Dichotomy for Holant* Problems of Boolean Domain

Jin-Yi Cai* Pinyan Lu† Mingji Xia‡

Abstract

Holant problems are a general framework to study counting problems. Both counting Constraint Satisfaction Problems (#CSP) and graph homomorphisms are special cases. We prove a complexity dichotomy theorem for $\mathrm{Holant}^*(\mathcal{F})$, where $\mathcal{F}$ is a set of constraint functions on Boolean variables and taking complex values. The constraint functions need not be symmetric functions. We identify four classes of problems which are polynomial time computable; all other problems are proved to be #P-hard. The main proof technique and indeed the formulation of the theorem use holographic algorithms and reductions. By considering these counting problems over the complex domain, we discover surprising new tractable classes, which are associated with isotropic vectors, i.e., a (non-zero) vector whose inner product with itself is zero.

*University of Wisconsin-Madison and Beijing University. `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. `mingji@ios.ac.cn`

1 Introduction

Many graph counting problems can be formulated as computing *partition functions*. For example INDEPENDENT SET can be formulated as follows: Given a graph $G = (V, E)$, attach to every edge $e \in E$ the NAND function f_e. For any vertex assignment $\sigma : V \to \{0, 1\}$, define the weight function $\mathbf{wt}(\sigma) = \prod_{e=\{u,v\}\in E} f_e(\sigma(u), \sigma(v))$. Then $\mathbf{wt}(\sigma) \neq 0$ iff $\sigma^{-1}(1)$ is an independent set. The counting problem is to compute the partition function of spin-system $\mathbf{Z}(G) = \sum_\sigma \mathbf{wt}(\sigma)$. By varying the edge functions f_e, other problems can be stated in a uniform way, e.g., VERTEX COVER corresponds to the Boolean OR function, and 3-COLORING corresponds to the DISEQUALITY function on domain size 3. The functions f_e need not be 0-1 valued. Nonnegative values are the most natural combinatorially, but negative or complex values are also interesting. E.g., let $H = \left(\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}\right)$ be the Hadamard matrix, which defines a function $H(0, 0) = H(0, 1) = H(1, 0) = 1$ and $H(1, 1) = -1$. The weight function $\mathbf{wt}(\sigma) = \pm 1$, and is -1 precisely when the induced subgraph on $\sigma^{-1}(1)$ has an odd number of edges. Therefore, $(2^n - \mathbf{Z}(G))/2$ is the number of induced subgraphs with an odd number of edges. We will demonstrate in this paper that, at a deeper level, by considering general complex valued functions[1] we gain a more structural understanding mathematically.

When every edge is attached the same symmetric edge function it is called a graph homomorphism problem [29, 20]. There is also a long history in statistical physics community in the study of partition functions. Ever since Wilhelm Lenz asked his student Ernst Ising [21] to work on what's now known as the Ising model, physicists have studied so-called "Exactly Solved Models" [2, 31]. In computer science language, physicists' notion of an "exactly solvable" system corresponds to systems with polynomial time computable partition functions. Many physicists (Ising, Onsager, Fisher, Temperley, Kasteleyn, C. N. Yang, T. D. Lee, Baxter, Lieb, Wilson e.t.c. [21, 32, 43, 44, 27, 36, 24, 25, 2, 28, 42]) contributed to this intellectual edifice. But the physicists lacked a formal notion of what it means to be not "exactly solvable", which should correspond to #P-hardness. Great progress has been made on the complexity of partition functions, giving classification theorems [16, 4, 19, 37, 6, 15, 5] in terms of polynomial time tractability or #P-hardness. A major further research direction is when a #P-hard partition function can be approximated [22, 14, 12, 23, 30, 33, 18].

Now consider the problem of counting perfect matchings. Given a graph $G = (V, E)$, attach a local constraint function f_v to every vertex $v \in V$. For perfect matchings, let f_v be the EXACT-ONE function. We now consider edges to be variables. For any assignment $\sigma : E \to \{0, 1\}$, let $\mathbf{wt}(\sigma) = \prod_{v\in V} f_v(\sigma \mid_{E(v)})$, where $E(v)$ are the incident edges at v. For $f_v =$ EXACT-ONE, the weight function $\mathbf{wt}(\sigma) \neq 0$ iff $\sigma^{-1}(1)$ is a perfect matching. We define $\mathrm{Holant}(G) = \sum_{\sigma:E\to\{0,1\}} \mathbf{wt}(\sigma)$. Given a choice of local constraint functions, a Holant problem on G is to evaluate $\mathrm{Holant}(G)$.

Holant problems were defined in [8], and the name was inspired by the introduction of *Holographic Algorithms* by L. Valiant [41, 40] (who first used the term Holant). It is easy to simulate a partition function by Holant. In fact Holant problems can simulate all #CSP problems. A #CSP problem is specified by a bipartite graph $G = (V, W, E)$ where each $v \in V$ is a variable, each $w \in W$ has a constraint function f_w, and $N(w) = \{v \in V \mid (v, w) \in E\}$ is the (ordered) set of variables f_w applies to. The computational problem of a #CSP instance is to evaluate $\sum_\sigma \prod_w f_w(\sigma \mid_{N(w)})$, a sum, over all assignments σ on V, of the products of all function evaluations f_w on $N(w)$. The partition function of a spin system is a special case of #CSP where every $w \in W$ has degree 2. On the other hand, given any #CSP instance, if we assign EQUALITY functions at every $v \in V$, and consider E as variables, then the #CSP problem on G is reduced to a Holant problem[2].

[1] To avoid any difficulties with models of computation, we restrict to functions taking algebraic numbers in $\mathbb{C}$.

[2] On the other hand, Freedman, Lovász, and Schrijver [17] showed that counting perfect matchings cannot be expressed as graph homomorphisms.

To study which counting problems are computable in polynomial time (tractable) and which are not (intractable), we try to characterize this by the function sets used as local constraints. An ideal outcome in this line of research is to be able to classify, within a broad class of functions, *every* function set either leads to tractable problems or is #P-hard. This is called a dichotomy theorem [34, 10, 35] (By an analogue of Ladner's theorem [26], such a dichotomy is *false* for the whole #P, unless $\mathrm{P} = \mathrm{P}^{\#\mathrm{P}}$.) Dichotomy theorems have been obtained for counting graph homomorphisms for successively broader class of functions [16, 4, 19, 37, 6, 15, 5]. A sweeping dichotomy theorem for all #CSP with 0-1 constraint functions over any finite domain was given by Bulatov [3]. An alternative proof is given in [11]. It can be extended to functions taking non-negative rational values [1]. However in general when negative values are allowed, cancelations occur, and this could lead to surprising P-time algorithms. Holographic Algorithms precisely take advantages of such cancelations. By operating without restriction to non-negative values, some deeper underlying mathematical structures become visible (cf. [19, 6]).

For any set of functions $\mathcal{F}$, we use Holant($\mathcal{F}$) to denote the class of Holant problems using $\mathcal{F}$. Similarly #CSP($\mathcal{F}$) is the class of #CSP problems using $\mathcal{F}$. Let $\mathcal{EQ} = \{=_k \mid k \geq 1\}$ denote the set of Equality functions. Then #CSP($\mathcal{F}$) = Holant($\mathcal{F} \cup \mathcal{EQ}$) (i.e., #CSP = Holant with $\mathcal{EQ}$ for free.)

It turns out that allowing Equality functions for free has a major influence on tractability. By making the presence of these Equality functions explicit, the Holant framework makes a finer complexity classification than #CSP. While #CSP is Holant with $\mathcal{EQ}$ for free, we can consider other special cases of Holant problems. It turns out that the set $\mathcal{U}$ of all unary functions are structurally important. Tensor products by unary functions constitute all *degenerate* functions, which have played a crucial role in many classification theorems. Holant* is the class of Holant problems where all unary functions are free, i.e., Holant*($\mathcal{F}$) = Holant($\mathcal{F} \cup \mathcal{U}$).

Previously we have studied Holant* problems for any set $\mathcal{F}$ of *symmetric* functions on Boolean variables [9]. This study led to a complexity dichotomy theorem for all #CSP($\mathcal{F}$), where $\mathcal{F}$ is any set of complex-valued constraint functions on Boolean variables [9]. This improves previously the strongest dichotomy for Boolean #CSP($\mathcal{F}$) by Dyer, Goldberg and Jerrum [13], which applies to nonnegative-valued constraint functions. The extension to complex-valued constraint functions not only extends the scope formally, it also discovers inherent structural properties not visible for nonnegative numbers.

The main result in this paper is to prove a dichotomy theorem for all Holant*($\mathcal{F}$), where $\mathcal{F}$ is any set of complex-valued functions on Boolean variables, and these functions need *not be symmetric*. This research is strongly influenced by the development of holographic algorithms and reductions [39, 40, 7, 8], Indeed, they not only provide the main proof techniques but also aid in the discovery and formulation of the theorem.

The theorem identifies four classes of functions $\mathcal{F}$ where Holant*($\mathcal{F}$) is polynomial time computable. These can be roughly described as follows: The first class $\mathcal{F}_1$ is tractable due to its arity, and the computation is done by matrix product and taking trace. The second tractable class $\mathcal{F}_2$ is a generalization of the so-called Fibonacci gates, denoted by $\mathscr{F}$ [8]. These are symmetric functions and Holant*($\mathscr{F}$) is tractable. $\mathcal{F}_2$ generalizes this to functions that are not necessarily symmetric. Here holographic transformations become crucial, which allow us to *discover* and to *express* this class in a succinct and elegant way. It is basically Fibonacci gates under an orthogonal transformation[3].

The third and fourth tractable classes $\mathcal{F}_3$, $\mathcal{F}_4$ depend even more fundamentally on holographic transformations. It is also here that the complex domain $\mathbb{C}$ becomes essential. Over $\mathbb{C}$ there are so-called *isotropic* vectors $v \neq 0$ which satisfy $v^{\mathrm{T}} v = 0$. (No nonzero real vector has this property.) $\mathcal{F}_3$ (resp. $\mathcal{F}_4$) are Fibonacci gates (resp. Matching gates, a class related to weighted matchings), after a holographic transformation correlated with isotropic vectors.

Our dichotomy here is a generalization of the dichotomy in [9] for symmetric Holant* Problems. The

[3] In this paper, we actually present it slightly differently, in order to give a more succinct proof.

symmetric dichotomy can be viewed as a special case of the dichotomy in this paper and on the other hand also servers as the starting point for our reduction. Furthermore, by seeing the whole picture, we also gain a deeper and clearer understanding of the tractable cases for the symmetric ones.

In Section 2, we give some formal definitions and state the main theorem. In Section 3 we prove the tractability results. Section 4 gives a proof outline. In Section 5 we prove some useful algebraic lemmas. In Sections 6 and 7 we prove that, assuming $\mathrm{P} \neq \mathrm{P}^{\#\mathrm{P}}$, we have found *all* the tractable $\mathrm{Holant}^*(\mathcal{F})$.

2 Definition and statement

A (constraint) function, or synonymously a signature, of arity $n \geq 0$, is a mapping from $\{0,1\}^n \rightarrow \mathbb{C}$. A function of arity 0 is a constant. A function of arity 1 is called a unary function. We use the same symbol F to denote the column vector indexed by $\{0,1\}^n$ as an expression of F, listing all its values. When we use it as a row vector we write F^{T}. Sometimes it is also convenient to partition the variable set into two parts $\{x_1, x_2, \ldots, x_n\} = I \cup J$, and write F as a matrix with rows indexed by $\{0,1\}^{|I|}$ and columns indexed by $\{0,1\}^{|J|}$. This is particularly useful for a binary function $F(x,y)$, whose matrix form $F = F_{x,y}$ is a 2×2 matrix, with row index x and column index y range over $\{0,1\}$. We also use this matrix form for functions of larger arities. For example, $F_{x_1 x_2, x_3}$ is a 4×2 matrix.

Suppose $c \in \mathbb{C}$ is a nonzero number. As constraint functions F and cF are equivalent in terms of the complexity of Holant problems they define. Hence we will consider functions F and cF to be interchangeable, denoted by $F \cong cF$. The notation $F \cong 0$ means that F is (identically) zero.

We denote by $=_k$ the Equality function 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), and $(=_2) = [1,0,1]$ ($= (1,0,0,1)$ in row vector form).

A *signature grid* $\Omega = (G, \mathcal{F}, \pi)$ consists of a graph $G = (V,E)$, and a labeling π of each vertex $v \in V$ with a function $f_v \in \mathcal{F}$. The Holant problem on instance Ω is to compute $\mathrm{Holant}_\Omega = \sum_{\sigma: E \rightarrow \{0,1\}} \prod_{v \in V} f_v(\sigma \mid_{E(v)})$. 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* $\mathrm{Holant}(\mathcal{F})$*:*
Input: A signature grid $\Omega = (G, \mathcal{F}, \pi)$;
Output: Holant_Ω.

We would like to characterize the complexity of Holant problems in terms of its signature sets[4].

Definition 2.2. *Let $\mathcal{U}$ denote the set of all unary signatures. Then* $\mathrm{Holant}^*(\mathcal{F}) = \mathrm{Holant}(\mathcal{F} \cup \mathcal{U})$.

In [9], we proved a dichotomy theorem when $\mathcal{F}$ is a set of symmetric signatures.

Theorem 2.1. *Let $\mathcal{F}$ be a set of* symmetric *signatures over $\mathbb{C}$. Then $\mathrm{Holant}^*(\mathcal{F})$ is computable in polynomial time in the following three* Classes. *In all other cases, $\mathrm{Holant}^*(\mathcal{F})$ 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 all signatures $[x_0, x_1, \ldots, x_n] \in \mathcal{F}$ one of the two conditions is satisfied: (1) for every $k = 0, 1, \ldots, n-2$, we have $a x_k + b x_{k+1} - a x_{k+2} = 0$; (2) $n = 2$ and the signature $[x_0, x_1, x_2]$ is of the form $[2a\lambda, b\lambda, -2a\lambda]$.*

[4]We allow $\mathcal{F}$ to be an infinite set. $\mathrm{Holant}(\mathcal{F})$ 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. $\mathrm{Holant}(\mathcal{F})$ is #P-hard means that there exists a finite subset of $\mathcal{F}$ for which the problem is #P-hard.

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 the form* $[\lambda, 0, \lambda]$.

The dichotomy is still true even if the inputs are restricted to planar graphs.

An $\mathcal{F}$-gate Γ, or a gadget, is a tuple $(H, \mathcal{F}, \pi)$, where $H = (V, E, D)$ is a graph with some dangling edges D. (See Figure 1 for one example.) Other than these dangling edges, an $\mathcal{F}$-gate is the same as a

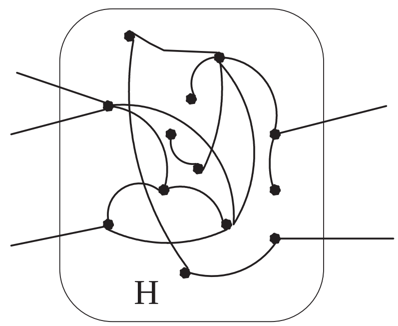

Figure 1: An example of an $\mathcal{F}$-gate with five dangling edges.

signature grid. The role of dangling edges is similar to that of external nodes in Valiant's notion [38, 41], however we allow more than one dangling edges for a node. In $H = (V, E, D)$ each node is assigned a function in $\mathcal{F}$ by the mapping π (we do not consider "dangling" leaf nodes at the end of a dangling edge among these), E is the set of regular edges, denoted as $1, 2, \ldots, m$, and D is the set of dangling edges, denoted as $m+1, m+2, \ldots, m+n$. 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.

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 say g is realized by the gadget Γ.

The most direct and general way to express a gadget and its function, is the graph of the gadget. But in order to reason about this function, we need some simple and intuitive notations, especially for two basic compositional constructions. The first operation is identifying two variables. We use $F^{x_i = x_j}$ to denote the function of arity $n-2$ realized by a function F of arity $n \geq 2$, such that the two dangling edges corresponding x_i and x_j are merged to become one (internal) edge. (See Figure 2 for one example.)

The second operation is called juxtaposition. Suppose F is a function of arity n and $\mathcal{I} = \{I_1, \ldots, I_k\}$ is a partition of $[n]$. If $F(X) = \prod_{j=1}^{k} F_j(X|_{I_j})$ for some functions $F_1, \ldots, F_k$, where $X = \{x_1, \ldots, x_n\}$ and $X|_{I_j} = \{x_s | s \in I_j\}$ (we also denote it by X_j), then we say F can be decomposed into type $\mathcal{I}$, or simply F has type $\mathcal{I}$. We denote such an F by $F = \bigotimes_{\mathcal{I}}(F_1, \ldots, F_k)$. If each F_j is the function of some gadget, then $\bigotimes_{\mathcal{I}}(F_1, \ldots, F_k)$ is the function of the gadget which is the disjoint union of these

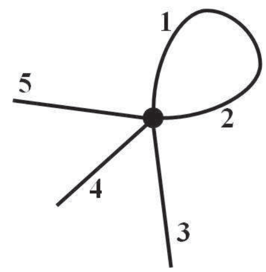

Figure 2: An example of $F^{x_1=x_2}$.

gadgets for F_j, with variables arranged according to $\mathcal{I}$. When the indexing is clear, we also use notation $F_1 \otimes \cdots \otimes F_k$. Note that this tensor product notation $\otimes$ is consistent with tensor product of matrices. (See Figure 3 for one example.)

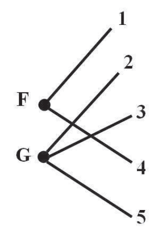

Figure 3: An example of juxtaposition $\bigotimes_{\mathcal{I}}(F, G) \in \mathcal{F}$, where $\mathcal{I} = \{\{1, 4\}, \{2, 3, 5\}\}$.

We use $F^{x_{j_1}=U_1,\ldots,x_{j_k}=U_k}$ to denote the function of arity $n-k$ realized by a function F of arity n such that its input variable x_{j_s} is connected with the unary function U_s, $s = 1, \ldots, k$. $F^{x_j=0}$, $F^{x_j=1}$ and FU are respectively abbreviations for $F^{x_j=[1,0]}$, $F^{x_j=[0,1]}$ and $F^{x_j=U}$ (where x_j is clear from the context). Note that $[1, 0]$ and $[0, 1]$ are two special unary functions.

We also use matrix multiplication, especially when gadgets are sequentially chained together. For example, suppose $A = A_{x_1,x_2}$, $B = B_{x_3,x_4}$ and $C = C_{x_5,x_6}$ are three binary functions. Then ABC expresses the function $(A \otimes B \otimes C)^{x_2=x_3,x_4=x_5}$, which has the matrix form exactly the matrix product ABC, indexed by x_1 and x_6. Note that $A_{\emptyset,x_1} B_{x_2,\emptyset}$ or $A^{\mathrm{T}}B$ is the inner product of unary functions A and B. Similarly, $A_{x_1,\emptyset} B_{\emptyset,x_2}$ or AB^{T} is the matrix form of the tensor product function $\bigotimes_{\{\{1\},\{2\}\}}(A, B)$ (or just $A \otimes B$) of unary functions A and B.

We say a function set $\mathcal{F}$ is closed under tensor product (or more precisely under juxtaposition), if for any $A, B \in \mathcal{F}$ and $\mathcal{I} = \{I_1, I_2\}$, $\bigotimes_{\mathcal{I}}(A, B) \in \mathcal{F}$. Tensor closure $\langle \mathcal{F} \rangle$ of a set $\mathcal{F}$ is the minimum set containing $\mathcal{F}$, closed under tensor product. This closure exists, being the set of all functions obtained by taking a finite sequence of tensor products from $\mathcal{F}$.

Next we define several important sets of functions on Boolean variables. $\mathcal{U}$ is the set of all unary functions. $\mathcal{E}$ is the set of all functions F such that F is zero except on two inputs $(a_1, \ldots, a_n)$ and $(\bar{a}_1, \ldots, \bar{a}_n) = (1 - a_1, \ldots, 1 - a_n)$. In other words, $F \in \mathcal{E}$ iff its support is contained in a pair of complementary points. We think of $\mathcal{E}$ as a generalized form of Equality. $\mathcal{M}$ is the set of all functions F such that F is zero except on $n+1$ inputs whose Hamming weight is at most 1, where n is the arity of F. The name $\mathcal{M}$ is given for *matching*. $\mathcal{T}$ is the set of all functions of arity at most 2. Note that $\mathcal{U}$ is a subset of $\mathcal{E}$, $\mathcal{M}$ and $\mathcal{T}$.

A binary function belongs to $\langle \mathcal{U} \rangle$ iff its matrix form is degenerate. A ternary function $F(x_1, x_2, x_3)$

belongs to $\langle \mathcal{T} \rangle$ iff $F^{x_j=U} \cong 0$ for some $1 \le j \le 3$ and some unary $U \not\cong 0$. If furthermore the ternary function $F(x_1, x_2, x_3)$ is symmetric, then the following statements are all equivalent: (1) $F \in \langle \mathcal{T} \rangle$; (2) $F \in \langle \mathcal{U} \rangle$; (3) $F = [a, b]^{\otimes 3}$ for some unary $[a, b]$; and (4) $FU \cong 0$ for some unary $U \not\cong 0$ (take $U = [b, -a]$ if $[a, b] \not\cong 0$, or *any* unary $U \not\cong 0$ if $[a, b] \cong 0$).

Suppose $\mathcal{F}$ is a function set and M is a 2×2 matrix. We use $M\mathcal{F}$ to denote the set consisting of all functions in $\mathcal{F}$ transformed by a matrix M,

$$M\mathcal{F} = \{M^{\otimes r_F} F | F \in \mathcal{F}, r_F = \text{arity}(F)\}.$$

If the transformation matrix M is an orthogonal matrix, then we denote it by H; if M is one of $Z_1 = \begin{pmatrix} 1 & 1 \\ i & -i \end{pmatrix}$ or $Z_2 = \begin{pmatrix} 1 & 1 \\ -i & i \end{pmatrix}$, we denoted it by Z. Note that $(1, \ \pm i)$ is *isotropic*.

The following sets of functions will play a pivotal role: $H\mathcal{E}$, $Z\mathcal{E}$ and $Z\mathcal{M}$. Our main theorem is the following complete classification of the complexity of Holant* problems.

Theorem 2.2. *Let $\mathcal{F}$ be any set of complex valued functions in Boolean variables. The problem Holant$^*(\mathcal{F})$ is polynomial time computable, if (1) $\mathcal{F} \subseteq \langle \mathcal{T} \rangle$, or (2) there exists an orthogonal matrix H such that $\mathcal{F} \subseteq \langle H\mathcal{E} \rangle$, or (3) there exists a matrix $Z \in \{Z_1 = \begin{pmatrix} 1 & 1 \\ i & -i \end{pmatrix}, Z_2 = \begin{pmatrix} 1 & 1 \\ -i & i \end{pmatrix}\}$ such that $\mathcal{F} \subseteq \langle Z\mathcal{E} \rangle$, or (4) there exists a matrix $Z \in \{Z_1, Z_2\}$ such that $\mathcal{F} \subseteq \langle Z\mathcal{M} \rangle$. In all other cases, Holant$^*(\mathcal{F})$ is #P-hard. The dichotomy is still true even if the inputs are restricted to planar graphs.*

3 Tractability

The tractability part is given by the following theorem.

Theorem 3.1. *Holant$^*(\langle \mathcal{T} \rangle)$, Holant$^*(\langle H\mathcal{E} \rangle)$, Holant$^*(\langle Z\mathcal{E} \rangle)$ and Holant$^*(\langle Z\mathcal{M} \rangle)$ are polynomial time computable.*

Proof. By "decoupling" a vertex v into several vertices according to its tensor product factors of the function at v, one can trivially reduce Holant$^*(\langle \mathcal{F} \rangle)$ to Holant$^*(\mathcal{F})$, for any $\mathcal{F}$.

Firstly, to show the tractability of Holant$^*(\mathcal{T})$, we consider any input graph G. G has maximum degree 2, so each connected component is either a path or a cycle. So we only need to compute some m steps of matrix multiplications and trace operations, where m is the number of edges in G. This is clearly polynomial time computable.

Secondly, we prove the tractability of Holant$^*(H\mathcal{E})$. We first reformulate it as a bipartite Holant problem Holant$(=_2 | H\mathcal{E})$ (since $\mathcal{U} = H\mathcal{U} \subset H\mathcal{E}$). Here the edges are replaced by the binary Equality function $(=_2) = [1, 0, 1]$. Now we perform a holographic reduction by the basis transformation H^{-1} on the RHS. This contravariant transformation on the RHS is accompanied by the covariant transformation $[1, 0, 1] \mapsto [1, 0, 1]H^{\otimes 2}$. One can verify that an orthogonal H keeps $[1, 0, 1]$ invariant, namely $[1, 0, 1]H^{\otimes 2} = [1, 0, 1]$. To wit: let $H = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, then

$$\begin{aligned} [1,0,1]H^{\otimes 2} &= \left((1,0)^{\otimes 2} + (0,1)^{\otimes 2}\right) H^{\otimes 2} \\ &= ((1,0)H)^{\otimes 2} + ((0,1)H))^{\otimes 2} \\ &= (a,b)^{\otimes 2} + (c,d)^{\otimes 2} \\ &= (a^2 + c^2, ab + cd, ab + cd, b^2 + d^2) \\ &= (1,0,0,1) = [1,0,1] \end{aligned}$$

Note that unary functions are transformed to unary functions. Hence, after a holographic reduction, our problem becomes Holant$^*(\mathcal{E})$. This is clearly polynomial time computable: If a unary function U

is connected to some $F \in \mathcal{E}$, we may absorb this U and use FU. Note that $FU \in \mathcal{E}$. If a unary U_1 is connected to another unary U_2, then they must form a connected component alone, and its value is trivially computed, which contributes a global factor. After eliminating all unaries, we have an instance of Holant$(\mathcal{E} - \mathcal{U})$, which can be computed on each connected component by uniquely propagating exactly two assignments on an edge. So, Holant$^*(H\mathcal{E})$ is polynomial time computable.

The third class is Holant$^*(Z\mathcal{E})$. Because $\mathcal{U} \subseteq Z\mathcal{E}$, it is a bipartite Holant problem Holant$(=_2 \mid Z\mathcal{E})$. We perform a holographic reduction by the basis transformation Z^{-1} on the RHS. This contravariant transformation on the RHS is accompanied by the covariant transformation $[1,0,1] \mapsto [1,0,1]Z^{\otimes 2} \cong [0,1,0]$. To verify the latter, we have

$$[1,0,1]Z^{\otimes 2} = \left((1,0)^{\otimes 2} + (0,1)^{\otimes 2}\right) Z^{\otimes 2} = ((1,0)Z)^{\otimes 2} + ((0,1)Z))^{\otimes 2} = (1,1)^{\otimes 2} + (i,-i)^{\otimes 2} \cong (0,1,1,0).$$

As an aside, for us in this paper, these holographic transformations demonstrate a main proof technology as well as a tool in the discovery and formulation of our dichotomy theorems. Just as the Equality function $=_2$ can be "factored" by an orthogonal H, and thus "contributes" an orthogonal H to the RHS in this holographic transformation:

$$\text{Holant}(=_2 \mid \mathcal{F}) \longrightarrow \text{Holant}(=_2 \mid H\mathcal{F}),$$

the binary Disequality function $\neq_2$ can be "factored" by $Z = Z_1$ in matrix form (same for $Z = Z_2$)

$$(\neq_2) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \cong Z_1^{\mathtt{T}} Z_1 = \begin{pmatrix} 1 & i \\ 1 & -i \end{pmatrix} \begin{pmatrix} 1 & 1 \\ i & -i \end{pmatrix}$$

and thus "contributes" a Z to the RHS in the following holographic transformation:

$$\text{Holant}(\neq_2 \mid \mathcal{F}) \longrightarrow \text{Holant}(=_2 \mid Z\mathcal{F}).$$

Hence, after a holographic reduction, our problem Holant$^*(Z\mathcal{E})$ becomes Holant$(\{\neq_2\} \mid \mathcal{E})$. (Note $\mathcal{U} \subset \mathcal{E}$.) However $(\neq_2) \in \mathcal{E}$, and thus we have reached a restriction of the tractable Holant$^*(\mathcal{E})$.

Finally we prove the tractability of the fourth class Holant$^*(Z\mathcal{M})$. After a holographic reduction by Z^{-1} on the RHS, it becomes Holant$(\{\neq_2\} \mid \mathcal{M})$. We first eliminate all unary functions as follows. A unary function $[x,y]$ connected with $\neq_2$ is simply another unary function $[y,x]$, which we will replace the pair $[x,y]$ and $\neq_2$. If $F \in \mathcal{M}$ and $U \in \mathcal{U}$, then $FU \in \mathcal{M}$, since the function value of FU on any input with Hamming weight ≥ 2 is certainly 0. A unary connected to another unary forms a trivial connected component and contributes a global factor. Recursively apply these replacement steps until there are no more unary functions left. Hence, we only need to show that Holant$(\{\neq_2\} \mid \mathcal{M} - \mathcal{U})$ is tractable. The input graph is a bipartite graph. Because all functions on the LHS vertex set are $\neq_2$, in order to have a non-zero evaluation, any assignment must have exactly half of edges assigned 0 and the other half of edges assigned 1. All functions on the RHS vertex set are from $\mathcal{M} - \mathcal{U}$. If there is a vertex of degree more than 2 belonging to the RHS vertex set, then this side requires that strictly less than half of edges are 1, so the value of this problem is 0. Thus we only need to calculate on graphs where all vertices have degree 2 (a cycle), which is tractable by matrix multiplication and taking trace. □

We remark that $\langle H\mathcal{E}\rangle$ is a proper generalization of Fibonacci gates $\mathscr{F}$ [8]. Recall that a (symmetric) signature $[f_0, f_1, \ldots, f_k]$ is called a Fibonacci gate of arity k if it satisfies $f_{i+2} = f_{i+1} + f_i$, for $0 \leq i \leq k-2$. Remarkably Holant$^*(\mathscr{F})$ is tractable [8]. E.g., the following counting problem is in P on 3-regular graphs G: Attach at every vertex the signature $[1,0,1,1] \in \mathscr{F}$. Then Holant$(G)$ is the number of edge 2-colorings (Blue or Green) such that every vertex does not have exactly one Blue incident edge.

Let $\phi = \frac{1+\sqrt{5}}{2}$ be the golden ratio, and $\bar{\phi} = \frac{1-\sqrt{5}}{2}$. Then

$$\begin{pmatrix} \frac{1}{\sqrt{1+\phi^2}} & \frac{1}{\sqrt{1+\bar{\phi}^2}} \\ \frac{\phi}{\sqrt{1+\phi^2}} & \frac{\bar{\phi}}{\sqrt{1+\bar{\phi}^2}} \end{pmatrix}^{\otimes k} \left[a \binom{1}{0}^{\otimes k} + b \binom{0}{1}^{\otimes k} \right] = a' \binom{1}{\phi}^{\otimes k} + b' \binom{1}{\bar{\phi}}^{\otimes k}$$

transforms the symmetric signature $a \binom{1}{0}^{\otimes k} + b \binom{0}{1}^{\otimes k} = [a, 0, \ldots, 0, b] \in \mathcal{E}$ to a Fibonacci gate $[f_0, f_1, \ldots, f_k] \in \mathscr{F}$. (Note that the matrix is orthogonal $(1,\ \phi) \cdot (1,\ \bar{\phi}) = 0$. The signature is $f_i = a'\phi^i + b'\bar{\phi}^i$, and satisfies $f_{i+2} = f_{i+1} + f_i$.) The theorem shows a far reaching generalization of Fibonacci gates $\mathscr{F}$ to asymmetric signatures $\langle H\mathcal{E} \rangle$. Our dichotomy theorem will say that this is *the correct* generalization.

4 Outline of the Hardness Proof

Starting from this section, we prove the hardness part of Theorem 2.2, that is, if $\mathcal{F} \not\subseteq \langle \mathcal{T} \rangle$, $\mathcal{F} \not\subseteq \langle H\mathcal{E} \rangle$, $\mathcal{F} \not\subseteq \langle Z\mathcal{E} \rangle$, $\mathcal{F} \not\subseteq \langle Z\mathcal{M} \rangle$, then Holant$^*(\mathcal{F})$ is #P-hard. The proof is quite involved and we give an outline in this section.

The main idea is to reduce the general Holant* problems to the symmetric ones, for which we already have a dichotomy theorem [9]. However, it is not easy to do that when functions have large arities. In Section 6, we first establish an arity reduction theorem. We show that starting from any $F \in \mathcal{F}$ which is not contained in any of the four tractable families $\mathcal{F}'$, we can construct a function Q such that (1) Holant$^*(\mathcal{F} \cup \{Q\}) \equiv_T$ Holant$^*(\mathcal{F})$, (2) $Q \notin \mathcal{F}'$, and (3) Q has a reduced arity. So assuming that the given set of functions is not a subset of any of the four tractable families (otherwise, we are done since it is tractable by Section 3), we can keep on doing arity reductions. This will eventually in a finite number of steps produce the following: In the case of $\langle \mathcal{T} \rangle$, we will end up with an arity 3 signature which is not in $\langle \mathcal{T} \rangle$. For the other three families $\langle H\mathcal{E} \rangle$, $\langle Z\mathcal{E} \rangle$, $\langle Z\mathcal{M} \rangle$, we can get a signature of arity 2 which is not in the respective family.

Having these functions with small arities (2 or 3) in hand, we can construct some simple gadgets to get symmetric functions, which we do in Section 7. The hope is that these symmetric functions are also out of various tractable families. However, we come cross some difficulties by doing this. For example, using a single function of arity 3 which is not in $\langle \mathcal{T} \rangle$, it seems not easy to construct a symmetric arity 3 function which is not in $\langle \mathcal{T} \rangle$ either. In our proof, we get help from other signatures. Namely, we not only use a signature of arity 3 which is not in $\langle \mathcal{T} \rangle$, but also some binary signatures which are not in $\langle H\mathcal{E} \rangle$, $\langle Z\mathcal{E} \rangle$ or $\langle Z\mathcal{M} \rangle$, respectively, to construct a symmetric signature of arity 3 which is not in $\langle \mathcal{T} \rangle$. This is proved in Lemma 7.3. Similarly, in Lemma 7.4, we prove that we can also construct binary symmetric signatures which are not in $\langle H\mathcal{E} \rangle$, $\langle Z\mathcal{E} \rangle$ or $\langle Z\mathcal{M} \rangle$, respectively. Then by the symmetric dichotomy, we know that this ternary signature is either already #P-hard or belongs to $\langle H\mathcal{E} \rangle$, $\langle Z\mathcal{E} \rangle$ or $\langle Z\mathcal{M} \rangle$. If it is #P-hard, then we are done. Otherwise, since we have a binary signature which is not in the same family, we also get the hardness result by the symmetric dichotomy [9]. We note that, all our starting problems for hardness are already hard for planar graphes and all the gadgets we use in the reduction are planar. As a result, our final dichotomy also holds for planar graphes. In the proofs later, we will not explicitly state this every time.

One technical lemma is used extensively in both Section 6 and 7, which substantially simplified the proof. We call it the Separation Lemma, which is stated and proved in Section 5.

5 Separation Lemma

In this section, we introduce a simple lemma which is used frequently in the proofs, and its main purpose is proof simplification. This lemma is applied in the following situation. We have identified a finite set of requirements, the violation of each requirement can be expressed as a system of polynomial equations. Then to show all these requirements can be simultaneously satisfied, we only need to prove each requirement can be individually satisfied, without regard to the consistency of the satisfying variable assignments.

The following lemma is well-known. For completeness we give a proof.

Lemma 5.1. *Suppose $\{P_1, P_2, \ldots, P_m\}$ is a finite set of nonzero polynomials in $\mathbb{F}[x_1, x_2, \ldots, x_n]$, where $\mathbb{F}$ is an infinite field. There exist values $a_1, a_2, \ldots, a_n \in \mathbb{F}$ such that $P_i(a_1, a_2, \ldots, a_n) \neq 0$ for all $1 \leq i \leq m$.*

Proof. For $n = 1$, the conclusion holds obviously.

Suppose the conclusion holds for $n-1$. Let $P_i = \sum_{j=0}^{m_i} p_{i,j}(x_1, \ldots, x_{n-1})x_n^j$. Because P_i is not the zero polynomial, we may assume p_{i,m_i} is a nonzero polynomial in $\mathbb{F}[x_1, x_2, \ldots, x_{n-1}]$. By induction, there exist values $a_1, a_2, \ldots, a_{n-1} \in \mathbb{F}$ such that $p_{i,m_i}(a_1, a_2, \ldots, a_{n-1}) \neq 0$, and $P_i(a_1, a_2, \ldots, a_{n-1}, x_n) \in \mathbb{F}[x_n]$ is a non-zero polynomial in x_n, where $1 \leq i \leq m$. It follows that there exists $a_n \in \mathbb{F}$ such that all $P_i(a_1, a_2, \ldots, a_n) \neq 0$. □

We will give various gadget constructions which use some unary functions $U_k = [x_k, y_k]$, $k = 1, 2, \ldots, m$. Technically the gadget is only defined when specific values for x_k, y_k have been chosen. A signature is expressed as an ordered set of values; this is true for the given constraint functions as well as the signature of the constructed gadget. The entries of the signature of the constructed gadget can be expressed as polynomials in x_k, y_k (the coefficients depend on the given constraint functions). Frequently we have a finite set of conditions, the negation of each condition is expressible as polynomial equations on x_k, y_k. A construction succeeds if we satisfy all these conditions. The following lemma lets us deal with these condition separately.

Lemma 5.2. *Let F be the function realized by a gadget construction, whose arity is n, and each entry of F is a polynomial in x_k and y_k, $k = 1, 2, \ldots, m$.*

Suppose $S_1, S_2, \ldots, S_N$ are sets of functions of arity n, where a function $K \in S_i$ iff the signature entries of K satisfy a finite system of polynomial equations $\{P_{i,1}, P_{i,1}, \ldots, P_{i,m_i}\}$.

Assume for every assignment σ of x_k and y_k, $k = 1, 2, \ldots, m$, there exists a $1 \leq i_\sigma \leq N$, such that $F \in S_{i_\sigma}$ under σ, then there exists a $1 \leq i \leq N$, such that for all σ, $F \in S_i$.

Proof. $F \in S_i$ is expressed by a finite set of polynomial equations $\{P_{i,1}, P_{i,1}, \ldots, P_{i,m_i}\}$ on x_k and y_k. If the conclusion of the lemma is false, then for all $1 \leq i \leq N$, there exists a polynomial $P_{i,s(i)}$ which is not the zero polynomial. By Lemma 5.1, there exist values of x_k and y_k, such that $P_{i,s(i)} \neq 0$ for all $1 \leq i \leq N$, contradicting the condition. □

The following lemma is another direct corollary of lemma 5.1.

Lemma 5.3. *Suppose a gadget construction using unary functions $U_k = [x_k, y_k]$, $k = 1, 2, \ldots, m$ succeeds if it satisfies a finite set of properties R_i, $i = 1, 2, \ldots, N$. Suppose violation of each property R_i is specified by a finite set of polynomial equations. If for each i we can find unary functions $U_k = [x_k, y_k]$ to satisfy property R_i, then we can find unary functions $U_k = [x_k, y_k]$ so that the construction succeeds.*

We call it the Separation Lemma in the proofs.

6 Arity reduction

The next two sections prove the hardness part of Theorem 2.2, that is, if $\mathcal{F} \not\subseteq \langle \mathcal{T} \rangle$, $\mathcal{F} \not\subseteq \langle H\mathcal{E} \rangle$, $\mathcal{F} \not\subseteq \langle Z\mathcal{E} \rangle$, $\mathcal{F} \not\subseteq \langle Z\mathcal{M} \rangle$, then Holant$^*(\mathcal{F})$ is #P-hard.

In this section, we show that starting from any $F \in \mathcal{F}$ which is not contained in any of the four tractable families $\mathcal{F}'$, we can construct a function Q such that (1) Holant$^*(\mathcal{F} \cup \{Q\}) \equiv_T$ Holant$^*(\mathcal{F})$, (2) $Q \notin \mathcal{F}'$, and (3) Q has a reduced arity.

Lemma 6.1. *Let $\mathcal{F}'$ be any one of $\langle \mathcal{T} \rangle$, or $\langle H\mathcal{E} \rangle$, or $\langle Z\mathcal{E} \rangle$, or $\langle Z\mathcal{M} \rangle$. Let $r = 3$ if $\mathcal{F}' = \langle \mathcal{T} \rangle$, and $r = 2$ in the other three cases. Suppose function $F \in \mathcal{F} - \mathcal{F}'$. If $r < \text{arity}(F)$, then we can realize a function Q by connecting F with some unary functions, such that (1) Holant*$(\mathcal{F} \cup \{Q\}) \equiv_T$ Holant*$(\mathcal{F})$; (2) $Q \notin \mathcal{F}'$ and (3) $r \leq \text{arity}(Q) < \text{arity}(F)$.*

The proof of this lemma is divided into following several lemmas. Firstly, we show that any *type specification* in a tensor product decomposition can be described by a system of polynomial equations.

Lemma 6.2. *For any type specification $\mathcal{I}$, there is a finite set of polynomial equations $E_{\mathcal{I}}$ in the entries of a signature F, such that F has type $\mathcal{I}$ iff F satisfies $E_{\mathcal{I}}$.*

Proof. If $\mathcal{I} = \{[n]\}$, there is no requirement on F to have type $\mathcal{I}$. We can use a trivial equation such as $0 = 0$.

Consider the case $\mathcal{I} = \{I_1, I_2\}$. Suppose F has type $\mathcal{I}$. Then obviously, for any two values a_1, b_1 of X_1 and any two values a_2, b_2 of X_2, $F(a_1, a_2)F(b_1, b_2) = F(a_1, b_2)F(b_1, a_2)$ (W.o.l.o.g., we assume the indices in I_1 all precede those of I_2.) Hence the collection of all these equations $E_{\{I_1,I_2\}}$ is a necessary condition that F has type $\mathcal{I}$. It is also a sufficient condition. Arrange the values of F into a matrix $F_{X_1,X_2} = F(X_1, X_2)$, where the row indices (resp. column indices) are all possible values of X_1 (resp. X_2). The condition $F(a_1, a_2)F(b_1, b_2) = F(a_1, b_2)F(b_1, a_2)$ for all a_1, b_1 and all a_2, b_2 implies that any 2 by 2 submatrix of F is singular, and so $\text{rank}(F_{X_1,X_2}) \leq 1$. Hence, F_{X_1,X_2} is the product of a column vector and a row vector, and F has type $\mathcal{I}$.

Now consider a general partition $\mathcal{I} = \{I_1, \ldots, I_k\}$, and again suppose F has type $\mathcal{I}$. It follows that for any $1 < i \leq k$, any fixed values $a_{i+1}, \ldots, a_k$ for $X_{i+1}, \ldots, X_k$, $F^{X_{i+1}=a_{i+1},\ldots,X_k=a_k}$ has type $\{\bigcup_{j=1}^{i-1} I_j, I_i\}$. We define the following set of equations: $\forall 1 < i \leq k$, and $\forall$ assignments $a_{i+1}, \ldots, a_k$ for $X_{i+1}, \ldots, X_k$, include the equations in $E_{\{\bigcup_{j=1}^{i-1} I_j, I_i\}}$. This is a finite set of polynomial equations. Obviously, this is a necessary condition for F has type $\mathcal{I}$.

We prove that it is also a sufficient condition. If F is the zero function, then F has type $\mathcal{I}$ trivially. Assume F is not the zero function. Let $i = k$, by what has been proved when $k = 2$, $F = \bigotimes_{\{\bigcup_{j=1}^{k-1} I_j, I_k\}}(P_{k-1}, F_k)$. Because F is not the zero function, there exists a value a_k for X_k such that $F_k(a_k) \neq 0$. The remaining conditions, for $1 < i \leq k-1$, yield a finite set of homogeneous equations for $F^{X_k=a_k} = P_{k-1}F_k(a_k)$. After canceling the non-zero factor $F_k(a_k)$, by induction, we obtain the necessary and sufficient conditions that P_{k-1} has type $\{I_1, \ldots, I_{k-1}\}$. Hence F has type $\mathcal{I}$. ∎

Next, we prove a property of this decomposition. This property is used throughout in the proof of Lemma 6.1.

Lemma 6.3. *Suppose there exists some type $\mathcal{I} = \{I_1, \ldots, I_k\}$ over $[n-1]$, such that for all unary functions $U = [x, y]$, $FU = xF^{x_n=0} + yF^{x_n=1}$ has the same type $\mathcal{I}$. Furthermore, suppose $F^{x_n=0} = \bigotimes_{\mathcal{I}}(F_1, F_2, \ldots, F_k)$ and $F^{x_n=1} = \bigotimes_{\mathcal{I}}(K_1, K_2, \ldots, K_k)$ are linearly independent as two vectors. Then there exists at most one index $i \in [k]$ such that F_i, K_i are linearly independent.*

Proof. For a contradiction, suppose there are two distinct indices $i \in [k]$, with the property that F_i, K_i are linearly independent. W.o.l.o.g., let $i = 1, 2$ respectively.

Because $F^{x_n=0}, F^{x_n=1}$ are linearly independent, F_j and K_j are not the zero function for any $j \in [k]$. For any $j \in [k] - \{1, 2\}$, by Lemma 5.2, there exist $|I_j|$ unary functions such that both F_j and K_j become nonzero constants when combined with them. After combining $F^{x_n=0}$ and $F^{x_n=1}$ with these unary functions, we obtain respectively the functions $c_0 F_1 \otimes F_2$ and $c_1 K_1 \otimes K_2$ over the variables in $I_1 \cup I_2$, where $c_0, c_1 \neq 0$.

Suppose $U = [x, y]$ and $xy \neq 0$. If we combine $FU = xF^{x_n=0} + yF^{x_n=1}$ with the same set of $\left|\bigcup_{j=3}^{k} I_j\right|$ many unary functions, the resulting function is $c_0 x F_1 \otimes F_2 + c_1 y K_1 \otimes K_2$. By the assumption on FU having the same type $\mathcal{I}$, this function has type $\{I_1, I_2\}$. However, we will show that, for any $xy \neq 0$, this function does not have type $\{I_1, I_2\}$. The matrix form (row index is $X|_{I_1}$, column index is $X|_{I_2}$) of this function is $\begin{pmatrix} F_1 & K_1 \end{pmatrix} \begin{pmatrix} c_0 x & 0 \\ 0 & c_1 y \end{pmatrix} \begin{pmatrix} F_2^{\mathrm{T}} \\ K_2^{\mathrm{T}} \end{pmatrix}$. Since F_1, K_1 are linearly independent, and F_2, K_2 are linearly independent, this matrix has rank two. If this function has type $\{I_1, I_2\}$, its matrix form would have rank at most one. This contradiction proves the Lemma. $\square$

Proof of Lemma 6.1: $\mathcal{F}' = \langle \mathcal{T} \rangle$

Suppose $F \in \mathcal{F} - \langle \mathcal{T} \rangle$, with arity$(F) > 3$. Being out of $\langle \mathcal{T} \rangle$, F is not the zero function. If for some unary function $U = [x, y]$, $FU \notin \langle \mathcal{T} \rangle$, then we are done by setting $Q = FU$. Hence we assume for any unary function $U = [x, y]$, $FU = xF^{x_n=0} + yF^{x_n=1}$ has some type $\mathcal{J}$, where each set $J_j \in \mathcal{J}$ has size at most 2. For the fixed arity(F), there are only finitely many such types, which are specifiable by a finite set of polynomial equations in x, y. It follows from Lemma 5.2 that, there exists some type $\mathcal{I} = \{I_1, \ldots, I_k\}$, where each $|I_j| \leq 2$, such that for all x, y, FU has the same type $\mathcal{I}$. In particular, both $F^{x_n=0}$ and $F^{x_n=1}$ have type $\mathcal{I}$.

If $F^{x_n=0}$ and $F^{x_n=1}$ are linearly dependent, then $F \in \langle \mathcal{T} \rangle$, having type $\mathcal{I} \cup \{\{n\}\}$.

So $F^{x_n=0} = \bigotimes_{\mathcal{I}}(F_1, F_2, \ldots, F_k)$ and $F^{x_n=1} = \bigotimes_{\mathcal{I}}(K_1, K_2, \ldots, K_k)$ are linearly independent. Being linearly independent, none of the tensor factors of $F^{x_n=0}$ and $F^{x_n=1}$ can be the zero function. By Lemma 6.3, there is at most one pair of linearly independent tensor factors, w.o.l.o.g., F_1 and K_1. Expressing K_i in terms F_i, for $i \geq 2$, there exists a nonzero constant c, such that $F^{x_n=1} = \bigotimes_{\mathcal{I}}(cK_1, F_2, \ldots, F_k)$. If $|I_1| = 1$, that is, F_1 and K_1 are unary functions, then $F \in \langle \mathcal{T} \rangle$, of type $\{I_1 \cup \{n\}, I_2, \ldots, I_k\}$. Thus, $|I_1| = 2$.

W.o.l.o.g., assume $I_1 = \{1, 2\}$. We can fix the variables of F in $I_2, \ldots, I_k$ to some values, such that $F_2, \ldots, F_k$ each contributes a nonzero factor. We get a ternary function Q in variables x_1, x_2, x_n. $F = Q \otimes F_2 \otimes \cdots \otimes F_k$. Then we claim that $Q \notin \langle \mathcal{T} \rangle$, for otherwise, if $Q \in \langle \mathcal{T} \rangle$, then $F \in \langle \mathcal{T} \rangle$.

Proof of Lemma 6.1: $\mathcal{F}' = \langle H\mathcal{E} \rangle$ **and** $\langle Z\mathcal{E} \rangle$

For any function F and invertible matrix M, $F \in \langle M\mathcal{E} \rangle$ iff $(M^{-1})^{\otimes n} F \in \langle \mathcal{E} \rangle$. (Note that $\langle M\mathcal{E} \rangle = M\langle \mathcal{E} \rangle$.) Hence we only need to prove for $\langle \mathcal{E} \rangle$. Suppose $F \notin \langle \mathcal{E} \rangle$, and arity$(F) = n > 2$. F is not the zero function. If for some unary function $U = [x, y]$, $FU \notin \langle \mathcal{E} \rangle$, we are done with $Q = FU$. Hence we assume for any unary function $U = [x, y]$, $FU = xF^{x_n=0} + yF^{x_n=1} \in \langle \mathcal{E} \rangle$.

For any partition $\mathcal{I} = \{I_1, \ldots, I_k\}$ of $[n]$, and any $\mathcal{A} = \{A_1, \ldots, A_k\}$, such that $A_j \in \{0, 1\}^{|I_j|}$, we define a set of functions $(\mathcal{I}, \mathcal{A})$. Each A_j is a 0-1 string of length $|I_j|$. What we use in the definition is the set $\{A_j, \bar{A}_j\}$, so we may normalize the first bit of A_j to be 0. A function P belongs to the set $(\mathcal{I}, \mathcal{A})$, iff P has type $\mathcal{I}$, that is, $P = \bigotimes_{\mathcal{I}}(P_1, P_2, \ldots, P_k)$, and for any $j \in [k]$, $P_j(X|_{I_j})$ is zero if $X|_{I_j} \notin \{A_j, \bar{A}_j\}$. Thus, $P_j \in \mathcal{E}$ for each $j \in [k]$.

Functions in $\langle \mathcal{E} \rangle$ of arity n is the union of these finitely many function sets $(\mathcal{I}, \mathcal{A})$. Obviously, functions in $(\mathcal{I}, \mathcal{A})$ can be described by a finite system of polynomial equations (Lemma 6.2). Since $FU \in \langle \mathcal{E} \rangle$ for all $U = [x, y]$, by Lemma 5.2, there must exist one $(\mathcal{I}, \mathcal{A})$, such that for any x, y, FU belongs to the same set $(\mathcal{I}, \mathcal{A})$.

If $F^{x_n=0}$ and $F^{x_n=1}$ are linearly dependent, then obviously, $F \in \langle \mathcal{E} \rangle$.

Let $F^{x_n=0} = \bigotimes_{\mathcal{I}}(F_1, F_2, \ldots, F_k)$ and $F^{x_n=1} = \bigotimes_{\mathcal{I}}(K_1, K_2, \ldots, K_k)$ be linearly independent. Being linearly independent, none of the tensor factors of $F^{x_n=0}$ and $F^{x_n=1}$ can be the zero function. By Lemma 6.3, there is at most one pair of linearly independent tensor factors, w.o.l.o.g., F_1 and K_1. They must be indeed linearly independent, otherwise $F^{x_n=0}$ and $F^{x_n=1}$ are also linearly independent. Expressing K_i in terms F_i, for $i \geq 2$, there exists a nonzero constant c, such that $F^{x_n=1} = \bigotimes_{\mathcal{I}}(cK_1, F_2, \ldots, F_k)$.

We can fix the variables of F in $I_2, \ldots, I_k$ to some values, such that $F_2, \ldots, F_k$ each contributes a nonzero factor. We obtain a function K. $K^{x_n=0} = F_1$ and $K^{x_n=1} = cK_1$, where $c \neq 0$. K evaluates to zero, except on possibly four inputs $\{A_1 0, \bar{A}_1 0, A_1 1, \bar{A}_1 1\}$. Combine the middle $|I_1| - 1$ variables of K (that is, except x_n and the first variable in I_1) with the function $[1, 1]$, we get a binary function in matrix form $Q = \begin{pmatrix} K(A_1 0) & K(A_1 1) \\ K(\bar{A}_1 0) & K(\bar{A}_1 1) \end{pmatrix}$, where we index the row by the first variable in I_1 and the column by x_n. Note that we have used the definition of $(\mathcal{I}, \mathcal{A})$. Because F_1 and K_1 are linearly independent, Q is nonsingular. We claim this $Q \notin \langle \mathcal{E} \rangle$. For otherwise, being non-degenerate, $Q \in \mathcal{E}$, and this implies that $K \in \mathcal{E}$, and hence $F = K \otimes F_2 \otimes \cdots \otimes F_k \in \langle \mathcal{E} \rangle$.

Proof of Lemma 6.1: $\mathcal{F}' = \langle Z\mathcal{M} \rangle$

Again we only need to prove for $\langle \mathcal{M} \rangle$. Suppose $F \notin \langle \mathcal{M} \rangle$, and arity$(F) = n > 2$. Again we may assume for any unary function $U = [x, y]$, $FU = xF^{x_n=0} + yF^{x_n=1} \in \langle \mathcal{M} \rangle$; otherwise, we are done.

For any $U = [x, y]$, $FU \in \langle \mathcal{M} \rangle$ has some type $\mathcal{I}$, and each tensor factor belongs to $\mathcal{M}$, that is, it is zero except on inputs of Hamming weight at most one. Each type $\mathcal{I}$ can be specified by a system of polynomial equations (Lemma 6.2). The requirement associated with each type $\mathcal{I}$ consists of these polynomial equations together with the zero requirements on all entries of each tensor factor whose Hamming weight is greater than one. (For example, if we require that $(a_{i,j})$ is a tensor product $(b_i) \otimes (c_j)$, where $1 \leq i \leq n, 1 \leq j \leq m$, and for some subset $B \subseteq [n]$, $C \subseteq [m]$, $[\forall i \in B, \forall j \in C, b_i = c_j = 0]$, then we include the equations from Lemma 6.2 together with $[a_{i,j} = 0, \forall i \in B, \textit{ or } \forall j \in C]$. The equations from Lemma 6.2 implies that a tensor factorization $(a_{i,j}) = (b_i) \otimes (c_j)$ exists. If for all $i \in [n]$, $b_i = 0$, then $a_{i,j}$ is identically 0. Similarly if for all $j \in [m]$, $c_j = 0$. On the other hand, if for some $i_0 \in [n]$, and $j_0 \in [m]$, $b_{i_0} \neq 0$ and $c_{j_0} \neq 0$, Then $b_i = a_{i,j_0}/c_{j_0} = 0$ for all $i \in B$. Similarly $c_j = 0$ for all $j \in C$.)

Applying Lemma 5.2, we conclude that there is one type $\mathcal{I}$ such that for any x, y, FU has the same decomposition associated with $\mathcal{I}$ with tensor factors from $\mathcal{M}$.

If $F^{x_n=0}$ and $F^{x_n=1}$ are linearly dependent, obviously, $F \in \langle \mathcal{M} \rangle$.

Let $F^{x_n=0} = \bigotimes_{\mathcal{I}}(F_1, F_2, \ldots, F_k)$ and $F^{x_n=1} = \bigotimes_{\mathcal{I}}(K_1, K_2, \ldots, K_k)$ be linearly independent. As before none of the tensor factors of $F^{x_n=0}$ and $F^{x_n=1}$ can be the zero function, and exactly one pair among F_i and K_i are linearly independent, w.o.l.o.g., F_1 and K_1. We can fix the variables of F in $I_2, \ldots, I_k$ to some values, such that $F_2, \ldots, F_k$ contribute a nonzero factor. We get a function in matrix form $K = \begin{pmatrix} F_1^{\mathtt{T}} \\ cK_1^{\mathtt{T}} \end{pmatrix}$, where $c \neq 0$. Here the first row is $K^{x_n=0} = F_1^{\mathtt{T}}$. The second row is $K^{x_n=1} = cK_1^{\mathtt{T}}$. Columns are indexed by $A \in \{0,1\}^{|I_1|}$. If the weight of A is greater than 1, then the Ath column is zero. Let S_0 denote $(0 \cdots 0)$th column, and S_i denote the Ath column, where only the ith bit of A is 1.

For simplicity of notations, assume $I_1 = \{1, 2, \ldots, m\}$. There exists a 0-1 string $A \in \{0,1\}^m$ of Hamming weight 1, such that $K_1(A) \neq 0$; otherwise, $K \in \mathcal{M}$ and $F \in \langle \mathcal{M} \rangle$. That is, there is a column S_i, $1 \leq i \leq m$, whose second entry is not zero. W.l.o.g., we assume this S_i is S_m. Because F_1 and K_1 are linearly independent, There exists a column S_j linearly independent with the nonzero column S_m. If S_0 is such a column, let $Q = K^{x_1=0,\ldots,x_{m-1}=0} = (S_0, S_m)$, since the index for the column S_m is $A = (0 \cdots 01) \in \{0,1\}^m$. Otherwise, S_0 is linearly dependent with S_m. Then for some $1 \leq j \leq m-1$, S_j is linearly independent with S_m. Let $Q = K^{x_1=0,\ldots,x_{j-1}=0,x_j=[x,y],x_{j-1}=0,\ldots,x_{n-2}=0} = (xS_0 + yS_j, xS_m)$, where $x \neq 0$ and $y \neq 0$. We have obtained our Q such that Q is not degenerate and $Q(1,1) \neq 0$, i.e., $Q \notin \langle \mathcal{M} \rangle$.

7 From asymmetric to symmetric

In this section, we show how to get a symmetric function from some asymmetric functions, keeping the property that F does not belong to any one of the four tractable classes, $\langle\mathcal{T}\rangle$, $\langle H\mathcal{E}\rangle$, $\langle Z\mathcal{E}\rangle$ or $\langle Z\mathcal{M}\rangle$.

Suppose $c \in \mathbb{C}$ is nonzero. We may consider functions F and cF being the same function, i.e., $F \cong cF$. In the following lemma, when we count the number of solutions, we count in terms of equivalence classes under $\cong$ (i.e., we count in projective space.)

Lemma 7.1. *For any ternary function F, $F^{x_3=U} \cong 0$ for some $U \ncong 0$ iff F has type $\{\{1,2\},\{3\}\}$.*

Let a ternary function $F \notin \langle\mathcal{T}\rangle$. Then $F^{x_3=U} \ncong 0$ for any nonzero unary function U, and there exist exactly one or two nonzero $U = [x, y]$ such that $F^{x_3=U}$ is degenerate.

Proof. If F has type $\{\{1,2\},\{3\}\}$, $F = T \otimes [a,b]$. If $a = b = 0$ then F is identically 0, and $F^{x_3=U} \cong 0$ for any unary U. If $[a,b] \ncong 0$, then $F^{x_3=U} \cong 0$ for $U = [b,-a] \ncong 0$. Conversely, if $F^{x_3=U} \cong 0$ for some $U \ncong 0$, then $F^{x_3=0}$ and $F^{x_3=1}$ are linearly dependent, and hence F has type $\{\{1,2\},\{3\}\}$. It follows that if $F \notin \langle\mathcal{T}\rangle$, then $F^{x_3=U} \ncong 0$ for any nonzero unary function U.

$F^{x_3=U}$ is degenerate iff $FU(0,0)FU(1,1) = FU(0,1)FU(1,0)$. Let $U = [x,y]$, then the entries of FU are linear homogeneous polynomials of x, y, so the equation is a quadratic homogeneous equation. Either it has one or two solutions, or it is identically zero and all x, y are solutions. We only need to prove the latter case contradicts $F \notin \langle\mathcal{T}\rangle$.

Suppose FU is degenerate for all U, then in particular $F[1,0] = F_1 \otimes F_2$ and $F[0,1] = K_1 \otimes K_2$. If $F[1,0]$ and $F[0,1]$ are linearly dependent, then $F \in \langle\mathcal{T}\rangle$. If $F[1,0]$ and $F[0,1]$ are linearly independent, then by Lemma 6.3, at most one of the two pairs functions F_1, K_1 and F_2, K_2 are linearly independent. W.o.l.o.g., suppose F_2, K_2 are linearly dependent. Then F is the tensor product of F_2 and one binary function, so $F \in \langle\mathcal{T}\rangle$. ☐

For any ternary function $F(x_1, x_2, x_3) \notin \langle\mathcal{T}\rangle$, the conclusion of Lemma 7.1 certainly applies to all three variables. There is a simple relationship, among $1 \le i \le 3$, between the nonzero unary functions U_i such that $F^{x_i=U_i}$ is degenerate. Suppose $F^{x_1=U_1}$ is degenerate, where $U_1 \ncong 0$, then $F^{x_1=U_1} = L \otimes R$, where L and R are unary functions on x_2 and x_3 respectively. Since $F^{x_1=U_1} \ncong 0$, both L and $R \ncong 0$. It follows that the decomposition $L \otimes R$ is unique under $\cong$. If we define $[x,y]^{\perp} = [y,-x]$, then $F^{x_1=U_1, x_2=U_2}$ is identically 0, where $U_2 = L^{\perp}$. Thus $F^{x_2=U_2}$ is degenerate. Similarly for x_3. This mapping from $U_1 \mapsto U_2 = L^{\perp}$ is well-defined under $\cong$. It is also 1-1: Suppose $F^{x_1=U_1, x_2=U_2}$ and $F^{x_1=U_1', x_2=U_2}$ are both identically 0. Then $F^{x_2=U_2}$ is degenerate, and expressible as $A(x_1) \otimes B(x_3)$, where A and B are nonzero unary functions. It follows that both $U_1 \cong A^{\perp}$ and $U_1' \cong A^{\perp}$. Thus $U_1 \cong U_1'$. By symmetry the inverse map is also well-defined.

We summarize this in the following lemma. Suppose a ternary function $F \notin \langle\mathcal{T}\rangle$. Let

$$\mathcal{U}_i = \{U \mid F^{x_i=U} \text{ is degenerate}\}$$

for $1 \le i \le 3$.

Lemma 7.2. *There is a one-to-one correspondence between $\mathcal{U}_1$, $\mathcal{U}_2$, and $\mathcal{U}_3$, as follows. For $\{i,j,k\} = \{1,2,3\}$, each $U_i \in \mathcal{U}_i$ gives a unique factorization $F^{x_i=U_i} = V_j(x_j) \otimes V_k(x_k)$, where $V_j^{\perp} \in \mathcal{U}_j$ and $V_k^{\perp} \in \mathcal{U}_k$. In particular $|\mathcal{U}_1| = |\mathcal{U}_2| = |\mathcal{U}_3| = 1$ or 2.*

Now we will prove a crucial lemma for the hardness part of Theorem 2.2.

Lemma 7.3. *Suppose in Holant*$^*(\mathcal{F})$*, we can realize the following functions*

1. *$F \notin \langle\mathcal{T}\rangle$ of arity 3;*

2. For any orthogonal matrix H, some $P_H \notin \langle H\mathcal{E}\rangle$ of arity 2;

3. For both $Z = Z_1$ or Z_2, some $P_Z \notin \langle Z\mathcal{E}\rangle$ of arity 2; and

4. For both $Z = Z_1$ or Z_2, some $S_Z \notin \langle Z\mathcal{M}\rangle$ of arity 2.

Then we can realize a symmetric *ternary function $Q \notin \langle \mathcal{T}\rangle$ in Holant*$(\mathcal{F})$.*

Proof. We use the gadget shown in Figure 4 to realize a symmetric ternary function Q. (In some cases we will need to modify it to define Q; this will be discussed later.) This gadget consists of 9 copies of the function F, 3 copies of a unary function U_1 and 3 copies of a unary function U_2. The unary functions are to be determined later. Each shaded triangle labeled with F in a central inner triangle represents the function $F(x_1, x_2, x_3) \notin \langle \mathcal{T}\rangle$. The labels $1, 2, 3$ inside the shaded triangle indicate which edge corresponds to variables x_1, x_2, x_3. This gadget remain unchanged if we rotate it $\frac{2\pi}{3}$ or $\frac{4\pi}{3}$. Hence, $Q(x_1, x_2, x_3) = Q(x_2, x_3, x_1) = Q(x_3, x_1, x_2)$. It follows that Q is symmetric (notice that this conclusion uses the fact that each variable x_i is a Boolean variable).

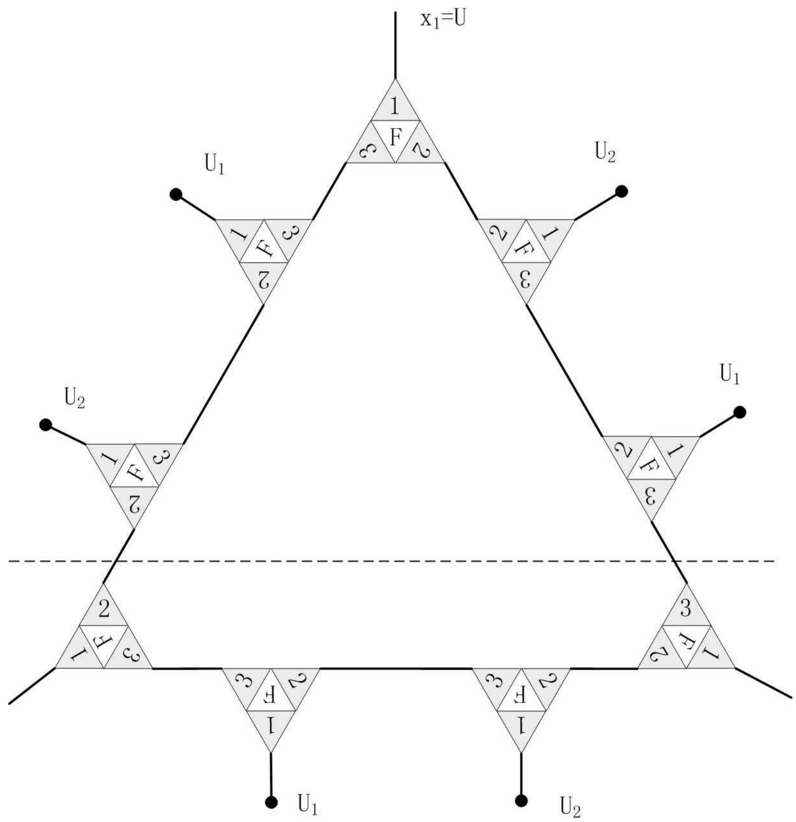

Figure 4: Gadget to realize a symmetric ternary function.

Our goal is to prove that there exist nonzero unary functions U_1 and U_2, such that $Q \notin \langle \mathcal{T}\rangle$. Since Q is symmetric, this is equivalent to: there exists no nonzero unary function U satisfying $Q^{x_1=U} \cong 0$, by Lemma 7.1.

To prove this, we divide the gadget into two parts, as shown by the dashed line in Figure 4. We establish two properties, one property for each part respectively. The upper part is a ternary function, denoted by S. The first property is that if $U \ncong 0$, then $S^{x_1=U} \ncong 0$. The matrix form of $S^{x_1=U}$ is the matrix product $F^{x_1=U_2}_{x_2,x_3}F^{x_1=U_1}_{x_2,x_3}F^{x_1=U}_{x_3,x_2}F^{x_1=U_2}_{x_2,x_3}F^{x_1=U_1}_{x_2,x_3}$, where $F^{x_1=*}_{x_2,x_3}$ denotes the matrix form of $F^{x_1=*}$ with row index x_2 and column index x_3. Because $F \notin \langle \mathcal{T} \rangle$, if $U \not\cong 0$, then $F^{x_1=U} \not\cong 0$, by Lemma 7.1. To satisfy this property on S, we only need some U_1 and U_2 such that $F^{x_1=U_1}$ and $F^{x_1=U_2}$ are nonsingular. By Lemma 7.1, there exist such U_1 and U_2.

The lower part is a function of arity 4, denoted by P. Two inputs of P are the original inputs x_2, x_3 of Q, corresponding to the lower left and lower right corders of the gadget respectively. The other two inputs correspond to edges connecting P with S, denoted by y_2, y_3 respectively. The second property is that the 4×4 matrix $P_{x_2x_3,y_2y_3}$ is nonsingular.

If there exist U_1 and U_2 such that both properties hold, then for any nonzero function U, the vector form of $Q^{x_1=U}$ is $P_{x_2x_3,y_2y_3}S^{x_1=U}$, where $S^{x_1=U}$ takes its vector form as a vector of dimension 4. Hence $Q^{x_1=U}$ is not the zero function, because $S^{x_1=U}$ is a nonzero column vector (the first property) and $P_{x_2x_3,y_2y_3}$ is a nonsingular matrix (the second property). This proves $Q \notin \langle \mathcal{T} \rangle$.

To establish the two properties, we can apply the Separation Lemma 5.3, and prove the two properties individually. We have proved the first one. Now we prove the second one. (The Separation Lemma allows us to choose unary functions U_1 and U_2 separately for the two parts in order to satisfy the two properties, even though in the actual gadget construction the 3 occurrences of U_1 must be the same, and similarly for U_2, in order to produce a symmetric Q.)

The idea for the proof of the second property on P will be counter intuitive. Our goal is to choose unary functions U_1 and U_2 such that the function P has a full-rank matrix. We will do this by a nonzero unary function U_1 such that $F^{x_1=U_1}$ has a *singular* matrix. (This could be surprising as we seem to go the opposite direction.) However once $F^{x_1=U_1}$ is degenerate, this effectively severs the bottom path in this gadget P. (This entanglement on the bottom makes it difficult to analyze P.) Consequently the matrix $P_{x_2x_3,y_2y_3}$ become a tensor product of two matrices $A_{x_2,y_2} \otimes B_{x_3,y_3}$. We then aim to guarantee that both A_{x_2,y_2} and B_{x_3,y_3} are nonsingular 2×2 matrices.

Since $F \notin \langle \mathcal{T} \rangle$, by Lemma 7.1 there exists a nonzero U_1 such that $F^{x_1=U_1}$ is degenerate, and $F^{x_1=U_1} = L_L \otimes R_L$, or in more detail, $F^{x_1=U_1}(z_3, z_2) = L_L(z_3)R_L(z_2)$. L_L and R_L are not the zero function. We also want the matrix form A_{x_2,y_2} of $F^{x_3=L_L}$ to be nonsingular. In the notation of Lemma 7.2, by the one-to-one correspondence from $\mathcal{U}_1$ to $\mathcal{U}_3$, $U_1 \in \mathcal{U}_1$ gives L_L and then gives a corresponding $L_L^{\perp} \in \mathcal{U}_3$. Thus we want some unary $U \in \mathcal{U}_3$, such that $U^{\perp} = L_L \notin \mathcal{U}_3$. Such a $U \in \mathcal{U}_3$, by the inverse map of the one-to-one correspondence, gives us the desired $U_1 \in \mathcal{U}_1$.

We have the similar requirement for U_2 and $B = F^{x_2=R_R}$, on the right half of the gadget P: $U_2 \in \mathcal{U}_1$, $F^{x_1=U_2} = L_R \otimes R_R$, and $R_R \notin \mathcal{U}_2$. In fact, writing in matrix form for P from left to right (Figure 4), $P = F_{x_1x_2,x_3}(L_L R_L^{\mathtt{T}})(L_R R_R^{\mathtt{T}})F_{x_2,x_1x_3}$. Taking out the inner product value $R_L^{\mathtt{T}}L_R$ (a scalar), the remainder of the function can be written as $F^{x_3=L_L} \otimes F^{x_2=R_R}$. Now writing P in the matrix form with rows indexed by the original inputs x_2, x_3 of Q and columns indexed by the edges y_2, y_3 connecting P to S, the 4×4 matrix $P_{x_2x_3,y_2y_3}$ is $(R_L^{\mathtt{T}}L_R)A_{x_2,y_2} \otimes B_{x_3,y_3}$, where A_{x_2,y_2} and B_{x_3,y_3} are the matrix form for $F^{x_3=L_L}$ and $F^{x_2=R_R}$. (Figure 5). So we also require $R_L^{\mathtt{T}}L_R \neq 0$.

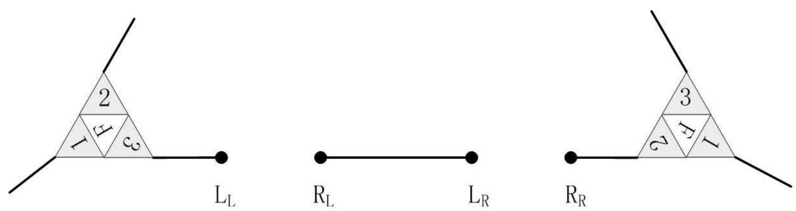

Figure 5: Replace $F^{x_1=U_1}$ by $L_L \otimes R_L$, and $F^{x_1=U_2}$ by $L_R \otimes R_R$.

To summarize for P, for the second property, we identify three conditions whose conjunction is sufficient. Condition (1): $F^{x_1=U_1} = L_L \otimes R_L$ is degenerate and $F^{x_3=L_L}$ is nondegenerate. Condition (2): $F^{x_1=U_2} = L_R \otimes R_R$ is degenerate and $F^{x_2=R_R}$ is nondegenerate. Condition (3): $R_L^{\mathtt{T}} L_R \neq 0$.

There are three cases, depending on $\mathcal{U}_3$, where one cannot pick U_1, U_2 to satisfy Condition (1).

a. $|\mathcal{U}_3| = 1$ and for the unique $U \in \mathcal{U}_3$, it also holds that $U^\perp \in \mathcal{U}_3$.

b. $|\mathcal{U}_3| = 2$ and for both $U \in \mathcal{U}_3$, it also holds that $U^\perp = U \in \mathcal{U}_3$.

c. $|\mathcal{U}_3| = 2$ and $\mathcal{U}_3 = \{U, U^\perp\}$.

In case (a.): $U^\perp \cong U$, and thus $U \cong [1, i]$ or $[1, -i]$.

If $U \cong [1, i]$ (resp. $[1, -i]$), we show that $S_{Z_1}U$ (resp. $S_{Z_2}U$) does not belong to $\mathcal{U}_3$. Because the binary $S_{Z_1} \notin \langle Z_1 \mathcal{M} \rangle$ in matrix form $S_{Z_1} = Z_1 T Z_1^{\mathtt{T}}$ for some $T \notin \mathcal{M}$, that is $T(1,1) \neq 0$. Then $U^{\mathtt{T}} S_{Z_1} U = U^{\mathtt{T}} Z_1 T Z_1^{\mathtt{T}} U \cong (0,1) T \begin{pmatrix} 0 \\ 1 \end{pmatrix} = T(1,1) \neq 0$. Hence, $S_{Z_1}U \not\cong [1, i]$, and $S_{Z_1}U \not\cong [0, 0]$. In this case, $F^{x_3 = S_{Z_1}U}$ is nondegenerate, and we will modify the construction in Figure 4 by adding the binary gadget with signature S_{Z_1} to replace the three edges whose endpoints are triangles marked by 3 in the gadget. (This change does not affect what has been proved for S, since S_{Z_1} is nondegenerate. The same is true for case (b.) and (c.)) The proof for $S_{Z_2}U$ is similar.

In case (b.), $\mathcal{U}_3 = \{[1, i], [1, -i]\}$.

We show in this case, one of $P_{Z_1}[1, i]$ or $P_{Z_1}[1, -i] \notin \mathcal{U}_3$. Because $P_{Z_1} \notin \langle Z_1 \mathcal{E} \rangle$, in matrix form $P_{Z_1} = Z_1 T Z_1^{\mathtt{T}}$ for some $T \notin \langle \mathcal{E} \rangle$. One of the two columns of T is $\begin{pmatrix} e \\ f \end{pmatrix}$, such that $e \neq 0$ and $f \neq 0$. If the first (resp. second) column has this property, $P_{Z_1}[1, -i]$ (resp. $P_{Z_1}[1, i]$) does not belong to $\mathcal{U}_3$. In this case, $F^{x_3 = P_{Z_1}[1,-i]}$ (resp. $F^{x_3 = P_{Z_1}[1,i]}$) is nondegenerate, and we will modify the construction in Figure 4 by adding the binary gadget with signature P_{Z_1} to replace the three edges whose endpoints are triangles marked by 3 in the gadget.

In case (c.), As $U \not\cong U^\perp$, U and $U^\perp$ are linearly independent. Then $U \not\perp U$, otherwise $U \cong 0$. Hence the inner product $U^{\mathtt{T}} U \neq 0$, and w.o.l.o.g, we assume $U = [a, b]$ and $U^\perp = [c, d]$ are unit vectors. Let $H = \begin{pmatrix} a & c \\ b & d \end{pmatrix}$, H is an orthogonal matrix. We show that one of $P_H[a, b]$ or $P_H[c, d]$ does not belong to $\mathcal{U}_3$. The proof is similar with case (b.). In case (c.) we will modify the construction in Figure 4 by adding the binary gadget with signature P_H to replace the three edges whose endpoints are triangles marked by 3 in the gadget.

The proof for Condition (2) is similar to Condition (1). The replacement in the construction of Figure 4 happens at the three edges connecting the copy of F with U_2 and the corner F.

Now consider Condition (3). If $R_L = L_R = [1, i]$ or $[1, -i]$, then $R_L^{\mathtt{T}} S_{Z_1} L_R \neq 0$ or $R_L^{\mathtt{T}} S_{Z_2} L_R \neq 0$. If $R_L = [a, b] \perp L_R = [c, d]$ and $R_L \notin \{[1, i], [1, -i]\}$, then one of $R_L^{\mathtt{T}} P_H L_R$ and $R_L^{\mathtt{T}} P_H^{\mathtt{T}} L_R$ is not zero, where $H = \begin{pmatrix} a & c \\ b & d \end{pmatrix}$. For Condition (3), the replacement in the construction of Figure 4 happens at the three edges connecting the copy of F with U_1 with the copy of F with U_2.

If Conditions (1) (2) (3) all hold, then the gadget satisfies the second property, and the lemma is proved. For each condition, if it does not hold, all possible cases are analyzed, and some binary function is added to rectify the construction, which are available by the conditions of the lemma. With these modifications to the construction in Figure 4, the proof of the lemma is complete.

□

We will prove the hardness part of Theorem 2.2 by appealing to our dichotomy theorem [9] for symmetric Holant* problems. For that purpose we need to construct appropriate *symmetric* binary signatures.

Lemma 7.4. *Let $\mathcal{F}$ denote any one of the function sets $\langle H\mathcal{E}\rangle$ (for an orthogonal matrix H), $\langle Z\mathcal{E}\rangle$ or $\langle Z\mathcal{M}\rangle$ (for the matrix $Z = Z_1$ or Z_2). Suppose we can realize a symmetric ternary function $F \in \mathcal{F} - \langle\mathcal{T}\rangle$ and a binary function $P \notin \mathcal{F}$. Then we can realize a* symmetric *binary function $Q \notin \mathcal{F}$.*

Proof. The matrix form of the *symmetric* binary function Q is $PF^{x_1=U}P^{\mathrm{T}}$, for some unary function U. Q is realizable by a gadget linking P followed by $F^{x_1=U}$ and then followed by P.

The essence of the proof is an appropriate holographic transformation. Denote $M = H$ or Z, suppose $P = MP_1M^{\mathrm{T}}$, and $F = M^{\otimes 3}F_1$. $F \notin \langle\mathcal{T}\rangle$ implies that $F_1 \notin \langle\mathcal{T}\rangle$. We also have $F^{x_1=U} = MF_1^{x_1=MU}M^{\mathrm{T}}$. Then the matrix form of $Q = MP_1M^{\mathrm{T}}MF_1^{x_1=MU}M^{\mathrm{T}}MP_1^{\mathrm{T}}M^{\mathrm{T}}$. Let $Q = MQ_1M^{\mathrm{T}}$.

Case (1) $\mathcal{F}$ is $\langle H\mathcal{E}\rangle$.

We take $M = H$. Since H is orthogonal, $Q = HP_1F_1^{x_1=HU}P_1^{\mathrm{T}}H^{\mathrm{T}}$. We have $F_1 \notin \langle\mathcal{T}\rangle$. $P_1 \notin \langle\mathcal{E}\rangle$, since $P \notin \langle H\mathcal{E}\rangle$. Also by $F \in \langle H\mathcal{E}\rangle$, we have $F_1 \in \langle\mathcal{E}\rangle$. By the nondegeneracy condition $F_1 \notin \langle\mathcal{T}\rangle$, we have $F_1 \in \mathcal{E}$. Being symmetric and nondegenerate, $F_1 = [u, 0, 0, v]$, where $u \neq 0$ and $v \neq 0$. We only need to prove $Q_1 = P_1F_1^{x_1=HU}P_1^{\mathrm{T}} \notin \langle\mathcal{E}\rangle$, which is the same as $Q \notin \langle H\mathcal{E}\rangle$. Because we can pick any $U' = HU$, for any x, y, we can realize $F_1^{x_1=HU} = [x, 0, y]$. Suppose $P_1 = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. Then $Q_1 = \begin{pmatrix} a^2x + b^2y & acx + bdy \\ acx + bdy & c^2x + d^2y \end{pmatrix}$.

We need $Q_1 \notin \langle\mathcal{E}\rangle$. That is, Q_1 is nonsingular, Q_1 is not of $[*, 0, *]$ form, and Q_1 is not of $[0, *, 0]$ form. By the Separation Lemma, we only need to prove that there is some $[x, y]$ to satisfy each property. If $x \neq 0$ and $y \neq 0$, $F_1^{x_1=HU} = [x, 0, y]$ and P_1 are both nonsingular. Thus Q_1 is nonsingular. Because $P_1 \notin \langle\mathcal{E}\rangle$, either $ac \neq 0$ or $bd \neq 0$. There exists some $[x, y]$ such that $acx + bdy \neq 0$. Similarly, we can prove there exists some $[x, y]$ such that $a^2x + b^2y \neq 0$ and $c^2x + d^2y \neq 0$ simultaneously.

Case (2) $\mathcal{F}$ is $\langle Z\mathcal{E}\rangle$.

Take $M = Z$. Note that $Z^{\mathrm{T}}Z \cong (\neq_2)$. $Q \cong ZP_1(\neq_2)F_1^{x_1=ZU}(\neq_2)P_1^{\mathrm{T}}Z^{\mathrm{T}} = ZQ_1Z^{\mathrm{T}}$. We have $F_1 \notin \langle\mathcal{T}\rangle$, $F_1 \in \mathcal{E}$, and $P_1 \notin \langle\mathcal{E}\rangle$. We only need to prove $Q_1 \notin \langle\mathcal{E}\rangle$. For any x, y, we can pick U to realize $(\neq_2)F_1^{x_1=ZU}(\neq_2) = [x, 0, y]$. This is seen by the fact that $(\neq_2)[x, 0, y](\neq_2) = [y, 0, x]$. The remaining proof is the same as Case (1) for $\langle H\mathcal{E}\rangle$.

Case (3) $\mathcal{F}$ is $\langle Z\mathcal{M}\rangle$.

Take $M = Z$. Since $Z^{\mathrm{T}}Z \cong (\neq_2)$, we have $Q_1 \cong P_1(\neq_2)F_1^{x_1=ZU}(\neq_2)P_1^{\mathrm{T}}$.

We have $F_1 \notin \langle\mathcal{T}\rangle$, and $F_1 \in \mathcal{M}$. Being symmetric and nondegenerate, F_1 has the form $F_1 \cong [f, 1, 0, 0]$. Let $P_1 = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, and $P_1 \notin \langle\mathcal{M}\rangle$. For any x, there is some U, such that $F_1^{x_1=ZU} = [x, 1, 0]$. We only need to prove $Q_1 = P_1(\neq_2)F_1^{x_1=ZU}(\neq_2)P_1^{\mathrm{T}} \notin \langle\mathcal{M}\rangle$.

Because $P_1 \notin \langle\mathcal{M}\rangle$, Q_1 is not singular. Also $d \neq 0$. There exists x such that $Q_1(1, 1) = 2cd + d^2x \neq 0$. Hence, $Q_1 \notin \langle\mathcal{M}\rangle$. ∎

Proof of Theorem 2.2 (hardness part)

If $\mathcal{F} \nsubseteq \langle\mathcal{T}\rangle$, $\mathcal{F} \nsubseteq \langle H\mathcal{E}\rangle$, $\mathcal{F} \nsubseteq \langle Z\mathcal{E}\rangle$, and $\mathcal{F} \nsubseteq \langle Z\mathcal{M}\rangle$, by Lemma 6.1, we can realize functions of arity of 2 or 3 not belonging to these function sets respectively.

The conditions in Lemma 7.3 are satisfied, so we can realize a symmetric ternary function $Q_3 \notin \langle\mathcal{T}\rangle$ (with the help of those binary functions). If Holant$^*(\{Q_3\})$ is hard, then the theorem is proved. Otherwise, by the dichotomy theorem for the symmetric case [9], Q_3 belongs to one of the special function families from Theorem 2.1. It can be shown that these are precisely restrictions of $H\mathcal{E}$, $Z\mathcal{E}$ or $Z\mathcal{M}$ to symmetric signatures. By Lemma 7.4, we can realize a symmetric binary function Q_2 not in this set. By the dichotomy theorem for the symmetric case [9], Holant$^*(\mathcal{F})$ is #P-hard.

References

[1] Leslie Ann Goldberg Markus Jalsenius Mark Jerrum Andrei Bulatov, Martin Dyer and David Richerby. The complexity of weighted and unweighted #csp. *manuscript*, 2010.

[2] R.J. Baxter. *Exactly solved models in statistical mechanics*. Academic press London, 1982.

[3] Andrei A. Bulatov. The complexity of the counting constraint satisfaction problem. In Luca Aceto, Ivan Damgård, Leslie Ann Goldberg, Magnús M. Halldórsson, Anna Ingólfsdóttir, and Igor Walukiewicz, editors, *ICALP (1)*, volume 5125 of *Lecture Notes in Computer Science*, pages 646–661. Springer, 2008.

[4] Andrei A. Bulatov and Martin Grohe. The complexity of partition functions. In Josep Díaz, Juhani Karhumäki, Arto Lepistö, and Donald Sannella, editors, *ICALP*, volume 3142 of *Lecture Notes in Computer Science*, pages 294–306. Springer, 2004.

[5] Jin-Yi Cai and Xi Chen. A decidable dichotomy theorem on directed graph homomorphisms with non-negative weights. *to appear in FOCS*, 2010.

[6] Jin-Yi Cai, Xi Chen, and Pinyan Lu. Graph homomorphisms with complex values: A dichotomy theorem. *To appear in ICALP*.

[7] 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.

[8] 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.

[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] Nadia Creignou and Miki Hermann: Complexity of Generalized Satisfiability Counting Problems. Inf. Comput. 125(1): 1-12 (1996)

[11] Martin Dyer and David Richerby. On the complexity of #CSP. *STOC*, 2010.

[12] Martin E. Dyer, Leslie Ann Goldberg, Markus Jalsenius, and David Richerby. The complexity of approximating bounded-degree boolean #csp. In Jean-Yves Marion and Thomas Schwentick, editors, *STACS*, volume 5 of *LIPIcs*, pages 323–334. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2010.

[13] Martin E. Dyer, Leslie Ann Goldberg, and Mark Jerrum. The complexity of weighted boolean #csp. *CoRR*, abs/0704.3683, 2007.

[14] Martin E. Dyer, Leslie Ann Goldberg, and Mark Jerrum. An approximation trichotomy for boolean #csp. *J. Comput. Syst. Sci.*, 76(3-4):267–277, 2010.

[15] Martin E. Dyer, Leslie Ann Goldberg, and Mike Paterson. On counting homomorphisms to directed acyclic graphs. *J. ACM*, 54(6), 2007.

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

[17] M. Freedman, L. Lovász, and A. Schrijver. Reflection positivity, rank connectivity, and homomorphism of graphs. *J. AMS*, 20:37–51, 2007.

[18] Leslie Ann Goldberg and Mark Jerrum. Approximating the partition function of the ferromagnetic potts model. *CoRR*, abs/1002.0986, 2010.

[19] L. A. Goldberg, M. Grohe, M. Jerrum, and M. Thurley. A complexity dichotomy for partition functions with mixed signs. In *Proceedings of the 26th International Symposium on Theoretical Aspects of Computer Science*, pages 493–504, 2009.

[20] P. Hell and J. Nešetřil. On the complexity of h-coloring. *Journal of Combinatorial Theory, Series B*, 48(1):92–110, 1990.

[21] E. Ising. Beitrag zur theorie des ferromagnetismus. *Zeitschrift für Physik A Hadrons and Nuclei*, 31(1):253–258, 1925.

[22] M.R. Jerrum and A.J. Sinclair. Polynomial-time approximation algorithms for the Ising model. *SIAM Journal on Computing*, 22: 10871116, 1993.

[23] M.R. Jerrum and A.J. Sinclair, The Markov Chain Monte Carlo Method: An Approach to Approximate Counting and Integration. Approximation Algorithms for NP-Hard Problems, D.S. Hochbaum, ed., PWS Publishing, 1997, pp. 482520.

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

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

[26] Richard E. Ladner. On the structure of polynomial time reducibility. *J. ACM*, 22(1):155–171, 1975.

[27] T.D. Lee and C.N. Yang. Statistical theory of equations of state and phase transitions. II. Lattice gas and Ising model. *Physical Review*, 87(3):410–419, 1952.

[28] E.H. Lieb and A.D. Sokal. A general Lee-Yang theorem for one-component and multicomponent ferromagnets. *Communications in Mathematical Physics*, 80(2):153–179, 1981.

[29] L. Lovász. Operations with structures. *Acta Math. Hung.*, 18:321–328, 1967.

[30] N. Madras and D. Randall. Markov chain decomposition for convergence rate analysis. Annals of Applied Probability, 12: 581606, 2002.

[31] B.M. McCoy and T.T. Wu. *The two-dimensional Ising model*. Harvard University Press Cambridge, 1973.

[32] L. Onsager. Crystal statistics. I. A two-dimensional model with an order-disorder transition. *Physical Review*, 65(3-4):117–149, 1944.

[33] D. Randall. Mixing. The 44th Symposium on Foundations of Computer Science (FOCS) : 4-15.

[34] T. Schaefer. The Complexity of Satisfiability Problems. STOC 1978. pp. 216226.

[35] N. Creignou, S. Khanna, and M. Sudan. Complexity Classifications of Boolean Constraint Satisfaction Problems. Siam Monographs On Discrete Mathematics And Applications. 2001. ISBN:0-89871-479-6.

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

[37] Marc Thurley. The complexity of partition functions on hermitian matrices. *CoRR*, abs/1004.0992, 2010.

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

[39] Leslie G. Valiant. Holographic algorithms (extended abstract). In *FOCS '04: Proceedings of the 45th Annual IEEE Symposium on Foundations of Computer Science*, pages 306–315, Washington, DC, USA, 2004. IEEE Computer Society.

[40] 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.

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

[42] DJA Welsh. *Complexity: knots, colourings and counting.* Cambridge Univ Pr, 1993.

[43] C.N. Yang. The spontaneous magnetization of a two-dimensional Ising model. *Physical Review*, 85(5):808–816, 1952.

[44] C.N. Yang and T.D. Lee. Statistical theory of equations of state and phase transitions. I. Theory of condensation. *Physical Review*, 87(3):404–409, 1952.