Holographic Algorithms with Matchgates Capture Precisely Tractable Planar #CSP

Jin-Yi Cai* Pinyan Lu† Mingji Xia‡

Abstract

Valiant introduced matchgate computation and holographic algorithms. A number of seemingly exponential time problems can be solved by this novel algorithmic paradigm in polynomial time. We show that, in a very strong sense, matchgate computations and holographic algorithms based on them provide a universal methodology to a broad class of counting problems studied in statistical physics community for decades. They capture precisely those problems which are #P-hard on general graphs but computable in polynomial time on planar graphs.

More precisely, we prove complexity dichotomy theorems in the framework of counting CSP problems. The local constraint functions take Boolean inputs, and can be arbitrary real-valued symmetric functions. We prove that, *every* problem in this class belongs to precisely three categories: (1) those which are tractable (i.e., polynomial time computable) on general graphs, or (2) those which are #P-hard on general graphs but tractable on planar graphs, or (3) those which are #P-hard even on planar graphs. The classification criteria are explicit. Moreover, problems in category (2) are tractable on planar graphs precisely by holographic algorithms with matchgates.

1 Introduction

Given a set of functions $\mathcal{F}$, the Counting Constraint Satisfaction Problem #CSP($\mathcal{F}$) is the following problem: An input instance consists of a set of *variables* $X = \{x_1, x_2, \ldots, x_n\}$ and a set of *constraints* where each constraint is a function $f \in \mathcal{F}$ applied to some variables in X. The output is the sum, over all assignments to X, of the products of these function evaluations. This sum-of-product evaluation is called the partition function. In the special case where $f \in \mathcal{F}$ outputs values in $\{0, 1\}$ it counts the number of satisfying assignments. But constraint functions taking real or complex values are also interesting, called (real or complex) weighted #CSP. Our $\mathcal{F}$ consists of real or complex valued functions in general. There is a deeper reason for allowing this generality: The theory of *holographic reductions* is a powerful tool which operates naturally over $\mathbb{C}$, even if the original problem has only 0-1 valued functions.

A closely related framework for locally constrained counting problems is called Holant Problems [7, 9]. This framework is inspired by the introduction of *Holographic Algorithms* by L. Valiant [22, 21]. In two ground-breaking papers [20, 22] Valiant introduced matchgates and holographic algorithms based on matchgates to solve a number of problems in polynomial time, which appear to require exponential time. At the heart of these exotic algorithms is a tensor transformation from a given problem to the problem of counting (complex) weighted perfect matchings over planar graphs. The latter problem has a remarkable P-time algorithm (FKT-algorithm) [16, 12, 13]. Planarity is crucial, as counting perfect matchings over general graphs is #P-hard [?]. Most of these holographic algorithms use a suitable linear basis to realize locally a *symmetric* function with at most 3 Boolean variables on a matchgate. This work has been extended in [5]. In particular we have obtained a complete characterization of all realizable symmetric functions by matchgates over the complex field $\mathbb{C}$.

The study of "tractable #CSP" type problems has a much longer history in the statistical physics community (under different names). Ever since Wilhelm Lenz who invented what is now known as the Ising model, and asked his student Ernst Ising [?] to work on it, physicists have studied so-called "Exactly Solved Models" [?, ?]. In the language of modern complexity theory, physicists' notion of an "Exactly Solvable" system corresponds

*University of Wisconsin-Madison. jyc@cs.wisc.edu. Supported by NSF CCF-0830488 and CCF-0511679.
†Microsoft Research Asia. pinyanl@microsoft.com
‡Institute of Software, Chinese Academy of Sciences. xmjljx@gmail.com

to systems with polynomial time computable partition functions. This is captured completely by the computer science notion of "tractable #CSP". In Physics, many great researchers worked to build this intellectual edifice, with remarkable contributions by Ising, Onsager, Fisher, Temperley, Kasteleyn, C.N.Yang, T.D.Lee, Baxter, Lieb, Wilson etc. A central question is to identify what "systems" can be solved "exactly" and what "systems" are "difficult". The basic conclusion from physicists is that some "systems", including the Ising model, are "exactly solvable" for planar graphs, but they appear difficult for higher dimensions. There does not exist any rigorous or provable classification. This is partly because the notion of a "difficult" partition function had no rigorous definition in physics. However, in the language of complexity theory, it is natural to consider the classification problem. In this paper we do that, in the more general setting of #CSP with real valued constraint functions. This will also shed light on why the valiant efforts by physicists to generalize the "exactly solved" planar system to higher dimensions failed. (In the appendix we will give some more background.)

Now turning from Physics to CS proper, after Valiant introduced his holographic algorithms with matchgates, the following question can be raised: Do these novel algorithms capture all P-time tractable counting problems on planar graphs, *or* are there other more exotic algorithmic paradigms yet undiscovered? A suspicion (and perhaps an audacious proposition) is that they have indeed captured all tractable planar counting problems. If so it would provide a universal methodology to a broad class of counting problems studied in statistical physics and beyond. The results of this paper can be viewed as an affirmation of that suspicion. Within the framework of weighted Boolean #CSP problems our answer is YES, for *all* symmetric real valued functions.

While #CSP problems provide a natural framework to address this question, it turns out that the deeper reason comes from Holant problems, which can be described as follows: An input graph $G = (V, E)$ is given, where each $v \in V$ is attached a function $f_v \in \mathcal{F}$, mapping $\{0,1\}^{\deg(v)} \to \mathbb{R}$ or $\mathbb{C}$. We consider all edge assignments $\sigma : E \to \{0,1\}$. For each σ, f_v takes its input bits from the incident edges $E(v)$ at v, and evaluates to $f_v(\sigma\mid_{E(v)})$. The counting problem on instance G is to compute $\text{Holant}_G = \sum_\sigma \prod_{v\in V} f_v(\sigma\mid_{E(v)})$. In effect, in a Holant problem, edges are variables and vertices represent constraint functions. This framework is very natural; e.g., the problem of Perfect Matching corresponds to attaching the Exact-One function at each vertex, taking 0-1 inputs. The class of all Holant problems with function set $\mathcal{F}$ is denoted by $\text{Holant}(\mathcal{F})$.

Every #CSP problem can be simulated by a Holant problem. Represent any instance of a #CSP problem by a bipartite graph where LHS are labeled by variables and RHS are labeled by constraints. Denote by $=_k$: $\{0,1\}^k \to \{0,1\}$ the Equality function of arity k, which is 1 on 0^k and 1^k, and is 0 elsewhere. Then we can turn the #CSP instance to an input graph of a Holant problem, by replacing every variable vertex v on LHS by $=_{\deg(v)}$. In fact, $\#\text{CSP}(\mathcal{F})$ is exactly the same as $\text{Holant}(\mathcal{F} \cup \{=_k \mid k \geq 1\})$. Thus, #CSP problems can be viewed as Holant problems where all Equality functions are available for free, or assumed to be present. However, when we wish to discuss some restricted classes of counting problems, e.g., for 3-regular graphs, the framework of Holant problems is the more natural one. And as it turns out, the main technical breakthrough for our dichotomy theorem for planar #CSP comes from Holant problems.

In this paper we will only consider Boolean variables X. For a symmetric function on k variables, we denote it as $[f_0, f_1, \ldots, f_k]$, where f_i is the value of f on inputs of Hamming weight i. E.g., $(=_1) = [1,1], (=_2) = [1,0,1]$ and $(=_3) = [1,0,0,1]$ etc. When we relax Holant problems by allowing all Equality functions for free, we obtain #CSP. We can also consider other relaxations. Let $\mathbf{0} = [1,0]$ and $\mathbf{1} = [0,1]$ denote the constant 0 and 1 unary (arity 1) functions. Then Holant^c is the natural class of Holant problems where $\mathbf{0}$ and $\mathbf{1}$ are free. This amounts to computing Holant on input graphs where we can set 0 or 1 to some dangling edges (one end has degree 1). Another class of Holant problems is called Holant^* problems where we assume all unary functions $[u_0, u_1]$ are free.

In [9] we obtained a dichotomy theorem for (complex) Holant^* problems and (real) Holant^c problems. The dichotomy criterion for Holant^* problems is still valid for *planar graphs*. The proof of dichotomy theorems in this paper starts from there.

In Section 4, we prove that for any real-valued symmetric function set $\mathcal{F}$, the planar $\text{Holant}^c(\mathcal{F})$ problem is tractable (i.e., computable in P) iff either it is tractable over general graphs (for which we already have an effective dichotomy theorem [9]), or it is tractable because every function in $\mathcal{F}$ is realizable by a matchgate, in which case the planar $\text{Holant}^c(\mathcal{F})$ problem is computable by matchgates in P-time using FKT. In *all other cases* the problem is #P-hard.[1] A crucial ingredient of the proof is a cross-over construction whose validity is proved algebraically, which seems to defy any direct combinatorial justification.

[1]Strictly speaking, we must only consider $\mathcal{F}$ where functions take computable real numbers; this will be assumed implicitly.

Our second theorem (Section 5) is about planar #CSP problems. We prove that for any set of real-valued symmetric functions $\mathcal{F}$, the planar #CSP($\mathcal{F}$) problem is tractable iff either it is tractable as #CSP($\mathcal{F}$) without the planarity restriction (for which we have an effective dichotomy theorem [9]), or it is tractable because every function in $\mathcal{F}$ is realizable by a matchgate under a specific holographic basis transformation. Thus planar #CSP($\mathcal{F}$) is solvable by a holographic algorithm in the second case. For all other $\mathcal{F}$ the problem is #P-hard. The proof of this dichotomy theorem for planar #CSP is built on the one for planar Holant^c in Section 4.

Our third result is a dichotomy theorem for planar 2-3 regular bipartite Holant problems (Section 6). (This theorem deals with Holant problems without assuming unary **0** and **1**.) This includes Holant problems for 3-regular graphs as a special case. The tractability criterion is the same: Either it is tractable for general graphs (for which we also have an effective dichotomy theorem [4]), or it is tractable by a suitable holographic algorithm, which is a holographic reduction to FKT using matchgates. In all other cases the problem is #P-hard.

The three dichotomy theorems are not mutually subsumed by each other and are of independent interest. In each framework the respective theorem is a demonstration that holographic algorithms with matchgates capture precisely those #P-hard problems which become tractable for planar graphs.

2 Preliminaries

2.1 Problem and Definitions

Our functions take values in $\mathbb{C}$ by default. The framework of Holant problems is defined for functions mapping any $[q]^k \to \mathbb{C}$ for a finite q. Our results in this paper are for the Boolean case $q = 2$. So we give the following definitions only for $q = 2$ for notational simplicity.

A *signature grid* $\Omega = (H, \mathcal{F}, \pi)$ consists of a graph $H = (V, E)$, and a labeling π which labels each vertex with a function $f_v \in \mathcal{F}$. The Holant problem on instance Ω is to compute $\text{Holant}_\Omega = \sum_\sigma \prod_{v \in V} f_v(\sigma \mid_{E(v)})$, a sum over all edge assignments $\sigma : E \to \{0, 1\}$. A function f_v can be represented as a vector of length $2^{\deg(v)}$, or a tensor in $(\mathbb{C}^2)^{\otimes \deg(v)}$. A function $f \in \mathcal{F}$ is also called a *signature*. We denote by $=_k$ the EQUALITY signature of arity k. A symmetric function f on k Boolean variables can be expressed by $[f_0, f_1, \ldots, f_k]$, where f_i is the value of f on inputs of Hamming weight i. Thus, $(=_k) = [1, 0, \ldots, 0, 1]$ (with $k - 1$ zeros). A Holant problem is parameterized by a set of signatures.

Definition 2.1. *Given a set of signatures $\mathcal{F}$, we define a counting problem* Holant($\mathcal{F}$)*:*
Input: A signature grid $\Omega = (G, \mathcal{F}, \pi)$;
Output: Holant_Ω.

Planar Holant problems are Holant problems on planar graphs.

Definition 2.2. *Given a set of signatures $\mathcal{F}$, we define a counting problem* Pl-Holant($\mathcal{F}$)*:*
Input: A signature grid $\Omega = (G, \mathcal{F}, \pi)$, where G is a planar graph;
Output: Holant_Ω.

We would like to characterize the complexity of Holant problems in terms of its signature sets. [2] For some $\mathcal{F}$, it is possible that Holant($\mathcal{F}$) is #P-hard, while Pl-Holant($\mathcal{F}$) is tractable. These new tractable cases make dichotomies for planar Holant problems more challenging. This is also the focus of this work. Some special families of Holant problems have already been widely studied. For example, if $\mathcal{F}$ contains all EQUALITY signatures $\{=_1, =_2, =_3, \ldots\}$, then this is exactly the weighted #CSP problem. Pl-#CSP denotes the restriction of #CSP to planar structures, i.e., the standard bipartite graphs representing the input instances of #CSP are planar. In [9], we also introduced the following two special families of Holant problems by assuming some signatures are freely available.

Definition 2.3. *Let $\mathcal{U}$ denote the set of all unary signatures. Given a set of signatures $\mathcal{F}$, we use* $\text{Holant}^*(\mathcal{F})$ *(or* $\text{Pl-Holant}^*(\mathcal{F})$ *respectively) to denote* Holant($\mathcal{F} \cup \mathcal{U}$) *(or* Pl-Holant($\mathcal{F} \cup \mathcal{U}$) *respectively).*

[2]Usually our set of signatures $\mathcal{F}$ is a finite set, and the assertion of either Holant($\mathcal{F}$) is tractable or #P-hard has the usual meaning. However our dichotomy theorem is actually stronger: we allow $\mathcal{F}$ to be infinite, e.g., to include $\{=_1, =_2, =_3, \ldots\}$ or all unary signatures. Holant($\mathcal{F}$) is tractable means that it is computable in P even when we include the description of the signatures in the input Ω in the input size. Holant($\mathcal{F}$) is #P-hard means that there exists a finite subset of $\mathcal{F}$ for which the problem is #P-hard.

Definition 2.4. *Given a set of signatures $\mathcal{F}$, we use* $\text{Holant}^c(\mathcal{F})$ *(or* $\text{Pl-Holant}^c(\mathcal{F})$ *respectively) to denote* $\text{Holant}(\mathcal{F} \cup \{[1,0],[0,1]\})$ *(or* $\text{Pl-Holant}(\mathcal{F} \cup \{[1,0],[0,1]\})$ *respectively).*

Replacing a signature $f \in \mathcal{F}$ by a constant multiple cf, where $c \neq 0$, does not change the complexity of $\text{Holant}(\mathcal{F})$. So we view f and cf as the same signature. An important property of a signature is whether it is degenerate.

Definition 2.5. *A signature is degenerate iff it is a tensor product of unary signatures. In particular, a symmetric signature in $\mathcal{F}$ is degenerate iff it can be expressed as $\lambda[x,y]^{\otimes k}$.*

2.2 $\mathcal{F}$-Gate and Matchgate

A signature from $\mathcal{F}$ is a basic function which can be used at a vertex in an input graph. Instead of a single vertex, we can use graph fragments to generalize this notion. An $\mathcal{F}$-gate Γ is a tuple $(H, \mathcal{F}, \pi)$, where $H = (V, E, D)$ is a graph where the edge set consists of regular edges E and dangling edges D. Some nodes of degree 1 are designated as external nodes, and a dangling edge connects an internal node to an external node, while a regular edge connects two internal nodes. The labeling π assigns a function from $\mathcal{F}$ to each internal node. The dangling edges define variables for the $\mathcal{F}$-gate. (See Figure 1 for one example.) We denote the regular edges in E by

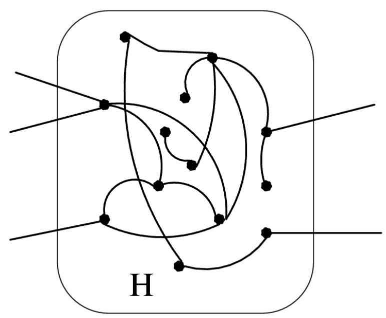

Figure 1: An $\mathcal{F}$-gate with 5 dangling edges.

$1, 2, \ldots, m$, and denote the dangling edges in D by $m+1, m+2, \ldots, m+n$. Then we can define a function for this $\mathcal{F}$-gate $\Gamma = (H, \mathcal{F}, \pi)$,

$$\Gamma(y_1, y_2, \ldots, y_n) = \sum_{x_1, x_2, \ldots x_m} H(x_1, x_2, \ldots, x_m, y_1, y_2, \ldots y_n),$$

where $(y_1, y_2, \ldots, y_n) \in \{0,1\}^n$ denotes an assignment on the dangling edges and $H(x_1, x_2, \ldots, x_m, y_1, y_2, \ldots, y_n)$ denotes the value of the signature grid on an assignment of all edges. We will also call this function the signature of the $\mathcal{F}$-gate Γ. An $\mathcal{F}$-gate can be used in a signature grid as if it is just a single node with the particular signature.

Using the idea of $\mathcal{F}$-gates, we can reduce one Holant problem to another. Let g be the signature of some $\mathcal{F}$-gate Γ. Then $\text{Holant}(\mathcal{F} \cup \{g\}) \leq_T \text{Holant}(\mathcal{F})$. The reduction is quite simple. Given an instance of $\text{Holant}(\mathcal{F} \cup \{g\})$, by replacing every appearance of g by an $\mathcal{F}$-gate Γ, we get an instance of $\text{Holant}(\mathcal{F})$. Since the signature of Γ is g, the values for these two signature grids are identical.

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. Matchgate signatures are an example. Matchgate is introduced by Valiant [20, 19, 22], whose definition is combinatorial in nature. Matchgates can be viewed as a special case of planar $\mathcal{F}$-gates, where $\mathcal{F}$ contains Exact-One functions of all arities and weight functions ($[1,0,w]$, $w \in \mathbb{C}$) on edges. The signature function Γ defined above for a matchgate is called a matchgate signature, or a standard signature. A signature function is realizable by a matchgate if it is the standard signature of that matchgate. (After a holographic transformation, a signature function is realizable under a basis if it is the transformed signature of a matchgate; see below.)

2.3 Holographic Reduction

To introduce the idea of holographic reductions, it is convenient to consider bipartite graphs. This is without loss of generality. For any general graph, we can make it bipartite by replacing each edge by a path of length two, and giving each new vertex the EQUALITY function $=_2$ on 2 inputs. (This is just the incident graph.)

We use Holant$(\mathcal{G}|\mathcal{R})$ to denote all counting problems, expressed as Holant problems on bipartite graphs $H = (U, V, E)$, where each signature for a vertex in U or V is from $\mathcal{G}$ or $\mathcal{R}$, respectively. An input instance for the bipartite Holant problem is a bipartite signature grid and is denoted as $\Omega = (H, \mathcal{G}|\mathcal{R}, \pi)$. Signatures in $\mathcal{G}$ are denoted by column vectors (or contravariant tensors); signatures in $\mathcal{R}$ are denoted by row vectors (or covariant tensors) [10].

One can perform (contravariant and covariant) tensor transformations on the signatures. We will define a simple version of holographic reductions, which are invertible. They are called holographic because they may produce exponential cancellations in the tensor space. Suppose Holant$(\mathcal{G}|\mathcal{R})$ and Holant$(\mathcal{G}'|\mathcal{R}')$ are two Holant problems defined for the same family of graphs, and $T \in \mathbf{GL}_2(\mathbb{C})$. We say that there is an (invertible) holographic reduction from Holant$(\mathcal{G}|\mathcal{R})$ to Holant$(\mathcal{G}'|\mathcal{R}')$, and T is the basis transformation, if the *contravariant* transformation $G' = T^{\otimes g}G$ and the *covariant* transformation $R = R'T^{\otimes r}$ map $G \in \mathcal{G}$ to $G' \in \mathcal{G}'$ and $R \in \mathcal{R}$ to $R' \in \mathcal{R}'$, and vice versa, where G and R have arity g and r respectively. (Notice the reversal of directions when the transformation $T^{\otimes n}$ is applied. This is the meaning of *contravariance* and *covariance*.)

Theorem 2.6 (Valiant's Holant Theorem [22])**.** *Suppose there is a holographic reduction from $\#\mathcal{G}|\mathcal{R}$ to $\#\mathcal{G}'|\mathcal{R}'$ mapping signature grid Ω to Ω', then* $\text{Holant}_\Omega = \text{Holant}_{\Omega'}$.

In particular, for invertible holographic reductions from Holant$(\mathcal{G}|\mathcal{R})$ to Holant$(\mathcal{G}'|\mathcal{R}')$, one problem is in P iff the other one is, and similarly one problem is #P-hard iff the other one is also.

In the study of Holant problems, we will commonly transfer between bipartite and non-bipartite settings. When this does not cause confusion, we do not distinguish signatures between column vectors (or contravariant tensors) and row vectors (or covariant tensors). Whenever we write a transformation as $T^{\otimes n}F$ or $T\mathcal{F}$, we view the signature or signatures as column vectors (or contravariant tensors); whenever we write a transformation as $FT^{\otimes n}$ or $\mathcal{F}T$, we view the signature or signatures as row vectors (or covariant tensors).

2.4 Some Known Dichotomy Results

In this subsection, we state some known dichotomy theorems. We first review three dichotomy theorems from [9].

Theorem 2.7. *Let $\mathcal{F}$ be a set of symmetric signatures over $\mathbb{C}$. Then* Holant$^*(\mathcal{F})$ *is computable in polynomial time in the following three cases. In all other cases,* Holant$^*(\mathcal{F})$ *is #P-hard.*

1. *Every signature in $\mathcal{F}$ is of arity no more than two;*

2. *There exist two constants a and b (not both zero, depending only on $\mathcal{F}$), such that for every signature $[x_0, x_1, \ldots, x_n] \in \mathcal{F}$ one of the two conditions is satisfied: (1) for every $k = 0, 1, \ldots, n-2$, we have $ax_k + bx_{k+1} - ax_{k+2} = 0$; (2) $n = 2$ and the signature $[x_0, x_1, x_2]$ is of form $[2a\lambda, b\lambda, -2a\lambda]$.*

3. *For every signature $[x_0, x_1, \ldots, x_n] \in \mathcal{F}$, one of the two conditions is satisfied: (1) For every $k = 0, 1, \ldots, n-2$, we have $x_k + x_{k+2} = 0$; (2) $n = 2$ and the signature $[x_0, x_1, x_2]$ is of form $[\lambda, 0, \lambda]$.*

The same dichotomy also holds for Pl-Holant$^*(\mathcal{F})$.

Theorem 2.8. *Let $\mathcal{F}$ be a set of* real *symmetric signatures, and let $\mathcal{F}_1, \mathcal{F}_2$ and $\mathcal{F}_3$ be three families of signatures defined as*

$$\begin{aligned}
\mathcal{F}_1 &= \{\lambda([1,0]^{\otimes k} + i^r[0,1]^{\otimes k}) | \lambda \in \mathbb{C}, k = 1, 2, \ldots, r = 0, 1, 2, 3\}; \\
\mathcal{F}_2 &= \{\lambda([1,1]^{\otimes k} + i^r[1,-1]^{\otimes k}) | \lambda \in \mathbb{C}, k = 1, 2, \ldots, r = 0, 1, 2, 3\}; \\
\mathcal{F}_3 &= \{\lambda([1,i]^{\otimes k} + i^r[1,-i]^{\otimes k}) | \lambda \in \mathbb{C}, k = 1, 2, \ldots, r = 0, 1, 2, 3\}.
\end{aligned}$$

Then Holant$^c(\mathcal{F})$ *is computable in polynomial time if (1) After removing unary signatures from $\mathcal{F}$, it falls in one of the three Classes of Theorem 2.7 (this implies* Holant$^*(\mathcal{F})$ *is computable in polynomial time) or (2) (Without removing any unary signature) $\mathcal{F} \subseteq \mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3$. Otherwise,* Holant$^c(\mathcal{F})$ *is #P-hard.*

Here we explicitly list all the real signatures in $\mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3$:

1. ($\mathcal{F}_1$): $[1, 0, 0, \ldots, 0, \pm 1]$,
2. ($\mathcal{F}_2$): $[1, 0, 1, 0, \ldots, 0/1]$,
3. ($\mathcal{F}_2$): $[0, 1, 0, 1, \ldots, 0/1]$,
4. ($\mathcal{F}_3$): $[1, 0, -1, 0, 1, 0, -1, 0, \ldots, 0/1/(-1)]$,
5. ($\mathcal{F}_3$): $[0, 1, 0, -1, 0, 1, 0, -1, \ldots, 0/1/(-1)]$,
6. ($\mathcal{F}_3$): $[1, 1, -1, -1, 1, 1, -1, -1, \ldots, 1/(-1)]$,
7. ($\mathcal{F}_3$): $[1, -1, -1, 1, 1, -1, -1, 1, \ldots, 1/(-1)]$.

Definition 2.9. *A k-ary function $f(x_1, \ldots, x_k)$ is affine if it has the form*

$$\chi_{[AX=0]} i^{\sum_{j=1}^{n} \langle \alpha_j, X \rangle}$$

where $X = (x_1, x_2, \ldots, x_k, 1)$, and χ is a 0-1 indicator function such that $\chi_{[AX=0]}$ is 1 iff $AX = 0$. Note that the inner product $\langle \alpha, X \rangle$ is calculated over $\mathbb{F}_2$, while the summation over j on the exponent of $i = \sqrt{-1}$ is over $\mathbb{F}_4$. We use $\mathcal{A}$ to denote the set of all affine functions.

We use $\mathcal{P}$ to denote the set of functions which can be expressed as a product of unary functions, binary equality functions ($[1, 0, 1]$ on some two variables) and binary disequality functions ($[0, 1, 0]$ on some two variables).

Theorem 2.10. *Suppose $\mathcal{F}$ is a set of functions mapping Boolean inputs to complex numbers. If $\mathcal{F} \subseteq \mathcal{A}$ or $\mathcal{F} \subseteq \mathcal{P}$, then #CSP($\mathcal{F}$) is computable in polynomial time. Otherwise, #CSP($\mathcal{F}$) is #P-hard.*

As we mentioned in [9], the class $\mathcal{A}$ is a natural generalization of the symmetric signatures family $\mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3$. It is easy to show that the set of symmetric signatures in $\mathcal{A}$ is exactly $\mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3$.

The following dichotomy for 2-3 regular graphs is from [15].

Theorem 2.11. ([15]) *The problem Holant($[y_0, y_1, y_2]|[1, 0, 0, 1]$) is #P-hard for all $y_0, y_1, y_2 \in \mathbb{C}$ except in the following cases, for which the problem is in P: (1) $y_1^2 = y_0 y_2$; (2) $y_0^{12} = y_1^{12}$ and $y_0 y_2 = -y_1^2$ ($y_1 \neq 0$) ; (3) $y_1 = 0$; (4) $y_0 = y_2 = 0$. If we restrict the input to planar graphs, then these four categories are tractable in P, as well as a fifth category $y_0^3 = y_2^3$, and the problem remains #P-hard in all other cases.*

2.5 Characterization of Realizable Signatures by Matchgates

A matchgate is called even (respectively odd) if it has an even (respectively odd) number of vertices. The following two lemmas are from [3].

Lemma 2.12. *A symmetric signature $[z_0, \ldots, z_m]$ is the standard signature of some even matchgate iff for all odd i, $z_i = 0$, and there exist r_1 and r_2 not both zero, such that for every even $2 \leq k \leq m$,*

$$r_1 z_{k-2} = r_2 z_k.$$

Lemma 2.13. *A symmetric signature $[z_0, \ldots, z_m]$ is the standard signature of some odd matchgate iff for all even i, $z_i = 0$, and there exist r_1 and r_2 not both zero, such that for every odd $3 \leq k \leq m$,*

$$r_1 z_{k-2} = r_2 z_k.$$

In [6], we characterized all symmetric signatures realizable by matchgates under a given basis. Here we state the theorem for a particular basis $\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$, which will be used in Theorem 5.1.

Theorem 2.14. *A symmetric signature $[x_0, x_1, \ldots, x_n]$ is realizable under the basis* $\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$ iff *it takes one of the following forms:*

- *Form 1: there exist constants λ, s, t and ϵ where $\epsilon = \pm 1$, such that for all $i, 0 \le i \le n$,*

$$x_i = \lambda[(s+t)^{n-i}(s-t)^i + \epsilon(s-t)^{n-i}(s+t)^i].$$

- *Form 2: there exist a constant λ, such that for all $i, 0 \le i \le n$,*

$$x_i = \lambda[(n-i)(-1)^i + i(-1)^{i-1}].$$

- *Form 3: there exist a constant λ, such that for all $i, 0 \le i \le n$,*

$$x_i = \lambda[(n-2)i].$$

3 Polynomial Interpolation

In this section, we discuss the interpolation method we will use in this paper. Polynomial interpolation is a powerful tool in the study of counting problems initiated by Valiant [18] and further developed by Vadhan, Dyer and Greenhill [17, 11] and others. The method we use here is essentially the same as Vadhan [17].

For some set of signatures $\mathcal{F}$, suppose we want to show that for all unary signatures $f = [x, y]$, we have $\text{Holant}(\mathcal{F} \cup \{[x, y]\}) \le_{\text{T}} \text{Holant}(\mathcal{F})$. Let $\Omega = (G, \mathcal{F} \cup \{[x, y]\}, \pi)$. We want to compute Holant_Ω in polynomial time using an oracle for $\text{Holant}(\mathcal{F})$.

Let V_f be the subset of vertices in G assigned f in Ω. Suppose $|V_f| = n$. We can classify all 0-1 assignments σ in the Holant sum according to how many vertices in V_f whose incident edge is assigned a 0 or a 1. Then the Holant value can be expressed as

$$\text{Holant}_\Omega = \sum_{0 \le i \le n} c_i x^i y^{n-i}, \tag{1}$$

where c_i is the sum over all edge assignments σ, of products of evaluations at all $v \in V(G) - V_f$, where σ is such that exactly i vertices in V_f have their incident edges assigned 0 (and $n - i$ have their incident edges assigned 1.) If we can evaluate these c_i, we can evaluate Holant_Ω.

Now suppose $\{G_s\}$ is a sequence of $\mathcal{F}$-gates, and each G_s has one dangling edge. Denote the signature of G_s by $f_s = [x_s, y_s]$, for $s = 0, 1, \ldots$. If we replace each occurrence of f by f_s in Ω we get a new signature grid Ω_s, which is an instance of $\text{Holant}(\mathcal{F})$, with

$$\text{Holant}_{\Omega_s} = \sum_{0 \le i \le n} c_i x_s^i y_s^{n-i}. \tag{2}$$

One can evaluate Holant_{Ω_s} by oracle access to $\text{Holant}(\mathcal{F})$. Note that the same set of values c_i occurs. We can treat c_i in (2) as a set of unknowns in a linear system. The idea of interpolation is to find a suitable sequence $\{f_s\}$ such that the evaluation of Holant_{Ω_s} gives a linear system (2) of full rank, from which we can solve all c_i.

In this paper, the sequence $\{G_s\}$ will be constructed recursively using suitable gadgetry. There are two gadgets in a recursive construction: one gadget has arity 1, giving the initial signature $g = [x_0, y_0]$; the other has arity 2, giving the recursive iteration. It is more convenient to use a 2×2 matrix A to denote it. So we can recursively connect them as in Figure 2 and get $\{G_s\}$.

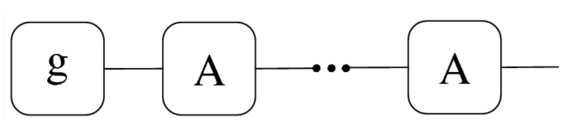

Figure 2: Recursive construction.

The signatures of $\{G_s\}$ have the following relation,

$$\begin{bmatrix} x_s \\ y_s \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_{s-1} \\ y_{s-1} \end{bmatrix}, \tag{3}$$

where $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$ and $g = \begin{bmatrix} x_0 \\ y_0 \end{bmatrix}$.

We call this gadget pair (A, g) a recursive construction. It follows from Lemma 6.1 in [17] that

Lemma 3.1. *Let α, β be the two eigenvalues of A. If the following three conditions are satisfied*

1. $\det(A) \neq 0$*;*
2. *g is not a column eigenvector of A (nor the zero vector);*
3. *α/β is not a root of unity;*

then the recursive construction (A, g) can be used to interpolate all unary signatures.

A similar interpolation method also works for signatures with larger arity but have two degrees of freedom. For example, all signatures of form $[0, x, 0, y]$. This is used in the proof of Lemma 4.9.

4 Dichotomy for Planar Holant^c Problems

Before presenting the main dichotomy theorem for planar Holant^c problems, we prove the following theorem, which plays a crucial role in the proof of the main theorem.

Theorem 4.1. *Let $a, b \in \mathbb{R}$.*

- *If $ab \neq 1$ then Pl-Holant$^c([a, 0, 1, 0, b])$ is #P-hard.*
- *If $ab = 1$ then Pl-Holant$^c([a, 0, 1, 0, b])$ is solvable in P.*

We first prove three lemmas which will be used in the proof of this theorem.

Lemma 4.2. *Let $a, b, x \in \mathbb{R}$, $ab \neq 0$ and $x \neq \pm 1$. Then Pl-Holant$^c(\{[a, 0, 0, 0, b], [0, 1, 0, x]\})$ is #P-hard.*

Proof. Firstly, we show how to realize $(=_6) = [1, 0, 0, 0, 0, 0, 1]$ by $[a, 0, 0, 0, b]$. $[a, 0, 0, 0, b]$ can be attached to a vertex of degree 4. We can connect 3 pairs of edges of two copies of $[a, 0, 0, 0, b]$ to realize the binary function $[a^2, 0, b^2]$.

If $a^2 = b^2$, then we connect one pair of edges from two copies of $[a, 0, 0, 0, b]$ to get $[a^2, 0, 0, 0, 0, 0, b^2]$. This is the same as $(=_6) = [1, 0, 0, 0, 0, 0, 1]$ after factoring out the non-zero factor $a^2 = b^2$.

If $a^2 \neq b^2$, then we connect $[a, 0, 0, 0, b]$ with a chain of $[a^2, 0, b^2]$ of length i to get $[a^{2i+1}, 0, 0, 0, b^{2i+1}]$. Because for any $i \neq j$, $a^{2i+1}/b^{2i+1} \neq a^{2j+1}/b^{2j+1}$, we can realize $(=_4) = [1, 0, 0, 0, 1]$ using polynomial interpolation, as follows. Consider any signature grid on a planar graph G with n occurrences of $=_4$ together with some other signatures. Let $x_{k,\ell}$ be the sum, over all 0-1 edge assignments σ, of the products of all other vertex function values in G except at n vertices with $=_4$, where $k, \ell \geq 0$ and $k + \ell = n$, *and* in σ exactly k occurrences of $=_4$ have input 0, and exactly ℓ occurrences of $=_4$ have input 1. The Holant value is $\sum_{k+\ell=n} x_{k,\ell}$. Now substitute each occurrence of $=_4$ by $[a^{2i+1}, 0, 0, 0, b^{2i+1}]$. The new signature grid has Holant value $\sum_{k+\ell=n} x_{k,\ell}(a^k b^\ell)^{2i+1}$. This gives a Vandermonde system from which we solve for $x_{k,\ell}$. Now we have $=_4$. Then we connect two copies of $=_4$ on one pair of edges to get $=_6$.

Take a vertex of degree 6 in a planar graph attached with $=_6$, where the 6 incident edges are its variables. We will bundle two adjacent variables to form 3 bundles of 2 edges each. Then if the inputs are restricted to $\{(0, 0), (1, 1)\}$ on each bundle, then the function takes value 1 on $((0, 0), (0, 0), (0, 0))$ and $((1, 1), (1, 1), (1, 1))$, and takes value 0 elsewhere. Thus if we restrict the domain to $\{(0, 0), (1, 1)\}$, it is the ternary Equality function $=_3$.

Let $F = [0, 1, 0, x]$ and let $H(x_1, x_2, y_1, y_2) = \sum_{z=0,1} F(x_1, y_1, z)F(x_2, y_2, z)$. This H is realizable by connecting one pair of edges of two copies of F. (See Figure 3.) We will consider H as a function in (x_1, x_2) and (y_1, y_2). However we will only connect H externally by connecting (x_1, x_2) and (y_1, y_2) to some bundle of two adjacent edges of some $=_6$. Since $=_6$ enforces the values on the bundle to be either $(0, 0)$ or $(1, 1)$, we will only be interested in the restriction of H to the domain $\{(0, 0), (1, 1)\}$. On this domain, H is a *symmetric* function of arity 2, and can be denoted as $[1, 1, x^2]$. (Note that H is *not* a symmetric function of arity 4 on $\{0, 1\}$, as $H(0, 1, 0, 1) = x$.)

Now we have reduced Pl-Holant$^c(\{[1, 0, 0, 1], [1, 1, x^2]\})$ to Pl-Holant$^c(\{[a, 0, 0, 0, b], [0, 1, 0, x]\})$.

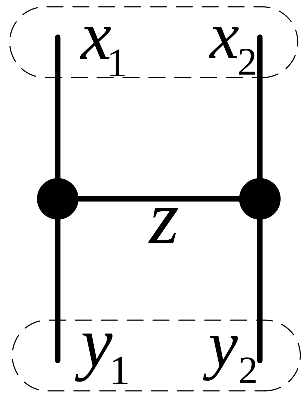

Figure 3: The gadget for function H.

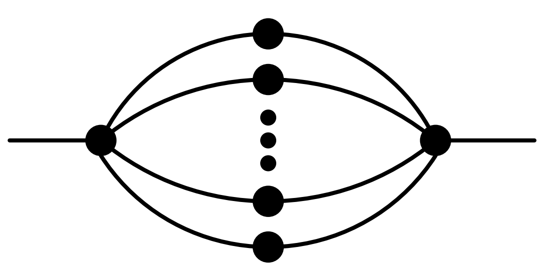

Figure 4: The gadget for function H.

Using $(=_3) = [1, 0, 0, 1]$, we can realize the EQUALITY function $=_k$ of any arity $k \geq 3$. Then we can realize $[1, 1, x^{2k}]$, for all $k \geq 1$. (See Figure 4.) If $x = 0$, then we already have $[1, 1, 0]$. Suppose $x \neq 0$. Because $x^2 \neq 1$ and being a positive real number, we can realize $[1, 1, 0]$ by interpolation. Now we have reduced the problem Pl-Holant$([1, 0, 0, 1] \mid [1, 1, 0])$ to Pl-Holant$^c(\{[1, 0, 0, 1], [1, 1, x^2]\})$. The bipartite problem Pl-Holant$([1, 0, 0, 1] \mid [1, 1, 0])$ is #P-hard since it is counting VERTEX COVERS on planar 3-regular graphs [23]. □

The following lemma handles a special case of Theorem 4.1. The proof uses Lemma 4.2.

Lemma 4.3. *Pl-Holant*$^c([0, 0, 1, 0, 0])$ *is #P-hard.*

Proof. We construct a reduction from Pl-Holant$^c([1, 0, 0, 0, 1], [0, 1, 0, 0])$, which is #P-hard by Lemma 4.2, to Pl-Holant$^c([0, 0, 1, 0, 0])$ by polynomial interpolation.

Let $F = [0, 0, 1, 0, 0]$. There is a series of planar gadgets (a chain of F) realizing the following sequence of functions:

$$H_2(x_1, x_2, y_1, y_2) = \sum_{x_3, x_4 = 0,1} F(x_1, x_2, x_3, x_4) F(y_1, y_2, x_3, x_4),$$

and for $i \geq 1$,

$$H_{2i+2}(x_1, x_2, y_1, y_2) = \sum_{x_3, x_4 = 0,1} H_{2i}(x_1, x_2, x_3, x_4) H_2(y_1, y_2, x_3, x_4).$$

The gadget for H_{2i} is composed of $2i$ functions F. As an example, the gadget for H_4 is shown in Figure 5.

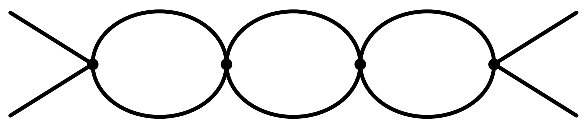

Figure 5: The gadget for H_4.

By calculation, $H_{2i}(0, 0, 0, 0) = H_{2i}(1, 1, 1, 1) = 1$, and $H_{2i}(0, 1, 0, 1) = H_{2i}(0, 1, 1, 0) = H_{2i}(1, 0, 0, 1) = H_{2i}(1, 0, 1, 0) = 2^{2i-1}$, and H_{2i} is zero on other inputs. Again we will consider the inputs to H_{2i} as bundled into (x_1, x_2) and (y_1, y_2).

Given a planar graph G as an instance of Pl-Holant$^c([1, 0, 0, 0, 1], [0, 1, 0, 0])$, suppose there are n vertices in G attached with the function $(=_4) = [1, 0, 0, 0, 1]$. For $i = 1, 2, \ldots, n + 1$, we construct an instance G_i of Pl-Holant$^c([0, 0, 1, 0, 0])$ as follows: Replace each occurrence of $=_4$ by a copy of H_{2i}, and replace each occurrence of $[0, 1, 0, 0]$ by $[0, 0, 1, 0, 0]$ connected with a $[0, 1]$, which exactly realizes $[0, 1, 0, 0]$. Note that by replacing $=_4$ with H_{2i}, we have bundled two adjacent edges together (in the planar embedding) for each vertex attached with $=_4$.

Let $x_{a,b}$ denote the summation, over all 0-1 edge assignments σ, of the products of all other vertex function values in G except at those n vertices with $=_4$, where $a, b \geq 0$ and $a + b = n$, *and* in σ exactly a occurrences of $=_4$ have inputs $\{0000, 1111\}$, and exactly b occurrences of $=_4$ have inputs $\{0101, 0110, 1001, 1010\}$.

Note that the Holant value on G_i is

$$\sum_{a+b=n} x_{ab} 1^a (2^{2i-1})^b.$$

On the other hand, the value of Pl-Holant$^c([1, 0, 0, 0, 1], [0, 1, 0, 0])$ on G is exactly $x_{n,0}$.

When we take $1 \leq i \leq n + 1$, we get a system of linear equations in x_{ab}, whose coefficient matrix is a full ranked Vandermonde matrix. Solving this Vandermonde system we obtain the value $x_{n,0}$.

□

The following result can be proved by interpolation as well.

Lemma 4.4. *Let* $a \notin \{-1, 0, 1\}$ *be a real number. Then we can interpolate all* $[x, 0, y, 0]$ *and* $[0, y, 0, x]$ *for* $x, y \in \mathbb{C}$ *starting from either* $[0, 1, 0, a]$ *or* $[a, 0, 1, 0]$.

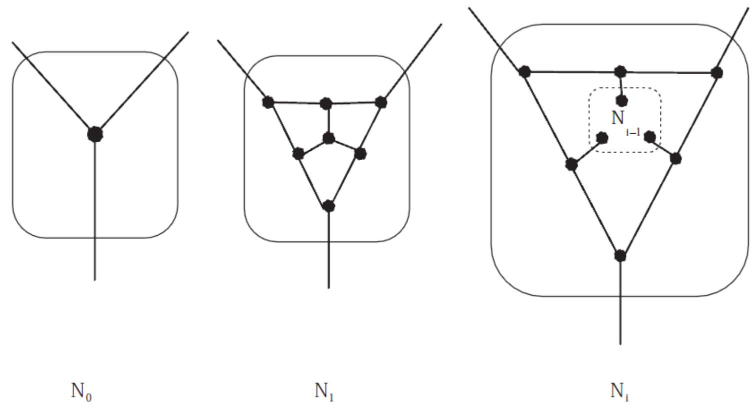

Figure 6: The recursive construction. The signature of every vertex in the gadget is $[0,1,0,a]$.

Proof. The recursive construction is depicted by Figure 6. By a simple parity argument, every $\mathcal{F}$-gate N_i has a signature of the form $[0,x_i,0,y_i]$. After some calculation, we see that they satisfy the following recursive relation:

$$\begin{bmatrix} x_{i+1} \\ y_{i+1} \end{bmatrix} = \begin{bmatrix} 3(a^2+1) & a^3+a \\ 3(a^3+a) & a^6+1 \end{bmatrix} \begin{bmatrix} x_i \\ y_i \end{bmatrix}.$$

The signatures we want to interpolate are of arity 3. But since all of them take the form $[0,x_i,0,y_i]$ with two degrees of freedom, we can use the interpolation method in Section 3. Now we verify that the conditions of that theorem are satisfied. Let $A = \begin{bmatrix} 3(a^2+1) & (a^3+a) \\ 3(a^3+a) & a^6+1 \end{bmatrix}$, then $(A,[1,a]^{\mathtt{T}})$ forms a recursive construction. Since $\det(A) = 3(a^4-1)^2 \neq 0$, the first condition holds. Its characteristic equation is $X^2-(a^6+3a^2+4)X+3(a^4-1)^2 = 0$. For this quadratic equation, the discriminant $\Delta = (a^6-3a^2-2)^2+12(a+a^3)^2 > 0$. So A has two distinct real eigenvalues. The sum of the two eigenvalues is $\mathrm{tr}A = a^6+3a^2+4 > 0$. So they are not opposite to each other. Therefore, the ratio of these two eigenvalues is not a root of unity and the third condition holds. Consider the second condition: if the initial vector $[1,a]^{\mathtt{T}}$ is a column eigenvector of A, then we have $A\begin{bmatrix} 1 \\ a \end{bmatrix} = \lambda \begin{bmatrix} 1 \\ a \end{bmatrix}$, where λ is an eigenvalue of A. From this, we will conclude that $a(a^2-1)(a^4-1) = 0$, which can not happen given $a \notin \{-1,0,1\}$. To sum up, this recursive relation satisfies all three conditions of Lemma 3.1 and can be used to interpolate all signatures of the form $[0,x,0,y]$. This completes the proof. □

Proof of Theorem 4.1 If $ab = 1$, then $[a,0,1,0,b]$ is realizable by some matchgate, by Lemma 2.12. This realizability also applies to the unary functions $[1,0]$ and $[0,1]$. Hence the problem Pl-Holant$^c([a,0,1,0,b])$ can be solved in polynomial time by matchgate computation via the FKT method [16, 12, 13]. In the following we assume that $ab \neq 1$ and prove that the problem is #P-hard. The case $a = b = 0$ is proved in Lemma 4.3. Now we can assume at least one of a and b is non-zero, and by symmetry we assume $a \neq 0$.

We know from our dichotomy for Holant^c problems [9] that Holant$^c([a,0,1,0,b])$ for general graphs is #P-hard unless $a = b = 1$ or $a = b = -1$, in which cases it is tractable. Both of these tractable cases are also included in the tractable cases ($ab = 1$) here. Therefore, if we can realize a *cross function* X with a planar gadget when $ab \neq 1$, we can reduce Holant$^c([a,0,1,0,b])$ for general graphs to Pl-Holant$^c([a,0,1,0,b])$ and finish the proof. Here a cross function X has 4 input bits, and satisfies $X_{0000} = X_{0101} = X_{1010} = X_{1111} = 1$ and $X_\alpha = 0$ for all other inputs $\alpha \in \{0,1\}^4$.

If $\{a,b\} \not\subset \{-1,0,1\}$, we can use Lemma 4.4 to interpolate all $[x,0,y,0]$, for $x,y \in \mathbb{C}$. If $\{a,b\} \subset \{-1,0,1\}$, then there are only four cases: $[1,0,1,0,-1]$, $[1,0,1,0,0]$, $[-1,0,1,0,1]$ and $[-1,0,1,0,0]$. In all four cases, it is easy to verify that we can realize a signature with a form $[c_1,0,c_2,0]$ where $c_1c_2 \neq 0$ and $c_1 \neq \pm c_2$ using the gadget in Figure 7. After factoring out a nonzero factor, we have $[c',0,1,0]$, where $c' \in \mathbb{R}$ and $c' \notin \{0,\pm 1\}$. As a result, we can also interpolate all $[x,0,y,0]$, where $x,y \in \mathbb{C}$.

Now we can use all signatures of the form $[x,0,y,0]$, for arbitrary $x,y \in \mathbb{C}$, to build new gadgets. We also have all $[x,0,y]$ by connecting $[x,0,y,0]$ to a $[1,0]$. By connecting a $[\sqrt[4]{t/a},0,\sqrt[4]{a/t}]$ to each edge of the signature

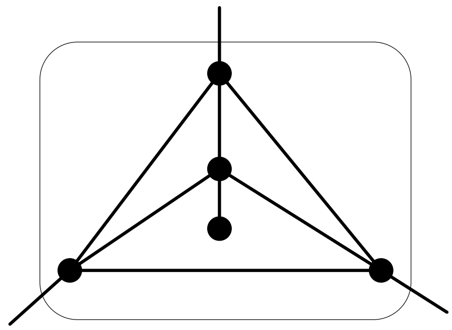

Figure 7: The signature of the degree 1 vertex in the gadget is $[1,0]$.

$[a,0,1,0,b]$, we get $[t,0,1,0,\frac{c}{t}]$ for all $t \neq 0$, where $c = ab \neq 1$. Using all these, we will build a planar gadget in Figure 8 to realize the cross function X. In the equations below x, y, t are three variables we can set to any complex numbers, with $t \neq 0$. The parameter c is given and not equal to 1.

(Of course we presumably could not build a cross function X if $c = 1$; this is *exactly* when the problem is in P, and this is also *exactly* when our construction of X fails. If a cross function X were to exist when $c = 1$ then P = #P would follow. However, it is still rather mysterious that algebraically $c = 1$ is *exactly* when our construction fails. This failure condition is by no means obvious from the equations below.)

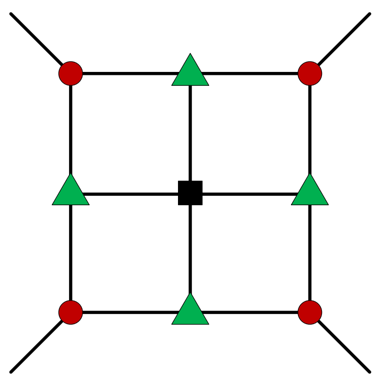

Figure 8: This gadget is to realize the Cross function. The signature for the center vertex (black and square) is $[t,0,1,0,\frac{c}{t}]$. The signature for the vertexes in the four corners (red and circle) is $[x,0,1,0]$. The signature for the vertexes in the middle of the boundaries (green and triangle) is $[y,0,1,0]$.

We can compute the signature of the gadget in Fig. 8. If the input has an odd number of 1s, the value is 0. For other inputs, we have

$$\begin{aligned}
X_{0000} &= x^4y^4t + t + 4x^3y^2 + 4x + 4x^2y + \frac{2cx^2}{t} \\
X_{1111} &= 2y^2t + 12y + \frac{2c}{t} \\
X_{0101} = X_{1010} &= 2xy^2t + 4x^2y^2 + 4 + 4xy + \frac{2cx}{t} \\
X_{0011} = X_{1001} = X_{1100} = X_{0110} &= x^2y^3t + yt + 3x^2y^2 + 3 + 6xy + \frac{2cx}{t}.
\end{aligned}$$

Here we prove that for any $c \neq 1$, we can assign suitable complex values to x, y and t, where $t \neq 0$, such that

$A = B = C \neq 0$, and $D = 0$, where A, B, C and D denote respectively the four functions of x, y and t listed in the four lines above.

Claim 1. *For any $c \neq 1$,*

$$(x-1)^2 = \frac{16}{c-1}$$

has a solution $x \notin \{0, +1, -1\}$. This x satisfies

$$\left(2 - \frac{x(x+3)}{x-1}\right)\left(\frac{x+3}{x-1}\right) + cx + 6 = 0. \tag{4}$$

Proof. Clearly $x = 1$ is not a solution to $(x-1)^2 = \frac{16}{c-1}$. Also the equation has two distinct roots. When $c = 17$ there is a solution $x = 2 \notin \{0, +1, -1\}$. When $c \neq 17$, we can verify $x = 0$ is not a solution. Hence the equation always has a solution other than $0, \pm 1$.

To verify (4) we have

$$\begin{aligned}
& (2x - 2 - x^2 - 3x)(x+3) + (cx+6)(x^2 - 2x + 1) \\
= \; & -(x^3 + 4x^2 + 5x + 6) + cx^3 + (6 - 2c)x^2 + (-12 + c)x + 6 \\
= \; & (c-1)x^3 - 2(c-1)x^2 + (c-17)x \\
= \; & (c-1)x[(x-1)^2 - 16/(c-1)] \\
= \; & 0.
\end{aligned}$$

□

Now we fix $x \notin \{0, +1, -1\}$ satisfying (4) for any given $c \neq 1$.

Claim 2. *For any $c \neq 1$, we can pick $z \neq \pm 1$ such that*

$$\frac{4z}{(1+z)^2} = \frac{x(x+3)}{x-1}. \tag{5}$$

Proof. We are given $x \neq 0, \pm 1$. If $x = -3$, we can pick $z = 0$. Now suppose $x \neq -3$. Consider the quadratic equation in z

$$4z(x-1) = x(x+3)(1+z)^2.$$

This is quadratic since $x(x+3) \neq 0$. We can check that $z = +1$ (and -1 respectively) is not a solution, as this would force $x = -1$ (and $+1$ respectively). However, any solution where $z \neq -1$ and $x \neq 1$ is equivalent to (5). Hence we have a solution $z \neq \pm 1$ to (5).

□

Now we further fix a $z \neq \pm 1$ satisfying (5), and let $y = z/x$ such that $xy \neq \pm 1$, for any $c \neq 1$.

Claim 3. *For any $c \neq 1$, there exist $x \notin \{0, +1, -1\}$ and y such that $xy \neq \pm 1$ satisfying*

$$\frac{2(1 + x^2y^2)}{(1+xy)^2} \cdot \frac{x+3}{x-1} + cx + 6 = 0. \tag{6}$$

Proof.

$$\begin{aligned}
& \frac{2(1 + x^2y^2)}{(1+xy)^2} \cdot \frac{x+3}{x-1} + cx + 6 \\
= \; & 2\left(1 - \frac{2z}{(1+z)^2}\right) \cdot \frac{x+3}{x-1} + cx + 6 \\
= \; & \left(2 - \frac{x(x+3)}{x-1}\right) \cdot \frac{x+3}{x-1} + cx + 6 \\
= \; & 0.
\end{aligned}$$

Here we used (5) and (4).

□

Now we will set $t = 4/(1+xy)^2$. Clearly $t \neq 0$. We next verify that $D = 0$.

By (5) and (6) we get

$$\frac{8y(1+x^2y^2)}{(1+xy)^4} + cx + 6 = 0.$$

Then

$$t^2y(1+x^2y^2) + 2cx + 3t(1+xy)^2 = 0.$$

Thus

$$D = yt(1+x^2y^2) + 3(1+xy)^2 + \frac{2cx}{t} = 0.$$

Next we show that $C = \frac{4(1-xy)^2}{1-x} \neq 0$.

By $D = 0$, we have

$$C = 2xy^2\frac{4}{(1+xy)^2} + 4(1+xy)^2 - 4xy + [-yt(1+x^2y^2) - 3(1+xy)^2].$$

Hence

$$\begin{aligned} C &= \frac{8xy^2}{(1+xy)^2} + (1+xy)^2 - 4xy - y\frac{4(1+x^2y^2)}{(1+xy)^2} \\ &= \frac{4y}{(1+xy)^2}\left[2xy - 1 - x^2y^2\right] + (1-xy)^2 \\ &= \left(\frac{-4y}{(1+xy)^2} + 1\right)(1-xy)^2 \\ &= \frac{4(1-xy)^2}{1-x} \neq 0, \end{aligned}$$

using (5).

The next task is to show $B = C$.

We have

$$C = 4(1-xy)^2 + xB.$$

Hence

$$B = \frac{1}{x}\left[\frac{4(1-xy)^2}{1-x} - 4(1-xy)^2\right] = \frac{4(1-xy)^2}{x}\left[\frac{1}{1-x} - 1\right] = \frac{4(1-xy)^2}{1-x} = C.$$

Finally we verify $A = C$ as well.

$$A = (x^4y^4+1)t + x[C - 2xy^2t] = C + (x-1)C - 2x^2y^2t + (x^4y^4+1)t = C - 4(1-xy)^2 + t(x^2y^2-1)^2 = C.$$

□

Now we come to the main dichotomy theorem for Pl-Holant^c problems.

Theorem 4.5. *Let $\mathcal{F}$ be a set of real symmetric signatures.* Pl-Holant$^c(\mathcal{F})$ *is #P-hard unless $\mathcal{F}$ satisfies one of the following conditions, in which case it is tractable:*

1. Holant$^c(\mathcal{F})$ *is tractable (for which we have an effective dichotomy [9]); or*
2. *Every signature in $\mathcal{F}$ is realizable by some matchgate (for which we have a complete characterization [3]).*

Before we give the proof, we do some normalization of the signature set $\mathcal{F}$. Since any degenerate signature $[x,y]^{\otimes k}$ can be replaced by the corresponding unary signature $[x,y]$ without changing the complexity of the problem, we always assume that all the signatures in $\mathcal{F}$, whose arity is greater than 1, are non-degenerate. Since $[1,0]$ and $[0,1]$ are freely available, we can construct any sub-signature of an original signatures as well as any signature realizable by some $\mathcal{F}$-gate.

The main idea of the proof is to interpolate all unary functions. If we can do that, we can reduce the problem Pl-Holant$^*(\mathcal{F})$ to Pl-Holant$^c(\mathcal{F})$ and finish the proof. We note that our dichotomy in [9] for Holant$^*(\mathcal{F})$ also holds for planar graphs. In some cases, we cannot interpolate all unary functions, then we prove the theorem separately, mainly using Lemma 4.2 and Theorem 4.1. The following lemma is for interpolation of unary functions.

Lemma 4.6. *If we can construct from $\mathcal{F}$ a gadget with signature $[a,b,c]$, where $b^2 \neq ac$, $b \neq 0$ and $a+c \neq 0$, then we can interpolate all unary functions. (Hence the conclusions of Theorem 4.5 hold.)*

Proof. we use the interpolation method as described in Section 3. We consider two recursive constructions $(\begin{bmatrix} a & b \\ b & c \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \end{bmatrix})$ and $(\begin{bmatrix} a & b \\ b & c \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix})$, and argue that at least one of them will succeed given the conditions on a,b,c. We use A to denote $\begin{bmatrix} a & b \\ b & c \end{bmatrix}$. Since $b^2 \neq ac$, A is non-degenerate, the first condition of Lemma 3.1 is satisfied for both recursive constructions. If both $[1,0]$ and $[0,1]$ are column eigenvectors of A, then $b=0$, a contradiction. So at least for one of the two recursive constructions, the second condition of Lemma 3.1 is satisfied. Since A is a real symmetric matrix, both its eigenvalues are real. If the ratio of two real numbers is a root of unity, they must be the same or opposite to each other. If the two eigenvalues are the same, we have $b=0$ and $a=c$, a contradiction. If the two eigenvalues are opposite to each other, then we have $a+c=0$, also a contradiction. Therefore, the third condition of Lemma 3.1 is also satisfied for both recursive constructions. To sum up, at least one of the two recursive constructions satisfies all the conditions of Lemma 3.1. This completes the proof. □

If we can construct from $\mathcal{F}$ a gadget with a binary symmetric signature $[a,b,c]$, which satisfies all the conditions in Lemma 4.6, then we are done. For most cases, we prove the theorem by interpolating all unary signatures. However, in some more delicate cases, we are not able to do that. For example, if all signatures from $\mathcal{F}$ have the parity condition, which includes a proper superset of matchgate signatures, then all unary signatures we can realize have form $[a,0]$ or $[0,a]$, so we can not interpolate all unary signatures. For these cases, our starting point is Theorem 4.1.

We define some families of symmetric signatures, which will be used in our proof.

$$
\begin{aligned}
\mathcal{G}_1 &= \{[a,0,0,\cdots,0,b] \mid ab \neq 0\} \\
\mathcal{G}_2 &= \{[x_0,x_1,\cdots,x_k] \mid \forall i \text{ is even}, x_i = 0 \text{ or } \forall i \text{ is odd}, x_i = 0\} \\
\mathcal{G}_3 &= \{[x_0,x_1,\cdots,x_k] \mid \forall i, x_i + x_{i+2} = 0\} \\
\mathcal{M} &= \{\ f \mid f \text{ is realizable by some matchgate }\}.
\end{aligned}
$$

We note that $\mathcal{G}_1$, $\mathcal{G}_2$ and $\mathcal{G}_3$ are supersets of $\mathcal{F}_1$, $\mathcal{F}_2$ and $\mathcal{F}_3$ respectively. Furthermore (the real part of) $\mathcal{F}_2 \subseteq \mathcal{M} \subseteq \mathcal{G}_2$. The conditions in $\mathcal{G}_2$ are called parity conditions. The following several lemmas all have the form "If $\mathcal{F} \not\subseteq \mathcal{A}$, then the conclusions of Theorem 4.5 hold." After proving each lemma, in subsequent lemmas, we only need to consider the case that $\mathcal{F} \subseteq \mathcal{A}$.

Lemma 4.7. *If $\mathcal{F} \not\subseteq \mathcal{G}_1 \cup \mathcal{G}_2 \cup \mathcal{G}_3$, then the conclusions of Theorem 4.5 hold.*

Proof. Since $\mathcal{F} \not\subseteq \mathcal{G}_1 \cup \mathcal{G}_2 \cup \mathcal{G}_3$, there exists an $f \in \mathcal{F}$ and $f \notin \mathcal{G}_1 \cup \mathcal{G}_2 \cup \mathcal{G}_3$. Since all unary signatures are in $\mathcal{G}_3$, the arity of f is greater than 1 and f is non-degenerate. There are two cases according to whether f has a zero entry or not.

(1) f has some zero entries. If there exists a sub-signature of f of the form $[0,a,b]$ or $[a,b,0]$, where $ab \neq 0$, then we are done by Lemma 4.6. Otherwise, we can conclude that there are no two successive non-zero entries. So the signature f has this form $[0^{i_0} x_1 0^{i_1} x_2 0^{i_2} \cdots x_k 0^{i_k}]$, where $k \geq 1$, $x_j \neq 0$ and for all $1 \leq j \leq k-1$, $i_j \geq 1$. If for all $1 \leq j \leq k-1$, i_j is odd, (including $k=1$), then $f \in \mathcal{G}_2$, a contradiction. Otherwise there exists a sub-signature of form $[x,0,0,\cdots,0,y]$, where $xy \neq 0$ and there are a positive even number of 0s between x and y. If this is the entire f, then $f \in \mathcal{G}_1$, a contradiction. So there is one 0 before x or after y. By symmetry, we assume there is a 0 before x, so we have a sub-signature $[0,x,0,0,\cdots,0,y]$, whose arity is even and at least 4. We label its dangling edges $1,2,\cdots,2k$. Then for every $i=1,2,\cdots,k-1$, we connect dangling edges $2i+1$ and $2i+2$ together to form a regular edge. After that, we have an $\mathcal{F}$-gate with arity 2, and its signature is $[0,x,y]$. Then we are done by Lemma 4.6.

(2) f has no zero entry. We only need to prove that we can construct a function $[a',b',c']$ satisfying the three conditions in Lemma 4.6. Suppose all sub-signatures of f with arity 2 do not satisfy all the three conditions. For each sub-signature $[a',b',c']$, either $a'+c'=0$, or $b'^2 = a'c'$. If all of them satisfy $a'+c'=0$, then $f \in \mathcal{G}_3$. A contradiction. If all of them satisfy $b'^2 = a'c'$, then f is degenerate. A contradiction. W.l.o.g., we can assume there is a sub-signature $[a,b,c,d]$ of f, such that $a+c=0$, $b+d \neq 0$, and $c^2 = bd$. We get this sub-signature $[a,b,c,d]$ by $[1,0]$ and $[0,1]$. Combining two $[a,b,c,d]$, we can get a function $[a',b',c'] = [a^2+2b^2+c^2, ab+2bc+cd, b^2+2c^2+d^2] =$

$[2(b^2+c^2), c(b+d), (b+d)^2]$. Then $b' = c(b+d) \neq 0$. $a'+c' > 0$. And $a'c' - b'^2 = (b+d)^2(2b^2+c^2) > 0$. We are done by Lemma 4.6. □

The following lemma uses Theorem 4.1 in an essential way, which in turns depends on the crossover.

Lemma 4.8. *If $\mathcal{F} \not\subseteq \mathcal{G}_1 \cup \mathcal{M} \cup \mathcal{G}_3$, then the conclusions of Theorem 4.5 hold.*

Proof. If $\mathcal{F} \not\subseteq \mathcal{G}_1 \cup \mathcal{G}_2 \cup \mathcal{G}_3$, then by Lemma 4.7, we are done. Otherwise, there exists a signature $f \in \mathcal{F} \subseteq \mathcal{G}_1 \cup \mathcal{G}_2 \cup \mathcal{G}_3$ and $f \notin \mathcal{G}_1 \cup \mathcal{M} \cup \mathcal{G}_3$. Then it must be the case that $f \in \mathcal{G}_2$. Note that every signature with arity at most 3 in $\mathcal{G}_2$ (this is called the parity condition) is also contained in $\mathcal{M}$, so f is of arity greater than 3. Let $f = [x_0, x_1, \cdots, x_n]$, for some $n \geq 4$. Suppose there exists some $i \in [2, 3, \cdots, n-2]$ such that $x_i \neq 0$. If $x_{i-2}x_{i+2} \neq x_i^2$, then we can get $[x_{i-2}, 0, x_i, 0, x_{i+2}]$ by $[1, 0]$ and $[0, 1]$ which restrict the signature to a sub-signature. Then the problem is #P-hard by Theorem 4.1 and we are done. Otherwise, we have $x_{i-2}x_{i+2} = x_i^2 \neq 0$. Then starting from $x_{i-2} \neq 0$ and if $i-2 \in [2, 3, \cdots, n-2]$, we can get $x_{i-4}x_i = x_{i-2}^2 \neq 0$. Similarly we can start with x_{i+2}. A signature satisfying the parity condition and is a geometric series on the alternate entries is realizable by a matchgate [?, 3], a contradiction.

Now we may assume $x_i = 0$ for all $i \in [2, 3, \cdots, n-2]$. Since $f \in \mathcal{G}_2 - (\mathcal{M} \cup G_1)$, we know that there are only three possible subcases: (1) n is odd, $n \geq 5$, $x_0x_{n-1} \neq 0$ and $x_1 = x_n = 0$; (2) n is odd, $n \geq 5$, $x_1x_n \neq 0$ and $x_0 = x_{n-1} = 0$; (3) $n \geq 6$ is even, $x_1x_{n-1} \neq 0$ and $x_0 = x_n = 0$. This uses the theory of matchgate realizability [?, 3]. Crucially, if n is even and $n < 6$, then $n = 4$ and the case $x_1x_{n-1} \neq 0$, $x_0 = x_n = 0$ belongs to $\mathcal{M}$. The subcases (1) and (2) are reversals of each other and (3) contains a signature in form (1) and (2). So after normalizing (and connecting pairs of edges together if $n > 5$), we will get a signature $[0, 1, 0, 0, 0, x]$ where $x \neq 0$. So we have both sub-signature $[0, 1, 0, 0]$ and $[1, 0, 0, 0, x]$. As we proved in Lemma 4.2, the problem is #P-hard and we are done. This finishes the proof. □

Lemma 4.9. *If $[0, 1, 0, x] \in \mathcal{F}$ (or $[1, 0, x, 0] \in \mathcal{F}$) where $x \in \mathbb{R}$, $x \neq \pm 1$, then the conclusions of Theorem 4.5 hold.*

Proof. If $x \neq 0$, we can use Lemma 4.4 to interpolate $[0, 1, 0, 0]$. So we assume we have $[0, 1, 0, 0]$ from $\mathcal{F}$. If $\mathcal{F} \not\subseteq \mathcal{G}_1 \cup \mathcal{M} \cup \mathcal{G}_3$, then by Lemma 4.8, we are done. If $\mathcal{F} \subseteq \mathcal{M}$, then the problem is tractable and we are done. Otherwise, there exists a signature $f \in \mathcal{F} \subseteq \mathcal{G}_1 \cup \mathcal{M} \cup \mathcal{G}_3$ and $f \notin \mathcal{M}$. That is $f \in (\mathcal{G}_1 \cup \mathcal{G}_3 - \mathcal{M})$.

If f has arity ≥ 1 and of the form $[x_0, x_1, -x_0, -x_1, x_0 \cdots] \in \mathcal{G}_3 - \mathcal{M}$, then we will have $x_0x_1 \neq 0$. Otherwise we would have $f \in \mathcal{M}$, a contradiction. Connecting one unary signature $[x_0, x_1]$ to $[0, 1, 0, 0]$, we get $[x_1, x_0, 0]$ which satisfies all the conditions in Lemma 4.6, and we are done.

Now we consider $f = [1, 0, 0, \cdots, 0, y] \in \mathcal{G}_1 - \mathcal{M}$, where $y \neq 0$. Since $f \notin \mathcal{M}$, its arity n is greater than 2. If n is odd, we can connect its edges except one to get a unary signature $[1, y]$. Then we can use a similar argument as above and we are done. If n is even, then it is at least 4, since $f \notin \mathcal{M}$. After connecting its edges except four, we can get $[1, 0, 0, 0, y]$. Together with $[0, 1, 0, 0]$, we know the problem is #P-hard by Lemma 4.2. This completes the proof. □

Lemma 4.10. *If $\mathcal{F} \not\subseteq \mathcal{G}_1 \cup \mathcal{F}_2 \cup \mathcal{G}_3$, then the conclusions of Theorem 4.5 hold.*

Proof. If $\mathcal{F} \not\subseteq \mathcal{G}_1 \cup \mathcal{G}_2 \cup \mathcal{G}_3$, then by Lemma 4.7, we are done. Otherwise, there exists a signature $f \in \mathcal{F} \subseteq \mathcal{G}_1 \cup \mathcal{G}_2 \cup \mathcal{G}_3$ and $f \notin \mathcal{G}_1 \cup \mathcal{F}_2 \cup \mathcal{G}_3$. Then it must be the case that $f \in \mathcal{G}_2$. Note that every signature with arity less than 3 in $\mathcal{G}_2$ is also contained in $\mathcal{G}_1 \cup \mathcal{G}_3$, so f is of arity greater than 2. Since $f \notin \mathcal{G}_1$, there is some non-zero in the middle of the signature f, after normalization, we can assume there is a sub-signature of form $[0, 1, 0, x]$ (or $[x, 0, 1, 0]$). If $x \neq \pm 1$, then by Lemma 4.9, we are done. Otherwise, for every such pattern, we have $x = \pm 1$. Since $f \notin \mathcal{F}_2$, then there is some sub-signature $[0, 1, 0, -1]$ and because $f \notin \mathcal{G}_3$, there is some sub-signature $[0, 1, 0, 1]$. Therefore, there is a sub-signature $[1, 0, 1, 0, -1]$ of f. Then by Theorem 4.1, we know that the problem is #P-hard and we are done. This completes the proof. □

Lemma 4.11. *If $\mathcal{F} \not\subseteq \mathcal{G}_1 \cup \mathcal{G}_3$, then the conclusions of Theorem 4.5 hold.*

Proof. If $\mathcal{F} \not\subseteq \mathcal{G}_1 \cup \mathcal{F}_2 \cup \mathcal{G}_3$, then by Lemma 4.10, we are done. Otherwise, there exists a signature $f \in \mathcal{F} \subseteq \mathcal{G}_1 \cup \mathcal{F}_2 \cup \mathcal{G}_3$ and $f \notin \mathcal{G}_1 \cup \mathcal{G}_3$. Then it must be the case that $f \in \mathcal{F}_2$. Note that every signature with arity less than 3 in $\mathcal{F}_2$ is also contained in $\mathcal{G}_1 \cup \mathcal{G}_3$, so f is of arity at least 3. Then f has a sub-signature $[1, 0, 1, 0]$ or $[0, 1, 0, 1]$. By symmetry, we assume it is $[1, 0, 1, 0]$. If $\mathcal{F} \subseteq \mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3$, then Theorem 4.5 trivially holds and there is nothing to prove. If not, there exists a signature $g \in \mathcal{F} - \mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3$. By $\mathcal{F} \subseteq \mathcal{G}_1 \cup \mathcal{F}_2 \cup \mathcal{G}_3$, either $g \in \mathcal{G}_1 - \mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3$ ($\subseteq \mathcal{G}_1 - \mathcal{F}_1$) or $g \in \mathcal{G}_3 - \mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3$ ($\subseteq \mathcal{G}_3 - \mathcal{F}_3$).

For the first case, $g \in (\mathcal{G}_1 - \mathcal{F}_1)$, after a scale, g is of form $[1, 0, 0, \cdots, b]$, where $b \notin \{-1, 0, 1\}$. If the arity of g is odd, we can realize $[1, b]$. (We connect every two adjacent dangling edges into one edge and leave one dangling edge.) Then connecting this unary signature to one dangling edge of $[1, 0, 1, 0]$, we can realize a binary signature $[1, b, 1]$. Then by Lemma 4.6, Theorem 4.5 holds. If the arity of g is even, we can realize $[1, 0, b]$ (leave two dangling edges). By connecting one of its dangling edge to one dangling edge of $[1, 0, 1, 0]$, we can have a new ternary signature $[1, 0, b, 0]$. By Lemma 4.9, we are done.

For the second case $g \in (\mathcal{G}_3 - \mathcal{F}_3)$, g has a sub-signature of form $[1, b]$, where $b \notin \{-1, 0, 1\}$. By the same argument as above, Theorem 4.5 holds. This completes the proof. □

Lemma 4.12. *If $\mathcal{F} \nsubseteq \mathcal{G}_1 \cup \mathcal{F}_3$, then the conclusions of Theorem 4.5 hold.*

Proof. If $\mathcal{F} \nsubseteq \mathcal{G}_1 \cup \mathcal{G}_3$, then by Lemma 4.11, we are done. Otherwise, there exists a signature $f \in \mathcal{F} \subseteq \mathcal{G}_1 \cup \mathcal{G}_3$ and $f \notin \mathcal{G}_1 \cup \mathcal{F}_3$. Then it must be the case that $f \in \mathcal{G}_3$, and f has a sub-signature of form $[1, a, -1]$, where $a \notin \{-1, 0, 1\}$.

If $\mathcal{F} \subseteq \{[1, 0, 1]\} \cup \mathcal{G}_3$, then $\text{Holant}^*(\mathcal{F})$ is polynomial time computable by Theorem 2.7 and as a result Theorem 4.5 trivially holds and we are done.

If not, there exists a signature $g \in \mathcal{F} \subseteq \mathcal{G}_1 \cup \mathcal{G}_3$ and $g \notin \{[1, 0, 1]\} \cup \mathcal{G}_3$. Then it must be the case that $g \in \mathcal{G}_1$. The arity of g is greater than 1, as $g \notin \mathcal{G}_3$.

If the arity of g is 2, then g is of form $[1, 0, b]$, where $b \notin \{-1, 0, 1\}$. Connecting two signatures $[1, 0, b]$ to both sides of one binary signature $[1, a, -1]$, we can get a new binary signature $[1, ab, -b^2]$. It satisfies all the conditions of Lemma 4.6, and we are done. If the arity of g is greater than 2, then we can always realize a signature $[1, 0, 0, b]$, where $b \neq 0$. (We connect the unary signature $[1, a]$ to all its dangling edges except the three ones.) Then we can use an $\mathcal{F}$-gate in Figure 9. Its signature is $[1, a^2 b, b^2]$, and by Lemma 4.6, we are done. This completes the

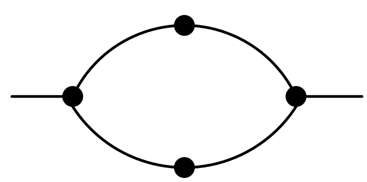

Figure 9: The function on degree 2 nodes is $[1, a, -1]$, and the function on degree 3 nodes is $[1, 0, 0, b]$.

proof. □

By the above lemmas, the only case left we have to handle is that $\mathcal{F} \subseteq \mathcal{G}_1 \cup \mathcal{F}_3$. This is done by the following lemma, which completes the proof of Theorem 4.5.

Lemma 4.13. *If $\mathcal{F} \subseteq \mathcal{G}_1 \cup \mathcal{F}_3$, then the conclusions of Theorem 4.5 hold.*

Proof. If $\mathcal{F} \subseteq \mathcal{F}_1 \cup \mathcal{F}_3$, then by Theorem 2.8 part (2), $\text{Holant}^c(\mathcal{F})$ is computable in polynomial time. Similarly, if $\mathcal{F} \subseteq \mathcal{U} \cup \mathcal{F}_3 \cup \{[1, 0, 1]\}$, then by Theorem 2.8 part (1), and then by Theorem 2.7 part (3), $\text{Holant}^c(\mathcal{F})$ is computable in polynomial time. Hence in these two cases, Theorem 4.5 holds. Now suppose $\mathcal{F} \nsubseteq \mathcal{F}_1 \cup \mathcal{F}_3$ and $\mathcal{F} \nsubseteq \mathcal{U} \cup \mathcal{F}_3 \cup \{[1, 0, 1]\}$.

There exists $f \in \mathcal{F} - \mathcal{F}_1 \cup \mathcal{F}_3$. Since $\mathcal{F} \subseteq \mathcal{G}_1 \cup \mathcal{F}_3$, such an $f \in \mathcal{G}_1$.

Now there are two cases. The first case is that we have such an $f \notin \mathcal{U}$, and so, $f \in \mathcal{F} \cap \mathcal{G}_1 - (\mathcal{F}_1 \cup \mathcal{F}_3 \cup \mathcal{U})$. The arity of f is greater than 1. By connecting its dangling edges together except two or three depends on the parity of the arity of f, we can assume f has form $[1, 0, a]$ or $[1, 0, 0, a]$, where $a \notin \{-1, 0, 1\}$.

The second case is every $f \in \mathcal{F} \cap \mathcal{G}_1 - (\mathcal{F}_1 \cup \mathcal{F}_3)$ is also in $\mathcal{U}$. By $\mathcal{F} \nsubseteq \mathcal{U} \cup \mathcal{F}_3 \cup \{[1, 0, 1]\}$, there exists $f_1 \in \mathcal{F} - (\mathcal{U} \cup \mathcal{F}_3 \cup \{[1, 0, 1]\})$. Since $\mathcal{F} \subseteq \mathcal{G}_1 \cup \mathcal{F}_3$, and $f_1 \notin \mathcal{F}_3$, we get $f_1 \in \mathcal{G}_1$. If $f_1 \notin \mathcal{F}_1$, we could use this f_1 as the f above, namely $f_1 \in \mathcal{F} \cap \mathcal{G}_1 - (\mathcal{F}_1 \cup \mathcal{F}_3 \cup \mathcal{U})$. A contradiction. Thus $f_1 \in \mathcal{F}_1$. Also we have some $f_2 \in \mathcal{F} - (\mathcal{F}_1 \cup \mathcal{F}_3)$. So $f_2 \in \mathcal{G}_1$, since $\mathcal{F} \subseteq \mathcal{G}_1 \cup \mathcal{F}_3$. Also since we are in this second case, certainly $f_2 \in \mathcal{U}$.

So we have $f_1, f_2 \in \mathcal{F} \cap \mathcal{G}_1$ such that $f_1 \in \mathcal{F}_1$ but $f_1 \notin \mathcal{U} \cup \mathcal{F}_3 \cup \{[1, 0, 1]\}$, and $f_2 \in \mathcal{U}$ but $f_2 \notin \mathcal{F}_1$. The arity of f_1 is at least 2. We claim it is greater than 2. Otherwise, f_1 being from $\mathcal{F}_1$ and not $[1, 0, 1]$, it would be

$f_1 = [1, 0, -1] \in \mathcal{F}_3$, a contradiction. So f_1 has form $[1, 0, 0, \ldots, \pm 1]$ of arity at least 3. f_2 is of form $[1, a']$, where $a' \notin \{-1, 0, 1\}$; this follows from $f_2 \in \mathcal{U} \cap \mathcal{G}_1 - \mathcal{F}_1$. By connecting all the dangling edges of f_1 except two with f_2, we can construct an $\mathcal{F}$-gate with signature of form $[1, 0, a]$, where $a \notin \{-1, 0, 1\}$. This is one of the above two forms after the first case. To sum up, in both cases, we have some f of the form $[1, 0, a]$ or $[1, 0, 0, a]$, where $a \notin \{-1, 0, 1\}$.

If $\mathcal{F} \subseteq \mathcal{G}_1 \cup \{[0, 1, 0]\} \cup \mathcal{U}$, then by Theorem 2.8 part (1), and then by Theorem 2.7 part (2) (with $a = 0$ and $b = 1$), Holant$^c(\mathcal{F})$ is computable in polynomial time and Theorem 4.5 holds. Otherwise, there exists $g \in \mathcal{F} \subseteq \mathcal{G}_1 \cup \mathcal{F}_3$, and $g \notin \mathcal{G}_1 \cup \{[0, 1, 0]\} \cup \mathcal{U}$. Then g must be in $\mathcal{F}_3$, and have one of the following sub-signatures: $[1, 1, -1], [1, -1, -1], [1, 0, -1, 0], [0, 1, 0, -1]$; this follows from a careful examination of the forms of $\mathcal{F}_3$. By symmetry (taking the reversal of both f and g), we only need to consider two cases $f = [1, 0, a]$ or $[1, 0, 0, a]$, where $a \notin \{-1, 0, 1\}$, and $g = [1, 1, -1]$ or $[1, 0, -1, 0]$.

According to f and g, we have four cases. If $f = [1, 0, a]$ and $g = [1, 1, -1]$, then connecting them together into a chain fgf, we can realize $[1, a, -a^2]$. By Lemma 4.6, we are done. If $f = [1, 0, a]$ and $g = [1, 0, -1, 0]$, for each dangling edge of g, we extend it by one copy of f. Then we can realize $[1, 0, -a^2, 0]$. So by Lemma 4.9, we are done. If $f = [1, 0, 0, a]$ and $g = [1, 1, -1]$, we can connect a unary signature $[1, 1]$ (sub-signature of g) to one dangling edge of f, and realize a binary signature $f = [1, 0, a]$. This reduces it to the first case, which has been proved. If $f = [1, 0, 0, a]$ and $g = [1, 0, -1, 0]$, we can realize a unary signature $[1, a]$ from f by connecting two of its dangling edges together, and then connect this unary signature to one dangling edge of g to realize $[1, -a, -1]$. Note that $[1, -a, -1] \notin \mathcal{G}_1 \cup \mathcal{F}_3$, by Lemma 4.12, we are done. □

5 Dichotomy for Planar Weighted #CSP

In this section, we prove a dichotomy for planar real weighted #CSP. Compared to the dichotomy for general real weighted #CSP, the new tractable cases for planar structures are precisely those which can be computed by holographic algorithms with matchgates. Since all the equality functions are assumed to be available, the only possible basis used in holographic algorithms is $\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$ (this can be computed by the characterization in [5]). Now we present the dichotomy theorem for planar weighted #CSP.

Theorem 5.1. *Let $\mathcal{F}$ be a set of real symmetric functions.* Pl-#CSP($\mathcal{F}$) *is #P-hard unless $\mathcal{F}$ satisfies one of the following conditions, in which case it is tractable:*

1. #CSP($\mathcal{F}$) *is tractable (for which we have an effective dichotomy [9]); or*
2. *Every function in $\mathcal{F}$ is realizable by some matchgate under basis* $\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$ *(for which we have a complete characterization [3]).*

The main proof idea is to reduce Pl-Holant^c problems to Pl-#CSP problems. Pl-#CSP($\mathcal{F}$) is exactly the same as planar Holant with all the Equality functions, i.e., Pl-Holant($\mathcal{F} \cup \{[1, 1], [1, 0, 1], [1, 0, 0, 1], [1, 0, 0, 0, 1], \ldots\}$). We can use a holographic reduction under the basis $H = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$. Under this transformation, the problem is transformed to, and hence has the same complexity as Pl-Holant($H\mathcal{F} \cup \{[1, 0], [1, 0, 1], [1, 0, 1, 0], [1, 0, 1, 0, 1], \ldots\}$). Since this holographic reduction gives us $[1, 0]$ (from $[1, 1]$), if we can further realize (or interpolate) $[0, 1]$, we can view the problem as a Pl-Holant^c problem and apply Theorem 4.5 to $H\mathcal{F} \cup \{[1, 0, 1], [1, 0, 1, 0], [1, 0, 1, 0, 1], \ldots\}$ to get a proof of Theorem 5.1. In the following, we show how to realize (or interpolate) $[0, 1]$. Once we have $[0, 1]$, the translation of the criterion of Theorem 4.5 to Theorem 5.1 is straightforward.

It turns out that to realize (or interpolate) $[0, 1]$ in some cases is difficult. The following lemma says that it is also sufficient if we can realize (or interpolate) $[0, 0, 1]$. $[0, 0, 1]$ can be viewed as two copies of $[0, 1]$, as $[0, 0, 1] = [0, 1] \otimes [0, 1]$. Intuitively, we will use one copy of $[0, 0, 1]$ to replace two occurrences of $[0, 1]$. However, there are two technical difficulties. One is that there may be an odd number of occurrences of $[0, 1]$ used in the input instance; the second difficulty, which is more subtle, is that we have to pair up two copies of $[0, 1]$ while maintaining planarity of the instance.

Lemma 5.2. *Pl-Holant($\mathcal{F} \cup \{[1, 0], [0, 0, 1], [1, 0, 1], [1, 0, 1, 0], [1, 0, 1, 0, 1], \ldots\}$) is #P-hard (or in P) if and only if Pl-Holant^c($\mathcal{F} \cup \{[0, 0, 1], [1, 0, 1], [1, 0, 1, 0], [1, 0, 1, 0, 1], \ldots\}$) is #P-hard (or in P).*

Proof. There is one more function $[0,1]$ in the second signature set than the first, so obviously the first one can be reduced to the second one. Hence if the second problem is in P, so is the first. We have already proved a dichotomy theorem for Pl-Holant^c problems. So now we may assume the second problem is #P-hard, and show that the first problem is also #P-hard.

We observe that all the proofs in this paper and [9], when the second problem for any signature set is proved to be #P-hard, one of the following three problems: (a) Pl-Holant($[1,0,0,1]|[1,1,0]$), (b) Pl-Holant($[1,1,0,0]$), or (c) Holant$[0,1,0,0]$ (respectively counting Vertex Cover, Matching for planar 3-regular graphs, or Perfect Matching for general 3-regular graphs) is reduced to it by a chain of reductions. There are only three reduction methods in this reduction chain, direct gadget construction, polynomial interpolation, and holographic reduction.

Given an instance G of Pl-Holant($[1,0,0,1]|[1,1,0]$), Pl-Holant($[1,1,0,0]$), or Holant$[0,1,0,0]$, we consider the graph $G \cup G$, which denotes the disjoint union of two copies of G.

Notice that the value of Pl-Holant($[1,0,0,1]|[1,1,0]$), Pl-Holant($[1,1,0,0]$), or Holant$[0,1,0,0]$ on the instance G is a non-negative integer, and the value on $G \cup G$ is its square. So we can compute the value on G uniquely from its square. Suppose the reduction chain on the instance G produced instances $G_1, G_2, \ldots, G_m$ of the second problem. The same reduction applied to $G \cup G$ produces instances of the form $G_1 \cup G_1, G_2 \cup G_2, \ldots, G_{m'} \cup G_{m'}$. (We note that the reduction on $G \cup G$ may produce polynomially more instances than on G because of polynomial interpolation.)

Now we only need to show how to transform instances $G_1 \cup G_1, G_2 \cup G_2, \ldots, G_{m'} \cup G_{m'}$ in the second problem, to instances of the first problem with the same values (replacing all occurrences of the signature $[0,1]$ by some $[0,0,1]$). $G_i \cup G_i$ is a planar graph with zero or more vertices of degree one attached with the function $[0,1]$. We want to use one copy of $[0,0,1]$ to replace one pair of $[0,1]$, while maintaining planarity.

Take a spanning tree of the dual graph of G_i. Let the outer face be the root. Choose an arbitrary leaf of this tree, which corresponds to a face C of G_i. Suppose C' is the face corresponding to the parent of C in the tree. If there are an even number of vertices of degree one attached with $[0,1]$ in face C, we can perfectly match them and realize them using $[0,0,1]$ while maintaining planarity in this face. This can be done by matching these dangling vertices of degree one in a clockwise fashion on this face C. If there are an odd number of $[0,1]$ in face C, we choose one edge e between C and C', and add a new vertex v_e on e, and connect two new vertices of degree one to v_e. The two new vertices are attached $[0,1]$, and v_e has degree 4 and is attached $[1,0,1,0,1]$. The effect of $[1,0,1,0,1]$ connected by two $[0,1]$ is the same as the function $[1,0,1]$, which is exactly the same as the edge e itself. We put one new vertex with $[0,1]$ in face C, and the other one in face C'. Now, there are an even number of $[0,1]$ in face C, and we can replace them by $[0,0,1]$ in C, as before. We may repeat this process, until we reach the root in the dual graph of G_i. If we do the same for the two G_i in $G_i \cup G_i$, we will have an even number of $[0,1]$ in the common outer face and can at last perfectly match the $[0,1]$ vertices and realize them by $[0,0,1]$. In the end we get an instance of the first problem, which has the same value. □

To sum up the above discussion, and apply Theorem 4.5, we have the following lemma, which is the starting point of our proof of Theorem 5.1.

Lemma 5.3. *If we can realize (or interpolate)* $[0,1]$ *or* $[0,0,1]$ *from* $H\mathcal{F} \cup \{[1,0],[1,0,1],[1,0,1,0],[1,0,1,0,1],\ldots\}$, *then the conclusion of Theorem 5.1 holds.*

Next we give two lemmas which give a general condition to realize or interpolate $[0,1]$ or $[0,0,1]$.

Lemma 5.4. *Let* $a \in \mathbb{R}$. *If* $a \notin \{0,1,-1\}$, *then we can interpolate* $[0,1]$ *from* $(\{[1,a],[1,0],[1,0,1],[1,0,1,0],\ldots\})$.

Proof. For every $j \geq 1$, we can take a function $F_{j+1} = [1,0,1,0,1,\ldots]$ of arity $j+1$, and connect j functions $[1,a]$ to it. The row vector form of the function (i.e., a listing of its values) of arity j composed of j copies of $[1,a]$ is $(1,a)^{\otimes j}$.

The column vector form of F_{j+1} is $1/2\begin{pmatrix}1\\1\end{pmatrix}^{\otimes(j+1)} + 1/2\begin{pmatrix}1\\-1\end{pmatrix}^{\otimes(j+1)}$. The $2^j \times 2$ matrix form of F_{j+1} is $1/2\begin{pmatrix}1\\1\end{pmatrix}^{\otimes j} \otimes (1,1) + 1/2\begin{pmatrix}1\\-1\end{pmatrix}^{\otimes j} \otimes (1,-1)$.

Our gadget realizes

$$(1,a)^{\otimes j}\left[1/2\begin{pmatrix}1\\1\end{pmatrix}^{\otimes j} \otimes (1,1) + 1/2\begin{pmatrix}1\\-1\end{pmatrix}^{\otimes j} \otimes (1,-1)\right] = \frac{(1+a)^j}{2}(1,1) + \frac{(1-a)^j}{2}(1,-1).$$

Because $a \in \mathbb{R}$ and $a \notin \{0, 1, -1\}$, $(1 + a)/(1 - a)$ is well defined and is neither zero nor a root of unity. We can interpolate any unary function $x(1, 1) + y(1, -1)$, in particular $[0, 1]$. $\square$

Lemma 5.5. *Let $a \in \mathbb{R}$. If $a \notin \{0, 1, -1\}$, then we can interpolate $[0, 0, 1]$ from $[1, 0, a]$.*

Proof. The function of a chain of length j composed of $[1, 0, a]$ is $[1, 0, a^j]$. Since the real number $a \notin \{0, 1, -1\}$, we can interpolate all $[x, 0, y]$, and in particular $[0, 0, 1]$, by polynomial interpolation. $\square$

Proof of Theorem 5.1. In this proof, we augment the class $\mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3$ to include those degenerate signatures which can be obtained from tensor products from unary signatures in $\mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3$.

If $H\mathcal{F} \subseteq \mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3$, then the problem is tractable even for general graphs and the conclusion of the theorem holds. Now we assume that there exists an $f \in H\mathcal{F} - (\mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3)$. In the following, we will prove that we can realize (or interpolate) $[0, 1]$ or $[0, 0, 1]$ from f and $\{[1, 0], [1, 0, 1], [1, 0, 1, 0], [1, 0, 1, 0, 1], \ldots\}$.

The general thrust of the proof is to squeeze all possible f into several standardized forms, and either prove #P-hardness or reach a contradiction. We assume for a contradiction that we cannot realize (or interpolate) $[0, 1]$ or $[0, 0, 1]$. Suppose $f = [f_0, f_1, \ldots, f_n]$. Since we have $[1, 0]$, we can always take an initial subsequence of an f we already have as the signature of a realizable function. Given a function g with arity $r > 1$, we often use the gadget composed of two copies of g such that $r - 1$ inputs of them are connected to each other. We call this the double gadget from g. We separate two cases according to whether $f_0 = 0$, or $f_0 \neq 0$ which we normalize to $f_0 = 1$.

1. $f_0 = 0$.

   As the constant 0 function is in $\mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3$, f is not identically 0, and thus $n \geq 1$. If $f_0 = 0$ and $f_1 \neq 0$, then we can connect $n - 1$ functions $[1, 0]$ to f to get $[0, f_1]$, which is $[0, 1]$ up to a factor.

   So we have $f_0 = f_1 = 0$, then $n \geq 2$. If $f_2 \neq 0$, then we can connect $n - 2$ functions $[1, 0]$ to f to get $[0, 0, f_2]$, which is $[0, 0, 1]$ up to a factor.

   So we have $f_0 = f_1 = f_2 = 0$, then $n \geq 3$. Let $m \leq n$ be the first nonzero, $f_0 = f_1 = f_2 = \cdots = f_{m-1} = 0$, $f_m \neq 0$, then we can connect a function $[1, 0, 1]$ to two dangling edges of f to get a function whose first nonzero entry is f_m at index $m - 2$. We can repeat this process until exactly one or two zeros are left at index 0 or at index 0 and 1, and we reach one of the two scenarios above.

2. $f_0 = 1$.

   By Lemma 5.4, we only need to consider $f_1 \in \{0, 1, -1\}$. Otherwise, we are done.

   (a) $f_0 = 1$ and $f_1 = \pm 1$.
   If $n = 1$, then $f = [f_0, f_1] = [1, \pm 1] \in \mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3$, a contradiction. Therefore we have $n \geq 2$, we can take its initial part $[1, f_1, f_2]$. Connecting one edge to $[1, f_1]$, we get $[1 + f_1^2, f_1 + f_1 f_2] = [2, \pm(1 + f_2)]$. By Lemma 5.4, we only need to consider $f_2 \in \{1, -1, -3\}$.
   We can construct another gadget which connects two inputs of $[1, 0, 1, 0]$ by $[1, f_1, f_2]$. This produces a unary signature $[1 + f_2, 2f_1]$. It follows that $f_2 \neq -1$, since otherwise we have $[0, 1]$ after normalization.
   Next we rule out $f_2 = -3$. The double gadget of $[f_0, f_1, f_2] = [1, \pm 1, f_2]$ has signature $[2, -2, 10]$ and $[2, 2, 10]$. After normalizing, this gives $[1, \pm 1, 5]$ and $5 \notin \{1, -1, -3\}$. Hence, we may assume $f_2 = 1$.
   Our goal in this case 2.(a) of $f_0 = 1$ and $f_1 = \pm 1$ is to extend this pattern $[1, \pm 1, 1, \ldots]$. Assume we have proved that $f_j = 1$ (respectively $f_j = (-1)^j$) for $j = 0, 1, \ldots, m$ ($m \geq 2$). Since $f \notin \mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3$, the arity $n > m$. We can connect $[f_0, f_1]$ to $[f_0, f_1, \ldots, f_{m+1}]$, to get a function $[f_0^2 + f_1^2, f_0 f_1 + f_1 f_2, f_0 f_2 + f_1 f_3, \ldots]$ of arity m, which is $[2, 2, \ldots, 1 + f_{m+1}]$ (respectively $[2, -2, \ldots, f_{m-1} - f_m, f_m - f_{m+1}]$). By what has been proved inductively, $f_{m+1} = 1$ (respectively $f_{m+1} = f_{m-1}$). So in this case we showed that either $f \in \mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3$, which is a contradiction, or #P-hardness.

   (b) $f_0 = 1$ and $f_1 = 0$.
   Since $[1, 0] \in \mathcal{F}_2$, and $f \notin \mathcal{F}_2$, we have $n > 1$. If f were degenerate it would be $[1, 0]^{\otimes n} = [1, 0, \ldots, 0]$, which would belong to the augmented class of $\mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_2$. But f does not. So f is non-degenerate, in particular there is some nonzero entry other than f_0. Suppose f_m is the first nonzero starting from $m > 1$. Because we can connect some $[1, 0, 1]$ to $[f_0, 0, \ldots, 0, f_m]$ to get $[1, f_m]$ or $[1, 0, f_m]$, we have $f_m = \pm 1$ by Lemma 5.4 and 5.5. Since $f \notin \mathcal{F}_1$ we have $n > m$.

We prove m is an even number. Otherwise, we can get $[1, f_m, f_{m+1}]$. By the proof for the case 2.(a), we get $f_{m+1} = 1$. We can also get $[1, 0, 0, f_m, f_{m+1}]$, since $m > 1$, whose double gadget has the signature $[1 + f_m^2, f_m f_{m+1}, 3f_m^2 + f_{m+1}^2] = [2, \pm 1, 4]$. This gives #P-hardness.

Now we know m must be even. Next we show that in fact $m = 2$. Otherwise, $m \geq 4$ and we can get $[1, 0, 0, 0, f_m, f_{m+1}]$, whose double gadget has the signature $[1 + f_m^2, f_m f_{m+1}, 4f_m^2 + f_{m+1}^2] = [2, \pm f_{m+1}, 4 + f_{m+1}^2]$. By what has been proved so far this also leads to #P-hardness.

Now we have reached $[1, 0, \pm 1, f_3]$, whose double gadget has the signature $[2, \pm f_3, 2 + f_3^2]$, so $f_3 = 0$.

Again since $f \notin \mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3$ we have $n > 3$. Hence we have $[1, 0, \pm 1, 0, f_4]$. For $[1, 0, -1, 0, f_4]$, by connecting two edges with $[1, 0, 1]$, we get $[0, 0, -1 + f_4]$, and we must have $f_4 = 1$, or else we have the signature $[0, 0, 1]$. For $[1, 0, 1, 0, f_4]$, by connecting two edges with $[1, 0, 1]$, we get $[2, 0, 1 + f_4]$, and it follows from Lemma 5.5 that $f_4 \in \{1, -1, -3\}$. Connecting three edges of $[1, 0, 1, 0, f_4]$ to three edges of $[1, 0, 1, 0, 1]$, we get $[4, 0, 3 + f_4]$, which rules out $f_4 = -1$, by Lemma 5.5 again. The double gadget of $[1, 0, 1, 0, f_4]$ gives $[4, 0, 3 + f_4^2]$, which rules out $f_4 = -3$. To sum up, we get $f_4 = 1$.

We have reached $[1, 0, \pm 1, 0, 1, \ldots]$. The rest of the proof is similar to the induction proof for the case 2.(a) but by skipping all entries with an odd index. Assume we have proved that f_j are of the proper form, for $j = 0, 1, \ldots, m$. More precisely, $f_j = 0$ for all odd $j \leq m$, and, either $f_{2j} = 1$ for all $j \leq m/2$, or $f_{2j} = (-1)^j$ for all $j \leq m/2$. Since $f \notin \mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3$, the arity $n > m$. We connect $[f_0, f_1, f_2]$ to $[f_0, f_1, \ldots, f_{m+1}]$, to get a function g of arity $m - 1$. If m is even, $f_{m-1} = 0$, and f_{m+1} is added to or subtracted from f_{m-1}, namely $f_{m-1} \pm f_{m+1}$, to form the last entry in g at index $m - 1$. This entry should be zero by induction, so $f_{m+1} = 0$. If m is odd, we can repeat the proof in the case 2.(a), but we ignore all zero entries at odd indexed locations, then the induction can be completed as before. This completes the proof.

□

6 Dichotomy for Planar 2-3 Regular Graphs

In this section we prove a dichotomy for Holant on planar 2-3 regular graphs. This setting is very interesting for at least two reasons. From dichotomy theorem point of view, this is the simplest nontrivial setting and always serves as the starting point of more general dichotomy theorems as in [9, 4]. This was also a focus of several previous work [7, 14, 8, 15], whose result is the starting point of this theorem. From the holographic algorithms point of view, most of the known holographic algorithms [22, 21] are essentially for planar 2-3 regular graphs. The dichotomy theorem here explains the reason why they are special and why many variations of them are #P-hard. In the previous two dichotomies for Pl-Holant^c and Pl-#CSP, the new tractable cases for planar are also done by holographic algorithms with matchgates. However, only special basis transformations are used since we assume some signatures are freely available. In this planar 2-3 regular graphs setting, no additional signatures are assumed to be freely available. Therefore all possible bases can be used in tractable cases.

Theorem 6.1. *Let $[y_0, y_1, y_2]$ and $[x_0, x_1, x_2, x_3]$ be two complex symmetric signatures with arity 2 and 3 respectively. Then* Pl-Holant$([y_0, y_1, y_2]|[x_0, x_1, x_2, x_3])$ *is #P-hard unless $[y_0, y_1, y_2]$ and $[x_0, x_1, x_2, x_3]$ satisfy one of the following conditions, in which case it is tractable:*

1. Holant$([y_0, y_1, y_2]|[x_0, x_1, x_2, x_3])$ *is tractable (for which we have an effective dichotomy [4]); or*
2. *There exists a basis T such that both $[y_0, y_1, y_2](T^{-1})^{\otimes 2}$ and $T^{\otimes 3}[x_0, x_1, x_2, x_3]$ are realizable by some matchgates (for which we have a complete characterization [5]).*

Proof. If $[x_0, x_1, x_2, x_3]$ or $[y_0, y_1, y_2]$ is degenerate, the problem is tractable, even for the non-planar case, and so this falls in condition *1.* Now we assume that they are both non-degenerate. As proved in [9], we can choose an invertible T_1 such that $[x_0, x_1, x_2, x_3]$ (or its reversal, which is similar and we omit that case) can be written as $T_1^{\otimes 3}[1, 0, 0, 1]$ or $T_1^{\otimes 3}[1, 1, 0, 0]$. Therefore by a holographic reduction, we can always reduce the problem equivalently to one of the following two problems: (1) Pl-Holant$([z_0, z_1, z_2]|[1, 0, 0, 1])$ and (2) Pl-Holant$([z_0, z_1, z_2]|[1, 1, 0, 0])$. So it is sufficient to prove the theorem for these two cases.

For Pl-Holant$([z_0, z_1, z_2]|[1, 0, 0, 1])$, a dichotomy theorem proved in [15] is also valid for planar structures. By that theorem, the only case which is hard for general graphs and tractable for planar graphs is $(z_0)^3 = (z_2)^3$.

This condition is exactly the same as the condition that there exists a basis T such that both $[y_0, y_1, y_2](T^{-1})^{\otimes 2}$ and $T^{\otimes 3}[1,0,0,1]$ are realizable by some matchgates. This proves Theorem 6.1 for case (1).

Now we consider Pl-Holant($[z_0, z_1, z_2]|[1,1,0,0]$). If $z_0 = 0$, the problem is trivially tractable even for general graphs. This can be seen by a simple counting argument: in a bipartite graph the LHS vertices all have the signature $[0, z_1, z_2]$ and thus at least half the edges must be 1, while the RHS vertices all have the signature $[1,1,0,0]$ and thus less than half the edges are 1. This is also the only case where the problem is not #P-hard for general graphs when the RHS has $[1,1,0,0]$ by [4]. Now we assume $z_0 \neq 0$. Then it is sufficient to prove that either the problem is #P-hard or there exists a basis transformation T such that $[1,1,0,0]T^{\otimes 3}$ and $(T^{-1})^{\otimes 2}[z_0, z_1, z_2]$ are realizable by some matchgates. Let $T = \begin{bmatrix} \sqrt{z_0} & 0 \\ z_1/\sqrt{z_0} & \sqrt{(z_0 z_2 - (z_1)^2)/z_0} \end{bmatrix}$. Note that T is well defined and invertible since $z_0 \neq 0$ and $[z_0, z_1, z_2]$ is non-degenerate (i.e., $z_0 z_2 - (z_1)^2 \neq 0$). Then we can verify that

$$[1,1,0,0]T^{\otimes 3} = [\sqrt{z_0}(z_0 + 3z_1), \sqrt{z_0(z_0 z_2 - (z_1)^2)}, 0, 0] \quad \text{and} \quad (T^{-1})^{\otimes 2}[z_0, z_1, z_2] = [1,0,1].$$

We note that $\sqrt{z_0(z_0 z_2 - (z_1)^2)} \neq 0$. If $\sqrt{z_0}(z_0 + 3z_1) = 0$, then both $[\sqrt{z_0}(z_0 + 3z_1), \sqrt{z_0(z_0 z_2 - (z_1)^2)}, 0, 0]$ and $[1,0,1]$ can be realized by matchgates and the problem for planar graphs is tractable. We denote $v = \frac{\sqrt{z_0}(z_0+3z_1)}{\sqrt{z_0(z_0 z_2 - (z_1)^2)}} \neq 0$. Then the problem is equivalent to (non-bipartite) Pl-Holant($[v,1,0,0]$). Now it is sufficient to prove the following claim:

Claim: Let $v \neq 0$ be a complex number. Then Pl-Holant($[v,1,0,0]$) is #P-hard.

We can realize $[v^3 + 3v, v^2 + 1, v, 1]$ by connecting 3 copies of $[v,1,0,0]$'s as illustrated in Figure 10. If we

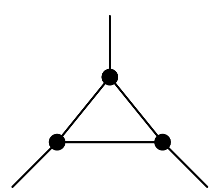

Figure 10: All vertex signatures are $[v,1,0,0]$.

can prove that Pl-Holant($[v^3 + 3v, v^2 + 1, v, 1]$) is #P-hard, then we are done. In tensor product notation this signature is

$$[v^3 + 3v, v^2 + 1, v, 1]^{\mathtt{T}} = \frac{1}{2}\left(\begin{bmatrix} v+1 \\ 1 \end{bmatrix}^{\otimes 3} + \begin{bmatrix} v-1 \\ 1 \end{bmatrix}^{\otimes 3} \right).$$

Then the following reduction chain holds:

$$\begin{aligned} \text{Pl-Holant}([v^3 + 3v, v^2 + 1, v, 1]) &\equiv_{\mathtt{T}} \text{Pl-Holant}([1,0,1]|[v^3 + 3v, v^2 + 1, v, 1]) \\ &\equiv_{\mathtt{T}} \text{Pl-Holant}([v^2 + 2v + 2, v^2, v^2 - 2v + 2]|[1,0,0,1]), \end{aligned}$$

where the second step is a holographic reduction using $\begin{bmatrix} v+1 & v-1 \\ 1 & 1 \end{bmatrix}$. This transforms the problem to our first case where the RHS all have $[1,0,0,1]$. The only possible exceptional case happens when $(v^2 + 2v + 2)^3 = (v^2 - 2v + 2)^3$. Since $(v^2 + 2v + 2)^3 - (v^2 - 2v + 2)^3 = 4v(3v^4 + 16v^2 + 12)$ and $v \neq 0$, we will have proved the claim as long as $3v^4 + 16v^2 + 12 \neq 0$. There are four roots for the equation $3v^4 + 16v^2 + 12 = 0$, and for these four exceptional values of v, we prove it separately as follows.

In addition to the gadget in Figure 10, we can construct a gadget in Figure 11 with a binary signature $[v^2 + 2, v, 1]$. Now it is enough to prove that Pl-Holant($[v^2 + 2, v, 1]|[v^3 + 3v, v^2 + 1, v, 1]$) is #P-hard. Under the same basis $\begin{bmatrix} v+1 & v-1 \\ 1 & 1 \end{bmatrix}$, we will get an equivalent problem Pl-Holant($[X, Y, Z] \mid [1,0,0,1]$), where $X = (v^2+2)(v^2+2v+1) + 2v(v+1) + 1$, $Y = (v^2+2)(v^2-1) + 2v^2 + 1$, and $Z = (v^2+2)(v^2-2v+1) + 2v(v-1) + 1$. Again this transforms the problem to our first case, and, it is easy to verify that any root of $3v^4 + 16v^2 + 12 = 0$ is not a tractable case here. This completes the proof of the claim and also the proof of the theorem. $\square$

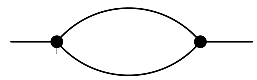

Figure 11: All vertex signatures are $[v, 1, 0, 0]$.

Appendix: Some Connections to Statistical Physics

In this section we describe some background and connections from Statistical Physics. Our discussion is necessarily a superficial one, both due to our limited knowledge and limitation on space. The purpose is to illustrate that, even at such a superficial level, a strong connection exists, and that our complexity results may shed some light on the venerable question from physics: Exactly what "systems" can be solved "exactly" and what "systems" are "difficult".

The Ising model was named after Ernst Ising. Wilhelm Lenz invented this model and gave it to his student Ising to work on it. The model consists of a discrete set of variables, called spins, that can be assigned one of two values (states). These spins are usually placed on a lattice structure or a graph, and each spin interacts with its nearest neighbors.

Denoting the values each spin i can take as $\sigma_i = +1$ and -1, the energy (the Hamiltonian) of the Ising model is $H(\sigma) = -\sum_{\text{edge}\{i,j\}} J_{i,j}\sigma_i\sigma_j$. The interaction between spins i and j is called ferromagnetic if $J_{i,j} > 0$, antiferromagnetic if $J_{i,j} < 0$, and noninteracting if $J_{i,j} = 0$. E.g., if all the spins are placed on a one-dimensional lattice, then the antiferromagnetic one-dimensional Ising model (with the same value $J_{i,j} = J < 0$) has the energy function $H = \sum_i \sigma_i\sigma_{i+1}$, after normalization. The ferromagnetic two-dimensional Ising model on a square lattice (with the same value $J_{i,j} = J > 0$) has energy $H = -\sum_{i,j}(\sigma_{i,j}\sigma_{i,j+1} + \sigma_{i,j}\sigma_{i+1,j})$. The Ising model may be modified by magnetic fields which amounts to a unary function at each spin $H = -\sum_{\text{edge}\{i,j\}} J_{i,j}\sigma_i\sigma_j - \sum_i h_i\sigma_i$.

The model is a statistical model. The central premise of statistical physics is that the probability of each configuration σ is given by the Boltzmann distribution, $e^{-H(\sigma)/kT} / \sum_\sigma e^{-H(\sigma)/kT}$, where k is Boltzmanns constant and T is the (absolute) temperature. This focuses attention on the partition function

$$Z = \sum_\sigma e^{-H(\sigma)/kT}.$$

Note that the exponential $e^{-H(\sigma)/kT}$ turns this into a sum-of-product functions exactly as we discussed in #CSP.

In 1925, Ising solved the one-dimensional Ising model. The 2-dimensional square lattice Ising model with zero magnetic field was solved by Onsager in 1944. Onsager announced the formula for the spontaneous magnetization for the two-dimensional model in 1949 but did not give a derivation. C.N.Yang (1952) gave the first published proof of this formula, using a limit formula for Fredholm determinants, proved in 1951 by Szegö in direct response to Onsager's work. There are many extensions to the basic Ising model.

Another landmark achievement is the exact computation of the number of perfect matchings (dimer problem) on any planar graph using Pfaffians. This was independently discovered by Kasteleyn and by Fisher and Temperley [**?**, **?**]. This problem can also be nicely expressed by a partition function in our Holant framework; where this time the Boolean variables are the edges (to include an edge or not), and the local constraint function at each vertex is the Exact-One function. Freedman, Lovász and Schrijver [**?**] recently proved that this partition function *cannot* be expressed as a graph homomorphism function, where the vertices are variables as in the Ising model. However in the framework of Holant problems we can find a unity for all these problems.

We note the following. In the paper [**?**] we gave a complete characterization of matchgate realizable symmetric signatures. The following lemma is proved [**?**]:

Lemma 6.2. *The set of bases under which the signature* $[x_0, x_1, x_2]$ *is realizable as a recognizer signature by some matchgate is*

$$\left\{ \left[\begin{pmatrix} n_0 \\ n_1 \end{pmatrix}, \begin{pmatrix} p_0 \\ p_1 \end{pmatrix} \right] \in \mathcal{M} \;\middle|\; \begin{array}{c} x_0p_1^2 - 2x_1p_1n_1 + x_2n_1^2 = 0, x_0p_0^2 - 2x_1p_0n_0 + x_2n_0^2 = 0 \\ \text{or} \quad x_0p_0p_1 - x_1(n_0p_1 + n_1p_0) + x_2n_0n_1 = 0 \end{array} \right\}.$$

This has the consequence that under the basis $\left[\binom{n_0}{n_1}, \binom{p_0}{p_1}\right] = \left[\binom{1}{1}, \binom{1}{-1}\right]$, the signature $[x, y, x]$ is realizable by a matchgate, for all values x and y. In terms of the Ising model, when two interacting spins i and j take the same assignment value $\sigma_i = \sigma_j = \pm 1$, the contribution to the Hamiltonian is $-J_{i,j}$, and when they take the opposite assignment $\sigma_i = -\sigma_j = \pm 1$, the contribution is $J_{i,j}$. Translating this to the contributions to the partition function we get exactly the local constraint evaluation $x = e^{J_{i,j}/kT}$ when inputs are 00 or 11, and $y = e^{-J_{i,j}/kT}$ when inputs are 01 and 10.

Then, the theory of Holographic Algorithms tells us that for planar graphs, this Ising model is exactly solvable by a holographic reduction to the FKT algorithm.

The present paper, especially Theorem 5.1, tells us why this is exactly where physicists stopped, and attempts to generalize this to non-planar systems have not been successful in the past 85 years.

Sorin Istrail [**?**] showed that computing the free energy of an arbitrary subgraph of an Ising model on a lattice of dimension three or more is NP-hard; see a nice article by Barry Cipra in the SIAM News [**?**]. A very partial list of a great deal of research on this and related models, from a computational complexity perspective, can be found in [**?**].

References

[1] Michael Alekhnovich, Allan Borodin, Joshua Buresh-Oppenheim, Russell Impagliazzo, Avner Magen, and Toniann Pitassi. Toward a model for backtracking and dynamic programming. In *CCC '05: Proceedings of the 20th Annual IEEE Conference on Computational Complexity*, pages 308–322, Washington, DC, USA, 2005. IEEE Computer Society.

[2] Allan Borodin. Towards a theory of algorithms. In Frank K. H. A. Dehne, Alejandro López-Ortiz, and Jörg-Rüdiger Sack, editors, *WADS*, volume 3608 of *Lecture Notes in Computer Science*, page 1. Springer, 2005.

[3] Jin-Yi Cai, Vinay Choudhary, and Pinyan Lu. On the theory of matchgate computations. In *CCC '07: Proceedings of the Twenty-Second Annual IEEE Conference on Computational Complexity*, pages 305–318, Washington, DC, USA, 2007. IEEE Computer Society.

[4] Jin-Yi Cai, Sangxia Huang, and Pinyan Lu. From holant to #csp and back: Dichotomy for holant^c problems. *submitted*, 2009.

[5] Jin-Yi Cai and Pinyan Lu. Holographic algorithms: from art to science. In *STOC '07: Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*, pages 401–410, New York, NY, USA, 2007. ACM.

[6] Jin-Yi Cai and Pinyan Lu. On symmetric signatures in holographic algorithms. In Wolfgang Thomas and Pascal Weil, editors, *STACS*, volume 4393 of *Lecture Notes in Computer Science*, pages 429–440. Springer, 2007.

[7] Jin-Yi Cai, Pinyan Lu, and Mingji Xia. Holographic algorithms by fibonacci gates and holographic reductions for hardness. In *FOCS '08: Proceedings of the 49th Annual IEEE Symposium on Foundations of Computer Science*, Washington, DC, USA, 2008. IEEE Computer Society.

[8] Jin-Yi Cai, Pinyan Lu, and Mingji Xia. A computational proof of complexity of some restricted counting problems. In Jianer Chen and S. Barry Cooper, editors, *TAMC*, volume 5532 of *Lecture Notes in Computer Science*, pages 138–149. Springer, 2009.

[9] Jin-Yi Cai, Pinyan Lu, and Mingji Xia. Holant problems and counting csp. In Michael Mitzenmacher, editor, *STOC*, pages 715–724. ACM, 2009.

[10] C. T. J. Dodson and T. Poston. *Tensor Geometry.* Graduate Texts in Mathematics 130. Springer-Verlag, New York, 1991.

[11] Martin E. Dyer and Catherine S. Greenhill. The complexity of counting graph homomorphisms (extended abstract). In *SODA*, pages 246–255, 2000.

[12] P. W. Kasteleyn. The statistics of dimers on a lattice. *Physica*, 27:1209–1225, 1961.

[13] P. W. Kasteleyn. Graph theory and crystal physics. In (F. Harary, editor, *Graph Theory and Theoretical Physics*, pages 43–110. Academic Press, London, 1967.

[14] Michael Kowalczyk. Classification of a class of counting problems using holographic reductions. In Hung Q. Ngo, editor, *COCOON*, volume 5609 of *Lecture Notes in Computer Science*, pages 472–485. Springer, 2009.

[15] Michael Kowalczyk and Jin-Yi Cai. Holant problems for regular graphs with complex edge functions. *submitted*, 2009.

[16] H. N. V. Temperley and M. E. Fisher. Dimer problem in statistical mechanics c an exact result. *Philosophical Magazine*, 6:1061C 1063, 1961.

[17] Salil P. Vadhan. The complexity of counting in sparse, regular, and planar graphs. *SIAM J. Comput.*, 31(2):398–427, 2001.

[18] Leslie G. Valiant. The complexity of enumeration and reliability problems. *SIAM J. Comput.*, 8(3):410–421, 1979.

[19] Leslie G. Valiant. Expressiveness of matchgates. *Theor. Comput. Sci.*, 289(1):457–471, 2002.

[20] Leslie G. Valiant. Quantum circuits that can be simulated classically in polynomial time. *SIAM J. Comput.*, 31(4):1229–1254, 2002.

[21] Leslie G. Valiant. Accidental algorthims. In *FOCS '06: Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science*, pages 509–517, Washington, DC, USA, 2006. IEEE Computer Society.

[22] Leslie G. Valiant. Holographic algorithms. *SIAM J. Comput.*, 37(5):1565–1594, 2008.

[23] Mingji Xia, Peng Zhang, and Wenbo Zhao. Computational complexity of counting problems on 3-regular planar graphs. *Theor. Comput. Sci.*, 384(1):111–125, 2007.