Theory of Computing Systems (2020) 64:1362–1391
https://doi.org/10.1007/s00224-020-09983-8

Dichotomy for Holant* Problems on the Boolean Domain

Jin-Yi Cai[1] · Pinyan Lu[2] · Mingji Xia[3]

Published online: 22 June 2020

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 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 with the broader scope that allows complex-valued constraint functions, we discover surprising new tractable classes, which are associated with isotropic vectors, i.e., a (non-zero) vector whose dot product with itself is zero.

Keywords #P-hardness · Polynomial time algorithms · Dichotomy theorems · Holant problems · Constraint satisfaction problems · Edge coloring models

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

A preliminary version of this paper appeared in ACM-SIAM Symposium on Discrete Algorithms (SODA) 2011 [12]. Jin-Yi Cai is supported by NSF CCF-1714275. Pinyan Lu is supported by Science and Technology Innovation 2030 "New Generation of Artificial Intelligence" Major Project No.(2018AAA0100903), NSFC grant 61922052 and 61932002, Program for Innovative Research Team of Shanghai University of Finance and Economics (IRTSHUFE) and the Fundamental Research Funds for the Central Universities. Mingji Xia is supported by NSFC grant 61932002, and the Youth Innovation Promotion Association, CAS.

✉ Jin-Yi Cai
jyc@cs.wisc.edu

Extended author information available on the last page of the article.

$\sigma : V \to \{0, 1\}$, 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 vertex 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 $H = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$ 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, even for 0-1 valued constraint functions.

When every edge is attached the same symmetric edge function it is called a graph homomorphism problem [29, 39]. 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 [30] to work on what is now known as the Ising model, physicists have studied so-called "Exactly Solved Models" [3, 41]. In computer science language, physicists' notion of an "exactly solvable" system corresponds roughly 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. [3, 30, 33, 34, 36, 37, 42, 46, 50–52]) 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 [6, 7, 9, 23, 24, 27] in terms of polynomial time tractability or #P-hardness. A major further research direction is when a #P-hard partition function can be approximated [20, 22, 28, 31, 32, 40, 43].

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 $\text{Holant}(G) = \sum_{\sigma:E\to\{0,1\}} \mathbf{wt}(\sigma)$. Given a choice of local constraint functions, a Holant problem on G is to evaluate $\text{Holant}(G)$.

Holant problems were defined in [16], and the name was inspired by the introduction of *Holographic Algorithms* by L. Valiant [48, 49] (who first used the term Holant) [2]. Partition functions $\mathbf{Z}(G)$ of the type discussed above are for vertex

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

[2] B. Szegedy [45] studied an *edge coloring model*, which is identical to Holant problems on a general domain, with a single symmetric constraint function per each arity.

models (spin systems) where assignments σ are on vertices. Holant is a partition function of edge models where assignments σ are on edges. It is easy to simulate a partition function of a vertex model 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. We note that Freedman, Lovász, and Schrijver [26] proved that counting perfect matchings cannot be expressed as a real-valued graph homomorphisms. This impossibility was extended to the complex-valued graph homomorphisms in [15].

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 [18, 19, 44] (By an analogue of Ladner's theorem [35], such a dichotomy is *false* for the whole #P, unless P = $\text{P}^{\#\text{P}}$.) Dichotomy theorems have been obtained for counting graph homomorphisms for successively broader class of functions [6, 7, 9, 23, 24, 27]. A sweeping dichotomy theorem for all #CSP with 0-1 constraint functions over any finite domain was given by Bulatov [4]. An alternative proof is given by Dyer and Richerby [25]. It can be extended to functions taking non-negative rational values [5]. 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. [9, 27]).

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. (The function $=_k$ takes k inputs and output 1 if all inputs are equal, and output 0 otherwise.) 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 the so-called *degenerate* functions, which are particularly weak, and have played a crucial role in many classification theorems. Holant* is the class of Holant problems where all unary functions are free, i.e., $\text{Holant}^*(\mathcal{F}) = \text{Holant}(\mathcal{F} \cup \mathcal{U})$.

Previously we have studied Holant* problems for any set $\mathcal{F}$ of *symmetric* functions on Boolean variables [17]. 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 [17]. This improves previously the strongest dichotomy for Boolean #CSP($\mathcal{F}$) by Dyer, Goldberg and Jerrum [21], 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 [11, 16, 48, 49]. 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 introduced in [16], and denoted by $\mathscr{F}$. 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 $\mathcal{F}_3$ and $\mathcal{F}_4$ depend even more fundamentally on holographic transformations. It is also here that the complex field $\mathbb{C}$ 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. a class related to weighted matchings that we call Matching gates), after a holographic transformation correlated with isotropic vectors.

Our dichotomy here is a generalization of the dichotomy in [17] for symmetric Holant* Problems. The symmetric dichotomy can be viewed as a special case of the dichotomy in this paper and on the other hand also serves as the starting point for our reduction. Furthermore, by proving a dichotomy theorem in this more general setting, 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 an outline of the hardness proof. 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 Holant$^*(\mathcal{F})$.

Subsequent development: The complexity dichotomy theorem in this paper has been included in the book [8]. It also have been generalized and extended since its conference publication [12]. Miriam Backens [1] first gave an extension of this dichotomy to Holant$^+$ problems. These are Holant problems that allow only four

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

unary auxiliary functions for free instead of all unary functions as in Holant*. These four unary auxiliary functions are inspired by connections to quantum computing. This dichotomy [1] is a stronger theorem. Jiabao Lin and Hanpin Wang [38] proved a dichotomy for Holant problems on the Boolean domain where constraint functions are not necessarily symmetric, but take nonnegative values. We [14] proved a dichotomy for Holant^c problems on the Boolean domain where constraint functions are not necessarily symmetric, but take real values. Holant^c problems are Holant problems that allow only two unary auxiliary functions (the pinning functions). These last two results [14, 38] are incomparable in scope, and do not strictly strengthen the dichotomy in the present paper. However, Backens [2] has achieved a generalization of [14] to complex-valued constraint functions. This strengthens not only [14], but also the dichotomy in the present paper as well as [1]. However, these subsequent results are proved using the results here, and thus they are logically dependent on the present paper.

Going beyond the Boolean domain, there have been relatively few dichotomies proved for Holant problems. In [13] using the dichotomy proved here we proved a very restricted dichotomy for Holant* problems on domain size 3. This domain 3 dichotomy makes essential use of the dichotomy of this paper. One other Holant dichotomy for higher domain problems is [10], where counting edge colorings is a special case.

2 Definition and Statement

A (constraint) function F, or synonymously a signature, of arity $n \geq 0$, is a mapping from $\{0, 1\}^n$ to $\mathbb{C}$. 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 in lexicographic order. 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 both range over $\{0, 1\}$. We also use this matrix form for functions of larger arities. For example, $F_{x_1x_2,x_3}$ 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).

Fix a set of signatures $\mathcal{F}$. We allow $\mathcal{F}$ to be infinite for the convenience of expressing some theorems and proofs; see below. A *signature grid* $\Omega = (G, \mathcal{F}, \pi)$ consists of a graph $G = (V, E)$, and a labeling π which maps each vertex $v \in V$ to a function

$f_v \in \mathcal{F}$ of arity $\deg(v)$ with input variables correspond to an ordered list of incident edges $E(v)$ of v. The Holant problem on instance Ω is to compute

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

A Holant problem is parameterized by a set of signatures $\mathcal{F}$.

Definition 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$_\Omega$.

When $\mathcal{F}$ is infinite, to account for the input size, we require that the labeling π include the specification of any function f_v used.

Bipartite Holant problems Holant($\mathcal{F} \mid \mathcal{G}$) are similarly defind, where signature grids have bipartite underlying graphs (U, V, E), and vertices in U and V are assigned signatures from $\mathcal{F}$ and $\mathcal{G}$ respectively.

We would like to characterize the complexity of Holant problems in terms of its signature sets. We say Holant($\mathcal{F}$) is tractable, if it is computable in P. Note that for an infinite $\mathcal{F}$ the input size includes the description of the signatures in the input instance Ω. We say Holant($\mathcal{F}$) is #P-hard if there exists a finite subset of $\mathcal{F}$ for which the problem is #P-hard.

Definition 2 Let $\mathcal{U}$ denote the set of all unary signatures. Then Holant$^*(\mathcal{F})$ = Holant$(\mathcal{F} \cup \mathcal{U})$.

A degenerate signature is a tensor product of unary signatures. Since all unary signatures can be used for free in Holant$^*(\mathcal{F})$, we may assume the arity of every signature in $\mathcal{F}$ is greater than one. And since any degenerate signature can be decomposed to unary signatures, we also assume that every signature in $\mathcal{F}$ is non-degenerate.

In [17], we proved a dichotomy theorem when $\mathcal{F}$ is a set of symmetric signatures.

Theorem 1 *Let $\mathcal{F}$ be a set of non-degenerate symmetric signatures over $\mathbb{C}$. Then Holant$^*(\mathcal{F})$ is computable in polynomial time in the following three Classes. 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 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 $ax_k + bx_{k+1} - ax_{k+2} = 0$; (2) $n = 2$ and the signature $[x_0, x_1, x_2]$ is of the form $[2a\lambda, b\lambda, -2a\lambda]$.*
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 valid 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 Fig. 1 for one example.)

Other than these dangling edges, an $\mathcal{F}$-gate is the same as a signature grid. The role of dangling edges is to provide input/output variables. This is similar to the notion of external nodes for matchgates in Valiant's definition [47, 49], 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 $\Gamma = (H, \mathcal{F}, \pi)$,

$$\Gamma(y_1, \ldots, y_n) = \sum_{x_1,\ldots,x_m \in \{0,1\}} H(x_1, \ldots, x_m, y_1, \ldots y_n),$$

where $(y_1, y_2, \ldots, y_n) \in \{0, 1\}^n$ 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 Holant($\mathcal{F} \cup \{g\}$) $\leq_T$ Holant($\mathcal{F}$). The reduction is quite simple. Given an instance of Holant($\mathcal{F} \cup \{g\}$), by replacing every appearance of g by an $\mathcal{F}$-gate Γ, we get an instance of 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 Fig. 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

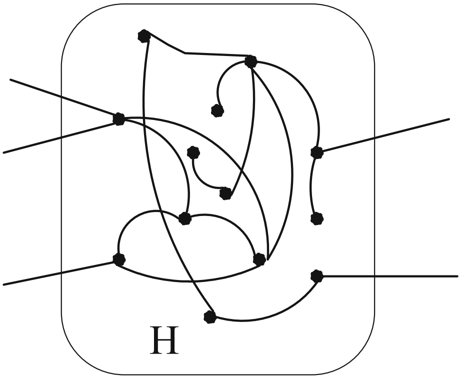
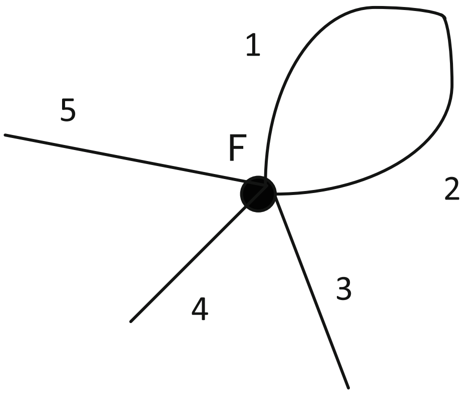


Fig. 1 An example of an $\mathcal{F}$-gate with five dangling edges

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



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 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 Fig. 3 for one example.) This definition of $F = \bigotimes_{\mathcal{I}}(F_1, \ldots, F_k)$ can be easily generalized to the case where $\mathcal{I} = \{I_1, \ldots, I_k\}$ is a partition of an arbitrary finite set of indices.

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 (for $s = 1, \ldots, k$). $F^{x_j=0}$, $F^{x_j=1}$ 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 for FU). 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 (for row) and x_6 (for column). Note that $A_{\emptyset,x_1} B_{x_2,\emptyset}$ or $A^{\mathrm{T}}B$ is the dot 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.

When we discuss function sets $\mathcal{F}$, whenever a function $f(X) \in \mathcal{F}$, where $|X| = n$, we may change the names of the variables, i.e., we consider $f(X')$ also belongs to $\mathcal{F}$, where X' is another set of variables, $|X'| = |X|$. We say a function set $\mathcal{F}$ is closed under tensor product (or more precisely under juxtaposition), if for any functions A and B on two disjoint sets of variables indexed by I and J respectively, $A, B \in \mathcal{F}$ implies that $\bigotimes_{\mathcal{I}}(A, B) \in \mathcal{F}$, where $\mathcal{I} = \{I, J\}$. 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

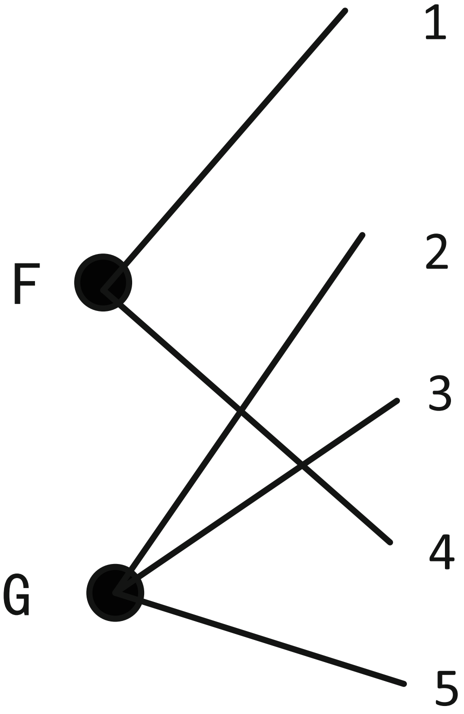


Fig. 3 An example of juxtaposition $\bigotimes_{\mathcal{I}}(F, G)$, where $\mathcal{I} = \{\{1, 4\}, \{2, 3, 5\}\}$

(possibly) 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 (possibly) 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 $\mathcal{T}$.

The class $\langle\mathcal{U}\rangle$ is called the degenerate signatures. A binary function belongs to $\langle\mathcal{U}\rangle$ iff its matrix form is singular. 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 $M \circ \mathcal{F}$ to denote the set consisting of all functions in $\mathcal{F}$ transformed by a matrix M,

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

$M^{\otimes r_F} F$ is called a holographic transformation of the function F. In the notation $M^{\otimes r_F} F$ we write F as a column vector of dimension 2^{r_F}. If $f(x_1, \ldots, x_n, y_1, \ldots, y_m)$ is a function on $n + m$ variables, and if we use F to denote its representation as a column vector of dimension 2^{n+m}, and use $F_{n,m}$ to denote its representation as a

$2^n \times 2^m$ matrix, then the holographic transformation $M^{\otimes(n+m)}F$ has its representation in matrix notation $M^{\otimes n}F_{n,m}(M^{\mathrm{T}})^{\otimes m}$. In particular if f is a binary function with a 2×2 matrix form $F_{x,y}$, then its holographic transformation by M is $MF_{x,y}M^{\mathrm{T}}$ in matrix form.

Define

$$Z_1 = \begin{pmatrix} 1 & 1 \\ \mathfrak{i} & -\mathfrak{i} \end{pmatrix} \quad \text{and} \quad Z_2 = \begin{pmatrix} 1 & 1 \\ -\mathfrak{i} & \mathfrak{i} \end{pmatrix}.$$

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

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

Theorem 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 $\mathcal{F} \subseteq \langle H \circ \mathcal{E} \rangle$, or (3) there exists a matrix $Z \in \{Z_1, Z_2\}$ such that $\mathcal{F} \subseteq \langle Z \circ \mathcal{E} \rangle$, or (4) there exists a matrix $Z \in \{Z_1, Z_2\}$ such that $\mathcal{F} \subseteq \langle Z \circ \mathcal{M} \rangle$. In all other cases, Holant$^*(\mathcal{F})$ is #P-hard. The dichotomy is still valid even if the inputs are restricted to planar graphs.*

3 Tractability

The tractability part is given by the following theorem.

Theorem 3 *The following classes of Holant* problems are polynomial time computable.*

- *Holant$^*(\langle \mathcal{T} \rangle)$*
- *Holant$^*(\langle H \circ \mathcal{E} \rangle)$;*
- *Holant$^*(\langle Z \circ \mathcal{E} \rangle)$; and*
- *Holant$^*(\langle Z \circ \mathcal{M} \rangle)$*

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$^*(H \circ \mathcal{E})$. We first reformulate it as a bipartite Holant problem Holant$(=_2 \ | H \circ \mathcal{E})$ (since $\mathcal{U} = H \circ \mathcal{U} \subset H \circ \mathcal{E}$, we can drop the $*$ notation in Holant). 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$^*(H \circ \mathcal{E})$ is polynomial time computable.

The third class is Holant$^*(Z \circ \mathcal{E})$. Because $\mathcal{U} \subseteq Z \circ \mathcal{E}$, it is a bipartite Holant problem Holant($=_2$ $|Z \circ \mathcal{E}$). 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

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

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:

$$\text{Holant}(=_2 |H \circ \mathcal{F}) \longleftrightarrow \text{Holant}(=_2 |\mathcal{F}),$$

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

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

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

$$\text{Holant}(=_2 |Z \circ \mathcal{F}) \longleftrightarrow \text{Holant}(\neq_2 |\mathcal{F}).$$

Hence, after a holographic reduction, our problem Holant$^*(Z \circ \mathcal{E})$ becomes Holant$(\{\neq_2\}|\mathcal{E})$. (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$^*(Z \circ \mathcal{M})$. After a holographic reduction by Z^{-1} on the RHS, it becomes Holant$(\{\neq_2\}|\mathcal{M})$. 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$(\{\neq_2\}|\mathcal{M} - \mathcal{U})$ 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 all edges assigned 0 and the other half 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 $\langle H \circ \mathcal{E} \rangle$ is a proper generalization of Fibonacci gates defined in [16] and denoted by $\mathscr{F}$. 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 [16]. 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

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

transforms the symmetric signature $a \begin{pmatrix} 1 \\ 0 \end{pmatrix}^{\otimes k} + b \begin{pmatrix} 0 \\ 1 \end{pmatrix}^{\otimes k} = [a, 0, \ldots, 0, b] \in \mathcal{E}$ 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 $\langle H \circ \mathcal{E} \rangle$. 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, that is, if $\mathcal{F} \not\subseteq \langle \mathcal{T} \rangle$, $\mathcal{F} \not\subseteq \langle H \circ \mathcal{E} \rangle$, $\mathcal{F} \not\subseteq \langle Z \circ \mathcal{E} \rangle$, and $\mathcal{F} \not\subseteq \langle Z \circ \mathcal{M} \rangle$, 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 [17]. 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, for any one of the four tractable families $\mathcal{F}'$, starting from any funtion F of "hign arity which is not contained in $\mathcal{F}'$, we can construct a function Q such that (1) Holant$^*(\mathcal{F} \cup \{Q\}) \equiv_T$ Holant$^*(\mathcal{F} \cup \{F\})$, (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), and it contains a signature of "high" arity outside a particular tractable family, we can keep on doing arity reductions. Specifically, in a finite number of steps this will 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 $\langle H \circ \mathcal{E} \rangle$, $\langle Z \circ \mathcal{E} \rangle$, $\langle Z \circ \mathcal{M} \rangle$, 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 across 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 that is not in $\langle \mathcal{T} \rangle$, but also some binary signatures that are not in $\langle H \circ \mathcal{E} \rangle$, $\langle Z \circ \mathcal{E} \rangle$ or $\langle Z \circ \mathcal{M} \rangle$, respectively, to construct a symmetric signature of arity 3 that is not in $\langle \mathcal{T} \rangle$. This is proved in Theorem 4. Similarly, in Theorem 5, we prove that we can also construct binary symmetric signatures that are not in $\langle H \circ \mathcal{E} \rangle$, $\langle Z \circ \mathcal{E} \rangle$ or $\langle Z \circ \mathcal{M} \rangle$, respectively. Then by the symmetric dichotomy, we know that either this ternary signature already defines a #P-hard problem or it belongs to $\langle H \circ \mathcal{E} \rangle$, $\langle Z \circ \mathcal{E} \rangle$ or $\langle Z \circ \mathcal{M} \rangle$. If it is #P-hard, then we are done. Otherwise, since we have a binary signature that is not in the same family, we also get the hardness result by the symmetric dichotomy [17]. We note that, all our starting problems for hardness are already hard for planar graphs and all the gadgets we use in the reduction are planar. As a result, our final dichotomy also holds for planar graphs. In the proof later, we will not explicitly state this every time.

One technical lemma is used extensively in both Sections 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 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, for all $1 \leq i \leq m$. It follows that there exists $a_n \in \mathbb{F}$ such that $P_i(a_1, a_2, \ldots, a_n) \neq 0$ for all $1 \leq i \leq m$. □

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 conditions separately.

Lemma 2 *Let F be the signature of a gadget construction using unary functions $U_k = [x_k, y_k]$, $k = 1, 2, \ldots, m$.*

Suppose $S_1, S_2, \ldots, S_N$ are sets of functions, where a function $K \in S_i$ iff the signature entries of K satisfy a finite system of polynomial equations $\{P_{i,1} = 0, P_{i,2} = 0, \ldots, P_{i,m_i} = 0\}$.

If for every $1 \leq i \leq N$, there exists an assignment σ of x_k and y_k ($k = 1, 2, \ldots, m$), such that F is shown to be not in S_i under σ, then there exists an assignment σ' of x_k and y_k ($k = 1, 2, \ldots, m$), such that F is shown under σ' to be not in all S_i ($1 \leq i \leq N$).

Equivalently, by contrapositive, suppose for every assignment σ' of x_k and y_k ($k = 1, 2, \ldots, m$), there exists $1 \leq i \leq N$, such that F satisfies the condition for being in S_i under σ', then there exists $1 \leq i \leq N$, such that for all assignments σ of x_k and y_k ($k = 1, 2, \ldots, m$), F satisfiyes the condition for being in S_i under σ.

Proof Every signature entry of F is expressible as a polynomial in x_k and y_k, $k = 1, 2, \ldots, m$. Substituting these expressions in $\{P_{i,1} = 0, P_{i,2} = 0, \ldots, P_{i,m_i} = 0\}$ we can express the condition $F \in S_i$ by a finite set of polynomial equations $\{P'_{i,1} = 0, P'_{i,2} = 0, \ldots, P'_{i,m_i} = 0\}$ on x_k and y_k. If for every $1 \leq i \leq N$, there exists an assignment σ of x_k and y_k ($k = 1, 2, \ldots, m$), such that $F \notin S_i$ under σ, then for every $1 \leq i \leq N$, there exists some $1 \leq j_i \leq m_i$ such that P'_{i,j_i} is not identically 0 as a polynomial on x_k and y_k ($k = 1, 2, \ldots, m$). Therefore there is an assignment σ' on x_k and y_k ($k = 1, 2, \ldots, m$), such that for all $1 \leq i \leq N$, $P'_{i,j_i}|_{\sigma'} \neq 0$, hence $F \notin S_i$ under σ'. □

The following lemma is another direct corollary of lemma 1.

Lemma 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 what follows.

6 Arity Reduction

In the next two sections we prove the hardness part of Theorem 2, that is, if $\mathcal{F} \not\subseteq \langle\mathcal{T}\rangle$, $\mathcal{F} \not\subseteq \langle H \circ \mathcal{E}\rangle$, $\mathcal{F} \not\subseteq \langle Z \circ \mathcal{E}\rangle$, and $\mathcal{F} \not\subseteq \langle Z \circ \mathcal{M}\rangle$, then Holant$^*(\mathcal{F})$ is #P-hard.

In this section, we show that for any one of the four tractable families $\mathcal{F}'$, starting from any funtion $F \in \mathcal{F}$, if $F \notin \mathcal{F}'$ then 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 4 *Let $\mathcal{F}'$ be any one of $\langle\mathcal{T}\rangle$, or $\langle H \circ \mathcal{E}\rangle$, or $\langle Z \circ \mathcal{E}\rangle$, or $\langle Z \circ \mathcal{M}\rangle$. 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 $r <$ 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) $r \le$ arity$(Q) <$ arity(F).

The proof of this lemma is divided into the following several lemmas. Recall that we say a signature F has (respectively, a set of signatures $\mathcal{F}$ all have) a type $\mathcal{I}$ if F (respectively, every signature in $\mathcal{F}$) can be expressed as a tensor product of functions on variable sets from the partition $\mathcal{I}$. Note that the type of a signature is not unique; e.g., if $\mathcal{J}$ is a refinement of $\mathcal{I}$, then a signature having type $\mathcal{J}$ also has type $\mathcal{I}$.

A type specification is given by a *type* $\mathcal{I}$, and is the requirement that a signature (or a set of signatures all) have the given type. Firstly, we show that any *type specification* in a tensor product decomposition can be described by a system of polynomial equations.

Lemma 5 *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 $\mathcal{I} = \{[n]\}$, the trivial partition that consists of a single set $[n] = \{1, \ldots, n\}$, 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 $\mathcal{I}$. Recall that when a function F has type $\{I_1, \ldots, I_k\}$, we use X_j to denote the subsequence $X|_{I_j}$ of the input variable sequence X of F. Then obviously, for any two values a_1, b_1 of X_1 and any two values a_2, b_2 of X_2, $F(a_1, a_2)F(b_1, b_2) = F(a_1, b_2)F(b_1, a_2)$. (In this equation, for the simplicity of expression, we write it in such a way that assumes all indices in I_1 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 by the following argument: Arrange the values of F into a matrix $F_{X_1,X_2} = (F(a_1, a_2))$, where the row indices (respectively column indices) are all possible values of X_1 (respectively X_2). The conditions $F(a_1, a_2)F(b_1, b_2) = F(a_1, b_2)F(b_1, a_2)$ for all a_1, b_1 and all a_2, b_2 imply that any 2 by 2 submatrix of F_{X_1,X_2} is singular, and so $\mathrm{rank}(F_{X_1,X_2}) \leq 1$. Hence, F_{X_1,X_2} is the product of a column vector and a row vector. It follows that 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: for all $1 < i \leq k$, and for all 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 to have 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 $k = 2$, $F = \bigotimes_{\{\bigcup_{j=1}^{k-1} I_j, I_k\}}(Q_{k-1}, F_k)$, where Q_{k-1} and F_k are functions on the respective sets of variables $\bigcup_{j=1}^{k-1} X_j$ and X_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 $F^{X_k=a_k} = Q_{k-1}F_k(a_k)$. After canceling the non-zero factor $F_k(a_k)$, by induction, we obtain the necessary and sufficient conditions that Q_{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 4.

Lemma 6 *Suppose there exists some type $\mathcal{I} = \{I_1, \ldots, I_k\}$ over $[n-1]$, such that for all unary functions $U = [x, y]$, $F^{x_n=U} = FU = xF^{x_n=0} + yF^{x_n=1}$ 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 exactly one index $i \in [k]$ such that F_i and K_i are linearly independent.*

Proof Obviously F_i and K_i cannot be linearly dependent for all $i \in [k]$, for otherwise $F^{x_n=0}$ and $F^{x_n=1}$ would be linearly dependent, contrary to assumption. Now for a contradiction, suppose there are two distinct indices $i \in [k]$, such that F_i and K_i are linearly independent. Without loss of generality, 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 1, there exist $|I_j|$ many unary functions such that both F_j and K_j become nonzero constants when combined with them. This is because we can write a finite system of polynomial equations Φ_j in a_k, b_k for the $|I_j|$ many unary functions $U_k = [a_k, b_k]$, that expresses the condition that F_j becomes 0 when connecting variables x_k to U_k, for all $k \in I_j$. Since F_j is not the zero function, some 0-1 assignment violates some equation in Φ_j. Hence at least one polynomial P_1 in Φ_j is not identically zero. Similarly there is a finite system of polynomial equations Ψ_j for K_j and some polynomial P_2 is not identically zero. Then by Lemma 1 there is a common assignment to a_k, b_k ($k \in I_j$) such that both F_j and K_j are nonzero. After combining $F^{x_n=0}$ and $F^{x_n=1}$ with these unary functions, for all $j \in [k] - \{1, 2\}$, 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, FU has type $\mathcal{I}$, then this function has type $\{I_1, I_2\}$. However, we 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 the $2^{|I_1|} \times 2^{|I_2|}$ matrix

$$\begin{bmatrix} F_1 & K_1 \end{bmatrix} \begin{bmatrix} c_0 x & 0 \\ 0 & c_1 y \end{bmatrix} \begin{bmatrix} F_2^{\mathrm{T}} \\ K_2^{\mathrm{T}} \end{bmatrix},$$

where F_1, K_1, F_2, K_2 are column vectors. Since F_1 and K_1 are linearly independent, and F_2 and 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. □

Proof of Lemma 4: Case $\mathcal{F}' = \langle \mathcal{T} \rangle$ □

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 directly follows from the last statement of Lemma 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 we assume $F^{x_n=0} = \bigotimes_{\mathcal{I}}(F_1, F_2, \ldots, F_k)$ 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, there is exactly one pair of linearly independent tensor factors, without loss of generality, 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, $|I_1| = 2$.

Without loss of generality, 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. By this we get a ternary function Q in variables x_1, x_2, x_n, and $F = Q \otimes F_2 \otimes \cdots \otimes F_k$. If $Q \in \langle \mathcal{T} \rangle$, then $F \in \langle \mathcal{T} \rangle$, contrary to assumption. Hence $Q \notin \langle \mathcal{T} \rangle$.

Proof of Lemma 4: Case $\mathcal{F}' = \langle H \circ \mathcal{E} \rangle$ or $\langle Z \circ \mathcal{E} \rangle$ □

For any function F of arity n and invertible matrix M, $F \in \langle M \circ \mathcal{E} \rangle$ iff $(M^{-1})^{\otimes n} F \in \langle \mathcal{E} \rangle$. (Note that $\langle M \circ \mathcal{E} \rangle = M \circ \langle \mathcal{E} \rangle$.) Since we will realize our function Q of a lower arity than F by connecting F with some unary functions, and since unary functions are transformed to other unary functions under any invertible holographic transformation, we only need to prove for $\langle \mathcal{E} \rangle$. Suppose $F \notin \langle \mathcal{E} \rangle$, and $\text{arity}(F) = n > 2$. 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 $S_{(\mathcal{I},\mathcal{A})}$. Each A_j is a 0-1 string of length $|I_j|$. In the definition for $S_{(\mathcal{I},\mathcal{A})}$ below we use the set $\{A_j, \bar{A}_j\}$ where A_j and its complement $\bar{A}_j$ play symmetric roles, and so we may normalize the first bit of A_j to be 0. We say a function R belongs to the set $S_{(\mathcal{I},\mathcal{A})}$ iff R has type $\mathcal{I}$, and $R = \bigotimes_{\mathcal{I}}(R_1, R_2, \ldots, R_k)$ for some functions $R_1, R_2, \ldots, R_k$ such that for any $j \in [k]$, $R_j(X|_{I_j})$ is zero if $X|_{I_j} \notin \{A_j, \bar{A}_j\}$. Thus, $R_j \in \mathcal{E}$ for each $j \in [k]$, and any function in $S_{(\mathcal{I},\mathcal{A})}$ belongs to $\langle \mathcal{E} \rangle$.

The set of functions in $\langle \mathcal{E} \rangle$ of arity n is the union of these finitely many function sets $S_{(\mathcal{I},\mathcal{A})}$. Obviously, functions in $S_{(\mathcal{I},\mathcal{A})}$ can be described by a finite system of polynomial equations (Lemma 5). Since $FU \in \langle \mathcal{E} \rangle$ for all $U = [x, y]$, by Lemma 2, there must exist one $S_{(\mathcal{I},\mathcal{A})}$, such that for any x, y, FU belongs to the same set $S_{(\mathcal{I},\mathcal{A})}$. In particular, both $F^{x_n=0}$ and $F^{x_n=1}$ belong to $S_{(\mathcal{I},\mathcal{A})}$.

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, there is exactly one pair of linearly independent tensor factors, without loss of generality, F_1 and K_1. Expressing K_i in terms F_i, for $i \geq 2$, there exists a nonzero constant c, such that $F^{x_n=1} = \bigotimes_{\mathcal{I}}(cK_1, F_2, \ldots, F_k)$.

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 $|I_1| - 1$ variables of K other than 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 $S_{(\mathcal{I},\mathcal{A})}$. For example, in the sum defining $Q(0, 0)$, only one (possibly) nonzero term $K(A_1 0)$ is involved because the first and last bits are both set to 0. The first column in the 2×2 matrix form of Q is F_1 at entries $A_1, \bar{A}_1$, and similarly the second column is from cK_1. Because F_1 and K_1 are linearly independent, and these are the only possible nonzero entries of K, we have $\det Q \neq 0$. We claim that $Q \notin \langle \mathcal{E} \rangle$. For otherwise, being non-degenerate, $Q \in \mathcal{E}$, then the support of K is contained in either $\{A_1 0, \bar{A}_1 1\}$ or $\{A_1 1, \bar{A}_1 0\}$. 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 4: Case $\mathcal{F}' = \langle Z\mathcal{M} \rangle$ □

Again we only need to prove for $\langle \mathcal{M} \rangle$. Suppose $F \notin \langle \mathcal{M} \rangle$, and $\text{arity}(F) = n > 2$. 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, each tensor factor is zero except on inputs of Hamming weight at most one. Each type $\mathcal{I}$ can be specified by a finite system of polynomial equations $E_{\mathcal{I}}$ by Lemma 5. Now we put $E_{\mathcal{I}}$ together with the requirement that for each tensor factor, all entries of Hamming weight greater than one are 0. This requirement can also be stated as a finite system of polynomial equations. We illustrate this point by the following simple case. Suppose we require that $(a_{i,j})$ is a tensor product $(b_i) \otimes (c_j)$, where $1 \le i \le n, 1 \le j \le m$, and for some subsets $B \subseteq [n]$, $C \subseteq [m]$, we require that $\forall i \in B, \forall j \in C, b_i = c_j = 0$. Then we include the equations for the type specification from Lemma 5 together with the following equations: $a_{i,j} = 0$ for all (i, j) such that $i \in B$ or $j \in C$. Note that these equations are on the entries of $a_{i,j}$. The equations from Lemma 5 imply that a tensor factorization $(a_{i,j}) = (b_i) \otimes (c_j)$ exists. If for all $i \in [n]$, $b_i = 0$, or if for all $j \in [m]$, $c_j = 0$, then $a_{i,j}$ is identically 0, and thus $a_{i,j}$ trivially has a tensor factorization $(b'_i) \otimes (c'_j)$ that satisfies the requirement $\forall i \in B, \forall j \in C, b'_i = c'_j = 0$. 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 $j \in C$. Thus the factorization $(a_{i,j}) = (b_i) \otimes (c_j)$ satisfies the requirement. (The salient point is that the polynomial equations are on the entries of $a_{i,j}$, through which we express a vanishing property on the desired tensor factors, on which we cannot directly specify using a polynomial.)

Applying Lemma 2, we conclude that there is one type $\mathcal{I}$ such that for any $U = [x, y]$, FU has a decomposition having the same type specified by $\mathcal{I}$ with tensor factors from $\mathcal{M}$.

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

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, say, 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 $K = \begin{bmatrix} F_1^{\mathrm{T}} \\ cK_1^{\mathrm{T}} \end{bmatrix}$, where the row index is $x_n = 0, 1$, columns are indexed by $\{0, 1\}^{|I_1|}$, and $c \neq 0$. Here the first row is $K^{x_n=0} = F_1^{\mathrm{T}}$. The second row is $K^{x_n=1} = cK_1^{\mathrm{T}}$. Columns are indexed by $A \in \{0, 1\}^{|I_1|}$. If the weight of A is greater than 1, then the column indexed by A is zero because $F_1, K_1 \in \mathcal{M}$. Let S_0 denote the column indexed by $0 \cdots 0 \in \{0, 1\}^m$, and S_i denote the column indexed by the bit sequence $A \in \{0, 1\}^m$, where only the ith bit of A is 1.

For simplicity of notation, 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, the only nonzero entry for K all have Hamming weight at most 1, and so $K \in \mathcal{M}$. This would imply $F \in \langle \mathcal{M} \rangle$, contrary to assumption. Hence there is a column S_i, $1 \le i \le m$, whose second entry is not zero. Without loss of generality 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, then $Q = K^{x_1=0,\ldots,x_{m-1}=0}$ is a binary function on $\{x_n, x_m\}$, and has the matrix form $Q = [S_0, S_m]$. Note that the index for the column S_m is $A = 0\cdots 01 \in \{0,1\}^m$. If S_0 is linearly dependent with S_m, then for some $1 \le j \le m-1$, S_j is linearly independent with S_m. Let $Q = K^{x_1=0,\ldots,x_{j-1}=0,x_j=[x,y],x_{j-1}=0,\ldots,x_{m-1}=0} = \left[xS_0 + yS_j, xS_m\right]$, where $x \neq 0$ and $y \neq 0$. (Here in Q, the row index is by x_n and column index is by x_m.) 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 of *not* belonging to any one of the four tractable classes, $\langle \mathcal{T} \rangle$, $\langle H \circ \mathcal{E} \rangle$, $\langle Z \circ \mathcal{E} \rangle$ or $\langle Z \circ \mathcal{M} \rangle$.

Recall that $F \cong cF$ for $c \neq 0$. In the following lemma, when we count the number of solutions, we count in terms of equivalence classes under $\cong$.

Lemma 7 *Suppose F is a ternary function. Then $F^{x_3=U} \cong 0$ for some $U \not\cong 0$ iff F has type $\{\{1,2\},\{3\}\}$. Suppose $F \notin \langle \mathcal{T} \rangle$. Then $F^{x_3=U} \not\cong 0$ 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 $\{\{1,2\},\{3\}\}$, then $F = T \otimes [a,b]$. If $a = b = 0$ then F is identically 0, and $F^{x_3=U} \cong 0$ for any unary U. If $[a,b] \not\cong 0$, then $F^{x_3=U} \cong 0$ for $U = [b, -a] \not\cong 0$. Conversely, if $F^{x_3=U} \cong 0$ for some $U \not\cong 0$, 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 $F^{x_3=U} \not\cong 0$ for any nonzero unary function U.

Let $F' = F^{x_3=U}$. Then F' is degenerate iff $\det \begin{bmatrix} F'(0,0) & F'(0,1) \\ F'(1,0) & F'(1,1) \end{bmatrix} = 0$. Let $U = [x, y]$, then the entries of F' are linear homogeneous polynomials of x and y, so the determinant equation is a quadratic homogeneous equation. It has either one or two solutions $U \not\cong 0$, or it is identically zero. We only need to prove the latter case contradicts $F \notin \langle \mathcal{T} \rangle$.

Suppose F' is degenerate for all U, then we have in particular $F^{x_3=0} = F_1 \otimes F_2$ and $F^{x_3=1} = K_1 \otimes K_2$. If $F^{x_3=0}$ and $F^{x_3=1}$ are linearly dependent, then $F \in \langle \mathcal{T} \rangle$, and so $F^{x_3=0}$ and $F^{x_3=1}$ are linearly independent. Then by Lemma 6, exactly one of the two pairs of functions $\{F_1, K_1\}$ and $\{F_2, K_2\}$ is linearly independent, say, $\{F_1, K_1\}$ is linearly dependent. Then F is the tensor product of F_1 and one binary function on the remaining two variables $\{x_2, x_3\}$, and 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 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 $U_1 \not\cong 0$, then $F^{x_1=U_1} = L \otimes R$, where L and R are unary functions on x_2 and x_3 respectively. Since $F^{x_1=U_1} \not\cong 0$, both $L \not\cong 0$ and $R \not\cong 0$. The matrix form of $F^{x_1=U_1}$ has rank exactly one. It follows that the decomposition $L \otimes R$ is unique under $\cong$. If we define $[x, y]^{\perp} = [y, -x]$, and let $U_2 = L^{\perp}$, then $F^{x_1=U_1,x_2=U_2}$ is identically 0. Given U_1 such that $F^{x_1=U_1}$ is degenerate but not identically 0, the property that $F^{x_1=U_1,x_2=U_2} \cong 0$ also uniquely determines a nonzero U_2 as $U_2 \cong L^{\perp}$, where L is uniquely determined by U_1. Because $(F^{x_2=U_2})^{x_1=U_1} \cong 0$, we have $F^{x_2=U_2}$ is degenerate. 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 for $U_1 \not\cong 0$ and $U_1' \not\cong 0$. 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.

The same statement holds for x_3.

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

$$\mathcal{U}_i = \{U \not\cong 0 \mid F^{x_i=U} \text{ is degenerate}\}, \qquad 1 \leq i \leq 3.$$

Lemma 8 *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 theorem for the hardness part of Theorem 2.

Theorem 4 *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 $P_H \notin \langle H \circ \mathcal{E} \rangle$ of arity 2;*
3. *For both $Z = Z_1$ or Z_2, some $P_Z \notin \langle Z \circ \mathcal{E} \rangle$ of arity 2; and*
4. *For both $Z = Z_1$ or Z_2, some $S_Z \notin \langle Z \circ \mathcal{M} \rangle$ 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 Fig. 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 remains unchanged if we rotate it $\frac{2\pi}{3}$. 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).

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.

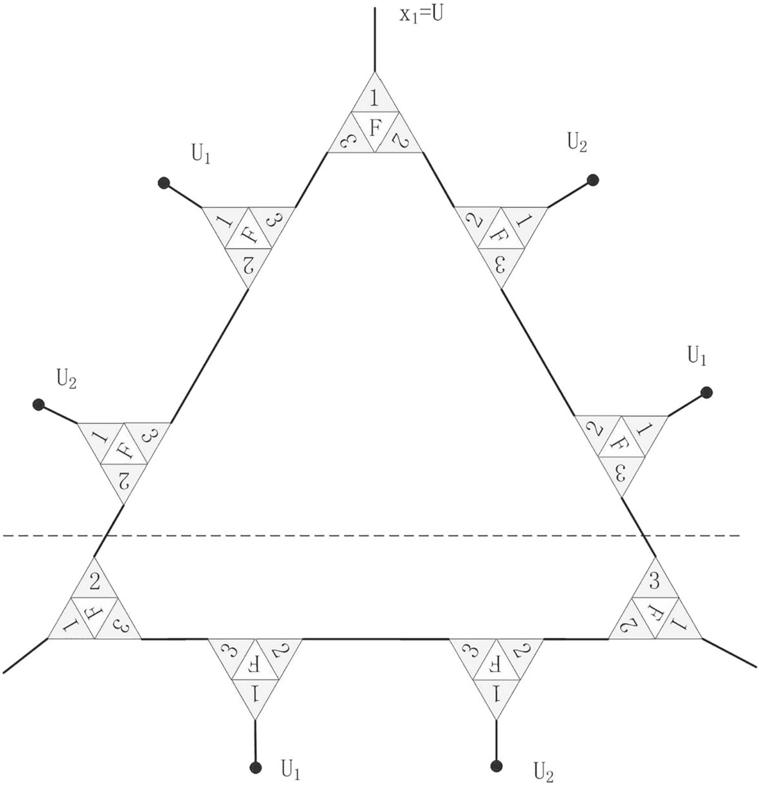


Fig. 4 Gadget to realize a symmetric ternary function

To prove this, we divide the gadget into two parts, as shown by the dashed line in Fig. 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 $U \not\cong 0$, then $S^{x_1=U} \not\cong 0$. 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 $F^{x_1=U'}_{x_2,x_3}$ denotes the 2×2 matrix form of $F^{x_1=U'}$ 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. 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 non-degenerate. By Lemma 7, 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 corners 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 non-singular.

If there exist U_1 and U_2 such that both properties hold, then for any nonzero unary function U, the vector form of $Q^{x_1=U}$ is the matrix-vector product $P_{x_2x_3,y_2y_3} S^{x_1=U}$, 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 non-singular matrix (the second property). This proves $Q \notin \langle \mathcal{T} \rangle$.

To establish the two properties, we can apply the Separation Lemma 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 should 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 path connecting the two copies of F on the lower two corners 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 non-singular 2×2 matrices.

Since $F \notin \langle \mathcal{T} \rangle$, by Lemma 7 there exists $U_1 \not\equiv 0$ 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 non-singular. In the notation of Lemma 8, by the 1-to-1 correspondence from $\mathcal{U}_1$ to $\mathcal{U}_3$, $U_1 \in \mathcal{U}_1$ gives L_L and then gives a corresponding $L_L^{\perp} \in \mathcal{U}_3$. Each one of (at most two) $U_1 \in \mathcal{U}_1$ gives a unique $L_L^{\perp} \in \mathcal{U}_3$. We want to choose a $U_1 \in \mathcal{U}_1$ such that its corresponding $L_L \notin \mathcal{U}_3$. By the 1-1 correspondence, this is equivalent to choosing some unary $U \in \mathcal{U}_3$, such that $U^{\perp} \notin \mathcal{U}_3$. Such a $U \in \mathcal{U}_3$, by the inverse map of the 1-1 correspondence, gives us the desired $U_1 \in \mathcal{U}_1$.

We have a 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 $R_R \notin \mathcal{U}_2$. Suppose $F^{x_1=U_1} = L_L \otimes R_L$ and $F^{x_1=U_2} = L_R \otimes R_R$, then we can replace them by unary functions, and combine $R_L^{\mathrm{T}} L_R$ to get a scalar factor c. Then $P_{x_2x_3,y_2y_3}$ is $cA_{x_2,y_2} \otimes B_{x_3,y_3}$, where A_{x_2,y_2} and B_{x_3,y_3} are the matrix forms for $F^{x_3=L_L}$ and $F^{x_2=R_R}$ respectively (Fig. 5). So we also want $c = R_L^{\mathrm{T}} L_R \neq 0$, in addition to A_{x_2,y_2} and B_{x_3,y_3} being non-singular.

Note that if we write in matrix form for P from left to right (see Figs. 4 and 5), we have another 4×4 matrix form for P as $P_{y_2x_2,y_3x_3} = F_{x_2x_1,x_3}(L_L R_L^{\mathrm{T}})(L_R R_R^{\mathrm{T}})F_{x_2,x_3x_1}$. Taking out the dot product value $c = R_L^{\mathrm{T}} L_R$ (a scalar), the remainder of the function is $(F_{x_2x_1,x_3} L_L)(R_R^{\mathrm{T}} F_{x_2,x_3x_1})$, a product of a 4×1 matrix with a 1×4 matrix. This matrix has rank 1. However this matrix form

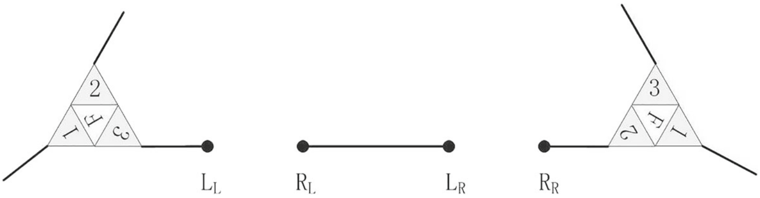


Fig. 5 Replace $F^{x_1=U_1}$ by $L_L \otimes R_L$, and $F^{x_1=U_2}$ by $L_R \otimes R_R$

for P is not the same as $P_{x_2x_3,y_2y_3}$, which is a rotated version. It is $P_{x_2x_3,y_2y_3}$ that we want to ensure that it has rank 4.

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 non-degenerate.
Condition (2): $F^{x_1=U_2} = L_R \otimes R_R$ is degenerate and $F^{x_2=R_R}$ is non-degenerate.
Condition (3): $R_L^{\mathrm{T}} L_R \neq 0$.

There are three cases, depending on $\mathcal{U}_3$, where one cannot pick U_1 to satisfy Condition (1). (We will deal with Conditions (2) and (3) separately.)

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

Now we will use the given binary signatures $P_H \notin \langle H \circ \mathcal{E}\rangle$, $P_Z \notin \langle Z \circ \mathcal{E}\rangle$, and $S_Z \notin \langle Z \circ \mathcal{M}\rangle$.

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

If $U \cong [1, \mathfrak{i}]$ (resp. $[1, -\mathfrak{i}]$), we show that $S_{Z_1}U$ (resp. $S_{Z_2}U$) does not belong to $\mathcal{U}_3$. Because the binary $S_{Z_1} \notin \langle Z_1 \circ \mathcal{M}\rangle$, in matrix form $S_{Z_1} = Z_1 T Z_1^{\mathrm{T}}$ for some $T \notin \langle \mathcal{M}\rangle$. Note that $\mathcal{U} \subset \mathcal{M}$, so that if the binary signature T is degenerate, then $T \in \langle \mathcal{M}\rangle$, a contradiction. Hence not only $T \notin \langle \mathcal{M}\rangle$, but also $T \notin \mathcal{M}$, that is $T(1,1) \neq 0$. Note that the matrix-vector product $Z_1^{\mathrm{T}} U = \begin{bmatrix} 1 & \mathfrak{i} \\ 1 & -\mathfrak{i} \end{bmatrix}\begin{bmatrix} 1 \\ \mathfrak{i} \end{bmatrix} \cong \begin{bmatrix} 0 \\ 1 \end{bmatrix}$. Then we calculate $U^{\mathrm{T}} S_{Z_1} U = U^{\mathrm{T}} Z_1 T Z_1^{\mathrm{T}} U \cong \begin{bmatrix} 0 & 1 \end{bmatrix} T \begin{bmatrix} 0 \\ 1 \end{bmatrix} = T(1,1) \neq 0$. Hence, $S_{Z_1}U \ncong [1, \mathfrak{i}]$, and $S_{Z_1}U \ncong [0, 0]$. In this case, $F^{x_3=S_{Z_1}U}$ is non-degenerate, and we will modify the construction in Fig. 4 by adding the binary gadget with signature S_{Z_1} to replace the three edges whose endpoints are small triangles both marked by 3 in the gadget. (This modification does change two edges in the construction of S; but it does not affect what has been proved for S, since S_{Z_1} is non-degenerate. The same is true for case (b.) and (c.) below.) The proof for $S_{Z_2}U$ is similar.

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

We show in this case, one of $P_{Z_1}[1, \mathfrak{i}]$ or $P_{Z_1}[1, -\mathfrak{i}] \notin \mathcal{U}_3$. Because $P_{Z_1} \notin \langle Z_1 \circ \mathcal{E}\rangle$, in matrix form $P_{Z_1} = Z_1 T Z_1^{\mathrm{T}}$ for some $T \notin \langle \mathcal{E}\rangle$. We claim that at least one of the two columns $\begin{bmatrix} e \\ f \end{bmatrix}$ of T have both entries nonzero. Otherwise, either T is degenerate or $T \in \mathcal{E}$, in either case $T \in \langle \mathcal{E}\rangle$, a contradictoion. If the first (resp. second) column has this property, $P_{Z_1}[1, -\mathfrak{i}]$ (resp. $P_{Z_1}[1, \mathfrak{i}]$) does not belong to $\mathcal{U}_3$. Indeed, if it is the first case, $P_{Z_1}\begin{bmatrix} 1 \\ -\mathfrak{i} \end{bmatrix} = Z_1 T Z_1^{\mathrm{T}} \begin{bmatrix} 1 \\ -\mathfrak{i} \end{bmatrix} \cong Z_1 T \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ \mathfrak{i} & -\mathfrak{i} \end{bmatrix}\begin{bmatrix} e \\ f \end{bmatrix} \ncong [1, \pm\mathfrak{i}]$. In this case, $F^{x_3=P_{Z_1}[1,-\mathfrak{i}]}$ (resp. $F^{x_3=P_{Z_1}[1,\mathfrak{i}]}$) is non-degenerate, and we will modify the construction in Fig. 4 by adding the binary gadget with signature P_{Z_1} to replace the three edges whose endpoints are small triangles both marked by 3 in the gadget.

In case (c.), we have $|\mathcal{U}_3| = 2$ and $\mathcal{U}_3 = \{U, U^{\perp}\}$. Hence $U \ncong U^{\perp}$, and U and $U^{\perp}$ are linearly independent. Then $U \not\perp U$, otherwise $U \cong 0$. Hence the dot product

$U^{\mathrm{T}}U \neq 0$, and we may assume $U = [a, b]$ and $U^{\perp} = [b, -a]$ are unit vectors: $a^2+b^2 = 1$. Let $H = \begin{bmatrix} a & b \\ b & -a \end{bmatrix}$, then H is an orthogonal matrix. Then it follows that one of $P_H[a, b]$ or $P_H[b, -a]$ does not belong to $\mathcal{U}_3$. The proof is similar with case (b.). In case (c.) we will modify the construction in Fig. 4 by adding the binary gadget with signature P_H to replace the three edges whose endpoints are small triangles both marked by 3 in the gadget.

The proof for Condition (2) is similar to Condition (1). The replacement in the construction of Fig. 4 happens at the three edges connecting the copy of F with U_2 and the corner F (the three edges whose endpoints are small triangles both marked by 2 in the gadget).

Now consider Condition (3). Suppose $R_L^{\mathrm{T}} L_R = 0$. We separate out the case $R_L \in \{[1, \mathfrak{i}], [1, -\mathfrak{i}]\}$, or not. If $R_L \in \{[1, \mathfrak{i}], [1, -\mathfrak{i}]\}$, and $R_L^{\mathrm{T}} L_R = 0$, then $R_L = L_R = [1, \mathfrak{i}]$ or $[1, -\mathfrak{i}]$. In this case $R_L^{\mathrm{T}} S_{Z_1} L_R \neq 0$ or $R_L^{\mathrm{T}} S_{Z_2} L_R \neq 0$ by a simple calculation as before. If $R_L \notin \{[1, \mathfrak{i}], [1, -\mathfrak{i}]\}$, and $R_L^{\mathrm{T}} L_R = 0$, then $L_R \notin \{[1, \mathfrak{i}], [1, -\mathfrak{i}]\}$ as well. We can assume $R_L = [a, b]$ and $L_R = [b, -a]$ and $a^2 + b^2 = 1$. Then one of $R_L^{\mathrm{T}} P_H L_R$ and $R_L^{\mathrm{T}} P_H^{\mathrm{T}} L_R$ is not zero, also by a simple calculation as before, where $H = \begin{bmatrix} a & b \\ b & -a \end{bmatrix}$. For Condition (3), the replacement in the construction of Fig. 4 happens at the three edges connecting the copy of F with U_1 with the copy of F with U_2 (the three edges whose endpoints are small triangles marked by 2 and 3 respectively in the gadget).

If Conditions (1) (2) (3) all hold, then the gadget satisfies the second property, and the theorem is proved. For each condition, if it does not hold, we have modified the gadget construction by adding some binary functions to rectify the construction, and these binary functions are available by the conditions of the theorem. With these modifications to the construction in Fig. 4, the proof of the theorem is complete. □

Remark In the proof of Theorem 4 we used the Separation Lemma to satisfy various conditions in isolation. In particular for the second property (for the arity 4 signature P in the lower part of the gadget), we argued that we *could* satisfy the property by choosing unary functions U_1 and U_2 to make $F^{x_1=U_1}$ and $F^{x_1=U_2}$ degenerate. When the final gadget is produced by simultaneously satisfying both the first condition (for S) and the second condition (for P), there is no expectation that this degeneracy will persist. In fact, the we proved that the first condition (for S) *could* be satisfied by choosing U_1 and U_2 to make $F^{x_1=U_1}$ and $F^{x_1=U_2}$ non-degenerate. There is some non-constructiveness by using the Separation Lemma to argue for the success of such constructions. If necessary, one can make it constructive (in the sense of Turing computability), by looking more closely at the sets of polynomial equations (which could be large, but specifiable by at most polynomially many bits). The reduction proved to exist is a polynomial time reduction, but the proof does not give it explicitly.

We will prove the #P-hardness part of Theorem 2 by appealing to Theorem 1 for symmetric Holant* problems. For that purpose we need to construct appropriate *symmetric* binary signatures as well.

Theorem 5 *Let $\mathcal{F}$ denote any one of the function sets $\langle H \circ \mathcal{E}\rangle$ (for an orthogonal matrix H), $\langle Z \circ \mathcal{E}\rangle$ or $\langle Z \circ \mathcal{M}\rangle$ (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 We define the *symmetric* binary function $Q = PF^{x_1=U}P^{\mathrm{T}}$ in matrix form, for some unary function U. This Q is realizable by a gadget consisting of a path of three signatures P, $F^{x_1=U}$ and $P^{\ \mathrm{T}}$. At both ends we have a copy of P and we connect the second variable of each copy of P to the two remaining variables of the symmetric $F^{x_1=U}$.

The essence of the proof is an appropriate holographic transformation. We are given $F \in \langle H\mathcal{E}\rangle$ (for an orthogonal matrix H), or $\langle Z \circ \mathcal{E}\rangle$ or $\langle Z \circ \mathcal{M}\rangle$. Let $M = H$ or Z depending on the cases. Let $F = M^{\otimes 3}F_1$ and $P = MP_1M^{\mathrm{T}}$ in matrix form. $F \notin \langle \mathcal{T}\rangle$ implies that $F_1 \notin \langle \mathcal{T}\rangle$. We also have $F^{x_1=U} = MF_1^{x_1=MU}M^{\mathrm{T}}$ in matrix form for the binary $F^{x_1=U}$. Let $Q_1 = P_1M^{\mathrm{T}}MF_1^{x_1=MU}M^{\mathrm{T}}MP_1^{\mathrm{T}}$, then

$$Q = MQ_1M^{\mathrm{T}} = MP_1M^{\mathrm{T}}MF_1^{x_1=MU}M^{\mathrm{T}}MP_1^{\mathrm{T}}M^{\mathrm{T}}.$$

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

We take $M = H$. Since H is orthogonal, $Q = HP_1F_1^{x_1=HU}P_1^{\mathrm{T}}H^{\mathrm{T}}$. We have $F_1 \notin \langle \mathcal{T}\rangle$. We also have $P_1 \notin \langle \mathcal{E}\rangle$, since $P \notin \langle H\mathcal{E}\rangle$. But by $F \in \langle H\mathcal{E}\rangle$, we have $F_1 \in \langle \mathcal{E}\rangle$. By the condition $F_1 \notin \langle \mathcal{T}\rangle$, we must actually have $F_1 \in \mathcal{E}$, because F_1 only has arity 3 and therefore any tensor factorization will put F_1 in $\langle \mathcal{T}\rangle$. Being symmetric and non-degenerate, $F_1 = [u, 0, 0, v]$, 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 $Q \notin \langle H \circ \mathcal{E}\rangle$. Because we can pick any $U' = HU$, for any x, y, we can realize $F_1^{x_1=HU} = [x, 0, y]$. Suppose $P_1 = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. Then $Q_1 = \begin{bmatrix} a^2x + b^2y & acx + bdy \\ acx + bdy & c^2x + d^2y \end{bmatrix}$.

We need $Q_1 \notin \langle \mathcal{E}\rangle$. This is translated into 3 conditions: (1) Q_1 is non-degenerate, (2) Q_1 is not of the form $[*, 0, *]$, and (3) Q_1 is not of the form $[0, *, 0]$. By the Separation Lemma, we only need to prove that there is some $[x, y]$ to satisfy each condition individually. If $x \neq 0$ and $y \neq 0$, then $F_1^{x_1=HU} = [x, 0, y]$ in non-degenerate. Also P_1 is non-degenerate because $P_1 \notin \langle \mathcal{E}\rangle$. Thus Q_1 is non-degenerate. Again because $P_1 \notin \langle \mathcal{E}\rangle$, either $ac \neq 0$ or $bd \neq 0$. There exists some $[x, y]$ such that $acx + bdy \neq 0$, thus Q_1 is not of the form $[*, 0, *]$. Similarly, it is easy find $[x, y]$ such that Q_1 is not of the form $[0, *, 0]$.

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

Take $M = Z$. Note that $Z^{\mathrm{T}}Z \cong (\neq_2)$. We have $Q \cong ZP_1(\neq_2)F_1^{x_1=ZU}(\neq_2)P_1^{\mathrm{T}}Z^{\mathrm{T}} = ZQ_1Z^{\mathrm{T}}$. 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 $\langle H \circ \mathcal{E}\rangle$.

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

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 $F_1 \in \langle \mathcal{M}\rangle$. Again because F_1 has arity 3, from these two conditions we conclude $F_1 \in \mathcal{M}$, since any tensor factorization for F_1 would place it in $\langle\mathcal{T}\rangle$. Being symmetric and non-degenerate, F_1 has the form $F_1 \cong [f, 1, 0, 0]$. Let $P_1 = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, and we have $P_1 \notin \langle\mathcal{M}\rangle$, since $P \notin \langle Z \circ \mathcal{M}\rangle$. For any x, there is some U, such that $F_1^{x_1=ZU} = [x, 1, 0]$. We only need to prove that $Q_1 = P_1(\neq_2)F_1^{x_1=ZU}(\neq_2)P_1^{\mathrm{T}} \notin \langle\mathcal{M}\rangle$.

Because $P_1 \notin \langle\mathcal{M}\rangle$, certainly P_1 is non-degenerate. Therefore $\det Q_1 \neq 0$, and Q_1 is non-degenerate. Also the entry $P_1(1, 1) = d \neq 0$, because $P_1 \notin \mathcal{M}$. It follows that there exists x such that $Q_1(1, 1) = 2cd + d^2x \neq 0$. Hence, $Q_1 \notin \langle\mathcal{M}\rangle$. □

Now we are ready to finish the proof Theorem 2.

Proof of Theorem 2 (#P-hardness part) Suppose $\mathcal{F} \not\subseteq \langle\mathcal{T}\rangle$, $\mathcal{F} \not\subseteq \langle H \circ \mathcal{E}\rangle$ for all orthogonal H, $\mathcal{F} \not\subseteq \langle Z \circ \mathcal{E}\rangle$ and $\mathcal{F} \not\subseteq \langle Z \circ \mathcal{M}\rangle$ for $Z = Z_1$ and Z_2. By Lemma 4, we can realize functions of arity of 2 or 3 *not* belonging to these function sets respectively.

The conditions in Theorem 4 are satisfied, so we can realize a symmetric ternary function $Q_3 \notin \langle\mathcal{T}\rangle$ (with the help of those binary functions). Certainly Q_3 is non-degenerate. If Holant$^*(\{Q_3\})$ is #P-hard, then the theorem is proved. Otherwise, by Theorem 1 for symmetric Holant* problems, Q_3 belongs to one of the special function families listed in the theorem. It can be shown that these are precisely restrictions of $\langle H \circ \mathcal{E}\rangle$, $\langle Z \circ \mathcal{E}\rangle$ or $\langle Z \circ \mathcal{M}\rangle$ to symmetric signatures. By Theorem 5, we can realize a symmetric binary function Q_2 not in this family. By Theorem 1 again, Holant$^*(\{Q_3, Q_2\})$ is #P-hard, and therefore Holant$^*(\mathcal{F})$ is also #P-hard. □

Acknowledgements We sincerely thank Miriam Backens, Xi Chen, Martin Dyer, Leslie Ann Goldberg, Zhiguo Fu, Heng Guo, Michael Kowalczyk, Les Valiant and Tyson Williams for discussions and comments. We also thank the anonymous referees of SODA 2011 who read our submission and made very constructive comments. Last but not least we sincerely thank the anonymous referees for the journal submission version. Many comments have been incorporated in this revised version with hopefully an improved presentation.

References

1. Backens, M.: A New Holant Dichotomy Inspired by Quantum Computation. In: 44Th International Colloquium on Automata, Languages, and Programming, ICALP 2017, Warsaw, pp. 16:1–16:14 (2017)
2. Backens, M.: A Complete Dichotomy for Complex-Valued Holant^c. In: 45Th International Colloquium on Automata, Languages, and Programming, ICALP 2018, pp. 12:1–12:14 (2018)
3. Baxter, R.J.: Exactly solved models in statistical mechanics. Academic Press, London (1982)
4. Bulatov, A.A.: The complexity of the counting constraint satisfaction problem. J. ACM **60**(5), 34:1–34:41 (2013)
5. Bulatov, A.A., Dyer, M.E., Goldberg, L.A., Jalsenius, M., Jerrum, M., Richerby, D.: The complexity of weighted and unweighted #CSP. J. Comput. Syst. Sci. **78**(2), 681–688 (2012)
6. Bulatov, A.A., Grohe, M.: The complexity of partition functions. Theor. Comput. Sci. **348**(2-3), 148–186 (2005)
7. Cai, J., Chen, X.: A Decidable Dichotomy Theorem on Directed Graph Homomorphisms with Non-Negative Weights. In: 51Th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2010, pp. 437–446. IEEE Computer Society, Las Vegas (2010)

8. Cai, J., Chen, X.: Complexity dichotomies for counting problems. Cambridge University Press, Cambridge (2017)
9. Cai, J., Chen, X., Lu, P.: Graph homomorphisms with complex values: a dichotomy theorem. SIAM J. Comput. **42**(3), 924–1029 (2013)
10. Cai, J., Guo, H., Williams, T.: The Complexity of Counting Edge Colorings and a Dichotomy for Some Higher Domain Holant Problems. In: 55Th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2014, pp. 601–610, Philadelphia (2014)
11. Cai, J., Lu, P.: Holographic algorithms: From art to science. J. Comput. Syst. Sci. **77**(1), 41–61 (2011)
12. Cai, J., Lu, P., Xia, M.: Dichotomy for Holant* problems of Boolean domain. In: Proceedings of the Twenty-Second Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2011, pp. 1714–1728, San Francisco (2011)
13. Cai, J., Lu, P., Xia, M.: Dichotomy for Holant* problems with domain size 3. In: Proceedings of the Twenty-Fourth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2013, pp. 1278–1295, New Orleans (2013)
14. Cai, J., Lu, P., Xia, M.: Dichotomy for real Holant^c problems. In: Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2018, pp. 1802–1821, New Orleans (2018)
15. Cai, J.Y., Govorov, A.: Perfect matchings, rank of connection tensors and graph homomorphisms. In: Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA ’19, pp. 476–495. Society for Industrial and Applied Mathematics, Philadelphia (2019)
16. Cai, J.Y., Lu, P., Xia, M.: 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. IEEE Computer Society, Washington (2008)
17. Cai, J.Y., Lu, P., Xia, M.: Holant Problems and Counting CSP. In: Mitzenmacher, M. (ed.) STOC, pp. 715–724. ACM (2009)
18. Creignou, N., Hermann, M.: Complexity of generalized satisfiability counting problems. Inf. Comput. **125**(1), 1–12 (1996)
19. Creignou, N., Khanna, S., Sudan, M.: Complexity Classifications of Boolean Constraint Satisfaction Problems. Siam Monographs On Discrete Mathematics And Applications (2001)
20. Dyer, M.E., Goldberg, L.A., Jalsenius, M., Richerby, D.: The Complexity of Approximating Bounded-Degree Boolean #CSP. In: Marion, J., Schwentick, T. (eds.) STACS, LIPIcs, vol. 5, pp. 323–334. Schloss Dagstuhl - Leibniz-Zentrum für Informatik (2010)
21. Dyer, M.E., Goldberg, L.A., Jerrum, M.: The complexity of weighted Boolean #CSP. SIAM J. Comput. **38**(5), 1970–1986 (2009)
22. Dyer, M.E., Goldberg, L.A., Jerrum, M.: An approximation trichotomy for Boolean #CSP. J. Comput. Syst. Sci. **76**(3-4), 267–277 (2010)
23. Dyer, M.E., Goldberg, L.A., Paterson, M.: On counting homomorphisms to directed acyclic graphs. J. ACM **54**(6), 27:1–27:23 (2007)
24. Dyer, M.E., Greenhill, C.S.: The complexity of counting graph homomorphisms. Random Struct. Algorithms **17**(3-4), 260–289 (2000)
25. Dyer, M.E., Richerby, D.: On the complexity of #CSP. In: Schulman, L.J. (ed.) Proceedings of the 42nd ACM Symposium on Theory of Computing, STOC 2010, pp. 725–734. ACM, Cambridge (2010)
26. Freedman, M., Lovász, L., Schrijver, A.: Reflection positivity, rank connectivity, and homomorphism of graphs. J. AMS **20**, 37–51 (2007)
27. Goldberg, L.A., Grohe, M., Jerrum, M., Thurley, M.: A complexity dichotomy for partition functions with mixed signs. SIAM J. Comput. **39**(7), 3336–3402 (2010)
28. Goldberg, L.A., Jerrum, M.: Approximating the partition function of the ferromagnetic potts model. J. ACM **59**(5), 25:1–25:31 (2012)
29. Hell, P., Nešetřil, J.: On the complexity of H-coloring. J. Combin. Theory Ser. B **48**(1), 92–110 (1990)
30. Ising, E.: Beitrag zur theorie des ferromagnetismus. Z. Phys. Hadrons Nucl. **31**(1), 253–258 (1925)
31. Jerrum, M., Sinclair, A.: Polynomial-time approximation algorithms for the ising model. SIAM J. Comput. **22**(5), 1087–1116 (1993)
32. Jerrum, M., Sinclair, A.: The Markov Chain Monte Carlo Method: an Approach to Approximate Counting and Integration. In: Approximation Algorithms for NP-Hard Problems, pp. 482–520. PWS Publishing (1996)
33. Kasteleyn, P.W.: The statistics of dimers on a lattice. Physica **27**, 1209–1225 (1961)

34. Kasteleyn, P.W.: Graph Theory and Crystal Physics. In: Harary, F. (ed.) Graph Theory and Theoretical Physics, pp. 43–110. Academic Press, London (1967)
35. Ladner, R.E.: On the structure of polynomial time reducibility. J. ACM **22**(1), 155–171 (1975)
36. Lee, T., Yang, C.: Statistical theory of equations of state and phase transitions. II. Lattice gas and Ising model. Phys. Rev. **87**(3), 410–419 (1952)
37. Lieb, E., Sokal, A.: A general Lee-Yang theorem for one-component and multicomponent ferromagnets. Commun. Math. Phys. **80**(2), 153–179 (1981)
38. Lin, J., Wang, H.: The Complexity of Holant Problems over Boolean Domain with Non-Negative Weights. In: 44Th International Colloquium on Automata, Languages, and Programming, ICALP 2017, pp. 29:1–29:14, Warsaw (2017)
39. Lovász, L.: Operations with structures. Acta Math. Hung. **18**, 321–328 (1967)
40. Madras, N., Randall, D.: Markov chain decomposition for convergence rate analysis. Ann. Appl. Probab. **12**(2), 581–606 (2002)
41. McCoy, B., Wu, T.: The two-dimensional Ising model. Harvard University Press, Cambridge (1973)
42. Onsager, L.: Crystal statistics. i. a two-dimensional model with an order-disorder transition. Phys. Rev. **65**(3-4), 117–149 (1944)
43. Randall, D.: Mixing. In: 44Th Symposium on Foundations of Computer Science (FOCS 2003), 11-14 October 2003, Cambridge, MA, USA, Proceedings, pp. 4–15 (2003)
44. Schaefer, T.J.: The complexity of satisfiability problems. In: Proceedings of the 10th Annual ACM Symposium on Theory of Computing, pp. 216–226, San Diego (1978)
45. Szegedy, B.: Edge coloring models and reflection positivity. J. Amer. Math. Soc. **20**, 969–988 (2007)
46. Temperley, H.N.V., Fisher, M.E.: Dimer problem in statistical mechanics c an exact result. Philos. Mag. **6**, 1061 C 1063 (1961)
47. Valiant, L.G.: Quantum circuits that can be simulated classically in polynomial time. SIAM J. Comput. **31**(4), 1229–1254 (2002)
48. Valiant, L.G.: Accidental algorthims. In: FOCS '06: Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science, pp. 509–517. IEEE Computer Society, Washington (2006). https://doi.org/10.1109/FOCS.2006.7
49. Valiant, L.G.: Holographic algorithms. SIAM J. Comput. **37**(5), 1565–1594 (2008). https://doi.org/10.1137/070682575
50. Welsh, D.: Complexity: knots, colourings and counting. Cambridge University Press, Cambridge (1993)
51. Yang, C.: The spontaneous magnetization of a two-dimensional Ising model. Phys. Rev. **85**(5), 808–816 (1952)
52. Yang, C., Lee, T.: Statistical theory of equations of state and phase transitions. I. Theory of condensation. Phys. Rev. **87**(3), 404–409 (1952)

Affiliations

Jin-Yi Cai[1] **· Pinyan Lu**[2] **· Mingji Xia**[3]

Pinyan Lu
lu.pinyan@mail.shufe.edu.cn

Mingji Xia
mingji@ios.ac.cn

[1] Computer Sciences Department, University of Wisconsin-Madison, Madison, WI, 53706, USA

[2] School of Information Management and Engineering, Shanghai University of Finance and Economics, Yangpu District, Shanghai, China

[3] Stake Key Laboratory of Computer Science, Institute of Software, Chinese Academy of Sciences, University of Chinese Academy of Sciences, Beijing, China