Holographic Algorithms with Matchgates Capture Precisely Tractable Planar #CSP

Jin-Yi Cai
University of Wisconsin-Madison, and Peking University.
jyc@cs.wisc.edu

Pinyan Lu
Microsoft Research Asia.
pinyanl@microsoft.com

Mingji Xia
Institute of Software, Chinese Academy of Sciences.
xmjljx@gmail.com

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

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

I. 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 a *partition function*. In the special case where $f \in \mathcal{F}$ outputs values in $\{0, 1\}$ it counts the number of satisfying assignments. But constraint functions taking real or complex values are also interesting, called (real or complex) weighted #CSP. Our $\mathcal{F}$ consists of real or complex valued functions in general. There is a deeper reason for allowing this generality: The theory of *holographic reductions* is a powerful tool which operates naturally over $\mathbb{C}$, even if the original problem has only 0-1 valued functions.

A closely related framework for locally constrained counting problems is called Holant Problems [7], [9]. This framework is inspired by the introduction of *Holographic Algorithms* by L. Valiant [28], [27]. In two ground-breaking papers [26], [28] Valiant introduced matchgates and holographic algorithms based on matchgates to solve a number of problems in polynomial time, which appear to require exponential time. At the heart of these exotic algorithms is a tensor transformation from a given problem to the problem of counting (complex) weighted perfect matchings over planar graphs. The latter problem has a remarkable P-time algorithm (FKT-algorithm) [23], [15], [16]. Planarity is crucial, as counting perfect matchings over general graphs is #P-hard [24]. Most of these holographic algorithms use a suitable linear basis to realize locally a *symmetric* function with at most 3 Boolean variables on a matchgate. This work has been extended in [5]. In particular we have obtained a complete characterization of all realizable symmetric functions by matchgates over the complex field $\mathbb{C}$.

The study of "tractable #CSP" type problems has a much longer history in the statistical physics community (under different names). Ever since Wilhelm Lenz who invented what is now known as the Ising model, and asked his student Ernst Ising [12] to work on it, physicists have studied so-called "Exactly Solved Models" [1], [21]. In the language of modern complexity theory, physicists' notion of an "Exactly Solvable" system corresponds to systems with polynomial time computable partition functions. This is captured completely by the computer science notion of "tractable #CSP". Many great researchers in physics made remarkable contributions to this intellectual edifice, including Ising, Onsager, C.N.Yang, T.D.Lee, Fisher, Temperley, Kasteleyn, Baxter, Lieb, Wilson etc [12], [22], [30], [31], [19], [23], [15], [16], [1], [20]. 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 in terms of intrinsic solvability, except that the known method does not work in higher dimensions. 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 shed light on why the valiant efforts by physicists to generalize the "exactly solved" planar systems to higher dimensions failed.

Now turning from Physics to CS proper, Istrail [13] 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. A classical paper by Jerrum and Sinclair gave a randomized algorithm for an arbitrary ferromagnetic Ising system [14]. After Valiant introduced his holographic algorithms with matchgates, the following question can be raised: Do these novel algorithms capture all P-time tractable counting problems on planar graphs, *or* are there other more exotic algorithmic paradigms yet undiscovered? A suspicion (and perhaps an audacious proposition) is that they have indeed captured all tractable planar counting problems. If so it would provide a universal methodology to a broad class of counting problems studied in statistical physics and beyond. The results of this paper can be viewed as an affirmation of that suspicion. Within the framework of weighted Boolean #CSP problems our answer is YES, for *all* symmetric real valued functions.

While #CSP problems provide a natural framework to address this question, it turns out that a deeper understanding 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 LHS are labeled by variables and RHS are labeled by constraints. Denote by $=_k$: $\{0,1\}^k \to \{0,1\}$ the EQUALITY function of arity k, which is 1 on 0^k and 1^k, and is 0 elsewhere. Then we can turn the #CSP instance to an input graph of a Holant problem, by replacing every variable vertex v on LHS by $=_{\deg(v)}$. In fact, #CSP($\mathcal{F}$) is exactly the same as Holant($\mathcal{F} \cup \{=_k \mid k \geq 1\}$). Thus, #CSP problems can be viewed as Holant problems where all EQUALITY functions are available for free, or assumed to be present. It turns out that the main technical breakthrough for our dichotomy theorem for planar #CSP comes from Holant problems.

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

In [9] we obtained a dichotomy theorem for (complex) Holant* problems and (real) Holant^c problems for symmetric local constraint functions. The dichotomy criterion for Holant* problems is still valid for *planar graphs*. The proof of dichotomy theorems in this paper starts from there. We note crucially that in Holant* one cannot differentiate perfect matchings from general matchings, with the former being tractable and the latter remain hard for planar graphs.

In Section III, 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) iff either it is tractable over general graphs (for which we already have an effective dichotomy theorem [9]), or it is tractable because every function in $\mathcal{F}$ is realizable by a matchgate, in which case the planar Holant$^c(\mathcal{F})$ problem is computable by matchgates in P-time using FKT. In *all other cases* the problem is #P-hard.[1] A crucial ingredient of the proof is a crossover construction whose validity is proved algebraically, which seems to defy any direct combinatorial justification. Many additional ideas are used, including a successive "squeeze" that eventually isolates the set of *matchgates* precisely.

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

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

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

The three dichotomy theorems are not mutually subsumed by each other and are of independent interest. In each framework the respective theorem is a demonstration that holographic algorithms with matchgates capture precisely those #P-hard problems which become tractable for planar graphs. Many proof details are omitted for lack of space; see the full paper [10].

II. Preliminaries

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 $H = (V, E)$ with each vertex labeled by a function $f_v \in \mathcal{F}$. The Holant problem on instance Ω is to compute $\mathrm{Holant}_\Omega = \sum_\sigma \prod_{v \in V} f_v(\sigma\ |_{E(v)})$, a sum over all edge assignments $\sigma : E \to \{0, 1\}$. A function f_v is given by a vector in $\mathbb{C}^{2^{\deg(v)}}$ listing its values, or a tensor in $(\mathbb{C}^2)^{\otimes \deg(v)}$. A function $f \in \mathcal{F}$ is also called a *signature*. A Holant problem is parameterized by a set of signatures.

Definition II.1. *Given a set of signatures $\mathcal{F}$, we define a counting problem* $\mathrm{Holant}(\mathcal{F})$ *(reps.* Pl-Holant$(\mathcal{F})$*):*

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

Output: Holant_Ω.

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

Definition II.2. *Let $\mathcal{U}$ denote the set of all unary signatures. Given a set of signatures $\mathcal{F}$, we use* $\mathrm{Holant}^*(\mathcal{F})$ *(or* Pl-Holant$^*(\mathcal{F})$ *respectively) to denote* $\mathrm{Holant}(\mathcal{F} \cup \mathcal{U})$ *(or* Pl-Holant$(\mathcal{F} \cup \mathcal{U})$ *respectively).*

[2]Usually our set of signatures $\mathcal{F}$ is a finite set, and the assertion of either $\mathrm{Holant}(\mathcal{F})$ 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. $\mathrm{Holant}(\mathcal{F})$ is tractable means that it is computable in P even when we include the description of the signatures in the input Ω in the input size. $\mathrm{Holant}(\mathcal{F})$ is #P-hard means that there exists a finite subset of $\mathcal{F}$ for which the problem is #P-hard.

Definition II.3. *Given a set of signatures $\mathcal{F}$, we use* $\mathrm{Holant}^c(\mathcal{F})$ *(or* Pl-Holant$^c(\mathcal{F})$ *respectively) to denote* $\mathrm{Holant}(\mathcal{F} \cup \{[1, 0], [0, 1]\})$ *(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 $\mathrm{Holant}(\mathcal{F})$. So we view f and cf as the same signature. An important property of a signature is whether it is degenerate. 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 $\lambda[x, y]^{\otimes k}$.

A signature from $\mathcal{F}$ at a vertex is considered a basic realizable function. 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 with some dangling edges D. Other than these dangling edges, an $\mathcal{F}$-gate is the same as a signature grid. The role of dangling edges is similar to that of external nodes in Valiant's notion [26], [28], however we allow more than one dangling edges for a node. In $H = (V, E, D)$ each node is assigned a function in $\mathcal{F}$ by the mapping π (we do not consider "dangling" leaf nodes at the end of a dangling edge among these), E is the set of regular edges, denoted as $1, 2, \ldots, m$, and D is the set of dangling edges, denoted as $m + 1, m + 2, \ldots, m + n$. Then we can define a function for this $\mathcal{F}$-gate $\Gamma = (H, \mathcal{F}, \pi)$, $\Gamma(y_1, y_2, \ldots, y_n) = \sum_{x_1 x_2 \cdots x_m} H(x_1 x_2 \cdots x_m y_1 y_2 \cdots y_n)$, where $(y_1, y_2, \ldots, y_n) \in \{0, 1\}^n$ denotes an assignment on the dangling edges and $H(x_1 x_2 \cdots x_m y_1 y_2 \cdots y_n)$ denotes the value of the signature grid on an assignment of all edges. We will also call this function the signature of the $\mathcal{F}$-gate Γ. An $\mathcal{F}$-gate can be used in a signature grid as if it is just a single node with the particular signature.

Using the idea of $\mathcal{F}$-gates, we can reduce one Holant problem to another. Let g be the signature of some $\mathcal{F}$-gate Γ. Then $\mathrm{Holant}(\mathcal{F} \cup \{g\}) \leq_T \mathrm{Holant}(\mathcal{F})$. The reduction is quite simple. Given an instance of $\mathrm{Holant}(\mathcal{F} \cup \{g\})$, by replacing every appearance of g by an $\mathcal{F}$-gate Γ, we get an instance of $\mathrm{Holant}(\mathcal{F})$. 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 [26], [25], [28], whose definition is combinatorial in nature. Matchgates can be viewed as a special case of planar $\mathcal{F}$-gates, where $\mathcal{F}$ contains Exact-One functions of all arities and weight functions ($[1, 0, w]$, $w \in \mathbb{C}$) on edges. The signature function Γ defined above for a matchgate is called a matchgate signature, or a standard signature. A signature function is realizable by a matchgate if it is the standard signature of that matchgate. We have a complete characterization of realizable signatures by matchgates [3]. For symmetric signatures, the characterization is simple. It satisfies the parity condition, i.e, all the even indexed values are zero or all the odd indexed values are zero. After removing these zero entries, the remaining sequence is a geometric sequence. (After a holographic transformation, a signature function is realizable under a basis if it is the transformed signature of a matchgate; see below. We also have a complete characterization of all symmetric signatures realizable by matchgates under a holographic transformation [6].)

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 adding an additional vertex on each edge, and giving each new vertex the EQUALITY function $=_2$ on 2 inputs.

We use $\text{Holant}(\mathcal{G}|\mathcal{R})$ 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) [11].

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 $\text{Holant}(\mathcal{G}|\mathcal{R})$ and $\text{Holant}(\mathcal{G}'|\mathcal{R}')$ are two Holant problems defined for the same family of graphs, and $T \in \mathbf{GL}_2(\mathbb{C})$ is a basis. We say that there is an (invertible) holographic reduction from $\text{Holant}(\mathcal{G}|\mathcal{R})$ to $\text{Holant}(\mathcal{G}'|\mathcal{R}')$, 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*.) Suppose there is a holographic reduction from $\#\mathcal{G}|\mathcal{R}$ to $\#\mathcal{G}'|\mathcal{R}'$ mapping signature grid Ω to Ω', then $\text{Holant}_\Omega = \text{Holant}_{\Omega'}$ [28]. 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 non-bipartite settings. When this does not cause confusion, we do not distinguish signatures between column vectors (or contravariant tensors) and row vectors (or covariant tensors). Whenever we write a transformation as $T^{\otimes n}F$ or $T\mathcal{F}$, we view the signature or signatures as column vectors (or contravariant tensors); whenever we write a transformation as $FT^{\otimes n}$ or $\mathcal{F}T$, we view them as row vectors (or covariant tensors).

III. DICHOTOMY FOR PLANAR HOLANT^c PROBLEMS

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

Theorem III.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 III.2. *Let $a, b, x \in \mathbb{R}$, $ab \neq 0$ and $x \neq \pm 1$. Then Pl-Holant$^c(\{[a, 0, 0, 0, b], [0, 1, 0, x]\})$ is #P-hard.*

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

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

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

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

Let $F = [0, 1, 0, x]$ and let $H(x_1, x_2, y_1, y_2) = \sum_{z=0,1} F(x_1, y_1, z)F(x_2, y_2, z)$. This H is realizable by connecting one pair of edges of two copies of F. (Fig. 1.)

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

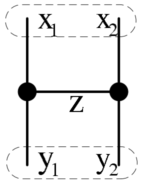

Figure 1. The gadget for function H.

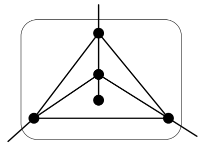

Figure 2. The gadget for function $[1, 1, x^{2k}]$.

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 $k \geq 3$. Then we can realize $[1, 1, x^{2k}]$, for all $k \geq 1$. (See Figure 2.) If $x = 0$, then we already have $[1, 1, 0]$.

Suppose $x \neq 0$. Because $x^2 \neq 1$ and being a positive real number, we can realize $[1, 1, 0]$ by interpolation. Now we have reduced the problem Pl-Holant$([1, 0, 0, 1] \mid [1, 1, 0])$ to Pl-Holant$^c(\{[1, 0, 0, 1], [1, 1, x^2]\})$. The bipartite problem Pl-Holant$([1, 0, 0, 1] \mid [1, 1, 0])$ is #P-hard since it is counting VERTEX COVERS on planar 3-regular graphs [29]. ■

The following lemma handles a special case of Theorem III.1. The proof uses Lemma III.2.

Lemma III.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 III.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 3.

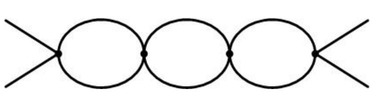

Figure 3. The gadget for H_4.

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

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

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

Note that the Holant value on G_i is

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

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

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

The following result can be proved by interpolation as well, whose proof is omitted here and can be found in the full paper.

Lemma III.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* $[a, 0, 1, 0]$.

Proof of Theorem III.1 If $ab = 1$, then $[a, 0, 1, 0, b]$ is realizable by some matchgate [3]. This realizability also applies to the unary functions $[1, 0]$ and $[0, 1]$. Hence the problem Pl-Holant$^c([a, 0, 1, 0, b])$ can be solved in polynomial time by matchgate computation via the FKT method [23], [15], [16]. 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 III.3. Now we can assume at least one of a and b is non-zero, and by symmetry we assume $a \neq 0$.

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

If $\{a, b\} \not\subset \{-1, 0, 1\}$, we can use Lemma III.4 to interpolate all $[x, 0, y, 0]$, for $x, y \in \mathbb{C}$. If $\{a, b\} \subset \{-1, 0, 1\}$, then there are only four cases: $[1, 0, 1, 0, -1]$, $[1, 0, 1, 0, 0]$, $[-1, 0, 1, 0, 1]$ and $[-1, 0, 1, 0, 0]$. In all four cases, it is easy to verify that we can realize a signature with a form $[c_1, 0, c_2, 0]$ where $c_1 c_2 \neq 0$ and $c_1 \neq \pm c_2$ using the gadget in Figure 4. (For $[1, 0, 1, 0, -1]$, we get $[8, 0, 4, 0]$ by using $[1, 0]$ 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 $[c', 0, 1, 0]$, where $c' \in \mathbb{R}$ and $c' \notin \{0, \pm 1\}$. As a result, we can also interpolate all $[x, 0, y, 0]$, where $x, y \in \mathbb{C}$.

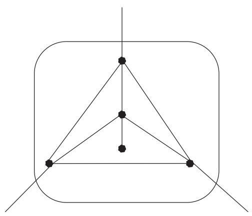

Figure 4. The signature of the degree 1 vertex in the gadget is $[1, 0]$ or $[0, 1]$.

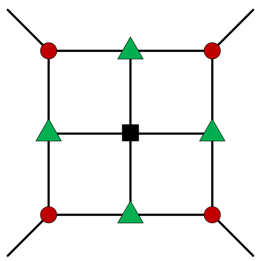

Figure 5. This gadget is to realize the Cross function. The signature for the center vertex (square) is $[t, 0, 1, 0, \frac{c}{t}]$. The signature for the vertices in the four corners (circle) is $[x, 0, 1, 0]$. The signature for the vertices on the four sides (triangle) is $[y, 0, 1, 0]$.

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 5 to realize the cross function X. In the equations below x, y, t are three variables we can set to any complex numbers, with $t \neq 0$. The parameter c is given and not equal to 1.

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

We can compute the signature of the gadget in Fig. 5. If the input has an odd number of 1s, the value is 0. For other inputs, we have

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

What we need is that $X_{0000} = X_{1111} = X_{0101} \neq 0$ and $X_{0011} = 0$. In the full paper, we show that we can always satisfy these conditions by choosing suitable $x, y, t \in \mathbb{C}$ and $t \neq 0$, for all $c \neq 1$. Details are omitted here. □

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

Theorem III.5. *Let $\mathcal{F}$ be a set of real symmetric signatures.* Pl-Holant$^c(\mathcal{F})$ *is #P-hard unless $\mathcal{F}$ satisfies one of the following conditions, in which case it is tractable:*

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

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

The main idea of the proof is to interpolate all unary functions. If we can do that, we can reduce the problem Pl-Holant$^*(\mathcal{F})$ to Pl-Holant$^c(\mathcal{F})$ and finish the proof. We note that our dichotomy in [9] for Holant$^*(\mathcal{F})$ also holds for planar graphs. In some cases, we cannot interpolate all unary functions, then we prove the theorem separately, mainly using Lemma III.2 and Theorem III.1. The following lemma is for interpolation of unary functions; the proof is omitted.

Lemma III.6. *If we can construct from $\mathcal{F}$ a gadget with signature $[a, b, c]$, where $b^2 \neq ac$, $b \neq 0$ and $a + c \neq 0$, then we can interpolate all unary functions. (Hence the conclusions of Theorem III.5 hold.)*

If we can construct from $\mathcal{F}$ a gadget with a binary symmetric signature $[a, b, c]$, which satisfies all the conditions in Lemma III.6, then we are done. For most cases, we prove the theorem by interpolating all unary signatures. However, in some more delicate cases, we are not able to do that. For example, if all signatures from $\mathcal{F}$ have the parity condition, which includes a proper superset of matchgate signatures, then all unary signatures we can realize have form $[a, 0]$ or $[0, a]$, so we can not interpolate all unary signatures. For these cases, our starting point is Theorem III.1.

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

$$\mathcal{G}_1 = \{[a, 0, 0, \cdots, 0, b] \mid ab \neq 0\}$$

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

The following several lemmas all have the form "If $\mathcal{F} \not\subseteq \mathcal{A}$, then the conclusions of Theorem III.5 hold." After proving each lemma, in subsequent lemmas, we only need to consider the case that $\mathcal{F} \subseteq \mathcal{A}$.

Lemma III.7. *If $\mathcal{F} \not\subseteq \mathcal{G}_1 \cup \mathcal{G}_2 \cup \mathcal{G}_3$, then the conclusions of Theorem III.5 hold.*

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

(1) f has some zero entries. If there exists a sub-signature of f of the form $[0, a, b]$ or $[a, b, 0]$, where $ab \neq 0$, then we are done by Lemma III.6. Otherwise, we can conclude that there are no two successive non-zero entries. So the signature f has this form $[0^{i_0} x_1 0^{i_1} x_2 0^{i_2} \cdots x_k 0^{i_k}]$, where $k \geq 1$, $x_j \neq 0$ and for all $1 \leq j \leq k-1$, $i_j \geq 1$. If for all $1 \leq j \leq k-1$, i_j is odd, (including $k = 1$), then $f \in \mathcal{G}_2$, a contradiction. Otherwise there exists a sub-signature of form $[x, 0, 0, \cdots, 0, y]$, where $xy \neq 0$ and there are a positive even number of 0s between x and y. If this is the entire f, then $f \in \mathcal{G}_1$, a contradiction. So there is one 0 before x or after y. By symmetry, we assume there is a 0 before x, so we have a sub-signature $[0, x, 0, 0, \cdots, 0, y]$, whose arity is even and at least 4. We label its dangling edges $1, 2, \cdots, 2k$. Then for every $i = 1, 2, \cdots, k-1$, we connect dangling edges $2i+1$ and $2i+2$ together to form a regular edge. After that, we have an $\mathcal{F}$-gate with arity 2, and its signature is $[0, x, y]$. Then we are done by Lemma III.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 III.6. Suppose all sub-signatures of f with arity 2 do not satisfy all the three conditions. For each sub-signature $[a', b', c']$, either $a' + c' = 0$, or $b'^2 = a'c'$. If all of them satisfy $a' + c' = 0$, then $f \in \mathcal{G}_3$. A contradiction. If all of them satisfy $b'^2 = a'c'$, then f is degenerate. A contradiction. W.l.o.g., we can assume there is a sub-signature $[a, b, c, d]$ of f, such that $a + c = 0$, $b + d \neq 0$, and $c^2 = bd$. We get this sub-signature $[a, b, c, d]$ by $[1, 0]$ and $[0, 1]$. Combining two $[a, b, c, d]$, we can get a function $[a', b', c'] = [a^2 + 2b^2 + c^2, ab + 2bc + cd, b^2 + 2c^2 + d^2] = [2(b^2 + c^2), c(b + d), (b + d)^2]$. Then $b' = c(b + d) \neq 0$. $a' + c' > 0$. And $a'c' - b'^2 = (b + d)^2(2b^2 + c^2) > 0$. We are done by Lemma III.6. ∎

The following lemma uses Theorem III.1 in an essential way, which in turns depends on the crossover.

Lemma III.8. *If $\mathcal{F} \not\subseteq \mathcal{G}_1 \cup \mathcal{M} \cup \mathcal{G}_3$, then the conclusions of Theorem III.5 hold.*

Proof: If $\mathcal{F} \not\subseteq \mathcal{G}_1 \cup \mathcal{G}_2 \cup \mathcal{G}_3$, then by Lemma III.7, we are done. Otherwise, there exists a signature $f \in \mathcal{F} \subseteq \mathcal{G}_1 \cup \mathcal{G}_2 \cup \mathcal{G}_3$ and $f \notin \mathcal{G}_1 \cup \mathcal{M} \cup \mathcal{G}_3$. Then it must be the case that $f \in \mathcal{G}_2$. Note that every signature with arity at most 3 in $\mathcal{G}_2$ (this is called the parity condition) is also contained in $\mathcal{M}$, so f is of arity greater than 3. Let $f = [x_0, x_1, \cdots, x_n]$, for some $n \geq 4$. Suppose there exists some $i \in [2, 3, \cdots, n-2]$ such that $x_i \neq 0$. If $x_{i-2}x_{i+2} \neq x_i^2$, then we can get $[x_{i-2}, 0, x_i, 0, x_{i+2}]$ by $[1, 0]$ and $[0, 1]$ which restrict the signature to a sub-signature. Then the problem is #P-hard by Theorem III.1 and we are done. Otherwise, we have $x_{i-2}x_{i+2} = x_i^2 \neq 0$. Then starting from $x_{i-2} \neq 0$ and if $i - 2 \in [2, 3, \cdots, n-2]$, we can get $x_{i-4}x_i = x_{i-2}^2 \neq 0$. Similarly we can start with x_{i+2}. A signature satisfying the parity condition and is a geometric series on the alternate entries is realizable by a matchgate [2], [3], a contradiction.

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

Lemma III.9. *If $[0, 1, 0, x] \in \mathcal{F}$ (or $[1, 0, x, 0] \in \mathcal{F}$) where $x \in \mathbb{R}$, $x \neq \pm 1$, then the conclusions of Theorem III.5 hold.*

Proof: If $x \neq 0$, we can use Lemma III.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 III.8, we are done. If $\mathcal{F} \subseteq \mathcal{M}$, then the problem is tractable and we are done. Otherwise, there exists a signature $f \in \mathcal{F} \subseteq \mathcal{G}_1 \cup \mathcal{M} \cup \mathcal{G}_3$ and $f \notin \mathcal{M}$. That is $f \in (\mathcal{G}_1 \cup \mathcal{G}_3 - \mathcal{M})$.

If f has arity ≥ 1 and of the form $[x_0, x_1, -x_0, -x_1, x_0 \cdots] \in \mathcal{G}_3 - \mathcal{M}$, then we will have $x_0x_1 \neq 0$. Otherwise we would have $f \in \mathcal{M}$, a contradiction. Connecting one unary signature $[x_0, x_1]$ to $[0, 1, 0, 0]$, we get $[x_1, x_0, 0]$ which satisfies all the conditions in Lemma III.6, and we are done.

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

The proof of Theorem III.5 is continued with 4 more lemmas, making successive "squeezes" on the class $\mathcal{F}$. They are omitted here and can be found in the full paper.

IV. Dichotomy for Planar Weighted #CSP

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

Theorem IV.1. *Let $\mathcal{F}$ be a set of real symmetric functions.* Pl-#CSP($\mathcal{F}$) *is #P-hard unless $\mathcal{F}$ satisfies one of the following conditions, in which case it is tractable:*

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

The main proof idea is to reduce Pl-Holant^c problems to Pl-#CSP problems. Pl-#CSP($\mathcal{F}$) is exactly the same as planar Holant with all the Equality functions, i.e., Pl-Holant($\mathcal{F}\cup\{[1,1],[1,0,1],[1,0,0,1],[1,0,0,0,1],\ldots\}$). We can use a holographic reduction under the basis $H = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$. Under this transformation, the problem is transformed to, and hence has the same complexity as Pl-Holant($H\mathcal{F} \cup \{[1,0],[1,0,1],[1,0,1,0],[1,0,1,0,1],\ldots\}$). Since this holographic reduction gives us $[1,0]$ (from $[1,1]$), if we can further realize (or interpolate) $[0,1]$, we can view the problem as a Pl-Holant^c problem and apply Theorem III.5 to $H\mathcal{F} \cup \{[1,0,1],[1,0,1,0],[1,0,1,0,1],\ldots\}$ to get a proof of Theorem IV.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 III.5 to Theorem IV.1 is straightforward.

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

Lemma IV.2. *The planar holant problem Pl-Holant($\mathcal{F} \cup \{[1,0],[0,0,1],[1,0,1],[1,0,1,0],[1,0,1,0,1],\ldots\}$) is #P-hard (or in P) if and only if the problem Pl-Holant^c($\mathcal{F} \cup \{[0,0,1],[1,0,1],[1,0,1,0],[1,0,1,0,1],\ldots\}$) is #P-hard (or in P).*

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

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

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

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

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

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

To sum up the above discussion, and apply Theorem III.5, we know that if we can realize (or interpolate) $[0,1]$ or $[0,0,1]$ from $H\mathcal{F} \cup \{[1,0],[1,0,1],[1,0,1,0],[1,0,1,0,1],\ldots\}$, then the conclusion of Theorem IV.1 holds. The main effort of the remaining proof is to do that, which is omitted here due to space limitation. The general thrust of the proof is to squeeze all possible f into several standardized forms, and either prove #P-hardness or reach a contradiction. The detail of the proof can be found in the full paper.

V. Dichotomy for Planar 2-3 Regular Graphs

In this section we prove a dichotomy for Holant on planar 2-3 regular graphs. This setting is very interesting for at least two reasons. From dichotomy theorem point of view, this is the simplest nontrivial setting and always serves as the starting point of more general dichotomy theorems as in [9], [4]. This was also a focus of several previous work [7], [17], [8], [18], whose result is the starting point of this theorem. From the holographic algorithms point of view, most of the known holographic algorithms [28], [27] 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 V.1. *Let $[y_0,y_1,y_2]$ and $[x_0,x_1,x_2,x_3]$ be two complex symmetric signatures with arity 2 and 3 respectively. Then* Pl-Holant$([y_0,y_1,y_2]|[x_0,x_1,x_2,x_3])$ *is #P-hard unless $[y_0,y_1,y_2]$ and $[x_0,x_1,x_2,x_3]$ satisfy one of the following conditions, in which case it is tractable:*

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

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

For Pl-Holant$([z_0,z_1,z_2]|[1,0,0,1])$, a dichotomy theorem proved in [18], is also valid for planar structures. By that theorem, the only case which is hard for general graphs and tractable for planar graphs is $z_0^3 = z_2^3$. This condition is exactly the same as the condition that there exists a basis T such that both $[y_0,y_1,y_2](T^{-1})^{\otimes 2}$ and $T^{\otimes 3}[1,0,0,1]$ are realizable by some matchgates. This proves Theorem V.1 for case (1).

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

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

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

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

Acknowledgments

Jin-Yi Cai is supported by NSF CCF-0830488 and CCF-0511679. Mingji Xia is supported by the Grand Challenge Program "Network Algorithms and Digital Information" of the Institute of Software, CAS and partially supported by NSFC 60970003. We thank the following colleagues for their interests 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 also thank the anonymous referees for their helpful comments.

References

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

[2] Jin-Yi Cai and Vinay Choudhary. Some results on matchgates and holographic algorithms. In *ICALP (1)*, volume 4051 of *Lecture Notes in Computer Science*, pages 703–714. 2006.

[3] Jin-Yi Cai, Vinay Choudhary, and Pinyan Lu. On the theory of matchgate computations. In *CCC '07*, pages 305–318, 2007.

[4] Jin-Yi Cai, Sangxia Huang, and Pinyan Lu. From holant to #CSP and back: Dichotomy for holant^c problems. *arXiv 1004.0803*, 2010.

[5] Jin-Yi Cai and Pinyan Lu. Holographic algorithms: from art to science. In *STOC '07*, pages 401–410, 2007.

[6] Jin-Yi Cai and Pinyan Lu. On symmetric signatures in holographic algorithms. *Theory Comput. Syst.*, 46(3):198–415, 2010.

[7] Jin-Yi Cai, Pinyan Lu, and Mingji Xia. Holographic algorithms by fibonacci gates and holographic reductions for hardness. In *FOCS '08*, pages 644-653, 2008.

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

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

[10] Jin-Yi Cai, Pinyan Lu, and Mingji Xia. Holographic algorithms with matchgates capture precisely tractable planar #CSP. *arXiv 1008.0683*, 2010.

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

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

[13] Sorin 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 *STOC*, pages 87–96, 2000.

[14] Mark Jerrum and Alistair Sinclair. Polynomial-time approximation algorithms for the ising model. *SIAM J. Comput.*, 22(5):1087–1116, 1993.

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

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

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

[18] Michael Kowalczyk and Jin-Yi Cai. Holant problems for regular graphs with complex edge functions. In *STACS*, pages 525-536, 2010.

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

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

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

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

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

[24] Leslie G. Valiant. The complexity of computing the permanent. *Theor. Comput. Sci.*, 8:189–201, 1979.

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

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

[27] Leslie G. Valiant. Accidental algorthims. In *FOCS '06*, pages 509–517, 2006.

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

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

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

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