

Fine-Grained Dichotomies for the Tutte Plane and Boolean #CSP*

Cornelius Brand¹, Holger Dell², and Marc Roth³

- 1 Saarland University, Saarbrücken, Germany; and Cluster of Excellence “Multimodal Computing and Interaction” (MMCI), Saarbrücken, Germany
- 2 Saarland University, Saarbrücken, Germany; and Cluster of Excellence “Multimodal Computing and Interaction” (MMCI), Saarbrücken, Germany
- 3 Saarland University, Saarbrücken, Germany; and Cluster of Excellence “Multimodal Computing and Interaction” (MMCI), Saarbrücken, Germany

Abstract

Jaeger, Vertigan, and Welsh [15] proved a dichotomy for the complexity of evaluating the Tutte polynomial at fixed points: The evaluation is #P-hard almost everywhere, and the remaining points admit polynomial-time algorithms. Dell, Husfeldt, and Wahlén [9] and Husfeldt and Taslaman [12], in combination with the results of Curticapean [7], extended the #P-hardness results to tight lower bounds under the counting exponential time hypothesis #ETH, with the exception of the line $y = 1$, which was left open. We complete the dichotomy theorem for the Tutte polynomial under #ETH by proving that the number of all acyclic subgraphs of a given n -vertex graph cannot be determined in time $\exp(o(n))$ unless #ETH fails.

Another dichotomy theorem we strengthen is the one of Creignou and Hermann [6] for counting the number of satisfying assignments to a constraint satisfaction problem instance over the Boolean domain. We prove that all #P-hard cases cannot be solved in time $\exp(o(n))$ unless #ETH fails. The main ingredient is to prove that the number of independent sets in bipartite graphs with n vertices cannot be computed in time $\exp(o(n))$ unless #ETH fails.

In order to prove our results, we use the block interpolation idea by Curticapean [7] and transfer it to systems of linear equations that might not directly correspond to interpolation.

1998 ACM Subject Classification F.2.1 [Analysis of Algorithms and Problem Complexity] Numerical Algorithms and Problems, G.2.1 [Discrete Mathematics] Combinatorics

Keywords and phrases Computational complexity, counting problems, Tutte polynomial, exponential time hypothesis, forests, independent sets

Digital Object Identifier 10.4230/LIPIcs.IPEC.2016.9

1 Introduction

Counting combinatorial objects is at least as hard as detecting their existence, and often it is harder. Valiant [20] introduced the complexity class #P to study the complexity of counting problems and proved that counting the number of perfect matchings in a given bipartite graph is #P-complete. By a theorem of Toda [19], we know that $\text{PH} \subseteq \text{P}^{\#\text{P}}$ holds; in particular, for every problem in the entire polynomial-time hierarchy, there is a

* This work was done while the authors were visiting the Simons Institute for the Theory of Computing.



polynomial-time algorithm that is given access to an oracle for counting perfect matchings. This theorem suggests that counting is much harder than decision.

When faced with a problem that is NP-hard or #P-hard, the area of exact algorithms strives to find the fastest exponential-time algorithm for a problem, or find reasons why faster algorithms might not exist. For example, the fastest known algorithm for counting perfect matchings in n -vertex graphs [1] runs in time $2^{n/2} \cdot \text{poly}(n)$. It has been hypothesized that no $O(1.99^{n/2})$ -time algorithm for the problem exists, but we do not know whether such an algorithm has implications for the strong exponential time hypothesis, which states that for all $\varepsilon > 0$, there is some k such that the problem of deciding satisfiability of boolean formulas in k -CNF on n variables does not have an algorithm running in time $(2 - \varepsilon)^n$. However, we know by [8] that the term $O(n)$ in the exponent is asymptotically tight, in the sense that a $2^{o(n)}$ -time algorithm for counting perfect matchings would violate the (randomized) exponential time hypothesis (ETH) by Impagliazzo and Paturi [13]. Using the idea of block interpolation, Curticapean [7] strengthened the hardness by showing that a $2^{o(n)}$ -time algorithm for counting perfect matchings would violate the (deterministic) counting exponential time hypothesis (#ETH).

Our main results are hardness results under #ETH for 1) the problem of counting all forests in a graph, that is, its acyclic subgraphs, and 2) the problem of counting the number of independent sets in a bipartite graph. If #ETH holds, then neither of these problems has an algorithm running in time $\exp(o(n))$ even in simple n -vertex graphs of bounded maximum degree. We use these results to lift two known “FP vs. #P-hard” dichotomy theorems to their more refined and asymptotically tight “FP vs. #ETH-hard” variants. Here FP is the class of functions computable in polynomial time. Note that #ETH is weaker than ETH, so that our results could also be stated under ETH.

1.1 The Tutte polynomial under #ETH

The Tutte polynomial of a graph G with $G = (V, E)$ is the bivariate polynomial $T(G; x, y)$ defined via

$$T(G; x, y) = \sum_{A \subseteq E} (x - 1)^{k(A) - k(E)} (y - 1)^{k(A) + |A| - |V|}, \quad (1)$$

where $k(A)$ is the number of connected components of the graph (V, A) . The Tutte polynomial captures many combinatorial properties of a graph in a common framework, such as the number of spanning trees, forests, proper colorings, and certain flows and orientations, but also less obvious connections to other fields, such as link polynomials from knot theory, reliability polynomials from network theory, and (perhaps most importantly) the Ising and Potts models from statistical physics. We make no attempt to survey the literature or the different applications for the Tutte polynomial, and instead refer to the upcoming CRC handbook on the Tutte polynomial [10].

Since $T(G; -2, 0)$ corresponds to the number of proper 3-colorings of G , we cannot hope to compute all coefficients of $T(G; x, y)$ in polynomial time. Instead, the literature and this paper focus on the complexity of evaluating the Tutte polynomial at fixed evaluation points. That is, for each $(x, y) \in \mathbf{Q}^2$, we consider the function $T_{x,y}$ defined as $G \mapsto T(G; x, y)$. Jaeger, Vertigan, and Welsh [15] proved that this function is either #P-hard to compute or has a polynomial-time algorithm. In particular, if (x, y) satisfies $(x - 1)(y - 1) = 1$, then $T_{x,y}$ corresponds to the 1-state Potts model and has a polynomial-time algorithm, and if (x, y) is one of the four points $(1, 1)$, $(-1, -1)$, $(0, -1)$, or $(-1, 0)$, it also has a polynomial-time

algorithm; the most interesting point here is $T(G; 1, 1)$, which corresponds to the number of spanning trees in G and happens to admit a polynomial-time algorithm.

A trivial algorithm to compute the Tutte polynomial runs in time $2^{O(m)}$, where m is the number of edges. Björklund et al. [2] proved that there is an algorithm running in time $\exp(O(n))$, where n is the number of vertices. Dell et al. [8] proved for all hard points, except for points with $y = 1$, that an $\exp(o(n/\log^3 n))$ -time algorithm for $T_{x,y}$ on simple graphs would violate #ETH. Distressingly, this result not only left open one line, but also left a gap in the running time. Curticapean [7] introduced the technique of block interpolation to close the running time gap: Under #ETH, there does not exist an $\exp(o(n))$ -time algorithm for $T_{x,y}$ on simple graphs at any hard point (x, y) with $y \neq 1$.¹

Our contributions: We resolve the complexity of the missing line $y = 1$ under #ETH. On this line, the Tutte polynomial counts forests weighted in some way, and the main result is the following theorem.

► **Theorem 1** (Forest counting is hard under #ETH). *If #ETH holds, then there exist constants $\varepsilon, C > 0$ such that no $O(\exp(\varepsilon n))$ -time algorithm can compute the number of all forests in a given simple n -vertex graph with at most $C \cdot n$ edges.*

The fact that the problem remains hard even on simple sparse graphs makes the theorem stronger. The previously best known lower bound under #ETH was that forests cannot be counted in time $O(\exp(n^\delta))$ where $\delta > 0$ is some constant depending on the instance blow-up caused by the known #P-hardness reductions for forest counting; the thesis of Taslamani [18] shows a detailed proof for $\delta = \frac{1}{8}$. Our approach yields a #P-hardness proof for forest counting that is simpler than the proofs we found in the literature.²

Combined with all previous results [15, 8, 7], we can now formally state a complete #ETH dichotomy theorem for the Tutte polynomial over the reals.

► **Corollary 2** (Dichotomy for the real Tutte plane under #ETH). *Let $(x, y) \in \mathbf{Q}^2$. If (x, y) satisfies*

$$(x - 1)(y - 1) = 1 \text{ or } (x, y) \in \{(1, 1), (-1, -1), (0, -1), (-1, 0)\},$$

then $T_{x,y}$ can be computed in polynomial time. Otherwise $T_{x,y}$ is #P-hard and, if #ETH is true, then there exists $\varepsilon > 0$ such that $T_{x,y}$ cannot be computed in time $\exp(\varepsilon n)$, even for simple graphs.

The result also holds for sparse simple graphs. We stated the results only for rational numbers in order to avoid discussions about how real numbers should be represented.

For the proof of Theorem 1, we establish a reduction chain that starts with the problem of counting perfect matchings on sparse graphs, which is known to be hard under #ETH. As an intermediate step, we find it convenient to work with the multivariate forest polynomial as defined, for example, by Sokal [17]. After a simple transformation of the graph, we are able to extract the number of perfect matchings of the original graph from the multivariate forest polynomial of the transformed graph, even when only two different variables are used. Subsequently, we use Curticapean’s idea of block interpolation [7] to reduce the problem of

¹ The conference version of [7] does not handle the case $y = 0$, but the full version (to appear) does.

² For the #P-hardness of forest counting, [15] refers to private communication with Mark Jerrum as well as the PhD thesis of Vertigan [21]. A self-contained (but involved) proof appears in “Complexity of Graph Polynomials” by Steven D. Noble, chapter 13 of [11].

computing all coefficients of the bivariate forest polynomial to the problem of evaluating the univariate forest polynomial on multigraphs where all edge multiplicity are bounded by a constant. Finally, we replace parallel edges with parallel paths of constant length to reduce to the problem of evaluating the univariate forest polynomial on simple graphs.

1.2 #CSP over the Boolean domain under #ETH

In the second part of this paper, we consider constraint satisfaction problems (CSPs) over the Boolean domain, which are a natural generalization of the satisfiability problem for k -CNF formulas. A *constraint* is a relation $R \subseteq \{0, 1\}^k$ for some $k \in \mathbf{N}$, and a set Γ of constraints is a *constraint language*. $\text{CSP}(\Gamma)$ is the constraint satisfaction problem where all constraints occurring in the instances are of a type contained in Γ , and $\#\text{CSP}(\Gamma)$ is the corresponding counting version, which wants to compute the number of satisfying assignments. If all constraints happen to be *affine*, that is, they are linear equations over $\text{GF}(2)$, then the number of solutions can be determined in polynomial time by applying Gaussian elimination and determining the size of the solution space. Creignou and Hermann [6] prove that, as soon as you allow even just one constraint type that is not affine, the resulting CSP problem is #P-hard.

Our contributions: We prove that all Boolean #CSPs that are #P-hard are also hard under #ETH. The #P-hardness is established in [6] by reductions from counting independent sets in bipartite graphs, which we prove to be hard in the following theorem.

► **Theorem 3** (Counting bipartite independent sets is hard under #ETH). *If #ETH holds, then there exist constants $\varepsilon > 0$ and $D \in \mathbf{N}$ such that no $O(\exp(\varepsilon n))$ -time algorithm can compute the number of independent sets in bipartite n -vertex graphs of maximum degree at most D .*

The fact that the problem is hard even on graphs of bounded degree makes the theorem stronger. We remark that the number of independent sets in bipartite graphs has a prominent role in counting complexity. Currently, the complexity of *approximating* this number is unknown, and many problems in approximate counting turn out to be polynomial-time equivalent to approximately counting independent sets in bipartite graphs. Theorem 3 shows that this mysterious situation does not occur for the exact counting problem in the exponential-time setting: it is just as hard as counting satisfying assignments of 3-CNFs.

The #P-hardness of counting independent sets in bipartite graphs was established by Provan and Ball [16]. The main ingredient in their proof is a system of linear equations that does not seem to correspond to polynomial interpolation directly. We prove Theorem 3 by transferring the block interpolation idea from [7] to this system of linear equations, which we do using a Kronecker power of the original system.

Theorem 3 combined with existing reductions in [6] yields the fine-grained dichotomy.

► **Corollary 4** (Creignou and Hermann under #ETH). *Let Γ be a finite constraint language. If every constraint in Γ is affine, then $\#\text{CSP}(\Gamma)$ has a polynomial-time algorithm. Otherwise $\#\text{CSP}(\Gamma)$ is #P-complete, and if #ETH holds, it cannot be computed in time $\exp(o(n))$ where n is the number of variables.*

We consider Corollary 4 to be a first step towards understanding the fine-grained complexity of technically much more challenging dichotomies, such as the ones for counting CSPs with complex weights of Cai and Chen [3], or the dichotomy for Holant problems with symmetric signatures over the Boolean domain of Cai, Lu and Cia [4].

2 Preliminaries

Given a matrix A of size $m_1 \times n_1$ and a matrix B of size $m_2 \times n_2$ their *Kronecker product* $A \otimes B$ is a matrix of size $m_1 m_2 \times n_1 n_2$ given by

$$A \otimes B = \begin{bmatrix} a_{11}B & \dots & a_{1n}B \\ \vdots & \ddots & \vdots \\ a_{m1}B & \dots & a_{mn}B \end{bmatrix}.$$

Let $A^{\otimes n}$ be the matrix defined by $A^{\otimes 1} = A$ and $A^{\otimes n+1} = A \otimes A^{\otimes n}$. Furthermore, if A and B are quadratic matrices of size n_a and n_b , respectively, it is known that $\det(A \otimes B) = \det(A)^{n_b} \cdot \det(B)^{n_a}$ holds.

The exponential time hypothesis (ETH) by Impagliazzo and Paturi [13] is that satisfiability of 3-CNF formulas cannot be computed substantially faster than by trying all possible assignments. The counting version of this hypothesis [8], which is a weaker assumption (clearly, counting the number of solutions entails deciding existence of a solution), reads as follows:

(#ETH) *There is a constant $c > 0$ such that no deterministic algorithm can compute #3-SAT in time $\exp(c \cdot n)$, where n is the number of variables.*

A different way of formulating #ETH is to say no algorithm can compute #3-SAT in time $\exp(o(n))$. The latter statement is clearly implied by the formal statement, and it will be more convenient for discussion to use this form.

The sparsification lemma by Impagliazzo, Paturi, and Zane [14] is that every k -CNF formula φ can be written as the disjunction of $2^{\varepsilon n}$ formulas in k -CNF, each of which has at most $c(k, \varepsilon) \cdot n$ clauses. Moreover, this disjunction of sparse formulas can be computed from φ and ε in time $2^{\varepsilon n} \cdot \text{poly}(m)$. The density $c = c(k, \varepsilon)$ is the *sparsification constant*, and the best known bound is $c(k, \varepsilon) = (k/\varepsilon)^{3k}$ [5]. It was observed [8] that the disjunction can be made so that every assignment satisfies at most one of the sparse formulas in the disjunction, and so the sparsification lemma applies to #ETH as well. In particular, #ETH implies that #3-SAT cannot be computed in time $\exp(o(m))$, where m is the number of clauses.

We also make use of the following result, whose proof is based on block interpolation.

► **Theorem 5** (Curticapean [7]). *If #ETH holds, then there are constants $\varepsilon, D > 0$ such that neither of the following problems have $O(2^{\varepsilon n})$ -time algorithms n -vertex graphs G , even if G is simple and of maximum degree at most D :*

- *Computing the number of perfect matchings of G .*
- *Computing the number of independent sets of G .*

In the proof, we recover a univariate graph polynomial of high degree, and a classical approach to do so is polynomial interpolation. Interpolation works by evaluating a polynomial at different points and then solving some linear equations. In our case, however, we can evaluate the univariate polynomial only at certain discrete points, which involves constructing graphs with a number of edges that grows roughly as the square of the number of evaluation points. Since we need roughly as many points as the (large) degree of the polynomial, the conditional lower bounds we can achieve are not optimal.

While univariate interpolation appears to be insufficient for obtaining tight bounds under ETH, the univariate polynomials we consider have natural multilinear variants. The univariate polynomial can be reconstructed from the multilinear one, but since the multilinear polynomial has linearly many variables, we would need exponentially many evaluation

points to reconstruct all coefficients of the multilinear polynomial. This is not feasible in a subexponential-time reduction.

Block interpolation is able to find a sweet spot between univariate interpolation, where the constructed graphs would grow super-linearly, and multilinear interpolation, where the number of evaluation points would be exponential. For this, we arbitrarily partition the variables of the multilinear polynomial into (almost) equal-size parts (or *blocks*) and identify any variables in the same part. By choosing large but constant-sized parts, the individual degree of each variable is constant, but their number is sub-linear, which allows a subexponential-time reduction to work. At the same time, the graphs that we need to query are larger by at most by a constant factor. After the block interpolation, we can recover the univariate polynomial as desired.

3 Counting forests is #ETH-hard

Let $\mathcal{F}(G)$ be the set of all *forests* of G , that is, edge subsets $A \subseteq E(G)$ such that the graph $(V(G), A)$ is acyclic. For $y = 1$, only the terms with $k(A) + |A| - |V| = 0$ survive, and we get the following:

$$T(G; x, 1) = \sum_{A \in \mathcal{F}(G)} (x - 1)^{k(A) - k(E)}. \quad (2)$$

We want to prove that, for every fixed $x \neq 1$, computing the value $T(G; x, 1)$ for a given graph G is hard under #ETH. In particular, this is true for $T(G; 2, 1)$, which is the number of forests in G . The goal of this section is to prove the following theorem.

► **Theorem 6.** *Let $x \in \mathbf{R} \setminus \{1\}$. If #ETH holds, then there exist $\varepsilon, C > 0$ such that the function that maps simple n -vertex graphs G with at most $C \cdot n$ edges to the value $T(G; x, 1)$ cannot be computed in time $2^{\varepsilon n}$.*

Theorem 6 yields Theorem 1 as its special case with $x = 2$.

3.1 The multivariate forest polynomial

A *weighted graph* is a graph G in which every edge $e \in E(G)$ is endowed with a weight w_e , which is an element of some ring. We use the *multivariate forest polynomial*, defined e.g. by Sokal [17, (2.14)] as follows:

$$F(G; w) = \sum_{A \in \mathcal{F}(G)} \prod_{e \in A} w_e.$$

Projecting all weights w_e onto a single variable x yields the *univariate forest polynomial*:

$$F(G; x) = \sum_{A \in \mathcal{F}(G)} x^{|A|} = \sum_{k=0}^{|E(G)|} a_k(G) x^k,$$

where $a_k(G)$ is the number of forests with k edges in G . For all $x \in \mathbf{R} \setminus \{1\}$, the formal relation between $T(G; x, 1)$ and the univariate forest polynomial is given by the identity

$$T(G; x, 1) = (x - 1)^{|V| - k(E)} \sum_{A \in \mathcal{F}(G)} (x - 1)^{-|A|} = (x - 1)^{|V| - k(E)} \cdot F\left(G; \frac{1}{x - 1}\right). \quad (3)$$

The first equality follows from (2) and the fact that $k(A) + |A| - |V| = 0$ holds if and only if A is a forest.

In particular, evaluating the forest polynomial and evaluating the Tutte polynomial for $y = 1$ are polynomial-time equivalent.

For a forest $A \in \mathcal{F}(G)$, let $\mathcal{C}(A)$ be the family of all sets $U \subseteq V(G)$ such that $U \neq \emptyset$ and U is a maximal connected component in A ; clearly, each such U is the vertex set of a tree in the forest, where we also allow trees with $|U| = 1$.

► **Lemma 7 (Adding an apex).** *Let G be a weighted graph, and let G' be obtained from G by adding a new vertex a and joining it with each vertex $v \in V(G)$ using edges of weight z_v . Then*

$$F(G') = \sum_{A \in \mathcal{F}(G)} \prod_{e \in A} w_e \cdot \prod_{U \in \mathcal{C}(A)} \left(1 + \sum_{u \in U} z_u \right). \quad (4)$$

Moreover, when we set $z_v = -1$ for all $v \in V(G)$ and $w_e = w$ for all $e \in E(G)$, we have that the coefficient of $w^{n/2}$ in $F(G')$ is equal to the number of perfect matchings in G .

Proof. We first define a projection ϕ that maps a forest A' in the graph G' to the forest $A = \phi(A')$ in the original graph G . In particular, ϕ simply removes all edges added in the construction of G' , that is, we define $\phi(A') = E(G) \cap A'$ for all $A' \in \mathcal{F}(G')$. Now $\phi(A')$ is a forest in G .

Next we characterize the forests A' that map to the same A under ϕ . Let A be a fixed forest in G . Then a forest A' in G' maps to A under ϕ if and only if set $X := A' \setminus A$ satisfies the following property:

(P) For all trees $U \in \mathcal{C}(A)$, at most one edge of X is incident to a vertex of U .

The forward direction of this claim follows from the fact that A' is a forest, and so in addition to any tree $U \in \mathcal{C}(A)$ it can contain at most one edge connecting U to a ; otherwise the tree and the two edges to a would contain a cycle in A' . For the backward direction of the claim, observe that adding a set X with the property (P) to A cannot introduce a cycle.

Finally, we calculate the weight contribution of all A' that map to the same A . Let A' be a forest in G' , let $A = \phi(A')$ and $X = A' \setminus A$. The weight contribution of A' in the definition of $F(G')$ is $\prod_{e \in A'} w'_e$. For all $e \in A$, we have $w'_e = w_e$. For each $e \in X$, we let $v_e \in V(G)$ be the vertex with $e = \{a, v_e\}$, and we have $w'_e = z_{v_e}$. Thus the overall weight contribution of all A' with $\phi(A) = A'$ is

$$\sum_{\substack{A' \in \mathcal{F}(G') \\ \phi(A')=A}} \prod_{e \in A'} w'_e = \prod_{e \in A} w_e \cdot \sum_X \prod_{e \in X} z_{v_e} = \prod_{e \in A} w_e \cdot \prod_{U \in \mathcal{C}(A)} \left(1 + \sum_{u \in U} z_u \right). \quad (5)$$

The sum in the middle is over all X with the property (P), and the first equality follows from the bijection between forests A' and sets X with property (P). For the second equality, we use property (P) and the distributive law. We obtain (4) by taking the sum of equations (5) over all $A \in \mathcal{F}(G)$.

For the moreover part of the lemma, note that the stated settings of the edge weights for G' yields

$$F(G') = \sum_{A \in \mathcal{F}(G)} w^{|A|} \prod_{U \in \mathcal{C}(A)} (1 - |U|).$$

The coefficient of $w^{n/2}$ in $F(G')$ satisfies

$$[w^{n/2}]F(G') = \sum_{\substack{A \in \mathcal{F}(G) \\ |A|=n/2}} \prod_{U \in \mathcal{C}(A)} (1 - |U|). \quad (6)$$

Since $(V(G), A)$ is an acyclic graph with exactly $n/2$ edges, it is either a perfect matching or it contains an isolated vertex. If it contains an isolated vertex v , then we have $\{v\} \in \mathcal{C}(A)$ and thus the product in (6) is equal to zero. It follows that A does not contribute to the sum if it is not a perfect matching. On the other hand, if A is a perfect matching, then we have $|U| = 2$ for all $U \in \mathcal{C}(A)$, so the product in (6) is equal to 1 or -1 , depending on the parity of $n/2$. Overall, we obtain that $[w^{n/2}]F(G')$ is equal in absolute value to the number of perfect matchings of G . ◀

Lemma 7 shows that computing the multivariate forest polynomial is at least as hard as counting perfect matchings; moreover, this is true even if at most two different edge weights are used. Next we argue how to reduce from the multivariate forest polynomial with at most two distinct weights to the problem of evaluating the univariate polynomial in multigraphs. We do so via an oracle self-reduction, whose queries are sparse multigraphs in which each edge has at most a constant number of parallel edges.

► **Lemma 8** (From two weights to small weights using block interpolation). *Let x and y be two variables, and let $z \in \mathbf{R} \setminus \{0\}$ be fixed. There is an algorithm as follows:*

1. *Its input is a weighted graph (G, w) with $w_e \in \{x, y\}$ for all $e \in E(G)$, and a real $\varepsilon > 0$.*
2. *It outputs all coefficients of the bivariate polynomial $F(G; w)$.*
3. *It runs in time $2^{\varepsilon|E(G)|} \cdot \text{poly}(|G|)$.*
4. *It has access to an oracle that computes $F(G; z \cdot w')$, where w' can be any weight function that assigns integer weights w'_e satisfying $0 \leq w'_e \leq C_\varepsilon$ for some constant C_ε that only depends on ε .*

We remark that non-negative integer multiples of z , say $z \cdot w'_e$, can be thought of as w'_e parallel edges of weight z in a multigraph. The quantity $F(G'; z)$ for this multigraph G' is then equal to the value $F(G; z \cdot w')$ of the weighted forest polynomial of G at $z \cdot w'$. In particular, $F(G'; 1)$ is the number of forests in G' .

Proof. Let (G, w) with $w_e \in \{x, y\}$ for all $e \in E$ and $\varepsilon > 0$ be given as input, and let $m := |E(G)|$. We define $C_\varepsilon \in \mathbf{N}$ as a large enough constant to be determined later. The algorithm assigns a new weight $z \cdot w'_e$ to each edge e , where each w'_e is chosen from the set of indeterminates $X \cup Y$ with $X = \{x_1, \dots, x_{m/C_\varepsilon}\}$ and $Y = \{y_1, \dots, y_{m/C_\varepsilon}\}$ in the following way: If $w_e = x$, choose $w'_e \in X$, and if $w_e = y$, choose $w'_e \in Y$. We further demand that the number of edges sharing the same weight is at most C_ε for each weight in $X \cup Y$. Among all such assignments $z \cdot w'$, we pick an arbitrary one. We now consider the polynomial $F(G; z \cdot w')$. It has at most $2m/C_\varepsilon$ variables and the maximum degree of each variable is at most C_ε , so $F(G; z \cdot w')$ has at most $(C + 1)^{2m/C_\varepsilon}$ monomials. The coefficients of this polynomial can be reconstructed when its values are given for all evaluation points in $z \cdot [0, C_\varepsilon]^{2m/C_\varepsilon}$, which is an $2m/C_\varepsilon$ -dimensional grid dilated by a factor z (that is, for any two grid points, the distance between any two entries in the same coordinate is z).

Since each evaluation point only uses non-negative integer multiples of z between 0 and $z \cdot C_\varepsilon$, we can obtain the values at these evaluation points by querying the oracle for $F(G; z \cdot w')$ that we are given. The number of evaluation points in the grid is equal to $(C_\varepsilon + 1)^{2m/C_\varepsilon}$. The claim on the running time follows since the interpolation can be

performed in time $\text{poly}\left((C_\varepsilon + 1)^{2m/C_\varepsilon}\right)$, which is at most $C_\varepsilon \cdot 2^{\varepsilon m}$ when C_ε is chosen large enough depending on ε .

In order to obtain the coefficient of $x^i y^j$ in $F(G; w)$, we compute the image of $F(G; z \cdot w')$ under the projection that maps all variables in X to x/z and all variables in Y to y/z . That is, we sum up the coefficients of $F(G; z \cdot w')$ corresponding to the same monomial $x^i y^j$, and divide by the factor z^{i+j} . ◀

The combination of Lemma 7 and Lemma 8 shows, for all fixed $x \neq 0$, that it is hard to evaluate $F(G; x)$ for multigraphs with at most a constant number of parallel edges. Next we apply a stretch to make the graphs simple. To this end, we calculate the effect of a k -stretch on the univariate forest polynomial of a graph.

► **Lemma 9** (The forest polynomial under a k -stretch). *Let G be a multigraph with m edges, where every edge is weighted with $w \in \mathbf{R}$ and let k be a positive integer such that the number $g_k(w)$ with*

$$g_k(w) = \frac{w^k}{(w+1)^k - w^k}$$

is well-defined. Let G' be the simple graph obtained from G by replacing every edge by a path of k edges. Then we have

$$F(G'; w) = ((w+1)^k - w^k)^m \cdot F(G; g_k(w)).$$

Proof. We define a mapping ϕ that maps forests in G' to forests in G as follows: We add an edge $e \in E(G)$ to $A = \phi(A')$ if and only if A' contains all k edges of G' that e got stretched into. That is, subgraphs A' that only differ by edges in “incomplete paths” are mapped to the same multigraph A by ϕ .

Clearly, ϕ partitions $\mathcal{F}(G')$ into sets of forests with the same image under ϕ . Let A be a forest in G , and let us describe a way to generate all A' with $\phi(A') = A$. First, for each $e \in A$, we add its corresponding path in G' of length k to A' . Moreover, for each edge $e \in E(G) \setminus A$, we can add to A' any proper subset of edges from the k -path in G' that corresponds to e . Therefore, at each $e \in E(G) \setminus A$ independently, there are $\binom{k}{i}$ ways to extend A' by i edges to a forest in G' . A forest A' can be obtained in this fashion if and only if $\phi(A') = A$ holds.

For a fixed A , let us consider all summands $w^{|A'|}$ in $F(G'; w)$ with $\phi(A') = A$. By the above considerations, the total weight contribution of these summands is $w^{k \cdot |A|} \cdot \left(\sum_{i=0}^{k-1} \binom{k}{i} w^i\right)^{m-|A|}$, which equals $w^{k \cdot |A|} \cdot ((w+1)^k - w^k)^{m-|A|}$ by the binomial theorem. These remarks justify the following calculation for the forest polynomial:

$$\begin{aligned} F(G'; w) &= \sum_{A \in \mathcal{F}(G)} \sum_{\substack{A' \in \mathcal{F}(G') \\ \phi(A')=A}} w^{|A'|} = \sum_{A \in \mathcal{F}(G)} w^{k \cdot |A|} \cdot \left((w+1)^k - w^k\right)^{m-|A|} \\ &= \left((w+1)^k - w^k\right)^m \cdot \sum_{A \in \mathcal{F}(G)} \left(\frac{w^k}{(w+1)^k - w^k}\right)^{|A|}. \end{aligned}$$

Since the sum in the last line is equal to $F(G; g_k(w))$, this concludes the proof. ◀

We are now in position to formally prove the main theorem of this section.

Proof of Theorem 6. Let $x \in \mathbf{R} \setminus \{1\}$. Suppose that, for all $\varepsilon > 0$, there exists an algorithm B to compute the mapping $G \mapsto T(G; x, 1)$ in time $2^{\varepsilon n}$ for given simple graphs G

with at most $C'_\varepsilon n$ edges, where C'_ε will be chosen later. By (3), algorithm B can be used to compute values $F(G; (x-1)^{-1})$ with no relevant overhead in the running time, so let $t = (x-1)^{-1}$. Given such an algorithm (or family of algorithms), we devise a similar algorithm for counting perfect matchings, which together with Theorem 5 implies that #ETH is false.

Let G be a simple n -vertex graph with at most $C \cdot n$ edges. Let G' be the graph obtained from G as in Lemma 7 by adding an apex, labeling the edges incident to the apex with the indeterminate z , and all other edges with the indeterminate w . By Lemma 7, the coefficients of the corresponding bivariate forest polynomial of G' are sufficient to extract the number of perfect matchings of G , so it remains to compute these coefficients.

To obtain the coefficients, we use Lemma 8. The reduction guaranteed by the lemma produces $2^{\varepsilon m}$ multigraphs H , all with the same vertex set $V(G')$. Moreover, each H has at most $C_\varepsilon |E(G')| = C_\varepsilon (|E(G)| + n) \leq O(C_\varepsilon n)$ edges, and the multiplicity of each edge is at most C_ε . Finally, each edge of each H is assigned the same weight z , which we will choose later.

The reduction makes one query for each H , where it asks for the value $F(H; z)$. Our assumed algorithm however only works for simple graphs, so we perform a 3-stretch to obtain a simple graph H' with at most $3|E(H)| \leq O(C_\varepsilon n)$ edges. Lemma 9 allows us to efficiently compute the value $F(H; z)$ when we are given the value $F(H'; t)$ and $z = g_3(t)$ holds. Since g_k is a total function whenever k is a positive odd integer, and 3 is indeed odd, the value $g_3(t)$ is well-defined, and we set $z = g_3(t)$.

We set C'_ε large enough so that $E(H') \leq C'_\varepsilon \cdot n$ holds. Tracing back the reduction chain, we can use algorithm B to compute $T(H'; x, 1)$ in time $2^{\varepsilon n}$ any $\varepsilon > 0$. Using (3), we get the value of $F(H'; t)$ since $x \neq 1$. This, in turn, yields the value of $F(H; z)$ since $(z+1)^k - z^k \neq 0$ and $g_3(t) = z$. We do this for each of the $2^{\varepsilon m}$ queries H that the reduction in Lemma 8 makes. Finally, the latter reduction outputs the coefficients of the bivariate forest polynomial of G' , which contains the information on the number of perfect matchings of G .

To conclude, assuming the existence of the algorithm family B , we are able to count perfect matching in time $\text{poly}(2^{\varepsilon m})$ for all $\varepsilon > 0$, which implies via Theorem 5 that #ETH is false. \blacktriangleleft

Note that the construction from the proof of Theorem 1 implies hardness of $T(G; x, 1)$ for tripartite G , and also in the bipartite case whenever $x \neq -1$.

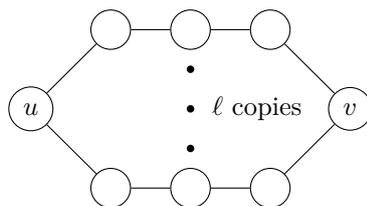
4 Counting solutions to Boolean CSPs under #ETH

In this section, we prove that the #P-hard cases of the dichotomy theorem for Boolean CSPs by Creignou and Hermann [6] are also hard under #ETH. The main difficulty is to establish #ETH-hardness of counting independent sets in bipartite graphs. We do so first, and afterwards observe that all other reductions in [6] can be used without modification.

4.1 Counting Independent Sets in Bipartite Graphs is #ETH-hard

We prove that the problem of counting independent sets in bipartite graphs admits no subexponential algorithm under #ETH, even for sparse and simple graphs.

Proof of Theorem 3. We reduce from the problem of counting independent sets in graphs of bounded degree; by Theorem 5, this problem does not have a subexponential-time algorithm. First we note that a set is an independent set if and only its complement is a vertex cover.



■ **Figure 1** The gadget H_ℓ of Provan and Ball [16] as used in the proof of Theorem 3. It corresponds to an ℓ -fattening of the edge $\{u, v\}$, followed by a 4-stretch of each of the ℓ parallel edges.

Hence their numbers are equal. We devise a subexponential-time oracle reduction family to reduce counting vertex covers in general to counting them in bipartite graphs.

Given a graph G with n vertices and m edges, and a running time parameter $d \in \mathbb{N}$, the reduction works as follows. We partition the edges into $\lfloor \frac{|E|}{d} \rfloor$ blocks of size at most d each. We denote the blocks by $B_1, \dots, B_{\frac{m}{d}}$. Next, for each $\vec{\ell} = (\ell_1, \dots, \ell_{\frac{m}{d}}) \in \mathbb{N}^{m/d}$, we denote $G_{\vec{\ell}}$ as the graph obtained from G by replacing each edge $e \in B_i$ with a copy of the gadget H_{ℓ_i} shown in Figure 1. Note that $G_{\vec{\ell}}$ is bipartite.

► **Observation 10** (Provan and Ball). *The number of vertex covers of H_ℓ containing neither u nor v is 2^ℓ , the number of vertex covers containing a particular one of u or v is 3^ℓ , and the number of vertex covers containing both u and v is 5^ℓ .*

We follow the proof of Provan and Ball, but do so in a block-wise fashion. To this end, let T be the set of all $(m/d) \times 3$ matrices with entries from $\{0, \dots, d\}$. The *type* of a set $S \subseteq V(G)$ is the matrix $t \in T$ such that, for all $i = 1 \dots \frac{m}{d}$,

1. t_{i1} is equal to the number of edges $e \in B_i$ with $|e \cap S| = 0$,
2. t_{i2} is equal to the number of edges $e \in B_i$ with $|e \cap S| = 1$, and
3. t_{i3} is equal to the number of edges $e \in B_i$ with $|e \cap S| = 2$.

Every set $S \subseteq V(G)$ has exactly one type. Let x_t be the number of all sets $S \subseteq V(G)$ that have type t .

We classify vertex covers $C \subseteq V(G_{\vec{\ell}})$ of $G_{\vec{\ell}}$ by their intersection with $V(G)$, so let $S = C \cap V(G)$ and let t be the type of S . By Observation 10 and the fact that all inserted gadgets act independently after conditioning on the intersection of the vertex covers of $G_{\vec{\ell}}$ with $V(G)$, there are exactly $\prod_{i=1}^{m/d} (2^{t_{i1}} 3^{t_{i2}} 5^{t_{i3}})^{\ell_i}$ vertex covers C' whose intersection with $V(G)$ is S . Moreover, the number of sets S of type t is equal to x_t . Hence the number $N_{\vec{\ell}}$ of vertex covers of $G_{\vec{\ell}}$ satisfies

$$N_{\vec{\ell}} = \sum_{t \in T} x_t \cdot \prod_{i=1}^{m/d} (2^{t_{i1}} 3^{t_{i2}} 5^{t_{i3}})^{\ell_i} \quad (7)$$

Since $G_{\vec{\ell}}$ is bipartite, our reduction can query the oracle to obtain the numbers $N_{\vec{\ell}}$ for all $\vec{\ell} \in [(d+1)^3]^{\frac{m}{d}}$. This yields a system of linear equations of type (7), where the x_t for $t \in T$ are the unknowns; note that we have exactly $|T|$ equations and unknowns. Let M be the corresponding $|T| \times |T|$ matrix, so that the system can be written as $N = M \cdot x$.

It remains to prove that M is invertible. For this, we observe that M can be decomposed into a tensor product of smaller matrices as follows. Let A be the $(d+1)^3 \times (d+1)^3$ where the row indices ℓ are from $[(d+1)^3]$, the column indices τ are from $\{0, \dots, d\}^3$, and the entries are defined via $A_{\ell\tau} = (2^{\tau_1} 3^{\tau_2} 5^{\tau_3})^\ell$. Provan and Ball, as well as the reader, observe that A is the transpose of a Vandermonde matrix. Due to the uniqueness of the prime factorization,

the evaluation points $2^{\tau_1}3^{\tau_2}5^{\tau_3}$ are distinct for distinct τ , and thus $\det(A) \neq 0$. Furthermore, we observe that $M = A^{\otimes \frac{n}{d}}$ holds, which implies $\det(M) \neq 0$ and that M is invertible.

Since M is invertible, we can solve the equation system $N = M \cdot x$ in time polynomial in its size, and compute x_t for all $t \in T$. Finally, we compute the sum of x_t over all matrices t whose first column contains only zeros. This yields the number of all sets $S \subseteq V(G)$ that intersect every edge of G at least once, that is, the number of vertex covers of G which equals, as mentioned above, the number of independent sets of G .

Assume that #ETH holds, and let $\varepsilon, D > 0$ be the constants from Theorem 5, which are such that no algorithm can count independent sets in general graphs of maximum degree D in time $2^{\varepsilon n}$. We apply our reduction to such a graph; it makes at most $2^{O(\log d \cdot m/d)}$ queries to the oracle. Since $m \leq Dn$ holds, and the running time for solving the linear equation system is polynomial in the number of queries, we can choose $d \in \mathbf{N}$ to be a large enough constant depending on $\varepsilon > 0$ to achieve an overall running time of $O(2^{\frac{1}{2}\varepsilon n})$ for the reduction. Also note that the queries to the oracle for bipartite graphs have degree at most $(d+1)^3 \cdot D$, which is a constant that only depends on ε . If there was an algorithm for counting independent sets in bipartite graphs that ran in time $O(2^{\frac{1}{2}\varepsilon n})$, we would get a combined algorithm for counting independent sets in general graphs that would be faster than the choice of ε and D would allow. Hence, under #ETH, there are constants $\varepsilon', D' > 0$ such that no $O(2^{\varepsilon' n})$ -time algorithm can count all independent sets on graphs of maximum degree at most D' . ◀

4.2 The Boolean CSP dichotomy

Instances of the constraint satisfaction problem #CSP(Γ) are conjunctions of relations in Γ applied to variables over the Boolean domain and the goal is to compute the number of satisfying assignments. A satisfying assignment is an assignment to the variables such that the formula evaluates to true, that is, every relation in the conjunction evaluates to true. A more detailed description of the problem can be found in the paper of Creignou and Hermann [6].

Creignou and Hermann prove Theorem 4 by reducing either from #Pos2Sat, the problem of counting satisfying assignments of a 2-CNF where every literal is positive, or from #Imp2Sat, the problem of counting satisfying assignments of a 2-CNF where every clause contains exactly one positive and one negative literal. A straightforward analysis of the construction reveals that the reductions only lead to a linear overhead. More precisely:

► **Observation 11.** *Given an instance of #Pos2Sat or #Imp2Sat with n variables and a set Γ of logical relations such that at least one of the relations is not affine, the construction of Creignou and Hermann results in an instance of #CSP(Γ) of size $c \cdot n$ where c only depends on the size of the largest non-affine relation in Γ .*

Therefore it suffices to establish that neither #Pos2Sat nor #Imp2Sat have a $2^{o(n)}$ -time algorithm. Since #Pos2Sat is identical to counting vertex covers in (general) graphs, Theorem 5 applies here. The #ETH-hardness of #Imp2Sat follows by a known reduction from counting independent sets in bipartite graphs, which we include here for completeness.

► **Lemma 12.** *Assuming #ETH, there is no algorithm that solves #Imp2Sat in time $2^{o(n)}$ where n is the number of variables.*

Proof. Given a bipartite graph $G = (V \dot{\cup} U, E)$ with constant degree we construct a 2-CNF F by adding a clause $(v \rightarrow u)$ for every edge $\{v, u\} \in E$. Now the number of independent sets in G equals the number of satisfying assignments of F . Furthermore the existence of an algorithm that solves #Imp2Sat in time $2^{o(n)}$ would imply the existence of an algorithm that

solves #BIS in time $2^{o(n)}$. Applying Theorem 3 we obtain that such an algorithm would refute #ETH. ◀

We sketch how to obtain the #ETH dichotomy theorem for Boolean CSPs.

Proof of Corollary 4. If every relation in Γ is affine then we can solve #CSP(Γ) in polynomial time using Gaussian elimination as in [6]. Otherwise, the problem is #P-hard by [6]. If, in addition, #ETH holds, #CSP(Γ) cannot be solved in time $2^{o(n)}$ as a subexponential algorithm could also be used to solve #Pos2Sat or #Imp2Sat (see Observation 11) in time $2^{o(n)}$ which is not possible assuming #ETH (by Theorem 5 and Lemma 12). ◀

Acknowledgments. We thank Radu Curticapean for various fruitful discussions and interactions, and also Leslie Ann Goldberg, Miki Hermann, Mark Jerrum, John Lapinskas, David Richerby, and all other participants of the “dichotomies” work group at the Simons Institute in the spring of 2016. Moreover, we thank Tyson Williams who in 2013 pointed the second author to [11], and Thore Husfeldt for encouraging us to pursue the publication of this manuscript.

References

- 1 Andreas Björklund. Counting perfect matchings as fast as Ryser. In *Proceedings of the 23rd Symposium on Discrete Algorithms, SODA 2012*, pages 914–921, 2012. doi:10.1137/1.9781611973099.73.
- 2 Andreas Björklund, Thore Husfeldt, Petteri Kaski, and Mikko Koivisto. Computing the Tutte polynomial in vertex-exponential time. In *Proceedings of the 47th annual IEEE Symposium on Foundations of Computer Science, FOCS 2008*, pages 677–686, 2008. doi:10.1109/FOCS.2008.40.
- 3 Jin-Yi Cai and Xi Chen. Complexity of counting CSP with complex weights. In *Proceedings of the 44th Symposium on Theory of Computing, STOC 2012*, pages 909–920, 2012. doi:10.1137/100814585.
- 4 Jin-Yi Cai, Pinyan Lu, and Mingji Xia. Computational complexity of holant problems. *SIAM Journal on Computing*, 40(4):1101–1132, 2011. doi:10.1137/100814585.
- 5 Chris Calabro, Russell Impagliazzo, and Ramamohan Paturi. A duality between clause width and clause density for SAT. In *Proceedings of the 21st Annual IEEE Conference on Computational Complexity, CCC’06*, pages 252–260, Washington, DC, USA, 2006. IEEE Computer Society. doi:10.1109/CCC.2006.6.
- 6 Nadia Creignou and Miki Hermann. Complexity of generalized satisfiability counting problems. *Information and Computation*, 125(1):1–12, 1996. doi:10.1006/inco.1996.0016.
- 7 Radu Curticapean. Block interpolation: A framework for tight exponential-time counting complexity. In *Proceedings of the 42nd International Colloquium on Automata, Languages and Programming, ICALP 2015*, pages 380–392. Springer, 2015. doi:10.1007/978-3-662-47672-7_31.
- 8 Holger Dell, Thore Husfeldt, Dániel Marx, Nina Taslaman, and Martin Wahlén. Exponential time complexity of the permanent and the Tutte polynomial. *ACM Transactions on Algorithms*, 10, 2014. doi:10.1145/2635812.
- 9 Holger Dell, Thore Husfeldt, and Martin Wahlén. Exponential time complexity of the permanent and the Tutte polynomial. In *Proceedings of the 37th International Colloquium on Automata, Languages and Programming, ICALP 2010*, pages 426–437, 2010. doi:10.1007/978-3-642-14165-2_37.
- 10 Joanna Ellis-Monaghan and Iain Moffatt. *CRC handbook on the Tutte polynomial and related topics*. CRC press, in preparation.

- 11 Geoffrey Grimmett and Colin McDiarmid, editors. *Combinatorics, Complexity, and Chance: A Tribute to Dominic Welsh*. Oxford Lecture Series in Mathematics and Its Applications (Book 34). Oxford University Press, 2007.
- 12 Thore Husfeldt and Nina Taslaman. The exponential time complexity of computing the probability that a graph is connected. In *Proceedings of the 5th International Symposium on Parameterized and Exact Computation, IPEC 2010*, pages 192–203, 2010. doi:10.1007/978-3-642-17493-3_19.
- 13 Russell Impagliazzo and Ramamohan Paturi. On the complexity of k -SAT. *Journal of Computer and System Sciences*, 62(2):367–375, 2001. doi:10.1006/jcss.2000.1727.
- 14 Russell Impagliazzo, Ramamohan Paturi, and Francis Zane. Which problems have strongly exponential complexity? *Journal of Computer and System Sciences*, 63(4):512–530, 2001. doi:10.1006/jcss.2001.1774.
- 15 François Jaeger, Dirk L. Vertigan, and Dominic J. A. Welsh. On the computational complexity of the Jones and Tutte polynomials. *Mathematical proceedings of the Cambridge Philosophical Society*, 108(1):35–53, 1990. doi:10.1017/S0305004100068936.
- 16 J. Scott Provan and Michael O Ball. The complexity of counting cuts and of computing the probability that a graph is connected. *SIAM Journal on Computing*, 12(4):777–788, 1983. doi:10.1137/0212053.
- 17 Alan D. Sokal. The multivariate Tutte polynomial (alias Potts model) for graphs and matroids. In *Surveys in Combinatorics*, volume 327 of *London Mathematical Society Lecture Note Series*, pages 173–226, 2005.
- 18 Nina Sofia Taslaman. *Exponential-Time Algorithms and Complexity of NP-Hard Graph Problems*. PhD thesis, IT-Universitetet i København, 2013.
- 19 Seinosuke Toda. PP is as hard as the polynomial-time hierarchy. *SIAM Journal on Computing*, 20(5):865–877, 1991. doi:10.1137/0220053.
- 20 Leslie G. Valiant. The complexity of computing the permanent. *Theoretical Computer Science*, 8(2):189–201, 1979. doi:10.1016/0304-3975(79)90044-6.
- 21 Dirk Llewellyn Vertigan. *On the computational complexity of Tutte, Jones, Homfly and Kauffman invariants*. PhD thesis, University of Oxford, 1991.