\text{sift}(W^{m-1},u,g) = \text{sift}(W^{m-1},v,h) = \bar{t}$$

Consider the remaining game on configuration $(u(x_{k+1}/g),v(x_{k+1}/h))$. Note that Player II has not yet lost. At the beginning of the next move, Player I will choose some pair of pebbles x_i and pick them up. Now we know that the remaining configurations have the same W^{m-1} color. It follows by induction that Player II wins the remaining $(m-1)$-move game. ∎

The following observation will be useful in the proof of our main theorem.

Observation 5.3. If Player I has a winning strategy for the m-move $\mathcal{C}_k$ game on G,H, then he has a winning strategy in which throughout the game he only chooses monochromatic sets A.

Proof. We saw in the above proof that whenever Player I chooses a set A, this set may be partitioned according to the k-tuple of color classes induced. Player II then answers separately for each k-tuple of colors. If Player II does not have the right number of elements in one of these classes then she will lose, and Player I need only have selected his elements from that class. Each of these classes is monochromatic. ∎

It is not hard to see using standard coloring algorithms, cf. [1, §4.13], that

Fact 5.4. [27] The stable colorings of k-tuples may be computed in $O(k^2 n^{k+1} \log n)$ steps on a RAM.

---

[8] For the case $k = 1$ we must explicitly consider adjacency and so the formula is $(\exists N x_2)(E(x_1,x_2) \wedge \psi_{t_1}(x_1/x_2))$.

[9] In the case $k=1$, Player II must choose h's that are adjacent to $v(x_1)$ iff the corresponding g's are adjacent to $u(x_1)$.

It then follows from Theorem 5.2 that graphs that are identified by the language $\mathscr{C}_{k+1}$ have a canonization algorithm that runs in time $O(k^2 n^{k+1} \log n)$.

Remark 5.5. It is interesting to note that the language $\mathscr{L}_{k+1}$ enjoys a relationship similar to that of Theorem 5.2 with a variant of the k-dim W–L algorithm with the same time bound. The only difference is that the computation of the new color treats the following as a set instead of a multiset:

$$\{\text{sift}(W^{m-1}, u, g) \mid g \in V_G\}$$

That is, after sorting the collection of k-tuples, we eliminate duplicates.

6. Construction

We construct our counterexample graphs by starting with low degree graphs having only linear size separators. We replace each vertex v of degree k in such a graph by the graph X_k, defined as follows: $X_k = (V_k, E_k)$, where

$$\begin{aligned} V_k &= A_k \cup B_k \cup M_k \text{ where } A_k = \{a_i \mid 1 \le i \le k\}, \\ B_k &= \{b_i \mid 1 \le i \le k\}, \text{ and} \\ M_k &= \{m_S \mid S \subseteq \{1, \ldots, k\}, |S| \text{ is even }\} \\ E_k &= \{(m_S, a_i) \mid i \in S\} \cup \{(m_S, b_i) \mid i \notin S\} \end{aligned}$$

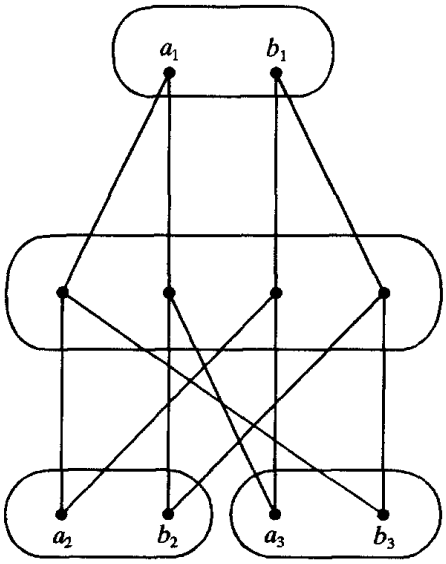


Fig. 3. The Graph X_3

Thus X_k consists of a set of 2^{k-1} vertices in the middle each connected to one vertex from each of the pairs $\{a_i, b_i\}$, $1 \leq i \leq k$. Furthermore, each of the middle vertices is connected to an even number of a_i's. (We will assume that the middle vertices (M_k) of X_k have a different color, say magenta, from the others ($A_k \cup B_k$). Furthermore, the pairs a_i and b_i should be able to recognize their mates. If necessary, add vertices c_i colored chartreuse, with edges to a_i and b_i.) See Figure 3 for a diagram of X_3.

The following lemma describes the relevant property of the graph X_k. The proof is immediate.

Lemma 6.1. *Suppose that we color the vertices a_i and b_i of graph X_k with the color i. (Thus all automorphisms of X_k must fix the sets $\{a_1, b_1\}, \ldots, \{a_k, b_k\}$.) Then there are exactly 2^{k-1} automorphisms of X_k. Each is determined by interchanging a_i and b_i for each i in some subset S of $\{1, \ldots, k\}$ of even cardinality.*

Let G be a finite, connected, undirected graph such that every vertex of G has degree at least two. Define the graph $X(G)$ ("X of G") as follows. For each vertex v of G, we replace v by a copy of X_k, call it $X(v)$, where k is the degree of v. To each edge (v, w) of v we associate one of the pairs $\{a_i, b_i\}$ from $X(v)$, call this pair $a(v, w)$ and $b(v, w)$. Finally, we connect the a vertices and the b vertices at each end of each edge, that is we draw the edges $(a(u, v), a(v, u))$ and $(b(u, v), b(v, u))$. If G is a colored graph, then each vertex in $X(v)$ should inherit the color of v. Next, define the graph $\tilde{X}(G)$ ("X twist of G") as follows: In the above construction of $X(G)$ arbitrarily choose one edge (v, w) and twist it, that is reverse the connections, drawing edges $(a(u, v), b(v, u))$ and $(b(u, v), a(v, u))$. In the next lemma we show some relevant properties of $X(G)$ and $\tilde{X}(G)$, including the fact that $\tilde{X}(G)$ is well defined.

Lemma 6.2. *Let G be any finite, connected graph such that every vertex of G has degree at least two. Let $X(G)$ and $\tilde{X}(G)$ be as above. Let $\hat{X}(G)$ be constructed like $X(G)$, but with exactly t of its edges twisted. Then $\hat{X}(G)$ is isomorphic to $X(G)$ iff t is even, and $\hat{X}(G)$ is isomorphic to $\tilde{X}(G)$ iff t is odd.*

Proof. First observe the following fact about $\hat{X}(G)$. Let v be any vertex of G, and let $(x, v), (y, v)$ be any two edges incident at v. If in $\hat{X}(G)$ we twist both (x, v) and (y, v), then the resulting graph is isomorphic to $\hat{X}(G)$. (This is immediate from Lemma 6.1.)

Now suppose that the number of twists in t is greater than or equal to two. The above observation lets us move the twists towards each other until they overlap and cancel each other out. Thus if t is even then $\hat{X}(G)$ is isomorphic to $X(G)$, otherwise it is isomorphic to $\tilde{X}(G)$.

It remains to show that $X(G)$ is not isomorphic to $\tilde{X}(G)$. Assume for the sake of a contradiction that φ is an isomorphism from $X(G)$ to $\tilde{X}(G)$. Consider the action of φ on any pair $\{a(v, w), b(v, w)\} \subset X(v)$, for (v, w) an edge of G. Because of the colorings in the definition of X_k, φ must map the pair $\{a(v, w), b(v, w)\}$ to some $\{a(v', w'), b(v', w')\}$ in $\tilde{X}(G)$, and thus φ also maps $\{a(w, v), b(w, v)\}$ to $\{a(w', v'), b(w', v')\}$. Define $\oplus\varphi$ to be the sum mod 2 over all such pairs in $X(G)$ of the number of times φ maps an a to a b. Clearly if we consider the two pairs corresponding to every edge (x, y) in G, the number of such switches is either zero

or two, except for the unique edge chosen in the construction of $\tilde{X}(G)$, where the number is one. Hence $\oplus\varphi$ is one mod 2. Now let's consider the mod 2 sum in another way, namely in terms of each copy of X_k in $X(G)$. By Lemma 6.1, it is immediate that $\oplus\varphi$ is zero mod 2. This contradiction proves the lemma. ∎

Definition 6.3. A *separator* of a graph $G=(V,E)$ is a subset $S \subset V$ such that the induced subgraph on $V-S$ has no connected component with more than $|V|/2$ vertices.

We now prove our main theorem:

Theorem 6.4. *Let T be a graph such that every separator of T has at least $s+1$ vertices. Then*

$$X(T) \equiv_{\mathscr{C}_s} \tilde{X}(T).$$

Proof. By Theorem 5.2, it suffices to give a winning strategy for Player II in the $\mathscr{C}_s$ game on $X(T)$ and $\tilde{X}(T)$. We will assume that the original graph T has color class size one. The graphs $X(T)$ and $\tilde{X}(T)$ inherit these colors and so have color class size 2^{k-1}, where k is the maximum degree of any vertex in T. This only makes life more difficult for Player II.

We know by Lemma 6.2 that if we add a twist to any edge of $X(T)$, then the resulting graph is isomorphic to $\tilde{X}(T)$. After the rth move of the game, let Q_r be the largest connected component in $T-P_r$ where P_r is the set of vertices $g \in T$ such that just after the rth move there is a pebble on a vertex of $X(g)$ in $X(T)$. Since T has no s separator, we know that Q_r contains over half the vertices of T. Player II's winning strategy will be to maintain the following property:

(*) For each vertex $g \in Q_r$, let $X^g(T)$ be $X(T)$ with an edge adjacent to g twisted. Then there exists an isomorphism $\alpha_{r,g}$ from $X^g(T)$ to $\tilde{X}(T)$, such that for all $i \leq s$, $\alpha_{r,g}$ maps the vertex under pebble x_i in $X(T)$ to the vertex under pebble x_i in $\tilde{X}(T)$.

The difference between $X(T)$ and $\tilde{X}(T)$ is that the latter graph has one twisted edge. An intuitive explanation of Player II's winning strategy is that she keeps this twisted edge inside of Q_r. With only s pebbles, Player I cannot break apart Q_r to expose the twist.

Clearly if Player II can maintain (*), then the map from the pebbled points in $X(T)$ to the corresponding pebbled points in $\tilde{X}(T)$ is an isomorphism, and she wins. We show by induction on r, that Player II can maintain (*). First let us make a remark about Player I's moves. By Observation 5.3, it always suffices for Player I to restrict himself to choosing a set of monochromatic points at each move. Notice that, if Player I chooses a vertex in $M(h)$, the middle of an $X(h)$, then all the other vertices in that $X(h)$ are determined. Furthermore, since one point in $M(h)$ determines all of $M(h)$, it suffices for Player I to choose only a single point at a time. (Thus counting does not help at all in distinguishing $X(T)$ from $\tilde{X}(T)$!)

Player II's inductive strategy can now be stated. Assume (*) holds, and suppose that on move $r+1$ Player I picks up pebble x_i and puts it down on a vertex in $M(w)$. Note that a new largest component Q_{r+1} is determined. Let g be a vertex in $Q_r \cap Q_{r+1}$. Player II's response is to answer Player I's move according to the isomorphism $\alpha_{r,g}$. To maintain (*), let $\alpha_{r+1,g} = \alpha_{r,g}$. Since there is a pebble-free

path from g to every other vertex in Q_{r+1}, the proof of Lemma 6.2 shows us how to define all the other isomorphisms, $\alpha_{r+1,h}$, $h \in Q_{r+1}$. ∎

Corollary 6.5. *There exists a sequence of pairs of graphs* $\{G_n, H_n\}$, $n \in \mathbb{N}$ *admitting a linear time canonical labeling algorithm and having the following additional properties:*

1. G_n *and* H_n *have* $O(n)$ *vertices.*
2. G_n *and* H_n *have degree three and color class size four.*
3. $G_n \equiv_{\mathscr{C}_n} H_n$.
4. G_n *is not isomorphic to* H_n.

Proof. This follows immediately from Theorem 6.4 when we let $G_n = X(T_n)$ and $H_n = \tilde{X}(T_n)$ where the T_n's are a sequence of degree three graphs of separator size n, with each vertex of T_n colored a unique color. Such graphs are well known to exist, see for example [2]. ∎

7. Corollaries

A long time ago, one of us showed that there is a polynomial-time property of graphs that requires $\Omega(2^{\sqrt{\log n}})$ quantifiers to be expressed in first-order logic without ordering. That proof also used the graphs $X(D_n)$ and $\tilde{X}(D_n)$, for a certain sequence of degree three graphs $\{D_n\}$ [22, Theorem 7]. Now, Corollary 6.5 improves that lower bound to $\Omega(n)$ variables.[10] It also shows graphically that if we exclude the ordering relation from inductive first-order logic, then the addition of counting does not suffice to express all polynomial-time graph properties. In particular, we have the following:

Corollary 7.1. *Let* Γ *be the set of all graphs of the form* $X(G)$, *or* $\tilde{X}(G)$, *for all graphs* G *of degree at most three and color class size one. Then the isomorphism problems for graphs in* Γ *is expressible in first-order logic with ordering and sum mod 2, but it is not expressible by any sequence of first-order sentences from* $\mathscr{C}_{r(n)}$ *(without ordering), where* $r(n) = o(n)$.

Remark 7.2. In particular, inductive logic with counting, but without ordering does not contain all the polynomial-time computable graph properties. In fact, it does not even contain all such properties computable by a uniform sequence of bounded-depth, polynomial-size Boolean circuits that include parity gates, cf. [12].

Proof. We have seen in Corollary 6.5 that the graphs $X(T_n)$ and $\tilde{X}(T_n)$ are indistinguishable in $\mathscr{C}_{\epsilon n}$ for some constant $\epsilon > 0$. Suppose for the sake of a contradiction that there were a sentence $\sigma \in (\mathrm{FO}+\mathrm{LFP}+\mathrm{COUNT})$ that expresses the isomorphism property for graphs from Γ. That is for graphs $G, H \in \Gamma$,

$$(G, H) \models \sigma \Leftrightarrow G \cong H$$

[10] This is a major improvement because n is much bigger than $2^{\sqrt{\log n}}$, and because a sentence with q quantifiers can make use of at most q variables, but a sentence with v variables can make use of 2^{n^v} quantifiers.

Let k be the number of distinct variables occurring in σ. For graphs of size n, let σ_n be the unwinding of σ as follows. Rewrite any least fixed points of arity a, $(\mathrm{LFP}\varphi)$ as $\varphi^{(n^a)}(\emptyset)$. Next replace any quantified number variable $\exists i$ (respectively, $\forall i$) by a disjunction $\bigvee_{i=0}^{n-1}$ (respectively, by a conjunction $\bigwedge_{i=0}^{n-1}$). Note that $\sigma_n \in \mathscr{C}_k$ and is equivalent to σ for structures of size at most n.

Thus we have that σ_n distinguishes the pair $P=(X(T_n),\tilde{X}(T_n))$ from the pair $Q=(X(T_n),X(T_n))$. It follows that Player I wins the $\mathscr{C}_k$ game on these two pairs. Note that Player II can match any vertex in the first $X(T_n)$ from P with the same vertex in the first $X(T_n)$ from Q. Thus, Player I must have a winning strategy for the $\mathscr{C}_k$ game on $X(T_n)$ and $\tilde{X}(T_n)$. This contradiction shows that isomorphism for graphs from Γ is not expressible in $(\mathrm{FO}+\mathrm{LFP}+\mathrm{COUNT})$.

We next show that we can distinguish $X(G)$ from $\tilde{X}(G)$ in first-order logic with ordering and sum mod 2. This is easy. The ordering gives us a way to mark each of the pairs $a(g,h)$ and $b(g,h)$ in the graphs. Let $a(g,h)$ be the first of the pair, and $b(g,h)$ the second. (Note that since the vertices in $M(g)$ and $M(h)$ inherit unique colors from g and h, we are given as part of the input which pair of vertices is $a(g,h),b(g,h)$.) Now, given this assignment of a's and b's, a simple first-order sentence asserts that $X(g)$ is straight (i.e. isomorphic to X_3) or twisted (i.e. each vertex in $M(g)$ is adjacent to an odd number of a's). Now, the graph is isomorphic to $X(G)$ iff the sum mod 2 of the number of twisted vertices and edges is 0, and it's isomorphic to $\tilde{X}(G)$ iff the sum mod 2 is 1.

Of course, if $G \neq H$, then since these graphs have color class size one, $X(G)$ and $X(H)$ can be distinguished by a sentence in $\mathscr{L}_2$. Thus isomorphism for graphs from Γ is expressible in AC^0 plus parity gates, as claimed. ∎

The next result proves a straightforward upper bound that nearly matches our lower bound on the number of variables needed to identify a class Δ of graphs as a function of the separator size of members of Δ.

Proposition 7.3. *Let Δ be a set of graphs closed under induced subgraphs, such that every graph $G \in \Delta$ has a separator of size at most $s(n)$, where n is the number of vertices of G. Then Δ is identified by $\mathscr{C}_{V(n)}$ where*

$$V(n) = 3 + \sum_{i=0}^{\lfloor \log n \rfloor} s(\lfloor n2^{-i} \rfloor).$$

(In particular, $V(n) \leq s(n)\log n$, and if $s(n)=n^\alpha$, then $V(n)=O(s(n))$.)

Proof. We use induction on n, the number of vertices of G. Given G, we can first say that there exist vertices $x_1,\ldots,x_{s(n)}$ such that every connected component of $G-\{x_i \mid 1 \leq i \leq s(n)\}$ has size at most $\lfloor n/2 \rfloor$. This is expressible in $s(n)+3$ variables. Next we assert how many connected components of each isomorphism type there are. This requires $V(\lfloor n/2 \rfloor)$ variables, in addition to the $s(n)$ that we leave on $x_1,\ldots,x_{s(n)}$. ∎

8. Conclusions and Open Questions

1. We redirect the reader's attention to Questions 3.2 and 3.3. We have shown in Corollary 6.5 that first-order logic plus counting and least fixed point, but without ordering, fails badly. The question, "What besides counting must be added to FO + LFP to get all polynomial-time graph problems?" is worthy of much study, cf. [27,20].
2. Planar graphs have separators of size $O(\sqrt{n})$, and thus by Proposition 7.3 they can be identified in $\mathcal{C}_{\sqrt{n}}$. However, Theorem 6.4 does not give a matching lower bound because even if G is planar, the graph $X(G)$ need not be. We would like to know if $\Omega(\sqrt{n})$ variables are necessary to identify planar graphs.

Acknowledgements: Thanks to Sandeep Bhatt who improved our results by pointing out that the essential property of the counterexample graphs we were using was that their separators are large. Thanks to Laci Babai for informing us about the status of the research on the W–L method in the Soviet Union.

References

[1] A. V. AHO, J. E. HOPCROFT and J. D. ULLMAN: *The Design and Analysis of Computer Algorithms,* Addison-Wesley (1974).

[2] M. AJTAI: Recursive Construction for 3-Regular Expanders, *28th IEEE Symp. on Foundations of Computer Science* (1987), 295-304.

[3] L. BABAI: Monte Carlo Algorithms in Graph Isomorphism Testing, Tech. Rep. DMS 79-10, Université de Montréal, 1979.

[4] L. BABAI: On the Complexity of Canonical Labeling of Strongly Regular Graphs, *SIAM J. Computing* **9** (1980), 212-216.

[5] L. BABAI: Moderately Exponential Bound for Graph Isomorphism, *Proc. Conf. on Fundamentals of Computation Theory,* Lecture Notes in Computer Science, Springer, 1981, 34-50.

[6] L. BABAI: On the Order of Uniprimitive Permutation Groups, *Annals of Math.* **113** (1981), 553-568.

[7] L. BABAI: *Permutation Groups, Coherent Configurations, and Graph Isomorphism,* D. Sc. Thesis, Hungarian Acad. Sci., 1984 (in Hungarian).

[8] L. BABAI, P. ERDŐS, and S. M. SELKOW: Random Graph Isomorphism, *SIAM J. on Comput.* **9** (1980), 628-635.

[9] L. BABAI, W. M. KANTOR, and E. M. LUKS: Computational Complexity and the Classification of Finite Simple Groups, *24th IEEE Symp. on Foundations of Computer Science* (1983), 162-171.

[10] L. BABAI and L. KUČERA: Canonical Labelling of Graphs in Linear Average Time, *20th IEEE Symp. on Foundations of Computer Science* (1979), 39-46.

[11] L. BABAI and E. M. LUKS: Canonical Labeling of Graphs, *15th ACM Symposium on Theory of Computing* (1983), 171-183.

[12] D. M. BARRINGTON, N. IMMERMAN, and H. STRAUBING: On Uniformity Within NC^1, *J. Comput. System Sci.* **41**, No. 3 (1990), 274-306.

[13] L. BABAI and R. MATHON: Talk at the South-East Conference on Combinatorics and Graph Theory, 1980.

[14] P. J. CAMERON: 6-Transitive Graphs, *J. Combinat. Theory B* **28**, (1980), 168-179.

[15] A. CHANDRA and D. HAREL: Structure and Complexity of Relational Queries, *J. Comput. System Sci.* **25** (1982), 99-128.

[16] A. EHRENFEUCHT: An Application of Games to the Completeness Problem for Formalized Theories, *Fund. Math.* **49** (1961), 129-141.

[17] R. FRAÏSSÉ: Sur quelques classifications des systèms de relations, *Publ. Sci. Univ. Alger* **1** (1954), 35-182.

[18] M. FÜRER, W. SCHNYDER, and E. SPECKER: Normal Forms for Trivalent Graphs and Graphs of Bounded Valence, *15th ACM Symposium on Theory of Computing* (1983), 161-170.

[19] YA. YU. GOL'FAND and M. H. KLIN: On k-Regular Graphs, in: *Algorithmic Research in Combinatorics,* Nauka Publ., Moscow, 1978, 76-85.

[20] YU. GUREVICH: Logic and the Challenge of Computer Science, in: *Current Trends in Theoretical Computer Science,* ed. Egon Börger, Computer Science Press, 1988, 1-57.

[21] D. G. HIGMAN: Coherent Configurations I.: Ordinary Representation Theory, *Geometriae Dedicata* **4** (1975), 1-32.

[22] N. IMMERMAN: Number of Quantifiers is Better than Number of Tape Cells, *J. Comput. System Sci.* **22**, No. 3 (1981), 384-406.

[23] N. IMMERMAN: Upper and Lower Bounds for First Order Expressibility, *J. Comput. System Sci.* **25**, No. 1 (1982), 76-98.

[24] N. IMMERMAN: Relational Queries Computable in Polynomial Time, *Information and Control* **68** (1986), 86-104.

[25] N. IMMERMAN: Languages That Capture Complexity Classes, *SIAM J. Computing* **16**, No. 4 (1987), 760-778.

[26] N. IMMERMAN and D. KOZEN: Definability with Bounded Number of Bound Variables, *Information and Computation* **83** (1989), 121-139.

[27] N. IMMERMAN and E. S. LANDER: Describing Graphs: A First-Order Approach to Graph Canonization, in: *Complexity Theory Retrospective,* Alan Selman, ed., Springer-Verlag, 1990, 59-81.

[28] N. IMMERMAN, S. PATNAIK, and D. STEMPLE: The Expressiveness of a Family of Finite Set Languages, *Tenth ACM Symposium on Principles of Database Systems* (1991), 37-52.

[29] L. KUČERA: Canonical Labeling of Regular Graphs in Linear Average Time, *28th IEEE Symp. on Foundations of Computer Science* (1987), 271-279.

[30] M. H. KLIN, M. E. MUZICHUK, and I. A. FARADŽEV: Cellular Rings and Groups of Automorphisms of Graphs, Introductory Article to a Book to be Published by D. Reidel Publ. Co.

[31] M. CH. KLIN, R. PÖSCHEL, and K. ROSENBAUM: Angewandte Algebra, Vieweg & Sohn Publ., Braunschweig 1988.

[32] R. LIPTON: The Beacon Set Approach to Graph Isomorphism, Yale Dept. Comp. Sci. preprint No. 135, 1978.

[33] E. M. LUKS: Isomorphism of Graphs of Bounded Valence Can be Tested in Polynomial Time, *J. Comput. System Sci.* **25** (1982), 42-65.

[34] R. MATHON: A Note On the Graph Isomorphism Counting Problem, *Inform. Proc. Let.* **8** (1979), 131-132.

[35] B. D. MCKAY: Nauty User's Guide (Version 1.2), Tech. Rep. TR-CS-87-03, Dept. Computer Science, Austral. National Univ., Melbourne, 1987.

[36] G. L. MILLER: On the $n^{\log n}$ Isomorphism Technique, *10th ACM Symposium on Theory of Computing* (1978), 51-58.

[37] R. C. READ and D. G. CORNEIL: The Graph Isomorphism Disease, *J. Graph Theory* **1** (1977), 339-363.

[38] M. VARDI: Complexity of Relational Query Languages, *14th ACM Symposium on Theory of Computing* (1982), 137-146.

[39] B. WEISFEILER, ED.: *On Construction and Identification of Graphs,* Lecture Notes in Mathematics 558, Springer, 1976.

[40] B. WEISFEILER and A. A. LEHMAN: A Reduction of a Graph to a Canonical Form and an Algebra Arising during this Reduction, (in Russian), *Nauchno-Technicheskaya Informatsia, Seriya 2,* **9** (1968), 12-16.

[41] V. N. ZEMLYACHENKO, N. KORNIENKO, and R. I. TYSHKEVICH: *Graph Isomorphism Problem,* (in Russian), The Theory fo Computation I, Notes Sci. Sem. LOMI 118, 1982.

Jin-Yi Cai

Computer Science Dept.
Princeton University
Princeton, NJ 08540
jyc@princeton.edu

Martin Fürer

Computer Science Dept.
Pennsylvania State University
University Park, PA 16802
furer@cs.psu.edu

Neil Immerman

Computer Science Dept.
University of Massachusetts
Amherst, MA 01003
immerman@cs.umass.edu

Paper 12

JOURNAL OF COMPUTER AND SYSTEM SCIENCES 38, 68–85 (1989)

With Probability One, a Random Oracle Separates *PSPACE* from the Polynomial-Time Hierarchy

JIN-YI CAI*

Department of Computer Science, Yale University, New Haven, Connecticut 06520

Received September 11, 1986

We consider how much error a fixed depth Boolean circuit must make in computing the parity function. We show that with an exponential bound of the form $\exp(n^\lambda)$ on the size of the circuits, they make a 50% error on all possible inputs, asymptotically and uniformly. As a consequence, we show that a random oracle set A separates *PSPACE* from the entire polynomial-time hierarchy with probability one.

1. INTRODUCTION

The relationship between time and space, as complexity measures, has been one of the primary concerns in complexity theory research. It is well known that the entire polynomial-time hierarchy *PH* is contained in *PSPACE*. However, despite convincing heuristic evidence and persistent effort, no proof is yet available for separating the polynomial-time hierachy from polynomial space.

A proof that $PH \neq PSPACE$ would be an extremely strong separation of time and space. In this paper, we show that *PH* is properly contained in *PSPACE* in almost all relativized worlds.

THEOREM 1.1. *With probability one, a random oracle separates PSPACE from the entire polynomial-time hierarchy.*

The present work is a continuation of the work pioneered by Furst, Saxe, Sipser, and Yao. For the definitions of some basic notions we refer the reader to Refs. [FSS84; Sip83; Yao85].

In 1978 Furst, Saxe, and Sipser showed that the Boolean function *Parity* (see the definition below) cannot be computed in a fixed depth polynomial size Boolean circuit. They also observed that an exponential poly-logarithmic lower bound (i.e., bounded below by $\exp((\log)^k)$ for all k) would establish the existence of an oracle separating *PSPACE* from the polynomial hierarchy. Later Sipser extended this

* The research was supervised by Professor J. Hartmanis as part of this author's Ph.D. thesis. It was supported by a Sage Fellowship from Cornell University and NSF Grant DCR-8301766.

0022-0000/89 $3.00

work in [Sip83]. Finally, in 1985, a breakthrough came with the following theorem by Yao, which influenced our research immensely.

THEOREM 1.2 (Yao). *There exists an oracle A such that*

$$P^A \neq NP^A \neq \Sigma_2^{P,A} \neq \cdots \neq PH^A \neq PSPACE^A.$$

Our strong separation result is obtained by looking at *how much* error is present in the supposed circuit computation (instead of the existence of a *single* error). This question is interesting in its own right in the theory of circuit computation. As a nice byproduct of the proof of this strong separation we have the following corollary.

COROLLARY 1.3. *Fixed depth Boolean circuits with a bound of the form* $\exp(n^\lambda)$ *on the size, for some* λ, *make a* 50% *error, aymptotically and uniformly, when they compute the Boolean function Parity.*

The proof in this paper is organized as follows:

1. Use the alternating Turning machine model [CKS81] to reduce the problem to a Boolean circuit computation problem.

2. Employ certain probabilistic and game theoretic techniques to crack a shallow circuit.

3. Inductively prove a theorem in the general case and then adapt it to resolve the problem on circuit computation in step 1.

2. INITIAL REDUCTIONS AND TECHNIQUES

We proceed with some definitions. Let X be the set of n Boolean variables $\{x_1, x_2, ..., x_n\}$. A $\Sigma_{0,n}$-formula (circuit) is the constant 0, and a $\Pi_{0,n}$-formula (circuit) is the constant 1. A $\Sigma_{1,n}$-formula (circuit) is a sum of the form $\Sigma_k \overline{x_{i_k}} + \sum_k x_{j_k}$, where $x_{i_k}, x_{j_k} \in X$. Without loss of generality, we assume that the variables are distinct. The number of literals is its size. A $\Pi_{1,n}$-formula (circuit) is the negation of a $\Sigma_{1,n}$-formula, with the same size, i.e., a product of the form $\prod_k \overline{x_{i_k}} \cdot \prod_k x_{j_k}$.

For $k > 1$, a $\Sigma_{k,n}$-formula (circuit) H is a sum of $\Pi_{k-1,n}$-formulae, $\sum_i G_i$, with $\text{size}(H) = \sum_i \text{size}(G_i)$. A $\Pi_{k,n}$-formula is the negation of a $\Sigma_{k,n}$-formula, with the same size. Inductively, a subcircuit of H is H or any of the subcircuits of the G_i's. The depth of a $\Sigma_{k,n}$-formula (circuit) or a $\Pi_{k,n}$-formula is k. The bottom fan-in (*bfi*) of a Boolean circuit is the maximum size of the depth one subcircuits.

For any $\Pi_{2,n}$-formula G, $G = \prod_{i=1}^{l} C_i$, where

$$C_i = \overline{x_{i_1}} + \overline{x_{i_2}} + \cdots + \overline{x_{i_s}} + x_{i_{s+1}} + x_{i_{s+2}} + \cdots + x_{i_{s+t}} \qquad \text{and} \qquad s, t \geqslant 0.$$

We let $J_{i-} = \{i_1, i_2, ..., i_s\}$ $J_{i+} = \{i_{s+1}, i_{s+2}, ..., i_{s+t}\}$, and $J_i = J_{i-} \cup J_{i+}$.

A (partial) assignment of X is an n-tuple $\sigma=(\sigma_1, \sigma_2, ..., \sigma_n)\in\{0, 1, *\}^n$. If $\sigma\in\{0, 1\}^n$, then σ is a total assignement. Let F be a Boolean function on X; then $F|_\sigma$ denotes the Boolean function after the assignment σ, i.e., assign $x_i=0$, 1, or unassigned, if $\sigma_i=0$, 1, or $*$.

To generalize a bit, we also consider random assignments of X. For $0\leqslant p\leqslant 1$, let $\mathscr{R}_p$ denote the probability space $\{0, 1, *\}^\omega$ with a product measure v, where (independently) for each coordinate i, $1\leqslant i<\omega$, $v\{(a_1, ..., a_i, ...) \mid \alpha_i=\alpha\}=(1-p)/2$, if $\alpha=0$ or 1; and p if $\alpha=*$. That such a product measure exists is a well known result of probability theory. A random assignement is simply a point in the measure space $\mathscr{R}_p$. We will write it as $A=(a_1, ..., a_i, ...)$.

If F is a Boolean function on free variables $\{x_{i_1}, ..., x_{i_s}\}\subseteq X$, then a random assignment A taken from $\mathscr{R}_p$ (denoted as $A\in\mathscr{R}_p$) assigns the variables x_{i_j} to 0, 1, or leaves it unassigned, according to a_j of A.

We denote by $F\|_A$ the Boolean function that resulted from the assignement.

Similarly we define one-sided random assignements. A random B taken from $\mathscr{R}_p^+$ (denoted as $B\in\mathscr{R}_p^+$) assigns independently to each x_{i_j} in F to 1 with probability p, and leaves it unassigned with probability $1-p$, respectively. $\mathscr{R}_p^-$ is defined in the same way with 0 substituting for 1. Note that all random assignments affect only free variables, when they are applied to a formula.

Consider a sequence of random assignements $R_1, ..., R_s$. $F\|_{R_1, ..., R_s}$ is defined to be $(F\|_{R_1, ..., R_{s-1}})\|_{R_s}$. For instance, let R and S be two random assignements, $R=(a_1, ..., a_j, ...)$, $S=(b_1, ..., b_k, ...)$. Let $a_{j_1}, a_{j_2}, ...$ be those a_j in R which are equal to $*$. As before let F be a Boolean function on free variables $\{x_{i_1}, ..., x_{i_s}\}$. Then in $F\|_{RS}$, x_{i_j} is assigned a_j if a_j is not a $*$. Otherwise, suppose a_j is a_{j_k}, the k th $*$ in R; then x_{i_j} is assigned b_k, provided b_k is not a $*$. Finally if b_k is a $*$, then x_{i_j} is left unassigned. The successive random assignments act only on the variables left untouched by previous assignments. In what follows, when we make a statement such as "take two random assignements R and S from probability spaces $\mathscr{R}$ and $\mathscr{S}$, respectively, with probability p, event E occurs," we assert the product measure of the set $\{(R, S) \mid E \text{ occurs}\}\subseteq\mathscr{R}\times\mathscr{S}$ is p. We also denote $R_1\cdots R_s$ as $A=(a_1, ..., a_j, ...)$, where $a_j=0$, 1, or $*$, depending on whether $R_1\cdots R_s$ assigns the jth variable to 0, 1, or unassigned, respectively.

Note, however, that a partial assignement σ is applicable to $F\|_{R_1, ..., R_s}$ iff σ assigns 0 or 1 to only those variables that are $*$-valued by $R_1\cdots R_s$. In this case, we denote the resulting function by $(F\|_{R_1, ..., R_s})|_\sigma$.

Fix an alphabet $\{0, 1\}$ and an integer n. Define the parity function $Parity_n$:

$$Parity_n(x_1, x_2, ..., x_n)=\sum_{i=1}^{n} x_i \pmod 2.$$

We will consider circuit computation in relation to the parity function $Parity_n$. The parity function is chosen for the following property: the value of $Parity_n$ is *vitally* dependent on each variable x_i.

For $A \subseteq \{0, 1\}^*$, define the parity language

$$Parity^A = \{1^n \mid \text{there are odd number of strings of length } n \text{ in } A\}.$$

Clearly, we have $Parity^A \in PSPACE^A$, for all A.

We study separation of the polynomial-time hierarchy from *PSPACE* in almost all the relativized world. Intuitively, a random oracle set A is generated as follows. For each string $x \in \{0, 1\}^*$, we flip a fair coin, and depending on the outcome, we put x in A or not. Formally, we may represent each A by its characteristic function, and then map to a real number in the binary expansion $\in [0, 1]$. Now we define the probability measure μ on the oracle space to be the Lebesgue measure on $[0, 1]$. The readers may easily verify that the formal definition represents our intuitive notion of a random set as described above.

We aim to prove that $\mu\{A \mid Parity^A \notin PH^A\} = 1$. Surely this implies that a (random) oracle separates *PSPACE* from the entire polynomial-time hierarchy, with probability one.

There are only countably many levels $\Sigma_k^{P,A}$ in PH^A. Each level Σ_k^P has a recursive enumeration as the class of languages accepted by polynomial time alternating machines at that level [CKS81]. Let $M_1, M_2, \ldots$ be an enumeration of Σ_k^P alternating machines; then it is sufficient to show that

$$\forall i, \qquad [\mu\{A \mid Parity^A \neq L(M_i^A)\} = 1].$$

According to a theorem by Bennet and Gill [3], we only need to show that for each level k,

$$\exists \varepsilon_0 > 0, \forall i, \qquad [\mu\{A: Parity^A \neq L(M_i^A)\} > \varepsilon_0].$$

Now we reduce alternating machines to Boolean circuits. This reduction is from Furst, Saxe, and Sipser in [FSS84].

For a fixed alternating machine with oracle M_i^A at level Σ_k^P, consider its computation on 1^m. We claim that it is always possible to postpone the queries of strings. The trick is to guess the answers and verify them at a later stage. For example, at an existential stage, whenever a query is needed, we instead guess the answer and proceed until the succeeding universal stage. At the beginning of this universal stage we verify with the orable the guesses at the previous existential stage. If any of the guesses is wrong, we abort this path; otherwise we proceed. Similarly we may delay the queries of a universal stage and verify them at the suceeding existential stage. This time if any guess is wrong we simply accept. It is easily shown that this transformation preserves the notion of acceptance by the alternating Turing machine. By adding one more level of the alternation using the same method, we may obtain an equivalent polynomial-time bounded alternating machine that queries only at the bottom level and queries only once on every computation path. Thus, the computation tree structure is *independent* of the oracle and, therefore, can be frozen to yield a Σ_{k+1}-circuit, G.

The set of input Boolean variables of G corresponds to those strings that are queried by the modified alternating machine, on 1^m. (A queried string is in A *iff* the corresponding Boolean variable is set to be true.) Empirically, there should be precisely 2^m input variables, corresponding to 2^m strings of length m. This is because whether $1^m \in Parity^A$ is independent of any string of length unequal to m; furthermore, for any string x of length m, whether $1^m \in Parity^A$ depends on whether or not $x \in A$. Does the machine have to query precisely those strings of length m, no more and no less? We show that this is indeed the case, without loss of generality, in the sense that one can always replace the machine with one that does. (Strictly speaking, we only replace the machine with a nonuniform circuit family. Thus there is no uniformity concern.)

Suppose then for some x, $|x| = m$, and x is never queried. Then clearly the machine errs with probability 1/2 (under μ) at length m. That is enough.

Now suppose all x's with $|x| = m$ are queried, but so are $y_1, ..., y_l$ of length unequal to m.

Consider all 2^l many possible assignements σ for the y_i's. For any such σ, consider the $\Sigma_{k+1,2^m}$-circuit $G|_\sigma$. Pick the best σ_0, in the sense that $G|_{\sigma_0}$ makes the least error for parity. Clearly the original G makes no less error than that made by $G|_{\sigma_0}$, percentage-wise. Formally speaking,

$$\mu\{A \mid Parity^A(1^m) \neq M_i^A(1^m)\} \geqslant \text{error rate of } G|_{\sigma_0} \text{ for parity} \equiv \frac{|\{\tau \in \{0,1\}^n : G|_{\sigma_0}|_\tau \neq Parity_n|_\tau\}|}{2^n},$$

where $n = 2^m$.

A remark on the size: Since $M_i(1^m)$ runs at most $p(m)$ steps, for some polynomial $p(\cdot)$, the size of the circuit is bounded by an exponential polylog in n, $\exp(O(p(\log(n))))$, where n is the input size to the circuit.

We have shown that in order to prove Theorem 1.1, the following theorem on Boolean circuit computation would suffice.

THEOREM 2.1. *For all $k \geqslant 2$, there is a sequence α_n, where $\alpha_n \to \frac{1}{2}$ as $n \to \infty$, such that all depth k Boolean circuits with n inputs and size bounded by* $\exp(n^{1/4(k+1)})$ *err on more than $\alpha_n \cdot 2^n$ of the 2^n many inputs when computing $Parity_n$.*

We now define a notion that is central to our exposition.

A Boolean function G is given. Consider the following class of two-man games, played between a master and a player: the general mode of the game is a cycle; the master gives a Boolean variable (unassigned so far) and asks the player to assign it. The player may assign it either 0 or 1. The master may repeat the cycle zero or more times, until he declares the end of the game. The rule dictates that when the master declares the end of the game, the assignment made by the player so far makes G a constant.

A Boolean function G is *k-monochromatic* iff there is a two-man game of the defined class, in which the master has a winning strategy in the following sense: the master can declare the end of the game after no more than $\lceil k \rceil$ many variables are assigned.

To put it differently, it is guaranteed that, no matter how the player plays, the master can force the function G to be constant, after at most $\lceil k \rceil$ variables are assigned.

Here we emphasize two points:

1. The k variables are *not* given out in a batch; rather the master makes up his mind as to which variable to give next, depending on how the player has assigned the variables so far.

2. Even in play following a winning strategy, the master is (technically) *not* required to declare the end of the game at the earliest possible moment.

We finish the section with the following lemma, which essentially states that under a *monochromaticity* condition, a *conjunction* of a *disjunction* and a *disjunction* of a *conjunction* are interchangeable.

LEMMA 2.2. *If G is k-monochromatic, then G is equivalent to a $\Sigma_{2,n}$-circuit (as well as a $\Pi_{2,n}$-circuit) with bfi $\leqslant k$. (It is a constant if bfi $= 0$.)*

Proof. The proof is by induction on k. The case $k = 0$ is trivial. Suppose $k > 0$, and the lemma is true for all values less than k. Let G be k-monochromatic, but not $(k-1)$-monochromatic. Let us play the game; suppose x_i is the first variable the master puts out when following the strategy given by k-monochromaticity. Then $G = [x_i \wedge G|_{x_i = T}] \vee [\overline{x_i} \wedge G|_{x_i = F}]$, where T stands for *true* and F stands for *false*. Now both $G|_{x_i = T}$ and $G|_{x_i = F}$ are $(k-1)$-monochromatic; we apply the inductive hypothesis once more, and the result follows. Q.E.D.

3. DEPTH TWO CIRCUITS

We wish to prove a theorem concerning depth two Boolean circuits.

THEOREM 3.1. *Fix $0 < \varepsilon < \frac{1}{5}$. Then there exists a constant C, such that, for any $G \in \Pi_{2,n}$ with bfi $\leqslant n^{\varepsilon}$, and for any q with $0 \leqslant q \leqslant n^{-1.05\varepsilon}$, and a random $Q_1 \in \mathscr{R}^{-}_{1-q}$ and a random $Q_2 \in \mathscr{R}^{+}_{1-q}$, the probability that $G \| _{Q_1 Q_2}$ is n^{ε}-monochromatic is $1 - \varepsilon_n$, where $\varepsilon_n \leqslant Ce^{-n^{\varepsilon}}$.*

The idea of the proof is as follows: We will define a two-man game associated with the circuit $G \| _{Q_1 Q_2}$, for which we claim that the master most probably has a

winning strategy (cf. Section 2). The game is designed so that each play creates a *record* of how the game was played. In the rare case in which $G\|_{Q_1Q_2}$ is not n^ε-monochromatic, the *record* will be "large." Now we define another procedure, Recording $(Q_1Q_2, record)$, which will reproduce the game play. On the other hand, given a "large" *record*, the event that a random assignment R will *survive* the procedure Recording $(R, record)$ is so unlikely that even if the probability is summed over all "large" *records*, it is still of measure near 0.

3.1. *The Game and the Recording*

We denote an assignment Q_1Q_2 as $A=(a_1, ..., a_i, ...)$, as in Section 2. Suppose $G=C_1 \wedge \cdots \wedge C_l$. Let $N=\{1, ..., \lceil n^\varepsilon \rceil\}$, $\mathscr{X}=\{\langle Z, s\rangle \mid Z\subseteq N, \quad s\in\{0,1\}^{|Z|}\}$. Define $\|\langle Z, s\rangle\|=|Z|$, the cardinality of Z. A record $\mathscr{S}$ is a finite sequence $\langle X_1, ..., X_l\rangle$, where $X_i\in\mathscr{X}$. Define the norm $\|\mathscr{S}\|=\sum_{i=1}^{l}\|X_i\|$.

Intuitively, when a record element $X_i=\langle Z, s\rangle$ is generated in a certain round of the game, Z codes the variables to be assigned and s codes the assignment made by the player, in that round.

The coding scheme in Z is an indirect addressing. Specifically, if $J=\{i_1, i_2, ..., i_\alpha\}$, and $Z=\{z_1, z_2, ..., z_\beta\}$, where $\alpha, \beta\geqslant 0$, $1\leqslant i_1<\cdots<i_\alpha$, $1\leqslant z_1<\cdots<z_\beta$, then Z codes the subset of J:

$$\{i_{z_1}, ..., i_{z_\beta}\}, \qquad \text{if} \quad z_\beta\leqslant\alpha.$$

We denote this set as $J\downarrow Z$. If $Z=\varnothing$, then $J\downarrow Z=\varnothing$. If $z_\beta>\alpha$, $J\downarrow Z$ is undefined. Conversely for $\Delta=\{i_{z_1}, ..., i_{z_\beta}\}\subseteq J$, we denote

$$J\uparrow\Delta=\{z_1, ..., z_\beta\}.$$

Note that $J\uparrow\varnothing=\varnothing$. Clearly for $\Delta\subseteq J$, $J\downarrow(J\uparrow\Delta)=\Delta$.

Our game is played in rounds. The master executes the program, and in certain rounds, he asks the player to assign a few Boolean variables. Then the master continues, until the program halts. When the program halts, it halts in "result" or in "abort."

The procedure Recording is similar; for technical reasons, we first present Recording. A record $\mathscr{S}=\langle X_1, ..., X_l\rangle$ is given. There are four essential variables Θ, Y^+, Y^*, and N^*, respectively, representing our knowledge about the assignement A at any given point in the execution (more accurately, our knowledge about A which we can be *forced* to acknowledge). Here are some intuitive ideas behind the procedure (they should be taken as such only). The variable Y^+ will collect indices which correspond to variables that are assigned to be true by the given assignment Q_1Q_2. Similarly, Y^* and N^* will correspond to variables that are assigned to $*$ by the given assignment but assigned to be true and false, respectively, by the player recorded in $\mathscr{S}$. And Θ will collect subsets of indices which contain variables that are assigned to be false by the given assignment. Our goal is to show that an *authentic large record* that was produced by a game play rarely occurs.

Procedure Recording(A, $\mathcal{S}$)
0 $\Theta, Y^+, Y^*, N^* := \varnothing$; $t := 0$; $List := [C_1, ..., C_l]$;
Repeat
1 **if** $List = \varnothing$ **then** case 1: $t < l \Rightarrow$ "abort";
case 2: $t \geqslant l \Rightarrow$ "result"
fi
2 let C_i be on the top of $List$
3 **if** $(\exists u \in J_{i-}, a_u = 0)$ **then** $\Theta := \Theta \cup \{J_{i-}\}$, delete C_i from $List$
4 **else** [critical round]
5 $t := t + 1$
6 $F := J_i - (Y^+ \cup Y^* \cup N^*)$
7 get $X_t = \langle Z, s \rangle$ from $\mathcal{S}$, "abort" if nonexistent
8 $D := F \downarrow Z$, "abort" if undefined
9 **if** $(D \neq \{j \in F \mid a_j = *\})$ **then** "abort"
10 **else** $Y^+ := Y^+ \cup \{u \in F \mid a_u = 1\}$
$\Theta := \Theta \cup \{\{u\} \mid u \in F, a_u = 0\}$
$Y^* := Y^* \cup \{u \in D \mid s$ assigns x_u to $1\}$
$N^* := N^* \cup \{u \in D \mid s$ assigns x_u to $0\}$
fi
11 **if** $(D = \varnothing)$ **then**
12 **if** $(\forall u \in J_{i+} - N^*, a_u = 0)$ **then** "abort" **fi**
fi
13 Delete any C_k from $List$ with
$J_{k-} \cap N^* \neq \varnothing$ or $J_{k+} \cap (Y^+ \cup Y^*) \neq \varnothing$
fi
End[Repeat]

Some properties of Recording are easily verified. Define $List_{\text{out}}$ to be the set of j such that C_j has been deleted from $List$. (In the following, c.r. is shorthand for "critical round"). We have

LEMMA 3.2. (1) Θ, Y^+, Y^*, N^*, *and* $List_{\text{out}}$ *are monotonically non-decreasing.*

Every time the **Repeat** *loop is entered, the following are true*:

(2) $\forall K \in \Theta, \exists u \in K, a_u = 0$.

(3) $\forall u \in Y^+, a_u = 1$.

(4) $\forall u \in Y^* \cup N^*, a_u = *$.

(5) $t = \#$ *of c.r. completed so far.*

(6) $\forall j \in List, J_{j-} \cap N^* = \varnothing$ *or* $J_{j+} \cap (Y^+ \cup Y^*) = \varnothing$.

(7) $\forall j \in List_{\text{out}}, J_{j-} \in \Theta$ *or* $J_{j-} \cap N^* \neq$ *or* $J_{j+} \cap (Y^+ \cup Y^*) \neq \varnothing$.

(8) $Y^* \cap N^* = \varnothing$.

Proof. A straightforward check. Q.E.D.

Next we define our Game. The Game is very much like Recording, except that the record $\mathcal{S}$ is produced as we go along, one slot per critical round. Spicifically,

- In the initialization part (line 0), add $\mathcal{S} := \varnothing$.
- Change line 1 to: **if** $List = \varnothing$ **then** "result" S.
- Change lines 7 and 8 to:

  Create X_t as follows:

  $Z := F\uparrow \{j \in F \mid a_j = *\}$;

  $D := F\downarrow Z (= \{j \in F \mid a_j = *\})$;

  for $j \in D$ **do** let the player assign x_j, and record the assignment in $\mathcal{S}$ with a binary string s of length $|D|$ (in the obvious way). $X_t := \langle Z, s \rangle$.

Let ρ be the assignment made by the player: $\rho_d = 0, 1$, or $*$; if $d \in N^*$, Y^*, or otherwise.

We wish to prove the following:

LEMMA 3.3. *The Game will eventually halt. When the Game halts,* $G \|_A |_\rho \equiv 1$ *(at line* 1*),* $or \equiv 0$ *(at line* 12*). Furthermore, let* $\mathcal{S}$ *be the record it created; then Recording*$(A, \mathcal{S})$ *will run in precisely the same way as Game*$(A, \mathcal{S})$*, until halting.*

Proof. We claim that every completed round of the Game either deletes a clause or assigns a variable. The only nontrivial case is in a c.r. with $D = \varnothing$. If $D = \varnothing$ and the round is completed, the condition at line 12 must be false. Thus $\exists t \in J_{i+} - N^*$, $a_t = 1$ (there is no $*$ in F), where i is the index of the current clause C_i. But then C_i must be deleted at line 13.

Therefore the Game will eventually halt. Let $\mathcal{S}_0$ be the record created when the Game halts.

We prove by induction that Recording $(A, \mathcal{S}_0)$ will reproduce this play of the game. Suppose they both enter a new round with all the variables having the same value. (This is certainly true initially.) Also assume $t =$ the length of $\mathcal{S}_0$ constructed so far in the game.

If $List = \varnothing$, then the game halts as a "result." Since $t = l$, the length of $\mathcal{S}_0$, Recording$(A, \mathcal{S}_0)$ will also halt as a "result."

Suppose $List \neq \varnothing$. Then they get the same C_i and the same condition at line 3 (same A!). If the condition is satisfied, then the induction is completed. Suppose not. They come to line 7. The Game creates the next X_t. Since X_t is never altered later in the Game, it is what Recording obtains from $\mathcal{S}_0$. From the way X_t is created in the Game, Recording will not halt at lines 7, 8, and 9. Now for the rest of this round they have the same code. The induction is completed.

We have proved that Recording will reproduce the play of the game, and hence Lemma 3.2 applies to the procedure Game. In fact we proved something more, namely, that the Game can halt only at line 1 or line 12.

It follows from Lemma 3.2(4), 8) that ρ is a valid assignment to $G\|_A$. If the Game halts at line 1, then $G\|_A|_\rho \equiv 1$, by Lemma 3.2(2), 3), 7). If the Game halts at line 12, then we claim that $C_i\|_A|_\rho \equiv 0$, hence $G\|_A|_\rho \equiv 0$.

By (6) of Lemma 3.2, $F = [J_{i-} - (Y^+ \cup Y^*)] \cup [J_{i+} - N^*]$, at line 6. Since $D = \varnothing$, Y^* and N^* are unchanged at line 10. Clearly the only way to satisfy C_i is in F. But if $t \in J_{i-} - Y^*$, $a_t = 1$, by lines 3, 9, and 11; and if $t \in J_{i+} - N^*$, $a_t = 0$, by line 12. So $C_i\|_A|_\rho \equiv 0$. Q.E.D.

Let $\mathscr{A} = \{A: G\|_A$ is *not* n^ε-monochromatic$\}$. For a record $\mathscr{S}$, let $\mathscr{A}[\mathscr{S}] = \{A:$ Recording$(A, \mathscr{S})$ "results"$\}$.

If $A \in \mathscr{A}$, then for any game in particular for our Game, the master has no winning strategy. Hence there is a play in which the player assigned $\lceil n^\varepsilon \rceil$ many variables and still the circuit is not constant.

Because the circuit is not constantly 0, there is a satisfying assignment σ. Now for the rest of the Game, the player adopts the following strategy: assign any new variable according to σ. Since this strategy keeps the circuit satisfiable and the Game eventually halts, the Game must halt with the circuit equal to constant 1. Hence, the Game "results" with some $\mathscr{S}$, where $\|\mathscr{S}\| > \lceil n^\varepsilon \rceil$. Therefore,

$$\mathscr{A} \subseteq \bigcup \mathscr{A}[\mathscr{S}],$$

where the union is over all $\mathscr{S}$, with $\|\mathscr{S}\| > \lceil n^\varepsilon \rceil$.

3.2. *A Probability Analysis*

In this section, we will focus on Recording $(A, \mathscr{S})$. For a fixed $\mathscr{S}$ with $\|\mathscr{S}\| > \lceil n^\varepsilon \rceil$, we consider the probability that Recording$(A, \mathscr{S})$ results, where $A = Q_1 Q_2$, $Q_1 \in \mathscr{R}^-_{1-q}$, and $Q_2 \in \mathscr{R}^+_{1-q}$.

Define $\mathscr{A}^{m\gamma} = \{A:$ Recording$(A, \mathscr{S})$ will come to its mth c.r. with $(\Theta, Y^+, Y^*, N^*) = \gamma\}$, $\Gamma^m = \{\gamma: \mathscr{A}^{m\gamma} \neq \varnothing\}$, $\mathscr{A}^m = \bigcup_{\gamma \in \Gamma^m} \mathscr{A}^{m\gamma}$.

We first derive a condition for $A \in \mathscr{A}^{m\gamma}$.

LEMMA 3.4. *For any* $\gamma = (\gamma^\Theta, \gamma^+, \gamma^Y, \gamma^N) \in \Gamma^m$, *there exists* i_γ, *such that* $A \in \mathscr{A}^{m\gamma} \Leftrightarrow A$ *satisfies the following conditions*:

(I) $\forall K \in \gamma^\Theta$, $\exists u \in K$, $a_u = 0$.

(II) $\forall u \in \gamma^+$, $a_u = 1$.

(III) $\forall u \in \gamma^Y \cup \gamma^N$, $a_u = *$.

(IV) $\forall u \in J_{i_\gamma -}$, $a_u \neq 0$.

Let us prove the following lemma first: Pick any $A^0 \in \mathscr{A}^{m\gamma}$.

LEMMA 3.5. *A satisfies conditions* (I), (II), *and* (III) $\Rightarrow$ *Recording for* A^0 *and* A *will run precisely the same* (*with all the variables* Θ, Y^+, Y^*, N^*, *List, and* t *the same at corresponding moments*) *up to line* 1 *of the* m*th c.r. of* A^0.

Proof. By induction. Suppose they are at line 1 of the m_0th round of A^0 (including m' c.r. and $m' < m$), and so far they are all the same (trivially true for the base case $m_0 = 1$).

If this is the mth c.r. for A^0, then the induction is completed. Suppose it is not. Hence A^0 will complete this round without halting. Thus $List \neq \varnothing$ and they pick the same C_i.

If A^0 satisfies the condition at line 3, then $J_{i-} \in \gamma^{\Theta}$, since the mth c.r. of A^0 is yet to come. By (I), A satisfies the same condition at line 3.

If A^0 fails the condition at line 3, this is a c.r. of A^0, but not the mth yet. A^0 will successfully record all $u \in J_{i-} \cup J_{i+}$ in (Θ, Y^+, Y^*, N^*), which will later become γ. In particular, $\forall u \in J_{i-}$, $u \in \gamma^+ \cup \gamma^Y \cup \gamma^N$. By (II) and (III), A must also fail the condition at line 3.

Hence either A^0 and A both finish the current round at line 3, in which case the induction is completed; or they both advance to line 4. Suppose then that this is a c.r. for both. They must find (the same) D well defined. As we noted, A^0 will record all $u \in J_{i-} \cup J_{i+}$ which will appear in γ.

In particular, by (I), (II), and (III), A must also find D to be precisely the set of $*$'s in F, and thus update Y^* and N^* in exactly the same way. Similarly, A must update Θ and Y^+ in the same way that A^0 does, by (I) and (II).

Now if $D \neq \varnothing$, we are done. If $D = \varnothing$, then A^0 will find the condition at line 12 to be false; i.e., $\exists t_0 \in J_{i+} - N^*$, $a^0_{t_0} \neq 0$. But $D = \varnothing \Rightarrow a^0_{t_0} = 1$. Hence $t_0 \in Y^+$, which is the same for both A and A_0. Hence $a_{t_0} = 1$ as well. Therefore A will not halt there. The induction is completed. Q.E.D.

Proof of Lemma 3.4. Pick $A^0 \in \mathscr{A}^{m\gamma}$ and run Recording$(A^0, \mathscr{S})$; let C_{i_γ} be the clause under consideration in its mth c.r.

$\Rightarrow$ Since $A \in \mathscr{A}^{m\gamma}$, A satisfies (I), (II), and (III). By Lemma 3.5, A and A^0 will reach line 1 of the mth c.r. of A_0, with all the variables the same. Since A^0 comes to line 4, $List \neq \varnothing$, which is the same as for A; so they both pick up C_{i_γ}. Since this is the mth c.r. for A^0, A^0 will fail the condition at line 3, and enter its c.r. with Θ unchanged. Thus γ^{Θ} is the common value for Θ when A and A^0 entered the current round at line 1.

If A were to satisfy the condition at line 3, then this is not a c.r. for A, and its mth c.r. is yet to come. Since $J_{i_\gamma -}$ is now added to Θ by A, $J_{i_\gamma -} \in \gamma^{\Theta}$. In other words, $J_{i_\gamma -} \in \Theta$ when they entered at line 1. But then A^0 must have satisfied the condition at line 3. A contradiction. Therefore A satisfies (IV).

$\Leftarrow$ Again by Lemma 3.5, we can assume they arrive at line 1 of the mth c.r. of A^0, with exactly the same history.

$A^0 \in \mathscr{A}^{m\gamma} \Rightarrow List \neq \varnothing$ and A picks up C_{i_γ}. Then (IV) says that this is also a c.r. for A. Since this is the mth c.r. for A^0, $A^0 \in \mathscr{A}^{m\gamma}$, and so far they are the same; this is also the mth c.r. for A, with $(\Theta, Y^+, Y^*, N^*) = \gamma$. Hence $A \in \mathscr{A}^{m\gamma}$. Q.E.D.

Now we are ready to estimate the probability $\Pr(\mathscr{A}[\mathscr{S}])$. Let E^m denote the event that Recording$(A, \mathscr{S})$ completes its mth c.r. without halting:

$$\begin{aligned}\Pr(\mathscr{A}[\mathscr{S}]) &\leqslant \Pr(\mathscr{A}^1)\cdot \Pi_{1\leqslant m<l}\Pr(A\in\mathscr{A}^{m+1}\mid A\in\mathscr{A}^m)\cdot\Pr(E^l\mid A\in\mathscr{A}^l)\\ &\leqslant [\Pi_{1\leqslant m<l}\Pr(E^m\mid A\in\mathscr{A}^m)]\cdot\Pr(E^l\mid A\in\mathscr{A}^l)\\ &= \Pi_{1\leqslant m\leqslant l}\Pr(E^m\mid A\in\mathscr{A}^m).\end{aligned}$$

We show the following:

LEMMA 3.6. $\forall\gamma\in\Gamma^m$,

$$\begin{aligned}\Pr(E^m\mid A\in\mathscr{A}^{m\gamma}) &\leqslant q^{\|X_m\|} \qquad \text{if} \quad \|X_m\|\neq 0\\ &\leqslant q\cdot n^{\varepsilon} \qquad \text{otherwise.}\end{aligned}$$

Clearly, Lemma 3.6 implies the same bound for $\Pr(E^m\mid A\in\mathscr{A}^m)$, since it can be estimated as

$$\sum_{\gamma\in\Gamma^m}\Pr(E^m\mid A\in\mathscr{A}^{m\gamma})\cdot\Pr(A\in\mathscr{A}^{m\gamma}\mid A\in\mathscr{A}^m),$$

and

$$\sum_{\gamma\in\Gamma^m}\Pr(A\in\mathscr{A}^{m\gamma}\mid A\in\mathscr{A}^m)=1.$$

Hence,

COROLLARY 3.7. $\Pr(\mathscr{A}[\mathscr{S}])\leqslant q^{\|\mathscr{S}\|}\cdot(qn^{\varepsilon})^{l'}$, *where* $l'=\#$ *of* X_i *in* $\mathscr{S}$ *with* $\|X_i\|=0$.

We use Lemma 3.4 to prove Lemma 3.6.

Proof of Lemma 3.6. Assume $\|X_m\|\neq 0$. Consider a random assignment taken from $\mathscr{R}^-_{1-q}$, followed by one from $\mathscr{R}^+_{1-q}$, on the variables in $F=J_{i_\gamma}-(\gamma^+\cup\gamma^Y\cup\gamma^N)$. We refer to the procedure Recording. In order to survive the mth critical round, we must have $D=F\downarrow X_m=\{j\in F\mid a_j=*\}$. Clearly the conditions on the random assignment of Lemma 3.4 can be strengthened so that all variables in F are assigned $*$ by the first round $\mathscr{R}^-_{1-q}$ (since in order to remain $*$ after two sweeps, it must remain $*$ after the first.) Note that originally the conditions from Lemma 3.4 on F were with $\mathscr{R}^-_{1-q}$ only. For $\mathscr{R}^+_{1-q}$, a given $u\in F$ is assigned $*$ only with probability q. Thus we have the upper bound $q^{\|X_m\|}$.

In the case $\|X_m\|=0$, we estimate

$$\begin{aligned}&\Pr(D=\varnothing \text{ is all the } *\text{'s in } F\wedge\exists t\in J_{i_\gamma+}-\gamma^N, a_t\neq 0\mid A\in A^{m\gamma})\\ &\leqslant \Pr(\exists t\in J_{i_\gamma+}-\gamma^N, a_t=1\mid A\in\mathscr{A}^{m\gamma}).\end{aligned}$$

We consider two sweeps from $\mathscr{R}^-_{1-q}$ followed by one from $\mathscr{R}^+_{1-q}$, on $J_{i_\gamma+}-\gamma^N$. Conditions (II), (III), and (IV) are irrelevant now (using independence). And con-

dition (I) would only reduce the probability for a given $u \in J_{i_\gamma +} - \gamma^N$ to be assigned 1. Unconditionally, a given u is assigned $*$ by $\mathscr{R}_{1-q}^{-}$ with probability q, hence the upper bound qn^ε, where n^ε comes from the *bfi* condition $|J_{i_\gamma}| \leqslant N^\varepsilon$. Q.E.D.

Now we can finally estimate $\Pr(\mathscr{A})$. It is bounded above by

$$\sum_{\|\mathscr{S}\| > \lceil n \rceil} \Pr(\mathscr{A}[\mathscr{S}]),$$

which is bounded by

$$\sum_{N > \lceil n^\varepsilon \rceil} \sum_{l=1}^{N} \binom{N-1}{l-1} \sum_{l' \geqslant 0} \binom{l'+l}{l} (2\lceil n^\varepsilon \rceil)^N (qn^\varepsilon)^{l'} q^N,$$

where N runs through possible values of the norm of records, $l = \#$ of nonempty X_i in $\mathscr{S}$, and $l' = \#$ of empty X_i in $\mathscr{S}$.

Recall that $q \leqslant n^{-1.05\varepsilon}$. For a fixed ε, $\exists N_\varepsilon$, such that $\forall n > N_\varepsilon$, $16n^{-0.05\varepsilon} < 1/(2e)$. We get, for $n > N_\varepsilon$,

$$\begin{aligned}
\Pr(\mathscr{A}) &\leqslant \sum_{N > \lceil n^\varepsilon \rceil} 2^{N-1} (2\lceil n^\varepsilon \rceil)^N q^N \sum_{l=1}^{N} 2^l \sum_{l' \geqslant 0} 2^{l'} (qn^\varepsilon)^{l'} \\
&\leqslant 2 \sum_{N > \lceil n^\varepsilon \rceil} 2^N (2\lceil n^\varepsilon \rceil)^N q^N 2^N \\
&\leqslant 2 \sum_{N > \lceil n^\varepsilon \rceil} (16n^{-0.05\varepsilon})^N \\
&\leqslant e^{-n^\varepsilon}.
\end{aligned}$$

Hence $\Pr(\mathscr{A}) \leqslant Ce^{-n^\varepsilon}$, $\forall n$, where C only depends on ε. Theorem 3.1 is proven.

4. Depth k Circuits

Theorem 3.1 is proved under a "skewed" probabilistic assignment. We first "unskew" it:

Theorem 4.1. *Fix $0 < \varepsilon < \frac{1}{5}$. Then there exists a constant C, such that for any circuit $G \in \Pi_{2,n}$ (or $\Sigma_{2,n}$) with bfi $\leqslant n^\varepsilon$ and any p with $0 \leqslant p \leqslant n^{-2.2\varepsilon}$, and $Q \in \mathscr{R}_p$, $G\|_Q$ is n^ε-monochromatic, with probability $1 - \varepsilon_n$, where $\varepsilon_n \leqslant Ce^{-n^\varepsilon}$.*

Proof. Clearly we only need to prove the $\Pi_{2,n}$ case. For $0 \leqslant p \leqslant n^{-2.2\varepsilon}$, let

$$p' = 1 - \frac{1}{2} \frac{(1-p)^2}{1 - (p(2-p))^{1/2}},$$

$$q = \frac{(p(2-p))^{1/2} - p}{1-p}.$$

It is easy to verify that $0 \leqslant p' \approx \frac{1}{2} \leqslant 1$, $0 \leqslant q = O(p^{1/2}) \leqslant n^{-1.05\varepsilon}$. Take random $R \in \mathscr{R}_{p'}^{+}$, $Q_1 \in \mathscr{R}_{1-q}^{-}$, and $Q_2 \in \mathscr{R}_{1-q}^{+}$. It is straightforward to show that RQ_1Q_2 has the same distribution as $Q \in \mathscr{R}_p$.

Now we apply Theorem 3.1 to each $G \|_R$ and the result follows. Q.E.D.

THEOREM 4.2. *Let $k \geqslant 2$, $1 \leqslant j \leqslant k-1$. Let $p = n^{-1/k}$, and let*

$$0 < \frac{1}{4k} = \varepsilon_k < \varepsilon_{k-1} < \cdots < \varepsilon_1 = \frac{1}{3k}$$

be equally spaced.

For any $G \in \Pi_{j+1,n}$ (or $\Sigma_{j+1,n}$) with bfi $\leqslant n^{1/3k}$ and size$(G) \leqslant e^{n^{1/4k}}$, and random $A_1, \ldots, A_j$ from $\mathscr{R}_p$, $G \|_{A_1,\ldots,A_j}$ is $n^{1/3k}$-monochromatic with probability $1 - O(\exp(-n^{\varepsilon_j}))$.

Note. The constant in the O-notation depends only on k.

Proof. Fixing $k \geqslant 2$, we prove the theorem by induction on j. Base case $j = 1$. $G \in \Pi_{2,n}$, with $bfi \leqslant n^{1/3k}$. Taking $\varepsilon = 1/3k < \frac{1}{5}$, $p = n^{-1/k} \leqslant n^{-2.2\varepsilon}$ in Theorem 4.1, we have $G \|_{A_1}$ is $n^{1/3k}$-monochromatic with probability $1 - O(\exp(-n^{1/3k}))$. The proof is similar for $G \in \Sigma_{2,n}$.

Now suppose $j > 1$, and the theorem is true for $j-1$. We prove the theorem for the $G \in \Sigma_{j+1,n}$ case. The $\Pi_{j+1,n}$ case is dual.

Let $G = \Sigma_{i=1}^{l} K_i$, where $K_i \in \Pi_{j,n}$. Since G has $bfi \leqslant n^{1/3k}$, and $size(G) \leqslant e^{n^{1/4k}}$, $l \leqslant e^{n^{1/4k}}$, and every K_i inherits the condition on *bfi* and *size*. Let $B_i = K_i \|_{A_1,\ldots,A_{j-1}}$; then $G \|_{A_1,\ldots,A_{j-1}} = \sum_{i=1}^{l} B_i$.

By our inductive hypothesis, for any i fixed, we have B_i is $n^{1/3k}$-monochromatic, with probability $1 - O(\exp(-n^{\varepsilon_{j-1}}))$, where the constant is independent of B_i. Hence, with probability $1 - O(\exp(-n^{(\varepsilon_{j-1}+\varepsilon_j)/2}))$ all B_i are simultaneously $n^{1/3k}$-monochromatic. Again, the constant here depends only on k.

By Lemma 2.2, all B_i are equivalent to $\Sigma_{2,n}$-formulae, and thus with probability $1 - O(\exp(-n^{(\varepsilon_{j-1}+\varepsilon_j)/2}))$, $G \|_{A_1,\ldots,A_{j-1}}$ is equivalent to a Σ_2-formula with $bfi \leqslant n^{1/3k}$. Applying Theorem 4.1 once more, we get that $G \|_{A_1,\ldots,A_j}$ is $n^{1/3k}$-monochromatic with probability $1 - O(\exp(-n^{\varepsilon_j}))$. Q.E.D.

Taking $j = k-1$ in Theorem 4.2, we obtain:

COROLLARY 4.3. *Let $k \geqslant 2$, $p = n^{-(k-1)/k}$. For any $G \in \Pi_{k,n}$ (or $\Sigma_{k,n}$) with $bfi \leqslant n^{1/3k}$ and size$(G) \leqslant e^{n^{1/4k}}$, and a random $R \in \mathscr{R}_p$, $G \|_R$ is $n^{1/3k}$-monochromatic with probability $1 - o(1)$, uniformly.*

We note that the restriction on *bfi* is only technical; one may always extend one more level of alternation to have $bfi \leqslant 1$.

5. Circuits vs Parity

In this section we complete the proof of Theorem 2.1. By the remark at the end of last section, we need only prove:

Theorem 5.1. *Let $k \geq 2$. There exists a sequence $\{\alpha_n\}$, $\alpha_n \to \frac{1}{2}$, such that all depth k Boolean circuits, with n inputs, size $\leq \exp(n^{1/4k})$, and bfi $\leq n^{1/3k}$, when computing Parity$_n$, make errors on $\geq \alpha_n$ of all 2^n possible inputs.*

The strategy to prove Theorem 5.1 is the following: Fix $k \geq 2$. Consider any depth k circuit G satisfying the conditions. Randomly take a total assignement σ (all 2^n many assignements from $\{0, 1\}^n$ are equally likely). We wish to prove that $G|_\sigma \neq Parity_n|_\sigma$, with probability $\frac{1}{2} - o(1)$, where $o(1)$ may depend on k, but it is independent of G.

Now we pick σ in two stages: First, randomly pick a "single $*$" σ^*, so that all $\sigma^* \in A^* \equiv \{\sigma \in \{0, 1, *\}^n \mid \exists \text{ a unique } d, \sigma_d = *\}$ are equally likely. Then assign the unique $*$ in σ^* to 0 or 1 with equal probability, to obtain our radom σ.

Theorem 5.1 will be proved if we can show that $G|_{\sigma^*} \equiv$ constant, with probability $1 - o(1)$, since for any σ^*, the conditional probability for failure is

$$\Pr(G|_\sigma \neq Parity_n|_\sigma \mid G|_{\sigma^*} \equiv \text{constant}) = 50\%.$$

Now our strategy to generate a random $\sigma^* \in A^*$ is the following: Let $p = n^{-(k-1)/k}$ and $q = n^{-1/2.5k}$. For a nonempty finite set of variables S, an "A^*-uniform assignement" on S is a random assignment that randomly picks one variable in S as $*$ and uniformly assigns the others to 0 or 1.

Procedure Generate 1 (σ^*)
  Take a random $A \in \mathscr{R}_p$
  if (A leaves $\leq n^{1/2k}$ variables in X unassigned)
  then take a random σ^*
  else take a random $B \in \mathscr{R}_q$
    if (AB assigned every variable in X)
    then take a random σ^*
    else let $S = \{x_i \in X \mid AB$ assigns x_i to $*\}$,
      A^*-uniformly assign S,
      let σ^* be the result.
    fi
  fi Return (σ^*)

Clearly Generate 1 does generate every $\sigma^* \in A^*$ equally likely. Now we "realize" Generate 1 by the following procedure, which will complete our proof:

Procedure Generate 2 (σ^*, *tag*)
Take a random $A \in \mathscr{R}_p$
if (A leaves $\leqslant n^{1/2k}$ variables in X unassigned)
then *tag* := failure, take a random σ^*
else **if** ($G \|_A$ is not $n^{1/3k}$-monochromatic)
then *tag* := failure, take a random $B \in \mathscr{R}_q$
if (AB assigned every variable in X)
then take a random σ^*
else let $S = \{x_i \in X \mid AB$ assigns x_i to $*\}$,
A^*-uniformly assign S,
let σ^* be the result.
fi
else play the game (as the player),
assign any given variable with distribution $\mathscr{R}_q$
if (the master ever gets a $*$)
then stop the game, *tag* := failure,
run through $\mathscr{R}_q$ for the remaining variables,
let $S = \{x_i \in X \mid x_i$ is unassigned so far$\}$,
A^*-uniformly assign S,
Let σ^* be the result.
else when the game is finished, run through $\mathscr{R}_q$ for the rest,
if (no variable is assigned $*$)
then *tag* := failure, take a random σ^*
else *tag* := success,
let $S = \{x_i \in X \mid AB$ assigns x_i to $*\}$,
A^*-uniformly assign S,
Let σ^* be the result.
fi
fi
fi
fi Return (σ^*, *tag*)

Clearly if we ignore the *tag*, Generate 2 is the same as Generate 1. If Generate 2 returns (σ^*, success), then $G|_{\sigma^*} \equiv$ constant. Let F denote the event that Generate 2 returns with *tag* = failure. We claim:

$$\Pr(F) = o(1).$$

We only need to verify:

1. $\Pr(A \in \mathscr{R}_p$ leaves $\leqslant n^{1/2k}$ variables in X unassigned$) = o(1)$. This follows from Chebechev's inequality.

2. $\Pr(G \|_A$ is not $n^{1/3k}$-monochromatic$) = o(1)$. This is Corollary 4.3.

3. Pr(the master gets a $*$ under $\mathscr{R}_q \mid G \parallel_A$ is $n^{1/3k}$-monochromatic) $= o(1)$. This is because $n^{1/3k} \cdot n^{-1/2.5k} \to 0$.

4. Pr(AB leaves no $*$ in $X \mid A$ leaves $\geqslant n^{1/2k}$ variables unassigned in X) $= o(1)$. This is trivial.

6. Final Remarks

The result concerning circuit and parity is of interest independently of relativization. After all, one cannot do worse than 50% error for parity.

The following corollary is evident.

Corollary 6.1 (Yao). *There is a recursive oracle A separating PSPACE from the polynomial-time hierarchy.*

The proof is simple. Observe that with probability one the parity language $Parity^A$ is not in PH^A. Hence for those A, $L^A(M_i)$ differs from $Parity^A$ infinitely often for any PH machine M_i. By the definition of measure μ, any initial segment of A corresponds to a small interval of $[0, 1]$. Now suppose we are given an initial segment of A, the oracle constructed so far, and we want to diagonalize over M_i. What we do is simply look for an extension that kicks M_i out. The "brute force" method must succeed due to our probability one separation.

Shortly after this work, Hastad [9] obtained a simplification of Yao's proof, improving the bound on the circuit size from $\Omega(e^{n^{1/4k}})$ to $\Omega(e^{cn^{1/k}})$. Later Babai [Bab86] obtained the result in Theorem 1.1 by a short proof, assuming Yao's theorem and a result by Ajtai [Ajt83].

The following question is still open:

- Is it true that with probability one, a random oracle separates the polynomial-time hierachy PH into an infinite hierarchy?

Acknowledgments

The author expresses his sincere gratitude to Professor Juris Harmanis for his constant encouragement and inspiration, without which this work would have been impossible. The author also thanks Professors R. Book, N. Immerman, D. Joseph, S. Mahaney, Y. Moschovakis, R. Shore, and A. Yao and my fellow students L. Hemachandra, J. Johnstone, and S. Smith for very stimulating conversations. Thanks are also due to the anonymous referee, whose comments have enhanced the presentation.

References

[Ajt83] M. Ajtai, Σ_1^1-formulae on finite structures, *Ann. Pure Appl. Logic* **24** (1983), 1–48.

[Bab86] L. Babai, A random oracle separates *PSPACE* from polynomial hierarchy, Notes, 1986, to appear.

[BG81] C. BENNET AND J. GILL, Relative to a random oracle A, $P^A \neq NP^A$ with probability 1, *SIAM J. Comput.* **10** (1981), 96–113.

[BGS75] T. BAKER, J. GILL, AND R. SOLOVAY, Relativization of $P = ?NP$ question, *SIAM J. Comput.* **4** (1975), 431–442.

[BS79] T. BAKER AND A. SELMAN, A second step toward the Polynomial hierarchy, *Theoret. Comput. Sci.* **8** (1979), 177–187.

[CH86] J. CAI AND L. A. HEMACHANDRA, The Boolean hierarchy: Hardware over *NP*, *in* "Structure in Complexity Theory," pp. 105–124, Lecture Notes in Computer Science, Vol. 223, Springer-Verlag, New York/Berlin, 1986.

[CKS81] A. CHANDRA, D. KOZEN, AND L. STOCKMEYER, Alternation, *J. Assoc. Comput. Mach.* **26**, No. 1 (1981).

[FSS84] M. FURST, J. SAXE, AND M. SIPSER, Parity, circuits, and the polynomial-time hierarchy, *Math. Systems Theory* **17** (1984), 13–27.

[Has86] J. HASTAD, Almost optimal lower bounds for small depth circuits, *in* "Proceedings, ACM Symposium on Theory of Computation, 1986," pp. 6–20.

[Hu79] J. HOPCROFT AND J. ULLMAN, "Introduction to Automata Theory, Languages, and Computation," Addison–Wesley, Reading, MA, 1979.

[Sip83] M. SIPSER, Borel sets and circuit complexity, *in* "Proceedings, ACM Symposium on Theory of Computation, 1983," pp. 61–69.

[Sto77] L. STOCKMEYER, The polynomial-time hierarchy, *Theoret. Comput. Sci.* **3** (1977), 1–22.

[Yao85] A. YAO, Separating the polynomial-time hierarchy by oracles, *in* "Proceedings, IEEE Annual Symposium on Foundations of Computer Science, 1985," pp. 1–10.