# Counting and Correlation Decay in Spin Systems

*Piyush Srivastava*

Electrical Engineering and Computer Sciences
University of California at Berkeley

August 1, 2014

# Counting and Correlation Decay in Spin Systems

by

Piyush Srivastava

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Computer Science

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Alistair Sinclair, Chair
Professor Elchanan Mossel
Professor Umesh Vazirani
Professor Allan Sly

Fall 2014

**Counting and Correlation Decay in Spin Systems**

# Abstract

Counting and Correlation Decay in Spin Systems

by

Piyush Srivastava

Doctor of Philosophy in Computer Science

University of California, Berkeley

Professor Alistair Sinclair, Chair

Spin systems originated in statistical physics as tools for modeling phase transitions in magnets. However, they have since been used to model complex systems arising in several other fields of study (e.g., in Bayesian inference and in the theory of social networks), so that various computational problems associated with them have received much attention in the literature. A lot of progress in the study of these problems has relied upon ideas from statistical physics; one example of this interplay has been the discovery of tight connections between the computational complexity of these problems and the phenomenon of phase transitions that many spin systems exhibit. Conversely, algorithmic ideas have also helped in the study of the phase transition phenomenon.

This thesis presents two lines of work that fit into this theme. The first considers the approximation of the *partition function*, a pivotal quantity associated with spin systems. Here, we obtain—in various settings—deterministic polynomial time algorithms for approximating the partition function in the so called *uniqueness regime*, and thus strengthen the algorithmic side of the tight correspondence between phase transitions and the computational complexity of the partition function.

The second line of work is concerned with the complexity of exact computation of various natural *mean observables* of spin systems, e.g., the magnetization in the Ising model. We relate these questions to the location of the complex zeros of the partition function, which have been studied in statistical physics because of their connections to the existence of phase transitions, and have been the subject of various celebrated results (such as the Lee-Yang theorem and the Heilmann-Lieb theorem). By proving a novel extension of the Lee-Yang theorem, we show that the magnetization of the ferromagnetic Ising model is indeed as hard to compute as the partition function (i.e., #P-hard). We also obtain similar results for the monomer-dimer model.

नानाजी और दादीजी को समर्पित

*Dedicated to my grandparents*

# Contents

# List of Figures

# List of Tables

# Acknowledgments

I would like to begin by thanking my adviser, Alistair Sinclair. Collaborating with Alistair was a great learning experience in all aspects of research, and he was also a constant source of advice and encouragement all through my time at Berkeley.

Much of this thesis is based on joint work with my collaborators: Alistair Sinclair, Daniel Štefankovič, Mario Szegedy, Mark Thurley, and Yitong Yin. Doing research with them was as pleasant as it was edifying, and I am grateful for the chance. A special note of thanks is due to Daniel Štefankovič for his indispensable help in cross-checking the numerical results in Section 4.5. I would also like to thank Elchanan Mossel, Allan Sly and Umesh Vazirani for agreeing to be on my thesis committee and for their helpful comments at various stages of this work. Special thanks are also due to Mario Szegedy, Nisheeth Vishnoi and Dror Weitz for helpful discussions (indeed, a significant portion of Part I of this thesis could quite fairly be dubbed "footnotes to Weitz"). Part of the work appearing in this thesis was done at Microsoft Research, Bangalore, and IBM Research, Almaden: I thank these organizations and my hosts there (Nisheeth Vishnoi and T. S. Jayram respectively) for their hospitality.

Soda Hall provided a very pleasant and stimulating atmosphere for research, with ideas always "flying around", sometimes quite literally. I must thank all the very nice office-mates I had: Siu Man Chan, Siu On Chan, Urmila Mahadev, George Pierrakos, Jonah Sherman, Di Wang, and Chris Wilkens. Insights from other Soda Hall "inmates" were always around the corner, and I got to learn a lot from everybody in the group. However, I would like to specially thank Nima Anari, Anand Bhaskar, Antonio Blanca, Andrew Chan, Siu Man Chan, Siu On Chan, James Cook, Anindya De, Rafael Frongillo, Sagar Jain, Urmila Mahadev, Mohammad Moharrami, Aviad Rubinstein, Seung Woo Shin, Greg Valiant, Thomas Vidick, Thomas Watson and Sam Wong for various enlightening discussions. On the same note, I would also like to thank Satish Rao and Umesh Vazirani for their free-form "Toolkit" class, Zeph Landau for organizing the Hamiltonian Complexity reading groups, Christos Papadimitriou for his thought-provoking talks on everything from Economics to Evolution, and Prasad Raghavendra and James Lee for their lectures on hardness of approximation.

I was also lucky to have some very good friends at Berkeley outside the ever-so-pleasant Soda Hall. I must first thank all my excellent roommates over the years: Anindya, Ankush, Debkishore, Devi, Kartik and Vinay; and also Ajith, Arka, Himanshu, Kaushik, Momo, Raj and Shaunak for their joyful camaraderie. I must specially thank Anindya for helping me settle down in Berkeley during my first few weeks here, and Arka for rescuing my dissertation talk from a technical disaster (these happen in Soda hall too!).

My decision to even embark on a Ph. D. was likely much influenced by some very inspiring people I got a chance to work with while in college. I thank Manindra Agrawal, Neeraj Kayal, Piyush Kurur, Leonard Schulman and Chris Umans for supervising my first few "research" projects in computational complexity, and Somenath Biswas, Sumit Ganguly and Shashank Mehta for their infectious enthusiasm for the subject. Even before I came to college, I was quite fortunate to have some wonderful teachers at school; I must specially thank Dr. Aditi Mukhopadhyay for her passionate teaching of physics and her deep interest in the intellectual development of

# Chapter 1

# Introduction and preliminaries

Spin systems originated in statistical mechanics in the study of phase transitions in magnets [Isi25]. Since then, aside from their role in statistical mechanics, they have become important objects of study in both probability theory and computer science, especially under the name of *Markov random fields* or *graphical models*. Research on various natural computational problems associated with spin systems has therefore seen a rich interplay between fundamental notions arising in these diverse fields. The results in this thesis add to this line of work by giving two examples of how tools developed in statistical mechanics for the study of phase transitions can be used to study the computational complexity of counting problems.

In general, a spin system is a probability distribution over certain combinatorial structures defined on a given graph $G = (V, E)$. The distribution itself is defined in terms of local interactions, in such a way that the marginal distribution on a subgraph $H$ is independent of the distribution on $G - H$ given the state of the neighbors of $H$ in $G - H$ (this property motivates the name "Markov random field"). Perhaps the simplest such model is the *hard core model* (more descriptively, the *hard core lattice gas with pair interaction* [SS05]). Here the configurations are simply the independent sets of the graph. The local interactions have two components, the first of which is of course the constraint that the configuration must never contain two adjacent vertices. The second is given in terms of a positive real parameter $\lambda$ called the *vertex activity* or *fugacity*, which models the tendency of vertices to lie in an independent set. The spin system then assigns a *weight* $w(I) = \lambda^{|I|}$ to each configuration, which in turn defines the *Gibbs measure*

$$\mu(I) := \frac{1}{Z} w(I),$$

where the normalization factor

$$Z = Z_G(\lambda) := \sum_{I:\text{ independent set of } G} \lambda^{|I|}$$

is called the *partition function*. The partition function turns out to be a quantity of much interest for a variety of reasons. In computer science, the (exact) computation of the partition function is a natural generalization of counting problems (e.g., setting $\lambda = 1$ turns the partition function of

the hard core model into the number of independent sets). Indeed, the computation of partition functions of various spin systems (also known as the problem of counting graph homomorphisms) is one of the best understood classes of #P-hard problems (see, e.g., [CCL10]). Approximating the partition function is of importance in its own right, in part due to its tight connections with the problem of *sampling* from the Gibbs measure: for a wide class of problems (known as "self-reducible" problems) the two problems are essentially equivalent [JVV86]. (The exact nature of this equivalence depends upon technical conditions on the interactions and the kind of guarantees required from the sampling and approximation algorithms, and is not important for the purposes of the present discussion). The problem of approximating the partition function also turns out to have very strong connections with the question of the existence of "long range" correlations in the Gibbs measure: we will discuss these connections in greater detail below.

However, at an even more basic level, the partition function is a generating polynomial for combinatorial structures, and hence encodes information about several natural *observables* associated with these structures. For example, suppose we want to compute the average size $\langle |I| \rangle$ of an independent set sampled according to the Gibbs distribution. It is easy to verify that this can be written in terms of the partition function and its derivative:

$$\langle |I| \rangle = \frac{\lambda}{Z(\lambda)} \frac{\partial Z(\lambda)}{\partial \lambda}.$$

It is in fact a consequence of the definition of the Gibbs distribution that several natural physical observables can be written in terms of the partition function and its derivatives, and this property of the partition function is what makes it interesting in statistical physics. To explore this connection further, we consider a more classical example, that of the *Ising model* [Isi25], a simple model for magnetism that was the first spin system to be studied. Here, a magnetic material is modeled as a graph whose vertices are magnetic domains, and whose edges represent interactions between these domains. The configurations are assignments of "+" and "−" *spins* to the vertices; these spins are thought of as the polarities of the domains. The local interactions are defined in terms of a positive *edge activity* $\beta$ and a positive *vertex activity* $\lambda$. We can represent the configurations as function $\sigma : V \rightarrow \{+, -\}$, and the weight $w(\sigma)$ of a configuration $\sigma$ is given by

$$w(\sigma) = \lambda^{p(\sigma)} \beta^{d(\sigma)}, \tag{1.1}$$

where $p(\sigma)$ denotes the number of vertices assigned spin +, and $d(\sigma)$ denotes the number of edges $e = \{i, j\}$ such that $\sigma(i) \neq \sigma(j)$. The Gibbs measure and the partition function are defined exactly as before:

$$Z = Z_G(\beta, \lambda) := \sum_{\sigma : V \rightarrow \{+, -\}} w(\sigma); \quad \mu(\sigma) = \frac{1}{Z} w(\sigma).$$

Note that the distribution favors "disagreements" on edges when $\beta > 1$, in which case the model is called *anti-ferromagnetic*; the opposite occurs in the *ferromagnetic* case $\beta < 1$. The parameter $\lambda$ models an "(external) magnetic field": in the "positive field" setting ($\lambda > 1$), vertices prefer

to be assigned the "+" spin, while in the "zero field" setting ($\lambda = 1$), the two spins are entirely symmetric. The chief observable of interest in the Ising model is the *magnetization $p(\sigma)$*, which is the number of $+$ spins in the configuration. It is again easy to verify that the *mean magnetization $M$* under the Gibbs distribution is given by

$$M_G(\beta, \lambda) := \langle p \rangle = \frac{\lambda}{Z(\beta, \lambda)} \frac{\partial Z(\beta, \lambda)}{\partial \lambda}. \tag{1.2}$$

Interestingly, although the *approximate* computation of the mean observables is an important application of the approximate sampling algorithms described above, the determination of the computational complexity of their *exact* computation has not been explored in much detail. In contrast to the extensive results on the computational complexity of partition functions alluded to above, no such results for mean observables were known.

In this thesis, we will consider the following two broad classes of problems:

1. Under what conditions can the partition function be efficiently approximated?

2. Are mean observables such as those discussed above as hard to compute exactly as the corresponding partition functions?

The two questions turn out to have connections with two very different lines of work in the study of phase transitions. We now describe these connections.

## 1.1  Correlation decay and approximation algorithms

The Gibbs measure of the hard core model shows an interesting phase transition on the $d$-ary tree: there exists a criticial activity $\lambda_c(d) := \frac{d^d}{(d-1)^{d+1}}$ such that, when $\lambda < \lambda_c(d)$, point-to-set correlations in the Gibbs distribution decay exponentially with distance, while they do not decay at all when $\lambda > \lambda_c(d)$. More formally, suppose we consider the $d$-ary tree of depth $\ell$, and fix all leaves to be occupied (respectively, unoccupied) by the independent set. Conditioned on this fixing, the Gibbs measure induces a conditional probability $p_+(\ell)$ (respectively, $p_-(\ell)$) of the root being occupied. The phase transition manifests as follows: when $\lambda < \lambda_c(d)$, we have

$$|p_+(\ell) - p_-(\ell)| = \exp\left(-\Theta(\ell)\right),$$

while when $\lambda > \lambda_c(d)$,

$$|p_+(\ell) - p_-(\ell)| = \Theta(1),$$

even in the limit $\ell \to \infty$. This phase transition is referred to as the *uniqueness phase transition*, for reasons described in Chapter 3. In a seminal paper, Weitz [Wei06] established a very strong relationship between this phase transition and the complexity of approximating the partition function of the hard core model: he showed that when $\lambda < \lambda_c(d)$, there exists a fully polynomial time approximation scheme (FPTAS) for the partition function of the hard core model on all graphs of degree at most $d + 1$. This connection was further strengthened by Sly [Sly08] (see

also [SS12, GGŠ+14]) who showed that an FPTAS for $(d + 1)$-regular graphs for $\lambda > \lambda_c(d)$ would imply that NP=RP. Taken together, these results imply that the uniqueness phase transition on the $d$-ary tree exactly determines the complexity of approximating the partition function of the hard core model on graphs of degree at most $d + 1$.

In the first part of this thesis (Chapters 2 to 4), we explore this connection between the uniqueness phase transition on trees and the approximation of the partition function. After some discussion of the background in Chapter 2, we give an analog of Weitz's result for the anti-ferromagnetic Ising model in Chapter 3, where we show that the uniqueness phase transition for the anti-ferromagnetic Ising model on the $d$-ary tree also determines the complexity of approximating the partition function of the model on graphs of degree at most $d + 1$. In Chapter 4, we then extend Weitz's methods beyond bounded degree graphs to graphs of bounded *connective constant*, a strict superset of the class of bounded degree graphs. Our results in terms of the connective constant also allow us to study decay of correlations on graphs in terms of correlation decay on $d$-ary trees at a much finer level than that allowed by considerations of only the maximum degree of the graph.

## 1.2   Lee-Yang zeros and the complexity of mean observables

A more classical view in physics of a phase transition is as a discontinuity in the mean value of an observable with respect to a model parameter. We will use the Ising model as our running example to see what this view entails in terms of the properties of the partition function. We begin by considering the *free energy per unit volume* [Sim93], which may be written as a normalized logarithm of the partition function:

$$F_G(\beta, \lambda) := \frac{1}{|V|} \log Z_G(\beta, \lambda).$$

The use of the logarithm means that mean observables discussed above correspond directly to the derivatives of the free energy, without any further normalization by the partition function. For example, the (normalized) mean magnetization, which is simply the expected *fraction* of vertices assigned the "+" spin, can be written simply as

$$\frac{1}{|V|} M_G(\beta, \lambda) = z \frac{\partial F_G(\beta, z)}{\partial z} \bigg|_{z=\lambda}. \tag{1.3}$$

**Remark 1.1.** The normalization by a factor of $1/|V|$ (corresponding to the size of the graph) is necessary because as we shall soon see, we will need to consider limits of these observables on a sequence of graphs of increasing size in order to relate them to the above view of phase transitions. These limits, in turn, can be expected to exist only if we scale the observables by the size of the graph.

Since observables correspond to derivatives of the free energy, while phase transitions correspond to discontinuities of those observables, it follows that a phase transition in this formalism

can be defined in terms of the analyticity of the free energy. In other words, if the free energy $F(z)$ is an analytic function for $z \in D$, then, by definition, the system *cannot* have a phase transition in the region $D$ of the parameter space. However, since the only physically interesting region of the parameter space is $\beta, \lambda > 0$, and the free energy of any finite graph is clearly an analytic function when $\beta, \lambda > 0$, the above system cannot exhibit such a phase transition on a finite graph. One therefore needs to consider infinite graphs to find examples of phase transitions.

In order to define the free energy for such a graph, we take an increasing sequence of finite graphs $G_n$ which approach the given graph "in the limit" as $n \to \infty$, and define the free energy $F_G(\beta, \lambda)$ of $G$ as $\lim_{n \to \infty} F_{G_n}(\beta, \lambda)$. Let us consider the example of the two-dimensional integer lattice $\mathbb{Z}^2$, where the natural choice is to take $G_n$ to be squares of increasing side length centered at the origin.[1] Phase transitions can indeed occur in this setting; in a celebrated paper [Ons44] Onsager explicitly calculated the free energy (as a function of $\beta$) of the Ising model on $\mathbb{Z}^2$ in the zero field case ($\lambda = 1$), and showed that there is a phase transition in the $\beta$ parameter: the second derivative of the free energy (with respect to $\beta$) has a singularity.

Following Onsager's result, Yang and Lee [YL52] sought to identify regions where phase transitions cannot occur. They proved the following general result, which relates the analyticity of the free energy of an infinite graph to the location of the *complex* zeros of partition functions of finite graphs. Their theorem essentially confirms the intuition that non-analyticity of the free energy, which is simply the logarithm of the partition function, can only result if the partition function vanishes.

**Theorem 1.1 (Yang and Lee [YL52]).** *Let the free energy $F_G(\beta, z)$ of an infinite graph be defined as the limit $F_G(\beta, z) = \lim_{n \to \infty} F_{G_n}(\beta, z)$, where $G_n$ is a sequence of* finite *graphs. Let $S$ be an open region in the complex plane in which the limit exists, and which is such that the partition functions $Z_{G_n}(\beta, z)$ do not vanish when $z \in S$. Then $F_G(\beta, z)$ is analytic in $S$. In particular, there are no phase transitions on any open interval $R$ of the positive real line contained inside $S$.*

Theorem 1.1 provided an excellent physical motivation for the study of zeros of partition functions. Its first application appeared in a subsequent paper of Lee and Yang [LY52] where they proved the following striking result about the zeros of the partition function of the ferromagnetic Ising model.

**Theorem 1.2 (Lee and Yang [LY52]).** *Let $G$ be any finite graph, and let $Z_G(\beta, z)$ be the partition function of the ferromagnetic Ising model on $G$ with edge activity $\beta \in (0, 1)$. Then, all zeros of $Z_G(\beta, z)$ lie on the circle $|z| = 1$ in the complex plane.*

Combining Theorem 1.2 with Theorem 1.1 immediately shows that the ferromagnetic Ising model on $\mathbb{Z}^2$ exhibits *no* phase transitions except possibly at $\lambda = 1$ (since any open region of the positive real line not containing the point $\lambda = 1$ is contained in an open region of the complex plane not containing any partition function zeros).

---

[1]Yang and Lee [YL52] showed that for integer lattices, the value $F_G(\beta, \lambda)$ calculated in this way does not depend upon the sequence $G_n$ (and their result also holds for more general classes of graphs).

The Lee-Yang program for proving the absence of phase transitions has since been applied to several other models, e.g., in papers by Asano [Asa70], Heilmann and Lieb [HL70, HL72], Suzuki and Fisher [SF71] and Newman [New74]. Due to its connections with the stability theory of polynomials, it has also been studied in its own right (see, e.g., papers by Ruelle [Rue10] and Borcea and Brändén [BB09a]). In Part II of this thesis (Chapters 6 to 8), we give novel applications of the Lee-Yang program by using results on the distinctness of zeros of the partition function to prove #P-hardness results for mean observables in the ferromagnetic Ising and monomer-dimer models. (The latter model is described in the next section.)

## 1.3 Other models

Both the models discussed above can be seen as a special case of *general two-state spin systems* [GJP03], which are parametrized by a $(+, +)$ edge potential $\alpha_1$, a $(-, -)$ edge potential $\alpha_2$, and a vertex activity $\lambda$. Configurations $\sigma : V \to \{+, -\}$ are still assignments of "+" or "−" spins to the systems, and the weights $w(\sigma)$ are given by

$$w(\sigma) = \lambda^{p(\sigma)} \alpha_1^{e_+(\sigma)} \alpha_2^{e_-(\sigma)},$$

where $e_+(\sigma)$ (respectively, $e_-(\sigma)$) denotes the number of edges with '+' (respectively, '−') spin on both end-points, while $p(\sigma)$ denotes the number of vertices with '+' spin. The partition function $Z_G(\alpha_1, \alpha_2, \lambda)$ and the mean magnetization $M_G(\alpha_1, \alpha_2, \lambda)$ are given by

$$Z_G(\alpha_1, \alpha_2, \lambda) := \sum_{\sigma : V \to \{+, -\}} w_S(\sigma);$$

$$M_G(\alpha_1, \alpha_2, \lambda) := \langle p \rangle = \frac{\sum_\sigma p(\sigma) w_S(\sigma)}{Z_G(\alpha_1, \alpha_2, \lambda)},$$

and the Gibbs measure is defined in terms of the weights exactly as above. The system is said to be *ferromagnetic* when $\alpha_1 \alpha_2 > 1$ and *anti-ferromagnetic* when $\alpha_1 \alpha_2 < 1$. To see the relationship with the nomenclature we used above for the Ising model, note that the Gibbs measure of the Ising model with edge activity $\beta$ can be obtained by setting $\alpha_1 = \alpha_2 = 1/\beta$ in the above model. The hard core model discussed above is also a special case, obtained by setting $\alpha_1 = 0$ and $\alpha_2 = 1$ (and is therefore anti-ferromagnetic).

Another well known model that we will consider, however, does not fit the above framework. This is the monomer-dimer model [HL72], whose configurations are all matchings of a given graph $G = (V, E)$. For a specified *dimer activity* $\gamma > 0$, the model assigns a weight $w(M) = \gamma^{|M|}$ to each matching $M$ of the graph. As before, the weights define the Gibbs distribution $\mu(M) = \frac{1}{Z} w(M)$ over matchings, where

$$Z = Z_G(\gamma) := \sum_{M : \text{matching}} w(M)$$

is the partition function. Again, the problem of computing the partition function is a natural generalization of the problem of counting matchings, which corresponds to the special case $\gamma = 1$.

Unlike the hard core and anti-ferromagnetic Ising models, the monomer-dimer model has the interesting feature that it does *not* exhibit a uniqueness phase transition on the $d$-ary tree: the correlations on the $d$-ary tree for this model always decay exponentially for fixed $\gamma$ and $d$, though the rate goes to $1$ as either $\gamma$ or $d$ approach $\infty$.

## 1.4   Notions of approximation

Since our main concern in much of this thesis will be designing approximation algorithms for partition functions, we describe the notion of approximation that we will use. The algorithms that we discuss will mostly be *fully polynomial time approximation schemes* (FPTAS). Formally, given a positive quantity $A(G)$ which is defined as a function on graphs, an algorithm $\mathcal{A}$ for computing $A(G)$ is an FPTAS if, given a graph $G$ and an accuracy parameter $\epsilon > 0$, it produces in time $\mathrm{poly}\,(|G|, 1/\epsilon)$ an estimate $\hat{A}$ such that $(1 - \epsilon)A(G) \leq \hat{A} \leq (1 + \epsilon)A(G)$. A related notion is that of a *fully polynomial time randomized approximation scheme* (FPRAS). In this case, the algorithm $\mathcal{A}$ is allowed to be randomized. Given an accuracy parameter $\epsilon > 0$ and an error parameter $\delta > 0$, it is then required to produce with probability at least $1 - \delta$ an estimate $\hat{A}$ with the same guarantees as before in time $\mathrm{poly}\,(|G|, 1/\epsilon, \log(1/\delta))$.

## 1.5   Note on prior publication and collaboration

*What is said here, may be found elsewhere...*

*– Mahābhārata* (Canto 1: *Ādi Parva*)

The results in this thesis were derived in collaboration with other researchers and some have already been published elsewhere. The discussion in Chapters 2 to 4 is based on joint work with Alistair Sinclair and Mark Thurley [SST14], and with Alistair Sinclair, Daniel Štefankovič, and Yitong Yin [SSY13,SSŠY14]. Chapters 6 to 8 are based on a paper with Alistair Sinclair [SS14]. The alternative proof of Theorem 7.3 in Section 7.6 arose in joint work with Mario Szegedy. I express my gratitude towards all my collaborators for allowing the inclusion of their co-authored work in this thesis.

# Part I

# Correlation decay and approximate counting

# Chapter 2

# Approximate counting via correlation decay: the Weitz paradigm

In this chapter we describe Weitz's framework [Wei06] for designing approximation algorithms for the partition function of spin systems. The roots of the method go back to a beautiful result of Godsil [God81] for the monomer-dimer model, which we will discuss first. The next chapter will then give a simple example of Weitz's method to the approximation of the partition function of the anti-ferromagnetic Ising model on bounded degree graphs under an optimal range of parameters, while the following chapter will be devoted to extending these techniques to graphs of bounded *connective constant*, which form a strict superset of the class of bounded degree graphs.

## 2.1   From graphs to trees

Well known "self-reducibility" arguments (see, e.g., Section 2.3) imply that it is sufficient to obtain an FPTAS for appropriate marginals of Gibbs distributions in order to obtain an FPTAS for the partition function itself. Beautiful results of Godsil [God81] and Weitz [Wei06] show that the computations of such marginals on general graphs can in turn be transferred to the computation of a dynamic programming recurrence on the tree of *self-avoiding walks* of the graph. Before describing these results and their possible algorithmic consequences, we set up some notation for recurrences on trees.

For a vertex $v$ in a tree $T$, we will denote by $|v|$ the distance of $v$ from the root of the tree. Similarly, for a set $S$ of vertices, $\delta_S := \min_{v \in S} |v|$.

**Definition 2.1 (Cutset).** Let $T$ be any tree rooted at $\rho$. A *cutset* $C$ is a set of vertices in $T$ satisfying the following two conditions:

1. Any path from $\rho$ to a leaf $v$ with $|v| \geq \delta_C$ must pass through $C$.

2. The vertices in $C$ form an antichain, i.e., for any vertices $u$ and $v$ in $C$, neither vertex is an ancestor of the other in $T$.

A trivial example of a cutset is the set $L$ of all the leaves of $T$. Another example we will often need is the set $S_\ell$ of all vertices at distance $\ell$ from $\rho$ in $T$.

For a cutset $C$, we denote by $T_{\leq C}$ the subtree of $T$ obtained by removing the descendants of vertices in $C$ from $T$, and by $T_{<C}$ the subtree of $T$ obtained by removing the vertices in $C$ from $T_{\leq C}$. Further, for a vertex $u$ in $T$, we denote by $T_u$ the subtree of $T$ rooted at $u$, and by $T_{u,\leq C}$ and $T_{u,<C}$ the intersections of $T_u$ with $T_{\leq C}$ and $T_{<C}$ respectively.

**Definition 2.2 (Initial condition).** An *initial condition* $\sigma = (S, P)$ is a set $S$ of vertices in $T$ along with an assignment $P : S \to [0, b]$ of bounded positive values to vertices in $S$.

We are now ready to describe the tree recurrences. Given an initial condition $\sigma = (S, P)$ along with a default value $b_0$ for the leaves, a family of functions $f_d : [0, b]^d \to [0, b]$ for every positive integer $d \geq 1$, and a vertex $u$ in $T$, we let $F_u(\sigma)$ denote the value obtained at $u$ by iterating the tree recurrences $f$ on the subtree $T_u$ rooted at $u$ under the initial condition $\sigma$. Formally, we define $F_u(\sigma) = b_0$ when $u \notin S$ is a leaf, and

$$
F_u(\sigma) = \begin{cases} P(u) & \text{when } u \in S, \\ & \text{when } u \notin S \text{ is of arity} \\ f_d\left(F_{u_1}(\sigma), \ldots, F_{u_d}(\sigma)\right) & d \geq 1 \text{ and has children} \\ & u_1, u_2, \ldots, u_d. \end{cases} \tag{2.1}
$$

### 2.1.1 The self-avoiding walk tree and associated recurrences

Given a vertex $v$ in a graph $G$, one can define a rooted tree $T_{SAW}(v, G)$ of self-avoiding walks (called the *self-avoiding walk tree*, or *SAW tree*) starting at $v$, as follows: the root of the tree represents the trivial self-avoiding walk that ends at $v$, and given any node $u$ in the tree, its children represent all possible self-avoiding walks than can be obtained by extending the self-avoiding walk represented by $u$ by exactly one step. The importance of the self-avoiding walk tree for computation stems from the beautiful results of Godsil [God81] (for the monomer-dimer model) and Weitz [Wei06] (for the hard core model and other two-spin systems) referred to above, which allow the derivation of simple recurrences for the monomer probability $p_v(G)$ on general graphs. We begin with the case of the monomer-dimer model.

#### 2.1.1.1 Monomer-dimer model

**Theorem 2.1 (Godsil [God81]).** *Let $v$ be a vertex in a graph $G$, and consider the monomer-dimer model with dimer activity $\gamma > 0$ on the graphs $G$ and $T_{SAW}(v, G)$. We then have*

$$
p_v(G) = p_v(T_{SAW}(v, G)).
$$

The promised recurrence for $p_v(G)$ can now be derived using dynamic programming on the tree $T_{SAW}(v, G)$. In particular, let $T$ be any tree rooted at $\rho$, and let $\rho_i$, $1 \leq i \leq d$ be the children

of $\rho$. Denoting by $p_i$ the monomer probability $p_{\rho_i}(T_{\rho_i})$ at the root of the subtree $T_{\rho_i}$, one can then show that (see, e.g., [KK98])

$$p_\rho(T) = f_{d,\gamma}(p_1, p_2, \ldots, p_d) := \frac{1}{1 + \gamma \sum_{i=1}^d p_i}. \tag{2.2}$$

**Remark 2.1.** In what follows, we will often suppress the dependence of $f_{d,\gamma}$ on $\gamma$ for convenience of notation.

In terms of our notation for tree recurrences, we note that the actual computation of $p_\rho(T)$ corresponds to computing $F_\rho(\mathbf{1}_L)$, where the initial condition $\mathbf{1}_L$ assigns the value $1$ to all vertices in $L$, the cutset comprising all the leaves (and with the boundary value $b_0$ set to $1$), since the base case of the recurrence comprises a single vertex which has monomer probability $1$ by definition.

Note that the self-avoiding walk tree can be of exponential size, so that Godsil's reduction does not immediately yield an efficient algorithm for computing $p_\rho(G)$. In order to obtain an algorithm, we would need to consider truncated versions of the recurrence, obtained by specifying initial conditions on the cutset $S_\ell$ comprising all vertices at distance $\ell$ from $\rho$. Since $f_{d,\gamma}$ is monotonically decreasing in each of its arguments, we have

$$\begin{aligned}
F_\rho(\mathbf{0}_\ell) \leq p_\rho(T) \leq F_\rho(\mathbf{1}_\ell) \quad \text{when } \ell \text{ is even, and} \\
F_\rho(\mathbf{0}_\ell) \geq p_\rho(T) \geq F_\rho(\mathbf{1}_\ell) \quad \text{when } \ell \text{ is odd.}
\end{aligned} \tag{2.3}$$

Here, the initial condition $\mathbf{0}_\ell$ (respectively, $\mathbf{1}_\ell$) assigns the value $0$ (respectively, $1$) to every vertex in $S_\ell$. Given these conditions, it is sufficient to show that the difference between $F_\rho(\mathbf{0}_\ell)$ and $F_\rho(\mathbf{1}_\ell)$ decreases exponentially in $\ell$ in order to establish that truncated versions of the recurrence converge to the true answer $p_\rho(T)$ exponentially fast in the "truncation length" $\ell$.

### 2.1.1.2  Hard core model

Weitz [Wei06] proved a reduction similar to that of Godsil for the hard core model. However, in contrast to Godsil's reduction for the monomer-dimer model Weitz's reduction requires a boundary condition to be applied to the self avoiding walk tree.

**Theorem 2.2 (Weitz [Wei06]).** *Let $v$ be a vertex in a graph $G$, and consider the hard core model with vertex activity $\lambda > 0$ on the graphs $G$ and $T_{SAW}(v, G)$. Then, there exists an efficiently computable boundary condition $\mathcal{W}$ on $T_{SAW}(v, G)$ such that for any boundary condition $\sigma$ on $G$, we have*

$$R_v(\sigma, G) = R_v(\mathcal{W} \cup \sigma, T_{SAW}(v, G)), \tag{2.4}$$

*where (1) the boundary condition $\sigma$ on the right hand side denotes the natural translation of the boundary condition $\sigma$ on $G$ to $T_{SAW}(v, G)$, and (2) $\mathcal{W} \cup \sigma$ is the boundary condition obtained by first applying the boundary condition $\mathcal{W}$, and then $\sigma$ (that is, $\sigma$ overrides $\mathcal{W}$ on vertices on which $\sigma$ and $\mathcal{W}$ are both specified and disagree).*

We will often refer to a self-avoiding walk tree with Weitz's boundary condition as a "Weitz SAW tree".

**Remark 2.2.** Although we stated Weitz's result for the special case of the hard core model, it actually holds for *any* two-spin system. In fact, even the boundary conditions $\mathcal{W}$ do not depend upon the details of the interactions defining the spin system. In addition, Weitz's construction also allows for a significant amount of freedom in the choice of the boundary condition $\mathcal{W}$. The details of the construction itself are not important for our technical development at this stage, but will turn out to be advantageous when we tackle the problem of establishing decay of correlations for the hard core model on $\mathbb{Z}^2$ (see Section 4.5).

As in the case of the monomer-dimer model, the theorem allows the computation of $R_v(\sigma, G)$ using natural recurrences on the tree. Using the same notation as in the case of the monomer-dimer model, we denote $R_{\rho_i}(\sigma, T_{\rho_i})$ as $R_i$. It is well known (see, e.g., [Wei06]) that

$$R_\rho(\sigma, T) = f_{d,\lambda}(R_1, R_2, \ldots, R_d) := \lambda \prod_{i=1}^d \frac{1}{1 + R_i}. \qquad (2.5)$$

We now see that in terms of our notation for tree recurrences, the computation of $R_\rho(\sigma, T)$ corresponds to computing $F_\rho(\boldsymbol{\lambda}_L \cup \sigma)$ (with the boundary value $b_0$ for leaves set to $\lambda$), where the initial condition $\boldsymbol{\lambda}_L \cup \sigma$ assigns the value $\lambda$ to all vertices in the set $L$ of leaves, and then applies the boundary condition $\sigma$ (overriding previously assigned values). Note that the boundary condition $\sigma$ assigns $R_v = \infty$ for vertices $v$ which are set to occupied by $\sigma$, and hence, strictly speaking, violates the requirement that initial conditions should only assign bounded values. However, this can be fixed easily by observing that an initial condition which assigns $R_v = \infty$ is equivalent to one which assigns $R_u = 0$ to the parent $u$ of $v$. Thus, we may assume without loss of generality that our initial conditions only assign values from the interval $[0, \lambda]$.

Again, as in the case of the monomer-dimer model, we will need to work with truncated trees. As before, we consider initial condition specified on cutsets $S_\ell$ of vertices at distance $\ell$ from the root $\rho$, and use the fact that $f_{d,\lambda}$ is monotonically decreasing in each of its arguments to see that

$$\begin{aligned} F_\rho(\mathbf{0}_\ell \cup \sigma) \le R_\rho(\sigma, T) \le F_\rho(\boldsymbol{\lambda}_\ell \cup \sigma) \quad \text{when } \ell \text{ is even, and} \\ F_\rho(\mathbf{0}_\ell \cup \sigma) \ge R_\rho(\sigma, T) \ge F_\rho(\boldsymbol{\lambda}_\ell \cup \sigma) \quad \text{when } \ell \text{ is odd.} \end{aligned} \qquad (2.6)$$

Here, the initial condition $\mathbf{0}_\ell \cup \sigma$ (respectively, $\boldsymbol{\lambda}_\ell \cup \sigma$) assigns the value $0$ (respectively, $\lambda$) to every vertex in $S_\ell$, after which the boundary condition $\sigma$ is applied, possibly overriding the earlier assignments (from the previous discussion, we can assume that the effect of $\sigma$ is limited o setting some more vertices to $0$). As before, it is then sufficient to show that the difference between $F_\rho(\mathbf{0}_\ell \cup \sigma)$ and $F_\rho(\mathbf{1}_\ell \cup \sigma)$ decreases exponentially in $\ell$ in order to establish that truncated versions of the recurrence converge to the true answer $R_\rho(\sigma, T)$ exponentially fast in the "truncation length" $\ell$.

### 2.1.1.3 Anti-ferromagnetic Ising model

As discussed above, Weitz's reduction in Theorem 2.2 works for any two spin system. We consider here the case of the ant-ferromagnetic Ising model, which is developed in greater detail in the next chapter. Let $R_\rho(\sigma, T)$ denote the ratio between the probabilities that the root $\rho$ of a tree $T$ is assigned spin "+" or "−" in a configuration sampled according to the anti-ferromagnetic Ising model with edge activity $\beta$ and vertex activity $\lambda$. We then have the following analog of the recurrence in eq. (2.5):

$$R_v = f_{d,\beta,\lambda}(q_1, q_2, \ldots, q_d) := \lambda \prod_{i=1}^{d} \frac{R_i + \beta}{\beta R_i + 1}, \tag{2.7}$$

where $R_i$ denotes the corresponding ratios $R_{\rho_i}(\sigma, T_i)$ in the subtrees $T_i$ rooted at the children $\rho_i$ of $\rho$. As in the case of the hard core model, the computation of $R_\rho(\sigma, T)$ again corresponds to computing $F_\rho(\boldsymbol{\lambda}_L \cup \sigma)$ (according to the recurrences defined in eq. (2.7)). Further, recall that for the anti-ferromagnetic Ising model, we have $\beta \geq 1$; so that $f_{d,\beta,\lambda}$ is monotonically decreasing in each of its arguments, so that we again have

$$\begin{aligned}
F_\rho(\mathbf{0}_\ell \cup \sigma) \leq R_\rho(\sigma, T) \leq F_\rho(\boldsymbol{\lambda}_\ell \cup \sigma) \quad \text{when } \ell \text{ is even, and} \\
F_\rho(\mathbf{0}_\ell \cup \sigma) \geq R_\rho(\sigma, T) \geq F_\rho(\boldsymbol{\lambda}_\ell \cup \sigma) \quad \text{when } \ell \text{ is odd,}
\end{aligned} \tag{2.8}$$

so that it will again be sufficient to show that the difference between $F_\rho(\boldsymbol{\lambda}_\ell \cup \sigma)$ and $F_\rho(\mathbf{0}_\ell \cup \sigma)$ decreases exponentially in $\ell$ in order to obtain an algorithm for approximating $R_\rho(\sigma, T)$.

## 2.2 Correlation decay on trees

In all three examples we considered above, we concluded that what needs to be established is an exponential rate of decay (as a function of $\ell$) of the difference $|F_\rho(\sigma) - F_\rho(\tau)|$, where $\sigma$ and $\tau$ are boundary conditions that differ *only* at a distance $\ell$ from the root $\rho$. For various spin systems, such a property has been studied extensively in the literature, and is referred to as *strong spatial mixing* [MO94a, MO94b, Wei06]. For concreteness, we give a formal definition here for the hard core model; the definitions for the anti-ferromagnetic Ising model and the monomer-dimer model are exactly analogous. Our definition here closely follows the version used by Weitz [Wei06].

**Definition 2.3 (Strong Spatial Mixing).** The hard core model with a fixed vertex activity $\lambda > 0$ is said to exhibit *strong spatial mixing* on a family $\mathcal{F}$ of graphs if for any graph $G$ in $\mathcal{F}$, any vertex $v$ in $G$, and any two boundary conditions $\sigma$ and $\tau$ on $G$ which differ only at a distance of at least $\ell$ from $v$, we have

$$|R_v(\sigma, G) - R_v(\tau, G)| = O(c^\ell). \tag{2.9}$$

for some fixed constant $0 \leq c < 1$, which is called the *rate* of strong spatial mixing.

Thus, for all the above models our goal is to establish conditions under which strong spatial mixing holds on the self-avoiding tree. Further, it is easy to see that for all these models, strong spatial mixing is *subgraph-monotone*: i.e., if it holds for the infinite $d$-ary tree, then it holds also for the family of its subtrees, so that, in particular, it holds for the self-avoiding walk trees of graphs of degree at most $d + 1$.

For all three models, but specially for the hard core and the anti-ferromagnetic Ising models, it turns out to be fruitful to study a weaker notion of decay of correlations called *weak spatial mixing* on the infinite $d$-ary tree. Unlike the case of strong spatial mixing, where an exponential decay of the form given by eq. (2.9) is required to hold for pairs of boundary conditions only to be the same at vertices that are at distance less than $\ell$ from the root, weak spatial mixing only requires the condition in eq. (2.9) to hold for pairs of boundary conditions which only apply to vertices in the cutset $S_\ell$ of vertices at distance $\ell$ from the root.



(a) Unstable fixed point: $|f'(x)| > 1$

(b) Stable fixed point: $|f'(x)| < 1$
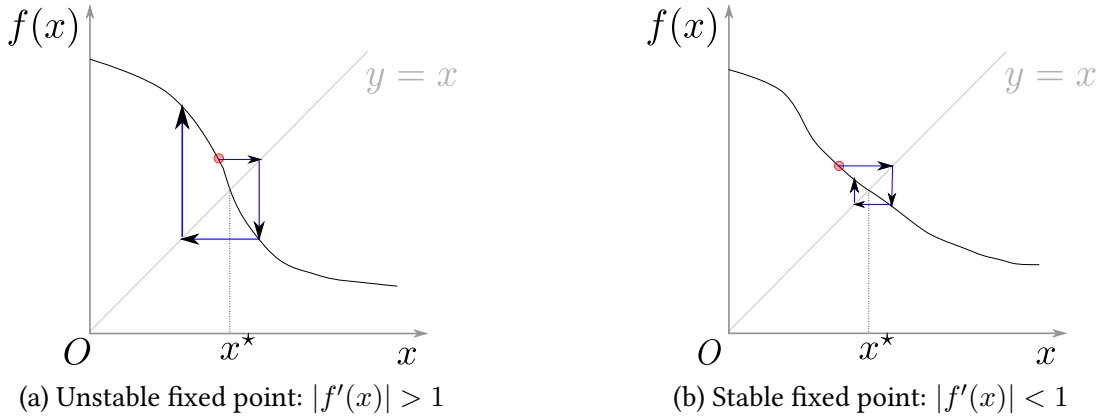
Figure 2.1: Slope at fixed point and spatial mixing

The fact that all the recurrences we consider are monotone decreasing thus allows weak spatial mixing on the infinite $d$-ary tree to be expressed far more succinctly as

$$\left| f_d^\ell(0) - f_d^\ell(\lambda) \right| = O(c^\ell) \text{ for } 0 \le c < 1, \tag{2.10}$$

where $f_d(x)$ denotes $f_d(x, x, \ldots, x)$, the natural univariate version of our multivariate recurrences $f_d$. Equation (2.10) naturally translates into a condition about the convergence of a fixed point computation as follows. Since $f_d$ is a decreasing map from $[0, \lambda]$ to itself, it has a unique fixed point, and eq. (2.10) holds only if one can converge to the fixed point by simply iterating $f_d$ starting at one of the two extremities of the interval. We are thus led to the following observation: a *necessary* condition for weak spatial mixing (and hence strong spatial mixing) to hold on the $d$-ary tree is that the unique fixed point $x^\star$ of $f_d$ must be *stable*; i.e., we must have

$$|f_d'(x^\star)| < 1, \tag{2.11}$$

and more generally, the best rate of strong spatial mixing we can expect on a $d$-ary tree is $|f_d'(x^\star)|$. In the next two chapters, we give several examples where one can indeed exhibit strong optimal

mixing with such an optimal rate. The first such result was by Weitz [Wei06], and concerned the hard core model on bounded degree graphs. As discussed earlier, a consequence of such results is an FPTAS for the partition function under the condition that weak spatial mixing holds on an appropriate $d$-ary tree. As we shall see later, the latter results turn out to be optimal in several important settings, in the sense that unless NP=RP, an FPTAS for the partition function cannot exist when weak spatial mixing does not hold on the same tree [Sly10, SS12, GGŠ+14].

## 2.3   From marginals to the partition function

We conclude this chapter by recalling some standard arguments for translating approximation algorithms for the marginal probabilities into approximation algorithms for the partition function (see, e.g, [Wei06, GK07]). We provide the calculations here for the case of the monomer-dimer model, and refer to Weitz [Wei06] for similar calculations for the hard core model.

Let $v_1, v_2, \ldots, v_n$ be any arbitrary ordering of the vertices of $G$. Since the monomer-dimer partition function of the empty graph is $1$, we then have

$$
\begin{aligned}
Z(G) &= \prod_{i=1}^{n} \frac{Z\left(G - \{v_1, \ldots, v_{i-1}\}\right)}{Z\left(G - \{v_1, \ldots, v_i\}\right)} \\
&= \prod_{i=1}^{n} \frac{1}{p_{v_i}\left(G - \{v_1, \ldots, v_{i-1}\}\right)}.
\end{aligned}
\tag{2.12}
$$

Suppose, we have an FPTAS for the probabilities $p_\rho$ which runs in time $t(n, 1/\epsilon)$ and produces an output $\hat{p}$ such that $p_\rho/(1+\epsilon) \leq \hat{p} \leq p_\rho$. Now, given $\epsilon \leq 1$, we use the FPTAS in time $t\left(n, 2n/\epsilon\right)$ to compute an approximation $\hat{p}_i$ to the $p_{v_i}\left(G - \{v_1, \ldots, v_{i-1}\}\right)$. We then have for each $i$

$$
\frac{1}{p_{v_i}\left(G - \{v_1, \ldots, v_{i-1}\}\right)} \leq \frac{1}{\hat{p}_i} \leq \frac{1 + \epsilon/(2n)}{p_{v_i}\left(G - \{v_1, \ldots, v_{i-1}\}\right)}.
$$

By multiplying these estimates. we obtain an estimate $\hat{Z}$ of the partition function which satisfies

$$
Z(G) \leq \hat{Z} \leq Z(G)\left(1 + \frac{\epsilon}{2n}\right)^n \leq Z(G)e^{\epsilon/2} \leq Z(G)(1+\epsilon),
$$

where we use the condition $\epsilon \leq 1$ in the last inequality. Thus, the total running time is $O\left(n \cdot t\left(n, 2n/\epsilon\right)\right)$, which is polynomial in $n$ and $1/\epsilon$ whenever $t$ is. Thus, it is sufficient to derive an FPTAS for the marginal probabilities in order to obtain an FPTAS for the partition function.

# Chapter 3

# Correlation decay on bounded degree graphs

In this chapter, we apply the framework introduced in the previous chapter to study the anti-ferromagnetic Ising model on bounded degree graphs: this is one of the simplest settings where weak spatial mixing can be used to deduce strong spatial mixing, and hence an FPTAS. We start with a concrete description of the threshold for weak spatial mixing of the anti-ferromagnetic Ising model on the infinite $d$-ary tree.

## 3.1 Weak spatial mixing for the anti-ferromagnetic Ising model

We first recall the standard result that the anti-ferromagnetic Ising model exhibits a phase transition with respect to weak spatial mixing on the $d$-ary tree. This phase transition is also referred to in the mathematical physics literature as the *uniqueness* phase transition, for reasons that we now describe. Recall that we defined the Gibbs measure for a finite graph $G = (V, E)$ in terms of a collection of *weights* $w(\sigma)$ over the $2^{|V|}$ configurations. However, this construction is clearly not valid for infinite graphs such as the infinite $d$-ary tree. For an infinite graph $\mathcal{G}$, we say that a measure $\mu$ is a Gibbs measure if the marginal distribution under $\mu$ on any finite subgraph $\mathcal{H}$, conditional on the configuration on $\mathcal{G} \backslash \mathcal{H}$, is given by the weights $w(\sigma)$. (Here the spins in $\mathcal{G} \backslash \mathcal{H}$ can be seen to act as a fixed *boundary condition* when determining the weights.) It is a well known result in the statistical physics literature (see, for example, [Geo88]) that at least one such measure $\mu$ can always be defined. However, for certain values of the parameters of the spin system there may be multiple solutions for $\mu$, in which case the Gibbs measure is said to be *non-unique.* On the $d$-ary tree, weak spatial mixing is essentially the same as uniqueness: the only caveat being that it is possible that uniqueness may hold even when the rate of decay in weak spatial mixing is merely exponential but not polynomial. However, the latter phenomenon occurs only at the boundary of the region of the parameters in which uniqueness holds, and we shall sidestep it by
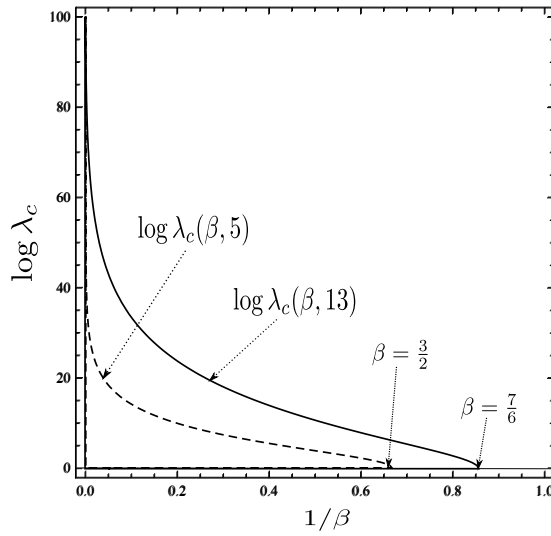
Figure 3.1: $\log \lambda_c(\beta, d)$ for $d = 5$ and $d = 13$. The curves intersect the $1/\beta$-axis at $\beta = \frac{3}{2}$ and $\beta = \frac{7}{6}$ respectively.

considering only the interior of the region. We then have the following results, which provide a description of the uniqueness region.

**Theorem 3.1 (Existence of a critical activity [Geo88, p. 254]).** *Consider the anti-ferromagnetic Ising model on an infinite $d$-ary tree with edge activity $\beta$ and vertex activity $\lambda$. If $1 \leq \beta \leq \frac{d+1}{d-1}$ then the Gibbs measure is unique for all values of $\lambda$. If $\beta > \frac{d+1}{d-1}$, then there exists a critical activity $\lambda_c(\beta, d) \geq 1$ such that the Gibbs measure is unique if and only if $|\log \lambda| \geq \log \lambda_c(\beta, d)$.*

The next result characterizes the relationship between the behavior of tree recurrences $f_d$ and the critical activity, and is basically a formalization of our informal discussion toward the end of Chapter 2.

**Theorem 3.2 ([Geo88]).** *For given values of $\beta$ and $\lambda$, the infinite $d$-ary tree has a unique Gibbs measure if and only if the two-step recurrence function $f \circ f$ has a unique fixed point. In particular, if the Gibbs measure is unique, and $(\beta, \lambda)$ are not on the boundary of the uniqueness region, then the unique fixed point $x^\star$ of $f$ satisfies*

$$|f'(x^\star)| < 1. \tag{3.1}$$

**Remark 3.1.** In [Geo88], it is claimed (implicitly) on the basis of numerical simulations that the condition (3.1) is also *sufficient* for uniqueness. To be precise, the expression for the critical activity $\lambda_c(\beta, d)$ given in [Geo88, p. 255] is exactly the same as that obtained by assuming that (3.1) is also a sufficient condition for uniqueness. While we believe this fact to be folklore, we have not been able to find a rigorous proof of it in the literature. With a slight abuse of terminology, we will henceforth refer to the set of $(\beta, \lambda)$ for which the fixed point $x^\star$ satisfies $|f'(x^\star)| < 1$ as the

"uniqueness region". We will justify this terminology later (see the Remark following the proof of Theorem 3.8) by proving that condition (3.1) does indeed imply uniqueness.

As discussed earlier, the remaining obstacle in going from weak spatial mixing to strong spatial mixing is the fact that strong spatial mixing requires an exponential decay of correlations to hold under arbitrary boundary conditions. In particular, strong spatial mixing will follow if we can show that there is a $c < 1$ such that for *any* two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ in the domain of $f_d$, we have

$$|f_d(\boldsymbol{x}) - f_d(\boldsymbol{y})| \leq c \cdot \|\boldsymbol{x} - \boldsymbol{y}\|_\infty, \tag{3.2}$$

which translates into the requirement that $\|\nabla f_d(\boldsymbol{x})\|_1 \leq c$ *everywhere* in the domain. However, weak spatial mixing guarantees this condition only at the point $\boldsymbol{x}^\star = (x^\star, x^\star, \dots, x^\star)$, where $x^\star$ is the unique fixed point of the (univariate) version of $f_d$. The solution to this obstacle is to look at recurrences for quantities other than just the occupation probabilities and occupation ratios. In the next section, we formalize this idea in terms of the notion of a *message*, that was first used in this context by Restrepo, Shin, Tetali, Vigoda and Yang [RST+13].

## 3.2 Messages for strong spatial mixing

**Definition 3.1 (Message).** A *message* is a strictly increasing continuously differentiable function $\phi : [0, b] \to \mathbb{R}$ whose derivative is bounded away from 0 on its domain.

Note that, being strictly increasing, a message is invertible on its range. Moreover, the inverse function $\phi^{-1}$, which we will denote by $\psi$, is also a continuously differentiable function with positive derivative.

Given a message $\phi$ and a set of recurrences $\{f_d\}_{d \geq 1}$ as described above, we define the functions $f_d^\phi$ as

$$f_d^\phi(x_1, x_2, \dots, x_d) := \phi\left(f_d\left(\psi(x_1), \psi(x_2), \dots, \psi(x_d)\right)\right).$$

The above definition clearly extends to the univariate version of the recurrences as well. A simple but important property of messages is that they do not change the derivative at the fixed point of the recurrence, so that our discussion of weak spatial mixing in terms of the stability of the fixed point remains valid even for arbitrary messages. We isolate this observation in the following lemma.

**Fact 3.3.** *For any message $\phi$, the parameters $(\beta, \lambda)$ are in the uniqueness region if and only if $|g'(p^\star)| < 1$ at the unique fixed point $p^\star$ of $g = f^\phi$.*

*Proof.* Notice that since $\phi$ is strictly increasing, and $f$ has a unique fixed point $x^\star$, $g = f^\phi$ also has a unique fixed point $p^\star = \phi(x^\star)$. Now, we notice that $g'(p^\star) = f'(x^\star)$, because

$$
\begin{aligned}
g'(p^\star) &= \phi'(f(\psi(p^\star)))f'(\psi(p^\star))\psi'(p^\star) \\
&= \phi'(f(x^\star))f'(x^\star)\frac{1}{\phi'(\psi(p^\star))} \\
&= \frac{\phi'(x^\star)}{\phi'(x^\star)}f'(x^\star) \\
&= f'(x^\star),
\end{aligned}
$$

where in the second line we used the facts that $\psi(p^\star) = x^\star$ and $\psi'(y) = \frac{1}{\phi'(\psi(y))}$, and in the third line the fact that $f(x^\star) = x^\star$. Thus, $(\beta, \lambda)$ are in the uniqueness region (as defined in the Remark following Theorem 3.2) if and only if $|g'(p^\star)| = |f'(x^\star)| < 1$. $\qquad\square$

Our strategy for showing strong spatial mixing is now to choose a message $\phi$ in such a way that we can enforce

$$\|\nabla g(\boldsymbol{x})\|_1 \leq g'(p^\star), \tag{3.3}$$

where $g$ denotes the function $f_d^\phi$ and $p^\star$ is the unique fixed point of $g$. This will establish strong spatial mixing via eq. (3.2). As a first step towards this goal, we will prove a univariate version of eq. (3.3). Before proceeding further, we make a slight change in notation. We will assume in the rest of this chapter that the recurrences $f_d$ are defined not for the occupation ratio $R_v$ but for the "inoccupation probability" $q_v$, which is defined as the probability of a vertex being assigned to be "$-$". (Thus, $R_v = \frac{1-q_v}{q_v}$). This is tantamount to modifying the original $f_d$ using a transformation $x \mapsto 1/(1+x)$. This transformation can also be thought of a (very simple) example of a message, and hence in light of Fact 3.3, all our discussions above pertaining to stability near the fixed point remain equally valid for this new set of recurrences, which we now describe in detail. We first define the function

$$h(x) \triangleq \frac{1 + (\beta - 1)x}{\beta - (\beta - 1)x}.$$

We can then write the recurrence as

$$f_d(q_{v_1}, q_{v_2}, \ldots, q_{v_d}) \triangleq \frac{1}{1 + \lambda \prod_{i=1}^d h(q_{v_i})}.$$

We will now prove that one can indeed choose a message which maximizes the slope at the fixed point.

**Theorem 3.4.** *Given $d, \beta$ and $\lambda$, there exists a message $\phi$ and a constant $c < 1$, such that the tree recurrence $g = f_d^\phi$ for the quantity $\phi(p_v)$ satisfies $|g'(x)| \leq c < 1$ for all $x$ in the domain of $g$, whenever $(\beta, \lambda)$ is in the uniqueness region for the $d$-ary tree.*

In light of Fact 3.3, the main technical challenge in the proof of the theorem is to come up with a message $\phi$ such that the quantity $|g'|$ is maximized at the unique fixed point of $g = f^\phi$. We now proceed to construct this message. We fix constants

$$A = d(\beta^2 - 1) + (1 - \beta)^2 \text{ and } D = \frac{\sqrt{A + 4\beta} - \sqrt{A}}{2\sqrt{A}},$$

and choose

$$\phi(x) = \log\left(\frac{x + D}{1 - x + D}\right).$$

Notice that $D > 0$, so $\phi$ is a continuously differentiable function with positive derivative on the interval $[0, 1]$. Using this message we are able to prove the following lemma (recall that $\psi$ denotes $\phi^{-1}$).

**Lemma 3.5.** *Consider the anti-ferromagnetic Ising model on a $d$-ary tree with edge activity $\beta$ and vertex activity $\lambda$. Then, using the shorthand notations $\alpha$ and $\eta$ for $\psi(x)$ and $f(\psi(x))$ respectively, we have*

$$\begin{aligned}
g''(x) = (\eta - \alpha)g'(x)\psi'(x) \\
\times \frac{d\beta(\beta^2 - 1)(2\beta + A\alpha\eta + A(1 - \alpha)(1 - \eta))}{(\beta + \alpha(1 - \alpha)(1 - \beta)^2)(\beta + A\eta(1 - \eta))(\beta + A\alpha(1 - \alpha))}.
\end{aligned} \quad (3.4)$$

The proof of Lemma 3.5 is somewhat technical and is deferred to Section 3.3. Note that the requirement that the derivative of the function $g = f_d^\phi$ should have its maximum magnitude at the unique fixed point of $g$ does not immediately lead to a solution for $\phi$, and thus some educated guesswork is needed for fixing the functional form of $\phi$. Our choice is guided by the intuition that, by analogy with the zero field case, where it is well known that the simple message $\phi(x) = \log\left(\frac{x}{1-x}\right)$ is sufficient, a log ratio of probabilities shifted by an additive constant $D$ to account for the field should be appropriate. The choice of $D$ is then determined by the above requirement.

Given Lemma 3.5, we can now show that $g$ indeed has its worst possible slope at its fixed point.

**Lemma 3.6.** *Let $g = f^\phi$, with the message $\phi$ defined above. Then $|g'(x)|$ is maximized at the unique positive fixed point of $g$.*

*Proof.* We simply need to use the expression for $g''(x)$ defined in Lemma 3.5. It is easy to see that, ignoring the factor $(\eta - \alpha)$, the rest of the right hand side of eq. (3.4) is negative: this is because $g$ is a decreasing function, while $\psi$, being the inverse of the increasing function $\phi(x)$, is increasing. Also, we have $\beta > 1$ (in the anti-ferromagnetic case) and $0 \le \alpha, \eta \le 1$ (since they are probabilities), so that the fractions appearing on the right hand side are positive.

Let $x^\star$ be the unique fixed point of the strictly decreasing function $g$. From the above discussion, it follows that the sign of $g''(x)$ is the opposite of the sign of $\eta - \alpha = f(\psi(x)) - \psi(x)$. Notice

that $\eta - \alpha$ is strictly positive for $x < x^\star$ and strictly negative for $x > x^\star$. This implies that $g'(x)$ is strictly decreasing for $x < x^\star$ and strictly increasing for $x > x^\star$. Since $g$ is strictly decreasing this shows that the magnitude of $g'$ is maximized at $x^\star$. □

Combining Lemma 3.6 with Fact 3.3, we immediately get Theorem 3.4. Lemma 3.6 further implies that the constant $c$ in the Theorem is $|g'(x^\star)|$, where $x^\star$ is the unique fixed point of $g$; this is what we sought in our earlier discussion. The remaining step in proving strong spatial mixing is to show that the univariate case is indeed the worst case, which we do in the following lemma.

**Lemma 3.7.** *Let $\eta = \psi(g(\boldsymbol{x}))$. Let $\bar{\eta}$ be the unique solution of $\psi(g(\bar{\eta})) = \eta$. Then $\|\nabla\left(g(\boldsymbol{x})\right)\|_1 \leq \|\nabla\left(g(\bar{\eta}, \bar{\eta}, \ldots, \bar{\eta})\right)\|_1 = |g'(\bar{\eta})|$.*

*Proof.* Set $\alpha_i = \psi(x_i)$ for $i = 1, 2, \ldots d$. We then have

$$\eta = \frac{1}{1 + \lambda \prod_{i=1}^{d} h(\alpha_i)} = \frac{1}{1 + \lambda h(\psi(\bar{\eta}))^d}. \tag{3.5}$$

Recalling the definitions of the quantities $A$ and $D$ given just before Lemma 3.5, we can now write $\|\nabla\left(G(\boldsymbol{x})\right)\|_1$ as

$$\|\nabla\left(G(\boldsymbol{x})\right)\|_1 = \frac{d\eta(1-\eta)(\beta^2-1)}{\beta + A\eta(1-\eta)}\left(1 + (\beta^2 - 1)\sum_{i=1}^{d}\frac{\alpha_i(1-\alpha_i)}{\beta + (1-\beta)^2\alpha_i(1-\alpha_i)}\right). \tag{3.6}$$

For notational convenience, we define the function $J(x) \triangleq \frac{x(1-x)}{\beta+(1-\beta)^2 x(1-x)}$. Note that maximizing the sum in (3.6) under the constraint (3.5) is the same as maximizing $\sum_{i=1}^{d} J(\alpha_i)$ under the constraint that $\prod_{i=1}^{d} h(\alpha_i) = \frac{1-\eta}{\lambda\eta}$. Since $h$ is positive and invertible, it is therefore sufficient to show that the function $K(x) \triangleq J(h^{-1}(e^x))$ is concave in order to show that all $\alpha_i$'s are equal at a maximum. We now show this by direct computation. After differentiating twice and simplifying, we have

$$K''(x) = -\frac{e^{-x}(1 + e^{2x})\beta}{(1-\beta^2)^2} < 0.$$

This shows that $K$ is concave, and hence it follows that the sum in equation (3.6) is maximized when all $\alpha_i$'s are equal. In conjunction with the condition that $\eta = \frac{1}{1+\prod_{i=1}^{d} h(\alpha_i)}$, this shows that

$$\|\nabla\left(g(\boldsymbol{x})\right)\|_1 \leq \|\nabla\left(g(\bar{\eta}, \bar{\eta}, \ldots, \bar{\eta})\right)\|_1.$$

Since, by definition, we have $g(x, x, \ldots, x) = g(x)$, we see that $\|\nabla\left(g(\bar{\eta}, \bar{\eta}, \ldots, \bar{\eta})\right)\|_1 = |g'(\bar{\eta})|$. □

Using Lemma 3.6 and the above lemma, we are now ready to prove our main theorems. The first of these establishes the promised equivalence between weak and strong spatial mixing.

**Theorem 3.8.** *For the anti-ferromagnetic Ising model with arbitrary field on the $d$-ary tree with $d \geq 2$, weak spatial mixing implies strong spatial mixing.*

*Proof.* Consider a setting of parameters $\beta$ and $\lambda$ such that the $d$-ary tree has weak spatial mixing. Let $x^\star$ be the unique fixed point of the function $g$. We will use only the property that the fixed point satisfies the condition (3.1) of Theorem 3.2. By Theorem 3.4 we have $\|g'\|_\infty = c < 1$. By Lemma 3.7, this implies that for all $\boldsymbol{x}$ in the domain of the multivariate function $g$, we have $\|\nabla g(\boldsymbol{x})\|_1 \le c$. Using the mean value theorem followed by Hölder's inequality, we then have

$$|g(\boldsymbol{x}) - g(\boldsymbol{y})| \le c\|\boldsymbol{x} - \boldsymbol{y}\|_\infty,$$

for all vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ in the domain of $G$. As noted above, this is sufficient for strong spatial mixing. We provide the details of the proof for completeness. Consider boundary conditions $\sigma_1$ and $\sigma_2$ on a set $S$ which differ only on a subset $T$ which is at distance $l$ from the root $\rho$ of the tree. Again, since both $\phi$ and $\phi^{-1}$ are continuously differentiable functions defined over compact sets, they are Lipschitz continuous, say with parameters $L_1$ and $L_2$ respectively. We define the quantity $q_i$ as

$$q_i \triangleq \max_{v:\delta(\rho,v)=l-i} |\phi(F_v(\sigma_1)) - \phi(F_v(\sigma_2))|.$$

Notice that $q_0 \le |\phi(1) - \phi(0)| \le L_1$. Also, since $g$ is the tree recurrence for $\phi(p_v)$, the above condition on the gradient implies that $q_{i+1} \le cq_i$ for $c < 1$. Thus, we get strong spatial mixing since

$$|F_\rho(\sigma_1) - F_\rho(\sigma_2)| \le L_2 q_l \le L_2 c^l q_0 \le L_1 L_2 c^l.$$

$\square$

**Remark 3.2.** We can now justify our use of the term "uniqueness region" as described in the Remark following Theorem 3.2. Notice that in the proof of Theorem 3.8 above, we used only the fact that weak spatial mixing implies that $(\beta, \lambda)$ is in the "uniqueness region" as defined in the aforementioned Remark. Thus, we see that whenever $(\beta, \lambda)$ is in the uniqueness region, we have strong spatial mixing, and hence, in particular, uniqueness. As stated earlier, this provides a rigorous proof of the claim in [Geo88] that the interior of the uniqueness region is equivalent to the condition (3.1).

Finally we can prove the following algorithmic consequence.

**Corollary 3.9.** *Let $d \ge 2$. Consider an anti-ferromagnetic Ising model with parameters $\beta$ and $\lambda$. For $\beta$ and $\lambda$ in the interior of the uniqueness region of the $d$-ary tree, every graph of degree at most $d + 1$ exhibits strong spatial mixing. Moreover, for such $\beta$ and $\lambda$, there is an FPTAS for the partition function of the associated spin system on graphs of degree at most $d + 1$.*

*Proof.* As observed earlier, in order to obtain an FPTAS for the partition function of the associated spin system, it is sufficient to give an FPTAS for approximating the occupation probability $p_\rho$ of a vertex $\rho$, under an arbitrary fixing of spin values for an arbitrary subset of vertices. Given a vertex $\rho$ in a graph $G$ of maximum degree $(d + 1)$, we start by constructing Weitz's self-avoiding walk (SAW) tree rooted at $\rho$. For non-leaf vertices (apart from $\rho$) in this tree which do not have $d$ children, we can create dummy children (so as to make the arity of the vertex $d$) all of which

independently have occupation probabilities of $1/2$. It is easy to see that this does not change the output of the tree recurrence at any vertex of the tree. As we saw in the proof of Theorem 3.8, we have strong spatial mixing on this SAW tree whenever $(\beta, \lambda)$ are in the uniqueness region of the $d$-ary tree. The corollary now follows using Weitz's reduction, since in order to obtain a $(1 \pm \epsilon/n)$ factor approximation to the $q_v$, we can truncate the tree at a depth of $O(\log(n/\epsilon))$, which takes time polynomial in both $n$ and $1/\epsilon$. $\qquad\square$

The range of validity of the FPTAS in Corollary 3.9 is optimal: it was proved by Sly and Sun [SS12] that an FPTAS on $(d + 1)$-regular graphs for parameters $(\beta, \lambda)$ that are *not* in the uniqueness region of the $d$-ary tree will imply that NP = RP. As pointed out earlier, similar tight connections between the complexity of approximating the partition function and the uniqueness phase transition were first proved for the hard core model [Wei06, Sly10, SS12]. In the next chapter, we go beyond the setting of bounded degree graphs, and extend the message approach discussed in this chapter to the deal with graphs of bounded *connective constant* [MS96, Ham57]. The latter class includes, for example, the class of graphs sampled from the Erdős–Rényi model $\mathcal{G}(n, d/n)$ (for a constant $d$), which are of unbounded degree with high probability.

## 3.3 Proof of Lemma 3.5

We now provide the proof of Lemma 3.5. The proof involves a few somewhat lengthy derivative computations, which we isolate in the following lemma. For ease of notation, we write $f_d$ as $f$.

**Lemma 3.10.** *With the notation used in Lemma 3.5 above, we have*

$$\frac{\phi''(x)}{\phi'(x)} = \frac{A(2x - 1)}{\beta + Ax(1 - x)}; \tag{3.7}$$

$$\frac{h'(x)}{h(x)} = \frac{\beta^2 - 1}{\beta + (1 - \beta)^2 x(1 - x)}; \tag{3.8}$$

$$\frac{h''(x)}{h'(x)} = \frac{2(\beta - 1)}{\beta - (\beta - 1)x}; \tag{3.9}$$

$$f'(x) = -df(x)(1 - f(x))\frac{h'(x)}{h(x)}; \tag{3.10}$$

$$\frac{f''(x)}{f'(x)} = \frac{f'(x)(1 - 2f(x))}{f(x)(1 - f(x))} + \frac{h''(x)}{h'(x)} - \frac{h'(x)}{h(x)}. \tag{3.11}$$

*Proof sketch.* All of these identities are easily verified by direct computation. In proving equation (3.7), one needs to keep in mind the definition of the constant $D$. $\qquad\square$

*Proof of Lemma 3.5.* To ease notation, we will suppress the dependence of the quantities $\eta$ and $\alpha$ on $x$. Using the chain rule, we have

$$g'(x) = \frac{\phi'(\eta)}{\phi'(\alpha)} f'(\alpha).$$

Here, we used the fact that since $\psi = \phi^{-1}$, $\psi'(x) = \frac{1}{\phi'(\psi(x))}$. After taking the logarithm, and noticing that the right hand side is more easily expressed as a function of $\alpha$ rather than of $x$, one can write the second derivative of $g$ as

$$\frac{1}{\psi'(x)} \frac{g''(x)}{g'(x)} = \frac{\phi''(\eta)}{\phi'(\eta)} \frac{\mathrm{d}\eta}{\mathrm{d}\alpha} - \frac{\phi''(\alpha)}{\phi'(\alpha)} + \frac{f''(\alpha)}{f'(\alpha)}. \tag{3.12}$$

We now consider each of the terms involved above. Recalling that $\eta = f(\alpha)$, and using equations (3.10) and (3.11) to expand the first and last terms in equation (3.12) above, we get

$$\frac{1}{\psi'(x)} \frac{g''(x)}{g'(x)} = T_1 - T_2, \tag{3.13}$$

where $T_1$ and $T_2$ are defined as

$$T_1 \triangleq \frac{h''(\alpha)}{h'(\alpha)} - \frac{h'(\alpha)}{h(\alpha)} - \frac{\phi''(\alpha)}{\phi'(\alpha)}, \text{ and}$$

$$T_2 \triangleq d\frac{h'(\alpha)}{h(\alpha)} \left[ \frac{\phi''(\eta)}{\phi'(\eta)} \eta(1-\eta) + 1 - 2\eta \right].$$

Notice that all terms containing $\eta$ are isolated in $T_2$. We now consider each of the terms separately. For $T_1$, we have

$$\frac{h''(\alpha)}{h'(\alpha)} - \frac{h'(\alpha)}{h(\alpha)} = \frac{2(\beta - 1)}{\beta - (\beta - 1)\alpha} - \frac{\beta^2 - 1}{\beta + (1 - \beta)^2\alpha(1 - \alpha)}$$

$$= \frac{(1 - \beta)^2(2\alpha - 1)}{\beta + (1 - \beta)^2\alpha(1 - \alpha)}.$$

Here, we used equations (3.9) and (3.8) in the first line. Now using equation (3.7), we have

$$T_1 = \frac{(2\alpha - 1)\left((1 - \beta)^2\left[\beta + A\alpha(1 - \alpha)\right] - A\left[\beta + (1 - \beta)^2\alpha(1 - \alpha)\right]\right)}{(\beta + (1 - \beta)^2\alpha(1 - \alpha))(\beta + A\alpha(1 - \alpha))}$$

$$= \frac{\beta(2\alpha - 1)((1 - \beta)^2 - A)}{(\beta + (1 - \beta)^2\alpha(1 - \alpha))(\beta + A\alpha(1 - \alpha))}$$

$$= \frac{-d\beta(2\alpha - 1)}{(\beta + A\alpha(1 - \alpha))} \frac{h'(\alpha)}{h(\alpha)}.$$

Here, we use $A = d(\beta^2 - 1) + (1 - \beta)^2$, followed by equation (3.8) in the last line.

We now consider $T_2$. Again using equation (3.7), we have

$$T_2 = d\frac{h'(\alpha)}{h(\alpha)} \left[ \frac{A(2\eta - 1)\eta(1 - \eta)}{\beta + A\eta(1 - \eta)} - (2\eta - 1) \right]$$

$$= \frac{-d\beta(2\eta - 1)}{\beta + A\eta(1 - \eta)} \frac{h'(\alpha)}{h(\alpha)}.$$

Notice that modulo the $\frac{h'(\alpha)}{h(\alpha)}$ factor, $T_1$ and $T_2$ have the same functional form as functions of $\alpha$ and $\eta$ respectively. In fact, the message $\phi$ is designed so as to make this possible. We can now substitute these values into equation (3.13) to get

$$
\begin{aligned}
g''(x) &= d\beta g'(x)\psi'(x)\frac{h'(\alpha)}{h(\alpha)}\left[\frac{2\eta - 1}{\beta + A\eta(1 - \eta)} - \frac{2\alpha - 1}{\beta + A\alpha(1 - \alpha)}\right] \\
&= d\beta g'(x)\psi'(x)\frac{h'(\alpha)}{h(\alpha)}\frac{(\eta - \alpha)(2\beta + A(\alpha\eta + (1 - \alpha)(1 - \eta)))}{(\beta + A\alpha(1 - \alpha))\,(\beta + A\eta(1 - \eta))} \\
&= (\eta - \alpha)g'(x)\psi'(x)\frac{d\beta(\beta^2 - 1)(2\beta + A\alpha\eta + A(1 - \alpha)(1 - \eta))}{(\beta + \alpha(1 - \alpha)(1 - \beta)^2)(\beta + A\eta(1 - \eta))(\beta + A\alpha(1 - \alpha))},
\end{aligned}
$$

where in the last step we used equation (3.8). $\qquad\square$

# Chapter 4

# Correlation decay and the connective constant

The last chapter gave one example of the tight relationship between the uniqueness phase transition on the infinite $d$-ary tree to decay of correlations on graphs of degree at most $d+1$. We now extend the techniques used there to show that these connections extend to the larger class of graphs of *connective constant* at most $d$, which also includes some graphs of unbounded degree. The specific models we will deal with in this chapter will be the hard core model and the monomer-dimer model, but we will first set up a general framework which will apply to both these and other models. We begin with a description of the connective constant.

## 4.1  The connective constant

The connective constant is a well studied notion of the average degree of a graph, which roughly speaking, measures the growth of the number of self-avoiding walks in the graph as a function of their length, and can in fact be quite accurately described as the average arity of the self-avoiding walk tree of the graph. We now present the formal definitions.

**Definition 4.1 (Connective constant: infinite graphs [MS96]).** Let $G = (V, E)$ be a locally finite infinite graph. The *connective constant* $\Delta(G)$ of $G$ is $\sup_{v \in V} \limsup_{\ell \to \infty} N(v, \ell)^{1/\ell}$.

**Remark 4.1.** The supremum over $v$ in the definition is clearly not required for vertex-transitive graphs such as Cartesian lattices. Further, in such graphs the $\limsup$ can be replaced by a limit [MS96].

The connective constant has been especially well studied for various infinite regular lattices, and rigorous upper and lower bounds on its value are known in several cases [MS96, Alm05]. It is a, however, a open problem in most cases to determine its value exactly; an exception being the hexagonal lattice $\mathcal{H}$ for which it was established to be $\sqrt{2 + \sqrt{2}}$ by Duminil-Copin and Smirnov [DCS12]. Note that in general, the best upper bound one can put on the value of the

connective constant of a lattice of maximum degree $d + 1$ is $d$. However, for small degree regular lattices such as $\mathcal{Z}^2$ and $\mathcal{H}$, much better upper bounds can be derived.

For algorithmic applications, we will need a suitable version of the above definition for finite graphs. The following definition, which closely follows the definition for infinite graphs given above, first appeared in [SSY13].

**Definition 4.2 (Connective constant: finite graphs [SSY13]).** Let $\mathcal{F}$ be a family of finite graphs. The connective constant of $\mathcal{F}$ is at most $\Delta$ if there exist constants $a$ and $c$ such that for any graph $G = (V, E)$ in $\mathcal{F}$ and any vertex $v$ in $G$, we have $\sum_{i=1}^{\ell} N(v, i) \leq c\Delta^{\ell}$ for all $\ell \geq a \log |V|$.

As before, it is easy to see that the connective constant of a graph of maximum degree $d + 1$ is at most $d$. However, the connective constant can be much smaller than the maximum degree, and in particular, can be bounded even for unbounded degree graphs. For example, though the maximum degree of a graph drawn from the Erdős–Rényi model $\mathcal{G}(n, d/n)$ is $\Theta(\log n / \log \log n)$ w.h.p, it is not hard to show that for any fixed $\epsilon > 0$, the connective constant of such a graph is at most $d(1 + \epsilon)$ w.h.p.

For both the hard core and monomer-dimer models, we will show that strong spatial mixing with rate $c$ on the infinite $d$-ary tree implies the same rate of strong spatial mixing on graph families of connective constant at most $d$. Our proofs will use the *message framework* discussed in the last chapter, but we will need a more fine-grained analysis of the step-wise decay than that carried out through Lemmas 3.5 and 3.6 in the last chapter. We now proceed to describe this general framework.

## 4.2 Decay of correlations on the SAW tree

In this section, we consider the general problem of proving decay of correlation results for the abstract tree recurrences $F_\rho$ defined in eq. (2.1), and show how the analysis of the message approach can be strengthened beyond that given in the last chapter. We will then instantiate our framework with appropriately chosen messages for the monomer-dimer and the hard core models in Sections 4.3 and 4.4.

We begin by fixing the boundary value $b_0$ for the leaves in our recurrence framework, and assume that the initial conditions specify values in the interval $[0, b]$. We assume that we have a set of tree recurrences $f_d : [0, b]^d \to [0, b]$ for every positive integer $d \geq 1$. The only constraints we put on the recurrences in this section are the following (both of which are trivially satisfied by the recurrences for all the models we have considered).

**Condition 4.1 (Consistency).** *We say that a set of recurrences $\{f_d\}_{d \geq 1}$, where $f_d$ is $d$-variate, are consistent if they obey the following two conditions:*

1. *If $\boldsymbol{x} \in \mathbb{R}^d$ is a permutation of $\boldsymbol{y} \in \mathbb{R}^d$, then $f_d(\boldsymbol{x}) = f_d(\boldsymbol{y})$.*

2. *If all but the first $k$ co-ordinates of $\boldsymbol{x} \in \mathbb{R}^d$ are $0$, then $f_d(\boldsymbol{x}) = f_k(x_1, x_2, x_3, \ldots, x_k)$.*

As before, given a message $\phi$ (and its inverse $\psi$), we define $f_d^\phi$ by

$$f_d^\phi(x_1, x_2, \ldots, x_d) := \phi\left(f_d\left(\psi(x_1), \psi(x_2), \ldots, \psi(x_d)\right)\right).$$

We then have the following simple consequence of the mean value theorem (a proof can be found in Section 4.6).

**Lemma 4.2 (Mean value theorem).** *Consider two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ in $\phi([0, B])^d$. Then there exists a vector $\boldsymbol{z} \in [0, \infty)^d$ such that*

$$\left| f_d^\phi(\boldsymbol{x}) - f_d^\phi(\boldsymbol{y}) \right| \leq \Phi\left(f_d(\boldsymbol{z})\right) \sum_{i=1}^d \frac{|y_i - x_i|}{\Phi(z_i)} \left| \frac{\partial f_d}{\partial z_i} \right|, \tag{4.1}$$

*where $\Phi := \phi'$ is the derivative of $\phi$, and by a slight abuse of notation we denote by $\frac{\partial f_d}{\partial z_i}$ the partial derivative of $f_d(R_1, R_2, \ldots, R_d)$ with respect to $R_i$ evaluated at $\boldsymbol{R} = \boldsymbol{z}$.*

So far our approach has been similar to that taken in the previous chapter (and indeed, in several other papers [RST$^+$13, SST14, LLY12, LLY13]) in that we use an appropriate message—along with the estimate in Lemma 4.2—to argue that the "distance" between two input message vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ at the children of a vertex shrinks by a constant factor at each step of the recurrence. However, in our analysis in the last chapter, we showed such a decay on some version of the $\ell_\infty$ norm of the "error" vector $\boldsymbol{x} - \boldsymbol{y}$: this was achieved by bounding the appropriate dual $\ell_1$ norm of the gradient of the recurrence (see, eq. (3.2) and Lemma 3.6). Our intuition is that in order to achieve a bound in terms of a global quantity such as the connective constant, it should be advantageous to use a more global measure of the error such as an $\ell_q$ norm for some $q < \infty$.

In line with this plan, we will attempt to bound the right hand side of eq. (4.1) in terms of $\|\boldsymbol{x} - \boldsymbol{y}\|_q$ for an appropriate value of $q < \infty$ by maximizing the sum while keeping $f_d(\boldsymbol{z})$ fixed. However, the convexity argument used in the proof of Lemma 3.6 in the last chapter, which showed that the "worst" case of the error is obtained for the univariate version of the recurrence $f_d$, does not work when we look at more general norms of the error vector. (However, as shown in Sinclair *et al.* [SSY13], a similar argument can be made to work also for the case $q = 2$, though this restriction leads to sub-optimal results). In order to get past this limitation of only being allowed certain fixed $\ell_q$ norms of the error vector, we will use a more flexible optimization than that used in the last chapter and in [SSY13]. To do this, we will seek to establish the following property for our messages (the exponent $a$ will be the Hölder conjugate of the value of $q$ that we eventually use).

**Definition 4.3.** Given a consistent family of recurrences $\{f_d\}_{d \geq 1}$, a message $\phi$ (with $\Phi := \phi'$) is said to be *symmetrizable with exponent $a$* with respect to the family if it satisfies the following two conditions:

1. Let $\mathcal{D}$ be the domain of the recurrence family. For every positive integer $d$ and every real $B > 0$ for which the program

$$\max \quad \sum_{i=1}^{d} \left( \frac{1}{\Phi(x_i)} \left| \frac{\partial f_d}{\partial x_i} \right| \right)^a, \qquad \text{where}$$

$$f_d(\boldsymbol{x}) = B$$

$$x_i \in \mathcal{D}, \qquad 1 \le i \le d$$

   is feasible, it also has a solution $\boldsymbol{x}$ in which all the non-zero entries of $\boldsymbol{x}$ are equal. (We assume implicitly that $0 \in \mathcal{D}$.)

2. $\lim_{x_i \to 0^+} \frac{1}{\Phi(x_i)} \left| \frac{\partial f_d}{\partial x_i} \right| = 0$ for all $d \ge 1$, and for any fixed values of the $x_j$, $j \ne i$.

   For symmetrizable messages, we will be able to bound the quantity $|f_d^\phi(\boldsymbol{x}) - f_d^\phi(\boldsymbol{y})|$ in terms of $\|\boldsymbol{x} - \boldsymbol{y}\|_q$, where $1/a + 1/q = 1$, and our improved correlation decay bounds will be based on the fact that symmetrizability can be shown to hold under a wider range of values of $q$ than that required by the concavity arguments we have hitherto discussed. Our bounds will be stated in terms of the following notion of decay (as before, given a $d$-variate function $f_d$ and a scalar $x$, we denote by $f_d(x)$ the quantity $f_d(x, x, \dots, x)$).
.

**Definition 4.4 (Decay factor $\alpha$).** Let $\phi$ be a message with derivative $\Phi$, and let $a$ and $q$ be positive reals such that $\frac{1}{a} + \frac{1}{q} = 1$. We define the functions $\Xi_{\phi,q}(d, x)$ and $\xi_{\phi,q}(d)$ as follows:

$$\Xi_{\phi,q}(d, x) := \frac{1}{d} \left( \frac{\Phi(f_d(x)) \, |f_d'(x)|}{\Phi(x)} \right)^q;$$

$$\xi_{\phi,q}(d) := \sup_{x \ge 0} \Xi_{\phi,q}(d, x).$$

The *decay factor* $\alpha$ is then defined as

$$\alpha := \sup_{d \ge 1} \xi_{\phi,q}(d). \tag{4.2}$$

Armed with the above definitions, we are now ready to prove Lemma 4.3, which provides the requisite decay bound for one step of the tree recurrence. The main technical step in applying this lemma is to find $a, q$ as in the definition and a message $\phi$ symmetrizable with exponent $a$ for which the decay factor $\alpha$ is small; Lemma 4.4 below then shows how the decay factor comes into play in proving exponential decay of correlations over the tree.

**Lemma 4.3.** *Let $\phi$ be a message with derivative $\Phi$, and let $a$ and $q$ be positive reals such that $\frac{1}{a} + \frac{1}{q} = 1$. If $\phi$ is symmetrizable with exponent $a$, then for any two vectors $\boldsymbol{x}, \boldsymbol{y}$ in $\phi([0, b])^d$, there exists an integer $k \le d$ such that*

$$\left| f_d^\phi(\boldsymbol{x}) - f_d^\phi(\boldsymbol{y}) \right|^q \le \xi_{\phi,q}(k) \|\boldsymbol{x} - \boldsymbol{y}\|_q^q.$$

*Proof.* We apply Lemma 4.2. Assuming $\boldsymbol{z}$ is as defined in that lemma, we have by Hölder's inequality

$$\left| f_d^\phi(\boldsymbol{x}) - f_d^\phi(\boldsymbol{y}) \right| \leq \Phi(f_d(\boldsymbol{z})) \sum_{i=1}^d \frac{|y_i - x_i|}{\Phi(z_i)} \left| \frac{\partial f_d}{\partial z_i} \right|$$

$$\leq \Phi(f_d(\boldsymbol{z})) \left( \sum_{i=1}^d \left( \frac{1}{\Phi(z_i)} \left| \frac{\partial f_d}{\partial z_i} \right| \right)^a \right)^{1/a} \|\boldsymbol{x} - \boldsymbol{y}\|_q.$$

Since $\phi$ is symmetrizable with exponent $a$, we can replace $\boldsymbol{z}$ in the above inequality with a vector $\tilde{\boldsymbol{z}}$ all of whose non-zero entries are equal to some fixed real $\tilde{z}$. Let $k \leq d$ be the number of non-zero entries in $\tilde{\boldsymbol{z}}$. Using the consistency condition, we then get

$$\left| f_d^\phi(\boldsymbol{x}) - f_d^\phi(\boldsymbol{y}) \right| \leq \Phi(f_k(\tilde{z})) \left( \sum_{i=1}^k \left( \frac{1}{k\Phi(\tilde{z})} |f_k'(\tilde{z})| \right)^a \right)^{1/a} \|\boldsymbol{x} - \boldsymbol{y}\|_q$$

$$= \frac{1}{k^{1-1/a}} \frac{\Phi(f_k(\tilde{z})) |f_k'(\tilde{z})|}{\Phi(\tilde{z})} \|\boldsymbol{x} - \boldsymbol{y}\|_q.$$

Raising both sides to the power $q$, and using $\frac{1}{a} + \frac{1}{q} = 1$ and the definitions of the functions $\Xi$ and $\xi$, we get the claimed inequality. □

Given a message $\phi$ satisfying the conditions of Lemma 4.3, we can easily prove the following lemma on the propagation of errors in locally finite infinite trees. Recall that $F_\rho(\sigma)$ denotes the value computed by the recurrence at the root $\rho$ under an initial condition $\sigma$. The lemma quantifies the dependence of $F_\rho(\sigma)$ on initial conditions $\sigma$ which are fixed everywhere except at some cutset $C$, in terms of the distance of $C$ from $\rho$.

**Lemma 4.4.** *Let $T$ be a finite tree rooted at $\rho$. Let $C$ be a cutset in $T$ at distance at least $1$ from the root which does not contain any leaves, and let $C'$ be the cutset consisting of the children of vertices in $C$. Consider two arbitrary initial conditions $\sigma$ and $\tau$ on $T_{\leq C'}$ which differ only on $C'$, and which assign values from the interval $[0, b]$. Given a recurrence family $\{f_d\}_{d \geq 1}$ , let $a$ and $q$ be positive reals such that $\frac{1}{a} + \frac{1}{q} = 1$ and suppose $\phi$ is a message that is symmetrizable with exponent $a$. We then have*

$$|F_\rho(\sigma) - F_\rho(\tau)|^q \leq \left( \frac{M}{L} \right)^q \sum_{v \in C} \alpha^{|v|},$$

*where $\alpha$ is as defined in eq. (4.2), and $L$ and $M$ are defined as follows:*

$$L := \inf_{x \in (0,b)} \phi'(x); \quad M := \max_{v \in C} |\phi(F_v(\sigma)) - \phi(F_v(\tau))| .$$

For a proof of this lemma, see Section 4.6. The next two sections are devoted to showing that appropriately chosen messages fit the framework described above in such a way that the lemma can be used to deduce optimal decay of correlation and algorithmic results.

## 4.3 Application: The hard core model

Recall that the uniqueness phase transition of the hard core model on the infinite $d$-ary tree is easily described: there exists a critical activity $\lambda_c(d) := \frac{d^d}{(d-1)^{d+1}}$ such that uniqueness of the Gibbs measure (and weak spatial mixing) holds when $\lambda < \lambda_c(d)$, but not when $\lambda > \lambda_c(d)$. As discussed above, Weitz [Wei06] showed that $\lambda < \lambda_c(d)$ also implies strong spatial mixing on the $d$-ary tree, and gave a beautiful argument to exploit this to design an FPTAS for the partition function on graphs of degree at most $d + 1$ whenever $\lambda < \lambda_c(d)$ holds. This range of applicability was later proved to be optimal by Sly [Sly10] (see also [SS12, GGŠ⁺14]), who showed that an FPTAS for $(d + 1)$-regular graphs when $\lambda > \lambda_c(d)$ would imply that NP = RP. We now apply the framework in Section 4.2 to extend Weitz's result to graphs of bounded connective constant.

**Theorem 4.5.** *Let $\mathcal{G}$ be a family of finite graphs of connective constant at most $\Delta$, and let $\lambda$ be such that $\lambda < \lambda_c(\Delta)$. Then there is an FPTAS for the partition function of the hard core model with vertex activity $\lambda$ for all graphs in $\mathcal{G}$. Further, even if $\mathcal{G}$ contains locally finite infinite graphs, the model exhibits strong spatial mixing on all graphs in $\mathcal{G}$.*

Note that Sly's result alluded to above also implies that the range of applicability of the above result is optimal, since $(\Delta + 1)$-regular graphs have connective constant at most $\Delta$. Before proceeding with the proof, we first note some consequences of the theorem. We first consider the hard core model on graphs sampled from $\mathcal{G}(n, d/n)$.

**Corollary 4.6.** *Let $\lambda < \lambda_c(d)$. Then, there is an algorithm for approximating the partition function of graphs drawn from $\mathcal{G}(n, d/n)$ which, with high probability over the random choice of the graph, runs in polynomial time.*

*Proof of Corollary 4.6.* Since $\lambda < \lambda_c(d)$, there exists an $\epsilon > 0$ such that $\lambda < \lambda_c(d(1 + \epsilon))$. Fix $\beta > 0$. In order to prove the corollary, we only need to show that graphs drawn from $\mathcal{G}(n, d/n)$ have connective constant at most $d(1 + \epsilon)$ with probability at least $1 - n^{-\beta}$.

Recall that $N(v, \ell)$ is the number of self-avoiding walks of length $\ell$ starting at $v$. Suppose $\ell \geq a \log n$, where $a$ is a constant depending upon the parameters $\epsilon$, $\beta$ and $d$ which will be specified later. We first observe that

$$\mathbb{E}\left[\sum_{i=1}^{\ell} N(v, i)\right] \leq \sum_{i=1}^{\ell} \left(\frac{d}{n}\right)^i n^i \leq d^\ell \frac{d}{d-1},$$

and hence by Markov's inequality, we have $\sum_{i=1}^{\ell} N(v, i) \leq d^\ell \frac{d}{d-1}(1 + \epsilon)^\ell$ with probability at least $1 - (1 + \epsilon)^{-\ell}$. By choosing $a$ such that $a \log(1 + \epsilon) \geq \beta + 2$, we see that this probability is at least $1 - n^{-(\beta+2)}$. By taking a union bound over all $\ell$ with $a \log n \leq \ell \leq n$ and over all vertices $v$, we see that the connective constant $\Delta$ is at most $d(1 + \epsilon)$ with probability at least $1 - n^{-\beta}$. We therefore see that with probability at least $1 - n^{-\beta}$, the conditions of Theorem 4.5 are satisfied. This completes the proof.  □

| Lattice | Max. degree | Previous SSM bound $\lambda$ | | Connective Constant $\Delta$ | | Our SSM bound $\lambda$ |
|---------|-------------|------------------------------|--|------------------------------|--|-------------------------|
| $\mathbb{T}$ | 6 | 0.762 | [Wei06] | 4.251 419 | [Alm05] | 0.961 |
| $\mathbb{H}$ | 3 | 4.0 | [Wei06] | 1.847 760 | [DCS12] | 4.976 |
| $\mathbb{Z}^2$ | 4 | 2.48 | [VVY13] | 2.679 193 | [PT] | 2.082 (2.538$^\star$) |
| $\mathbb{Z}^3$ | 6 | 0.762 | [Wei06] | 4.7387 | [PT] | 0.822 |
| $\mathbb{Z}^4$ | 8 | 0.490 | [Wei06] | 6.8040 | [PT] | 0.508 |
| $\mathbb{Z}^5$ | 10 | 0.360 | [Wei06] | 8.8602 | [PT] | 0.367 |
| $\mathbb{Z}^6$ | 12 | 0.285 | [Wei06] | 10.8886 | [Wei] | 0.288 |

$^\star$ See Section 4.5 for a description of how this improved bound is obtained.

Table 4.1: Strong spatial mixing bounds for various lattices. ($\mathbb{Z}^D$ is the $D$-dimensional Cartesian lattice; $\mathbb{T}$ and $\mathbb{H}$ denote the triangular and honeycomb lattices respectively.)

The second consequence is to the study of conditions under which strong spatial mixing holds on regular lattices. In Table 4.1, we show the best known upper bound for the connective constant and the strong spatial mixing (SSM) bounds we obtain using these values in Theorem 4.5. In the table, a value $\alpha$ in the "$\lambda$" column means that SSM is shown to hold for the appropriate lattice whenever $\lambda \leq \alpha$. As expected, improvements over results obtained using Weitz's bound in terms of the maximum degree are the most pronounced for lattices with smaller maximum degree.

The table shows that except in the case of the 2D integer lattice $\mathbb{Z}^2$, our general result immediately gives improvements on the best known SSM bounds for all lattices using only previously known estimates of the connective constant. Not unexpectedly, our bound for $\mathbb{Z}^2$ using the connective constant as a black-box still improves upon Weitz's bound but falls short of the bounds obtained by Restrepo *et al.* [RST$^+$13] and Vera *et al.* [VVY13] using numerically intensive methods tailored to this special case. However, any improvement in the bound on the connective constant would immediately yield an improvement in the SSM bound obtained using Theorem 4.5. In Section 4.5, we will use a tighter analysis of the connective constant of a suitably constructed self-avoiding walk tree of $\mathbb{Z}^2$ to show that SSM holds on this lattice whenever $\lambda < 2.538$, which improves upon the specialized bound $\lambda < 2.48$, obtained in *et al.* [VVY13, RST$^+$13].

We now proceed with the proofs. As before, the first step is to choose a message, and in this case we choose a message that has been used before [LLY13]:

$$\phi(x) := \sinh^{-1}\left(\sqrt{x}\right), \text{ so that } \Phi(x) := \phi'(x) = \frac{1}{2\sqrt{x(1+x)}}. \tag{4.3}$$

Notice that $\phi$ is a strictly increasing, continuously differentiable function on $(0, \infty)$, and also satisfies the technical condition that the derivative $\Phi$ be bounded away from zero on any finite interval, as required in the definition of a message. The first important technical step is to verify that this message fits our framework:

**Lemma 4.7.** *For any $a \geq 2$, the message $\phi$ as defined in eq. (4.3) is symmetrizable with exponent $a$ with respect to the tree recurrence $\{f_{d,\lambda}\}_{d \geq 1}$ of the hard core model.*

The proof of the above lemma is quite technical and is deferred to Section 4.7.2. As discussed above, the crucial advantage of the notion of symmetrizability is the flexibility it allows in the choice of exponent $a$. The decay factor $\alpha$ in Lemma 4.4 which governs the rate of decay in the tree recurrences depends upon the choice of the exponent, and as we now show, it is possible to obtain an optimal decay rate by choosing an appropriate exponent satisfying Lemma 4.7.

Our choice of $a$ and $q$ will depend upon the vertex activity $\lambda$, and to clarify this dependence, we first define the quantity $\Delta_c := \Delta_c(\lambda)$ as the unique solution of $\lambda_c(t) = \lambda$ (the existence and uniqueness of $\Delta_c$ follows from the well known fact that $\lambda_c$ is a strictly decreasing function and maps the interval $(1, \infty)$ onto $(0, \infty)$). Our choice of $a$ and $q$ will then enforce the following conditions:

1. $a \geq 2$, so that that $\phi$ is symmetrizable with exponent $a$ and hence Lemma 4.4 is applicable; and

2. for all $d > 0$, $\xi_{\phi,q}(d) \leq \frac{1}{\Delta_c}$, so that we get sufficient stepwise decay when Lemma 4.4 is applied.

The second condition is the key to making the proof work, and in order to enforce it we will need to analyze the function $\xi_{\phi,q}(d)$ in some detail. We begin with the following simple lemma, which shows how to perform one of the maximizations needed in the definition of the decay factor $\alpha$. In what follows, we drop the subscript $\phi$ for simplicity of notation.

**Lemma 4.8.** *Consider the hard core model with any fixed vertex activity $\lambda > 0$. For any $q \geq 1$ and with $\phi$ as defined in eq. (4.3), we have $\xi_q(d) = \Xi_q(d, \tilde{x}_\lambda(d))$, where $\tilde{x}_\lambda(d)$ is the unique solution to*

$$d\tilde{x}_\lambda(d) = 1 + f_{d,\lambda}(\tilde{x}_\lambda(d)). \tag{4.4}$$

*Proof.* Plugging in $\Phi$ from eq. (4.3) in the definition of $\Xi$, we get

$$\Xi_q(d, x) = d^{q-1} \left( \frac{x}{1+x} \frac{f_{d,\lambda}(x)}{1 + f_{d,\lambda}(x)} \right)^q.$$

Taking the partial derivative with respect to the second argument, we get

$$\Xi_q^{(0,1)}(d, x) = \frac{q\Xi_q(d, x)}{2x(1+x)(1 + f_{d,\lambda}(x))} \left[ 1 + f_{d,\lambda}(x) - dx \right].$$

For fixed $d$, the quantity outside the square brackets is always positive, while the expression inside the square brackets is strictly decreasing in $x$. Thus, any zero of the expression in the brackets will be a unique maximum of $\Xi_q$. The fact that such a zero exists follows by noting that the partial derivative is positive at $x = 0$ and negative as $x \to \infty$. Thus, $\Xi_q(d, x)$ is maximized at $\tilde{x}_\lambda(d)$ as defined above, and hence $\xi_q(d) = \Xi_q(d, \tilde{x}_\lambda(d))$, as claimed. $\square$

We now choose $a$ and $q$ as follows:

$$\frac{1}{q} = 1 - \frac{\Delta_c - 1}{2} \log\left(1 + \frac{1}{\Delta_c - 1}\right); \quad \frac{1}{a} = 1 - \frac{1}{q}. \tag{4.5}$$

Note that since $\log(1 + y) \le y$ for all $y \ge 0$, we get that $q \le 2$ (and hence $a \ge 2$) by using $y = \frac{1}{\Delta_c - 1}$, and noting that $\Delta_c > 1$. Thus, the first condition above is already satisfied. The values above are chosen to make sure that the second condition is satisfied as well, as we prove in the next lemma. Now that the exponents $a$ and $q$ are fixed, we define the function $\nu_\lambda(d)$ as follows in order to emphasize dependence upon $\lambda$:

$$\nu_\lambda(d) := \xi_q(d).$$

**Lemma 4.9.** *Fix $\lambda > 0$ and let $\Delta_c(\lambda) > 1$ be the unique solution to $\lambda_c(t) = \lambda$. The function $\nu_\lambda : \mathbb{R}^+ \to \mathbb{R}^+$ is maximized at $d = \Delta_c := \Delta_c(\lambda)$. Further,*

$$\nu_\lambda(\Delta_c(\lambda)) = \frac{1}{\Delta_c(\lambda)}.$$

The proof of the above lemma is somewhat technical, and is deferred to Section 4.7.1. The lemma shows that when $\lambda < \lambda_c(\Delta)$, the decay factor $\alpha < \frac{1}{\Delta}$. As we observed above, the main ingredient in the proof is the specific choice of the exponent $a$ in eq. (4.5), which in turn is allowed only because of the flexibility in the choice of $a$ allowed by Lemma 4.7.

Given the estimate on the decay factor in the above lemma, the proof of Theorem 4.5 requires only some standard arguments. However, we describe the steps in detail for future reference.

*Proof of Theorem 4.5.* Let $\mathcal{F}$ be any family of finite or infinite graphs with connective constant $\Delta$. We prove the result for any fixed $\lambda$ such that $\lambda < \lambda_c(\Delta)$. For such $\lambda$, we have $\Delta_c(\lambda) > \Delta$ (since $\lambda_c$ is a decreasing function). Using Lemma 4.9 we then see that there is an $\epsilon > 0$ such that $\nu_\lambda(d)\Delta \le 1 - \epsilon$ for all $d > 0$.

We first prove that the hard core model with these parameters exhibits strong spatial mixing on this family of graphs. Let $G$ be any graph from $\mathcal{F}$, $v$ any vertex in $G$, and consider any boundary conditions $\sigma$ and $\tau$ on $G$ which differ only at a distance of at least $\ell$ from $v$. We consider the Weitz self-avoiding walk tree $T_{SAW}(v, G)$ rooted at $v$. As before, we denote again by $\sigma$ (respectively, $\tau$) the translation of the boundary condition $\sigma$ (respectively, $\tau$) on $G$ to $T_{SAW}(v, G)$. From Weitz's theorem, we then have that $R_v(\sigma, G) = R_v(\mathcal{W} \cup \sigma, T_{SAW}(v, G))$ (respectively, $R_v(\tau, G) = R_v(\mathcal{W} \cup \tau, T_{SAW}(v, G))$).

Consider first the case where $G$ is infinite. Let $C_\ell$ denote the cutset in $T_{SAW}(v, G)$ consisting of all vertices at distance $\ell$ from $v$. Since $G$ has connective constant at most $\Delta$, it follows that for $\ell$ large enough, we have $|C_\ell| \le \Delta^\ell(1 - \epsilon/2)^{-\ell}$. Further, in the notation of Lemma 4.4, $\nu_\lambda(d)\Delta = 1 - \epsilon$ implies that the decay factor $\alpha$ (defined in eq. (4.2)) is at most $(1 - \epsilon)/\Delta$. We now apply Lemma 4.4. We first observe that given our message $\phi$, we can bound the quantities $L$ and $M$ in the lemma as

$$L = \frac{1}{2\sqrt{\lambda(1 + \lambda)}} \quad \text{and} \quad M = \sinh^{-1}(\sqrt{\lambda}).$$

The bounds on $L$ and $M$ follows from the fact that the values of the occupation ratio computed at any internal node of the tree lie in the range $[0, \lambda]$. Setting $c_0 = (L/M)^q$, we can the apply the lemma to get

$$|R_v(\sigma, G) - R_v(\tau, G)|^q = |R_v(\mathcal{W} \cup \sigma, T_{SAW}(v, G)) - R_v(\mathcal{W} \cup \tau, T_{SAW}(v, G))|^q$$

$$\leq c_0 \sum_{u \in C_\ell} \left(\frac{1 - \epsilon}{\Delta}\right)^\ell$$

$$\leq c_0 \left(\frac{1 - \epsilon}{1 - \epsilon/2}\right)^\ell, \text{ using } |C_\ell| \leq \Delta^\ell (1 - \epsilon/2)^{-\ell},$$

which establishes strong spatial mixing in $G$, since $1 - \epsilon < 1 - \epsilon/2$.

We now consider the case when $\mathcal{F}$ is a family of finite graphs, and $G$ is a graph from $\mathcal{F}$ of $n$ vertices. Since the connective constant of the family is $\Delta$, there exist constants $a$ and $c$ (not depending upon $G$) such that for $\ell \geq a \log n$, $\sum_{i=1}^\ell N(v, \ell) \leq c\Delta^\ell$. We now proceed with the same argument as in the infinite case, but choosing $\ell \geq a \log n$. The cutset $C_\ell$ is again chosen to be the set of all vertices at distance $\ell$ from $v$ in $T_{SAW}(v, G)$, so that $|C_\ell| \leq c\Delta^\ell$. As before, we then have for $\ell > a \log n$,

$$|R_v(\sigma, G) - R_v(\tau, G)|^q = |R_v(\mathcal{W} \cup \sigma, T_{SAW}(v, G)) - R_v(\mathcal{W} \cup \tau, T_{SAW}(v, G))|^q$$

$$\leq c_0 \sum_{u \in C_\ell} \left(\frac{1 - \epsilon}{\Delta}\right)^\ell$$

$$\leq c \cdot c_0 (1 - \epsilon)^\ell, \text{ using } |C_\ell| \leq c\Delta^\ell, \tag{4.6}$$

which establishes the requisite strong spatial mixing bound.

In order to prove the algorithmic part, we first recall that an FPTAS for the "non-occupation" probabilities $1 - p_v$ under arbitrary boundary conditions is sufficient to derive an FPTAS for the partition function. We further note that if the vertex $v$ is not already fixed by a boundary condition, then $1 - p_v = \frac{1}{1 + R_v} \geq \frac{1}{1 + \lambda}$, since $R_v$ lies in the interval $[0, \lambda]$ for any such vertex. Hence, an additive approximation to $R_v$ with error $\mu$ implies a multiplicative approximation to $1 - p_v$ within a factor of $1 \pm \mu(1 + \lambda)$. Thus, an algorithm that produces in time polynomial in $n$ and $1/\mu$ an *additive* approximation to $R_v$ with error at most $\mu$ immediately gives an FPTAS for $1 - p_v$, and hence, by Weitz's observation, also for the partition function. To derive such an algorithm, we again use the tree $T_{SAW}(v, G)$ considered above. Suppose we require an additive approximation with error at most $\mu$ to $R_v(\sigma, G) = R_v(\sigma, T_{SAW}(v, G))$. We notice first that $R_v = 0$ if and only if there is a neighbor of $v$ that is fixed to be occupied in the boundary condition $\sigma$. In this case, we simply return 0. Otherwise, we expand $T_{SAW}(v, G)$ up to depth $\ell$ for some $\ell \geq a \log n$ to be specified later. Notice that this subtree can be explored in time $O\left(\sum_{i=1}^\ell N(v, i)\right)$ which is $O(\Delta^\ell)$ since the connective constant is at most $\Delta$. We now consider two extreme boundary conditions $\sigma_+$ and $\sigma_-$ on $C_\ell$: in $\sigma_+$ (respectively, $\sigma_-$) all vertices in $C_\ell$ that are not already fixed by $\sigma$ are fixed to "occupied" (respectively, unoccupied). The form of the recurrence ensures that the true value

$R_v(\sigma)$ lies between the values $R_v(\sigma_+)$ and $R_v(\sigma_-)$. We compute the recurrence for both these boundary conditions on the tree. The analysis leading to eq. (4.6) ensures that, since $\ell \geq a \log n$, we have

$$|R_v(\sigma_+, G) - R_v(\sigma_-, G)| \leq M_1 \exp(-M_2 \ell)$$

for some fixed positive constants $M_1$ and $M_2$. Now, assume without loss of generality that $R_v(\sigma_+) \geq R_v(\sigma_-)$. By the preceding observations, we then have

$$R_v(\sigma) \leq R_v(\sigma_+) \leq R_v(\sigma) + M_1 \exp(-M_2 \ell).$$

By choosing $\ell = a \log n + O(1) + O(\log(1/\mu))$, we get the required $\pm \mu$ approximation. Further, by the observation above, the algorithm runs in time $O\left(\Delta^\ell\right)$, which is polynomial in $n$ and $1/\mu$ as required. □

## 4.4 Application: The monomer-dimer model

We now apply the framework of Section 4.2 to the monomer-dimer model. However, there are two crucial differences from the hard core model. First, in the case of the hard core model, the result in Theorem 4.5 is the best known even taking into account randomized algorithms based on Markov chain Monte Carlo (MCMC). In fact, no MCMC based algorithms for approximating the partition function of the hard core model have a range of applicability as wide as the results of Weitz [Wei06] or Theorem 4.5. In sharp contrast, for the monomer-dimer model, Jerrum and Sinclair [JS89] gave a *randomized* fully polynomial time approximation scheme (FPRAS) based on MCMC for the partition function which, for a fixed value of the edge activity $\gamma$, is valid for *all* graphs, without any restriction on the maximum degree. However, the best *deterministic* FPTAS for the problem is due to Bayati, Gamarnik, Katz, Nair and Tetali [BGK+07], who gave an algorithm which is polynomial time only on bounded degree graphs. Our goal in this section is to improve this result to give an FPTAS for graphs with bounded connective constant.

The second difference is that unlike the hard core and the anti-ferromagnetic Ising models, the monomer-dimer model does not show a phase transition with respect to weak spatial mixing on the $d$-ary tree, i.e., weak spatial mixing holds for all values of $\gamma$ at all arities $d$. However, the rate of weak spatial mixing does approach one if any one of $\gamma$ or $d$ are taken to infinity. In our results, we will establish that weak spatial mixing on the $d$-ary tree with rate $c$ implies the same rate on *all* graphs of connective constant at most $d$. Using this, we will prove the following result.

**Theorem 4.10.** *Let $\mathcal{G}$ be a family of graphs of connective constant at most $\Delta$, and let $\gamma > 0$ be any fixed edge activity. Then there is an FPTAS for the partition function of the monomer-dimer model with edge activity $\gamma$ for all graphs in $\mathcal{G}$. More specifically, the running time of the FPTAS for producing an $(1 \pm \epsilon)$ factor approximation is $(n/\epsilon)^{O(\sqrt{\gamma \Delta} \log \Delta)}$.*

As discussed above, the previous best deterministic approximation algorithm for the partition function of the monomer-dimer model was due to Bayati *et al.* [BGK+07], and ran in time $(n/\epsilon)^{O(\sqrt{\gamma d} \log d)}$ for graphs of degree at most $d + 1$. The above algorithm replaces the maximum

degree constraint of Bayati *et al.* by a corresponding constraint on the connective constant, without requiring any bounds on the maximum degree. In particular, for graphs such as $\mathcal{G}(n, d/n)$ which have bounded connective constant and unbounded degree, our analysis yields a polynomial time algorithm (for any fixed value of the edge activity $\gamma$) in contrast to the sub-exponential time algorithm obtained by Bayati *et al.* [BGK+07]. Using an observation of Kahn and Kim [KK98], Bayati *et al.* also pointed out that the $\sqrt{d}$ factor in the exponent of their running time is optimal for algorithms which are based on Weitz's framework and which use only the fact that the maximum degree of the graph is at most $d + 1$. A similar observation shows that the $\sqrt{\gamma\Delta}$ factor in the exponent of our running time is optimal for algorithms in the Weitz framework which use bounds on the connective constant; we will discuss this in more detail following the proof of Theorem 4.10.

Before proceeding with the proof, we again note that, exactly as in the case of the hard core model, Theorem 4.10 implies an analog of Corollary 4.6 for graphs drawn from the Erdős–Rényi model $\mathcal{G}(n, d/n)$. We now proceed with the proof. As in the case of the hard core model, the first step is to choose an appropriate message. Unfortunately, unlike the case of the hard core model where we could show that an already known message was sufficient, we need to find a new message function in this case. We claim that the following message works:

$$\phi(x) := \frac{1}{2}\log\left(\frac{x}{2-x}\right), \text{ so that } \Phi(x) := \phi'(x) = \frac{1}{x(2-x)}. \tag{4.7}$$

Note that $\phi$ is strictly increasing and continuously differentiable on the interval $(0, 1]$, and its derivative is bounded away from $0$ on that interval. Thus, $\phi$ satisfies the conditions required in the definition of a message (note that the bound $b$ used in the definition is $1$ in the case of the monomer-dimer model). Now, in order to apply Lemma 4.4, we first study the symmetrizability of $\phi$ in the following technical lemma.

**Lemma 4.11.** *Fix $r \in (1, 2]$. The message $\phi$ as defined in eq. (4.7) is symmetrizable with exponent $r$ with respect to the tree recurrences $\{f_{d,\gamma}\}_{d\geq 1}$ of the monomer-dimer model.*

We defer the proof of the above lemma to Section 4.8. As in the case of the hard core model, we will need to make a careful choice of the exponent $r$ in order to obtain an optimal decay factor. We begin with a technical lemma which characterizes the behavior of the function $\xi$ used in the definition of the decay factor. For ease of notation, we drop the subscript $\phi$ from the notation for $\xi$.

**Lemma 4.12.** *Consider the monomer-dimer model with edge activity $\gamma$, and let $\phi$ be the message chosen in (4.7). For any $q > 1$, we have $\xi_q(d) = \Xi_q(d, \tilde{p}_\gamma(d))$, where $\tilde{p}_\gamma(d)$ satisfies $\Xi_q^{(0,1)}(d, \tilde{p}_\gamma(d)) = 0$ and is given by*

$$\tilde{p}_\gamma(d) := \frac{\sqrt{1 + 4\gamma d} - 1}{2\gamma d}.$$

*Proof.* Plugging in $\Phi$ from eq. (4.7) in the definition of $\Xi$, we get

$$\Xi_q(d, x) = d^{q-1}\left(\frac{\gamma x(2-x)f_{d,\gamma}(x)}{2 - f_{d,\gamma}(x)}\right)^q = d^{q-1}\left(\frac{\gamma x(2-x)}{1 + 2\gamma dx}\right)^q, \text{ since } f_{d,\gamma}(x) = \frac{1}{1 + \gamma dx}.$$

Taking the partial derivative with respect to the second argument, we get

$$\Xi_q^{(0,1)}(d, x) = \frac{2q\Xi_q(d, x)}{x(2 - x)(1 + 2\gamma dx)} \left[1 - x - \gamma dx^2\right].$$

For fixed $d$, and $0 \leq x \leq 1$, the quantity outside the square brackets is always positive, while the expression inside the square brackets is strictly decreasing in $x$. Thus, any zero of the expression in the brackets in the interval $[0, 1]$ will be a unique maximum of $\Xi_q$. By solving the quadratic, we see that $\tilde{p}_\gamma(d)$ as defined above is such a solution. Thus, $\Xi_q(d, x)$ is maximized at $\tilde{p}_\gamma(d)$ as defined above, and hence $\xi_q(d) = \Xi_q(d, \tilde{p}_\gamma(d))$. □

Given the edge activity $\gamma$ and an upper bound $\Delta$ on the connective constant of the graph family being considered, we now choose $D > \max(\Delta, 3/(4\gamma))$. We claim that we can get the required decay factor by choosing

$$\frac{1}{r} = 1 - \frac{1}{\sqrt{1 + 4\gamma D}}; \qquad \frac{1}{q} = 1 - \frac{1}{r} = \frac{1}{\sqrt{1 + 4\gamma D}}. \tag{4.8}$$

Note that the choice of $D$ implies that $1 < r \leq 2$, so that $\phi$ is symmetrizable with respect to $r$. The following lemma shows that this choice of $r$ indeed gives us the required decay factor. As in the case of the hard core model, we emphasize the dependence of the decay parameter on the model parameters by setting

$$\nu_\gamma(d) := \xi_q(d), \text{ where } q \text{ is as chosen in eq. (4.8).}$$

**Lemma 4.13.** *Fix $\gamma > 0$ and $D > 3/4\gamma$, and let $q$ be as chosen in (4.8). Then the function $\nu_\gamma : \mathbb{R}^+ \to \mathbb{R}^+$ is maximized at $d = D$. Further, the decay factor $\alpha$ is given by*

$$\alpha = \nu_\gamma(D) = \frac{1}{D}\left(1 - \frac{2}{1 + \sqrt{1 + 4\gamma D}}\right)^q.$$

*Proof.* We consider the derivative of $\nu_\gamma(d)$ with respect to $d$. Recalling that $\nu_\gamma(d) = \xi(d) = \Xi(d, \tilde{p}_\gamma(d))$ and using the chain rule, we have

$$\begin{aligned}
\nu_\gamma'(d) &= \Xi^{(1,0)}(d, \tilde{p}) + \Xi^{(0,1)}(d, \tilde{p})\frac{d\tilde{p}}{dd} \\
&= \Xi^{(1,0)}(d, \tilde{p}), \text{ since } \Xi^{(0,1)}(d, \tilde{p}) = 0 \text{ by definition of } \tilde{p} \\
&= \frac{\Xi(d, \tilde{p})}{d(1 + 2\gamma d\tilde{p})}[q - 1 - 2\gamma d\tilde{p}] = \frac{\Xi(d, \tilde{p})}{d(1 + 2\gamma d\tilde{p})}\left[\sqrt{1 + 4\gamma D} - \sqrt{1 + 4\gamma d}\right],
\end{aligned} \tag{4.9}$$

where in the last line we substitute the values $\tilde{p}_\gamma(d) = (\sqrt{1 + 4\gamma d} - 1)/(2\gamma d)$ from Lemma 4.12 and $q = \sqrt{1 + 4\gamma D}$ from eq. (4.8). Now, we note that in eq. (4.9) the quantity outside the square brackets is always positive, while the quantity inside the square brackets is a strictly decreasing function of $d$ which is positive for $d < D$ and negative for $d > D$. It follows that $\nu_\gamma'(d)$ has a unique zero at $d = D$ for $d \geq 0$, and this zero is a global maximum of $\nu_\gamma$. □

We are now ready to prove our main result for the monomer-dimer model, Theorem 4.10. Given Lemmas 4.4 and 4.13, the proof needs some standard arguments similar to those used in the proof of Theorem 4.5 for the hard core model in Section 4.3.

*Proof of Theorem 4.10.* Let $\mathcal{F}$ be any family of finite graphs with connective constant at most $\Delta$. Given the vertex activity $\gamma$ of the monomer-dimer model, we choose $D = \max(\Delta, 3/(4\gamma))$. Using Lemma 4.13, we then see that the decay factor $\alpha$ appearing in Lemma 4.4 can be chosen to be

$$\alpha = \frac{1}{D}\left(1 - \frac{2}{1 + \sqrt{1 + 4\gamma D}}\right)^q.$$

Now, let $G$ be any graph (with $n$ vertices) from $\mathcal{F}$, and let $v$ be a vertex in $G$. As observed in Chapter 2, it is sufficient to construct an FPTAS for $p_v(G)$ in order to derive an FPTAS for the partition function.

Consider now the self-avoiding walk tree $T_{SAW}(v, G)$ rooted at $v$ (as defined in Section 2.1). From Godsil's theorem (Theorem 2.1), we know that $p_v(G) = p_v(T_{SAW}(v, G))$. Let $C_\ell$ denote the cutset in $T_{SAW}(v, G)$ consisting of all vertices at distance $\ell$ from $v$. Since $\mathcal{F}$ has connective constant at most $\Delta$, there exist constants $a$ and $c$ such that if $\ell \geq a \log n$, we have $\sum_{i=1}^{\ell} N(v, \ell) \leq c\Delta^\ell$. We will now apply Lemma 4.4 with $q$ as defined in eq. (4.8). We first observe that the quantities $L$ and $M$ in the lemma can be taken to be

$$L = 1, \quad \text{and,} \quad M = \frac{1}{2}\log(1 + 2\gamma n),$$

since the degree of any vertex in $G$ is at most $n$. Now, defining $c_0 := (M/L)^q$, and using $|C_\ell| \leq c\Delta^\ell$ and $D \geq \Delta$, we have

$$|F_v(\mathbf{0}_\ell) - F_v(\mathbf{1}_\ell)|^q \leq c_0 \sum_{u \in C_\ell} \alpha^\ell \leq c \cdot c_0 \cdot (\alpha\Delta)^\ell, \text{ using } |C_\ell| \leq c\Delta^\ell,$$

$$\leq c \cdot c_0 \cdot \left(1 - \frac{2}{1 + \sqrt{1 + 4\gamma D}}\right)^{q\ell}, \text{ using } D \geq \Delta \text{ after substituting for } \alpha.$$

$$(4.10)$$

Raising both sides to the power $1/q$ and substituting for $c_0$ and $q$, we then have

$$|F_v(\mathbf{0}_\ell) - F_v(\mathbf{1}_\ell)| \leq \frac{1}{2}c^{1/\sqrt{1+4\gamma D}} \cdot \log(1 + 2\gamma n) \cdot \left(1 - \frac{2}{1 + \sqrt{1 + 4\gamma D}}\right)^\ell. \qquad (4.11)$$

To analyze the running time, we note that in order to obtain a $(1 \pm \epsilon)$ multiplicative approximation to $p_v(G)$, it is sufficient to obtain a $\pm\epsilon/(1 + \gamma n)$ additive approximation; this is because $p_v(G) \geq 1/(1 + \gamma n))$ since the degree of each vertex in $G$ is at most $n$. Now, as observed in Section 2.1.1, $p_v(G)$ always lies between the quantities $F_v(\mathbf{0}_\ell)$ and $F_v(\mathbf{1}_\ell)$, so in order to obtain a $\pm\epsilon/(1 + \gamma n)$ approximation, it is sufficient to choose $\ell \geq a \log n$ large enough so that the right hand side of

eq. (4.11) is at most $\epsilon/(1 + \gamma n)$. Denoting by $\beta$ the quantity in the parenthesis on the right hand side of eq. (4.11), we can ensure this by choosing

$$\ell \geq \frac{1}{\log(1/\beta)} \left[ \log \frac{1 + \gamma n}{\epsilon} + \log \log \left( \sqrt{1 + 2\gamma n} \right) + \frac{1}{\sqrt{1 + 4\gamma D}} \log c \right].$$

Further, given such an $\ell$, the running time of the algorithm is $O(\sum_{i=1}^{\ell} N(v, \ell)) = O(\Delta^{\ell})$, since this is the time it takes to expand the self-avoiding walk tree up to depth $\ell$. Noting that $1/(\log(1/\beta)) = \sqrt{\gamma\Delta} + \Theta(1)$, we obtain an algorithm running in time

$$((1 + \gamma n)/\epsilon)^{O(\sqrt{\gamma\Delta} \cdot \log \Delta)}$$

which provides a $(1 \pm \epsilon)$ multiplicative approximation for $p_v(G)$. Recalling the arguments in Section 2.3, we see that this yields an algorithm for approximating the partition function up to a multiplicative factor of $(1 \pm \epsilon)$ with the same asymptotic exponent in the running time. This completes the proof. $\square$

**Remark 4.2.** Note that eq. (4.11) can be interpreted as showing that (when $\Delta > 1/(4\gamma)$), strong spatial mixing holds with rate $1 - \frac{2}{1+\sqrt{1+4\gamma\Delta}}$ on graphs of connective constant at most $\Delta$, and the factor $\sqrt{\gamma\Delta}$ in the exponent of our runtime is a direct consequence of this fact (in particular, strong spatial mixing at rate $c$ translates into the exponent being proportional to $\log c$). Recall that Bayati *et al.* [BGK⁺07] obtained an algorithm with the same runtime but only for graphs with *maximum degree* $\Delta + 1$ (which is a strict subset of the class of graphs with connective constant $\Delta$). This was due to the fact that their analysis essentially amounted to proving that eq. (4.11) holds for the special case of graphs of maximum degree $\Delta + 1$. Using an observation of Kahn and Kim [KK98] that the rate of spatial mixing on the infinite $d$-ary tree is $1 - 1/\Theta(\sqrt{\gamma\Delta})$, they concluded that such a runtime was the best possible for algorithms that use decay of correlations in this direct fashion. We note here that since the infinite $d$-ary tree also has connective constant exactly $d$, this observation also implies that the rate of strong spatial mixing obtained in eq. (4.11) is optimal for graphs of connective constant $\Delta$ (in fact, the rate of strong spatial mixing on the $d$-ary tree is exactly $1 - \frac{2}{1+\sqrt{1+4\gamma d}}$).

We now conclude this chapter by providing the technical details omitted from the previous sections.

## 4.5 Description of numerical results

> *Four more than hundred, multiplied by eight,*
> *And then added to sixty-two thousand,*
> *Approaches the circumference of a circle*
> *With diameter twenty thousand.*
>
> – Āryabhaṭa, *Āryabhaṭīya.* (c. 500 CE)

In this section, we describe the derivation of the numerical bounds in Table 4.1. As in [SSY13], all of the bounds are direct applications of Theorem 4.5 using published upper bounds on the connective constant for the appropriate graph (except for the starred bound of 2.538 for the case of $\mathbb{Z}^2$, which we discuss in greater detail below). The exact connective constant is not known for the Cartesian lattices $\mathbb{Z}^2, \mathbb{Z}^3, \mathbb{Z}^4, \mathbb{Z}^5$ and $\mathbb{Z}^6$, and the triangular lattice $\mathbb{T}$, and we use the rigorous upper and lower bounds available in the literature [MS96, Wei]. In contrast, for the honeycomb lattice, Duminil-Copin and Smirnov [DCS12] rigorously established the connective constant to be $\mathbb{H}$ is $\sqrt{2 + \sqrt{2}}$ in a celebrated breakthrough, and this is the bound we use for that lattice. In order to apply Theorem 4.5 for a given lattice of connective constant at most $\Delta$, we simply need to compute $\lambda_c(\Delta) = \frac{\Delta^\Delta}{(\Delta-1)^{(\Delta+1)}}$, and the monotonicity of $\lambda_c$ guarantees that the lattice exhibits strong spatial mixing as long as $\lambda < \lambda_c(\Delta)$.

We now consider the special case of $\mathbb{Z}^2$. As we pointed out in Section 4.3, any improvement in the connective constant of a lattice (or that of the Weitz SAW tree corresponding to the lattice) will immediately lead to an improvement in our bounds. In fact, as we discuss below, Weitz's construction allows for significant freedom in the choice of the SAW tree. We show here that using a tighter combinatorial analysis of the connective constant of a suitably chosen Weitz SAW tree of $\mathbb{Z}^2$, we can improve upon the bounds obtained by Restrepo *et al.* [RST$^+$13] and Vera *et al.* [VVY13] using sophistical methods tailored to the special case of $\mathbb{Z}^2$. Our basic idea is to exploit the fact that the Weitz SAW tree adds additional boundary conditions to the canonical SAW tree of the lattice. Thus, it allows a strictly smaller number of self-avoiding walks than the canonical SAW tree, and therefore can have a smaller connective constant than that of the lattice itself. Further, as in [SSY13], the proof of Theorem 4.5 only uses the Weitz SAW tree, and hence the bounds obtained there clearly hold if the connective constant of the Weitz SAW tree is used in place of the connective constant of the lattice.

The freedom in the choice of the Weitz SAW tree—briefly alluded to above—also offers the opportunity to incorporate another tweak which can potentially increase the effect of the boundary conditions on the connective constant. In Weitz's construction, the boundary conditions on the SAW tree are obtained in the following way (see Theorem 3.1 in [Wei06]). First, the neighbors of each vertex are ordered in a completely arbitrary fashion: this ordering need not even be consistent across vertices. Whenever a loop, say $v_0, v_1, \ldots, v_l, v_0$ is encountered in the construction of the SAW tree, the occurrence of $v_0$ which closes the loop is added to the tree along with a boundary condition which is determined by the ordering at $v_0$: if the neighbor $v_1$ (which "started" the loop) happens to be smaller than $v_l$ (the last vertex before the loop is discovered) in the ordering, then the last copy of $v_0$ appears in the tree fixed as "occupied", while otherwise, it appears as "unoccupied".

The orderings at the vertices need not even be fixed in advance, and different copies of the vertex $v$ appearing in the SAW tree can have different orderings, as long as the ordering at a vertex $v$ in the tree is a function only of the path from the root of the tree to $v$. We now specialize our discussion to $\mathbb{Z}^2$. The simplest such ordering is the "uniform ordering", where we put an ordering on the cardinal directions *north*, *south*, *east* and *west*, and order the neighbors at each vertex in accordance with this ordering on the directions. This was the approach used by Restrepo

*et al.* [RST+13].

However, it seems intuitively clear that it should be possible to eliminate more vertices in the tree by allowing the ordering at a vertex $v$ in the tree to depend upon the path taken from the origin to $v$. We use a simple implementation of this idea by using a "relative ordering" which depends only upon the last step of such a path. In particular, there are only three possible options available at a vertex $v$ in the tree (except the root): assuming the parent of $v$ in the tree is $u$: the first is to go *straight*, i.e., to proceed to the neighbor of $v$ (viewed as a point in $\mathbb{Z}^2$ which lies in the same direction as the vector $v - u$, where $v$ and $u$ are again viwed as points in $\mathbb{Z}^2$). Analogously, we can also turn *left* or *right* with respect to this direction. Our ordering simply stipulates that *straight > right > left*.

To upper bound the connective constant of the Weitz SAW tree, we use the standard method of *finite memory self-avoiding walks* [MS96]—these are walks which are constrained only to not have cycles of length up to some finite length $L$. Clearly, the number of such walks of any given length $\ell$ upper bounds $N(v, \ell)$. In order to bring the boundary conditions on the Weitz SAW tree into play, we further enforce the constraint that the walk is not allowed to make any moves which will land it in a vertex fixed to be "unoccupied" by Weitz's boundary conditions (note that a vertex $u$ can be fixed to be "unoccupied" also because one of its children is fixed to be "occupied": the independence set constraint forces $u$ itself to be "unoccupied" in this case, and hence leads to additional pruning of the tree by allowing the other children of $u$ to be ignored). Such a walk can be in one of a finite number $k$ (depending upon $L$) of states, such that the number of possible moves it can make to state $j$ while respecting the above constraints is some finite number $M_{ij}$. The $k \times k$ matrix $M = (M_{ij})_{i,j \in [k]}$ is called the *branching matrix* [RST+13]. We therefore get $N(v, \ell) \le \boldsymbol{e_1}^T M^\ell \boldsymbol{1}$, where $\boldsymbol{1}$ denotes the all 1's vector, and $\boldsymbol{e_1}$ denotes the co-ordinate vector for the state of the zero-length walk.

Since the entries of $M$ are non-negative, the Perron-Frobenius theorem implies that one of the maximum magnitude eigenvalues of the matrix $M$ is a positive real number $\gamma$. Using Gelfand's formula (which states that $\gamma = \lim_{\ell \to \infty} \|M^\ell\|^{1/\ell}$, for any fixed matrix norm) with the $\ell_\infty$ norm to get the last equality, we see that

$$\limsup_{\ell \to \infty} N(v, \ell)^{1/\ell} \le \limsup_{\ell \to \infty} (\boldsymbol{e_1}^T M^\ell \boldsymbol{1})^{1/\ell} \le \limsup_{\ell \to \infty} \|M^\ell\|_\infty^{1/\ell} = \gamma.$$

Hence, the largest real eigenvalue $\gamma$ of $M$ gives a bound on the connective constant of the Weitz SAW tree.

Using the matrix $M$ corresponding to walks in which cycles of length at most $L = 26$ are avoided, we get that the connective constant of the Weitz SAW tree is at most $2.433$ (we explicitly construct the matrix $M$ and then use Matlab to compute its largest eigenvalue). Using this bound for $\Delta$, and applying Theorem 4.5 as described above, we get the bound $2.529$ for $\lambda$ in the notation of the table, which is better than the bounds obtained by Restrepo *et al.* [RST+13] and Vera *et al.* [VVY13]. With additional computational optimizations we can go further and analyze self avoiding walks avoiding cycles of length at most $L = 30$. The first optimization is merging "isomorphic" states (this will decrease the number of states and hence the size of $M$ significantly, allowing computation of the largest eigenvalue): formally, the state of a SAW will be a suffix of

length $s$ such that the Manhattan distance between the final point and the point $s$ steps in the past is less than $L - s$ (note that the state of a vertex in the SAW tree can be determined from the state of its parent and the last step), and two states are isomorphic if they have the same neighbors at the next step of the walk. The second optimization is computing the largest eigenvalue using the power method. For $L = 30$ we obtain that the connective constant of the Weitz SAW tree is at most $2.429$, which on applying Theorem 4.5 yields the bound $2.538$ for $\lambda$, as quoted in Table 4.1.

## 4.6 Proofs omitted from Section 4.2

This section provides the proof of Lemmas 4.2 and 4.4, both of which require fairly standard arguments.

*Proof of Lemma 4.2.* Define $H(t) := f_{d,\lambda}^\phi(t\boldsymbol{x} + (1-t)\boldsymbol{y})$ for $t \in [0,1]$. By the scalar mean value theorem applied to $H$, we have

$$f_{d,\lambda}^\phi(\boldsymbol{x}) - f_{d,\lambda}^\phi(\boldsymbol{y}) = H(1) - H(0) = H'(s), \text{ for some } s \in [0,1].$$

Let $\psi$ denote the inverse of the message $\phi$: the derivative of $\psi$ is given by $\psi'(y) = \frac{1}{\Phi(\psi(y))}$, where $\Phi$ is the derivative of $\phi$. We now define the vector $\boldsymbol{z}$ by setting $z_i = \psi(sx_i + (1-s)y_i)$ for $1 \le i \le d$. We then have

$$\left| f_{d,\lambda}^\phi(\boldsymbol{x}) - f_{d,\lambda}^\phi(\boldsymbol{y}) \right| = |H'(s)| = \left| \left\langle \nabla f_{\lambda,d}^\phi(s\boldsymbol{x} + (1-s)\boldsymbol{y}), \boldsymbol{x} - \boldsymbol{y} \right\rangle \right|$$

$$= \Phi(f_{d,\lambda}(\boldsymbol{z})) \left| \sum_{i=1}^d \frac{x_i - y_i}{\Phi(z_i)} \frac{\partial f_{d,\lambda}}{\partial z_i} \right|, \quad \text{using the chain rule}$$

$$\le \Phi(f_{d,\lambda}(\boldsymbol{z})) \sum_{i=1}^d \frac{|y_i - x_i|}{\Phi(z_i)} \left| \frac{\partial f_{d,\lambda}}{\partial z_i} \right|, \quad \text{as claimed.}$$

We recall that for simplicity, we are using here the somewhat non-standard notation $\frac{\partial f}{\partial z_i}$ for the value of the partial derivative $\frac{\partial f}{\partial R_i}$ at the point $\boldsymbol{R} = \boldsymbol{z}$. $\qquad \square$

We now give the proof of Lemma 4.4.

*Proof of Lemma 4.4.* Recall that given a vertex $v$ in $T_{\le C}$, $T_v$ is the subtree rooted at $v$ and containing all the descendants of $v$, and $F_v(\sigma)$ is the value computed by the recurrence at the root $v$ of $T_v$ under the initial condition $\sigma$ restricted to $T_v$. We will denote by $C_v$ the restriction of the cutset $C$ to $T_v$.

By induction on the structure of $T_\rho$, we will now show that for any vertex $v$ in $T_\rho$ which is at a distance $\delta_v$ from $\rho$, and has arity $d_v$, one has

$$|\phi(F_v(\sigma)) - \phi(F_v(\tau))|^q \le M^q \sum_{u \in C_v} \alpha^{|u| - \delta_v}. \tag{4.12}$$

To see that this implies the claim of the lemma, we observe that since $F_\rho(\sigma)$ and $F_\rho(\tau)$ are in the interval $[0, b]$, we have $|F_v(\sigma) - F_v(\tau)| \leq \frac{1}{L}|\phi(F_v(\sigma)) - \phi(F_v(\tau))|$. Hence, taking $v = \rho$ in eq. (4.12), the claim of the lemma follows from the above observation.

We now proceed to prove eq. (4.12). The base case of the induction consists of vertices $v$ which are either of arity $0$ or which are in $C$. In the first case (which includes the case where $v$ is fixed by both the initial conditions to the same value), we clearly have $F_v(\sigma) = F_v(\tau)$, and hence the claim is trivially true. In the second case, we have $C_v = \{v\}$, and all the children of $v$ must lie in $C'$. Thus, in this case, the claim is true by the definition of $M$.

We now proceed to the inductive case. Let $v_1, v_2, \ldots v_{d_v}$ be the children of $v$, which satisfy eq. (4.12) by induction. In the remainder of the proof, we suppress the dependence of $\xi$ on $\phi$ and $q$. Applying Lemma 4.3 followed by the induction hypothesis, we then have for some positive integer $k \leq d_v$

$$|\phi(R_v(\sigma)) - \phi(R_v(\tau))|^q \leq \xi(k) \sum_{i=1}^{d_v} |\phi(R_{v_i}(\sigma)) - \phi(R_{v_i}(\tau))|^q \text{, using Lemma 4.3}$$

$$\leq M^q \xi(k) \sum_{i=1}^{d_v} \sum_{u \in C_{v_i}} \alpha^{|u| - \delta_{v_i}} \text{, using the induction hypothesis}$$

$$\leq M^q \sum_{u \in C_v} \alpha^{|u| - \delta_v} \text{, using } \xi(k) \leq \alpha \text{ and } \delta_{v_i} = \delta_v + 1.$$

This completes the induction. $\qquad\square$

## 4.7 Proofs omitted from Section 4.3

### 4.7.1 Maximum of $\nu_\lambda$

In this section, we prove Lemma 4.9.

*Proof of Lemma 4.9.* We first prove that given $\lambda$, $\tilde{x}_\lambda(d)$ is a *decreasing* function of $d$. For ease of notation, we suppress the dependence of $\tilde{x}_\lambda(d)$ on $d$ and $\lambda$. From Lemma 4.8, we know that $\tilde{x}$ is the unique positive solution of $d\tilde{x} = 1 + f_d(\tilde{x})$. Differentiating the equation with respect to $d$ (and denoting $\frac{d\tilde{x}}{dd}$ by $\tilde{x}'$), we have

$$\tilde{x} + d\tilde{x}' = -f_d(\tilde{x}) \left[ \frac{d\tilde{x}'}{1 + \tilde{x}} + \log(1 + \tilde{x}) \right]$$

which in turn yields

$$\tilde{x}' = -\frac{(1 + \tilde{x})\left[f_d(\tilde{x})\log(1 + \tilde{x}) + \tilde{x}\right]}{d(1 + d)\tilde{x}}. \tag{4.13}$$

Since $\tilde{x} \geq 0$, this shows that $\tilde{x}$ is a decreasing function of $d$.

We now consider the derivative of $\nu_\lambda(d)$ with respect to $d$. Recalling that $\nu_\lambda(d) = \xi(d) = \Xi(d, \tilde{x}_\lambda(d))$ and then using the chain rule, we have

$$
\begin{aligned}
\nu'_\lambda(d) &= \Xi^{(1,0)}(d, \tilde{x}) + \Xi^{(0,1)}(d, \tilde{x})\frac{d\tilde{x}}{dd} \\
&= \Xi^{(1,0)}(d, \tilde{x}), \text{ since } \Xi^{(0,1)}(d, \tilde{x}) = 0 \text{ by definition of } \tilde{x} \\
&= \Xi(d, \tilde{x})\left[\frac{q-1}{d} - \frac{q\log(1+\tilde{x})}{2(1 + f_{d,\lambda}(\tilde{x}))}\right] \\
&= \frac{q\Xi(d, \tilde{x})}{d}\left[1 - \frac{1}{q} - \frac{\log(1+\tilde{x})}{2\tilde{x}}\right].
\end{aligned}
\tag{4.14}
$$

Here, we use $1 + f_{d,\lambda}(\tilde{x}) = d\tilde{x}$ to get the last equality. We now note that the quantity inside the square brackets is an increasing function of $\tilde{x}$, and hence a decreasing function of $d$. Since $\Xi(d, \tilde{x})$ is positive, this implies that there can be at most one positive zero of $\nu_\lambda(d)$, and if such a zero exists, it is the unique maximum of $\nu_\lambda(d)$.

We now complete the proof by showing that $\nu'_\lambda(d) = 0$ for $d = \Delta_c(\lambda)$. At such a $d$, we have $\lambda = \lambda_c(d) = \frac{d^d}{(d-1)^{d+1}}$. We then observe that $\tilde{x}(d) = \frac{1}{d-1}$, since

$$
1 + f_d(\tilde{x}) = 1 + \frac{d^d}{(d-1)^{d+1}} \cdot \frac{(d-1)^d}{d^d} = \frac{d}{d-1} = d\tilde{x}.
$$

As an aside, we note that this is not a coincidence. Indeed, when $\lambda = \lambda_c(d)$, $\tilde{x}$ as defined above is well known to be the unique fixed point of $f_d$, and the potential function $\Phi$ was chosen in [LLY13] in part to make sure that at the critical activity, the fixed point is also the maximizer of (an analog of) $\Xi(d, \cdot)$.

We now substitute the value of $\frac{1}{q}$ and $\tilde{x}$ at $d = \Delta_c$ to verify that

$$
\nu_\lambda(\Delta_c) = \frac{q\Xi\left(\Delta_c, \frac{1}{\Delta_c-1}\right)}{2\Delta_c}\left[(\Delta_c - 1)\log\left(1 + \frac{1}{\Delta_c - 1}\right) - (\Delta_c - 1)\log\left(1 + \frac{1}{\Delta_c - 1}\right)\right] = 0,
$$

as claimed. Substituting these values of $d$ and $\tilde{x}$, along with the earlier observation that $f_{\Delta_c}(\tilde{x}) = \tilde{x} = \frac{1}{\Delta_c-1}$, into the definition of $\nu_\lambda$, we have

$$
\begin{aligned}
\nu_\lambda(\Delta_c) = \Xi\left(\Delta_c, \frac{1}{\Delta_c - 1}\right) &= \Delta_c^{q-1}\left(\frac{\tilde{x}}{1+\tilde{x}}\frac{f_{\Delta_c}(\tilde{x})}{1 + f_{\Delta_c}(\tilde{x})}\right)^{q/2} \\
&= \frac{1}{\Delta_c},
\end{aligned}
$$

which completes the proof. $\qquad\square$

### 4.7.2 Symmetrizability of the message

In this section, we prove Lemma 4.7. We start with the following technical lemma.

**Lemma 4.14.** *Let $r \geq 1$, $0 < A < 1$, $\gamma(x) := (1-x)^r$ and $g(x) := \gamma(Ax) + \gamma(A/x)$. Note that $g(x) = g(1/x)$, and $g$ is well defined in the interval $[A, 1/A]$. Then all the maxima of the function $g$ in the interval $[A, 1/A]$ lie in the set $\{1/A, 1, A\}$.*

Before proving the lemma, we observe the following simple consequence. Consider $0 \leq s_1, s_2 \leq 1$ such that $s_1 s_2$ is constrained to be some fixed constant $C < 1$. Then, applying the lemma with $A = \sqrt{C}$ we see that $\gamma(s_1) + \gamma(s_2)$ is maximized either when $s_1 = s_2$ or when one of them is 1 and the other is $C$.

*Proof of Lemma 4.14.* Note that when $r = 1$, $g(x) = 2 - A(x + 1/x)$, which is maximized at $x = 1$. We therefore assume $r > 1$ in the following.

We consider the derivative $g'(x) = Ar\left[(1 - A/x)^{r-1}\frac{1}{x^2} - (1 - Ax)^{r-1}\right]$. Note that $g(x) = g(1/x)$ and that $g'(x)$ and $g'(1/x)$ have opposite signs, so it is sufficient to study $g$ in the range $[1, 1/A]$. We now note that in the interior of the intervals of interest $g'$ always has the same sign as

$$h(x) := Ax^{t+1} - x^t + x - A,$$

where $t := \frac{r+1}{r-1} > 1$ for $r > 1$. We therefore only need to study the sign of $h$ in the interval $I := [1, 1/A]$. We note that $h(1) = 0$, and consider the derivatives of $h$.

$$h'(x) = A(t+1)x^t - tx^{t-1} + 1,$$
$$h''(x) = t(t+1)x^{t-2}\left[Ax - \frac{1}{r}\right].$$

Note that $h'(1) = (t+1)[A - 1/r]$. We now break the analysis into two cases.

**Case 1:** $A \geq 1/r$. In this case, we have $h''(x) > 0$ for $x$ in the interior of the interval $I$, and $h'(1) \geq 0$. This shows that $h'(x) > 0$ for $x$ in the interior of $I$, so that $h$ is strictly increasing in this interval. Since $h(1) = 0$, this shows that $h$ (and hence $g'$) are positive in the interior of $I$. Thus, $g$ is maximized in $I$ at $x = 1/A$.

**Case 2:** $A < 1/r$. We now have $h'(1) < 0$ and $h''(1) < 0$. Further, defining $x_0 = \frac{1}{Ar}$, we see that $h''$ is negative in $[1, x_0)$ and positive in $(x_0, 1/A]$ (and 0 at $x_0$). Since $h'(1) < 0$, this shows that $h'$ is negative in $[1, x_0]$, and hence can have no zeroes there. Further, we see that $h'$ is strictly increasing in $[x_0, 1/A]$, and hence can have at most one zero $x_1$ in $[x_0, 1/A]$.

If no such zero exists, then $h'$ is negative in $I$. In this case, we see that $h$ (and hence $g'$) is negative in the interior of $I$, and hence $g$ is maximized at $x = 1$. We now consider the case where there is a zero $x_1$ of $h'$ in $[x_0, 1/A]$. By the sign analysis of $h''$, we know that $h'$ is negative in $[1, x_1)$ and positive in $(x_1, 1]$. We thus see that $h$ is decreasing (and negative) in $(1, x_1)$ and increasing in $(x_1, 1/A)$. It can therefore have at most one zero $x_2$ in $(x_1, 1/A]$.

If no such zero exists, then $h$ (and hence $g'$) is negative in the interior of $I$, and hence $g$ is maximized at $x = 1$. If such a zero $x_2$ exists in $(x_1, 1/A]$, then—because $h$ is increasing in $(x_1, 1/A)$ and negative in $(1, x_1]$—$h$ (and hence $g'$) is negative in $(1, x_2)$ and positive in $(x_2, 1/A)$, which shows that $g$ is maximized at either $x = 1$ or at $x = 1/A$.

<div align="right">□</div>

We now prove Lemma 4.7.

*Proof of Lemma 4.7.* We first verify the second condition in the definition of symmetrizability:

$$\lim_{x \to 0^+} \frac{1}{(1+x)\Phi(x)} = \lim_{x \to 0^+} 2\sqrt{\frac{x}{1+x}} = 0.$$

We now recall the program used in the definition of symmetrizability, with the definitions of $\Phi$ and $f_d$ substituted, and with $r = a/2$:

$$\max \quad \sum_{i=1}^{d} \left(\frac{x}{1+x}\right)^r, \quad \text{where}$$

$$\lambda \prod_{i=1}^{d} \frac{1}{1+x_i} = B \tag{4.15}$$

$$x_i \geq 0, \quad 1 \leq i \leq d$$

Note that eq. (4.15) implies that $x_i \leq \lambda/B - 1$, so that the feasible set is compact. Thus, if the feasible set is non-empty, there is at least one (finite) optimal solution to the program. Let $\boldsymbol{y}$ be such a solution. Suppose without loss of generality that the first $k$ co-ordinates of $\boldsymbol{y}$ are non-zero while the rest are 0. We claim that $y_i = y_j \neq 0$ for all $1 \leq i \leq j$ and $y_i = 0$ for $i > k$.

To show this, we first define another vector $\boldsymbol{s}$ by setting $s_i = \frac{1}{1+x_i}$. Note that $s_i = s_j$ if and only if $x_i = x_j$ and $s_i = 1$ if and only if $x_i = 0$. Note that the constraint in eq. (4.15) is equivalent to

$$\prod_{i=1}^{d} s_i = B/\lambda. \tag{4.16}$$

Now suppose that there exist $i \neq j$ such that $y_i y_j \neq 0$ and $y_i \neq y_j$. We then have $s_i \neq s_j$ and $0 < s_1, s_2 < 1$. Now, since $r = a/2 \geq 1$ when $a \geq 2$, Lemma 4.14 implies that at least one of the following two operations, performed while keeping the product $s_i s_j$ fixed (so that the constraints in eqs. (4.15,4.16) are satisfied), will increase the value of the sum $\gamma(s_i) + \gamma(s_j) = \left(\frac{y_i}{1+y_i}\right)^r + \left(\frac{y_j}{1+y_j}\right)^r$:

1. Making $s_i = s_j$, or

2. Making $y_i = 0$ (so that $s_i = 1$).

Thus, if $\boldsymbol{y}$ does not have all its non-zero entries equal, we can increase the value of the objective function while maintaining all the constraints. This contradicts the fact that $\boldsymbol{y}$ is a maximum, and completes the proof.

<div align="right">□</div>

## 4.8  Symmetrizability for the monomer-dimer model

In this section, we prove Lemma 4.11. As in the case of the hard core model, we begin with an auxiliary technical lemma.

**Lemma 4.15.** *Let $r$ and $a$ satisfy $1 < r \le 2$ and $0 < a < 1$ respectively. Consider the functions $\gamma(x) := x^r(2-x)^r$ and $g(x) := \gamma(a-x) + \gamma(a+x)$. Note that $g$ is even and is well defined in the interval $[-A, A]$, where $A := \min(a, 1-a)$. Then all the maxima of the function $g$ in the interval $[-A, A]$ lie in the set $\{-a, 0, a\}$.*

The lemma has the following simple consequence. Let $0 \le s_1, s_2 \le 1$ be such that $(s_1 + s_2)/2$ is constrained to be some fixed constant $a \le 1$. Then, applying the lemma with $s_1 = a - x$, $s_2 = a + x$, we see that $\gamma(s_1) + \gamma(s_2)$ is maximized either when $s_1 = s_2 = a$ or when one of them is $0$ and the other is $2a$ (the second case can occur only when $a \le 1/2$).

*Proof of Lemma 4.15.* Since $g$ is even, we only need to analyze it in the interval $[0, A]$, and show that restricted to this interval, its maxima lie in $\{0, a\}$.

We begin with an analysis of the third derivative of $\gamma$, which is given by

$$\gamma'''(x) = -4r(r-1)(1-x)(1-(1-x)^2)^{r-2}\left[\frac{3-(2r-1)(1-x)^2}{1-(1-x)^2}\right]. \qquad (4.17)$$

Our first claim is that $\gamma'''$ is strictly increasing in the interval $[0, 1]$ when $1 < r \le 2$. In the case when $r = 2$, the last two factors in eq. (4.17) simplify to constants, so that $\gamma'''(x) = -12r(r-1)(1-x)$, which is clearly strictly increasing. When $1 < r < 2$, the easiest way to prove the claim is to notice that each of the factors in the product on the right hand side of is a strictly increasing non-negative function of $y = 1-x$ when $x \in [0, 1]$ (the fact that the second and third factors are increasing and non-negative requires the condition that $r < 2$). Thus, because of the negative sign, $\gamma'''$ itself is a strictly *decreasing* function of $y$, and hence a strictly *increasing* function of $x$ in that interval.

We can now analyze the behavior of $g$ in the interval $[0, A]$. We first show that when $a > 1/2$, so that $A = 1 - a \ne a$, $g$ does not have a maximum at $x = A$ when restricted to $[0, A]$. We will achieve this by showing that when $1 > a > 1/2$, $g'(1-a) < 0$. To see this, we first compute $\gamma'(x) = 2r(r-1)x^{r-1}(2-x)^{r-1}(1-x)$, and then observe that

$$g'(1-a) = \gamma'(1) - \gamma'(2a-1)$$
$$= -\gamma'(2a-1) < 0, \text{ since } 0 < 2a - 1 < 1.$$

We now start with the observation that $g'''(x) = \gamma'''(a+x) - \gamma'''(a-x)$, so that because of the strict monotonicity of $\gamma'''$ in $[0, 1]$ (which contains the interval $[0, A]$), we have $g'''(x) > 0$ for $x \in (0, A]$. We note that this implies that $g''(x)$ is strictly increasing in the interval $[0, A]$. We also note that $g'(0) = 0$. We now consider two cases.

**Case 1:** $g''(0) \geq 0$ Using the fact that $g''(x)$ is strictly increasing in the interval $[0, A]$ we see that $g''(x)$ is also positive in the $(0, A]$ in this case. This, along with the fact that $g'(0) = 0$, implies that $g'(x) > 0$ for $x \in (0, A]$, so that $g$ is strictly increasing in $[0, A]$ and hence is maximized only at $x = A$. As proved above, this implies that the maximum of $g$ must be attained at $x = a$ (in other words, the case $g''(0) \geq 0$ cannot arise when $a > 1/2$ so that $A = 1 - a \neq a$).

**Case 2:** $g''(x) < 0$ Again, using the fact that $g''(x)$ is strictly increasing in $[0, A]$, we see that there is at most one zero $c$ of $g''$ in $[0, A]$. If no such zero exists, then $g''$ is negative in $[0, A]$, so that $g'$ is strictly decreasing in $[0, A]$. Since $g'(0) = 0$, this implies that $g'$ is also negative in $(0, A)$ so that the unique maximum of $g$ in $[0, A]$ is attained at $x = 0$.

Now suppose that $g''$ has a zero $c$ in $(0, A]$. As before, we can conclude that $g'$ is strictly negative in $[0, c]$, and strictly increasing in $[c, A]$. Thus, if $g'(A) < 0$, $g'$ must be negative in all of $(0, A]$, so that $g$ is again maximized at $x = 0$ as in Case 1. The only remaining case is when there exists a number $c_1 \in (c, A]$ such that $g'$ is negative in $(0, c_1)$ and positive in $(c_1, A]$. In this case, we note that $g'(A) \geq 0$, so that—as observed above—we cannot have $A \neq a$. Further, the maximum of $g$ in this case is at $x = 0$ if $g(0) > g(A)$, and at $x = A$ otherwise. Since we already argued that $A$ must be equal to $a$ in this case, this shows that the maxima of $g$ in $[0, A]$ again lie in the set $\{0, a\}$. $\qquad \square$

We now prove Lemma 4.11.

*Proof of Lemma 4.11.* We first verify the second condition in the definition of symmetrizability:

$$\lim_{p_i \to 0} \frac{1}{\Phi(p_i)} \left| \frac{\partial f_{d,\gamma}}{\partial p_i} \right| = \lim_{p_i \to 0} \frac{\gamma p_i (2 - p_i)}{\left( 1 + \gamma \sum_{j=1}^{d} p_j \right)^2} = 0.$$

We now recall the program used in the definition of symmetrizability with respect to exponent $r$, with the definitions of $\Phi$ and $f_{d,\gamma}$ substituted:

$$\max \quad \gamma^r f_{d,\gamma}(\boldsymbol{p})^{2r} \sum_{i=1}^{d} p_i^r (2 - p_i)^r, \quad \text{where}$$

$$\frac{1}{1 + \gamma \sum_{i=1}^{d} p_i} = B$$

$$0 \leq p_i \leq 1, \quad 1 \leq i \leq d$$

Since we are only interested in the values of $\boldsymbol{p}$ solving the program, we can simplify the program

as follows:

$$\max \quad \sum_{i=1}^{d} p_i^r (2 - p_i)^r, \qquad \text{where}$$

$$\sum_{i=1}^{d} p_i = B' := \frac{1 - B'}{\gamma B}$$

$$0 \leq p_i \leq 1, \qquad 1 \leq i \leq d$$

We see that the feasible set is compact. Thus, if it is also non-empty, there is at least one (finite) optimal solution to the program. Let $\boldsymbol{y}$ be such a solution. Suppose without loss of generality that the first $k$ co-ordinates of $\boldsymbol{y}$ are non-zero while the rest are $0$. We claim that $y_i = y_j \neq 0$ for all $1 \leq i \leq j \leq k$.

For if not, let $i \neq j$ be such that $y_i y_j \neq 0$ and $y_i \neq y_j$. Let $y_i + y_j = 2a$. The discussion following Lemma 4.15 implies that at least one of the following two operations, performed while keeping the sum $y_i + y_j$ fixed and ensuring that $y_i, y_j \in [0, 1]$ (so that all the constraints in the program are still satisfied), will increase the value of the sum $\gamma(y_i) + \gamma(y_j) = y_i^r(2 - y_i)^r + y_j^r(2 - y_j)^r$:

1. Making $y_i = y_j$, or

2. Making $y_i = 0$ (so that $y_j = a$).

Thus, if $\boldsymbol{y}$ does not have all its non-zero entries equal, we can increase the value of the objective function while maintaining all the constraints. This contradicts the fact that $\boldsymbol{y}$ is a maximum, and completes the proof. $\qquad \square$

# Chapter 5

# Related work and open problems

We now conclude Part I with a discussion of previous research and open questions concerning the approximation of partition functions and decay of correlations in spin systems. Approximating the partition function has traditionally been studied in the framework of Markov Chain Monte Carlo (MCMC) methods. For the hard core model on bounded degree graphs, this line of work culminated in papers by Dyer and Greenhill [DG00a] and Vigoda [Vig01], who gave MCMC based FPRASs for $\lambda < 2/(d-1)$ for graphs of maximum degree at most $d+1$. Weitz [Wei06] (see also [BG08]) introduced a new paradigm by using correlation decay directly to design a *deterministic* FPTAS and gave an algorithm under the condition $\lambda < \lambda_c(d)$ for graphs of degree at most $d+1$; this range of applicability was later proved to be optimal by Sly [Sly10] (see also [SS12, GGŠ⁺14]). To date, no MCMC based algorithm is known to have a range of applicability as wide as Weitz's algorithm. The same is true for the anti-ferromagnetic Ising model, where again the correlation decay based algorithms described in Chapter 3 have the best (and in light of the results of Sly and Sun [SS12], optimal, unless NP = RP) range of applicability.

Interestingly, however, the situation is reversed if we consider the *ferromagnetic* Ising model, for which Jerrum and Sinclair [JS93] gave an MCMC based FPRAS which is valid for all graphs at all values of the edge and vertex activities. In contrast, the methods of the last few chapters cannot be used to get an FPTAS even in the uniqueness region, since, as observed by Weitz [Wei06], weak spatial mixing does not imply strong spatial mixing for the ferromagnetic Ising model (see [SST14] for a proof). A similar situation obtains for the case of the monomer-dimer model. Again, there is an MCMC based FPRAS for the partition function (due to Jerrum and Sinclair [JS89]) valid for all graphs for all values of the edge activity. However, the best known deterministic approximation algorithms are those described by Bayati *et al.* [BGK⁺07] and in Chapter 4, which require bounds on the maximum degree and the connective constant respectively.

Nevertheless, the main feature of Weitz's method—that of establishing strong spatial mixing on all graphs of degree $d+1$ whenever weak spatial mixing holds on the $d$-ary tree—is of great interest in its own right. For example, it improved upon the known range of the fugacity $\lambda$ under which the hard core model exhibits strong spatial mixing on specific lattices like $\mathbb{Z}^2$. The phenomenon of strong spatial mixing itself has its origin in the study of fast mixing of Markov chains [MO94a, MO94b], and Weitz's results automatically led to improvements in the range of $\lambda$

under which Markov chains for sampling from the hard core Gibbs measure on $\mathbb{Z}^2$ were known to exhibit fast mixing. Since $\mathbb{Z}^2$ has degree 4, Weitz's result implies that strong spatial mixing holds when $\lambda < \lambda_c(3) = 1.6875$. Restrepo, Shin, Tetali, Vigoda and Yang [RST$^+$13] extended Weitz's approach using sophisticated computational methods tailored to $\mathbb{Z}^2$ to improve the bound to $\lambda < 2.38$. Their methods were later extended by Vera, Vigoda and Yang [VVY13], who improved the bound further to $\lambda < 2.48$.

However, none of the above mentioned results could handle even special classes of unbounded degree graphs. Thus, the problem of sampling from the hard core model on graphs drawn from $G(n, d/n)$ (which have unbounded degree) was studied using MCMC methods by Mossel and Sly [MS09] and more recently by Efthymiou [Eft14], who gave a fast MCMC based sampler for $\lambda < 1/(2d)$. In contrast, Corollary 4.6 implies that a polynomial time correlation decay based sampler exists which works up to $\lambda < \lambda_c(d)$. Note that $\lambda_c(d) > e/d$, so that the above condition subsumes the bound $\lambda < e/d$ conjectured in [Eft14]. Nevertheless, it must be mentioned that in spite of their smaller range of (provable) applicability, MCMC based samplers often have the advantage that when fast mixing holds, their runtime for producing a sample that is at most $\epsilon$-far from the Gibbs distribution scales only as $\mathrm{poly}\left(\log(1/\epsilon)\right)$, where as that of correlation decay based samplers scales as $\mathrm{poly}\left(1/\epsilon\right)$.

In contrast to the case of the hard core model, much more progress has been made on relating spatial mixing to notions of average degree in the case of the zero field *ferromagnetic* Ising model. Lyons [Lyo89] demonstrated that on an arbitrary tree, a quantity similar in flavor to the connective constant, known as the *branching factor*, exactly determines the threshold for uniqueness of the Gibbs measure for this model. For the ferromagnetic Ising model on general graphs, Mossel and Sly [MS09, MS13] proved results analogous to our Theorem 4.5. An important ingredient in the arguments in both [Lyo89] and [MS09, MS13] relating correlation decay in the zero field Ising model to the branching factor and the connective constant is the symmetry of the "+" and "−" spins in the zero field case. In work related to [Lyo89], Pemantle and Steif [PS99] define the notion of a *robust phase transition (RPT)* and relate the threshold for RPT for various "symmetric" models such as the zero field Potts model and the Heisenberg model on general trees to the branching factor of the tree. Again, an important ingredient in their arguments seems to be the existence of a symmetry on the set of spins under whose action the underlying measure remains invariant. In contrast, in the hard core model, the two possible spin states of a vertex ("occupied" and "unoccupied") do not admit any such symmetry.

Since the publication of Weitz's paper (and of [BG08]), the correlation decay method has been applied to several other problems. In addition to the examples in the last few chapters, these include applications to general two spin systems [LLY13], to counting colorings [GK07, LY13] and to counting edge covers [LLL13]. Nevertheless, extending the tight connections between correlation decay on trees and algorithms for approximating the partition function on bounded degree graphs for spin systems with *three* or more spins remains open. The state of the problem of counting colorings illustrates the situation nicely. The relevant parameters here are the number of colors $q$, and the maximum degree $d$ of the graph to be colored. A fairly simple path-coupling argument gives an MCMC based FPRAS for the problem under the condition $q > 2d$ [Jer95]. A more elaborate MCMC based algorithm due to Vigoda [Vig00] gives a sampler under the

condition $q < 11d/6$. In contrast, the best known correlation decay based algorithms still require $q > 2.58d$ [LY13]. The main obstacle in getting close to the best known MCMC based bounds—and ideally, improving upon them—is that no crisp reductions to the self-avoiding walk tree of the kind afforded by the theorems of Godsil (Theorem 2.1) and Weitz (Theorem 2.2) are known for spin systems with more than two spins.

# Part II

# Hardness of computing averages

# Chapter 6

# Averages and zeros of polynomials

In contrast to the first part of this thesis, this chapter and the two following it deal with complexity theoretic questions about *exact* computation. More specifically, we will consider the question of exactly computing the mean of some natural observables of the ferromagnetic Ising and the monomer-dimer models. Our approach to these problems will be based on the beautiful results of Lee and Yang [LY52] and Heilmann and Lieb [HL72] characterizing the location of the complex zeros of the partition function. In this chapter, we set up the problems formally and introduce our basic framework for bringing the zeros of the partition function into play. The following two chapters then instantiate the approach for the Ising and monomer-dimer models respectively.

## 6.1   Mean observables

We begin by recalling our definitions of partition functions and observables, and then describing the specific examples that we will study. Recall that we introduced partition functions in the following framework. We are given as input a graph $G = (V, E)$ which implicitly defines a set $\Omega = \Omega(G)$ of combinatorial structures, or *configurations* (such as matchings in $G$, or $k$-colorings of its vertices). We further have a weight function $w : \Omega \to \mathbb{R}^+$ that assigns a positive weight to every element $\sigma \in \Omega$, giving rise to a probability distribution $\pi(\sigma) = w(\sigma)/Z$; the normalizing factor $Z := \sum_{\sigma \in \Omega} w(\sigma)$ is called the *partition function*.

An *observable* is a non-negative function $f : \Omega \to \mathbb{R}^+ \cup \{0\}$, such that $f(\sigma)$ is easily computable for each $\sigma \in \Omega$. Our goal is to compute the *mean* of $f$ with respect to $\pi$, i.e.,

$$\langle f \rangle := \sum_\sigma \pi(\sigma) f(\sigma) = \frac{\sum_\sigma w(\sigma) f(\sigma)}{Z}.$$

We now describe the specific observables that we will study, and their relationship with the partition function.

### 6.1.1 The ferromagnetic Ising model: Mean magnetization

Recall that in the ferromagnetic Ising model, the configurations are assignments of spin values $\{+, -\}$ to the vertices of $G$, i.e., $\Omega = \{+, -\}^V$. Recall that the weight of a configuration $\sigma$ is

$$w_I(\sigma) := \beta^{d(\sigma)} \lambda^{p(\sigma)}, \tag{6.1}$$

where $d(\sigma)$ is the number of *disagreements* in $\sigma$ (i.e., the number of edges $\{u, v\} \in E$ with $\sigma(u) \neq \sigma(v)$), and $p(\sigma)$ is the number of vertices $v \in V$ with $\sigma(v) = +$. Note that the model has two parameters: the *edge potential* $\beta$, satisfying $0 < \beta \leq 1$, which governs the strength of the interaction between neighboring spins; and the *vertex activity* $\lambda > 0$, which specifies the tendency for spins to be $+$. Thus, when $\beta < 1$, this Gibbs distribution favors agreement between neighboring spins, so that the interaction is indeed ferromagnetic. Similarly, the distribution favors '$+$' (respectively, '$-$') spins when $\lambda > 1$ (respectively, when $\lambda < 1$). The partition function $Z_I(G, \beta, z)$ is the sum of the weights in eq. (6.1).[1]

The observable we will be most concerned with here is the *magnetization* $p(\sigma)$, which is just the number of $+$-spins in $\sigma$. Its average, the *mean magnetization*, is a fundamental quantity in statistical physics:

$$\langle p \rangle := \frac{\sum_\sigma w_I(\sigma) p(\sigma)}{Z_I}.$$

Other widely studied averages include the *mean energy* $\langle d \rangle$ (the average size of the cut between $+$-spins and $-$-spins) and the *susceptibility* $\chi := \langle p^2 \rangle - \langle p \rangle^2$ (the variance of the magnetization).

### 6.1.2 The monomer-dimer model: Average monomer count

The configurations $\Omega$ in this case are all matchings (independent sets of edges) in $G$. We consider a more general definition of the weights than that used in previous chapters; the weight weight of a matching $\sigma$ is now given by

$$w_M(\sigma) := \lambda^{u(\sigma)} \prod_{e \in \sigma} \gamma_e, \tag{6.2}$$

where $u(\sigma)$ is the number of unmatched vertices (*monomers*) in $\sigma$. The parameter $\lambda > 0$ is the vertex weight (or *monomer activity*), while for each edge $e \in E$, $\gamma_e$ is an edge weight (or *dimer activity*). (In the formulation used in previous chapters, $\lambda$ was 1 and all $\gamma_e$ were equal.) The Gibbs distribution $\pi(\sigma) = w_M(\sigma)/Z_M$ is a natural weighted distribution on matchings, and the partition function $Z_M := Z_M(G, \{\gamma_e\}_{e \in E}, \lambda)$ is the weighted matching polynomial of $G$.

A natural observable here is $u(\sigma)$, the number of unmatched vertices (or monomers). Note that $(|V| - \langle u \rangle)/2$ is just the average size of a (weighted) matching in $G$ (or equivalently, the average number of dimers).

---

[1] In this chapter, we will often explicitly include the graph $G$ itself as a parameter to the partition function. Further, to distinguish between the partition functions of different models, we will rely on subscripts; i.e., $Z_I$ will denote the partition function of the Ising model, $Z_M$ that of the monomer-dimer model and $Z_S$ that of a general two spin system.

### 6.1.3 The partition function and mean observables

Part of the importance of the partition function comes from the fact that various important mean observables can be written in terms of its derivatives. This property will be of crucial importance in our technical development, so we describe it here in some detail.

Consider first the Ising model partition function $Z_I$ as a polynomial in $\lambda$ (actually a bivariate polynomial in $\lambda$ and $\beta$):

$$Z_I = \sum_{k=0}^{|V|} \alpha_k \lambda^k, \qquad \text{where } \alpha_k = \sum_{\sigma:p(\sigma)=k} \beta^{d(\sigma)}.$$

The mean magnetization then becomes

$$\langle p \rangle = \frac{\sum_k k\alpha_k \lambda^k}{Z_I} = \frac{\mathcal{D}Z_I}{Z_I}, \tag{6.3}$$

where $\mathcal{D}$ denotes the differential operator $\lambda\frac{\partial}{\partial\lambda}$. Similarly, the mean energy and susceptibility $\chi$ can be written

$$\langle d \rangle = \beta \frac{\frac{\partial}{\partial\beta}Z_I}{Z_I}; \qquad \chi := \langle p^2 \rangle - \langle p \rangle^2 = \frac{\mathcal{D}^2 Z_I}{Z_I} - \left(\frac{\mathcal{D}Z_I}{Z_I}\right)^2. \tag{6.4}$$

Similarly, for the monomer-dimer model the partition function $Z_M$ can be written as

$$Z_M = Z_M(G, (\gamma_e)_{e\in E}, \lambda) := \sum_{k=0}^{|V|} \alpha_k \lambda^k,$$

where $\alpha_k = \sum_{\sigma:u(\sigma)=k} \prod_{e\in\sigma} \gamma_e$ is a weighted sum over matchings with $k$ unmatched vertices. The average number of monomers is then

$$U = U(G, (\gamma_e)_{e\in E}, \lambda) := \langle u \rangle = \frac{\mathcal{D}Z_M}{Z_M}, \tag{6.5}$$

where $\mathcal{D}$ again denotes the differential operator $\lambda\frac{\partial}{\partial\lambda}$. Note that the *average dimer count* $D = D(G, (\gamma_e)_{e\in E}, \lambda)$ (equivalently, the average number of matchings) is simply given by

$$D(G, (\gamma_e)_{e\in E}, \lambda) = \frac{1}{2}\left(n - U(G, (\gamma_e)_{e\in E}, \lambda)\right),$$

where $n$ is the number of vertices in the graph $G$.

## 6.2 Partition function zeros and computational complexity

Our goal now is to prove that computing the mean magnetization and the average monomer count are both complexity theoretically hard. We begin with an informal overview of our approach.

It is tempting to argue that computing an average as in, say, (6.3) is at least as hard as computing the partition function $Z_I$, because (6.3) is a rational function and thus by evaluating it at a small number of points we could recover the numerator and denominator polynomials by rational interpolation. Since the partition function is #P-hard in almost all cases of interest (including $Z_I$ and $Z_M$ above at all but trivial values of the parameters), we would be done. In fact, such interpolation arguments are a staple of proofs of #P-hardness of partition functions.

However, in the context of averages, the problem with this argument is that, viewed as polynomials in the variable $\lambda$, $Z_I$ and its derivative $\mathcal{D}Z_I$ may have common factors (equivalently, viewed as polynomials in the complex variable $\lambda$, they may have common zeros); and in this case we are clearly not able to recover $Z_I$ by rational interpolation. Indeed it seems hard *a priori* to rule out the possibility that non-trivial interactions between $Z_I$ and its derivative could conspire to make the average much easier to compute than $Z_I$ itself. Thus we are naturally led to the following question:

**Question:** *Is it possible for the partition function $Z$ and its derivative to have common zeros?*[2]

If the answer is no, then we will be able to conclude that computing the average is as hard as computing $Z$ itself, and thus #P-hard in all interesting cases.

Our main goal will be to carry through this program using resolutions of the above question in several interesting cases. However, before proceeding, we mention a possible alternative approach to deal with the issue of repeated zeros. Since a generic polynomial does not have repeated zeros, one could try to argue that any given graph $G$ can be perturbed so that its partition function has distinct zeros, and so that the magnetization of the perturbed graph is close to the magnetization of the original graph. One could then perform the interpolation operations with respect to the perturbed graph, and hope that if the perturbations are small enough then the reduction still goes through. Indeed, this is our intuition for why the magnetization (and other averages) should be hard to compute.

However, it is not clear how to convert this intuition into a formal proof: in addition to a rather involved error analysis, this would require showing that the partition function of a "perturbed" Ising model behaves like a generic polynomial with respect to the structure of its zeros, which seems no easier than answering the Question above. Our approach sidesteps this issue by tackling the question directly, and in addition establishes a non-trivial property of the zeros of the partition function that may be of independent interest.

In line with the above program, we begin with two of the most well known results about location of zeros of partition functions. The first of these is the beautiful result of Lee and Yang [LY52] on the zeros of the partition function $Z_I(G, \beta, z)$ of the ferromagnetic Ising model. Before describing the result, we introduce a generalization of the Ising model in which the vertex

---

[2]Note that a common zero of $Z$ and $\mathcal{D}Z$ corresponds to a repeated zero of $Z$, so this question is equivalent to the question of whether $Z$ may have repeated zeros.

activities can vary across vertices of G. Suppose that the vertex activity at vertex $v$ is $z_v$. The weight of a configuration $\sigma$ is then defined as

$$w_I(\sigma) := \beta^{d(\sigma)} \prod_{v:\sigma(v)=+} z_v,$$

and the partition function is given by $Z_I(G, \beta, (z_v)_{v \in V}) = \sum_\sigma w_I(\sigma)$. We also take this opportunity to define the following related linear differential operator $\mathcal{D}_G$, which will be useful in the next chapter:

$$\mathcal{D}_G := \sum_{v \in V} z_v \frac{\partial}{\partial z_v}.$$

Note that using this operator, we can write the magnetization $M(G, \beta, (z_v)_{v \in V})$ in the generalized version of the Ising model exactly as in eq. (6.3):

$$M(G, \beta, (z_v)_{v \in V}) = \frac{\mathcal{D}_G Z_I(G, \beta, (z_v)_{v \in V})}{Z_I(G, \beta, (z_v)_{v \in V})}.$$

We now proceed to describe the Lee-Yang theorem.

**Theorem 6.1** ([LY52]). *Let $G$ be any undirected graph and suppose $0 < \beta \leq 1$. Then the complex zeros of $Z_I(G, \beta, z)$, considered as a polynomial in $z$, satisfy $|z| = 1$.*

Actually, Lee and Yang proved the following multivariate version of their theorem, the proof of which was later considerably simplified by Asano [Asa70].

**Theorem 6.2** ([LY52, Asa70]). *Let $G = (V, E)$ be a connected undirected graph, and suppose $0 < \beta < 1$. Suppose $(z_v)_{v \in V}$ is a set of complex valued vertex activities such that $|z_v| \geq 1$ for all $v \in V$, and $|z_u| > 1$ for at least one $u \in V$. Then $Z_I(G, \beta, (z_v)_{v \in V}) \neq 0$.*

Theorem 6.2 is readily seen to imply Theorem 6.1 by setting $z_v = z$ for all $v \in V$, and then using the symmetry $Z_I(G, \beta, z) = z^{|V|} Z_I(G, \beta, 1/z)$. We now consider the partition function of the monomer-dimer model. In [HL70], Heilmann and Lieb announced the following result (see [HL72] for the complete proof).

**Theorem 6.3** ([HL70, HL72]). *Let $G = (V, E)$ be any graph, and $(\gamma_e)_{e \in E}$ be a collection of positive real edge weights. The complex zeros of $Z_M(G, (\gamma_e)_{e \in E}, z)$, considered as a polynomial in $z$, satisfy $\Re(z) = 0$. Further, if $G$ contains a Hamiltonian path, all the zeros are simple. Here, $\Re(z)$ denotes the real part of the complex number $z$.*

We also record a standard fact about rational interpolation, which will be needed in our hardness reductions. While it is clear that it is not in general possible to determine all coefficients of a rational function given its values at any number of points, this can be done if we impose a few simple conditions, as stated in the following theorem.

**Theorem 6.4** ([**MD62**]). *Suppose* $R(x) = \frac{p(x)}{q(x)}$ *where* $\gcd(p(x), q(x)) = 1$ *and both* $p(x)$ *and* $q(x)$ *are of degree* $n$. *Suppose* $\tilde{p}(x)$ *and* $\tilde{q}(x)$ *are polynomials of degree at most* $n$ *satisfying*

$$\frac{\tilde{p}(x_i)}{\tilde{q}(x_i)} = R(x_i)$$

*for* $2n + 2$ *distinct values* $x_1, x_2, \ldots x_{2n+2}$. *Then there is a constant* $c$ *such that* $p(x) = c\tilde{p}(x)$ *and* $q(x) = c\tilde{q}(x)$.

Notice that given the evaluations at the points $x_i$ one can write down a system of $2n + 2$ homogeneous linear equations for the $2n + 2$ unknown coefficients of $p$ and $q$. The theorem then guarantees that this system has rank exactly $2n + 1$. Thus, since Gaussian elimination can be implemented to run in strongly polynomial time (see, e.g., [Edm67]), a polynomial time algorithm for evaluating $R$ immediately yields a polynomial time algorithm for determining *some* $\tilde{p}$ and $\tilde{q}$ satisfying the conditions of the above theorem. If we know at least one non-zero coefficient of $p$ or $q$, we can then determine the proportionality constant $c$, and hence $p$ and $q$ also, in time polynomial in $n$.

## 6.3 Conclusion

This chapter sets up the groundwork for the next two chapters, which will use the program introduced in Section 6.2 to prove hardness results for the mean magnetization of the Ising model and the average monomer-count of the monomer-dimer model respectively. In the former case, we will actually need to extend the Lee-Yang theorem (Theorem 6.1) to show that the partition function has no repeated zeros, while in the case of the monomer-dimer model, our reduction will be more combinatorial.

# Chapter 7

# The mean magnetization: Extending the Lee-Yang theorem

In this chapter, we prove the following hardness theorem for the exact computation of the mean magnetization of the ferromagnetic Ising model.

**Theorem 7.1.** *For any fixed $0 < \beta < 1$ and any fixed $\lambda \neq 1$, the problem of computing the mean magnetization of the Ising model on connected graphs is #P-hard. Moreover, the problem remains #P-hard even when the input is restricted to graphs of maximum degree at least $4$.*

Note that in the case $\lambda = 1$ the mean magnetization is trivially $|V|/2$ by symmetry. Theorem 7.1 confirms that in all non-trivial cases, the problem of computing the fundamental average quantity associated with the Ising model is as hard as it could possibly be. Furthermore, the result also holds for bounded degree graphs, which are relevant in the statistical physics setting. The result can also be extended to arbitrary ferromagnetic two-spin systems and to planar graphs: we discuss the details of these extensions in Section 7.4.

We will also prove a similar (but slightly weaker) result for the *susceptibility* of the Ising model.

**Theorem 7.2.** *For any fixed $0 < \beta < 1$, the problem of computing the susceptibility of the Ising model on connected graphs, when $\lambda$ is specified in unary[1], is #P-hard. Moreover, the problem remains #P-hard even when the input is restricted to graphs of maximum degree at least $3$.*

**Remark 7.1.** The requirement that $\lambda$ be part of the input seems to be an artifact of the rational interpolation operations we use in our proof. In particular, our proof of Theorem 7.1 shows hardness for *fixed* $\lambda$ by "simulating" different values of $\lambda$ by suitably modifying the graph. To adapt this reduction approach to prove hardness for susceptibility (at fixed values of $\lambda$) seems to require the polynomial time computation of magnetization as a subroutine. However, we conjecture that computing the susceptibility should be hard even for fixed values of $\lambda$ (including $\lambda = 1$).

---

[1]When $\lambda$ is a non-integer rational number, this means that the numerator and denominator of $\lambda$ are specified in unary.

The above complexity theoretic results will be obtained as almost immediate corollaries of our following extension of the Lee-Yang theorem.

**Theorem 7.3.** *Let $G = (V, E)$ be a connected graph, and suppose $0 < \beta < 1$. Then the zeros of the polynomial $\mathcal{D}Z_I(G, \beta, \lambda)$ (in $\lambda$) satisfy $|\lambda| < 1$.*

Since the Lee-Yang Theorem says that all zeros of $Z_I$ satisfy $|\lambda| = 1$, Theorem 7.3 immediately implies that $Z_I$ and $\mathcal{D}Z_I$ have no common zeros. Note that the restriction that $G$ be connected is needed: there exist disconnected graphs for which the conclusion of the theorem does not hold. A simple example is a graph consisting of two isomorphic disconnected subgraphs. For the same reason we require $\beta < 1$. We also note that standard facts from complex analysis (in particular, the Gauss-Lucas theorem) imply that the zeros of $\mathcal{D}Z_I$ lie in the convex hull of those of $Z_I$, and hence within the closed unit circle. The content of Theorem 7.3 is that they must lie in the *interior* of the circle. This refinement is of course crucial for our application.

In the next section, we give a proof of Theorem 7.3, and the proof of the complexity theoretic corollaries follow. However, before proceeding we first give an informal description of the structure of the proof. We will actually prove a more general result concerning the zeros of the multivariate partition function $Z_I(G, \beta, \{\lambda_v\}_{v \in V})$ which is analogous to the multivariate version of the Lee-Yang theorem (Theorem 6.2). Our proof is based on a delightful combinatorial proof of the Lee-Yang Theorem due to Asano [Asa70], which begins with the empty graph (which trivially satisfies the theorem) and builds the desired graph $G$ by repeatedly adding edges one at a time; by a careful induction one can show that the Lee-Yang property is preserved under each edge addition. Our proof follows a similar induction, but the argument is more delicate because we are working with the more complicated polynomial $\mathcal{D}Z_I$ rather than $Z_I$. In particular, in the inductive step we need to invoke a correlation inequality due to Newman [New74].

## 7.1 An extended Lee-Yang theorem

In this section we prove Theorem 7.3, our extension of the classical Lee-Yang theorem. Let $G = (V, E)$ be a connected graph with $|V| = n$ and $|E| = m$, with vertex activity $z_i$ at the $i$th vertex. When clear from the context , we will write $Z(G)$ and $M(G)$ for the partition function $Z_I(G, \beta, (z_v)_{v \in V})$ and the mean magnetization $M(G, \beta, (z_v)_{v \in V})$ of the Ising model on $G$. In terms of the linear operator $\mathcal{D}_G$ defined in Section 6.2, we then have $M(G) = \mathcal{D}_G Z(G)/Z(G)$.

For convenience, we will use the shorthand $Y' = \mathcal{D}_G Y$ (when $G$ is clear from the context) in this section. Notice that this is slightly non-standard, as this shorthand is usually used for the actual derivative. In particular, when all the $z_i$ are equal to $z$, we have $Y' = z \frac{\partial Y}{\partial z}$ with our notation. Also, observe that the operator $\mathcal{D}_G$ obeys the usual product rule: $(Y_1 Y_2)' = Y_1' Y_2 + Y_1 Y_2'$.

In our proof, we will also need the following generalization of the partition function. We call an assignment of positive integer valued *weights* $w : V \to \mathbb{Z}^+$ to the vertices of $G$ *legal* if $w(v)$ is at least equal to the degree of $v$, for all $v \in V$.

**Definition 7.1.** Let $w$ be a legal collection of weights for $G$. The *weighted partition function* $Z_w(G)$ is then defined as

$$Z_w(G) := \sum_{\sigma \in \{+,-\}^V} \beta^{d(\sigma)} \prod_{v:\sigma(v)=+} z_v^{w(v)}, \tag{7.1}$$

where, as before, $d(\sigma)$ is number of disagreeing edges in the configuration $\sigma$.

Notice that the multivariate Lee-Yang theorem (Theorem 6.2) holds also for the weighted partition function, since all the weights are positive integers and we are effectively just changing variables from $z_v$ to $z_v^{w(v)}$.

We will also need the following consequence of a correlation inequality of Newman [New74], whose proof can be found in Section 7.5.

**Theorem 7.4 ([New74, Theorem 3.2]).** *Let $G$ be any graph, and let $w$ be a legal collection of weights for $G$. Suppose $0 < \beta < 1$, and $|z_v| \geq 1$ for all $v \in V$ are such that $Z_w(G) \neq 0$. Then*[2]

$$\Re\left(M(G)\right) = \Re\left(\frac{Z_w'(G)}{Z_w(G)}\right) \geq n/2.$$

In the special case of real valued activities, the above theorem is equivalent to the well known Griffiths inequality [Gri67], which states the intuitive fact that in a ferromagnetic Ising model where all activities favor the $+$ spin, the magnetization must be at least $n/2$.

For ease of reference in the inductive proof, we give a name to the property we want to establish. Recall that when all the vertex activities are equal to $z$, the classical Gauss-Lucas theorem, together with the Lee-Yang theorem, implies that the zeros of the derivative $\mathcal{D}_G(Z(G))$ lie *on or inside* the unit circle. Our goal is to establish that they actually lie *inside* the unit circle. Accordingly, we use the following terminology:

**Definition 7.2 (Strict Gauss-Lucas property).** A graph $G = (V, E)$ has the *strict Gauss-Lucas property (SGLP)* if for every set of activities such that $|z_v| \geq 1$ for all $v \in V$, and every $0 < \beta < 1$, one has $\mathcal{D}_G Z(G) \neq 0$. The graph has the *weighted strict Gauss-Lucas property (WSGLP)* if for all legal weights $w$, $\mathcal{D}_G Z_w(G) \neq 0$ necessarily holds under the same conditions.

Note that WSGLP easily implies SGLP: we simply choose $w(v) = \Delta$ for all $v$, where $\Delta$ is the maximum degree of $G$. From WSGLP, we then have that whenever $|z_v| \geq 1$ and $0 < \beta < 1$, $\mathcal{D}_G Z_w(G) = \Delta \mathcal{D}_G Z(G) \neq 0$, and hence $\mathcal{D}_G Z(G) \neq 0$. Thus Theorem 7.3 is implied by the following more general statement.

**Theorem 7.5.** *Every connected graph has the weighted strict Gauss-Lucas property, and hence also the strict Gauss-Lucas property.*

---

[2]Recall that we are using here the slightly non-standard notation $Z_w'(G) = \mathcal{D}_G Z_w(G)$, as described at the beginning of this section.

We now proceed to prove Theorem 7.5, using induction on the number of edges in the graph $G$. We first consider the base case of a connected graph with a single edge.

**Lemma 7.6 (Base Case).** *Let $G$ be the graph consisting of a single edge connecting two vertices. Then $G$ has the weighted strict Gauss-Lucas property.*

*Proof.* In this case we have $Z_w(G) = z_1^{w_1} z_2^{w_2} + \beta(z_1^{w_1} + z_2^{w_2}) + 1$ and therefore $\mathcal{D}_G Z_w(G) = (w_1 + w_2) z_1^{w_1} z_2^{w_2} + \beta(w_1 z_1^{w_1} + w_2 z_2^{w_2})$, with $w_1, w_2 \geq 1$. When $|z_1|, |z_2| \geq 1$, the latter vanishes only if

$$w_1 + w_2 = \beta \left| \frac{w_2}{z_1^{w_1}} + \frac{w_1}{z_2^{w_2}} \right| \leq \beta \left( \frac{w_2}{|z_1^{w_1}|} + \frac{w_1}{|z_2^{w_2}|} \right) \leq \beta \left( w_1 + w_2 \right),$$

which cannot hold since $0 < \beta < 1$. $\qquad\square$

For the inductive case, we require two operations: adding a new vertex to the graph, and merging two existing vertices. These operations are formalized in the following lemmas.

**Lemma 7.7 (Adding a vertex).** *Suppose $G = (V, E)$ is a connected graph satisfying the weighted strict Gauss-Lucas property. Let $u$ be a vertex not in $V$. Then, the graph $G_1$ obtained by attaching the new vertex $u$ to any vertex (say $v_1$) of $G$ also has the weighted strict Gauss-Lucas property.*

**Lemma 7.8 (Merging vertices).** *Suppose $G = (V, E)$ is a connected graph satisfying the weighted strict Gauss-Lucas property. Consider any two vertices, say $v_1$ and $v_2$, in $G$ that are not connected by an edge. The graph $G_1$ obtained by merging $v_1$ and $v_2$ into a single vertex $v$ (while making all the edges incident on $v_1$ and $v_2$ incident on $v$) also has the weighted strict Gauss-Lucas property.*

**Remark 7.2.** Note that the merge operation produces a multigraph with parallel edges when the vertices $v_1$ and $v_2$ share common neighbors, since all the edges to the common neighbors now become incident on the new vertex $v$. However, as will become apparent in the proof of Theorem 7.5 below, the merge operation is never applied to vertices with common neighbors when the underlying graph being constructed is simple. Nevertheless, we note that the proof of Theorem 7.5 given below holds also when the graph $G$ is allowed to have parallel edges, and this latter fact will be useful in our extension to general two-state ferromagnetic spin systems in Section 7.4. We also note that self-loops can never be produced by the merge operation (since the merge operation is not applicable to vertices connected by an edge). Further, self-loops cannot make any non-trivial contribution to the weight of a configuration, and hence can be safely neglected for the purposes of computing the partition function and the magnetization.

Before proceeding with the proofs of the above lemmas, we show how to use them to prove Theorem 7.5.

*Proof of Theorem 7.5.* We will prove by induction on $m$ that any connected graph with at most $m$ edges satisfies WSGLP. By Lemma 7.6, this statement is true when $m = 1$. Now suppose that the statement is true when $m = k$, and consider any connected graph $G$ with $k + 1$ edges.

In case $G$ has a cycle, there exist vertices $u$ and $v$ such that the edge $\{u, v\}$ can be removed from $G$ to obtain a connected graph $H$. Since $H$ has at most $k$ edges, $H$ satisfies WSGLP by the

inductive hypothesis. Let $v_1$ be a vertex not in $G$. By Lemma 7.7, the graph $H \cup \{\{u, v_1\}\}$ satisfies WSGLP. We can now merge $v_1$ and $v$ to obtain $G$, which therefore satisfies WSGLP by Lemma 7.8.

In case $G$ is a tree, there exists an edge $\{u, v\}$ such that $v$ is of degree 1. Again, we obtain a connected graph $H$ with at most $k$ edges by removing the edge $\{u, v\}$. By the inductive hypothesis, $H$ satisfies WSGLP, and hence by Lemma 7.7, $G$ does too. This completes the induction. $\qquad\square$

We turn now to the proofs of Lemmas 7.7 and 7.8, for which we will need the following additional lemma.

**Lemma 7.9.** *Let $G$ be a connected graph. Fix any set $S$ of vertices of $G$, and let $Z_w^+(S)$ denote the partition function restricted to configurations on the subgraph $G - S$, with all the vertices in $S$ fixed to have spin $+$.[3] Consider any set of vertex activities satisfying $|z_v| \geq 1$ for $v \in G - S$. Then, for $0 < \beta < 1$ and any set of permissible weights on the vertices of $G$, we have $Z_w^+(S) \neq 0$ and $\Re\left(Z_w^+(S)'/Z_w^+(S)\right) \geq 0$. In particular, for any positive real $a$, we have $Z_w^+(S)' + aZ_w^+(S) \neq 0$.*

*Proof.* Observe that $Z_w^+(S)$ is proportional to the product of weighted partition functions on connected components of the graph $G - S$, where the activities on the vertices connected to $S$ in these components (of which there is at least one in each component) have *increased* in magnitude by a factor of at least $1/\beta > 1$. We can therefore conclude using Theorem 6.2 that $Z_w^+(S) \neq 0$. The second condition $\Re\left(Z_w^+(S)'/Z_w^+(S)\right) \geq 0$ then follows from Theorem 7.4 applied to $G - S$. $\qquad\square$

We first prove Lemma 7.8, since its proof is somewhat simpler.

*Proof of Lemma 7.8.* Consider any legal weight assignment on $G_1$. If the weight of $v$ in $G_1$ is $w_v$, we can write $w_v = w_1 + w_2$ such that the weight assignment giving weights $w_1$ and $w_2$ to $v_1$ and $v_2$ respectively is legal for $G$. By partitioning into four cases based on the spins of $v_1$ and $v_2$, we can write the corresponding weighted partition function $Z_w(G)$ and its derivative as

$$Z_w(G) = Az_1^{w_1}z_2^{w_2} + Cz_1^{w_1} + Dz_2^{w_2} + B; \tag{7.2}$$
$$Z_w(G)' = (A' + (w_1 + w_2)A)z_1^{w_1}z_2^{w_2} + (C' + w_1C)z_1^{w_1} + (D' + w_2D)z_2^{w_2} + B', \tag{7.3}$$

for polynomials $A$, $B$, $C$, $D$ in the remaining variables $z_i$. Notice that in the notation of Lemma 7.9, $A = Z_w^+(\{v_1, v_2\})$. Similarly, denoting the activity at the merged vertex by $z$, we have the following expressions for $G_1$:

$$Z_w(G_1) = Az^{w_1+w_2} + B; \tag{7.4}$$
$$Z_w(G_1)' = (A' + (w_1 + w_2)A)z^{w_1+w_2} + B', \tag{7.5}$$

with $A$ and $B$ as defined above. Now consider any fixing of the activities such that $|z_i| \geq 1$ for $i > 2$. Since $G$ satisfies the weighted strict Gauss-Lucas property, we get by setting $z_1 = z_2$ in eq. (7.3) that the (univariate) polynomial

$$(A' + (w_1 + w_2)A)z^{w_1+w_2} + (C' + w_1C)z^{w_1} + (D' + w_2D)z^{w_2} + B'$$

---

[3]Note that the terms in $Z_w^+(S)$ *include* the interactions between the (fixed) vertices in $S$ and the vertices in $G - S$, but *exclude* the vertex activities in $S$.

in $z$ has no zeros satisfying $|z| \geq 1$. Also, we know from Lemma 7.9 that $A' + (w_1 + w_2)A \neq 0$. Thus, we must have that the product of the zeros, $B'/(A' + (w_1 + w_2)A)$, satisfies

$$\left| \frac{B'}{A' + (w_1 + w_2)A} \right| < 1.$$

However, using eq. (7.5), this implies that if $|z_i| \geq 1$ for $i > 2$, then $Z_w(G_1)'$ can be zero only if $|z| < 1$, and hence $G_1$ satisfies the weighted strict Gauss-Lucas property.    □

Finally, we give the proof of Lemma 7.7.

*Proof of Lemma 7.7.* Note that any legal set of weights for $G_1$ can be obtained by adding one to the weight $w_1$ of $v_1$ in a legal set of weights $w$ of $G$, and then assigning $u$ an arbitrary weight $w_0 \geq 1$. With a slight abuse of notation, we denote these related weight assignments (one on $G$ and the other on $G_1$) by the same letter $w$. We now partition the terms in $Z_w(G)$ based on the spin of $v_1$ to get

$$\begin{aligned} Z_w(G) &= Az_1^{w_1} + B; \\ Z_w(G)' &= (A' + w_1 A)z_1^{w_1} + B', \end{aligned}$$

where $w_1$ is the weight of $v_1$ in $G$. Here, $A$, $B$ are polynomials in the remaining variables $z_i$, and $A$ is of the form $Z_w^+(\{v_1\})$ in the notation of Lemma 7.9. We again assume $0 < \beta < 1$ and $|z_i| \geq 1$ for $i > 1$. We now consider $G_1$. Denoting the activity at $u$ by $z$, we can write

$$\begin{aligned} Z_w(G_1) &= A(\beta + z^{w_0})z_1^{w_1+1} + B(1 + \beta z^{w_0}); \\ Z_w(G_1)' &= (A' + w_1 A)(\beta + z^{w_0})z_1^{w_1+1} + A(\beta + (w_0 + 1)z^{w_0})z_1^{w_1+1} + B'(1 + \beta z^{w_0}) \\ &\quad + w_0 \beta B z^{w_0}. \end{aligned}$$

Now suppose that $G_1$ does not satisfy the weighted strict Gauss-Lucas property, and hence $|z|$ and $|z_1|$ are both also at least 1, but $Z_w(G_1)' = 0$. It follows from Theorem 7.4 that we then also have $Z_w(G_1) = 0$. We now proceed to derive a contradiction to the above observations. For convenience, we denote $z_1^{w_1+1}$ by $y$ in what follows.

Using Lemma 7.9, we know that $A \neq 0$ and that $A' + w_1 A \neq 0$ for our setting of activities. By Theorem 6.2 applied to $Z_w(G)$ and the weighted strict Gauss-Lucas property applied to $Z_w(G)'$, we get

$$\left| \frac{B}{A} \right| \leq 1, \quad \text{and} \quad \left| \frac{B'}{A' + w_1 A} \right| < 1. \tag{7.6}$$

Also, since $Z_w(G_1) = 0$, we must have

$$y = -\frac{B}{A} \frac{1 + \beta z^{w_0}}{\beta + z^{w_0}}. \tag{7.7}$$

Notice that $y$ is well defined since $A \neq 0$, $|z| \geq 1$ and $\beta < 1$. Further, since $\beta < 1$, we have $|1 + \beta z^{w_0}| < |\beta + z^{w_0}|$ when $|z| > 1$ (respectively, $|1 + \beta z^{w_0}| \leq |\beta + z^{w_0}|$ when $|z| \geq 1$). Since

we also have $|B| \leq |A|$, it follows from the above consideration that either one of $|z| > 1$ or $|B| < |A|$ would imply that $|y| < 1$, which is a contradiction to our assumption that $|z_1| \geq 1$ (since $y = z_1^{w_1+1}$). Thus, we must have

$$|z| = 1 \quad \text{and} \quad \left| \frac{B}{A} \right| = 1. \tag{7.8}$$

Now, substituting the value of $y$ from eq. (7.7) into $Z_w(G_1)' = 0$, we get

$$B'(1 + \beta z^{w_0}) + \beta w_0 B z^{w_0} = ((A' + w_1 A)(\beta + z^{w_0}) + A(\beta + (w_0 + 1)z^{w_0})) \frac{B}{A} \frac{1 + \beta z^{w_0}}{\beta + z^{w_0}}.$$

Dividing through by $(A' + w_1 A)(1 + \beta z^{w_0})$, setting $c = A/(A' + w_1 A)$ and rearranging terms, we get

$$\begin{aligned}
\frac{B'}{A' + w_1 A} &= \frac{B}{A} \left( 1 + c + w_0 c \left( \frac{z^{w_0}}{\beta + z^{w_0}} + \frac{1}{1 + \beta z^{w_0}} - 1 \right) \right) \\
&= \frac{B}{A} \left\{ 1 + c + w_0 c \left( 2\Re \left( \frac{z^{w_0}}{\beta + z^{w_0}} \right) - 1 \right) \right\}, \text{ since } |z| = 1. \tag{7.9}
\end{aligned}$$

Notice that these divisions are well defined since $A' + w_1 A \neq 0$, and $\beta < 1$ and $|z| = 1$ implies that $(1 + \beta z^{w_0}) \neq 0$ as well. Note also that $c$ is of the form $1/(w_1 + c')$ where $\Re(c') = \Re(A'/A) \geq 0$ by Lemma 7.9 and our earlier observations about $A$: it therefore follows that $\Re(c) \geq 0$. However, we then calculate that for $|z| = 1$ and $\beta < 1$, $\Re(z/(\beta + z)) > 1/2$. Combining this with the inequality $\Re(c) \geq 0$ derived above, we see that the factor inside the braces in (7.9) has real part (and hence absolute value) at least 1. Using $|B|/|A| = 1$ from (7.8), we then see that the right hand side of (7.9) always has absolute value at least 1, which gives us the required contradiction to (7.6). This shows that $G_1$ satisfies the weighted strict Gauss-Lucas property. $\qquad \square$

This finishes the proof of Theorem 7.5, and hence, as discussed above, also that of Theorem 7.3. In Section 7.6 we give an alternative analytical proof of a version of Theorem 7.3 that is also sufficient for our purposes.

## 7.2 Hardness of computing the mean magnetization

We now use our extended Lee-Yang theorem (Theorem 7.3) to prove Theorems 7.1 and 7.2 via reductions from the problem of computing the partition function of the Ising model, which is known to be #P-hard even for bounded degree graphs [DG00b, BG05]. More specifically, we will use the following #P-hardness result.

**Theorem 7.10** ([**BG05, Theorem 1**],[**DG00b, Theorem 5.1**]). *Fix $\beta$ satisfying $0 < \beta < 1$. The problem of computing the partition function $Z_I(G, \beta, 1)$ of the Ising model on connected graphs of maximum degree $\Delta \geq 3$ is #P-hard.*

For simplicity, we prove first a version of Theorem 7.1 without the bounded degree constraint. The extension to bounded degree graphs requires some more combinatorial work and is deferred to Section 7.3.

*Proof of Theorem 7.1.* We assume $\lambda > 1$, since the case $\lambda < 1$ is symmetrical. For given $0 < \beta < 1$, suppose that we have an algorithm $\mathcal{A}$ which, given a connected graph $G$, outputs the mean magnetization $M(G, \beta, \lambda)$ in polynomial time. Let $G$ be a graph of $n$ vertices. Notice that as a rational function in $z$, $M(G, \beta, z)$ is a ratio of the two polynomials, $\mathcal{D}Z_I(G, \beta, z)$ and $Z_I(G, \beta, z)$, which are both of degree $n$. Further, since $G$ is connected, these polynomials are co-prime by Theorem 7.3. Thus, if we could efficiently evaluate $M(G, \beta, z)$ at $2n + 2$ distinct points $z$ using algorithm $\mathcal{A}$, we could uniquely determine the coefficients of $Z_I(G, \beta, z)$ by Theorem 6.4 (since we know that the constant term in $Z_I(G, \beta, z)$ is 1). We could then determine $Z_I(G, \beta, 1)$ in polynomial time. Theorem 7.10 would then imply that computing the mean magnetization for the given values of the parameters $\beta$ and $\lambda$ is #P-hard.

In order to evaluate $M(G, \beta, z)$ at $2n + 2$ distinct values, we consider the graph $G(k)$ obtained by attaching $k$ new neighbors to each vertex of $v$. We then have

$$Z_I(G(k), \beta, \lambda) = (1 + \beta\lambda)^{nk} Z_I(G, \beta, \lambda_k), \text{ and} \tag{7.10}$$

$$M(G(k), \beta, \lambda) = \frac{kn\beta\lambda}{1 + \beta\lambda} + \left[1 + \frac{k\lambda(1 - \beta^2)}{(1 + \beta\lambda)(\beta + \lambda)}\right] M(G, \beta, \lambda_k), \tag{7.11}$$

where $\lambda_k = \lambda\left(\frac{\beta + \lambda}{1 + \beta\lambda}\right)^k$. Notice that when $\beta < 1$, all the $\lambda_k$ are distinct, and further, $M(G, \beta, \lambda_k)$ can be easily determined given $M(G(k), \beta, \lambda)$. Therefore, we can evaluate $M(G(k), \beta, \lambda)$ for $0 \leq k \leq 2n + 1$ using the algorithm $\mathcal{A}$, and then using eqs. (7.10) and (7.11) we can determine $M(G, \beta, \lambda_k)$ in polynomial time. Since these evaluations are at distinct points, the reduction is complete. $\square$

*Proof of Theorem 7.2.* For a given $\beta$ as specified in the theorem, suppose that there is a polynomial time algorithm $\mathcal{A}$ which, given a graph $G$ of maximum degree $\Delta$, and a value of $\lambda$ in unary, outputs the susceptibility $\chi(G, \beta, \lambda)$. Notice that as a rational function in $z$, $\chi(G, \beta, z)$ is a ratio of the two polynomials $Z_I(G, \beta, z) \cdot \mathcal{D}^2 Z_I(G, \beta, z) - (\mathcal{D}Z_I(G, \beta, \lambda))^2$ and $Z_I(G, \beta, z)^2$, which are both of degree $2n$. Further, since $G$ is connected, these polynomials are co-prime by Theorem 7.3. To see this, notice that any common complex zero of these two polynomials must be a common zero of $Z_I(G, \beta, \lambda)$ and $\mathcal{D}Z_I(G, \beta, \lambda)$, which is prohibited by Theorem 7.3.

To complete the reduction, we notice that we can choose $4n + 2$ distinct values of $\lambda$ in the interval $(0, 1]$ all of which have a unary representation length of at most $5n$. Thus, using $\mathcal{A}$, we can efficiently evaluate $\chi(G, \beta, z)$ at $4n + 2$ distinct values of $z$. By Theorem 6.4 we can then use these evaluations to uniquely determine the coefficients of $Z_I(G, \beta, z)^2$ (since we already know that the constant coefficient is 1), and hence, $Z_I(G, \beta, 1)$, in polynomial time. Because of Theorem 7.10, this implies that the problem of computing the susceptibility at the given value of $\beta$ is #P-hard. $\square$

## 7.3   Hardness for bounded degree graphs

In our proof of Theorems 7.1 in Section 7.2, we realized different values of $\lambda$ required for the interpolation by attaching $k$ extra vertices to each vertex of $G$. This necessarily entails a large increase in the degree of $G$. In this section, we give an alternative way of realizing different values of $\lambda$ which entails an increase in degree of exactly one, and which therefore allows us to complete the proofs of the stronger, degree-bounded version of Theorem 7.1.

We denote by $P_k$ a path of $k$ vertices. Let $p_k^+$ (respectively, $p_k^-$) be the partition function $Z_I(P_k, \beta, \lambda)$ restricted to configurations in which the leftmost vertex of $P_k$ is fixed to spin '+' (respectively, '−'). We also set $r_k := \frac{p_k^+}{p_k^-}$. Notice that $p_1^- = 1$ and $p_1^+ = r_1 = \lambda$. The following recurrence relations show that, for fixed $\beta$ and $\lambda$, $p_k^+, p_k^-$, and $r_k$ can be computed in time polynomial in $k$:

$$p_k^+ = \lambda(\beta p_{k-1}^- + p_{k-1}^+); \tag{7.12}$$
$$p_k^- = \beta p_{k-1}^+ + p_{k-1}^-; \tag{7.13}$$
$$r_k = \lambda \frac{\beta + r_{k-1}}{1 + \beta r_{k-1}}; \tag{7.14}$$

Notice that $p_k^+, p_k^-$, and $r_k^+$ are all functions of $\lambda$. We note that values of their derivatives with respect to $\lambda$ can also be computed in time polynomial in $k$ via the following recurrence relations: $\dot{p}_1^- = 0, \dot{p}_1^+ = \dot{r}_1 = 1$, and

$$\dot{p}_k^+ = p_k^+/\lambda + \lambda(\beta \dot{p}_{k-1}^- + \dot{p}_{k-1}^+); \tag{7.15}$$
$$\dot{p}_k^- = \beta \dot{p}_{k-1}^+ + \dot{p}_{k-1}^-; \tag{7.16}$$
$$\dot{r}_k = \frac{\dot{p}_k^+ p_k^- - p_k^+ \dot{p}_k^-}{\left(p_k^-\right)^2}; \tag{7.17}$$

Here, we use the dot notation for the derivative with respect to $\lambda$. Using a simple induction, one can also show that when $\beta < 1$, $\dot{r}_k > 0$ for all $k$.

Now consider a connected graph $G$. For $k \geq 1$, we define $G(k)$ as the graph obtained by attaching to each vertex $v$ of $G$ a different instance of the path $P_k$, such that $v$ is connected to the "leftmost" vertex of $P_k$ via an edge. Notice that the maximum degree of $G(k)$ is one more than the maximum degree of $G$. We now consider the Ising model on the graphs $G(k)$. We have

$$Z_I(G(k), \beta, \lambda) = (p_{k+1}^-)^n Z_I(G, \beta, \lambda_k), \tag{7.18}$$

where $n$ is the number of vertices in $G$ and $\lambda_k = r_{k+1}$. Notice that when $0 < \beta < 1$, the sequence $\lambda_k$ is strictly increasing and greater than 1 (respectively, strictly decreasing and less than 1) when $\lambda > 1$ (respectively, when $\lambda < 1$): this follows from the observation that the right hand side of the recurrence (7.14) is a strictly increasing function of $r_{k-1}$, and that $r_2 > r_1$ (respectively, $r_2 < r_1$) when $\lambda > 1$ (respectively, when $\lambda < 1$). We also have

$$M(G(k), \beta, \lambda) = \frac{n\lambda \dot{p}_{k+1}^-}{p_{k+1}^-} + \frac{\lambda \dot{r}_{k+1}}{r_{k+1}} M(G, \beta, \lambda_k). \tag{7.19}$$

We can now complete the proof of Theorem 7.1.

*Proof of Theorem 7.1.* As in the proof for the case of unbounded degree graphs (Section 7.2), we assume $\lambda > 1$ (since the case $\lambda < 1$ is symmetrical) and suppose that we have a polynomial time algorithm $\mathcal{A}$ which, given a connected graph $G$ of maximum degree $\Delta \geq 4$, outputs the mean magnetization $M(G, \beta, \lambda)$ in polynomial time.

Now consider any connected graph $G$ of maximum degree $\Delta - 1 \geq 3$. As shown in the proof in Section 7.2, Theorem 7.3 implies that if we can efficiently evaluate $M(G, \beta, z)$ at $2n + 2$ distinct values of $z$ using our hypothetical algorithm $\mathcal{A}$, we can uniquely determine the coefficients of $Z_I(G, \beta, z)$, and hence also $Z_I(G, \beta, 1)$, in polynomial time. In view of Theorem 7.10, this would imply that the problem of computing the mean magnetization in graphs of maximum degree at least 4 for parameter values $\beta$ and $\lambda$ is #P-hard.

In order to evaluate $M(G, \beta, z)$ at $2n + 2$ distinct values, we evaluate $M(G(k), \beta, \lambda)$ for $1 \leq k \leq 2n + 2$ using our hypothetical algorithm $\mathcal{A}$. Notice that this can be done since the construction of the $G(k)$ (as given in this section) implies that its maximum degree is one more than the maximum degree of $G$. Using eqs. (7.12) to (7.17) and (7.19), and the fact that $\dot{r}_k > 0$ for all $k$, we can then determine $M(G, \beta, \lambda_k)$ in polynomial time. Since $\lambda_k$ is a strictly increasing sequence, these evaluations are at distinct points, and hence the reduction is complete. $\square$

## 7.4 Extensions: General two state ferromagnetic spin systems and planar graphs

We now show how to extend our results to general two-state ferromagnetic spin systems and planar graphs. Recall that a general two-state spin system [GJP03] is parametrized by a $(+, +)$ edge potential $\alpha_1$, a $(-, -)$ edge potential $\alpha_2$, and a vertex activity $\lambda$. As before, given a graph $G = (V, E)$, we define a probability distribution over the set of configurations $\sigma : V \to \{+, -\}$ via the weights $w_S(\sigma)$ given by

$$w_S(\sigma) = \lambda^{p(\sigma)} \alpha_1^{e_+(\sigma)} \alpha_2^{e_-(\sigma)},$$

where $e_+(\sigma)$ (respectively, $e_-(\sigma)$) denotes the number of edges with '+' (respectively, '−') spin on both end-points, while $p(\sigma)$ denotes the number of vertices with + spin. The partition function $Z_S(G, \alpha_1, \alpha_2, \lambda)$ and the magnetization $M_S(G, \alpha_1, \alpha_2, \lambda)$ are given by

$$Z_S(G, \alpha_1, \alpha_2, \lambda) := \sum_{\sigma \in \{+, -\}^V} w_S(\sigma);$$

$$M_S(G, \alpha_1, \alpha_2, \lambda) := \langle p \rangle = \frac{\sum_\sigma p(\sigma) w_S(\sigma)}{Z_S(G, \alpha_1, \alpha_2, \lambda)}.$$

**Remark 7.3.** The Ising model corresponds to the special case $\alpha_1 = \alpha_2 = 1/\beta$.

It is well known that general two-state spin systems can be represented in terms of an Ising model in which the activity at each vertex depends upon the degree of the vertex [GJP03]. In particular, if $G$ is a $\Delta$-regular graph then all vertex activities in the equivalent Ising model are the same, and one has

$$w_S(\sigma) = \alpha_2^{|E|} w_I(\sigma) \tag{7.20}$$

where the Ising model has an edge potential $\beta = 1/\sqrt{\alpha_1\alpha_2}$ and a vertex activity $\lambda' = \lambda(\alpha_1/\alpha_2)^{\Delta/2}$. A two-spin system is called *ferromagnetic* if the above translation produces a ferromagnetic Ising model, that is, when $\alpha_1\alpha_2 \geq 1$.

However, the above translation does not allow us to directly translate our hardness result for the ferromagnetic Ising model, since our results were not derived for regular graphs. We will instead do a reduction similar to the ones done in our earlier proof, but starting from the following somewhat stronger hardness result for the partition function.

**Remark 7.4.** In this section, we allow graphs to have parallel edges (in computing the degree of a vertex, we count each parallel edge separately). As observed in Remark 7.2 in Section 7.1, our extended Lee-Yang theorem (Theorem 7.3) holds also in this setting.

**Theorem 7.11 ([CK10, Theorem 1]).** *Fix $\alpha_1, \alpha_2 > 0$ with $\alpha_1\alpha_2 > 1$ and $\Delta \geq 3$. The problem of computing the partition function $Z_S(G, \alpha_1, \alpha_2, 1)$ on $\Delta$-regular graphs is #P-hard.*

We will then prove the following theorem. We will also show later in this section that the theorem can be strengthened so that the #P-hardness holds even when the input is restricted to planar graphs.

**Theorem 7.12.** *Fix $\alpha_1, \alpha_2$, and $\lambda > 0$, such that $\alpha_1\alpha_2 > 1$ . The problem of computing the magnetization $M_S(G, \alpha_1, \alpha_2, \lambda)$ on connected graphs of maximum degree at least $4$ is #P-hard, except when $\alpha_1 = \alpha_2$ and $\lambda = 1$, in which case it can be solved in polynomial time.*

**Remark 7.5.** Notice that when $\alpha_1\alpha_2 = 1$, the problem reduces to the case of a graph consisting of isolated vertices, and hence can be solved in polynomial time. Similarly, in the case $\alpha_1 = \alpha_2$ and $\lambda = 1$, the two spins are symmetric, and the magnetization is therefore $n/2$, where $n$ is the number of vertices in $G$.

Before proceeding with the proof of Theorem 7.12, we will need to analyze the model on graphs $G(k)$ defined in Section 7.3. As before, we begin by analyzing the model on the path $P(k)$. We denote by $p_k^+$ (respectively, $p_k^-$) the partition function $Z_S(P_k, \alpha_1, \alpha_1, \lambda)$ restricted to configurations in which the leftmost vertex is fixed to be '+' (respectively, '−'). We also define the ratio $r_k = p_k^+/p_k^-$. Similarly, we denote by $m_k^+$ (respectively, $m_k^-$) the average magnetization of the path $P_k$ conditioned on the leftmost vertex being fixed to '+' (respectively, '−'). We have

$p_1^+ = r_1 = \lambda, p_1^- = 1$ and $m_1^+ = 1, m_1^- = 0$, and the following recurrences for $k \geq 1$:

$$p_k^+ = \lambda(\alpha_1 p_{k-1}^+ + p_{k-1}^-) \tag{7.21}$$

$$p_k^- = \alpha_2 p_{k-1}^- + p_{k-1}^+ \tag{7.22}$$

$$r_k = \lambda \frac{\alpha_1 r_{k-1} + 1}{\alpha_2 + r_{k-1}} \tag{7.23}$$

$$m_k^+ = 1 + \frac{\alpha_1 m_{k-1}^+ p_{k-1}^+ + m_{k-1}^- p_{k-1}^-}{\alpha_1 p_{k-1}^+ + p_{k-1}^-} \tag{7.24}$$

$$m_k^- = \frac{\alpha_2 m_{k-1}^- p_{k-1}^- + m_{k-1}^+ p_{k-1}^+}{\alpha_2 p_{k-1}^- + p_{k-1}^+} \tag{7.25}$$

Under the condition $\alpha_1 \alpha_2 > 1$, one can prove using a simple induction that for all $k \geq 1$, $m_k^+ - m_k^- > 0$, and that when $(\alpha_1 - 1)\lambda - (\alpha_2 - 1) > 0$ (respectively, when $(\alpha_1 - 1)\lambda - (\alpha_2 - 1) < 0$), the $r_k$ form a strictly increasing (respectively, strictly decreasing) sequence, and hence are all distinct.

For reasons that will become clear shortly, we need the $r_k$ to be distinct, and hence we will need to handle the remaining case $(\alpha_1 - 1)\lambda - (\alpha_2 - 1) = 0$ specially. We observe that unless $\alpha_1 = \alpha_2 = 1$, or $\alpha_1 = \alpha_2$ and $\lambda = 1$, both of which are excluded in the hypothesis of the theorem, we cannot have both $(\alpha_1 - 1)\lambda - (\alpha_2 - 1) = 0$ and $(\alpha_1^2 - 1)\lambda - (\alpha_2^2 - 1) = 0$. To take advantage of this, we will modify $P(k)$ by replacing each edge in $P(k)$ by two parallel edges. We call the resulting graph $P(k)'$, and again define the quantities $p_k^+, p_k^-. r_k, m_k^+$ and $m_k^-$ by recursion on $P(k)'$. Notice that for $k = 1$, these quantities are the same as those for $P(k)$; however for $k \geq 2$, we now need to modify the recurrences above by replacing $\alpha_1$ and $\alpha_2$ by $\alpha_1^2$ and $\alpha_2^2$ respectively. As before, we have $m_k^+ - m_k^- > 0$ for all $k \geq 1$. Further, by our observation, the $r_k$ form a strictly monotone sequence. Thus, in the case $(\alpha_1 - 1)\lambda - (\alpha_2 - 1) = 0$, we redefine $G(k)$ to use the paths $P(k)'$ in place of $P(k)$. In what follows, we will assume that $G(k)$ are appropriately defined taking into account the values of $\lambda, \alpha_1$ and $\alpha_2$, and will not explicitly keep track of the above modification. Notice that the maximum degree of $G(k)$ is $\max(\Delta + 1, 3)$, where $\Delta$ is the maximum degree of $G$.

Given the above definition of $G(k)$, we have the relations

$$Z_S(G(k), \alpha_1, \alpha_2, \lambda) = (\alpha_2 p_k^- + p_k^+)^n Z_S(G, \alpha_1, \alpha_2, \lambda_k) \tag{7.26}$$

$$M_S(G(K), \alpha_1, \alpha_2, \lambda) = nm_k^- + (m_k^+ - m_k^-)M_S(G(k), \alpha_1, \alpha_2, \lambda_k), \tag{7.27}$$

where $\lambda_k = \lambda(\alpha_1 r_k + 1)/(\alpha_2 + r_k)$. Since the $r_k$ form a strictly monotone sequence, it follows that (since $\alpha_1 \alpha_2 > 1$) so do the $\lambda_k$. In particular, all the $\lambda_k$ are distinct.

*Proof of Theorem 7.12.*  Proceeding as in the proof of Theorem 7.10, we fix any $\lambda > 0$ and $\alpha_1, \alpha_2 > 0$ satisfying $\alpha_1 \alpha_1 > 1$, and suppose that there exists a polynomial time algorithm $\mathcal{B}$ which, given a connected graph $H$ of maximum degree $\Delta \geq 4$, outputs $M_S(G, \alpha_1, \alpha_2, \lambda)$.

Now consider any connected regular graph $G = (V, E)$ of degree $d := \Delta - 1 \geq 3$ on $n$ vertices. From the translation in eq. (7.20), we see that for any $\lambda > 0$,

$$Z_S(G, \alpha_1, \alpha_2, \lambda) = \alpha_2^{|E|} Z_I \left( G, \beta, \lambda \left( \frac{\alpha_1}{\alpha_2} \right)^{d/2} \right), \text{ and}$$

$$M_S(G, \alpha_1, \alpha_2, \lambda) = M_I \left( G, \beta, \lambda \left( \frac{\alpha_1}{\alpha_2} \right)^{d/2} \right),$$

where $\beta = 1/\sqrt{\alpha_1 \alpha_2} < 1$. Theorem 6.4 along with our main Theorem 7.3 then implies that if we can efficiently evaluate $M_S(G, \alpha_1, \alpha_2, z)$ at $2n + 2$ distinct values of $z$ using our hypothetical algorithm $\mathcal{B}$, we can uniquely determine the coefficients of $Z_S(G, \alpha_1, \alpha_2, z)$, and hence also the value of $Z_S(G, \alpha_1, \alpha_2, 1)$, in polynomial time. In view of Theorem 7.11, this would imply that the problem of computing the mean magnetization in graphs of maximum degree at least $4$ for parameter values $\alpha_1, \alpha_2$ and $\lambda$ is #P-hard.

In order to evaluate $M_S(G, \alpha_1, \alpha_2, z)$ at $2n + 2$ distinct values of $z$, we instead compute $M_S(G(k), \alpha_1, \alpha_2, \lambda)$, for $1 \leq k \leq 2n + 2$, using our hypothetical algorithm $\mathcal{B}$. Notice that this can be done since the construction of $G(k)$ implies that its maximum degree is one more than the maximum degree of $G$. Using eqs. (7.21) to (7.25) and (7.27), and the fact that $m_k^+ - m_k^- > 0$ for all $k$, we can then determine $M_S(G, \alpha_1, \alpha_2, \lambda_k)$ in polynomial time. Since $\lambda_k$ is a strictly monotone sequence as shown in the discussion above, these evaluations are at distinct values of $z$, and hence the reduction is complete.                                                                           $\square$

We now extend our results to the case of planar graphs of bounded degree. Our starting point is the following planar graph version of Theorem 7.11, again due to Cai and Kowalczyk [CK10].

**Theorem 7.13 ([CK10, Theorem 1]).** *Fix $\alpha_1, \alpha_2 > 0$ with $\alpha_1 \alpha_2 > 1$, $\alpha \neq \alpha_2$ and $\Delta \geq 3$. The problem of computing the partition function $Z_S(G, \alpha_1, \alpha_2, 1)$ on planar $\Delta$-regular graphs is #P-hard.*

In order to extend Theorem 7.12 to planar graphs, we consider the cases $\alpha_1 \neq \alpha_2$ and $\alpha_1 = \alpha_2 = \alpha$ separately. In case $\alpha_1 \neq \alpha_2$, we proceed exactly as in the proof of Theorem 7.12 given above, except that we start with a *planar* $d$-regular graph $G$ in the reduction, and use Theorem 7.13 instead of Theorem 7.11 as our starting hardness result. Since $G$ is planar, so are the $G(k)$, and hence we see that computing $M_S(H, \alpha_1, \alpha_2, \lambda)$ on planar graphs $H$, for $\alpha_1, \alpha_2$ and $\lambda$ satisfying the condition $\alpha_1 \neq \alpha_2$ in addition to the conditions of Theorem 7.12 is #P-hard.

We now turn to the case $\alpha_1 = \alpha_2 = \alpha > 1$ (with $\lambda \neq 1$). In this case, we start with the fact that computing $Z_S(G, 2\alpha, \frac{\alpha}{2}, 1)$ on planar $\Delta$-regular graphs is #P-hard (this is a direct corollary of Theorem 7.13) . We again proceed exactly as in proof of Theorem 7.12, starting with an arbitrary *planar* $(\Delta - 1)$-regular graph $G$ (for $\Delta \geq 4$), and noting that the $G(k)$ are planar too. Notice that the proof then shows that assuming the existence of a polynomial time algorithm to compute the magnetization in planar graphs of maximum degree $\Delta$, we can evaluate the coefficients of the polynomial $Z_s(G, \alpha, \alpha, z)$, and hence also the quantity $Z_S(G, \alpha, \alpha, 2^d)$. However we then use the

translation to the Ising model given above to see that

$$Z_S(G, \alpha, \alpha, 2^d) = \left(\frac{\alpha}{2}\right)^{|E|} Z_S\left(G, 2\alpha, \frac{\alpha}{2}, 1\right),$$

which shows that we can also evaluate $Z_S(G, 2\alpha, \frac{\alpha}{2}, 1)$ in polynomial time. This establishes the #P-hardness in the remaining case $\alpha_1 = \alpha_2$ (with $\lambda \neq 1$).

We thus see that in Theorem 7.12, the input graphs can be restricted to be planar, and the same hardness result still holds.

## 7.5 Proof of Theorem 7.4

As indicated earlier, Theorem 7.4 is an easy corollary of the following result of Newman [New74] (restated in our notation).

**Theorem 7.14 ([New74, Theorem 3.2 and eq. 3.5]).** *Consider the ferromagnetic Ising model on a graph $G = (V, E)$ on $n$ vertices, where we allow the edge potentials also to be variable, with the condition that the edge potential $\beta_{uv}$ on every edge $uv$ satisfies $0 < \beta_{uv} < 1$. Let $(z_v)_{v \in V}$ be a collection of complex vertex activities such that $|z_v| > 1$ for all $v \in V$. Then,*

$$\Re\left(M(G, (\beta_{uv})_{uv \in E}, (z_v)_{v \in V})\right) > \frac{n}{2}.$$

**Remark 7.6.** Note that the main difference between the above theorem and Theorem 7.4 is that the above theorem only applies when all the vertex activities have magnitudes *strictly* larger than 1, while Theorem 7.4 also allows vertex activities with magnitude 1.

We now proceed with the proof of Theorem 7.4.

*Proof of Theorem 7.4.* Let $w$ be any weight assignment (not necessarily legal) of positive integral weights to the vertices of $G$. Consider the graph $H$ obtained from $G$ by appending to each vertex $v$ of $G$ a chain $C_v$ of $w(v) - 1$ vertices. Further, we let the edge potential be $\beta$ on all edges of $H$ which were present in $G$, and $0 < \gamma < 1$ on all the edges which are either part of some $C_v$, or connect a vertex $v$ to its associated chain $C_v$. Let $(y_v)_{v \in V}$ be a set of vertex activities on $V$. Henceforth, we will drop the subscript and refer to this set of activities as $(y_v)$. With a slight abuse of notation, we also denote by $(y_v)$ the collection of activities on $H$ such that for any vertex $x$ in $H$ such that $x \in C_v$ for some $v \in V$, we have $y_x = y_v$.

Now consider any collection of activities such that $|y_v| > 1$ for all $v \in V$. From Theorem 7.14, we get that for any $\gamma \in (0, 1)$,

$$\Re\left(M(H, \{\beta, \gamma\}, (y_v))\right) > \frac{n}{2}. \tag{7.28}$$

Since $|y_v| > 1$ for all $v$, the Lee-Yang theorem[4] implies that both $Z_I(H, \{\beta, \gamma\}, (y_v))$ as well as $Z_w(G, \beta, (y_v))$ are non-zero (when $0 < \gamma < 1$). We can therefore take the limit $\gamma \to 0$ in eq. (7.28) to get

$$\frac{n}{2} \leq \lim_{\gamma \to 0} \Re\left(M(H, \{\beta, \gamma\}, (y_v))\right) = \Re\left(\frac{Z'_w(G, \beta, (y_v))}{Z_w(G, \beta, (y_v))}\right) = \Re\left(M(G, \beta, (y_v))\right). \qquad (7.29)$$

We now take a sequence $\left((y_v^\ell)\right)_{\ell=1}^\infty$ of activity assignments such that $\left|y_v^\ell\right| > 1$ for all $\ell$ and $v$ and such that $\lim y_v^\ell = z_v$. Since we assume in the hypothesis of the theorem that $Z_w(G, \beta, (z_v)_{v \in V}) \neq 0$, we can take the limit $\ell \to \infty$ in eq. (7.29) to get

$$\frac{n}{2} \leq \lim_{\ell \to \infty} \Re\left(M\left(G, \beta, (y_v^\ell)\right)\right) = \Re\left(M(G)\right),$$

which completes the proof. $\qquad \square$

## 7.6   An alternative analytical proof of Theorem 7.3

We conclude with an alternative, analytical proof of the following, slightly weaker version of Theorem 7.3. This section is based on joint work with Mario Szegedy.

**Theorem 7.15.** *Let $G = (V, E)$ be a connected undirected graph on $n$ vertices, and assume $0 < \beta < 1$. If $(y_v)_{v \in V}$ are complex numbers such that $|y_v| = 1$ for $v \in V$, then $\mathcal{D}_G Z(G, \beta, (y_v)_{v \in V}) \neq 0$.*

**Remark 7.7.** By setting all $y_v$ to be equal, we see from the above theorem that the derivative of the (univariate) partition function has no zeros on the unit circle. Combined with the Lee-Yang theorem, this establishes that the partition function and its derivative have no common zeros, and hence the above theorem is also sufficient for the purposes of our complexity theoretic results.

As before, we observe that given any vertex $u \in V$, we can decompose the partition function as

$$Z(G, \beta, (z_v)_{v \in V}) = \beta^{\deg(u)} Z\left(G - u, \beta, \left(z_v^+\right)_{v \in V - \{u\}}\right) z_u$$
$$+ \quad Z\left(G - \{u\}, \beta, \left(z_v^-\right)_{v \in V - \{u\}}\right), \qquad (7.30)$$

where $z_w^+ = z_w^- = z_w$ when $w \nsim u$ in $G$, and $z_w^+ = z_w/\beta$, $z_w^- = \beta z_w$ when $w \sim u$ in $G$. We will denote the linear and constant coefficients in the above decomposition as $A_u(G, \beta, (z_v)_{v \in V - \{u\}})$ and $B_u(G, \beta, (z_v)_{v \in V - \{u\}})$ respectively, so that $Z = A_u z_u + B_u$.

It then follows from the first part of Lemma 7.9 that for any connected undirected graph $G = (V, E)$, $0 < \beta < 1$, and a collection of vertex activities $(z_v)_{v \in V}$ such that $|z_v| = 1$ for all $v \in V$, the coefficient $A_u$ defined above is non-zero. We can now prove Theorem 7.15.

---

[4]Although we stated Theorem 6.2 only for uniform edge potentials, Asano's proof [Asa70] in fact supports our current conclusion with variable edge potentials and a possibly disconnected graph.

*Proof of Theorem 7.15.* Let $G$ and $\beta$ be as in the hypotheses of the theorem. Suppose now that there exists a point $(y_v)_{v \in V}$ such that $|y_v| = 1$ for all $v$, and $\mathcal{D}_G Z(G, \beta, (y_v)_{v \in V}) = 0$. We will show that this leads to a contradiction to the Lee-Yang theorem.

We begin by proving that that $Z(G, \beta, (y_v)_{v \in V}) = 0$ as well. To see this, define the univariate polynomial

$$f(t) := Z_G(G, \beta, (t y_v)_{v \in V}),$$

and observe that $f'(1) = \mathcal{D}_G Z(G, \beta, (y_v)_{v \in V}) = 0$. Further, the Lee-Yang theorem implies that $f(t) \neq 0$ when $|t| \neq 1$, so that all the zeros of $f$ lie on the unit circle. From the Gauss-Lucas lemma, we then see that the derivative $f'$ can be zero at a point $\alpha$ on the unit circle if and only if $f$ itself is 0 at the that point. Thus, since $f'(1) = 0$, we also have $f(1) = Z(G, \beta, (y_v)_{v \in V}) = 0$.

We now consider the function

$$g(t) := Z\left(G, \beta, \left(e^t y_v\right)_{v \in V}\right).$$

Note that $g'(t) = \mathcal{D}_G Z(G, \beta, (e^t y_v)_{v \in V})$, so that the preceding discussion implies that $g(0) = g'(0) = 0$. Since $g$ is entire and not identically zero, we can then conclude from the power series development of $g$ around 0 that there exist $\epsilon, \delta > 0$ such that for $0 < |t| \leq \epsilon$, we have $0 < |g(t)| \leq \delta |t|^2$.

Now, fix any vertex $u$ in $G$, and consider the decomposition of the partition function in terms of the polynomials $A_u$ and $B_u$ described above. From the remark just preceding the proof, we have $A_u(G, \beta, (y_v)_{v \in V - \{u\}}) \neq 0$. Since $A_u$ is a polynomial, continuity implies that there exist positive $\epsilon_1$ and $\delta_1$ such that when $|y_v - x_v| \leq \epsilon_1$ for all $v \in V - \{u\}$, we have

$$|A_u(G, \beta, (x_v)_{v \in V - \{u\}})| \geq \delta_1.$$

We now choose

$$t = \min\left(1, \epsilon, \frac{\epsilon_1}{2}, \frac{\delta_1}{2\delta}\right) > 0,$$

and let $x_v = e^t y_v$ for all $v \in V$. Note that $|x_v| = e^t \geq 1 + t$. We also have $|y_v - x_v| = e^t - 1 \leq 2t \leq \epsilon_1$ (since $t \leq \min(1, \epsilon_1/2)$). Denoting $A_u(G, \beta, (x_v)_{v \in V - \{u\}})$ by $\omega$, we see from the last paragraph that $\omega \geq \delta_1$. Further, we have $Z(G, \beta, (x_v)_{v \in V}) = g(t)$, so that if $(z_v)_{v \in V}$ is collection of vertex activities such that $z_v = x_v$ for $v \neq u$, the decomposition in eq. (7.30) implies that

$$Z(G, \beta, (z_v)_{v \in V}) = g(t) + \omega(z_u - x_u).$$

Thus, if we choose $z_u = x_u - g(t)/\omega$, we get a zero of the partition function. We now show that $|z_v| > 1$ for all $v$ so that this is a contradiction to the Lee-Yang theorem. When $v \neq u$, we have $|z_v| = |x_v| \geq 1 + t > 1$. For $z_u$, we have (using $t \leq \epsilon$ followed by $t \leq \delta_1/(2\delta)$)

$$|z_u| = \left|x_u - \frac{g(t)}{\omega}\right| \geq |x_u| - \frac{|g(t)|}{|\omega|} \geq 1 + t - \frac{\delta t^2}{\delta_1} \geq 1 + \frac{t}{2} > 1,$$

which completes the proof. $\qquad \square$

# Chapter 8

# The average monomer count: The Heilmann-Lieb theorem

This chapter carries forward the program introduced in Chapter 6 with a proof of the following hardness result for the average monomer count in the monomer-dimer model.

**Theorem 8.1.** *For any fixed $\lambda > 0$, the problem of computing the average number of monomers (equivalently, the average size of a matching) in the monomer-dimer model on connected graphs with edge weights in the set $\{1, 2, 3\}$ is #P-hard. Moreover, the problem remains #P-hard even when the input is restricted to graphs of maximum degree at least $5$.*

**Remark 8.1.** Note that our hardness result requires a small finite number (three) of different values for the edge weights. However, this requirement can be removed if $G$ is allowed to have parallel edges; the theorem then holds for any single fixed non-zero edge weight (including the uniform case in which all edge weights are $1$).

As in the case of the ferromagnetic Ising model, our strategy is to exploit results about the location of zeros of the partition function and its derivatives, the appropriate analog here being the Heilmann-Lieb theorem (Theorem 6.3). Unlike the case of the Lee-Yang theorem, the Heilmann-Lieb theorem already guarantees that the partition function and its derivative do not have any common zeros, but only does so for the special case of *Hamiltonian* graphs, i.e., graphs that contain a Hamiltonian path. In fact, Heilmann and Lieb [HL72] also present examples of *connected* graphs $G$ for which $Z_M$ has repeated zeros, so we cannot hope to prove an analog of Theorem 7.3 in this case. However, we will capitalize on their result for Hamiltonian graphs by adapting existing #P-hardness reductions for $Z_M$ in such a way that the instances of $Z_M$ that appear in the reduction always contain a Hamiltonian path. Specifically, we will give a reduction from the problem Monotone 2-SAT of counting satisfying assignments of a monotone 2-CNF formula to computing $Z_M$ in Hamiltonian graphs $G$. The reduction is an elaboration of Valiant's original #P-completeness proof for the permanent [Val79a], and is the subject of the rest of this chapter.

## 8.1 The monomer-dimer partition function in Hamiltonian graphs

We now prove the following hardness result for the exact computation of the monomer-dimer partition function in Hamiltonian graphs.

**Theorem 8.2.** *There exists a polynomial time algorithm $\mathcal{A}$ which, when given as input a* Monotone *2-SAT formula $\phi$, outputs a weighted graph $G$ with the following properties:*

1. *The weights in $G$ are drawn from the set $\{1, 2, 3\}$.*

2. *Suppose $\phi$ has $\nu$ variables and $\mu$ clauses. Then, given the total weight $W$ of perfect matchings in $G$, the number of satisfying assignments of $\phi$ can be determined in polynomial time from $W$, $\mu$, and $\nu$.*

3. *$G$ contains a Hamiltonian path.*

We observe here that Valiant's reduction from #3-SAT [Val79a] can be easily modified so that it satisfies properties 1 and 2. However, it is property 3 that is crucial for our purposes, since it allows the use of Theorem 6.3. We first show how Theorem 8.2 can be used to immediately prove a slightly weaker version of Theorem 8.1, which shows hardness only on general graphs. The proof showing hardness for bounded degree graphs is exactly analogous to that for the Ising model given in the last chapter, and can be found in Section 8.2.

*Proof of Theorem 8.1.* Fix any $\lambda > 0$, and suppose that there exists a polynomial time algorithm $\mathcal{B}$ which, given a connected graph $H$, with edge weights in the set $\{1, 2, 3\}$ outputs $U(H, (\gamma_e)_{e \in E}, \lambda)$. In the following, we suppress the dependence on edge weights $(\gamma_e)_{e \in E}$ for clarity of notation. Given a Monotone 2-SAT formula $\phi$, we can then produce the graph $G = \mathcal{A}(\phi)$ in polynomial time. Let $n$ be the number of vertices in $G$. Since $G$ contains a Hamiltonian path, Theorem 6.3 implies that $Z_M(G, z)$ and $\mathcal{D}Z_M(G, z)$ have no common zeros. Thus, being able to use algorithm $\mathcal{B}$ to evaluate $U(G, z)$ at $2n + 2$ different values of $z$ would allow us to uniquely determine the coefficients of $Z_M(G, z)$ in polynomial time by rational interpolation (Theorem 6.4), since we already know that the coefficient of $z^n$ is 1. This would allow us to obtain $W$ (which is the constant term in $Z_M(G, z)$), and hence, by property 2, also the number of satisfying assignments of $\phi$, in polynomial time. This would show that the problem of computing $U(G, \lambda)$ is #P-hard (since #Monotone 2-SAT is #P-hard [Val79b]).

However, $\mathcal{B}$ only allows us to evaluate $U(G, z)$ at $z = \lambda$. In order to "simulate" other values of $\lambda$, we consider the graph $G(k)$ obtained by attaching $k$ new vertices to each vertex of $G$ with unit weight edges. We then have

$$Z_M(G(k), \lambda) = \lambda^{nk} Z_M(G, \lambda_k); \tag{8.1}$$

$$U(G(k), \lambda) = nk + \frac{\lambda^2 - k}{\lambda^2 + k} U(G, \lambda_k), \tag{8.2}$$

where $\lambda_k = \lambda + k/\lambda$. Thus, by choosing $2n + 2$ different values of $k$, none of which is equal to $\lambda^2$, we can determine $U(G, z)$ at $2n + 2$ different values of $z$ by running $\mathcal{B}$ on $G(k)$ and using eq. (8.2). This completes the proof.    □

In the rest of this section, we prove Theorem 8.2 in a sequence of steps. For simplicity, we will describe our reduction in terms of cycle covers in a directed graph rather than perfect matchings in an undirected graph (this also allows us to directly compare our gadget construction with that of Valiant [Val79a] at various steps). Given a weighted directed graph $G = (V, E)$, we define the undirected bipartite graph $\mathrm{Bip}(G) = (V \times \{0, 1\}, E')$ where the edge $\{(x, 0), (y, 1)\}$ is in $E'$ with weight $\gamma_e$ if and only if $(x, y)$ is an edge in $E$ with the same weight. Note that a subset $S \subseteq E$ forms a cycle cover of weight $w$ in $G$ if and only if the corresponding subset of edges $S' = \{\{(x, 0), (y, 1)\} \mid (x, y) \in S\}$ forms a perfect matching of weight $w$ in $\mathrm{Bip}(G)$. In particular, the total weight of all perfect matchings in $\mathrm{Bip}(G)$ is the same as the total weight of all cycle covers of $G$.

Later, while arguing about the existence of Hamiltonian paths in graphs of the form $\mathrm{Bip}(G)$, we will find it convenient to use the following short-hand notation for simple paths in the graph $\mathrm{Bip}(G)$ in terms of the edges and vertices of $G$. Consider any simple path $(x_1, 1), (x_2, 0), (x_3, 1)$, $(x_4, 0), \ldots (x_l, 1)$, where we have assumed for simplicity that $l$ is odd. The edges corresponding to this path in $G$ are $x_1 \leftarrow x_2, x_2 \rightarrow x_3, x_3 \leftarrow x_4, \ldots x_{l-1} \rightarrow x_l$. Notice that alternate edges are traversed in reverse in this representation. The path can therefore be represented as $x_1 \leftarrow x_2 \rightarrow x_3 \leftarrow x_4 \rightarrow \ldots \rightarrow x_l$. Similarly for a path starting on the other side, say $(x_1, 0), (x_2, 1), (x_3, 0), (x_4, 1)$, we have the representation $x_1 \rightarrow x_2 \leftarrow x_3 \rightarrow x_4$. Notice that a path $p_2$ starting at a vertex $v$ in this notation can be appended to a path $p_1$ ending at $v$ if and only if the arrows at $v$ in $p_1$ and $p_2$ respectively are in opposite directions. We will refer to this notation as the *alternating path* representation. Further, given an alternating path representation of a path, we will refer to edges going right (such as $x_1 \rightarrow x_2$ in the last example) as *forward* edges, and edges going left (such as $x_2 \leftarrow x_3$ in the above example) as *backward* edges.

### 8.1.1   Overview of the reduction

We now look at the basic structure of our reduction, which is an elaboration of Valiant's reduction [Val79a] as modified by Papadimitriou [Pap94] and presented in [AB09]. Recall that given a MONOTONE 2-SAT formula $\phi$, the reduction needs to produce in polynomial time a directed graph $G$ such that the number of satisfying assignments of $\phi$ can be easily determined from the total weight of cycle covers of $G$, and such that $\mathrm{Bip}(G)$ has a Hamiltonian path. Our first step is to introduce a shared variable in all the clauses of $\phi$: this shared variable will be useful later in showing the existence of a Hamiltonian path through the gadget.

**Observation 8.3.** *Let $\phi = \bigwedge_{i=1}^{\mu} c_i$ be a MONOTONE 2-SAT formula with $\mu$ clauses, $\nu$ variables, and $s$ satisfying assignments. Let $\tau$ be a variable not appearing in $\phi$ and consider the 3-SAT formula*

$$\phi' = \bigwedge_{i=1}^{\mu} (\tau \vee c_i).$$

*The number of satisfying assignments of $\phi'$ is $s' := 2^\nu + s$.*

Notice that each clause in $\phi'$ has exactly three variables, and that the number of satisfying assignments of $\phi$ can be easily determined given the number of satisfying assignments of $\phi'$.

We start the construction of $G$ by creating a separate *variable gadget* (see Figure 8.1) for each of the variables $\tau$, $x_1$, $x_2$, ..., $x_\nu$ occurring in $\phi'$. This gadget has an external *dotted* edge for each appearance of the variable in the formula, and is designed so that any cycle cover must either use *all* the dotted edges in a particular gadget, or none of them.
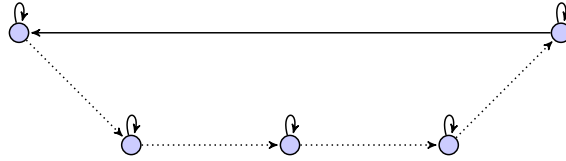


Figure 8.1: Variable gadget

As done in Valiant's reduction, we then introduce a *clause gadget* (see Figure 8.2) for each clause in $\phi'$. Each clause gadget has one external dotted edge for each literal in the clause, and is designed so that no cycle cover can include all the dotted edges; and so that for any other subset of the dotted edges, there is exactly one cycle cover including all the edges in the subset and no others. For each clause gadget, we label each of the three dotted edges in the gadget with one of the three literals appearing in the clause. However, in this step, we ensure that in each gadget the $b \to c$ dotted edge is the one labeled with the literal $\tau$, since this is needed to show that the final construction has a Hamiltonian path. We now "pair" each dotted edge appearing in a clause gadget with a dotted edge corresponding to the same literal in a variable gadget, so that each dotted edge appears in exactly one pair.

We first consider cycle covers which obey the constraint that they must choose *exactly* one edge from each such pair. We claim that the number of cycle covers satisfying this "pairing" constraint equals the number of satisfying assignments of $\phi'$. To see this, we associate a truth assignment with every cycle cover by setting the variable $v$ to *true* if the cycle cover uses all the dotted edges in the variable gadget for $v$, and to *false* if it uses none of the dotted edges. Notice that because of the pairing constraint, a cycle cover is uniquely determined by specifying its assignment. Further, given the above properties of the clause gadget, exactly those cycle covers are permitted whose associated assignments are satisfying assignments of $\phi'$.

We now enforce the "pairing" constraint referred to above using a gadget similar to Valiant's *XOR-gadget*. The XOR-gadget has two *ports* (labeled $a$ and $d$), each of which admits one incoming and one outgoing edge (see Figure 8.3). To ensure the "pairing" constraint for a pair of dotted edges $e_1 \to f_2$ and $e_2 \to f_2$, we replace them by the incoming-outgoing pair of a single XOR-gadget (see Figure 8.3b). The gadget has the property that after the replacement, the weight of every cycle cover which would have included exactly one of the two dotted edges $e_1 \to f_1$ and $e_2 \to f_2$ in the original graph gets multiplied by a factor of 2 (for each replacement made), while the weight
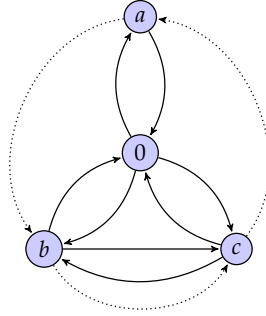
Figure 8.2: 3-SAT clause gadget

of any cycle covers not satisfying the pairing constraint becomes $0$ (see Section 8.3 for a proof). The total weight of all cycle covers in the final graph so obtained is therefore $2^l s'$, where $s'$ is the number of satisfying assignments and $l$ is the total number of literals in $\phi'$ (since one XOR-gadget is needed to replace the pair of dotted edges for each literal). Further, replacing a pair of edges by a XOR gadget does not change the in-degree or out-degree of any vertex already present.

Note that the XOR-gadget has edges of weight $-1$, which are not permitted in the monomer-dimer model. This can be remedied by replacing the $-1$ weight edges by a large chain of edges (of length, say, $m^2$ where $m$ is the number of edges in the original graph) of weight $2$, with individual vertices in the chain having self-loops (of weight $1$). The total weight of cycle covers in the new graph modulo $2^{m^2} + 1$ then gives the total weight of cycle covers in the original graph.

This last step of replacing the $-1$ edge by a long chain presents a challenge since we will need to include all the vertices in the chain in our Hamiltonian path (equivalently, all $-1$ weight edges must appear in the Hamiltonian path). For this reason, we cannot use Valiant's XOR-gadget directly. Our XOR-gadget, on the other hand, is such that the $-1$ weight edges can always be included in our Hamiltonian path. However, we have to be careful in the orientation of the XOR-gadgets in order to be able to construct a Hamiltonian path later: when replacing a pair of dotted edges one of which belongs to $\tau$'s variable gadget, we orient the XOR-gadget so that the incoming edge at vertex $a$ in the XOR-gadget comes from the variable gadget. At all other pairs, we orient the XOR-gadgets so that the incoming edge at the vertex $a$ comes from a clause gadget.

### 8.1.2 Analyzing the reduction

We now proceed to analyze the output of the reduction to complete the proof of Theorem 8.2. The use of XOR-gadgets to enforce the "pairing" constraint as described above introduces a factor of $2$ for each literal appearing in the clause, and therefore the total weight of cycle covers after this step is $2^{3\mu} s' \leq 2^{6\mu}$. To get rid of the $-1$ weight edges in the XOR-gadgets, we replace each such edge by a chain of $\kappa = 6\mu - 1$ vertices with self-loops of weight $1$ and connecting edges of weight $2$. We call this final graph $G$. Since the initial total weight of cycle covers was at most $2^{6\mu}$, the weight of cycle covers in $G$ (and hence the total-weight of all perfect matchings in $\mathrm{Bip}(G)$),

(a) XOR-gadget                              (b) Use of XOR-gadget
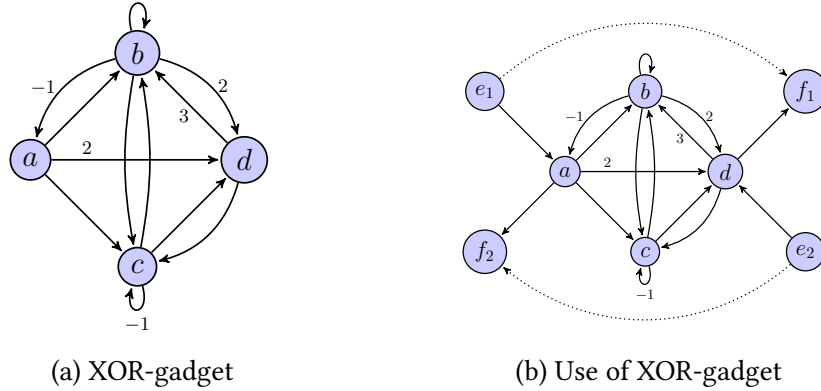
Figure 8.3: Replacing dotted edges $e_1 \to f_1$ and $e_2 \to f_2$

modulo $2^{\kappa+1} + 1$, is exactly $2^{3\mu}s'$. Since all steps in the construction of $\mathrm{Bip}(G)$ starting from $\phi$ can be done in time polynomial in the representation size of $\phi$, this proves parts 1 and 2 of Theorem 8.2.

We now proceed to prove part 3, that is, that $\mathrm{Bip}(G)$ has a Hamiltonian path. We will use the alternating path notation described above in order to keep our discussion in terms of the vertices and edges of $G$, and we will call this representation of a Hamiltonian path in $\mathrm{Bip}(G)$ an *alternating Hamiltonian path*. In an alternating Hamiltonian path, each vertex of $G$ is visited exactly twice, and the length of the alternating path between the two visits is odd. This is equivalent to saying that all vertices must appear exactly *twice* in an alternating Hamiltonian path, with all vertices except the first vertex in the path having one incoming *forward* edge, and one incoming *backward* edge.

Our gadgets so far are designed to have alternating Hamiltonian paths which can be pieced together to form an alternating Hamiltonian path for $G$, and hence, we only need to list these paths and show how to stitch them together. We begin with alternating Hamiltonian paths in the clause gadget.

**Observation 8.4.** *The clause gadget in Figure 8.2 has the alternating Hamiltonian path*

$$c \leftarrow 0 \to a \leftarrow c \to b \leftarrow a \to 0 \leftarrow b$$

*which uses all the dotted edges except the $b \to c$ dotted edge.*

Recall that in the construction of the reduction, we ensured that the new variable $\tau$ was associated with the $b \to c$ dotted edge in each clause gadget. This will be used to connect the above alternating Hamiltonian path in different clause gadgets via connections to the variable gadget for $\tau$ at the $b \to c$ edge. Also, in the final construction, the dotted edges in the alternating Hamiltonian path will be replaced by detours into the associated XOR-gadget.

We now consider the XOR-gadget in Figure 8.3a. It turns out that in some cases, we will need to traverse the XOR-gadget partially, so that a path enters at $a$ via a backward edge, uses the $a \to d$ edge, and then leaves via a backward edge at $d$. In order to cover the rest of the vertices,

we will then need to construct an alternating path that enters at $a$ and leaves at $d$ via *forward* edges, and covers all the vertices except $a$ and $d$ twice. Another complication with the XOR-gadget is the presence of two $-1$ weight edges which need to be replaced with chains of vertices with self-loops. However, this will not be a problem if we can ensure that both of the alternating paths described above use both the $-1$ weight edges, since an edge in an alternating path can always be replaced by a chain of vertices with self-loops. We now show that, as we claimed above, our modified XOR-gadget satisfies all of these conditions.

**Observation 8.5.** *The XOR-gadget in Figure 8.3a has the alternating Hamiltonian path*

$$a \leftarrow b \rightarrow b \leftarrow a \rightarrow d \leftarrow c \rightarrow c \leftarrow d.$$

*The gadget also has the following alternating path which enters at $a$ and leaves from $d$ using forward edges, but which does not otherwise visit these vertices:*

$$a \leftarrow b \rightarrow b \leftarrow c \rightarrow c \leftarrow d.$$

*Moreover, both these paths use the $-1$ weight edges $b \rightarrow a$ and $c \rightarrow c$.*

**Remark 8.2.** Since the XOR-gadget is connected to variable gadgets (except those for variable $\tau$) in $G$ via an outgoing edge at $a$ and an incoming edge at $d$, it will be possible to replace the $a \rightarrow d$ edge in the alternating Hamiltonian path above by a detour into the connected variable gadget when constructing an alternating Hamiltonian path in $G$. Similarly, it will be possible to use the $a \rightarrow d$ edge as a replacement for the dotted edge in the variable gadget that was replaced by the XOR-gadget, at the cost of visiting the vertices $a$ and $d$ once. The role of the second alternating path is to visit the remaining vertices in a XOR-gadget which has already been partially traversed in this way.

**Remark 8.3.** As pointed out above, it does not seem possible to include the two $-1$ weight edges in both the above alternating paths in Valiant's original construction. This necessitated the construction of our new XOR-gadget in which the $-1$ weight edges are part of both the paths. The edges in our construction are the same as those in Valiant's construction, but the weights have been chosen differently.

We now consider the variable gadget shown in Figure 8.1. We first work as if the dotted edges are present. In this case, for any vertex $v$ in the gadget except the leftmost vertex, we can construct an alternating Hamiltonian path which covers all the vertices in the gadget, uses all the dotted edges except the one between $v$ and its predecessor and can be appended to an alternating path that enters via a forward edge at $v$ and leaves via a forward edge at $v$'s predecessor. Now consider such a variable gadget used for a variable other than the special variable $\tau$ (we will see how to traverse the variable gadget for the special variable $\tau$ in the next paragraph). When the dotted edges are replaced by a XOR-gadget, this alternating path can still be traversed as described in the remarks following Observation 8.5, by instead following a forward edge to the $d$ vertex of the XOR-gadget, following the $a \rightarrow d$ edge in reverse, and then entering the variable gadget at the

successor of $v$ via the outgoing edge at the $a$ vertex of the XOR-gadget. It is for this reason that we enforced above the condition that when a XOR-gadget is connected to a variable gadget for a variable other that $\tau$, it is oriented so that the incoming external edge at its $d$ vertex comes from the variable gadget.

We now start constructing the alternating Hamiltonian path in $G$ starting at the left-most vertex in the variable gadget for $\tau$. If all the dotted edges were present, this gadget is just a chain of vertices, and hence there is an alternating Hamiltonian path that covers all its vertices. However, each dotted edge has been replaced by an outgoing edge to the $a$ vertex and an incoming edge from the $d$ vertex of a XOR-gadget. Thus, instead of following the dotted edges, our alternating path will take a detour into the corresponding XOR-gadget, and after traversing several other vertices, return via its $d$ vertex to visit the other vertices in the variable gadget for $\tau$. Thus, we need to show that these detours into the XOR-gadgets can be used to make the alternating path go through all the other vertices in $G$ twice while respecting the required parity constraints.

We consider one such detour. We suppose that the XOR-gadget in question connects to a clause gadget $C$ for the clause $\tau \vee v_1 \vee v_2$. While following the alternating path for the XOR-gadget, we bypass the $a \to d$ edge of the XOR-gadget and instead take a detour into the $c$ vertex of $C$. We then start following the alternating path in Observation 8.4. If the dotted edges $c \to a$ and $a \to b$ (corresponding to the literals $v_1$ and $v_2$, respectively) were present, we would be able to complete an alternating path covering all vertices in $C$ and then return via a forward edge from the $b$ vertex of $C$ into the $d$ vertex of the XOR-gadget. We could then complete the alternating Hamiltonian path in the XOR-gadget, and return via a forward edge into the variable gadget for $\tau$. However, since the dotted edges have been replaced by XOR-gadgets, we would need to take detours into the XOR-gadgets replacing them. Suppose, without loss of generality, that we are trying to replace the dotted edge $c \to a$, corresponding to the literal $v_1$ (the procedure for replacing the dotted edge $a \to b$ corresponding to the literal $v_2$ is exactly the same). At this point there can be two cases:

**Case 1**   Suppose that the vertices of the variable gadget for $v_1$ have still not been covered by our growing alternating Hamiltonian path. Consider the XOR-gadget $X$ replacing the $c \to a$ dotted edge of $C$. We consider the alternating Hamiltonian path in Observation 8.5 starting at the $a$ vertex of $X$. We follow this path until we need to use the $a \to d$ edge. At this point, we take a detour into the variable gadget for $v_1$ via a forward edge at vertex $a$ of $X$. The vertex $u$ we connect to in the variable gadget cannot be a leftmost vertex, since its predecessor $u'$ is connected to vertex $d$ of $X$ via a $u' \to d$ edge. As discussed above, we will therefore get an alternating Hamiltonian path for the vertex gadget which will leave the gadget through the $u' \to d$ edge (though this will end up using the $a \to d$ edges in all other XOR-gadgets corresponding to occurrences of $v_1$). We can then complete the alternating Hamiltonian path for $X$, and this gives us an alternating path starting with a backward edge at vertex $a$ of $X$, ending with a backward edge at vertex $d$ of $X$, and covering all vertices in $X$ and the variable gadget of $v_1$, while also using up the $a \to d$ edge in XOR-gadgets corresponding to all other occurrences of $v_1$. We then use the forward edge from vertex $c$ of $C$ to vertex $a$ of $X$ and the forward edge from vertex $d$ of $X$ to vertex $a$ of $C$ to replace the dotted $c \to a$ edge by the above alternating path.

**Case 2**  Suppose that the variable gadget for the vertex $v_1$ has already been covered by our growing Hamiltonian path. Then, as seen in Case 1, in the XOR-gadget $X$ corresponding to the $c \to a$ edge, the vertex $a$ has already been visited using a backward edge, while $d$ has already been visited via a forward edge. Consider the second alternating path in Observation 8.5. Traversing this alternating path from $a$ to $d$ will satisfy the remaining covering requirements for all the vertices in $X$. Thus, as in Case 1 above, we can replace the $c \to a$ dotted edge in the alternating Hamiltonian path for $C$ by an edge from vertex $c$ of $C$ to vertex $a$ of $X$ and an edge from vertex $d$ of $X$ to vertex $a$ of $X$. As before, this ensures that the vertices of this XOR-gadget are covered while traversing the alternating Hamiltonian path for $C$.

Observe that since each clause gadget is connected to the variable gadget for $\tau$, and since all other variable and XOR-gadgets are connected to at least one of the clause gadgets, the above alternating path eventually covers all of the individual gadgets. This completes the proof for the existence of the alternating Hamiltonian path in $G$, and hence the proof of Theorem 8.2.   □

## 8.2   Hardness for bounded degree graphs

As in the case of the Ising model, our proof of Theorem 8.1 realized different values of $\lambda$ required for the interpolation by attaching $k$ extra vertices to each vertex of $G$, which entailed a large increase in the degree of $G$. In this section, we show that the construction used in the case of the Ising model can me recycled to prove the stronger, degree-bounded version of Theorem 8.1.

As before, we define $P_k$ to be a path of $k$ vertices. We denote by $y_k$ the partition function $Z_M(P_k, \lambda)$, where we assume that all edges in $P_k$ have weight one, and suppress the dependence on edge weights for clarity of notation. Note that $y_1 = \lambda$. We further define $y_0 = 1$. We then have the following recurrence for $y_k$, which shows that it can be computed in time polynomial in $k$:

$$y_k = \lambda y_{k-1} + y_{k-2}. \tag{8.3}$$

Notice that $y_k$ is a function of $\lambda$. We now show that the value of its derivative with respect to $\lambda$ can also be computed in time polynomial in $k$ via the following recurrence: $\dot{y}_0 = 0, \dot{y}_1 = 1$, and

$$\dot{y}_k = y_{k-1} + \lambda \dot{y}_{k-1} + \dot{y}_{k-2}. \tag{8.4}$$

Here, we use the dot notation for the derivative with respect to $\lambda$.

As before, given a connected graph $G$, we define $G(k)$ as the graph obtained by attaching to each vertex $v$ of $G$ a different instance of the path $P_k$, such that $v$ is connected to the "leftmost" vertex of $P_k$ via an edge. The maximum degree of $G(k)$ is thus one more than the maximum degree of $G$. We now consider the monomer-dimer model on the graphs $G(k)$. We have

$$Z_M(G(k), \lambda) = y_k^n Z_M(G, \lambda_k), \tag{8.5}$$

where $n$ is the number of vertices in $G$ and $\lambda_k = y_{k+1}/y_k$. We also have

$$U(G(k), \lambda) = n\lambda t_k + \lambda(t_{k+1} - t_k)U(G, \lambda_k), \tag{8.6}$$

where $t_k = \dot{y}_k/y_k$. It turns out that the sequence $(\lambda_{2k})_{k\geq 0}$ is strictly increasing and hence consists of distinct values, and further that $t_{2k+1} - t_{2k} > 0$ for all $k$. This follows easily from the following explicit solutions for the $y_k$ and the $\lambda_k$:

$$y_k = \frac{\xi^{k+1} - \eta^{k+1}}{\xi - \eta}; \qquad \lambda_k = \frac{\xi^{k+2} - \eta^{k+2}}{\xi^{k+1} - \eta^{k+1}},$$

where

$$\xi = \frac{1}{2}\left(\lambda + \sqrt{\lambda^2 + 4}\right) > 0; \qquad \eta = \frac{1}{2}\left(\lambda - \sqrt{\lambda^2 + 4}\right) < 0.$$

Notice that $t_{k+1} - t_k > 0$ for even $k$ implies that for such $k$, we can determine $U(G, \lambda_k)$ given $U(G(k), \lambda)$, using equations (8.3), (8.4) and (8.6). We can now complete the proof of Theorem 8.1 for the bounded degree case.

*Proof of Theorem 8.1.* As in the proof for the case of unbounded degree graphs (see page 78), we fix any $\lambda > 0$, and suppose that there exists a polynomial time algorithm $\mathcal{B}$ which, given a connected graph $H$ with edge weights in the set $\{1, 2, 3\}$, and of maximum degree $\Delta \geq 5$, outputs $U(H, \lambda)$ (recall that we are suppressing explicit dependence on the edge weights for clarity of notation). Given a MONOTONE 2-SAT formula $\phi$, we then produce the graph $G = \mathcal{A}(\phi)$ in polynomial time. Notice that in the construction of $G$ as given in Section 8.1.1, each vertex has degree at most 4: this corresponds to the maximum of the in-degrees and the out-degrees over all vertices in the directed version of the reduction.

As argued in the proof for the case of unbounded degree graphs, Theorem 6.3 and the existence of a Hamiltonian path in $G$ together imply that if we could use algorithm $\mathcal{B}$ to evaluate $U(G, z)$ at $2n + 2$ different values of $z$, then we can determine the number of satisfying assignments of $\phi$ in polynomial time. This would in turn imply that computing $U(H, \lambda)$ for graphs $H$ of maximum degree at least 5 is #P-hard.

As before, in order to realize other values of $\lambda$, we consider the graphs $G(k)$ (as described in this section), for $k = 0, 2, 4, \ldots 4n + 4$. Notice that the maximum degree of $G(k)$ is one more than that of $G$, and hence is at most 5. Further, as argued in the remarks following eqs. (8.5) and (8.6), these choices of $k$ ensure that the values $\lambda_k$ are distinct, and that $U(G, \lambda_k)$ can be easily determined from $U(G(k), \lambda)$. We can therefore determine $U(G, z)$ at $2n + 2$ different values of $z$ by running $\mathcal{B}$ on the $G(k)$, as required. $\square$

## 8.3 The XOR-gadget

We conclude with a proof of the properties of the XOR-gadget used in the reduction in Section 8.1.1.

**Claim 8.6.** *The total weight of cycle covers of the XOR-gadget in Figure 8.3a is* 2 *when either*

- *$a$ is connected to an external incoming edge and $d$ is connected to an external outgoing edge;*

*or*

- *$a$ is connected to an external outgoing edge and $d$ is connected to an external incoming edge.*

*For all other external connections of $a$ and $d$, the total weight of cycle covers of the gadget is $0$.*

*Proof.* When the total number of incoming external edges at $a$ and $d$ is not equal to the total number of outgoing external edges, the XOR-gadget cannot admit a cycle cover due to parity constraints, and thus, the total weight of all cycle covers in these cases is trivially zero. A simple way to see this is that a cycle cover corresponds to a perfect matching in the natural undirected bipartite representation of the gadget discussed above. When the numbers of external incoming and outgoing edges are not equal, the bipartite graph does not remain balanced and hence cannot have a perfect matching. For all other configurations, in which the number of external incoming and outgoing edges are equal, the weights of all cycle covers can be shown to have the claimed value by exhaustive enumeration. □

# Chapter 9

# Related work and open problems

Extensive research has been done on the classification of partition functions based on the computational complexity of their exact computation. For several interesting general classes of spin systems, *dicohotomy theorems* are known which characterize the partition function of a given spin system as being either computable in polynomial time or #P-hard [CC12, GGJT10, CCL10, Bul06, BG05, DG00b]. However, there appear to be no analogous results on the complexity of computing averages such as the magnetization. In contrast, for both the mean magnetization of the ferromagnetic Ising model and for the average monomer count of the monomer-dimer model, randomized *approximation* algorithms (which give an additive approximation) follow from the corresponding MCMC based algorithms for sampling from the Gibbs distribution [JS89, JS93]. However, as we saw in Part I, the complexity of the sampling problem for anti-ferromagnetic spin systems such as the hard core model and the anti-ferromagnetic Ising model is closely related to the uniqueness phase transition [Wei06, Sly10, SS12, SST14]. It is an open questions whether hardness results analogous to those in [Sly10, SS12] can be proven for the approximate computation of averages related to these models.

Even the question of the exact computation of the mean observables for these spin systems remains open. The methods of the preceding chapters cannot be directly applied, since results completely characterizing the location of zeros of the partition function of these systems are not available. For example, the best known such result for the partition function of the hard core model—due to Chudnovsky and Seymour [CS07]—is only applicable to claw-free graphs (note that this later result is already a strict generalization of the Heilmann-Lieb theorem). It remains open whether results such as those of Chudnovsky and Seymour can be exploited to prove #P-hardness results for the mean observables of the hard core model; note that our strategy of reducing from a partition function computation via an interpolation operation is unlikely to work directly, since we do not expect the problem itself to be hard on claw-free graphs.

The study of the location of zeros of the partition function itself was initiated by Yang and Lee [YL52] in connection with the analysis of phase transitions. In the follow-up paper [LY52], they instantiated this approach for the ferromagnetic Ising model by proving the celebrated results that is now known as the Lee-Yang theorem and using it to conclude that the ferromagnetic Ising model can have at most one phase transition. The Lee-Yang approach has since become

a cornerstone of the study of phase transitions, and has been used extensively in the statistical physics literature: see, e.g., [Asa70, HL72, New74, SF71, BBCK04, BBC$^+$04] for specific examples, and Ruelle's book [Rue83] for background.

Zeros of partition functions have also been studied in a purely combinatorial setting without reference to the physical interpretation: see, for example, Choe *et al.* [COSW04] for a collection of such results about zeros of a general class of partition functions. Another important example is the work of Chudnovsky and Seymour [CS07] cited above, who show that the zeros of the independence polynomial of claw-free graphs lie on the real line. A result of a somewhat different flavor is that of Ruelle [Rue10], who provides a characterization of polynomials for which the Lee-Yang theorem holds. There have also been attempts to relate the Lee-Yang program to the Riemann hypothesis [New91]. However, we are aware of only two works which consider the multiplicity of the zeros of the Ising partition function: Heilmann and Lieb [HL72] and Biskup *et al.* [BBC$^+$04, BBCK04]. In [HL72], a theorem similar to our Theorem 7.3 is proven in the special case when the underlying graph $G$ has a Hamiltonian path and $\beta$ is close enough to $1$ (depending upon the graph $G$). Similarly, in the special case of the Ising model, the results of [BBC$^+$04] imply our result but only when $\beta$ is close to $0$, and only in the special case of lattice graphs [BBCK04, Bis12]. Note that neither of these results appears to be sufficient for the purposes of our hardness result.

Lee-Yang theorems have also been studied in mathematics in connection with the theory of *stability preserving* operators. The main problem underlying this area is the characterization of linear operators that preserve the class of polynomials, called $\Omega$-*stable* polynomials, which are guaranteed not to vanish when their arguments lie in some fixed set $\Omega$. This research area has its origins in the work of Laguerre [Lag82] and of Pólya and Schur [PS14], and also has connections to control theory [Cla92] and to electrical circuit theory [Bru31]. It has also seen considerable recent activity, especially through the breakthrough results of Borcea and Brändén, who completely characterize stability preserving operators for multivariate polynomials in various important settings [BB09a, BB09b].

Part of this activity stems from the fact that stability theory provides some very powerful tools for various combinatorial problems; we conclude with three different examples of its applications. As our first example, we refer to the work of Scott and Sokal [SS05], who show that $\mathcal{D}$-stability of the multivariate hard core partition function (where each vertex has a different fugacity) for $\mathcal{D}$ a product of origin-centered disks in the complex plane, has very strong connections to the Lovász Local Lemma. More formally, given an undirected graph $G = (V, E)$, the multivariate hard core partition function of $G$ is the polynomial

$$Z((z_v)_{v \in V}) := \sum_{I:\text{ independent set in } G} \prod_{v \in I} z_v.$$

Scott and Sokal show that this polynomial is stable with respect to the product of closed discs

$$\prod_{v \ in V} D(0, p_v)$$

of radii $p_v$ around the origin if and only if the Lovász Local Lemma holds on the graph $G$ with the probabilities $(p_v)_{v \in V}$ at the vertices.

Another example concerns the probability generating functions (pgf) of multivariate probability distributions. In this case, stability of the pgf with respect to the half plane $\Im(z) > 0$ implies *negative association* properties for the underlying distribution [BBL09]. An important example of a distribution satisfying such a property is the uniform distribution over the spanning trees of a graph. Coupled with the list of various algebraic operations which preserve stability and their natural probabilistic interpretations when the polynomial being studied is a pgf, this connection provides a set of very powerful tools for studying such distributions. An example of the use of these tools are the breakthrough results of Oveis Gharan, Saberi and Singh [OGSS11] for the traveling salesperson problem.

Our concluding example is the work of Marcus, Spielman and Srivastava [MSS13a], who developed techniques based on stability theory to study the eigenvalues of 2-lifts of graphs. These were then used to prove the existence of bipartite Ramanujan expanders of every given degree. Marcus *et al.* later extended their techniques to give a positive resolution of the Kadison-Singer conjecture [MSS13b].

# Bibliography

[AB09]       Sanjeev Arora and Boaz Barak. *Computational Complexity: A Modern Approach.* Cambridge University Press, 2009.

[Alm05]      Sven Erick Alm. Upper and lower bounds for the connective constants of self-avoiding walks on the Archimedean and Laves lattices. *J. Phys. A*, 38(10):2055–2080, March 2005.

[Asa70]      Taro Asano. Lee-Yang theorem and the Griffiths inequality for the anisotropic Heisenberg ferromagnet. *Phys. Rev. Lett.*, 24(25):1409–1411, June 1970.

[BB09a]      Julius Borcea and Petter Brändén. The Lee-Yang and Pólya-Schur programs. I. Linear operators preserving stability. *Invent. Math.*, 177(3):541–569, 2009.

[BB09b]      Julius Borcea and Petter Brändén. Pólya-Schur master theorems for circular domains and their boundaries. *Ann. Math.*, 170(1):465–492, July 2009.

[BBC$^+$04]  M. Biskup, C. Borgs, J. Chayes, L. Kleinwaks, and R. Kotecký. Partition function zeros at first-order phase transitions: A general analysis. *Commun. Math. Phys.*, 251(1):79–131, October 2004.

[BBCK04]     M. Biskup, C. Borgs, J. Chayes, and R. Kotecký. Partition function zeros at first-order phase transitions: Pirogov-Sinai theory. *J. Stat. Phys.*, 116(1):97–155, 2004.

[BBL09]      Julius Borcea, Petter Brändén, and Thomas Liggett. Negative dependence and the geometry of polynomials. *J. Amer. Math. Soc.*, 22(2):521–567, 2009.

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

[BG08]       Antar Bandyopadhyay and David Gamarnik. Counting without sampling: Asymptotics of the log-partition function for certain statistical physics models. *Random Struct. Algorithms*, 33(4):452–479, 2008.

[BGK$^+$07]  Mohsen Bayati, David Gamarnik, Dimitriy Katz, Chandra Nair, and Prasad Tetali. Simple deterministic approximation algorithms for counting matchings. In *Proc. 39th ACM Symp. Theory Comput. (STOC)*, page 122–127. ACM, 2007.

[Bis12]    Marek Biskup, February 2012. Personal Communication.

[Bru31]    Otto Brune. *Synthesis of a finite two-terminal network whose driving-point impedance is a prescribed function of frequency.* Thesis (Sc. D.), Dept. of Electrical Engineering, Massachusetts Institute of Technology, 1931.

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

[CC12]    Jin-Yi Cai and Xi Chen. Complexity of counting CSP with complex weights. In *Proc. 44th ACM Symp. Theory Comput. (STOC)*, pages 909–920, New York, NY, USA, 2012. ACM.

[CCL10]    Jin-Yi Cai, Xi Chen, and Pinyan Lu. Graph homomorphisms with complex values: A dichotomy theorem. In *Automata, Languages and Programming*, volume 6198 of *Lecture Notes in Computer Science*, pages 275–286. 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 *Theory and Applications of Models of Computation*, volume 6108 of *Lecture Notes in Computer Science*, pages 328–339. Springer, 2010.

[Cla92]    R.N. Clark. The Routh-Hurwitz stability criterion, revisited. *IEEE Control Syst.*, 12(3):119–120, June 1992.

[COSW04]    Young-Bin Choe, James G. Oxley, Alan D. Sokal, and David G. Wagner. Homogeneous multivariate polynomials with the half-plane property. *Adv. Appl. Math.*, 32(1–2):88–187, February 2004.

[CS07]    Maria Chudnovsky and Paul Seymour. The roots of the independence polynomial of a clawfree graph. *J. Comb. Theory Ser. B*, 97(3):350–357, May 2007.

[DCS12]    Hugo Duminil-Copin and Stanislav Smirnov. The connective constant of the honeycomb lattice equals $\sqrt{2 + \sqrt{2}}$. *Ann. Math.*, 175(3):1653–1665, May 2012.

[DG00a]    Martin Dyer and Catherine Greenhill. On Markov chains for independent sets. *J. Algorithms*, 35(1):17–49, April 2000.

[DG00b]    Martin E. Dyer and Catherine S. Greenhill. The complexity of counting graph homomorphisms. *Random Structures and Algorithms*, 17(3-4):260–289, 2000.

[Edm67]    Jack Edmonds. Systems of distinct representatitives and linear algebra. *J. Res. Natl. Bur. Stand. B Math. & Math. Phys.*, 71B(4):241–245, 1967.

[Eft14]    Charilaos Efthymiou. MCMC sampling colourings and independent sets of $\mathcal{G}(n, d/n)$ near uniqueness threshold. In *Proc. 25th ACM-SIAM Symp. Discret. Algorithms (SODA)*, pages 305–316. SIAM, 2014. Full version available at arXiv:1304.6666.

[Geo88]   Hans-Otto Georgii. *Gibbs Measures and Phase Transitions.* De Gruyter Studies in Mathematics. Walter de Gruyter Inc., October 1988.

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

[GGŠ⁺14]  Andreas Galanis, Qi Ge, Daniel Štefankovič, Eric Vigoda, and Linji Yang. Improved inapproximability results for counting independent sets in the hard-core model. *Random Struct. Algorithms*, 45(1):78–110, 2014.

[GJP03]   L. A. Goldberg, M. Jerrum, and M. Paterson. The computational complexity of two-state spin systems. *Random Struct. Algorithms*, 23:133–154, 2003.

[GK07]    David Gamarnik and Dmitriy Katz. Correlation decay and deterministic FPTAS for counting list-colorings of a graph. In *Proc. 18th ACM-SIAM Symp. Discret. Algorithms (SODA)*, pages 1245–1254. SIAM, 2007.

[God81]   C. D. Godsil. Matchings and walks in graphs. *J. Graph Th.*, 5(3):285–297, September 1981.

[Gri67]   Robert B. Griffiths. Correlations in Ising ferromagnets. I. *J. Math. Phys.*, 8(3):478–483, March 1967.

[Ham57]   J. M. Hammersley. Percolation processes II. The connective constant. *Math. Proc. Camb. Philos. Soc.*, 53(03):642–645, 1957.

[HL70]    Ole J. Heilmann and Elliott H. Lieb. Monomers and dimers. *Phys. Rev. Let.*, 24(25):1412–1414, June 1970.

[HL72]    Ole J. Heilmann and Elliott H. Lieb. Theory of monomer-dimer systems. *Comm. Math. Phys.*, 25(3):190–232, 1972.

[Isi25]   E. Ising. Beitrag zur Theorie des Ferromagnetismus. *Z. Phys.*, 31:253–258, February 1925.

[Jer95]   Mark Jerrum. A very simple algorithm for estimating the number of $k$-colorings of a low-degree graph. *Random Struct. Algorithms*, 7(2):157–165, September 1995.

[JS89]    M. Jerrum and A. Sinclair. Approximating the permanent. *SIAM J. Comput.*, 18(6):1149–1178, December 1989.

[JS93]    Mark Jerrum and Alistair Sinclair. Polynomial-time approximation algorithms for the Ising model. *SIAM J. Comput.*, 22(5):1087–1116, 1993.

[JVV86]   Mark Jerrum, Leslie G. Valiant, and Vijay V. Vazirani. Random generation of combinatorial structures from a uniform distribution. *Theor. Comput. Sci.*, 43:169–188, 1986.

[KK98]   Jeff Kahn and Jeong Han Kim. Random matchings in regular graphs. *Combinatorica*, 18(2):201–226, February 1998.

[Lag82]   Edmond Laguerre. Sur les fonctions du genre zéro et du genre un. *C. R. Séances Acad. Sci.*, 98:828–831, 1882.

[LLL13]   C. Lin, J. Liu, and P. Lu. A simple FPTAS for counting edge covers. In *Proc. 25th ACM-SIAM Symp. Discret. Algorithms (SODA)*, pages 341–348. SIAM, December 2013.

[LLY12]   Liang Li, Pinyan Lu, and Yitong Yin. Approximate counting via correlation decay in spin systems. In *Proc. 23rd ACM-SIAM Symp. Discret. Algorithms (SODA)*, pages 922–940. SIAM, 2012.

[LLY13]   Liang Li, Pinyan Lu, and Yitong Yin. Correlation decay up to uniqueness in spin systems. In *Proc. 24th ACM-SIAM Symp. Discret. Algorithms (SODA)*, pages 67–84. SIAM, 2013.

[LY52]   T. D. Lee and C. N. Yang. Statistical theory of equations of state and phase transitions. II. Lattice gas and Ising model. *Phys. Rev.*, 87(3):410–419, August 1952.

[LY13]   Pinyan Lu and Yitong Yin. Improved FPTAS for multi-spin systems. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, volume 8096 of *Lecture Notes in Computer Science*, pages 639–654. Springer, January 2013.

[Lyo89]   Russell Lyons. The Ising model and percolation on trees and tree-like graphs. *Commun. Math. Phys.*, 125(2):337–353, 1989.

[MD62]   Nathaniel Macon and D. E. Dupree. Existence and uniqueness of interpolating rational functions. *Am. Math. Mon.*, 69(8):751–759, October 1962.

[MO94a]   F. Martinelli and E. Olivieri. Approach to equilibrium of Glauber dynamics in the one phase region I. The attractive case. *Comm. Math. Phys.*, 161(3):447–486, April 1994.

[MO94b]   F. Martinelli and E. Olivieri. Approach to equilibrium of Glauber dynamics in the one phase region II. The general case. *Comm. Math. Phys.*, 161(3):487–514, April 1994.

[MS96]   Neal Madras and Gordon Slade. *The Self-Avoiding Walk*. Birkhäuser, 1996.

[MS09]   Elchanan Mossel and Allan Sly. Rapid mixing of Gibbs sampling on graphs that are sparse on average. *Random Struct. Algorithms*, 35(2):250–270, 2009.

[MS13]   Elchanan Mossel and Allan Sly. Exact thresholds for Ising–Gibbs samplers on general graphs. *Ann. Probab.*, 41(1):294–328, January 2013.

[MSS13a]   Adam Marcus, Daniel A. Spielman, and Nikhil Srivastava. Interlacing families i: Bipartite ramanujan graphs of all degrees. In *Proc. 54th IEEE Symp. Found. Comput. Sci. (FOCS)*, pages 529–537. IEEE Computer Society, 2013. Full version available at arXiv:1304.4132.

[MSS13b]   Adam Marcus, Daniel A. Spielman, and Nikhil Srivastava. Interlacing families II: Mixed characteristic polynomials and the Kadison-Singer problem, June 2013. Available at arXiv:1306.3969.

[New74]   Charles M. Newman. Zeros of the partition function for generalized Ising systems. *Commun. Pure Appl. Math.*, 27(2):143–159, March 1974.

[New91]   Charles M. Newman. The GHS inequality and the Riemann hypothesis. *Constr. Approx.*, 7(1):389–399, December 1991.

[OGSS11]   S. Oveis Gharan, A. Saberi, and M. Singh. A randomized rounding approach to the traveling salesman problem. In *Proc. 52nd IEEE Symp. Found. Comput. Sci. (FOCS)*, pages 550–559. IEEE Computer Society, 2011.

[Ons44]   Lars Onsager. Crystal statistics. I. A two-dimensional model with an order-disorder transition. *Phys. Rev. (2)*, 65:117–149, 1944.

[Pap94]   Christos H. Papadimitriou. *Computational Complexity*. Addison-Wesley Publishing Company, Reading, MA, 1994.

[PS14]   G. Pólya and I. Schur. Über zwei Arten von Faktorenfolgen in der Theorie der Algebraischen Gleichungen. *J. Reine Angew. Math.*, 144:89–113, 1914.

[PS99]   Robin Pemantle and Jeffrey E. Steif. Robust phase transitions for Heisenberg and other models on general trees. *Ann. Probab.*, 27(2):876–912, April 1999.

[PT]   André Pönitz and Peter Tittmann. Improved upper bounds for self-avoiding walks in $\mathbb{Z}^d$. *Electron. J. Comb.*, vol. 7, Research Paper 21, 2000.

[RST+13]   Ricardo Restrepo, Jinwoo Shin, Prasad Tetali, Eric Vigoda, and Linji Yang. Improved mixing condition on the grid for counting and sampling independent sets. *Probab. Theory Relat. Fields*, 156(1-2):75–99, June 2013.

[Rue83]   David Ruelle. *Statistical Mechanics*. W. A. Benjamin Inc., 1983.

[Rue10]   David Ruelle. Characterization of Lee-Yang polynomials. *Ann. Math.*, 171(1):589–603, March 2010.

[SF71]   Masuo Suzuki and Michael E Fisher. Zeros of the partition function for the Heisenberg, ferroelectric, and general Ising models. *J. Math. Phys.*, 12(2):235–246, February 1971.

[Sim93]     Barry Simon. *The Statistical Mechnaics of Lattice Gases*, volume 1. Princeton University Press, 1993.

[Sly08]     Allan Sly. Uniqueness thresholds on trees versus graphs. *Ann. Appl. Probab.*, 18(5):1897–1909, October 2008.

[Sly10]     Allan Sly. Computational transition at the uniqueness threshold. In *Proc. 51st IEEE Symp. Found. Comput. Sci. (FOCS)*, pages 287–296. IEEE Computer Society, 2010.

[SS05]      Alexander D. Scott and Alan D. Sokal. The repulsive lattice gas, the independent-set polynomial, and the Lovász local lemma. *J. Stat. Phys.*, 118(5-6):1151–1261, March 2005.

[SS12]      A. Sly and N. Sun. The computational hardness of counting in two-spin models on d-regular graphs. In *Proc. 53rd IEEE Symp. Found. Comput. Sci. (FOCS)*, pages 361–369, Oct 2012. Full version available at arxiv:1203.2602.

[SS14]      Alistair Sinclair and Piyush Srivastava. Lee–Yang theorems and the complexity of computing averages. *Comm. Math. Phys.*, 329(3):827–858, 2014.

[SSŠY14]    Alistair Sinclair, Piyush Srivastava, Daniel Štefankovič, and Yitong Yin. Spatial mixing and the connective constant: Optimal bounds. Submitted for review, 2014.

[SST14]     Alistair Sinclair, Piyush Srivastava, and Marc Thurley. Approximation algorithms for two-state anti-ferromagnetic spin systems on bounded degree graphs. *J. Stat. Phys.*, 155(4):666–686, 2014.

[SSY13]     Alistair Sinclair, Piyush Srivastava, and Yitong Yin. Spatial mixing and approximation algorithms for graphs with bounded connective constant. In *Proc. 54th IEEE Symp. Found. Comput. Sci. (FOCS)*, pages 300–309. IEEE Computer Society, 2013. Full version available at arXiv:1308.1762v1.

[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.

[Vig00]     Eric Vigoda. Improved bounds for sampling colorings. *J. Math. Phys.*, 41(3):1555–1569, March 2000.

[Vig01]     Eric Vigoda. A note on the Glauber dynamics for sampling independent sets. *Electron. J. Combin.*, 8(1):R8, January 2001.

[VVY13]   Juan C. Vera, Eric Vigoda, and Linji Yang. Improved bounds on the phase transition for the hard-core model in 2-dimensions. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, volume 8096 of *Lecture Notes in Computer Science*, pages 699–713. Springer Berlin Heidelberg, 2013.

[Wei]   Eric W. Weisstein.   Self-avoiding walk connective constant.   From *Math-World* – A Wolfram Web Resource. [http://mathworld.wolfram.com/Self-AvoidingWalkConnectiveConstant.html](http://mathworld.wolfram.com/Self-AvoidingWalkConnectiveConstant.html).

[Wei06]   Dror Weitz. Counting independent sets up to the tree threshold. In *Proc. 38th ACM Symp. Theory Comput. (STOC)*, pages 140–149. ACM, 2006.

[YL52]   C. N. Yang and T. D. Lee. Statistical theory of equations of state and phase transitions. I. Theory of condensation. *Phys. Rev.*, 87(3):404–409, August 1952.