SIAM J. COMPUT.
Vol. 46, No. 3, pp. 853–889

HOLOGRAPHIC ALGORITHMS WITH MATCHGATES CAPTURE PRECISELY TRACTABLE PLANAR #CSP*

JIN-YI CAI†, PINYAN LU‡, AND MINGJI XIA§

Abstract. Valiant introduced matchgate computation and holographic algorithms. A number of seemingly exponential time problems can be solved by this novel algorithmic paradigm in polynomial time. We show that, in a very strong sense, matchgate computations and holographic algorithms based on them provide a universal methodology to a broad class of counting problems studied in the statistical physics community for decades. They capture precisely those problems which are #P-hard on general graphs but computable in polynomial time on planar graphs. More precisely, we prove complexity dichotomy theorems in the framework of counting CSP problems. The local constraint functions take Boolean inputs and can be arbitrary real-valued symmetric functions. We prove that *every* problem in this class belongs to precisely three categories: (1) those which are tractable (i.e., polynomial time computable) on general graphs, or (2) those which are #P-hard on general graphs but tractable on planar graphs, or (3) those which are #P-hard even on planar graphs. The classification criteria are explicit. Moreover, problems in category (2) are tractable on planar graphs precisely by holographic algorithms with matchgates.

Key words. holographic algorithms, counting problems, #P, dichotomy theorems

AMS subject classification. 68Q17

DOI. 10.1137/16M1073984

1. Introduction.

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

A closely related framework for locally constrained counting problems is called Holant problems [18, 20]. This framework is inspired by the introduction of *holographic algorithms* by Valiant [45, 44]. In two groundbreaking papers [43, 45] Valiant introduced matchgates and holographic algorithms based on matchgates to solve a

*Received by the editors May 6, 2016; accepted for publication (in revised form) January 3, 2017; published electronically May 11, 2017. The main results in this paper appeared in *proceedings of FOCS*, 2010.

http://www.siam.org/journals/sicomp/46-3/M107398.html

Funding: The first author was supported by NSF CCF-0511679, CCF-0830488, CCF-0914969, and CCF-1217549. The third author was supported by China National 973 program 2014CB340300 and NSFC 61003030.

†Department of Computer Sciences, University of Wisconsin–Madison, Madison, WI 53706 (jyc@cs.wisc.edu).

‡Corresponding author. Shanghai University of Finance and Economics, Shanghai, China (lu.pinyan@mail.shufe.edu.cn).

§State Key Laboratory of Computer Science, Institute of Software, Chinese Academy of Sciences, University of Chinese Academy of Sciences, Beijing, China (xmjljx@gmail.com).

number of problems in polynomial time, which appear to require exponential time. At the heart of these exotic algorithms is a tensor transformation from a given problem to the problem of counting (complex) weighted perfect matchings over planar graphs. The latter problem has a remarkable P-time algorithm (FKT-algorithm) [38, 29, 30]. Planarity is crucial, as counting perfect matchings over general graphs is #P-hard [40]. Most of these holographic algorithms use a suitable linear basis to realize locally a *symmetric* function with at most three Boolean variables on a matchgate. This work has been extended in [14]. In particular we have obtained a complete characterization of all realizable symmetric functions by matchgates over the complex field $\mathbb{C}$.

The study of "tractable #CSP" type problems has a much longer history in the statistical physics community (under different names). Ever since Wilhelm Lenz, who invented what is now known as the Ising model, asked his student Ernst Ising [27] to work on it, physicists have studied so-called exactly solved models [1, 35]. In the language of modern complexity theory, physicists' notion of an exactly solvable system corresponds to systems with polynomial time computable partition functions in spirit. This is captured completely by the computer science notion of "tractable #CSP" and "tractable Holant problems." In physics, many great researchers worked to build this intellectual edifice, with remarkable contributions by Ising, Onsager, C. N. Yang, T. D. Lee, Fisher, Temperley, Kasteleyn, Baxter, Lieb, Wilson, etc. [27, 36, 48, 49, 33, 38, 29, 30, 1, 34, 46]. A central question is to identify what "systems" can be solved "exactly" and what "systems" are "difficult." The basic conclusion from physicists is that some "systems," including the Ising model, are "exactly solvable" for planar graphs, but they appear difficult for higher dimensions. There does not exist any rigorous or provable classification. This is partly because the notion of a "difficult" partition function had no rigorous definition in physics. However, in the language of complexity theory, it is natural to consider the classification problem. In this paper we do that, in the general setting of #CSP with real-valued symmetric constraint functions. This will also shed light on why the valiant efforts by physicists to generalize the "exactly solved" planar system to higher dimensions failed. (In the appendix we will give some more background.)

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

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

Every #CSP problem can be simulated by a Holant problem. Represent any instance of a #CSP problem by a bipartite graph where left-hand sides (LHS) are labeled by variables and RHS are labeled by constraints. Denote by $=_k$: $\{0,1\}^k \to \{0,1\}$ the EQUALITY function of arity k, which is 1 on 0^k and 1^k, and is 0 elsewhere. Then we can turn the #CSP instance to an input graph of a Holant problem, by replacing every variable vertex v on the LHS by $=_{\deg(v)}$. Thus #CSP($\mathcal{F}$) can be simulated by Holant($\mathcal{F} \cup \{=_k \mid k \geq 1\}$). In fact, #CSP($\mathcal{F}$) is exactly the same as Holant($\mathcal{F} \cup \{=_k \mid k \geq 1\}$). To see this, suppose we are given an instance for Holant($\mathcal{F} \cup \{=_k \mid k \geq 1\}$). If two adjacent vertices u and v are both labeled by EQUALITY, then we can contract the edge (u, v) and merge them by a single vertex, labeled by an EQUALITY of arity $\deg(u) + \deg(v) - 2$. Repeat this step until there are no more adjacent vertices with EQUALITY functions. Now move all vertices with EQUALITY functions to the LHS and rename as variables, and keep all other vertices with labels from $\mathcal{F}$ on the RHS, we get an equivalent #CSP($\mathcal{F}$) instance. Thus, #CSP problems can be viewed as Holant problems where all EQUALITY functions are available for free, or assumed to be present. However, when we wish to discuss some restricted classes of counting problems, e.g., for 3-regular graphs, the framework of Holant problems is the more natural one. And as it turns out, the main technical breakthrough for our dichotomy theorem for planar #CSP comes from Holant problems.

In this paper we will only consider constraint functions on Boolean variables. For a symmetric function f on k variables $X = \{x_1, \ldots, x_k\}$, we denote it as $[f_0, f_1, \ldots, f_k]$, where f_i is the value of f on inputs of Hamming weight i, e.g., $(=_1) = [1, 1]$, $(=_2) = [1, 0, 1]$, and $(=_3) = [1, 0, 0, 1]$, etc. When we relax Holant problems by allowing all EQUALITY functions for free, we obtain #CSP. We can also consider other relaxations. Let $\mathbf{0} = [1, 0]$ and $\mathbf{1} = [0, 1]$ denote the unary (arity 1) pinning functions that set a variable to the constant values 0 and 1, respectively. Then Holant^c is the class of Holant problems where $\mathbf{0}$ and $\mathbf{1}$ are freely available. This amounts to computing Holant on input graphs where we can set 0 or 1 to some dangling edges (each end has degree 1). Another class of Holant problems is called Holant* problems, where we assume all unary functions $[u_0, u_1]$ are freely available.

In [20] we obtained a dichotomy theorem for (complex) Holant* problems and (real) Holant^c problems for all symmetric constraints. The dichotomy criterion for Holant* problems is still valid for *planar graphs.* The proofs of dichotomy theorems in this paper start from there.

In section 4, we prove that for any real-valued symmetric function set $\mathcal{F}$, the planar Holant^c($\mathcal{F}$) problem is tractable (i.e., computable in P) if *either* it is tractable over general graphs (for which we already have an effective dichotomy theorem [20]) *or* it is tractable because every function in $\mathcal{F}$ is realizable by a matchgate, in which case the planar Holant^c($\mathcal{F}$) problem is computable by matchgates in P-time using FKT. In *all other cases* the problem is #P-hard.[1] Thus, assuming P $\neq$ #P, the tractability criterion is a necessary and sufficient condition, i.e., a characterization. A crucial ingredient of the proof is a crossover construction whose validity is proved algebraically, which seems to defy any direct combinatorial justification.

Our second theorem (section 5) is about planar #CSP problems. We prove that for any set of real-valued symmetric functions $\mathcal{F}$, the planar #CSP($\mathcal{F}$) problem is tractable if *either* it is tractable as #CSP($\mathcal{F}$) without the planarity restriction (for which we have an effective dichotomy theorem [20]) *or* it is tractable because every

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

function in $\mathcal{F}$ is realizable by a matchgate under a specific holographic basis transformation. Thus planar #CSP($\mathcal{F}$) is solvable by a holographic algorithm in the second case. For all other $\mathcal{F}$ the problem is #P-hard. Thus, the tractability criterion is again a characterization assuming P $\neq$ #P. The proof of this dichotomy theorem for planar #CSP is built on the one for planar Holant^c in section 4.

Our third result is a dichotomy theorem for planar 2–3 regular bipartite Holant problems (section 6). (This theorem deals with Holant problems without assuming unary **0** and **1**.) This includes Holant problems for 3-regular graphs as a special case. The tractability criterion is the same: *Either* it is tractable for general graphs (for which we also have an effective dichotomy theorem [12]) *or* it is tractable by a suitable holographic algorithm, which is a holographic reduction to FKT using matchgates. In all other cases the problem is #P-hard.

The three dichotomy theorems are not mutually subsumed by each other and are of independent interest. In each framework the respective theorem is a demonstration that holographic algorithms with matchgates capture precisely those problems expressible within the framework that are #P-hard in general but become polynomial time tractable on planar graphs. Following the suggestion of a referee, to help readers better understand, we present a roadmap in section 7 explaining how some earlier work led to these dichotomy theorems. In an appendix we discuss some connections to statistical physics.

2. Preliminaries.

2.1. Problem and definitions.

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

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

A Holant problem is parameterized by a set of signatures.

DEFINITION 2.1. *Given a set of signatures $\mathcal{F}$, we define a counting problem* Holant($\mathcal{F}$)*:*

Input: A signature grid $\Omega = (G, \mathcal{F}, \pi)$*;*
Output: Holant_Ω.

Planar Holant problems are Holant problems on planar graphs.

DEFINITION 2.2. *Given a set of signatures $\mathcal{F}$, we define a counting problem* Pl-Holant($\mathcal{F}$)*:*

Input: A signature grid $\Omega = (G, \mathcal{F}, \pi)$*, where G is a planar graph;*
Output: Holant_Ω.

We would like to characterize the complexity of Holant problems in terms of signature sets.[2] For some $\mathcal{F}$, it is possible that Holant($\mathcal{F}$) is #P-hard, while Pl-Holant($\mathcal{F}$) is tractable. These new tractable cases make dichotomies for planar Holant problems more challenging. This is also the focus of this work. Some special families of Holant problems have already been widely studied. For example, if $\mathcal{F}$ contains all EQUALITY signatures $\{=_1, =_2, =_3, \ldots\}$, then this is exactly the weighted #CSP problem. Pl-#CSP denotes the restriction of #CSP to planar structures, i.e., the standard bipartite graphs representing the input instances of #CSP are planar. In [20], we also introduced the following two special families of Holant problems by assuming some signatures are freely available.

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

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

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

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

2.2. $\mathcal{F}$-gate and matchgate. A signature from $\mathcal{F}$ is a basic function which can be used at a vertex in an input graph. Instead of a single vertex, we can use graph fragments to generalize this notion. An $\mathcal{F}$-gate Γ is a tuple $(H, \mathcal{F}, \pi)$, where $H = (V, E, D)$ is a graph where the edge set consists of regular edges E and dangling edges D. Some nodes of degree 1 are designated as external nodes, and all other nodes are internal nodes; a dangling edge connects an internal node to an external node, while a regular edge connects two internal nodes. The labeling π assigns a function from $\mathcal{F}$ to each internal node. The dangling edges define variables for the $\mathcal{F}$-gate. (See Figure 1 for one example.) We denote the regular edges in E by $1, 2, \ldots, m$ and denote the dangling edges in D by $m+1, m+2, \ldots, m+n$. Then we can define a function for this $\mathcal{F}$-gate $\Gamma = (H, \mathcal{F}, \pi)$,

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

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

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

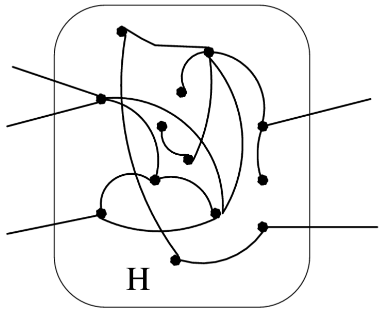

Fig. 1. *An $\mathcal{F}$-gate with five dangling edges.*

Using the idea of $\mathcal{F}$-gates, we can reduce one Holant problem to another in polynomial time. 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 note that even for a very simple signature set $\mathcal{F}$, the signatures for all $\mathcal{F}$-gates can be quite complicated and expressive. Matchgate signatures are an example. Matchgate is introduced by Valiant [43, 42, 45], whose definition is combinatorial in nature. Matchgates can be viewed as a special case of planar $\mathcal{F}$-gates, where $\mathcal{F}$ contains Exact-One functions of all arities and weight functions ($[1, 0, w]$, $w \in \mathbb{C}$) on edges. (Formally, we replace each matchgate edge of weight w by a path of length 2, and the new node on the path is assigned weight function $[1, 0, w]$.) The signature function Γ defined above for a matchgate is called a matchgate signature, or a standard signature. A signature function is realizable by a matchgate if it is the standard signature of that matchgate. (After a holographic transformation, a signature function is realizable under a basis if it is the transformed signature of a matchgate; see below.)

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

We use Holant$(\mathcal{G}|\mathcal{R})$ to denote all counting problems, expressed as Holant problems on bipartite graphs $H = (U, V, E)$, where each signature for a vertex in U or V is from $\mathcal{G}$ or $\mathcal{R}$, respectively. An input instance for the bipartite Holant problem is a bipartite signature grid and is denoted as $\Omega = (H, \mathcal{G}|\mathcal{R}, \pi)$. Signatures in $\mathcal{G}$ are denoted by column vectors (or contravariant tensors); signatures in $\mathcal{R}$ are denoted by row vectors (or covariant tensors) [22].

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

THEOREM 2.6 (Valiant's Holant theorem [45]). *Suppose there is a holographic reduction from* $\text{Holant}(\mathcal{G}|\mathcal{R})$ *to* $\text{Holant}(\mathcal{G}'|\mathcal{R}')$, *mapping a signature grid* Ω *to another* Ω', *then* $\text{Holant}_\Omega = \text{Holant}_{\Omega'}$.

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

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

2.4. Some known dichotomy results. In this subsection, we state some known dichotomy theorems. We first review three dichotomy theorems from [20].

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

1. *Every signature in* $\mathcal{F}$ *is of arity no more than two.*
2. *There exist two constants* a *and* b *(not both zero, depending only on* $\mathcal{F}$*) such that for every signature* $[x_0, x_1, \ldots, x_n] \in \mathcal{F}$ *one of the two conditions is satisfied:* (1) *for every* $k = 0, 1, \ldots, n-2$, *we have* $ax_k + bx_{k+1} - ax_{k+2} = 0$; (2) $n = 2$ *and the signature* $[x_0, x_1, x_2]$ *is of form* $[2a\lambda, b\lambda, -2a\lambda]$.
3. *For every signature* $[x_0, x_1, \ldots, x_n] \in \mathcal{F}$, *one of the two conditions is satisfied:* (1) *For every* $k = 0, 1, \ldots, n-2$, *we have* $x_k + x_{k+2} = 0$; (2) $n = 2$ *and the signature* $[x_0, x_1, x_2]$ *is of form* $[\lambda, 0, \lambda]$.

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

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

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

Then $\text{Holant}^c(\mathcal{F})$ *is computable in polynomial time if* (1) *After removing unary signatures from* $\mathcal{F}$, *it falls in one of the three classes of Theorem* 2.7 *(this implies* $\text{Holant}^*(\mathcal{F})$ *is computable in polynomial time) or* (2) *(without removing any unary signature but replacing each degenerate signature* $[x,y]^{\otimes k} \in \mathcal{F}$ *by* $[x,y]$*)* $\mathcal{F} \subseteq \mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3$. *Otherwise,* $\text{Holant}^c(\mathcal{F})$ *is #P-hard.*

Here we explicitly list all the real signatures in $\mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3$, up to an arbitrary scalar factor:

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

6. ($\mathcal{F}_3$): $[1, 1, -1, -1, 1, 1, -1, -1, \ldots, 1 \text{ (or } -1)]$,
7. ($\mathcal{F}_3$): $[1, -1, -1, 1, 1, -1, -1, 1, \ldots, 1 \text{ (or } -1)]$.

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

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

where $\lambda \in \mathbb{C}$, $X = (x_1, x_2, \ldots, x_k, 1)$, and χ is a 0–1 indicator function such that $\chi_{[AX=0]}$ is 1 iff $AX = 0$. Here the notation $\langle \alpha, X \rangle$ denotes an integer value $0, 1 \in \mathbb{Z}$ by performing a dot product on α and X over $\mathbb{Z}_2$. The sum over j on the exponent of $i = \sqrt{-1}$ is an integer sum in $\mathbb{Z}$, or $\mathbb{Z}_4$ (but not in $\mathbb{Z}_2$). We use $\mathcal{A}$ to denote the set of all affine functions.

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

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

As we mentioned in [20], the class $\mathcal{A}$ is a natural generalization of the family of symmetric signatures $\mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3$. Indeed one can show that $\mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3$ consists of precisely all scalar multiples of unary or nondegenerate symmetric signatures in $\mathcal{A}$.

The following dichotomy for 2–3 regular graphs is from [32].

Theorem 2.11. (see [32]) *The problem Holant($[y_0, y_1, y_2]|[1, 0, 0, 1]$) is #P-hard for all $y_0, y_1, y_2 \in \mathbb{C}$ except in the following cases, for which the problem is in P:* (1) *$y_1^2 = y_0 y_2$;* (2) *$y_0^{12} = y_1^{12}$ and $y_0 y_2 = -y_1^2$ ($y_1 \neq 0$);* (3) *$y_1 = 0$;* (4) *$y_0 = y_2 = 0$. If we restrict the input to planar graphs, then these four categories are tractable in P, as well as a fifth category $y_0^3 = y_2^3$, and the problem remains #P-hard in all other cases.*

2.5. Characterization of realizable signatures by matchgates. A signature f is said to satisfy the *even parity condition* if the value f at any input of odd Hamming weight is 0. It satisfies the *odd parity condition* if the value f at any input of even Hamming weight is 0. It satisfies the *parity condition* if it satisfies either the even parity condition or the odd parity condition.

A matchgate is called even (respectively, odd) if it has an even (respectively, odd) number of vertices. A matchgate signature satisfies the parity condition. The following two lemmas are from [9].

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

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

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

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

In particular, any symmetric signature of arity at most 3 that satisfies the parity condition is the standard signature of a matchgate.

In [17], we characterized all symmetric signatures realizable by matchgates under a given basis. Here we state the theorem for a particular basis $\left[\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}\right]$, which will be used in Theorem 5.1.

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

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

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

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

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

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

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

3. Polynomial interpolation. In this section, we discuss the interpolation method we will use in this paper. Polynomial interpolation is a powerful tool in the study of counting problems initiated by Valiant [41] and further developed by Vadhan [39], Dyer and Greenhill [24], and others. The method we use here is essentially the same as Vadhan's [39].

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

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

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

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

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

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

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

In this paper, the sequence $\{G_s\}$ will be constructed recursively using suitable gadgetry. There are two gadgets in a recursive construction: one gadget has arity 1, giving the initial signature $g = [x_0, y_0]$; the other has arity 2, giving the recursive

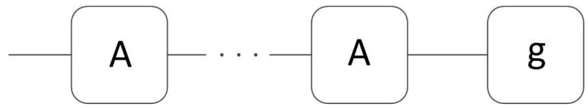

FIG. 2. *Recursive construction.*

iteration. It is more convenient to use a 2×2 matrix A to denote it. So we can recursively connect them as in Figure 2 and get $\{G_s\}$.

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

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

where $A = \left[\begin{smallmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{smallmatrix}\right]$ and $g = \left[\begin{smallmatrix} x_0 \\ y_0 \end{smallmatrix}\right]$.

We call this gadget pair (A, g) a recursive construction. The next lemma follows from Lemma 6.1 in [39].

LEMMA 3.1. *Let α, β be the two eigenvalues of A. If the three conditions*
1. $\det(A) \neq 0$*;*
2. *g is not a column eigenvector of A (nor the zero vector);*
3. *α/β is not a root of unity*

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

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

Starting from section 4 we will present the three dichotomy theorems of this paper. Following the suggestion of a referee, to aid the readers, we will present a "roadmap" in section 7, explaining how some earlier work led to these dichotomy theorems, and what would help the readers to understand more easily the details of what is done here. We will present a "revisionist" perspective, taking into account some understanding that has been gained only *after* the results of this paper appeared in [13]; some of this understanding is influenced by the present paper itself. In order not to interrupt the flow we will present this "roadmap" after the three dichotomy theorems are proved. However, readers may wish to look at that before reading through the proofs.

4. Dichotomy for planar Holant^c problems.

Before presenting the dichotomy theorem for planar Holant^c problems of this paper, Theorem 4.5, we prove the following theorem, which is a special case of Theorem 4.5 and plays a crucial role in the proof of the more general Theorem 4.5.

THEOREM 4.1. *Let $a, b \in \mathbb{R}$.*
- *If $ab \neq 1$ then Pl-Holant$^c([a, 0, 1, 0, b])$ is #P-hard.*
- *If $ab = 1$ then Pl-Holant$^c([a, 0, 1, 0, b])$ is solvable in P.*

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

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

Proof. Let $f = [a, 0, 0, 0, b]$. First, we show how to realize $(=_6) = [1, 0, 0, 0, 0, 0, 1]$ by f. The signature f has arity 4 and can be attached to a vertex of degree 4. We can take two copies of f; let it be $f(x_1, x_2, x_3, x_4)$ and $f(y_1, y_2, y_3, y_4)$. We connect three pairs of edges to realize a binary function $g(x,y) = \sum_{u,v,w=0,1} f(x,u,v,w)f(w,v,u,y)$. This is a planar gadget construction, resulting in a symmetric binary signature $g = [a^2, 0, b^2]$.

If $a^2 = b^2$, then we connect one pair of edges from two copies of f and get $\sum_{z=0,1} f(x_1,x_2,x_3,z)f(z,y_1,y_2,y_3)$, which is the symmetric signature $[a^2,0,0,0,0,0,b^2]$. This is the same as $(=_6) = [1,0,0,0,0,0,1]$ after factoring out the nonzero factor $a^2 = b^2$.

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

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

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

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

Using $(=_3) = [1,0,0,1]$, we can realize the EQUALITY function $=_k$ of any arity $k \geq 3$ by a tree gadget composed of $k-2$ many $(=_3)$ nodes. Then we can realize $[1,1,x^{2k}]$ for all $k \geq 1$ as in Figure 3. There are k parallel paths in the figure. The signature at each vertex of degree 2 on the parallel paths is $[1,1,x]$ and the signatures at the two vertices on both sides are $=_{k+1}$. If $x = 0$, then we already have $[1,1,0]$.

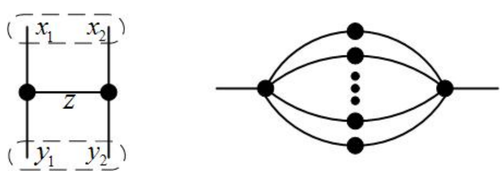

FIG. 3. *The gadget for function H and $[1,1,x^{2k}]$.*

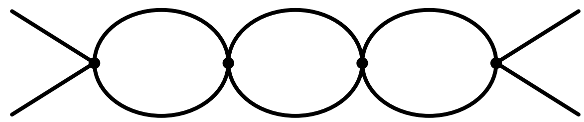

FIG. 4. *The gadget for* H_4.

Suppose $x \neq 0$. Because $x^2 \neq 1$ and being a positive real number, we can realize $[1,1,0]$ by interpolation. Now we have reduced the problem Pl-Holant($[1,0,0,1] \mid [1,1,0]$) to Pl-Holant^c($\{[1,0,0,1],[1,1,x^2]\}$). The bipartite problem Pl-Holant($[1,0,0,1] \mid [1,1,0]$) is #P-hard since it is counting VERTEX COVERS on planar 3-regular graphs [47]. (This is also a consequence of Theorem 2.11.) ∎

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

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

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

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

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

and for $i \geq 1$,

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

The gadget for H_{2i} is composed of $2i$ functions F. As an example, the gadget for H_4 is shown in Figure 4.

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

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

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

Note that the Holant value on G_i is

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

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

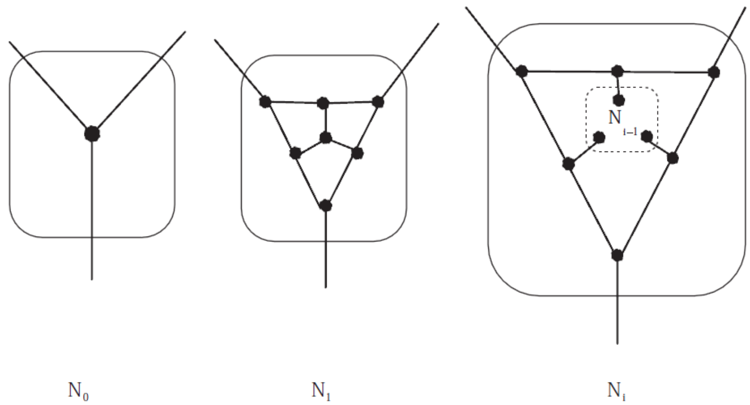

FIG. 5. *The recursive construction. The signature of every vertex in the gadget is* $[0, 1, 0, a]$.

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

The following result can be proved by interpolation as well.

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

Proof. From either $[0, 1, 0, a]$ or $[a, 0, 1, 0]$, we can get $[0, 1, 0]$ by pinning. Using three copies of the binary DISEQUALITY function $(\neq_2) = [0, 1, 0]$ we can flip all three variables of a ternary function f, and transform between $[f_0, f_1, f_2, f_3]$ and its reversal $[f_3, f_2, f_1, f_0]$. Hence, we only need to prove how to interpolate $[0, y, 0, x]$ from $[0, 1, 0, a]$. The recursive construction is depicted by Figure 5. By a simple parity argument, every $\mathcal{F}$-gate N_i has a signature of the form $[0, x_i, 0, y_i]$. After some calculation, we see that they satisfy the following recursive relation:

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

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

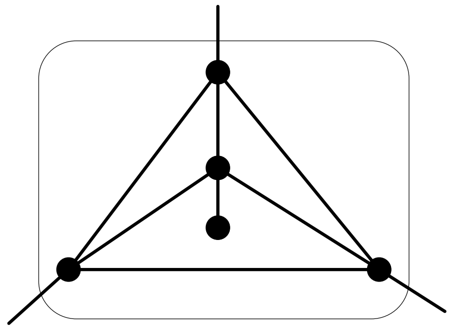

FIG. 6. *The signature of the degree* 1 *vertex in the gadget is* $[1,0]$ *or* $[0,1]$.

Lemma 3.1 and can be used to interpolate all signatures of the form $[0,x,0,y]$. This completes the proof. ∎

Proof of Theorem 4.1. If $ab=1$, then $[a,0,1,0,b]$ is realizable by some matchgate, by Lemma 2.12. This realizability also holds for the unary functions $[1,0]$ and $[0,1]$. Hence the problem Pl-Holant$^c([a,0,1,0,b])$ can be solved in polynomial time by matchgate computation via the FKT method [38, 29, 30]. In the following we assume that $ab \neq 1$ and prove that the problem is #P-hard. The case $a=b=0$ is proved in Lemma 4.3. Now we can assume at least one of a and b is nonzero, and by symmetry we assume $a \neq 0$. ∎

We know from our dichotomy for Holant^c problems [20], Theorem 2.8, that Holant$^c([a,0,1,0,b])$ for general graphs is #P-hard unless $a=b=1$ or $a=b=-1$, in which cases it is tractable. Both of these tractable cases are also included in the tractable cases ($ab=1$) here for Pl-Holant^c. Therefore, if we can realize a *cross function* X of arity 4 with a planar gadget when $ab \neq 1$, we can reduce Holant$^c([a,0,1,0,b])$ for general graphs to Pl-Holant$^c([a,0,1,0,b])$ and finish the proof. Here a cross function X satisfies

$$X_{0000}=X_{0101}=X_{1010}=X_{1111}=1 \quad \text{and} \quad X_\alpha = 0 \text{ for all other inputs } \alpha \in \{0,1\}^4.$$

If $\{a,b\} \not\subset \{-1,0,1\}$, we can apply the pinning functions $[1,0]$ and $[0,1]$ on $[a,0,1,0,b]$, and use Lemma 4.4 to interpolate all $[x,0,y,0]$, for $x,y \in \mathbb{C}$. If $\{a,b\} \subset \{-1,0,1\}$, then there are only four cases because $a \neq 0$ and $ab \neq 1$: $[1,0,1,0,-1]$, $[1,0,1,0,0]$, $[-1,0,1,0,1]$, and $[-1,0,1,0,0]$. In all four cases, it is easy to verify that we can realize a signature of the form $[c_1,0,c_2,0]$ where $c_1c_2 \neq 0$ and $c_1 \neq \pm c_2$ using the gadget in Figure 6. (For $[1,0,1,0,-1]$, we get $[8,0,4,0]$ by using $[1,0]$ for the degree 1 vertex in the gadget; for $[1,0,1,0,0]$, we get $[8,0,5,0]$ by using $[1,0]$; for $[-1,0,1,0,1]$, we get $[0,4,0,8]$ by using $[0,1]$; and for $[-1,0,1,0,0]$, we get $[0,1,0,3]$ by using $[0,1]$.) After factoring out a nonzero factor, we have a signature of the form $[c',0,1,0]$, or its reversal, where $c' \in \mathbb{R}$ and $c' \notin \{0,\pm 1\}$. As a result, by Lemma 4.4, we can also interpolate all $[x,0,y,0]$, where $x,y \in \mathbb{C}$.

Now we can use all signatures of the form $[x,0,y,0]$, for arbitrary $x,y \in \mathbb{C}$, to build new gadgets. We also have all $[x,0,y]$ by connecting $[x,0,y,0]$ to a $[1,0]$. By connecting a $[\sqrt[4]{t/a}, 0, \sqrt[4]{a/t}]$ to each edge of the signature $[a,0,1,0,b]$, we get $[t,0,1,0,\frac{c}{t}]$ for all $t \neq 0$, where $c = ab \neq 1$. Using all these, we will build a planar gadget in Figure 7 to realize the cross function X. In the equations below x,y,t are

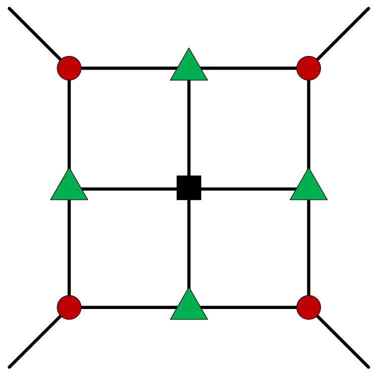

FIG. 7. *This gadget is used to realize the cross function. The signature for the center vertex (black square) is* $[t, 0, 1, 0, \frac{c}{t}]$. *The signature for the vertices in the four corners (red circles) is* $[x, 0, 1, 0]$. *The signature for the vertices in the middle of the boundaries (green triangles) is* $[y, 0, 1, 0]$.

three variables we can set to any complex numbers, with $t \neq 0$. The parameter c is given and not equal to 1.

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

We can compute the signature of the gadget in Figure 7. If the input (the four bits given to the four external dangling edges) has an odd number of 1s, then the value is 0. This is because all nine signatures satisfy the even parity condition. For other inputs, note that the gadget is rotationally symmetric. A calculation shows that

$$X_{0000} = x^4y^4t + t + 4x^3y^2 + 4x + 4x^2y + \frac{2cx^2}{t}, \tag{A}$$

$$X_{1111} = 2y^2t + 12y + \frac{2c}{t}, \tag{B}$$

$$X_{0101} = X_{1010} = 2xy^2t + 4x^2y^2 + 4 + 4xy + \frac{2cx}{t}, \tag{C}$$

$$X_{0011} = X_{1001} = X_{1100} = X_{0110} = x^2y^3t + yt + 3x^2y^2 + 3 + 6xy + \frac{2cx}{t}. \tag{D}$$

The construction of the cross function X will succeed if we can prove the following:

> For any $c \neq 1$, we can assign suitable complex values to x, y, and t, where $t \neq 0$, such that $A = B = C \neq 0$ and $D = 0$, where A, B, C, and D denote respectively the four functions of x, y, and t listed in the four lines above.

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

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

has a solution $x \in \mathbb{C}$ such that $x \notin \{0, +1, -1\}$. This x satisfies

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

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

To verify (4) we have

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

∎

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

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

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

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

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

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

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

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

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

Proof.

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

Here we used (5) and (4). ∎

Now we will set $t = 4/(1+xy)^2$. Since $z = xy \neq -1$, this t is well defined. Clearly $t \neq 0$. We next verify that $D = 0$. By (5) and (6) we get

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

Then

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

Thus

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

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

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

Hence

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

where the last step uses (5).

The next task is to show that $B = C$. We have

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

Hence

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

Finally we verify $A = C$ as well:

$$\begin{aligned} A &= (x^4y^4+1)t + x[C - 2xy^2t] = C + (x-1)C - 2x^2y^2t + (x^4y^4+1)t \\ &= C - 4(1-xy)^2 + t(x^2y^2-1)^2 = C. \end{aligned}$$

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

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

1. Holant$^c(\mathcal{F})$ *is tractable (for which we have an effective dichotomy* [20]*, Theorem* 2.8*); or*
2. *Every signature in $\mathcal{F}$ is realizable by some matchgate (for which we have a complete characterization* [9]*).*

Before giving the proof of Theorem 4.5, we normalize the signature set $\mathcal{F}$. If $\mathcal{F}$ contains any identically zero signature we can remove it from $\mathcal{F}$. This is because if it appears in any signature grid then the Holant value is 0, and the removal of it from $\mathcal{F}$ does not change the complexity of Pl-Holant$^c(\mathcal{F})$, nor the validity of the tractability criterion. So we may assume no signature in $\mathcal{F}$ is identically zero. Since $[1, 0]$ and $[0, 1]$ are freely available, we can construct any subsignature of a given signature. From any degenerate (but not identically zero) signature $[x, y]^{\otimes k}$, we can get the unary subsignature $x^{k-1}[x, y]$ and $y^{k-1}[x, y]$ by pinning. Either x or y is nonzero, hence we can get $[x, y]$. On the other hand, one can easily get $[x, y]^{\otimes k}$ from $[x, y]$; therefore we can replace any degenerate signature $[x, y]^{\otimes k}$ in $\mathcal{F}$ by $[x, y]$, without changing the complexity of Holant$^c(\mathcal{F})$ and Pl-Holant$^c(\mathcal{F})$ or the validity of the tractability criterion. So in the following we assume that all signatures in $\mathcal{F}$ of arity greater than 1 are nondegenerate.

The main idea of the proof is to interpolate all unary functions. If we can do that, we can reduce the problem Pl-Holant$^*(\mathcal{F})$ to Pl-Holant$^c(\mathcal{F})$ and finish the proof. We note that the dichotomy criterion in Theorem 2.7 for Holant$^*(\mathcal{F})$ is also valid for planar graphs. In some cases, we cannot interpolate all unary functions; then we prove the theorem separately, mainly using Lemma 4.2 and Theorem 4.1. The following lemma is for interpolation of unary functions.

LEMMA 4.6. *If* Pl-Holant$^c(\mathcal{F} \cup \{[a, b, c]\}) \le_{\mathrm{T}}$ Pl-Holant$^c(\mathcal{F})$*, where the binary signature* $[a, b, c]$ *satisfies the condition* $b^2 \neq ac$*,* $b \neq 0$*, and* $a + c \neq 0$*, then we can interpolate all unary functions,*

$$\text{Pl-Holant}^c(\mathcal{F} \cup \mathcal{U}) \le_{\mathrm{T}} \text{Pl-Holant}^c(\mathcal{F}).$$

Remark. If the condition of Lemma 4.6 holds, then Pl-Holant$^c(\mathcal{F}) \equiv_{\mathrm{T}}$ Pl-Holant^c $(\mathcal{F} \cup \mathcal{U}) \equiv_{\mathrm{T}}$ Pl-Holant$^*(\mathcal{F})$. Since the dichotomy criterion in Theorem 2.7 for Holant* problems also holds for Pl-Holant* problems, we have proved a dichotomy for Pl-Holant$^c(\mathcal{F})$ under this condition. We argue that the dichotomy criterion stated in Theorem 4.5 is indeed the correct criterion in this case. If P = #P, then stating that Pl-Holant$^c(\mathcal{F})$ is in P or is #P-hard are the same thing, and Theorem 4.5 is formally valid. Now we suppose P $\neq$ #P. We have the equivalence Pl-Holant$^c(\mathcal{F}) \equiv_{\mathrm{T}}$ Pl-Holant$^*(\mathcal{F})$ by Lemma 4.6. We can apply Theorem 2.7, the Holant* dichotomy. If $\mathcal{F}$ satisfies the tractability criterion of Theorem 2.7 then $\mathcal{F}$ also satisfies item 1 of Theorem 4.5 (which refers to Theorem 2.8 of which its item (1) is satisfied.) If $\mathcal{F}$ does not satisfy the tractability criterion of Theorem 2.7 then Pl-Holant$^c(\mathcal{F})$ is #P-hard by the equivalence and Theorem 2.7. Then we claim that $\mathcal{F}$ must not satisfy the tractability criterion of Theorem 4.5. Otherwise, by the tractability part of Theorem 4.5, we would have concluded that Pl-Holant$^c(\mathcal{F})$ is in P, a contradiction since P $\neq$ #P.

There is a subtlety in the argument above. Suppose $\mathcal{F}$ fails item (1) of Theorem 2.8 (the tractability criterion of Theorem 2.7). It is true that Pl-Holant$^c(\mathcal{F})$ $\equiv_{\mathrm{T}}$ Pl-Holant$^c(\mathcal{F} \cup \mathcal{U})$, and $\mathcal{U} \not\subseteq \mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3$. Thus the signature set $\mathcal{F} \cup \mathcal{U}$ also fails item (2) of Theorem 2.8. As well, not all of $\mathcal{U}$ are matchgates realizable, and so $\mathcal{F} \cup \mathcal{U}$ fails item 2 of Theorem 4.5. But this argument is deficient, because it applies the criteria of Theorem 4.5 on the set $\mathcal{F} \cup \mathcal{U}$, but not on $\mathcal{F}$ alone. Ideally one should prove directly that when we can construct or interpolate $[a, b, c]$ in Pl-Holant$^c(\mathcal{F})$ satisfying the conditions of Lemma 4.6, the set $\mathcal{F}$ cannot be a subset of $\mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3$ nor matchgates realizable. But we do not have such a direct proof. What we presented is an "end run" around it, based on a complexity theoretic argument accounting for both logical alternatives of the P = #P question.

Proof of Lemma 4.6. We use the interpolation method as described in section 3. Consider the recursive construction $\left(\left[\begin{smallmatrix} a & b \\ b & c \end{smallmatrix}\right], \left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right]\right)$. We use A to denote $\left[\begin{smallmatrix} a & b \\ b & c \end{smallmatrix}\right]$. Since $b^2 \neq ac$, A is nondegenerate, the first condition of Lemma 3.1 is satisfied. If $\left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right]$ is a column eigenvector of A, then $b = 0$, a contradiction. So it satisfies the second condition of Lemma 3.1. Since A is a real symmetric matrix, both its eigenvalues are real. If the ratio of two real numbers is a root of unity, they must be the same or opposite to each other. If the two eigenvalues are the same, since a real symmetric matrix is diagonalizable, this would imply that A is a scalar matrix, and we have $b = 0$ and $a = c$, a contradiction. If the two eigenvalues are opposite to each other, then we have the trace $a + c = 0$, also a contradiction. Therefore, the third condition of Lemma 3.1 is also satisfied. This completes the proof. ∎

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

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

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

All signatures here have arity at least 1. The signatures in $\mathcal{G}_1$ are the (nondegenerate) Generalized Equalities. The set $\mathcal{G}_2$ consists of signatures satisfying the parity condition. The signatures in $\mathcal{G}_3$ are those that satisfy the second order linear recurrence $x_{i+2} + x_i = 0$, which is the same recurrence satisfied by the signatures in $\mathcal{F}_3$, having characteristic equation $X^2 + 1 = 0$ with roots $\pm\sqrt{-1}$.

We note that $\mathcal{G}_1$, $\mathcal{G}_2$, and $\mathcal{G}_3$ are supersets of (not identically 0 signatures of) $\mathcal{F}_1$, $\mathcal{F}_2$, and $\mathcal{F}_3$, respectively. (Here we only consider the real part of $\mathcal{F}_2$; the nonreal part of $\mathcal{F}_2$ is not a subset of $\mathcal{G}_2$.) Furthermore the matchgate signature set $\mathcal{M}$ is sandwiched between (the real part of) $\mathcal{F}_2$ and $\mathcal{G}_2$, i.e., (the real part of) $\mathcal{F}_2 \subseteq \mathcal{M} \subseteq \mathcal{G}_2$. The following several lemmas are what we call a "squeeze." They successively narrow the scope of possible $\mathcal{F}$ for which we have to prove Theorem 4.5. They all have the form, for some class $\mathcal{C}$, "If $\mathcal{F} \not\subseteq \mathcal{C}$, then the conclusions of Theorem 4.5 hold." After proving each lemma, in subsequent lemmas, we only need to consider the case that $\mathcal{F} \subseteq \mathcal{C}$.

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

Proof. Since $\mathcal{F} \not\subseteq \mathcal{G}_1 \cup \mathcal{G}_2 \cup \mathcal{G}_3$, there exists an $f \in \mathcal{F}$ and $f \notin \mathcal{G}_1 \cup \mathcal{G}_2 \cup \mathcal{G}_3$. Since all unary signatures are in $\mathcal{G}_3$, the arity of f is greater than 1, and by our normalization of $\mathcal{F}$, f is nondegenerate, in particular not identically zero. There are two cases depending on whether f has a zero entry:

(1) f has some zero entries. If there exists a subsignature of f of the form $[0, a, b]$ or $[a, b, 0]$, where $ab \neq 0$, then we are done by Lemma 4.6. Otherwise, there

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

(2) f has no zero entry. We only need to prove that we can construct a function $[a', b', c']$ satisfying the three conditions in Lemma 4.6. Suppose all subsignatures of f with arity 2 do not satisfy all three conditions. Then for each sub-signature $[a', b', c']$, either $a' + c' = 0$, or $b'^2 = a'c'$. If all of them satisfy $a' + c' = 0$, then $f \in \mathcal{G}_3$, a contradiction. If all of them satisfy $b'^2 = a'c'$, then f is degenerate, a contradiction. Without loss of generality, we can assume there is a subsignature $[a, b, c, d]$ of f, such that $a + c = 0$, $b + d \neq 0$, and $c^2 = bd$. We get this subsignature $[a, b, c, d]$ by $[1, 0]$ and $[0, 1]$. Connecting two copies of $[a, b, c, d]$ by two edges, we can get a binary function $[a', b', c'] = [a^2 + 2b^2 + c^2, ab + 2bc + cd, b^2 + 2c^2 + d^2] = [2(b^2 + c^2), c(b + d), (b + d)^2]$. Since f has no zero entries, we have $b' = c(b + d) \neq 0$, $a' + c' > 0$, and $a'c' - b'^2 = (b + d)^2(2b^2 + c^2) > 0$. We are done by Lemma 4.6. ∎

The following lemma "squeezes" $\mathcal{G}_2$ down to $\mathcal{M}$. It uses Theorem 4.1 in an essential way, which in turn depends on the *cross function*.

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

Proof. If $\mathcal{F} \not\subseteq \mathcal{G}_1 \cup \mathcal{G}_2 \cup \mathcal{G}_3$, then by Lemma 4.7, we are done. Otherwise, there exists a signature $f \in \mathcal{F} \subseteq \mathcal{G}_1 \cup \mathcal{G}_2 \cup \mathcal{G}_3$ and $f \notin \mathcal{G}_1 \cup \mathcal{M} \cup \mathcal{G}_3$. Then it must be the case that $f \in \mathcal{G}_2$. Note that every signature with arity at most 3 in $\mathcal{G}_2$ (thus satisfying the parity condition) is also contained in $\mathcal{M}$, so f has arity greater than 3. Let $f = [x_0, x_1, \ldots, x_n]$ for some $n \geq 4$. In particular f is nondegenerate. Suppose there exists some $i \in \{2, 3, \ldots, n-2\}$ such that $x_i \neq 0$. If $x_{i-2}x_{i+2} \neq x_i^2$, then we can get $[x_{i-2}, 0, x_i, 0, x_{i+2}]$ by $[1, 0]$ and $[0, 1]$ as a subsignature. Then the problem is #P-hard by Theorem 4.1 and we are done. Otherwise, we have $x_{i-2}x_{i+2} = x_i^2 \neq 0$. Then starting from $x_{i-2} \neq 0$ and if $i - 2 \in \{2, 3, \ldots, n-2\}$, we can get $x_{i-4}x_i = x_{i-2}^2 \neq 0$. Similarly we can start with $x_{i+2} \neq 0$ and get $x_i x_{i+4} = x_{i+2}^2 \neq 0$ if $i + 2 \in \{2, 3, \ldots, n-2\}$. A signature satisfying the parity condition and that is a geometric series on the alternate entries is realizable by a matchgate, by Lemmas 2.12 and 2.13. This is a contradiction.

Now we may assume $x_i = 0$ for all $i \in \{2, 3, \ldots, n-2\}$, and f has the form $[x_0, x_1, 0, \ldots, 0, x_{n-1}, x_n]$. If n is odd then $n \geq 5$, and since $f \in \mathcal{G}_2$, either $x_0 = x_{n-1} = 0$ or $x_1 = x_n = 0$. In the former case, if $x_1 = 0$ then f would be degenerate, and if $x_n = 0$ then $f \in \mathcal{M}$. Since f is nondegenerate and $f \in \mathcal{G}_2 - \mathcal{M}$, we have $x_1 x_n \neq 0$. In the latter case that $x_1 = x_n = 0$, by the same argument we have $x_0 x_{n-1} \neq 0$. If n is even, then since $f \in \mathcal{G}_2$, either $x_0 = x_n = 0$ or $x_1 = x_{n-1} = 0$. The latter case is impossible because f is nondegenerate and $f \in \mathcal{G}_2 - \mathcal{G}_1$. In the former case, it must be that $x_1x_{n-1} \neq 0$, for otherwise $f \in \mathcal{M}$. Furthermore in fact $n \geq 6$ if n is even, because $[0, x_1, 0, x_3, 0] \in \mathcal{M}$, for any x_1, x_3.

Therefore, given that $f \in \mathcal{G}_2 - (\mathcal{M} \cup \mathcal{G}_1)$, there are only three possible subcases: (1) n is odd, $n \geq 5$, $x_1x_n \neq 0$, and $x_0 = x_{n-1} = 0$; (2) n is odd, $n \geq 5$, $x_0x_{n-1} \neq 0$, and $x_1 = x_n = 0$; (3) $n \geq 6$ is even, $x_1x_{n-1} \neq 0$, and $x_0 = x_n = 0$. The subcases (1) and (2) are reversals of each other and (3) contains a subsignature of form (1). By pinning we can get $(\neq_2) = [0, 1, 0]$, and therefore we can get reversals. So after normalizing (and connecting pairs of edges together if $n > 5$), we can get a signature $[0, 1, 0, 0, 0, z]$ of arity 5, where $z \neq 0$. So we have both subsignatures $[0, 1, 0, 0]$ and $[1, 0, 0, 0, z]$. As we proved in Lemma 4.2, the problem is #P-hard and we are done. This finishes the proof. ∎

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

Proof. By pinning we can get $(\neq_2) = [0, 1, 0]$ in either case, so we can get reversal, and therefore we only need to consider the case $[0, 1, 0, x] \in \mathcal{F}$. If $x \neq 0$, we can use Lemma 4.4 to interpolate $[0, 1, 0, 0]$. So we assume we have $[0, 1, 0, 0]$ from $\mathcal{F}$. If $\mathcal{F} \not\subseteq \mathcal{G}_1 \cup \mathcal{M} \cup \mathcal{G}_3$, then by Lemma 4.8, we are done. If $\mathcal{F} \subseteq \mathcal{M}$, then the problem is tractable and we are done. Otherwise, there exists a signature $f \in \mathcal{F} \subseteq \mathcal{G}_1 \cup \mathcal{M} \cup \mathcal{G}_3$ and $f \notin \mathcal{M}$. Thus $f \in \mathcal{G}_1 \cup \mathcal{G}_3 - \mathcal{M}$.

If $f \in \mathcal{G}_3$, then f has the form $[x_0, x_1, -x_0, -x_1, \ldots]$ having arity ≥ 1. If x_0 or $x_1 = 0$, we would have $f \in \mathcal{M}$ by Lemmas 2.12 and 2.13, a contradiction. Hence $x_0x_1 \neq 0$. We can get the unary signature $[x_0, x_1]$ by pinning. Connecting $[x_0, x_1]$ to $[0, 1, 0, 0]$ we get $[x_1, x_0, 0]$, which satisfies all the conditions in Lemma 4.6, and we are done.

Now suppose $f \in \mathcal{G}_1$; then it has the form $[1, 0, \ldots, 0, y]$ after normalization, where $y \neq 0$. If it has arity 1, then connecting $[1, y]$ to $[0, 1, 0, 0]$ we can also finish the proof by Lemma 4.6. If it has arity 2, then $[1, 0, y] \in \mathcal{M}$, a contradiction. Hence f has arity $n \geq 3$. If n is odd, we can connect its edges in pairs except one to get a unary signature $[1, y]$. Then we can use the same argument as above and we are done. If n is even, then it is at least 4. After connecting its edges in pairs except four of them, we can get $[1, 0, 0, 0, y]$. Together with $[0, 1, 0, 0]$, we know the problem is #P-hard by Lemma 4.2. This completes the proof. ∎

Now we resume the "squeeze" from $\mathcal{M}$ down to (the real part of) $\mathcal{F}_2$.

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

Proof. If $\mathcal{F} \not\subseteq \mathcal{G}_1 \cup \mathcal{G}_2 \cup \mathcal{G}_3$, then by Lemma 4.7, we are done. Otherwise, there exists a signature $f \in \mathcal{F} \subseteq \mathcal{G}_1 \cup \mathcal{G}_2 \cup \mathcal{G}_3$ and $f \notin \mathcal{G}_1 \cup \mathcal{F}_2 \cup \mathcal{G}_3$. Then it must be the case that $f \in \mathcal{G}_2$. Note that every unary or nondegenerate binary signature in $\mathcal{G}_2$ is also contained in $\mathcal{G}_1 \cup \mathcal{G}_3$, so f has arity at least 3. Since f is nondegenerate and $f \notin \mathcal{G}_1$, there is some nonzero in the middle of the signature f. Moreover, by $f \in \mathcal{G}_2$ (the parity condition), there is a 0 entry just before this nonzero entry as well as a 0 entry just after it. After normalization, we can assume there is a subsignature of the form $[0, 1, 0, x]$ (or $[x, 0, 1, 0]$). By pinning we can get $[0, 1, 0]$ and take reversal, and so we have $[0, 1, 0, x]$. If $x \neq \pm 1$, then by Lemma 4.9, we are done. Otherwise, for every such pattern, we have $x = \pm 1$. Thus f consists of alternatingly zero and nonzero entries, and all nonzero entries are ± 1. The nonzero entries cannot all be of the same sign, because $f \notin \mathcal{F}_2$. Also all nonzero entries of f cannot always strictly alternate between $+1$ and -1 because $f \notin \mathcal{G}_3$. Hence we can get a subsignature $[1, 0, 1, 0, -1]$ or $[1, 0, -1, 0, -1]$ of f, up to a nonzero scalar, by pinning. By taking a reversal, if necessary, we get $[1, 0, 1, 0, -1]$. Then by Theorem 4.1, we know that the problem is #P-hard and we are done. This completes the proof. ∎

The next lemma "squeezes" by dropping $\mathcal{F}_2$.

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

Proof. If $\mathcal{F} \not\subseteq \mathcal{G}_1 \cup \mathcal{F}_2 \cup \mathcal{G}_3$, then by Lemma 4.10, we are done. Otherwise, there exists a signature $f \in \mathcal{F} \subseteq \mathcal{G}_1 \cup \mathcal{F}_2 \cup \mathcal{G}_3$ and $f \notin \mathcal{G}_1 \cup \mathcal{G}_3$. Then it must be the case that $f \in \mathcal{F}_2$. Note that every (real-valued) signature with arity at most 2 in $\mathcal{F}_2$ is also contained in $\mathcal{G}_1 \cup \mathcal{G}_3$, so f has arity at least 3. Then f has a subsignature $[1, 0, 1, 0]$ or $[0, 1, 0, 1]$. By pinning we get $[0, 1, 0]$, and then taking reversal, we assume it is $[1, 0, 1, 0]$. If $\mathcal{F} \subseteq \mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3$, then Theorem 4.5 trivially holds and there is nothing to prove. If not, there exists a signature $g \in \mathcal{F} - \mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3$. By $\mathcal{F} \subseteq \mathcal{G}_1 \cup \mathcal{F}_2 \cup \mathcal{G}_3$, either $g \in \mathcal{G}_1 - \mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3$ $(\subseteq \mathcal{G}_1 - \mathcal{F}_1)$ or $g \in \mathcal{G}_3 - \mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3$ $(\subseteq \mathcal{G}_3 - \mathcal{F}_3)$.

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

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

The next lemma "squeezes" $\mathcal{G}_3$ down to $\mathcal{F}_3$.

LEMMA 4.12. *If $\mathcal{F} \not\subseteq \mathcal{G}_1 \cup \mathcal{F}_3$, then the conclusions of Theorem* 4.5 *hold.*

Proof. If $\mathcal{F} \not\subseteq \mathcal{G}_1 \cup \mathcal{G}_3$, then by Lemma 4.11, we are done. Otherwise, there exists a signature $f \in \mathcal{F} \subseteq \mathcal{G}_1 \cup \mathcal{G}_3$ and $f \notin \mathcal{G}_1 \cup \mathcal{F}_3$. Then it must be the case that $f \in \mathcal{G}_3$, and so f has the form $[x_0, x_1, -x_0, -x_1, \ldots]$. If either x_0 or $x_1 = 0$, then $f \in \mathcal{F}_3$. After normalizing $x_0 = 1$, we must have $x_1 \neq \pm 1$ for otherwise $f \in \mathcal{F}_3$. The arity of f is at least 2, for otherwise $f \in \mathcal{G}_1$. Hence f has a subsignature of the form $[1, a, -1]$, where $a \notin \{-1, 0, 1\}$.

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

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

If the arity of g is 2, then g has the form $[1, 0, b]$ by $g \in \mathcal{G}_1$. Here $b \neq 0$ by the definition of $\mathcal{G}_1$, and $b \neq -1$ by $g \notin \mathcal{G}_3$. That $b \neq 1$ is given by $g \notin \{[1, 0, 1]\} \cup \mathcal{G}_3$. Hence $b \notin \{-1, 0, 1\}$. Connecting two copies of the signature $[1, 0, b]$ to both sides of one binary signature $[1, a, -1]$, we can get a new binary signature $[1, ab, -b^2]$. It satisfies all the conditions of Lemma 4.6, and we are done. If the arity of g is greater than 2, then we can always realize a signature $[1, 0, 0, b]$, where $b \neq 0$ (by pinning we obtain $[1, a]$ from $[1, a, -1]$, and then connecting the unary signature $[1, a]$ to all dangling edges of g except three of them). Then we can use an $\mathcal{F}$-gate in Figure 8. Its signature is $[1, a^2 b, b^2]$, and by Lemma 4.6, we are done. This completes the proof. ∎

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

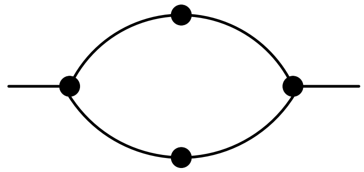

FIG. 8. *The function on degree* 2 *nodes is* $[1, a, -1]$, *and the function on degree* 3 *nodes is* $[1, 0, 0, b]$.

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

Proof. If $\mathcal{F} \subseteq \mathcal{F}_1 \cup \mathcal{F}_3$, then by Theorem 2.8, part (2), $\text{Holant}^c(\mathcal{F})$ is computable in polynomial time. Similarly, if $\mathcal{F} \subseteq \mathcal{U} \cup \mathcal{F}_3 \cup \{[1,0,1]\}$, then by Theorem 2.8, part (1), and then by Theorem 2.7, part (3), $\text{Holant}^c(\mathcal{F})$ is computable in polynomial time. Hence in these two cases, Theorem 4.5 holds. Now suppose $\mathcal{F} \not\subseteq \mathcal{F}_1 \cup \mathcal{F}_3$ and $\mathcal{F} \not\subseteq \mathcal{U} \cup \mathcal{F}_3 \cup \{[1,0,1]\}$.

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

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

The second case is every $f \in \mathcal{F} \cap \mathcal{G}_1 - (\mathcal{F}_1 \cup \mathcal{F}_3)$ is also in $\mathcal{U}$. By $\mathcal{F} \not\subseteq \mathcal{U} \cup \mathcal{F}_3 \cup \{[1,0,1]\}$, there exists $f_1 \in \mathcal{F} - (\mathcal{U} \cup \mathcal{F}_3 \cup \{[1,0,1]\})$. Since $\mathcal{F} \subseteq \mathcal{G}_1 \cup \mathcal{F}_3$, and $f_1 \notin \mathcal{F}_3$, we get $f_1 \in \mathcal{G}_1$. If $f_1 \notin \mathcal{F}_1$, we could use this f_1 as the f above, namely, $f_1 \in \mathcal{F} \cap \mathcal{G}_1 - (\mathcal{F}_1 \cup \mathcal{F}_3 \cup \mathcal{U})$, a contradiction. Thus $f_1 \in \mathcal{F}_1$. Also we still have some $f_2 \in \mathcal{F} - (\mathcal{F}_1 \cup \mathcal{F}_3)$. So $f_2 \in \mathcal{G}_1$, since $\mathcal{F} \subseteq \mathcal{G}_1 \cup \mathcal{F}_3$. Also since we are in this second case, certainly $f_2 \in \mathcal{U}$.

So we have $f_1, f_2 \in \mathcal{F} \cap \mathcal{G}_1$ such that $f_1 \in \mathcal{F}_1$ but $f_1 \notin \mathcal{U} \cup \mathcal{F}_3 \cup \{[1,0,1]\}$, and $f_2 \in \mathcal{U}$ but $f_2 \notin \mathcal{F}_1$. The arity of f_1 is at least 2. We claim it is greater than 2. Otherwise, f_1 being from $\mathcal{F}_1$ and not $[1,0,1]$, it would be $f_1 = [1, 0, -1] \in \mathcal{F}_3$, a contradiction. So f_1 has the form $[1, 0, \ldots, 0, \pm 1]$ of arity at least 3. f_2 has the form $[1, a']$, where $a' \notin \{-1, 0, 1\}$; this follows from $f_2 \in \mathcal{U} \cap \mathcal{G}_1 - \mathcal{F}_1$. By connecting all dangling edges of f_1 except two with f_2, we can construct an $\mathcal{F}$-gate with signature of the form $[1, 0, a]$, where $a \notin \{-1, 0, 1\}$. This is one of the desired forms we want.

To sum up, in both cases, we have some f of the form $[1, 0, a]$ or $[1, 0, 0, a]$, where $a \notin \{-1, 0, 1\}$.

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

According to f and g, we have four cases. If $f = [1, 0, a]$ and $g = [1, 1, -1]$, then connecting them together into a chain of first f then g and then f again, we can realize $[1, a, -a^2]$. By Lemma 4.6, we are done. If $f = [1, 0, a]$ and $g = [1, 0, -1, 0]$, for each dangling edge of g, we extend it by one copy of f. Then we can realize $[1, 0, -a^2, 0]$.

So by Lemma 4.9, we are done. If $f = [1, 0, 0, a]$ and $g = [1, 1, -1]$, we can connect a unary signature $[1, 1]$ (a subsignature of g) to one dangling edge of f and realize a binary signature $f = [1, 0, a]$. This reduces it to the first case, which has been proved. If $f = [1, 0, 0, a]$ and $g = [1, 0, -1, 0]$, we can realize a unary signature $[1, a]$ from f by connecting two of its dangling edges together and then connect this unary signature to one dangling edge of g to realize $[1, -a, -1]$. Note that $[1, -a, -1] \notin \mathcal{G}_1 \cup \mathcal{F}_3$, and by Lemma 4.12, we are done. ∎

5. Dichotomy for planar weighted #CSP. In this section, we prove a dichotomy for planar real weighted #CSP. Compared to the dichotomy for general real weighted #CSP, the new tractable cases for planar structures are precisely those problems that can be computed by holographic algorithms with matchgates.

Let $\mathcal{EQ} = \{=_k \mid k \geq 1\} = \{[1, 1], [1, 0, 1], [1, 0, 0, 1], \ldots\}$ be the set of EQUALITY signatures of all arities. A #CSP problem is just a Holant problem with $\mathcal{EQ}$ assumed to be freely available, i.e., #CSP($\mathcal{F}$) $\equiv_{\mathrm{T}}$ Holant($\mathcal{F} \cup \mathcal{EQ}$), and Pl-#CSP($\mathcal{F}$) $\equiv_{\mathrm{T}}$ Pl-Holant($\mathcal{F} \cup \mathcal{EQ}$). One can show using the signature theory developed in [14] that the only holographic transformation that is relevant in transforming a planar #CSP problem to be computable by matchgates is the Hadamard transformation $H = \left[\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}\right]$. (This fact will not be used in the proof of this paper in a logical sense. However, it is certainly what led to the formulation of Theorem 5.1. The signature theory developed in [14] says that in order to transform all $\mathcal{EQ}$ to be realizable by matchgate signatures, H can accomplish that, and any other transformation that accomplishes that is essentially equivalent to H.)

Now we present the dichotomy theorem for planar weighted #CSP.

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

1. #CSP($\mathcal{F}$) *is polynomial time tractable (for which we have an effective dichotomy* [20]*, Theorem* 2.10*); or*
2. *Every function in $\mathcal{F}$ is realizable by some matchgate under basis $\left[\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}\right]$ (for which we have a complete characterization* [14], Theorem 2.14*).*

The main proof idea is to reduce Pl-Holant^c problems to Pl-#CSP problems. Pl-#CSP($\mathcal{F}$) is exactly the same as planar Holant with all the EQUALITY functions, i.e., Pl-Holant($\mathcal{F} \cup \mathcal{EQ}$). We can use a holographic reduction H. Note that $\frac{1}{\sqrt{2}}H$ is an orthogonal matrix, thus $(=_2)(\frac{1}{\sqrt{2}}H)^{\otimes 2} = (=_2)$ is invariant. Pl-Holant($\mathcal{F} \cup \mathcal{EQ}$) can be written as a bipartite Holant problem Pl-Holant($=_2 \mid \mathcal{F} \cup \mathcal{EQ}$). Under the transformation by H we get the transformed bipartite Holant problem Pl-Holant($=_2 \mid H\mathcal{F} \cup \widehat{\mathcal{EQ}}$), where $\widehat{\mathcal{EQ}} = \{[1, 0], [1, 0, 1], [1, 0, 1, 0], [1, 0, 1, 0, 1], \ldots\}$ is the set of transformed signatures from $\mathcal{EQ}$ by H. Thus the problem Pl-Holant($H\mathcal{F} \cup \widehat{\mathcal{EQ}}$) has the same complexity as Pl-#CSP($\mathcal{F}$).

This holographic transformation gives us $[1, 0] \in \widehat{\mathcal{EQ}}$ (from $[1, 1]$). If we can further realize (or interpolate) $[0, 1]$, then we can view the problem as a Pl-Holant^c problem and apply Theorem 4.5 to $H\mathcal{F} \cup \widehat{\mathcal{EQ}}$ to get a proof of Theorem 5.1. In the following, we show how to realize (or interpolate) $[0, 1]$. Once we have $[0, 1]$, the translation of the criterion of Theorem 4.5 to Theorem 5.1 is straightforward.

It turns out that to realize (or interpolate) $[0, 1]$ in some cases is difficult. The following lemma says that it is also sufficient if we can realize (or interpolate) $[0, 0, 1] = [0, 1]^{\otimes 2}$. The signature $[0, 0, 1]$ can be viewed as two copies of $[0, 1]$. Intuitively, we will use one copy of $[0,0,1]$ to replace two occurrences of $[0,1]$. However, there are two technical difficulties. One is that there may be an odd number of occurrences of $[0,1]$ used in the input instance; the second difficulty, which is more subtle, is that we have to pair up two copies of $[0,1]$ while maintaining planarity of the instance.

For any set of signatures $\mathcal{G}$, let Σ denote the problem Pl-Holant$^c(\mathcal{G} \cup \widehat{\mathcal{EQ}})$, and let Π denote the problem Pl-Holant$(\mathcal{G} \cup \widehat{\mathcal{EQ}} \cup \{[0,1]^{\otimes 2}\})$.

Lemma 5.2. *If* Σ *is in P, then so is* Π. *If* Σ *is #P-hard, then so is* Π.

Proof. Both problems have the signature $[1,0]$. Problem Σ has $[0,1]$, which can obtain $[0,1]^{\otimes}$ easily. So clearly $\Pi \leq_T \Sigma$. Thus if Σ is in P, so is Π.

We have already proved a dichotomy theorem for Pl-Holant^c problems. So if Σ is not in P, then it is #P-hard, and there is a reduction guaranteed by Theorem 4.5. In this case we prove that Π is also #P-hard.

We observe that in all the proofs that lead to Theorem 4.5 asserting Σ is #P-hard, there is a chain of reductions that ultimately comes from following three problems: (a) Pl-Holant$([1,0,0,1]|[1,1,0])$, (b) Pl-Holant$([1,1,0,0])$, or (c) Holant$[0,1,0,0]$ (respectively counting Vertex Cover, Matching for planar 3-regular graphs, or Perfect Matching for general 3-regular graphs). There are only three reduction methods in this reduction chain: direct gadget construction, polynomial interpolation, and holographic reduction. We further observe that these reduction methods all share a certain *local* property described below.

Given an instance G of Pl-Holant$([1,0,0,1]|[1,1,0])$, Pl-Holant$([1,1,0,0])$, or Holant$[0,1,0,0]$, notice that the value of Pl-Holant$([1,0,0,1]|[1,1,0])$, Pl-Holant$([1,1,0,0])$, or Holant$[0,1,0,0]$ on the instance G is a nonnegative integer. We consider the graph $G \cup G$, which denotes the disjoint union of two copies of G. The value on $G \cup G$ is the square of the value on G. So we can compute the value on G uniquely from its square. Suppose the reduction chain on the instance G produced instances $G_1, G_2, \ldots, G_m$ of the problem Σ. The same reduction applied to $G \cup G$ produces instances of the problem Σ of the form $G_1 \cup G_1, G_2 \cup G_2, \ldots, G_{m'} \cup G_{m'}$. (This is the *local* property of the reduction. We note that the reduction on $G \cup G$ may produce polynomially more instances than on G because of polynomial interpolation.)

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

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

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

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

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

LEMMA 5.4. *Let* $a \in \mathbb{R}$. *If* $a \notin \{0,1,-1\}$*, then we can interpolate* $[0,1]$ *from* $\widehat{\mathcal{EQ}} \cup \{[1,a]\}$.

Proof. For every $j \geq 1$, we can take a function $F_{j+1} = [1,0,1,0,1,\ldots]$ of arity $j+1$ and connect j functions $[1,a]$ to it.

The row vector form of the function (i.e., a listing of its values) of arity j composed of j copies of $[1,a]$ is $(1,a)^{\otimes j}$. The column vector form of F_{j+1} is $1/2\left[\begin{smallmatrix}1\\1\end{smallmatrix}\right]^{\otimes(j+1)} + 1/2\left[\begin{smallmatrix}1\\-1\end{smallmatrix}\right]^{\otimes(j+1)}$. The $2^j \times 2$ matrix form of F_{j+1} is $1/2\left[\begin{smallmatrix}1\\1\end{smallmatrix}\right]^{\otimes j} \otimes (1,1) + 1/2\left[\begin{smallmatrix}1\\-1\end{smallmatrix}\right]^{\otimes j} \otimes (1,-1)$.

Our gadget realizes

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

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

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

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

Proof of Theorem 5.1. In this proof, we augment the class $\mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3$ to include those degenerate signatures that can be obtained from tensor products from unary signatures in $\mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3$. The augmented class is precisely all symmetric signatures of the affine class $\mathcal{A}$. For any set of symmetric constraint functions $\mathcal{F}$, to be contained in the augmented class of $\mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3$ is the same as to be contained in $\mathcal{A}$; however the former statement is easier to verify by an inspection of the explicit list in subsection 2.4. Note that for a set of symmetric signatures $\mathcal{F}$, the transformed set $H\mathcal{F}$ is also symmetric. ∎

If $H\mathcal{F} \subseteq \mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3$ (in the augmented sense), then the problem #CSP($\mathcal{F}$) is tractable (even for general graphs). In this case, in fact $\mathcal{F} \subseteq \mathcal{A}$, since $H^{-1}(\mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3) = \mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3$, and the conclusion of the theorem holds. Now we assume that there exists an $f \in H\mathcal{F} - (\mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3)$. In the following, we will prove that we can realize or interpolate either $[0, 1]$ or $[0, 0, 1]$ from f and $\widehat{\mathcal{EQ}}$ in Pl-Holant.

The general thrust of the proof is to squeeze all possible f into several standardized forms and prove the ability to directly realize or interpolate either $[0, 1]$ or $[0, 0, 1]$, using Lemma 5.4 or Lemma 5.5. Suppose $f = [f_0, f_1, \ldots, f_n]$. Since we have $[1, 0] \in \widehat{\mathcal{EQ}}$, we can always take any subsignature that is an initial segment of a signature we already have. Given a symmetric signature g with arity $r > 1$, we often use two copies of g such that $r - 1$ inputs of them are connected to each other. We call this the double gadget from g, which creates a binary symmetric signature. We separate two cases according to whether $f_0 = 0$ or $f_0 \neq 0$, which we normalize to $f_0 = 1$.

1. $f_0 = 0$.
   As the identically 0 function is in $\mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3$, f is not identically 0, and thus for some $i \geq 1$, $f_i \neq 0$. If $f_0 = 0$ and $f_1 \neq 0$, then we can connect $n - 1$ functions $[1, 0]$ to f to get $[0, f_1]$, which is $[0, 1]$ up to a nonzero factor.
   So we may assume $f_0 = f_1 = 0$, then $n \geq 2$. If $f_2 \neq 0$, then we can connect $n - 2$ functions $[1, 0]$ to f to get $[0, 0, f_2]$, which is $[0, 0, 1]$ up to a nonzero factor.
   So we may assume $f_0 = f_1 = f_2 = 0$, then $n \geq 3$. Let $m \leq n$ be the first nonzero, $f_0 = f_1 = f_2 = \cdots = f_{m-1} = 0$, $f_m \neq 0$, then $m \geq 3$, and we can first get $[f_0, f_1, \ldots, f_{m-1}, f_m] = [0, 0, \ldots, 0, f_m]$, which is $[0, 1]^{\otimes m}$, up to a nonzero factor f_m. Depending on the parity of m, by connecting $[1, 0, 1]$ to two dangling edges of it repeatedly we can get either $[0, 1]$ or $[0, 0, 1]$.
2. $f_0 = 1$.
   By Lemma 5.4, we only need to consider $f_1, \in \{0, 1, -1\}$. Otherwise, we are done.
   (a) $f_0 = 1$ and $f_1 = \pm 1$.
   If $n = 1$, then $f = [f_0, f_1] = [1, \pm 1] \in \mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3$, a contradiction. So $n \geq 2$, we can take its initial part $[1, \pm 1, f_2]$.

   Claim 1. If we have a signature of the form $g = [g_0, g_1, g_2] = [1, \pm 1, g_2]$, then either $g_2 = 1$ or we can get $[0, 1]$.

   To prove Claim 1, we connect one edge of g to $[g_0, g_1] = [1, \pm 1]$. This gives us a unary signature $[g_0^2 + g_1^2, g_0 g_1 + g_1 g_2] = [2, \pm(1 + g_2)]$. By Lemma 5.4 either we can get $[0, 1]$ or $g_2 \in \{1, -1, -3\}$. We can construct another gadget which connects two inputs of $[1, 0, 1, 0] \in \widehat{\mathcal{EQ}}$ by $g = [g_0, g_1, g_2]$. This produces a unary signature $[g_0 + g_2, 2g_1] = [1 + g_2, \pm 2]$. It follows that if $g_2 = -1$ then we can get $[0, 1]$. So we may assume $g_2 \neq -1$. Next consider the double gadget of g, which has signature matrix $\left[\begin{smallmatrix} g_0 & g_1 \\ g_1 & g_2 \end{smallmatrix}\right]^2 = \left[\begin{smallmatrix} 2 & \pm(1+g_2) \\ \pm(1+g_2) & 1+g_2^2 \end{smallmatrix}\right]$. If $g_2 = -3$, then this signature is $[2, \mp 2, 10]$, which is a multiple of $[1, \mp 1, 5]$. By the first part of the proof of Claim 1, since $5 \notin \{1, -1, -3\}$, we can get $[0, 1]$. So the only possibility left is $g_2 = 1$. We have proved Claim 1.
   By applying Claim 1 to $[f_0, f_1, f_2]$, we conclude that $[f_0, f_1, f_2] = [1, \pm 1, 1]$. Since $f \notin \mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3$, the arity n of f is greater than 2, and we have an initial segment $[1, \pm 1, 1, f_3]$. Our goal in this case 2(a) is either to get $[0, 1]$ or to extend this pattern $[1, \pm 1, 1, \ldots]$. Since $f \notin \mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3$, this pattern for f cannot be extended indefinitely, and then we will have proved that we can get $[0, 1]$.

Claim 2. If we have a signature of the form $g = [g_0, g_1, g_2, \ldots, g_m, g_{m+1}] = [1, \pm 1, 1, \ldots, g_{m+1}]$, where $m \geq 1$ and either (1) $g_j = 1$ for all $j = 0, 1, \ldots, m$, or (2) $g_j = (-1)^j$ for all $j = 0, 1, \ldots, m$, then either $g_{m+1} = 1$ in case (1) or $g_{m+1} = (-1)^{m+1}$ in case (2), or we can get $[0, 1]$.

We prove Claim 2 by induction on m. The base case $m = 1$ has already been proved in Claim 1. Now suppose $m \geq 2$. We connect one edge of g to $[g_0, g_1] = [1, \pm 1]$ to get $[g_0^2 + g_1^2, g_0 g_1 + g_1 g_2, g_0 g_2 + g_1 g_3, \ldots, g_0 g_m + g_1 g_{m+1}]$, which is either (1) $[2, 2, \ldots, 1 + g_{m+1}]$ or (2) $[2, -2, \ldots, g_{m-1} - g_m, g_m - g_{m+1}]$, both of arity m. In case (1), by the inductive hypothesis, we get $g_{m+1} = 1$, or we can get $[0, 1]$. In case (2), the entries starting from $2, -2, \ldots$ up to $g_{m-1} - g_m = \pm 2$ indexed at $m-1$ strictly alternate. Therefore by induction hypothesis, we also either can get $[0, 1]$, or have $g_m - g_{m+1} = -(g_{m-1} - g_m) = 2g_m$, hence $g_{m+1} = -g_m = (-1)^{m+1}$. Claim 2 is proved.

We now apply Claim 2 to f, and since $f \notin \mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3$, this pattern cannot be indefinitely extended, and therefore we can get $[0, 1]$.

(b) $f_0 = 1$ and $f_1 = 0$.

Since $[1, 0] \in \mathcal{F}_2$, and $f \notin \mathcal{F}_2$, we have $n > 1$. If $f = [1, 0]^{\otimes n} = [1, 0, \ldots, 0]$, it would belong to the augmented class of $\mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_2$. But f does not. So f has another nonzero entry other than f_0. Suppose f_m is the first nonzero f_i other than f_0. Then $m > 1$. By taking the initial segment we can get $[f_0, 0, \ldots, 0, f_m]$. We can connect $[1, 0, 1]$ to $[f_0, 0, \ldots, 0, f_m]$ to get $[1, f_m]$ or $[1, 0, f_m]$ depending on the parity of m. Then we may assume $f_m = \pm 1$ by Lemmas 5.4 and 5.5; otherwise we are done. Since $f \notin \mathcal{F}_1$ we have $n > m$.

Next we prove that we may assume m is even, or else we are done. If m is odd, we can get $[1, f_m, f_{m+1}]$ by connecting some $[1, 0, 1]$. Since $f_m = \pm 1$, Claim 1 applies. So $f_{m+1} = 1$, or else we are done. We can also get $[1, 0, 0, f_m, f_{m+1}]$, since $m > 1$, whose double gadget has the signature $[1 + f_m^2, f_m f_{m+1}, 3f_m^2 + f_{m+1}^2] = [2, \pm 1, 4]$. Then we can get $[2, \pm 1]$, and we are done by Lemma 5.4.

Now we know m must be even. Next we show that in fact we may assume $m = 2$. Otherwise, $m \geq 4$ and we can get $[1, 0, 0, 0, f_m, f_{m+1}]$ by connecting some $[1, 0, 1]$ to f. The double gadget of this has the signature $[1 + f_m^2, f_m f_{m+1}, 4f_m^2 + f_{m+1}^2] = [2, \pm f_{m+1}, 4 + f_{m+1}^2]$. From this we can get $[2, \pm f_{m+1}]$. By Lemma 5.4, we can get $[0, 1]$ unless $f_{m+1} \in \{0, \pm 2\}$. If $f_{m+1} = 0$, we have $[2, 0, 4]$. By Lemma 5.5 we can get $[0, 0, 1]$. If $f_{m+1} = \pm 2$, then we apply Claim 1 to $[2, \pm f_{m+1}, 4 + f_{m+1}^2]$ and conclude that we get $[0, 1]$. Hence $m = 2$, or we are done.

We have $m = 2$ and have reached $[1, 0, \pm 1, f_3]$, whose double gadget has the signature $[2, \pm f_3, 2 + f_3^2]$. Again by Lemma 5.4 applied to $[2, \pm f_3]$, and by Claim 1, we conclude that $f_3 = 0$, or we are done.

Now we have the initial segment of f being $[1, 0, \pm 1, 0]$. Again since $f \notin \mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3$ we have $n > 3$. Hence we have $[1, 0, \pm 1, 0, f_4]$. For $[1, 0, -1, 0, f_4]$, by connecting two edges with $[1, 0, 1]$, we get $[0, 0, f_4 - 1]$, and we must have $f_4 = 1$, or else we have the signature $[0, 0, 1]$, after normalization. For $[1, 0, 1, 0, f_4]$, by connecting two edges with $[1, 0, 1]$, we get $[2, 0, 1 + f_4]$, and it follows from Lemma 5.5 that $f_4 \in \{1, -1, -3\}$. Connecting three edges of $[1, 0, 1, 0, f_4]$ to three edges of $[1, 0, 1, 0, 1] \in$

$\widehat{\mathcal{EQ}}$, we get $[4, 0, 3 + f_4]$, which rules out $f_4 = -1$, by Lemma 5.5 again. The double gadget of $[1, 0, 1, 0, f_4]$ gives $[4, 0, 3 + f_4^2]$, which rules out $f_4 = -3$. To sum up, we get $f_4 = 1$, or else we are done.

Since $f \notin \mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3$, we have $n > 4$. We have reached $[1, 0, \pm 1, 0, 1, f_5, \ldots]$, from which we can get an initial segment $[1, 0, \pm 1, 0, 1, f_5]$. The rest of the proof is similar to the induction proof for case 2(a) but by skipping all entries with an odd index. Formally we have the next claim.

Claim 3. Suppose we have a signature of the form $g = [g_0, g_1, g_2, \ldots, g_m, g_{m+1}] = [1, 0, \pm 1, 0, 1, \ldots, g_{m+1}]$, where $m \geq 2$, $g_j = 0$ for all odd $1 \leq j \leq m$, and either (1) $g_{2j} = 1$ for all $0 \leq j \leq m/2$ or (2) $g_{2j} = (-1)^j$ for all $0 \leq j \leq m/2$. Then either we can get one of $[0, 1]$ or $[0, 0, 1]$ or we can conclude the following hold: If m is even then $g_{m+1} = 0$, and if m is odd then either $g_{m+1} = 1$ in case (1) or $g_{m+1} = (-1)^{m+1}$ in case (2).

Another way to state the pattern is that the entries of the signature g satisfy the second order recurrence $g_i = g_{i-2}$ for all $i \geq 2$, or the recurrence $g_i = -g_{i-2}$ for all $i \geq 2$, with initial values $g_0 = 1$ and $g_1 = 0$. The claim is that from this pattern holding up to $i \leq m$, it must also hold for the entry at g_{m+1}, or else we can get or interpolate $[0, 1]$ or $[0, 0, 1]$.

We prove Claim 3 by induction on m. The base cases $m = 2$ and $m = 3$ are proved by repeating the argument above on f applied to g. Inductively assume $m \geq 4$, and for all smaller values of m Claim 3 holds. First consider case (2), where the even indexed entries of g alternate up to g_m. We connect $[1, 0, 1]$ to two inputs of $[g_0, g_1, \ldots, g_{m+1}]$ to get a signature of arity $m-1 \geq 3$. This signature has the form $[0, \ldots, 0, g_{m-1} + g_{m+1}]$. If $g_{m-1} + g_{m+1} \neq 0$, we can get either $[0, 1]$ or $[0, 0, 1]$. Hence we may assume $g_{m+1} = -g_{m-1}$. By the induction hypothesis, this shows that $g_{m+1} = 0$ if m is even, or $g_{m+1} = (-1)^{(m+1)/2}$ if m is odd,

Now consider case (1). We also connect $[1, 0, 1]$ to two inputs of $[g_0, g_1, \ldots, g_{m+1}]$ to get a signature of arity $m - 1 \geq 3$. The last entry of this signature at index $m - 1$ is $g_{m-1} + g_{m+1}$. Before this last entry, the signature entries alternate between 2 and 0, starting with $g_0 + g_2 = 2$. If m is even, then $g_{m-1} = 0$, and this signature has the form $[2, 0, 2 \ldots, 0, 2, g_{m+1}]$. By the induction hypothesis, we have $g_{m+1} = 0$ or we can get one of $[0, 1]$ or $[0, 0, 1]$. If m is odd, then this signature has the form $[2, 0, 2 \ldots, 2, 0, g_{m-1} + g_{m+1}]$. Again by the induction hypothesis, we have $g_{m-1} + g_{m+1} = 2$ or we are done. It follows that $g_{m+1} = 1$, completing the induction.

Then we apply Claim 3 to the signature $f \notin \mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3$, and since this pattern cannot continue indefinitely for f, we conclude that we can get $[0, 1]$ or $[0, 0, 1]$.

This completes the proof.

6. Dichotomy for planar 2–3 regular graphs. In this section we prove a dichotomy for Holant on planar 2–3 regular graphs. This setting is very interesting for at least two reasons. From a dichotomy theorem point of view, this is the simplest nontrivial setting and always serves as the starting point of more general dichotomy theorems as in [20, 12]. This was also a focus of several previous works [18, 31, 19, 32], whose result is the starting point of this theorem. From the holographic algorithms point of view, most of the known holographic algorithms [45, 44] are essentially for planar 2–3 regular graphs. The dichotomy theorem here explains the reason why they are special and why many variations of them are #P-hard. In the previous two dichotomies for Pl-Holant^c and Pl-#CSP, the new tractable cases for planar are also done by holographic algorithms with matchgates. However, only special basis transformations are used since we assume some signatures are freely available. In this planar 2–3 regular graphs setting, no additional signatures are assumed to be freely available. Therefore all possible bases can be used in tractable cases.

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

1. Holant$([y_0, y_1, y_2]|[x_0, x_1, x_2, x_3])$ *is tractable (for which we have an effective dichotomy* [12]*); or*
2. *There exists a basis T such that both* $[y_0, y_1, y_2](T^{-1})^{\otimes 2}$ *and* $T^{\otimes 3}[x_0, x_1, x_2, x_3]$ *are realizable by some matchgates (for which we have a complete characterization* [14]*).*

Proof. If $[x_0, x_1, x_2, x_3]$ or $[y_0, y_1, y_2]$ is degenerate, the problem is tractable, even for the nonplanar case, and so this falls in condition 1. Now we assume that they are both nondegenerate. As proved in [20], we can choose an invertible T_1 such that $[x_0, x_1, x_2, x_3]$ (or its reversal, which is similar and we omit that case) can be written as $T_1^{\otimes 3}[1, 0, 0, 1]$ or $T_1^{\otimes 3}[1, 1, 0, 0]$. Therefore by a holographic reduction, we can always reduce the problem equivalently to one of the following two problems: (1) Pl-Holant$([z_0, z_1, z_2]|[1, 0, 0, 1])$ and (2) Pl-Holant$([z_0, z_1, z_2]|[1, 1, 0, 0])$. So it is sufficient to prove the theorem for these two cases.

For Pl-Holant$([z_0, z_1, z_2]|[1, 0, 0, 1])$, by [32], Theorem 2.11, the only case which is hard for general graphs and tractable for planar graphs is $z_0^3 = z_2^3$. This condition is exactly the same as the condition that there exists a basis T such that both $[y_0, y_1, y_2](T^{-1})^{\otimes 2}$ and $T^{\otimes 3}[1, 0, 0, 1]$ are realizable by some matchgates. (This statement follows from the explicit transformation formulae from the signature theory developed for matchgates in [14].) This proves Theorem 6.1 for case (1).

Now we consider Pl-Holant$([z_0, z_1, z_2]|[1, 1, 0, 0])$. If $z_0 = 0$, the problem is trivially tractable even for general graphs. This can be seen by a simple counting argument: in a bipartite graph the LHS vertices all have the signature $[0, z_1, z_2]$ and thus at least half the edges must be 1, while the RHS vertices all have the signature $[1, 1, 0, 0]$ and thus less than half the edges are 1. This is also the only case where the problem is not #P-hard for general graphs when the RHS has $[1, 1, 0, 0]$ by [12]. Now we assume $z_0 \neq 0$. Then it is sufficient to prove that either the problem is #P-hard or there exists a basis transformation T such that $[1, 1, 0, 0]T^{\otimes 3}$ and $(T^{-1})^{\otimes 2}[z_0, z_1, z_2]$ are realizable by some matchgates. Let $T = \begin{bmatrix} \sqrt{z_0} & 0 \\ z_1/\sqrt{z_0} & \sqrt{(z_0 z_2 - (z_1)^2)/z_0} \end{bmatrix}$. Note that T is well defined and invertible since $z_0 \neq 0$ and $[z_0, z_1, z_2]$ is nondegenerate (i.e., $z_0 z_2 - (z_1)^2 \neq 0$). Then we can verify that

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

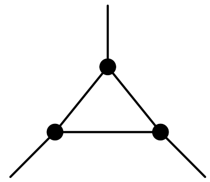

FIG. 9. *All vertex signatures are* $[v, 1, 0, 0]$.

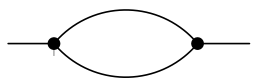

FIG. 10. *All vertex signatures are* $[v, 1, 0, 0]$.

We note that $\sqrt{z_0(z_0z_2 - (z_1)^2)} \neq 0$. If $\sqrt{z_0}(z_0 + 3z_1) = 0$, then both $[\sqrt{z_0}(z_0 + 3z_1), \sqrt{z_0(z_0z_2 - (z_1)^2)}, 0, 0]$ and $[1, 0, 1]$ can be realized by matchgates and the problem for planar graphs is tractable. We denote $v = \frac{\sqrt{z_0}(z_0+3z_1)}{\sqrt{z_0(z_0z_2-(z_1)^2)}} \neq 0$. Then the problem is equivalent to (nonbipartite) Pl-Holant($[v, 1, 0, 0]$). Now it is sufficient to prove the following claim.

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

We can realize $[v^3 + 3v, v^2 + 1, v, 1]$ by connecting three copies of $[v, 1, 0, 0]$'s as illustrated in Figure 9. If we can prove that Pl-Holant($[v^3+3v, v^2+1, v, 1]$) is #P-hard, then we are done. In tensor product notation this signature is

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

Then the following reduction chain holds:

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

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

In addition to the gadget in Figure 9, we can construct a gadget in Figure 10 with a binary signature $[v^2 + 2, v, 1]$. Now it is enough to prove that Pl-Holant($[v^2 + 2, v, 1] | [v^3 + 3v, v^2 + 1, v, 1]$) is #P-hard. Under the same basis $\left[\begin{smallmatrix} v+1 & v-1 \\ 1 & 1 \end{smallmatrix}\right]$, we will get an equivalent problem Pl-Holant($[X, Y, Z] \mid [1, 0, 0, 1]$), where $X = (v^2 + 2)(v^2 + 2v + 1) + 2v(v + 1) + 1$, $Y = (v^2 + 2)(v^2 - 1) + 2v^2 + 1$, and $Z = (v^2 + 2)(v^2 - 2v + 1) + 2v(v - 1) + 1$. Again this transforms the problem to our first case, and it is easy to verify that any root of $3v^4 + 16v^2 + 12 = 0$ is not a tractable case here. This completes the proof of the claim and also the proof of the theorem. ∎

7. A roadmap. Holographic algorithms using matchgates were introduced by Valiant [45, 44]. Initially these novel algorithms appeared rather mysterious, and their success appeared to be equal parts miraculous and coincidental, as perhaps suggested by the title of the paper [44].

Intrigued by this beautiful development, some of us embarked on a systematic study to understand the power of these new algorithms. Perhaps our first result was not a new result in a traditional theoretical computer science sense [8]. This paper mainly examined the breakthrough in [45] and reformulated it in a framework of covariant and contravariant tensor transformations, and then gave a tensor theoretic proof of Valiant's Holant theorem. While there are few "new" tangible results in a traditional sense, casting the theory in this new perspective sets the stage for a deeper understanding of holographic transformations and algorithms.

After this, the first technical inroads were made (1) to capture what exactly matchgates can express and (2) to capture under a holographic transformation what matchgates can be. After the initial work in [42], task (1) was largely accomplished in [7, 9] with a characterization of matchgate signatures by the so-called matchgates identities (also see [10] for a simplified and more streamlined treatment). Building on that, and after the preliminary work in [17], task (2) was largely accomplished in [14] (with extension in [16]). This signature theory is the basis for a more systematic understanding of the power of holographic algorithms with matchgates. It gives a more "explanatory" account for the success in [45, 44].

These results paved the way for classification theorems such as those in this paper. Inspired by the work in [45, 44], we introduced Fibonacci gates [18], which give another class of holographic algorithms different from the ones based on matchgates. In [18, 20] we introduced the Holant framework and proved some dichotomy theorems for Holant* problems (for complex-valued symmetric signature sets—Theorem 2.7), for Holant^c problems (for real-valued symmetric signature sets—Theorem 2.8), and for #CSP problems (for complex-valued not necessarily symmetric signature sets—Theorem 2.10). Two tractable families were isolated, the signatures of product type $\mathcal{P}$ and the signatures of affine type $\mathcal{A}$. It turns out that Fibonacci gates are essentially what are transformable to $\mathcal{P}$ by an orthogonal holographic transformation. Also the symmetric part of $\mathcal{A}$ is captured by $\mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3$ (augmented by tensor powers of unary members of $\mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3$).

The form of the Holant* dichotomy stated in Theorem 2.7 is presented in [20] and is very concrete. This has the advantage of being easy to apply in specific cases. In later work we have come to realize that perhaps a better conceptual understanding is as follows [11]: Problems of tractable class 1 of Theorem 2.7 are those with arity at most 2 as stated, and they are computable by matrix product and taking trace. Problems of tractable class 2 are those orthogonally transformable to $\mathcal{P}$, plus an additional type that will be identified as belonging to the so-called vanishing signatures [26, 11]. Problems of tractable class 3 are those transformable to $\mathcal{P}$ by the transformation $Z = \left[\begin{smallmatrix} 1 & 1 \\ i & -i \end{smallmatrix}\right]$. Theorem 2.10 gives a dichotomy for #CSP, where the tractable classes are $\mathcal{P}$ and $\mathcal{A}$. Theorem 2.8 gives a dichotomy for Holant^c problems where the tractable classes are those for Holant* and those belonging to the affine class $\mathcal{A}$.

Our results in sections 4 and 5 depend on these results. Theorem 2.11 is not used in sections 4 and 5 but is used for the results in section 6. In subsequent work this portion of the research (Holant problems on regular graphs) also plays an important role. This line of results started in [19]. But for the understanding of a significant fraction of the results, e.g., for results in sections 4 and 5, this portion is not needed.

The crucial technical advance for the dichotomy of sections 4 and 5 is Theorem 4.1. The proof is a combinatorial gadget that is not terribly tricky, followed by a purely algebraic demonstration of the remarkable fact that, *precisely* when the parameters are right for being matchgates realizable, the construction for the cross function *fails*. When the cross function can be constructed, there is no difference in the complexity of planar versus nonplanar counting problems. When it is matchgates realizable, then the FKT algorithm can solve the problem in polynomial time for planar instances of the problem that is #P-hard in general. This theorem was our first glimpse that matchgates could be *universal* for #CSP type problems that are #P-hard in general but polynomial time computable on planar graphs. Moreover it gave us the courage that such a theorem might be provable with the knowledge and tools already at hand.

We also feel that the algebraic proof of Theorem 4.1, and many subsequent similar results, suggests something more fundamental. These algebraic proofs seem to be devoid of any intuitive combinatorial explanation; the success at such a sharp demarcation between P and #P-hard seems most unreal if it were the case that #P collapses to P. If #P were equal to P, then it is most mysterious why these algebraic systems would "know" exactly where to stop being solvable (which is where we happen to know a clever algorithm, e.g., FKT), so that we cannot prove #P-hardness for a problem known in P, although by #P = P, every such problem has a polynomial time algorithm. It is as if some all-powerful adversary conspires to maintain this illusion: wherever we know an algorithm the algebraic system will not be solvable such that, e.g., our construction of the cross function fails, which stops us from proving #P-hardness in this case. And yet everywhere else when we do not know an algorithm the adversary will allow us to prove #P-hardness. This is the import and consequence of a dichotomy theorem. It seems the only "rational" explanation is that while we can't prove it, #P is indeed different from P. The sharp boundary of success and failure algebraically in these proofs is a manifestation of this reality, of which the closest nature has allowed us so far is this glimpse of its silhouette.

In the proofs of these dichotomies, we have often used the conjecture P $\neq$ #P as a guide in predicting which algebraic system has a solution and which one does not and have designed proof strategies accordingly, much like one reasons about the behavior of primes assuming the Riemann Hypothesis. Perhaps this aspect will develop to be a concrete contribution by computer science back to mathematics, as is often said to be the case with physics.

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

The Ising model was named after Ernst Ising [27]. Wilhelm Lenz invented this model and gave it to his student Ising to work on. The model consists of a discrete set of variables, called spins, that can be assigned one of two values (states). These spins are usually placed on a lattice structure or a graph, and each spin interacts with its nearest neighbors.

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

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

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

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

In 1925, Ising solved the one-dimensional Ising model [27]. The two-dimensional square lattice Ising model with zero magnetic field was solved by Onsager in 1944 [36]. Onsager announced the formula for the spontaneous magnetization for the two-dimensional model in 1949 but did not give a derivation [37]. Yang (in 1952) gave the first published proof of this formula [48], using a limit formula for Fredholm determinants, proved in 1951 by Szegö in direct response to Onsager's work. There are many extensions to the basic Ising model [35, 1].

Another landmark achievement is the exact computation of the number of perfect matchings (dimer problem) on any planar graph using Pfaffians. This was independently discovered by Kasteleyn [29, 30] and by Temperley and Fisher [38]. This problem can also be nicely expressed by a partition function in our Holant framework, where this time the Boolean variables are the edges (to include an edge or not), and the local constraint function at each vertex is the Exact-One function. Freedman, Lovász, and Schrijver [25] recently proved that this partition function *cannot* be expressed as a graph homomorphism function, where the vertices are variables as in the Ising model. However, in the framework of Holant problems we can find a unity for all these problems.

We note the following. In the paper [14] we gave a complete characterization of matchgate realizable symmetric signatures. The following lemma is proved [14].

Lemma A.1. *The set of bases under which the signature* $[x_0, x_1, x_2]$ *is realizable as a signature by some matchgate is*

$$\left\{ \left[\binom{n_0}{n_1}, \binom{p_0}{p_1} \right] \in \mathbf{GL}_2(\mathbb{C}) \;\middle|\; \begin{array}{c} x_0p_1^2 - 2x_1p_1n_1 + x_2n_1^2 = 0, x_0p_0^2 - 2x_1p_0n_0 + x_2n_0^2 = 0 \\ \text{or} \quad x_0p_0p_1 - x_1(n_0p_1 + n_1p_0) + x_2n_0n_1 = 0 \end{array} \right\}.$$

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

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

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

Istrail [28] showed that computing the free energy of an arbitrary subgraph of an Ising model on a lattice of dimension three or more is NP-hard; see a nice article by Cipra in the *SIAM News* [21]. A very partial list of a great deal of research on this and related models, from a computational complexity perspective, can be found in [2, 3, 4, 7, 17, 20, 19, 23, 31, 5, 6, 16, 15, 14].

Acknowledgments. We thank many colleagues for their interest and helpful comments: Xi Chen, Martin Dyer, Alan Frieze, Sean Hallgren, Leslie Goldberg, Sorin Istrail, Richard Lipton, Jason Morton, Dana Randall, Leslie Valiant, and Santosh Vempala. We want to thank particularly the anonymous referees and the editor for informative comments and many suggestions that helped us improve the presentation. Finally we thank Dr. Zhiguo Fu for a careful reading of the revised version for us and insightful comments on the presentation.

REFERENCES

[1] R. J. Baxter, *Exactly Solved Models in Statistical Mechanics*, Academic Press, London, 1982.

[2] A. A. Bulatov, *The complexity of the counting constraint satisfaction problem*, in Proceedings of ICALP, L. Aceto, I. Damgård, L. A. Goldberg, M. M. Halldórsson, A. Ingólfsdóttir, and I. Walukiewicz, eds., Lecture Notes in Comput. Sci. 5125, Springer, New York, 2008, pp. 646–661.

[3] A. A. Bulatov and V. Dalmau, *Towards a dichotomy theorem for the counting constraint satisfaction problem*, in Proceedings of FOCS, IEEE, 2003, pp. 562–571.

[4] A. A. Bulatov and M. Grohe, *The complexity of partition functions*, in Proceedings of ICALP, J. Díaz, J. Karhumäki, A. Lepistö, and D. Sannella, eds., Lecture Notes in Comput. Sci. 3142, Springer, New York, 2004, pp. 294–306.

[5] A. A. Bulatov and M. Grohe, *The complexity of partition functions*, Theoret. Comput. Sci., 348 (2005), pp. 148–186.

[6] J.-Y. Cai, X. Chen, and P. Lu, *Graph homomorphisms with complex values: A dichotomy theorem*, SIAM J. Comput., 42 (2013), pp. 924–1029.

[7] J.-Y. Cai and V. Choudhary, *Some results on matchgates and holographic algorithms*, in Proceedings of ICALP, M. Bugliesi, B. Preneel, V. Sassone, and I. Wegener, eds., Lecture Notes in Comput. Sci. 4051, Springer, New York, 2006, pp. 703–714.

[8] J.-Y. Cai and V. Choudhary, *Valiant's holant theorem and matchgate tensors*, Theoret. Comput. Sci., 384 (2007), pp. 22–32.

[9] J.-Y. Cai, V. Choudhary, and P. Lu, *On the theory of matchgate computations*, in CCC '07: Proceedings of the 22nd Annual IEEE Conference on Computational Complexity, Washington, DC, 2007, pp. 305–318, https://doi.org/10.1109/CCC.2007.22.

[10] J.-Y. Cai and A. Gorenstein, *Matchgates revisited*, Theory Comput., 10 (2014), pp. 167–197, http://theoryofcomputing.org/articles/v010a007/.

[11] J.-Y. Cai, H. Guo, and T. Williams, *A complete dichotomy rises from the capture of vanishing signatures: Extended abstract*, in Proceedings of the Symposium on Theory of Computing Conference, Palo Alto, CA, 2013, pp. 635–644.

[12] J.-Y. Cai, S. Huang, and P. Lu, *From holant to #CSP and back: Dichotomy for holant^c problems*, Algorithmica, 64 (2012), pp. 511–533.

[13] J.-Y. Cai, P. Lu, and M. Xia, *Holographic algorithms with matchgates capture precisely tractable planar_#CSP*, in Proceedings of the 51th Annual IEEE Symposium on Foundations of Computer Science, 2010, pp. 427–436.

[14] J.-Y. Cai and P. Lu, *Holographic algorithms: from art to science*, in Proceedings of the 39th Annual ACM Symposium on Theory of Computing, New York, ACM, 2007, pp. 401–410, https://doi.org/10.1145/1250790.1250850.

[15] J.-Y. Cai and P. Lu, *Bases collapse in holographic algorithms*, Comput. Complexity, 17 (2008), pp. 254–281.
[16] J.-Y. Cai and P. Lu, *Holographic algorithms with unsymmetric signatures*, in Proceedings of the 19th Annual ACM-SIAM Symposium on Discrete Algorithms, Philadelphia, 2008, pp. 54–63.
[17] J.-Y. Cai and P. Lu, *On symmetric signatures in holographic algorithms*, Theory Comput. Syst., 46 (2010), pp. 198–415.
[18] J.-Y. Cai, P. Lu, and M. Xia, *Holographic algorithms by fibonacci gates and holographic reductions for hardness*, in Proceedings of the 49th Annual IEEE Symposium on Foundations of Computer Science, Washington, DC, IEEE, 2008.
[19] J.-Y. Cai, P. Lu, and M. Xia, *A computational proof of complexity of some restricted counting problems*, in Proceedings of TAMC, J. Chen and S. B. Cooper, eds., Lecture Notes in Comput. Sci. 5532, Springer, New York, 2009, pp. 138–149.
[20] J.-Y. Cai, P. Lu, and M. Xia, *Holant problems and counting CSP*, in Proceedings of STOC, M. Mitzenmacher, ed., ACM, New York, 2009, pp. 715–724.
[21] B. A. Cipra, *The Ising model is NP-complete*, SIAM News, 33 (2000).
[22] C. T. J. Dodson and T. Poston, *Tensor Geometry*, Grad. Texts in Math. 130, Springer-Verlag, New York, 1991.
[23] M. E. Dyer, L. A. Goldberg, and M. Paterson, *On counting homomorphisms to directed acyclic graphs*, J. ACM, 54 (2007), Article 27.
[24] M. E. Dyer and C. S. Greenhill, *The complexity of counting graph homomorphisms*, Random Structures Algorithms, 17 (2000), pp. 260–289.
[25] M. Freedman, L. Lovász, and A. Schrijver, *Reflection positivity, rank connectivity, and homomorphism of graphs*, J. AMS, 20 (2007), pp. 37–51.
[26] H. Guo, P. Lu, and L. G. Valiant, *The complexity of symmetric boolean parity holant problems*, SIAM J. Comput., 42 (2013), pp. 324–356, https://doi.org/10.1137/100815530.
[27] E. Ising, *Beitrag zur theorie des ferromagnetismus*, Z. Phys. A Hadrons Nuclei, 31 (1925), pp. 253–258.
[28] S. Istrail, *Statistical mechanics, three-dimensionality and NP-completeness:* I. *Universality of intracatability for the partition function of the ising model across non-planar surfaces (extended abstract)*, in Proceedings of STOC, 2000, pp. 87–96.
[29] P. W. Kasteleyn, *The statistics of dimers on a lattice*, Physica, 27 (1961), pp. 1209–1225.
[30] P. W. Kasteleyn, *Graph theory and crystal physics*, in Graph Theory and Theoretical Physics, F. Harary, ed., Academic Press, London, 1967, pp. 43–110.
[31] M. Kowalczyk, *Classification of a class of counting problems using holographic reductions*, in Proceedings of COCOON, H. Q. Ngo, ed., Lecture Notes in Comput. Sci. 5609, Springer, New York, 2009, pp. 472–485.
[32] M. Kowalczyk and J.-Y. Cai, *Holant problems for regular graphs with complex edge functions*, in Proceeding of STACS, 2010, pp. 525–536.
[33] T. D. Lee and C. N. Yang, *Statistical theory of equations of state and phase transitions.* II. *Lattice gas and Ising model*, Phys. Rev., 87 (1952), pp. 410–419.
[34] E. H. Lieb and A. D. Sokal, *A general Lee-Yang theorem for one-component and multicomponent ferromagnets*, Commun. Math. Phys., 80 (1981), pp. 153–179.
[35] B. M. McCoy and T. T. Wu, *The Two-Dimensional Ising Model*, Harvard University Press, Cambridge, MA, 1973.
[36] L. Onsager, *Crystal statistics.* I. *A two-dimensional model with an order-disorder transition*, Phys. Rev., 65 (1944), pp. 117–149.
[37] L. Onsager, *Discussion*, Nuovo Cimento (suppl.), 2 (1949), p. 261.
[38] H. N. V. Temperley and M. E. Fisher, *Dimer problem in statistical mechanics—an exact result*, Philosophical Mag., 6 (1961), pp. 1061–1063.
[39] S. P. Vadhan, *The complexity of counting in sparse, regular, and planar graphs*, SIAM J. Comput., 31 (2001), pp. 398–427.
[40] L. G. Valiant, *The complexity of computing the permanent*, Theoret. Comput. Sci., 8 (1979), pp. 189–201.
[41] L. G. Valiant, *The complexity of enumeration and reliability problems*, SIAM J. Comput., 8 (1979), pp. 410–421.
[42] L. G. Valiant, *Expressiveness of matchgates*, Theoret. Comput. Sci., 289 (2002), pp. 457–471.
[43] L. G. Valiant, *Quantum circuits that can be simulated classically in polynomial time*, SIAM J. Comput., 31 (2002), pp. 1229–1254.
[44] L. G. Valiant, *Accidental algorthims*, in Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science, Washington, DC, 2006, pp. 509–517, https://doi.org/10.1109/FOCS.2006.7.

[45] L. G. Valiant, *Holographic algorithms*, SIAM J. Comput., 37 (2008), pp. 1565–1594, https://doi.org/10.1137/070682575.
[46] D. Welsh, *Complexity: Knots, Colourings and Counting*, Cambridge University Press, Cambridge, UK, 1993.
[47] M. Xia, P. Zhang, and W. Zhao, *Computational complexity of counting problems on* 3*-regular planar graphs*, Theoret. Comput. Sci., 384 (2007), pp. 111–125.
[48] C. N. Yang, *The spontaneous magnetization of a two-dimensional Ising model*, Phys. Rev., 85 (1952), pp. 808–816.
[49] C. N. Yang and T. D. Lee, *Statistical theory of equations of state and phase transitions.* I. *Theory of condensation*, Phys. Rev., 87 (1952), pp. 404–409.