Electronic Colloquium on Computational Complexity, Report No. 93 (2011)

Complexity Dichotomies of Counting Problems

Pinyan Lu
Microsoft Research Asia.
pinyanl@microsoft.com

Abstract

In order to study the complexity of counting problems, several interesting frameworks have been proposed, such as Constraint Satisfaction Problems (#CSP) and Graph Homomorphisms. Recently, we proposed and explored a novel alternative framework, called Holant Problems. It is a refinement with a more explicit role for constraint functions. Both graph homomorphism and #CSP can be viewed as special sub-frameworks of Holant Problems. One reason such frameworks are interesting is because the language is expressive enough so that they can express many natural counting problems, while specific enough so that it is possible to prove complete classification theorems on their complexity, which are called dichotomy theorems. From the unified prospective of a Holant framework, we summarize various dichotomies obtained for counting problems and also proof techniques used. This survey presents material from the talk given by the author at the 4-th International Congress of Chinese Mathematicians (ICCM 2010).

1 Introduction

The complexity of counting problems is a fascinating subject. Valiant defined the class #P to capture most of these counting problems [Val79b]. Beyond the complexity of individual problems, there has been a great deal of interest in proving complexity dichotomy theorems which state that for a wide class of counting problems, every problem in the class is either computable in polynomial time (tractable) or #P-hard. One such framework is called counting Constraint Satisfaction Problems (#CSP) [CH96, BD03, DGJ07, Bul08, CLX09b, DR10b, CCL11]. Another well studied framework is called Graph Homomorphisms or H-coloring problems, which can be viewed as a special case of #CSP problems [DG00, BG05, DGP07, GGJT10, CCL10, CC10]. One reason such frameworks are interesting is because the language is *expressive* enough so that they can express many natural counting problems, while *specific* enough so that it is possible to prove complete classification theorems on their complexity [CKS01]. According to a theorem of Ladner [Lad75], if P $\neq$ NP, or P $\neq$ #P, then such a dichotomy for NP or #P is false.

ISSN 1433-8092

The study of "tractable #CSP" type problems has a much longer history in the statistical physics community (under different names). Ever since Wilhelm Lenz invented what is now known as the Ising model, and asked his student Ernst Ising [Isi25] to work on it, physicists have studied so-called "Exactly Solved Models" [Bax82, MW73]. 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 [Isi25, Ons44, Yan52, YL52, LY52, TF61, Kas61, Kas67, Bax82, LS81]. A central question is to identify what "systems" can be solved "exactly" and what "systems" are "difficult".

The natural counting problems which can be expressed as graph homomorphism problems include counting the number of vertex covers, the number of k-colorings in a graph, and many others. However, there are some natural and important counting problems, which cannot be expressed as a graph homomorphism problem. In [FLS07], it is proved that counting the number of perfect matchings in a graph *cannot* be expressed as a graph homomorphism function. Additionally, sometimes a problem can be expressed in the existing framework, such as #CSP, but only with some contrived restrictions. Recently, we proposed and explored an alternative framework to study the complexity of counting problems, called Holant Problems. This notion is motivated by holographic reductions proposed by Valiant [Val08, Val06]. Compared to #CSP, it is a refinement with a more explicit role for the constraint functions. Both graph homomorphism and #CSP can be viewed as special cases of Holant Problems. We give a brief description here and a more formal definition is given in Section 2. A *signature grid* $\Omega = (G, \mathcal{F}, \pi)$ is a tuple, where $G = (V, E)$ is a graph, $\mathcal{F}$ is a set of functions, and π maps each $v \in V(G)$ to a function $f_v \in \mathcal{F}$. Edges are variables, and we consider all edge assignments. An assignment σ for every $e \in E$ gives an evaluation $\prod_{v \in V} f_v(\sigma \mid_{E(v)})$, where $E(v)$ denotes the incident edges of v, and f_v is evaluated on the restriction of σ on $E(v)$. The counting problem on an input instance Ω is to compute

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

For example, consider the Perfect Matching problem on G. This problem corresponds to attaching the Exact-One function at every vertex of G. Consider all 0-1 edge assignments σ. The product $\prod_{v \in V} f_v(\sigma \mid_{E(v)})$ evaluates to 0 or 1, and is 1 iff $\sigma^{-1}(1) \subseteq E$ is a perfect matching. Hence in this case, Holant_Ω counts the number of perfect matchings. If we use the At-Most-One function at every vertex, then we are counting all (not necessarily perfect) matchings. So this new framework can express some natural counting problems which are not expressible as graph homomorphisms.

Our Holant Problem framework is strongly influenced by the development of holographic algorithms and holographic reductions [Val08, Val06, CL11, CLX08].

Indeed, we use and develop holographic reductions as one of the primary techniques, which has not previously been used in the studied of #CSP. One advantage of our new framework is that one can naturally consider new subclasses of counting problems as special cases of Holant problems other than #CSP problems. Indeed, there are many dimensions of the framework. #CSP can be viewed as a special sub-framework of Holant by assuming that all equality functions are freely available. This gives one dimension of the framework. By assuming other freely available functions, we can define other interesting sub-frameworks. Graph Homomorphism can be viewed as a further special case of #CSP, whose function set contains only one single binary function. To study a single function or to study a set of functions is the second dimension we discuss in this survey. The third dimension is the domain size. For example, Boolean #CSP restricts #CSP to a domain of size 2. We can also restrict the functions in the set $\mathcal{F}$. For example, the value range of the functions can be $\{0, 1\}$, non-negative, real or complex. And we can also restrict to symmetric functions. In addition, we can restrict the input instants. For example, we can consider only inputs with a planar structure. From this unified perspective, most of the known dichotomies of counting can be viewed as dichotomy for a sub-framework of Holant by restricting its dimensions to a certain setting. For example, real weighted symmetric Boolean planar #CSP is a sub-framework of Holant that assumes that all equality functions are freely available, the domain size is 2, all functions are symmetric and real valued and the inputs are of planar structure. In this survey, we analyze these important dimensions of the framework, and summarize various known dichotomies and main proof techniques along the way.

Organization of the Survey

In Section 2, we formally define the framework of Holant Problems and some other basic notations. Section 3 summarizes some reduction techniques in this framework, which are also main proof approaches used in various results. Section 4 is the main section. We carefully discuss many dimensions of the framework, summarize various known complexity dichotomies, and propose many open questions.

2 Definitions and Background

A *signature grid* $\Omega = (H, \mathcal{F}, \pi)$ is a tuple. $H = (V, E)$ is a graph. $\mathcal{F}$ is a set of functions and a function $F \in \mathcal{F}$ with arity k is a mapping $[q]^k \to \mathbb{C}$. π is a mapping from the vertex set V to $\mathcal{F}$, satisfying that the arity of $\pi(v)$ is the same as the degree of v for any $v \in V$. We use F_v to denote the function $\pi(v)$. An assignment σ is a mapping $E \to [q]$ and gives an evaluation $\prod_{v \in V} F_v(\sigma \mid_{E(v)})$, where $E(v)$ denotes the incident edges of v. The counting problem on the instance Ω is to compute

$$\text{Holant}_\Omega = \sum_\sigma \prod_{v \in V} F_v(\sigma\mid_{E(v)}).$$

The term Holant was first introduced by Valiant in [Val08] to denote a related exponential sum. Cai, Xia and the current author first formally introduced this framework of counting in [CLX08, CLX09b]. We can view each function F_v as a truth table, and then we can represent it by a vector in $\mathbf{F}^{q^{d(v)}}$, or a tensor in $(\mathbf{F}^q)^{\otimes d(v)}$. This is called a *signature*. When we say "function", we slightly emphasize that it is a mapping. When we say "signature", we slightly emphasize that we view it as one objectee and that it is ready to do some linear transformations. We do not really distinguish "function" and "signature" in this survey.

A Holant problem is parameterized by a set of functions.

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

The main goal here is to characterize what kind of function set $\mathcal{F}$ makes the problem Holant($\mathcal{F}$) tractable (hard).

We use following notations to denote some special functions. $=_k$ denotes the equality function of arity k. Δ_s denotes the unary function which gives value 1 on inputs $s \in [q]$, and 0 on all other inputs.

A function is symmetric iff to apply a permutation of its input will not change value of the function. A symmetric function F on Boolean variables can be expressed by $[f_0, f_1, \ldots, f_k]$, where f_j is the value of F on inputs of weight j. For Boolean domain $[2] = \{0, 1\}$, $=_k = [1, 0, \ldots, 0, 1]$ and $\Delta_0 = [1, 0]$.

Replacing a signature $F \in \mathcal{F}$ by its scale cF, where $c \neq 0$, will not change the complexity of Holant($\mathcal{F}$). So we always view F and cF as the same signature. Another important property of signatures is degeneracy.

Definition 2.2. *A signature is called degenerate iff it can be decomposed into a tensor product of unary signatures.*

In particular, a symmetric signature over a Boolean domain is degenerate iff it can be expressed as $\lambda[x, y]^{\otimes k}$.

3 Reductions Among Problems

To prove dichotomy theorems for Holant problems, the main approach is to build reductions among different problems. In this section, we highlight several important methods to establish reductions.

3.1 Gadget Construction

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 [Val02, Val08], however we allow more than one dangling edge 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 $1, 2, \dots, m$, and D is the set of dangling edges, denoted as $m+1, m+2, \dots, m+n$. Then we can define a function for this $\mathcal{F}$-gate $\Gamma = (H, \mathcal{F}, \pi)$,

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

where $(y_1, y_2, \dots, y_n) \in \{0, 1\}^n$ 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

$$\text{Holant}(\mathcal{F} \cup \{g\}) \leq_T \text{Holant}(\mathcal{F}).$$

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.

3.2 Polynomial Interpolation

Polynomial interpolation is a powerful tool in the study of counting problems initiated by Valiant [Val79b] and further developed by Vadhan, Dyer and Greenhill [Vad01, DG00] and others [XZZ07].

We will use unary signatures $f = [x, y]$ over a Boolean domain as an example to introduce the idea of polynomial interpolation.

For some set of signatures $\mathcal{F}$, suppose we want to show that for all unary signatures $f = [x, y]$, we have

$$\text{Holant}(\mathcal{F} \cup \{[x, y]\}) \leq_{\mathrm{T}} \text{Holant}(\mathcal{F}).$$

Let $\Omega = (G, \mathcal{F} \cup \{[x, y]\}, \pi)$. We want to compute Holant_Ω in polynomial time using an oracle for Holant($\mathcal{F}$).

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

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

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

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

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

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

3.3 Holographic Reduction

To introduce the idea of holographic reduction, it is convenient (but not necessary) to consider bipartite graphs. We note that this is without loss of generality. For any general graph, we can make it bipartite by adding an additional vertex on each edge, and for each new vertex by giving it the Equality function on 2 inputs $=_2$.

We use $\text{Holant}(\mathcal{G}|\mathcal{R})$ to denote all the counting problem, expressed as Holant problem 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 of the Holant problem is a signature grid and is denoted as $\Omega = (H, \mathcal{G}|\mathcal{R}, \pi)$. Signatures in $\mathcal{G}$ are called generators, which are denoted by column vectors (or contravariant tensors); signatures in $\mathcal{R}$ are called recognizers, which are denoted by row vectors (or covariant tensors) [DP91].

One can perform (contravariant and covariant) tensor transformations on the signatures, which may produce exponential cancelations in tensor spaces. We will define a simple version of holographic reductions, which are invertible. 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 $T \in \mathbf{GL}(\mathbb{C})$ is a basis. We say that there is a 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 $R' \in \mathcal{R}'$, 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 3.1 (Holant Theorem)**.** *Suppose there is a holographic reduction from* $\text{Holant}(\mathcal{G}|\mathcal{R})$ *to* $\text{Holant}(\mathcal{G}'|\mathcal{R}')$ *mapping signature grid* Ω *to* Ω'*, then*

$$\text{Holant}_\Omega = \text{Holant}_{\Omega'}.$$

This theorem is due to Valiant [Val08]. The proof of this theorem follows from general principles of contravariant and covariant tensors [DP91], and a complete proof can be found in [CC07].

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

The following theorem is very useful as a way to normalize the given signature set $\mathcal{F}$.

Theorem 3.2. *Let* $\mathcal{F}$ *be a set of signatures and* M *be a* $q \times q$ *orthogonal matrix, i.e.,* $MM^{\mathrm{T}} = I_q$*. For any signature grid* $\Omega = (G, \mathcal{F}, \pi)$*, replacing every signature* $F \in \mathcal{F}$ *by* $M^{\otimes n}F$*, where* n *is the arity of* F*, we can get a new signature grid* Ω'*. Then*

$$\text{Holant}_\Omega = \text{Holant}_{\Omega'}.$$

Proof. First we use the standard technique to reformulate the signature grid $\Omega = (G, \mathcal{F}, \pi)$. We insert a new vertex at each edge of G with signature $=_2$. This will not change the value of the signature grid. Then for the new bipartite signature grid $\mathcal{F} \mid =_2$, we apply a holographic reduction on basis M. This will map a signature $F \in \mathcal{F}$ to $M^{\otimes n}F$, where n is the arity of F. It is an algebraic fact that the $=_2$ will map to itself. Then we view these (new) $=_2$ as an edge and ignore these vertices. This gives the signature grid Ω' as required. Due to the Holant theorem, its value is the same as Ω. $\square$

3.4 Existential Argument

This method is quite different from the above three. In the above, the reductions are quite constructive. The existential argument goes as follows. For a certain function set $\mathcal{F}$, if there exist some $\mathcal{F}$-gates which do not have some desired property, then we can prove #P-hardness or build a reduction. Otherwise, if all $\mathcal{F}$-gates satisfy the desired property, then we can make use of this fact and this desired property to give a polynomial time algorithm for $\text{Holant}(\mathcal{F})$. We will give an example in Section 4 when we discuss dichotomies proved by such an argument.

This kind of existential argument shares commonalities with the probabilistic method. Often, it is very powerful and we can find no proof of the dichotomy without using such an argument. However, this type of proof has some drawbacks. An existential argument that proves a dichotomy only states that there is a classification of the problems. However, the criterion of this classification is not clear and may not even be decidable for a given set of functions. When all $\mathcal{F}$-gates satisfy the desired property, we have a polynomial time algorithm. The correctness of the algorithm is proved given the premise that all $\mathcal{F}$-gates satisfy the desired property. However, the algorithms itself may not be able to verify the fact that the promise is satisfied.

4 Many Dimensions and Dichotomy Theorems

A complete dichotomy for the whole Holant is most desirable but on the other hand very difficulty. The main reason is that there are many dimensions of the framework. In this section, we carefully discuss these dimensions and many know dichotomy theorems for subclasses of Holant problems.

4.1 Freely available functions

Freely available functions are a set of functions which are always assumed to be in the set $\mathcal{F}$. More precisely, let $\mathcal{A}$ be a freely available set. Then, we can define a sub-framework Holant$^{\mathcal{A}}$ of Holant as

$$\text{Holant}^{\mathcal{A}}(\mathcal{F}) = \text{Holant}(\mathcal{F} \cup \mathcal{A}).$$

This is a sub-framework since Holant$^{\mathcal{A}}$ can only express and classify the function set which contains $\mathcal{A}$. To make the sub-framework meaningful, we should have that Holant($\mathcal{A}$) is tractable. Otherwise, all problems in Holant$^{\mathcal{A}}$ are hard.

This dimension of the framework usually defines a sub-framework of the Holant. Most importantly, if all equality functions are assumed to be available, it is exactly the #CSP problem.

$$\#CSP(\mathcal{F}) = \text{Holant}(\mathcal{F} \cup Equalities).$$

The usual definition of weighted #CSP goes as follows. Let $[q]$ be a domain set. A *weighted* constraint language $\mathcal{F}$ over the domain $[q]$ is a finite set of functions $\{f_1, \ldots, f_h\}$ in which $f_i : [q]^{r_i} \to \mathbb{C}$ is an r_i-ary function over $[q]$ for some $r_i \geq 1$. Then $\#CSP(\mathcal{F})$ is the following problem.

1. Let $\mathbf{x} = (x_1, \ldots, x_n) \in [q]^n$ be a set of n variables over $[q]$. The input is then a collection I of m tuples $(f, i_1, \ldots, i_r)$ in which f is an r-ary function in $\mathcal{F}$ and $i_1, \ldots, i_r \in [n]$.

2. The output of the problem is the following sum:

$$Z(I) \stackrel{\text{def}}{=} \sum_{\mathbf{x} \in [q]^n} \prod_{(f, i_1, \ldots, i_r) \in I} f(x_{i_1}, \ldots, x_{i_r}).$$

We show that every #CSP problem can be simulated by a Holant problem. Represent an instance of a #CSP problem by a bipartite graph where the Left-Hand-Side (LHS) is labeled by variables and the Right-Hand-Side (RHS) is labeled by constraints (functions) . Now the signature grid Ω on this bipartite graph is as follows: Every variable node on LHS is attached an Equality function and every constraint node on RHS has the given constraint function. Then

Holant_Ω is exactly the answer to the #CSP problem. In effect, the EQUALITY function on each variable node forces the incident edges to take the same value; this effectively reduces edge assignments to vertex assignments assigning values to each variable on LHS as in #CSP. It follows that #CSP problems are precisely the special case of Holant problems on bipartite graphs where every node on LHS is attached an EQUALITY function. It is easy to show that the class of #CSP problems is equivalent to Holant problems where all EQUALITY functions (of arbitrary arities) are always assumed to be freely available, and implicitly so.

To specify other freely available function sets, we can have other sub-frameworks of Holant problems. Well-studied ones include

Definition 4.1. *let $\mathcal{U}$ denote the set of all unary signatures. Given a set of signatures $\mathcal{F}$, we use* $\text{Holant}^*(\mathcal{F})$ *to denote* $\text{Holant}(\mathcal{F} \cup \mathcal{U})$.

Definition 4.2. *Given a set of signatures $\mathcal{F}$, we use* $\text{Holant}^c(\mathcal{F})$ *to denote* $\text{Holant}(\mathcal{F} \cup \{\Delta_1, \Delta_2, \cdots, \Delta_q\})$.

With few or no freely available constraint functions, the framework is more expressive and contains more interesting tractable cases, while on the other hand, it is more challenging to prove a dichotomy theorem for. To see an example, let NTW_3 be the NOT_TWO function of arity 3 over Boolean domain defined as $NTW_3(x_1x_2x_3) = 0$ if $x_1x_2x_3$ has exact two ones; otherwise the function value is 1. Then #CSP(NTW_3) is #P-complete. However without all equality functions, Holant(NTW_3) is in P. To see the algorithm for Holant(NTW_3), please ref to [CLX08].

By definition, it is clear that Holant^c is a super framework of Holant^* since it has fewer freely available functions. In the general setting, Holant^c can be also viewed as a super framework of CSP. This is due to the following pinning lemma from [DGJ07].

Lemma 4.3.

$$\#CSP(\mathcal{F} \cup \{\Delta_1, \Delta_2, \cdots, \Delta_q\}) \leq_T \#CSP(\mathcal{F}).$$

One can also study other sub-frameworks by allowing other freely available function sets. For example, one can define $\#\text{CSP}^*$ by allowing both equality and unary functions to be available. A dichotomy for $\#\text{CSP}^*$ would be very interesting.

4.2 Single function or a set of functions

In the definition of Holant($\mathcal{F}$), $\mathcal{F}$ is a set of functions. Sometimes, the setting that $\mathcal{F}$ only contains one single function (except maybe the freely available functions) is already very interesting. The most well studied setting for a single function is in the #CSP sub-framework containing a single binary function (except all equality functions). These are known as the counting Graph Homomorphisms problems.

A binary function can be written as a $q \times q$ matrix H. When all the matrix entries are $\{0, 1\}$, it can be viewed as a graph. An instance of the problem can also be viewed as a graph: The variables are vertices and the constraints are edges. The final value of the partition function can be viewed as the number of graph homomorphisms from the instance graph to the constraint graph. This is the reason why it is called graph homomorphisms.

In the #CSP framework, a single high arity function has also been studied by Dyer, Goldberg and Jerrum, called Hypergraph Homomorphisms [DGJ10].

For a Holant framework without equality functions, if the set of functions only contains binary functions, then the problem is always tractable. The simplest non-trivial setting starts with a ternary function. A dichotomy for a single symmetric ternary function was first proved in the Holant* framework [CLX09b], then in a pure Holant framework [CLX09a, KC10, CHL10]. These dichotomies for a single function serve as the starting points to prove general dichotomies for a set of functions. The Holant problem with one single high arity function was also studied by Cai and Kowalczyk [CK10].

4.3 Domain Size

The case $q = 1$ is meaningless since the final summation only contains one term. The case $q = 2$, which is usually called the Boolean domain, is already very non-trivial, interesting and important. Many combinatorial problems are in the Boolean domain. For the SAT problem, the Boolean domain is “True” or ”False”; for many graph problems like matching, vertex cover and independent set , the Boolean domain corresponds to an edge or a vertex is “chosen” or “not chosen”.

The complexity of the Holant problem on the Boolean domain is relatively well understood. For many special subfamilies, like Boolean #CSP, Boolean Holant* and symmetric Boolean Holant^c, a dichotomy was known. However, a complete dichotomy for Boolean Holant remains open.

When we move from the Boolean domain to an arbitrary finite domain, the difficulty of the problem increases significantly. Even extending known results of the Boolean domain to a domain of size 3 is already highly non-trivial [Bul06]. For example, proving a dichotomy for Holant* over a domain of size 3 is an interesting open question.

To attack the problem of a large domain, certain domain reduction techniques were invented. The simplest one is to decompose the domain into connected components then we need only focus on each component separately. A very clever twin reduction was proposed by Goldberg, Grohe, Jerrum ,Thurley [GGJT10], which can combine domain items that behave similarly into one and reduce the size of the domain. The idea was further extended to Cyclotomic Reduction in complex field by Cai, Chen and the current author. In a recent paper to prove the dichotomy for directed graph homomorphisms by Cai and Chen, a recursive domain reduction is the key of their proof [CC10].

To sum up the previous three subsections, we can draw a picture (Figure 1) to show the relation of several well studied sub-frameworks of Holant.

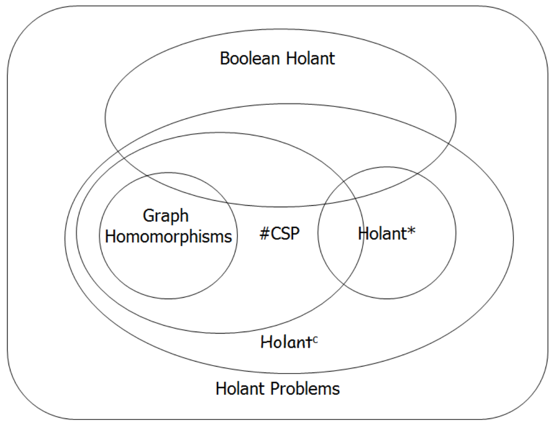

Figure 1: Relations of sub-frameworks of Holant Problems.

4.4 Value Range

The value range of the functions is another important dimension. If all functions take values from $\{0,1\}$, they can be viewed as relations or constraints. This is usually called the unweighted setting. Beyond the unweighted setting, we can consider non-negative weighted, real weighted and finally complex weighted settings. Beyond the unweighted setting, we can consider non-negative weighted, real weighted and finally complex weighted settings. From unweighted to complex weighted, the framework becomes more and more general while the proof of a dichotomy becomes more and more difficult. For unweighted setting, one can use some logical or combinatorial tools which may not be available to the weighted settings. Starting from the real weighted case, the appearance of possible cancelation offers more polynomial algorithms.

Another possibility is consider the problem over finite fields, and using modular arithmetic, which is interesting for complexity class $\oplus$P,Mod_k P, or $\#_k$ P.

4.4.1 Graph Homomorphisms

The first dichotomy theorem for counting unweighted undirected graph homomorphisms was proved by Dyer, Greenhill [DG00].

Theorem 4.4. *The counting graph homomorphisms problem of an undirected graph H is in P if every connected component of H is complete or complete bipartite. Otherwise, the problem is #P-complete.*

This was extended to the non-negatively weighted setting by Bulatov and Grohe [BG05]. Basically, the completeness in the unweighted setting is translated to rank one condition for the non-negatively weighted setting. For a weighted matrix, we define its unweighed version (support) by viewing all its non-zero entries as 1s. A connected component for a matrix is defined as the connected component of its support.

Theorem 4.5. *Let H be a symmetric non-negative matrix. The graph homomorphisms problem of H is in P if every connected component of H is of rank one or bipartite rank 2. Otherwise, the problem is #P-hard.*

This criterion is no longer true for real or complex weights. For example, the Hadamard matrix $\mathbf{H} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$ presents an obstacle. In general, unlike the non-negative weighted setting, when there are both positive and negative entries, there can be substantial cancelations in the exponential summation, which may yield surprisingly efficient computations. This is not dissimilar to monotone versus non-monotone complexity. Indeed, the Hadamard matrix turns out to be one such case. This is the starting point of the next great chapter by Goldberg, Grohe, Jerrum and Thurley. In a paper comprising 73 pages of beautiful proofs of both exceptional depth and conceptual vision, Goldberg, Jerrum, Grohe, and Thurley [GGJT10] proved a complexity dichotomy theorem for all real valued symmetric matrices. Their result is too intricate to give a short and accurate summary here, but essentially it states that the problem of computing graph homomorphisms for any real symmetric matrix is either in P or is #P-hard.

In the new tractable problems of real weighted graph homomorphisms, "-1" plays an important role. The reason is that "-1" is a non-trivial root of unity. When moving further to complex field, there are infinite many other roots of unity and indeed all of them bring new tractable problems. For example, the following complex matrix is tractable

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & \omega & \omega^2 \\ 1 & \omega^2 & \omega \end{pmatrix}, \qquad \text{where } \omega \text{ is a primative third root of unity.}$$

Finally, the dichotomy for all complex weighted symmetric matrices was obtained by Cai, Chen and the current author [CCL10].

Theorem 4.6. *Let H be a symmetric complex matrix. Then counting graph Homomorphisms to H either can be computed in polynomial time or is #P-hard.*

4.4.2 Boolean #CSP

An unweighted function can be viewed as a relation. A relation $R \subseteq \{0,1\}^k$ being affine means it is the affine linear subspace composed of the solutions of a system of affine linear equations over $\mathbb{F}_2$. Clearly, if all the relations are affine then the Boolean #CSP can be solved in polynomial time, since the problem is exactly asking the number of solutions for a linear system in the field $\mathbb{F}_2$. Creignou and Hermann proved that these are the only tractable problems for unweighted Boolean #CSP [CH96].

For non-negatively weighted Boolean #CSP, one obvious generalization of affine relations to functions is a globe scale of an unweighted affine relation. If the scale factor is c, then the function takes value c if the input is in the affine space and 0 otherwise. These scaled affine functions are called pure affine in [DGJ07]. In their paper, Dyer, Goldberg, Jerrum also introduced another family as follows, which is called product type.

$\mathscr{P}$ denotes the class of functions which can be expressed as a product of unary functions, binary equality functions and binary disequality functions.

Then they prove that

Theorem 4.7. *Let $\mathcal{F}$ be a set of non-negative functions over Boolean domain. Then #CSP($\mathcal{F}$) is #P-hard unless all the functions in $\mathcal{F}$ are pure affine or all the functions are of the product type $\mathscr{P}$, in which case the problem is in P.*

When looking at the complex field, there are more interesting tractable problems. Let X denote the $k+1$ dimensional column vector $(x_1, x_2, \ldots, x_k, 1)$ over the Boolean field $\mathbb{F}_2$. Suppose A is a Boolean matrix. χ_{AX} denotes the affine relation on inputs $x_1, x_2, \ldots, x_k$, whose value is 1 if AX is the zero vector, 0 if AX is not the zero vector.

$\mathscr{A}$ denotes all functions which have the form $\chi_{AX} i^{L_1(X)+L_2(X)+\cdots+L_n(X)}$, where $i = \sqrt{-1}$, L_j is a 0-1 indicator function $\chi_{\langle \alpha_j, X\rangle}$, where α_j is a $k+1$ dimensional vector, the inner product $\langle \cdot, \cdot \rangle$ is over $\mathbb{Z}_2$. The additions among $L_j X$ are just the usual addition in $\mathbb{Z}$. (Since we ignore global constant, all functions that are constant multiples of these functions are also in this class.)

Cai, Xia and the current author proved that

Theorem 4.8. *([CLX09b]) Suppose $\mathscr{F}$ is a class of functions mapping Boolean inputs to complex numbers. If $\mathscr{F} \subseteq \mathscr{A}$ or $\mathscr{F} \subseteq \mathscr{P}$, then #CSP($\mathscr{F}$) is computable in polynomial time. Otherwise, #CSP($\mathscr{F}$) is #*P*-hard.*

Independently, a dichotomy for a real weighted Boolean #CSP was given by Bulatov, Dyer, Goldberg, Jalsenius and Richerby [BDG$^+$09].

4.4.3 #CSP

Moving from Boolean #CSP to #CSP over arbitrary finite domains, the problem becomes much more difficult. Even for the unweighted setting, the problem was open for a long time. In a recent breakthrough result, Bulatov [Bul08] proved a sweeping dichotomy theorem. He gave a criterion, *congruence singularity*, and showed that for any finite set of constraint predicates Γ over any finite domain D, if Γ satisfies this condition, then #CSP(Γ) is solvable in P; otherwise it is #P-complete. His proof uses deep structural theorems from universal algebra [BS81]. Indeed this approach using universal algebra has been one of the most exciting developments in the study of the complexity of CSP in recent years, and has been called the *Algebraic Approach*. A more careful discussion of this Algebraic Approach is beyond the scope of this survey.

In [DR10b] Dyer and Richerby showed an alternative proof of the dichotomy theorem for unweighted #CSP. Their proof is considerably more direct, and uses no universal algebra other than the notion of a Mal'tsev polymorphism. A key idea is a data structure called a *frame*, which is a succinct representation of a *strongly rectangular* relation (and is similar to the "compact representation" of Bulatov and Dalmau [BD06]). Dyer and Richerby gave a combinatorial criterion, *strong balance*, and showed that this criterion is equivalent to Bulatov's *congruence singularity*, and determines the tractability of #CSP(Γ). They also showed that this criterion is a decidable criterion [DR10a, DR11]. Furthermore, by treating rational weights as integral multiples of a common denominator, the dichotomy theorem can be extended to include positive rational weights [BDG$^+$10].

Recently, the result was improved by Cai, Chen and the current author to non-negatively weighted #CSP($\mathcal{F}$) [CCL11]. Surprisingly, our tractability criterion is simpler than the previous criteria for the more restricted classes of problems, although when specialized to those cases, they are logically equivalent. Here is our dichotomy:

Theorem 4.9. *([CCL11]) Let $\mathcal{F}$ be a set of non-negative functions over domain $[q]$. The problem is polynomial computable if the following condition satisfied: For any input instance I over n variables of #CSP($\mathcal{F}$) and for any integers $a, b : 1 \leq a < b \leq n$, the following $q^a \times q^{b-a}$ matrix $\mathbf{M}$ is block-rank-1: The rows of $\mathbf{M}$ are indexed by $\mathbf{u} \in D^a$ and the columns are indexed by $\mathbf{v} \in D^{b-a}$, and*

$$M(\mathbf{u}, \mathbf{v}) = \sum_{\mathbf{w} \in D^{n-b}} F(\mathbf{u}, \mathbf{v}, \mathbf{w}), \qquad \text{for all } \mathbf{u} \in D^a \text{ and } \mathbf{v} \in D^{b-a}.$$

For the special case when $b = n$, we have that $M(\mathbf{u}, \mathbf{v}) = F(\mathbf{u}, \mathbf{v})$ is block-rank-1. Otherwise, the problem is #P-hard.

This is an example of the existential argument discussed in Section 3.4. The desired property here is block-rank-1. To extend these work to real weighted or complex weighted is a major open question here.

4.4.4 Boolean Holant

For Boolean symmetric Holant*, a dichotomy was proved for the complex weighted case directly by Cai, Xia and the current author [CLX09b].

Theorem 4.10. *Let $\mathcal{F}$ be a set of non-degenerate symmetric signatures over $\mathbb{C}$. Then Boolean* Holant$^*(\mathcal{F})$ *is computable in polynomial time in the following three* Classes. *In all other cases,* Holant$^*(\mathcal{F})$ *is #P-hard.*

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

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

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

For Boolean symmetric Holant^c, a dichotomy was first proved for the real weighted case in the same paper [CLX09b]. The main idea is to interpolate all the unary functions and reduce the Holant* problem to it.

Theorem 4.11. *Let $\mathcal{F}$ be a set of* real *symmetric signatures, and let $\mathcal{F}_1, \mathcal{F}_2$ and $\mathcal{F}_3$ be three families of signatures defined as*

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

Then Holant$^c(\mathcal{F})$ *is computable in polynomial time if (1)* Holant$^*(\mathcal{F})$ *is computable in polynomial time or (2) $\mathcal{F} \subseteq \mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3$. Otherwise,* Holant$^c(\mathcal{F})$ *is #P-hard.*

The above three families $\mathcal{F}_1, \mathcal{F}_2$ and $\mathcal{F}_3$ are exactly the functions in $\mathscr{A}$ (defined in Section 4.4.2) after being restricted to symmetric ones. This dichotomy was obtained before the one for Boolean #CSP and it suggested and helped the prove for the dichotomy for Boolean #CSP. The main difficulty to further extend it to the complex weighted setting is that some desirable properties of real numbers to make the interpolation work do not hold for the complex field. Finally, we solved this problem but used a very different approach. We in turn make use of the dichotomy of Boolean #CSP and reduce the Boolean #CSP to Holant^c problems. In [CHL10], Cai, Huang and the current author proved that

Theorem 4.12. *Let $\mathscr{F}$ be a set of complex symmetric signatures.* Holant$^c(\mathscr{F})$ *is #P-hard unless $\mathscr{F}$ satisfies one of the following conditions, in which case it is tractable:*

1. Holant$^*(\mathscr{F})$ *is tractable (for which we have an effective dichotomy in [CLX09b]);*
2. *There exists a $T \in \mathscr{T}$ such that $\mathscr{F} \subseteq T\mathscr{A}$, where*

$$\mathscr{T} \triangleq \{T \mid [1,0,1]T^{\otimes 2}, [1,0]T, [0,1]T \in \mathscr{A}\}$$

In Figure 1, we can see that both Holant* and #CSP can be viewed as a sub-framework of Holant^c. In the above dichotomy, we can see that essentially the tractable problems for Holant^c are the union of these for Holant* and #CSP. Also, the dichotomy for Holant* and #CSP plays an important role in the proof of dichotomy for Holant^c.

4.4.5 Counting Mod k

In this section, we discuss the counting problem modulo some integer k. This is the Holant problem where the functions take values in a finite field or ring.

There does exist several problems for which counting the number of solutions is #P-complete whereas computing it modulo some integer k is polynomial time computable. One prime example is computing the permanent of a 0/1 matrix, which is #P-complete [Val79a]. The parity version of this problem corresponds to computing the permanent modulo 2, which is the same as the determinant modulo 2, and is therefore computable in polynomial time via linear algebra computations. Some more such tractable parity problems were recently given by Valiant [Val10]. Furthermore, the characteristic of the finite field may affect the tractability. For example, Valiant showed that $\#_7$Pl-Rtw-Mon-3CNF (counting the number of satisfying assignments of a planar read-twice monotone 3CNF formula, modulo 7) is solvable in P by a holographic algorithm [Val06], while the parity or general version of the same problem is $\oplus$P-hard or #P-hard, respectively.

For the main reduction techniques summarized in Section 3, as pointed out by Valiant [Val10], for finite fields, holographic transformations and interpolation both appear to offer less flexibility than they do for general counting problems.

These two facts (some useful techniques cannot be adopted in finite fields and there exist some more complicated tractable cases) make it more challenging to obtain a dichotomy for $\#_k$Holant problems.

In [Fab08], Faben obtained a dichotomy theorem for unweighted Boolean $\#_k$CSP.

Theorem 4.13. [Fab08] *Given an unweighted function set $\mathcal{F}$, and an integer k, $\#_{\mathrm{k}}\mathrm{CSP}(\mathcal{F})$ is computable in polynomial time if all the relations in $\mathcal{F}$ are affine, or if $k = 2$ and all functions in $\mathcal{F}$ are closed under complement. Otherwise it is $\#_k$P-hard .*

Essentially, there is no additional tractable case in his dichotomy theorem (except one obvious case). However, when we allow functions to take weights in the ring $\mathbb{Z}_k$, some new non-trivial tractable cases do emerge, which are similar to weighted vs. unweighted #CSP without a modulus. As noted before, roots of unity play essential roles [GGJT10, CLX09b, BDG$^+$09, CCL10]. In finite fields, interesting cancelations do appear and every nonzero element is a root of unity. For general k, which may not be a prime, another subtlety is that the computation is performed in a ring $\mathbb{Z}_k$ rather than a field, where some desirable property of a field no longer holds.

In [GHLX11], Guo, Huang, Xia and the current author proved a dichotomy for Weighted Boolean #CSP Mod k, for any integer k. Our result starts from the finite field case, where the modulus k is an odd prime. In this case, the final result is algebraically the same as the dichotomy for complex weighted #CSP. The imaginary unit $i = \sqrt{-1}$ plays an important role in the dichotomy for the complex weighted #CSP [CLX09b]. Here by "algebraically", we mean that we view i as a fourth primitive root of unity which is also well defined in a finite field (or its extension). Then the dichotomy for $\#_k$CSP is identical to that for complex weighted #CSP. Some of the proof techniques are fairly similar to those in the proof for the complex weighted case [CLX09b], while others are completely different. For example, polynomial interpolation is one of the most important techniques in [CLX09b], but it is not available for the finite field. Such kinds of similarity and difference between fields with zero and finite characteristic p is one of the main themes of algebraical geometry [Har77]. It is interesting to observe similar phenomena in complexity theory.

For general k, let $k = p_1^{r_1} p_2^{r_2} \cdots p_m^{r_m}$, where p_i's are distinct primes, be the prime factorization of k. By the Chinese Remainder Theorem, solving the problem of $\#_k\mathrm{CSP}(\mathscr{F})$ is equivalent to solving all $\#_{p_i^{r_i}}\mathrm{CSP}(\mathscr{F})$. For $\#_{p^r}$CSP and p being an odd prime, we prove a surprising result that states that $\#_{p^r}\mathrm{CSP}(\mathscr{F})$ is tractable iff $\#_p\mathrm{CSP}(\mathscr{F})$ is, assuming #P is not equal to P. One direction is trivial, namely if $\#_{p^r}\mathrm{CSP}(\mathscr{F})$ can be solved in polynomial time, so can $\#_p\mathrm{CSP}(\mathscr{F})$. The reduction in the other direction is not of the black box style. We need the dichotomy for $\#_p\mathrm{CSP}(\mathscr{F})$ to state all the tractable cases, assuming #P is not equal to P, and we also need to explicitly use algorithms to solve such tractable cases. The algorithm for $\#_{p^r}\mathrm{CSP}(\mathscr{F})$ has a time complexity which is n^r times larger than that of the algorithm for $\#_p\mathrm{CSP}(\mathscr{F})$. We use a different treatment to solve the case that $p = 2$.

To sum up, we have

Theorem 4.14. *Let $k > 1$ and $\mathscr{F}$ be a set of functions. Then $\#_k CSP(\mathscr{F})$ is either in P or $\#_p$P-hard for some $p|k$.*

The $p = 2$ is of particular interest since $\oplus$P is an important complexity class. Guo, Valiant and the current author proved a complete dichotomy for symmetric Boolean Holant problems. We note that this is for the whole Holant framework without assuming any freely available functions. Such a dichotomy is open without mod.

Theorem 4.15. *([GLV11]) Let $\mathcal{F}$ be a set of symmetric signatures. Then the parity problem $\oplus\mathrm{Holant}(\mathcal{F})$ is either computable in polynomial time or $\oplus$P-complete.*

4.5 Symmetric or not

In this section, we discuss another property of the functions: Are the functions symmetric of not? We recall that a function is symmetric iff to apply a permutation of its input will not change the value of the function. Symmetric functions are important since they have clear combinatorial meanings. Many combinatorial counting problems can be formalized as Holant problems only with symmetric functions such as graph coloring, (perfect) matching, independent set and so on. In terms of proving dichotomy theorems, in many cases, to restrict the functions to be symmetric makes the problem significantly easier. On the other hand, to prove a dichotomy for general functions is more desirable and sometimes provides a more complete picture of the underlying structure.

4.5.1 Graph Homomorphisms

In section 4.4.1, all the dichotomies discussed are for an undirected graph which corresponds to symmetric functions. To extend these results to directed graph even for unweighted case, was an open question for a long time. To see one superficial reason why the problem become much difficult is that the definition of "connected component" is not as well defined as in undirected setting while to decompose into connected components is the first step for the dichotomy for undirected graph Homomorphisms

In [DGP07], Dyer, Goldberg, Paterson proved a dichotomy for a family of directed graphs which are acyclic. The problem was recently solved by Cai and Chen for all directed graphs even with non-negative weight [CC10]. They introduced some nice new domain reduction techniques. Since Graph Homomorphisms is a sub-framework of CSP, we note that these results were subsumed by recent dichotomy theorems of weighted or unweighted #CSP. To extend the weight to real or complex numbers remains open even for Graph Homomorphisms.

4.5.2 Boolean Holant* Problems

For graph homomorphisms, both the statement of dichotomy criterion and proof techniques for directed graphs are quite different from those for undirected graphs. Now we show another example of extending symmetric dichotomy to asymmetric dichotomy which suggests a clear relation between them. In Section 4.4.4, we stated a dichotomy for the symmetric Boolean Holant* problem from [CLX09b]. We now show an extension of this to general Boolean function set by Cai, Xia and the current author [CLX11].

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

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

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

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

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

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

Theorem 4.16. *Let $\mathcal{F}$ be any set of complex valued functions in Boolean variables. The problem Holant$^*(\mathcal{F})$ is polynomial time computable, if (1) $\mathcal{F} \subseteq \langle \mathcal{T} \rangle$, or (2) there exists an orthogonal matrix H such that $\mathcal{F} \subseteq \langle H \circ \mathcal{E} \rangle$, or (3) there exists a matrix $Z \in \{Z_1 = \begin{pmatrix} 1 & 1 \\ i & -i \end{pmatrix}, Z_2 = \begin{pmatrix} 1 & 1 \\ -i & i \end{pmatrix}\}$ such that $\mathcal{F} \subseteq \langle Z \circ \mathcal{E} \rangle$, or (4) there exists a matrix $Z \in \{Z_1, Z_2\}$ such that $\mathcal{F} \subseteq \langle Z \circ \mathcal{M} \rangle$. In all other cases, Holant$^*(\mathcal{F})$ is #P-hard.*

To prove an asymmetric dichotomy for Boolean Holant^c is an interesting open question. For its two special sub families Boolean CSP and Holant*, asymmetric dichotomies were known. A plausible conjecture could be that the union of these two tractable families is the tractable family for Holant^c.

4.6 Special Family of Graphs

In a general setting, most of the Holant problems are #P-hard. Sometimes, we would like to give further classification among all the hard problems. There are at least two different angles for this. One is to decrease the requirement of the algorithms. For example, we can satisfy with an approximation solution. Approximate counting is a fantastic subject with many good results. However, a discussion of these are out of the scope of this survey and we omit these here. Another angle is to restrict the input. For example, the counting perfect matching problem is #P-complete for general graphs but is polynomial time computable for planar graphs by the famous FKT method [Kas61, TF61, Kas61]. Recently, Valiant proposed a beautiful theory of holographic algorithms, which reduce other problems to the planar perfect matching problems. Using holographic algorithms, Valiant gave polynomial time algorithms for several interesting problems. After a sequence of papers developing a structural theory [CC06, CL10, CCL09, CL08, CL09], Cai and the current author give a more systemical characterization of what kind of problems can be solved by Holographical algorithms [CL11].

In [CLX10], Cai, Xia and the current author prove that holographic algorithms capture precisely those problems which are #Phard on general graphs but computable in polynomial time on planar graphs. More precisely, we prove three trichotomies of three different sub-frameworks of Holant.

Theorem 4.17. *Let $\mathcal{F}$ be a set of real symmetric signatures.* Planar Holant$^c(\mathcal{F})$ *is #P-hard unless $\mathcal{F}$ satisfies one of the following conditions, in which case it is tractable:*

1. Holant$^c(\mathcal{F})$ *(for general graphs) is tractable (for which we have an effective dichotomy [CLX09b]); or*

2. *Every signature in $\mathcal{F}$ is realizable by some matchgate (for which we have a complete characterization [CCL09]).*

Theorem 4.18. *Let $\mathcal{F}$ be a set of real symmetric functions.* Planar #CSP($\mathcal{F}$) *is #P-hard unless $\mathcal{F}$ satisfies one of the following conditions, in which case it is tractable:*

1. #CSP($\mathcal{F}$) *(for general instants) is tractable (for which we have an effective dichotomy [CLX09b]); 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 [CCL09]).*

Theorem 4.19. *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* Planar 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($[y_0, y_1, y_2] | [x_0, x_1, x_2, x_3]$) *(for general graph) is tractable (for which we have an effective dichotomy [CHL10]); 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 [CL11]).*

5 Conclusion

This survey summarizes various dichotomies obtained for counting problems from the unified framework of Holant problems. Every such dichotomy can be viewed as a dichotomy for certain sub-framework of Holant by restricting some of its dimensions. Such unified perspective does not only offer a language to summarize these results but also suggests interesting relations among different dichotomies. To extend any known dichotomy along any dimension of the framework is an interesting open question.

References

[Bax82] R.J. Baxter. *Exactly solved models in statistical mechanics.* Academic press London, 1982.

[BD03] Andrei A. Bulatov and Víctor Dalmau. Towards a dichotomy theorem for the counting constraint satisfaction problem. In *FOCS*, pages 562–571. IEEE Computer Society, 2003.

[BD06] A.A. Bulatov and V. Dalmau. A simple algorithm for Mal'tsev constraints. *SIAM Journal on Computing*, 36(1):16–27, 2006.

[BDG+09] A. Bulatov, M. Dyer, L.A. Goldberg, M. Jalsenius, and D. Richerby. The complexity of weighted boolean #CSP with mixed signs. *Theoretical Computer Science*, 410(38-40):3949–3961, 2009.

[BDG+10] A.A. Bulatov, M.E. Dyer, L.A. Goldberg, M. Jalsenius, M.R Jerrum, and D. Richerby. The complexity of weighted and unweighted #CSP. *arXiv:1005.2678*, 2010.

[BG05] Andrei A. Bulatov and Martin Grohe. The complexity of partition functions. *Theor. Comput. Sci.*, 348(2-3):148–186, 2005.

[BS81] S. Burris and H.P. Sankappanavar. *A course in universal algebra*, volume 78 of Graduate Texts in Mathematics. Springer-Verlag, New York-Berlin, 1981.

[Bul06] Andrei A. Bulatov. A dichotomy theorem for constraint satisfaction problems on a 3-element set. *J. ACM*, 53(1):66–120, 2006.

[Bul08] Andrei A. Bulatov. The complexity of the counting constraint satisfaction problem. In Luca Aceto, Ivan Damgård, Leslie Ann Goldberg, Magnús M. Halldórsson, Anna Ingólfsdóttir, and Igor Walukiewicz, editors, *ICALP (1)*, volume 5125 of *Lecture Notes in Computer Science*, pages 646–661. Springer, 2008.

[CC06] Jin-Yi Cai and Vinay Choudhary. Some results on matchgates and holographic algorithms. In Michele Bugliesi, Bart Preneel, Vladimiro Sassone, and Ingo Wegener, editors, *ICALP (1)*, volume 4051 of *Lecture Notes in Computer Science*, pages 703–714. Springer, 2006.

[CC07] Jin-Yi Cai and Vinay Choudhary. Valiant's holant theorem and matchgate tensors. *Theor. Comput. Sci.*, 384(1):22–32, 2007.

[CC10] Jin-Yi Cai and Xi Chen. A decidable dichotomy theorem on directed graph homomorphisms with non-negative weights. In *Proceedings of the 51st Annual IEEE Symposium on Foundations of Computer Science*, 2010.

[CCL09] Jin-Yi Cai, Vinay Choudhary, and Pinyan Lu. On the theory of matchgate computations. *Theory Comput. Syst.*, 45(1):108–132, 2009.

[CCL10] Jin-Yi Cai, Xi Chen, and Pinyan Lu. Graph homomorphisms with complex values: A dichotomy theorem. In Samson Abramsky, Cyril Gavoille, Claude Kirchner, Friedhelm Meyer auf der Heide, and Paul G. Spirakis, editors, *ICALP (1)*, volume 6198 of *Lecture Notes in Computer Science*, pages 275–286. Springer, 2010.

[CCL11] Jin-Yi Cai, Xi Chen, and Pinyan Lu. Non-negatively weighted #CSPs: An effective complexity dichotomy. *to appear in CCC 2011. Available at http://arxiv.org/abs/1012.5659*, 2011.

[CH96] Nadia Creignou and Miki Hermann. Complexity of generalized satisfiability counting problems. *Inf. Comput.*, 125(1):1–12, 1996.

[CHL10] Jin-Yi Cai, Sangxia Huang, and Pinyan Lu. From holant to #CSP and back: Dichotomy for holant problems. In Otfried Cheong, Kyung-Yong Chwa, and Kunsoo Park, editors, *ISAAC (1)*, volume 6506 of *Lecture Notes in Computer Science*, pages 253–265. Springer, 2010.

[CK10] Jin-Yi Cai and Michael Kowalczyk. A dichotomy for k-regular graphs with $\{0, 1\}$-vertex assignments and real edge functions. In Jan Kratochvíl, Angsheng Li, Jirí Fiala, and Petr Kolman, editors, *TAMC*, volume 6108 of *Lecture Notes in Computer Science*, pages 328–339. Springer, 2010.

[CKS01] N. Creignou, S. Khanna, and M. Sudan. *Complexity classifications of boolean constraint satisfaction problems*. SIAM Monographs on Discrete Mathematics and Applications, 2001.

[CL08] Jin-Yi Cai and Pinyan Lu. Basis collapse in holographic algorithms. *Computational Complexity*, 17(2):254–281, 2008.

[CL09] Jin-Yi Cai and Pinyan Lu. Holographic algorithms: The power of dimensionality resolved. *Theor. Comput. Sci.*, 410(18):1618–1628, 2009.

[CL10] Jin-Yi Cai and Pinyan Lu. On symmetric signatures in holographic algorithms. *Theory Comput. Syst.*, 46(3):398–415, 2010.

[CL11] Jin-Yi Cai and Pinyan Lu. Holographic algorithms: From art to science. *J. Comput. Syst. Sci.*, 77(1):41–61, 2011.

[CLX08] Jin-Yi Cai, Pinyan Lu, and Mingji Xia. Holographic algorithms by fibonacci gates and holographic reductions for hardness. In *FOCS '08: Proceedings of the 49th Annual IEEE Symposium on Foundations of Computer Science*, Washington, DC, USA, 2008. IEEE Computer Society.

[CLX09a] Jin-Yi Cai, Pinyan Lu, and Mingji Xia. A computational proof of complexity of some restricted counting problems. In Jianer Chen and S. Barry Cooper, editors, *TAMC*, volume 5532 of *Lecture Notes in Computer Science*, pages 138–149. Springer, 2009.

[CLX09b] Jin-Yi Cai, Pinyan Lu, and Mingji Xia. Holant problems and counting CSP. In Michael Mitzenmacher, editor, *STOC*, pages 715–724. ACM, 2009.

[CLX10] Jin-Yi Cai, Pinyan Lu, and Mingji Xia. Holographic algorithms with matchgates capture precisely tractable planar #CSP. In *FOCS*, pages 427–436. IEEE Computer Society, 2010.

[CLX11] Jin-Yi Cai, Pinyan Lu, and Mingji Xia. Dichotomy for holant* problems of boolean domain. In *SODA '11: Proceedings of the nineteenth annual ACM-SIAM symposium on Discrete algorithms*, 2011.

[DG00] M.E. Dyer and C. Greenhill. The complexity of counting graph homomorphisms. In *Proceedings of the 9th International Conference on Random Structures and Algorithms*, pages 260–289, 2000.

[DGJ07] Martin E. Dyer, Leslie Ann Goldberg, and Mark Jerrum. The complexity of weighted boolean #CSP. *CoRR*, abs/0704.3683, 2007.

[DGJ10] Martin E. Dyer, Leslie Ann Goldberg, and Mark Jerrum. A complexity dichotomy for hypergraph partition functions. *Computational Complexity*, 19(4):605–633, 2010.

[DGP07] Martin E. Dyer, Leslie Ann Goldberg, and Mike Paterson. On counting homomorphisms to directed acyclic graphs. *J. ACM*, 54(6), 2007.

[DP91] C. T. J. Dodson and T. Poston. *Tensor Geometry.* Graduate Texts in Mathematics 130. Springer-Verlag, New York, 1991.

[DR10a] M.E. Dyer and D.M. Richerby. An effective dichotomy for the counting constraint satisfaction problem. *arXiv:1003.3879*, 2010.

[DR10b] M.E. Dyer and D.M. Richerby. On the complexity of #CSP. In *Proceedings of the 42nd ACM symposium on Theory of computing*, pages 725–734, 2010.

[DR11] M.E. Dyer and D.M. Richerby. The #CSP dichotomy is decidable. In *Proceedings of the 28th Symposium on Theoretical Aspects of Computer Science*, 2011.

[Fab08] John Faben. The complexity of counting solutions to generalised satisfiability problems modulo k. *CoRR*, abs/0809.1836, 2008.

[FLS07] M. Freedman, L. Lovász, and A. Schrijver. Reflection positivity, rank connectivity, and homomorphism of graphs. *J. AMS*, 20:37–51, 2007.

[GGJT10] Leslie Ann Goldberg, Martin Grohe, Mark Jerrum, and Marc Thurley. A complexity dichotomy for partition functions with mixed signs. *SIAM J. Comput.*, 39(7):3336–3402, 2010.

[GHLX11] Heng Guo, Sangxia Huang, Pinyan Lu, and Mingji Xia. The complexity of weighted boolean #CSP modulo k. In Jean-Yves Marion and Thomas Schwentick, editors, *STACS*, volume 5 of *LIPIcs*, pages 249–260. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2011.

[GLV11] Heng Guo, Pinyan Lu, and Leslie Valiant. A dichotomy theorem for parity holant problems. *to appear in ICALP*, 2011.

[Har77] R. Hartshorne. *Algebraic geometry.* springer Verlag, 1977.

[Isi25] E. Ising. Beitrag zur theorie des ferromagnetismus. *Zeitschrift für Physik A Hadrons and Nuclei*, 31(1):253–258, 1925.

[Kas61] P. W. Kasteleyn. The statistics of dimers on a lattice. *Physica*, 27:1209–1225, 1961.

[Kas67] P. W. Kasteleyn. Graph theory and crystal physics. In (F. Harary, editor, *Graph Theory and Theoretical Physics*, pages 43–110. Academic Press, London, 1967.

[KC10] Michael Kowalczyk and Jin-Yi Cai. Holant problems for regular graphs with complex edge functions. *In the proceeding of STACS*, 2010.

[Lad75] Richard E. Ladner. On the structure of polynomial time reducibility. *J. ACM*, 22(1):155–171, 1975.

[LS81] 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.

[LY52] 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.

[MW73] B.M. McCoy and T.T. Wu. *The two-dimensional Ising model.* Harvard University Press Cambridge, 1973.

[Ons44] L. Onsager. Crystal statistics. I. A two-dimensional model with an order-disorder transition. *Physical Review*, 65(3-4):117–149, 1944.

[TF61] H. N. V. Temperley and M. E. Fisher. Dimer problem in statistical mechanics c an exact result. *Philosophical Magazine*, 6:1061C 1063, 1961.

[Vad01] Salil P. Vadhan. The complexity of counting in sparse, regular, and planar graphs. *SIAM J. Comput.*, 31(2):398–427, 2001.

[Val79a] Leslie G. Valiant. The complexity of computing the permanent. *Theor. Comput. Sci.*, 8:189–201, 1979.

[Val79b] Leslie G. Valiant. The complexity of enumeration and reliability problems. *SIAM J. Comput.*, 8(3):410–421, 1979.

[Val02] Leslie G. Valiant. Quantum circuits that can be simulated classically in polynomial time. *SIAM J. Comput.*, 31(4):1229–1254, 2002.

[Val06] Leslie G. Valiant. Accidental algorthims. In *FOCS '06: Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science*, pages 509–517, Washington, DC, USA, 2006. IEEE Computer Society.

[Val08] Leslie G. Valiant. Holographic algorithms. *SIAM J. Comput.*, 37(5):1565–1594, 2008.

[Val10] Leslie G. Valiant. Some observations on holographic algorithms. In Alejandro López-Ortiz, editor, *LATIN*, volume 6034 of *Lecture Notes in Computer Science*, pages 577–590. Springer, 2010.

[XZZ07] 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.

[Yan52] C.N. Yang. The spontaneous magnetization of a two-dimensional Ising model. *Physical Review*, 85(5):808–816, 1952.

[YL52] 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.

ECCC ISSN 1433-8092

http://eccc.hpi-web.de