

Spatial mixing and the connective constant: optimal bounds

Alistair Sinclair¹ · Piyush Srivastava² ·
Daniel Štefankovič³ · Yitong Yin⁴

Received: 10 July 2015 / Revised: 15 March 2016 / Published online: 14 July 2016
© Springer-Verlag Berlin Heidelberg 2016

Abstract We study the problem of deterministic approximate counting of matchings and independent sets in graphs of bounded *connective constant*. More generally, we consider the problem of evaluating the partition functions of the *monomer-dimer model* (which is defined as a weighted sum over all matchings where each matching is given a

Slightly weaker versions of some of the results in this paper appeared in an extended abstract in *Proceedings of the IEEE Symposium on the Foundations of Computer Science (FOCS)*, 2013, pp. 300–309 [44]. This version strengthens the main result (Theorem 1.3) of [44] to obtain an optimal setting of the parameters, and adds new results for the monomer-dimer model. An extended abstract of the current results appeared in the *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms*, 2015, pp. 1549–1563.

AS was supported in part by NSF grant CCF-1016896 and by the Simons Institute for the Theory of Computing. PS was supported by NSF grant CCF-1319745 and NSF grant CCF-1016896. DŠ was supported by NSF grant CCF-1016896 and was visiting the Simons Institute for the Theory of Computing when this work was done. YY was supported by NSFC Grants 61272081 and 61321491 and did part of this work while he was visiting UC Berkeley.

✉ Piyush Srivastava
piyushsri@gmail.com

Alistair Sinclair
sinclair@cs.berkeley.edu

Daniel Štefankovič
stefanko@cs.rochester.edu

Yitong Yin
yinyt@nju.edu.cn

- ¹ Computer Science Division, UC Berkeley, Berkeley, USA
- ² Center for the Mathematics of Information, Caltech, Pasadena, USA
- ³ Department of Computer Science, University of Rochester, Rochester, USA
- ⁴ State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing, China

weight $\gamma^{|V|-2|M|}$ in terms of a fixed parameter γ called the *monomer activity*) and the *hard core model* (which is defined as a weighted sum over all independent sets where an independent set I is given a weight $\lambda^{|I|}$ in terms of a fixed parameter λ called the *vertex activity*). The *connective constant* is a natural measure of the average degree of a graph which has been studied extensively in combinatorics and mathematical physics, and can be bounded by a constant even for certain unbounded degree graphs such as those sampled from the sparse Erdős–Rényi model $\mathcal{G}(n, d/n)$. Our main technical contribution is to prove the best possible rates of decay of correlations in the natural probability distributions induced by both the hard core model and the monomer-dimer model in graphs with a given bound on the connective constant. These results on decay of correlations are obtained using a new framework based on the so-called *message* approach that has been extensively used recently to prove such results for bounded degree graphs. We then use these optimal decay of correlations results to obtain fully polynomial time approximation schemes (FPTASs) for the two problems on graphs of bounded connective constant. In particular, for the monomer-dimer model, we give a deterministic FPTAS for the partition function on all graphs of bounded connective constant for any given value of the monomer activity. The best previously known deterministic algorithm was due to Bayati et al. (Proc. 39th ACM Symp. Theory Comput., pp. 122–127, 2007), and gave the same runtime guarantees as our results but only for the case of bounded degree graphs. For the hard core model, we give an FPTAS for graphs of connective constant Δ whenever the vertex activity $\lambda < \lambda_c(\Delta)$, where $\lambda_c(\Delta) := \frac{\Delta^\Delta}{(\Delta-1)^{\Delta+1}}$; this result is optimal in the sense that an FPTAS for any $\lambda > \lambda_c(\Delta)$ would imply that NP=RP (Sly and Sun, Ann. Probab. 42(6):2383–2416, 2014). The previous best known result in this direction was in a recent manuscript by a subset of the current authors (Proc. 54th IEEE Symp. Found. Comput. Sci., pp 300–309, 2013), where the result was established under the sub-optimal condition $\lambda < \lambda_c(\Delta + 1)$. Our techniques also allow us to improve upon known bounds for decay of correlations for the hard core model on various regular lattices, including those obtained by Restrepo et al. (Probab Theory Relat Fields 156(1–2):75–99, 2013) for the special case of \mathbb{Z}^2 using sophisticated numerically intensive methods tailored to that special case.

Mathematics Subject Classification 82B20 · 60J10 · 68W25 · 68W40

1 Introduction

1.1 Background

This paper studies the problem of approximately counting independent sets and matchings in sparse graphs. We consider these problems within the more general formalism of *spin systems*. In this setting, one first defines a natural probability distribution over *configurations* (e.g., independent sets or matchings) in terms of local *interactions*. The counting problem then corresponds to computing the normalization constant, known as the *partition function* in the statistical physics literature. The partition function can

also be seen as a generating function of the combinatorial structures being considered and is an interesting graph polynomial in its own right.

The first model we consider is the so called *hard core model*, which is defined as follows. We start with a graph $G = (V, E)$, and specify a *vertex activity* or *fugacity* parameter $\lambda > 0$. The configurations of the hard core model are the independent sets of the graph, and the model assigns a *weight* $w(I) = \lambda^{|I|}$ to each independent set I in G . The weights in turn determine a natural probability distribution $\mu(I) = \frac{1}{Z}w(I)$ over the independent sets known as the *Gibbs distribution*. Here,

$$Z = Z(\lambda) := \sum_{I:\text{independent set}} w(I)$$

is the *partition function*. Clearly, the problem of counting independent sets is the special case $\lambda = 1$.

Our next model is the *monomer-dimer model*, which has as its configurations 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

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

is the partition function. The problem of counting matchings again corresponds to the special case $\gamma = 1$.

The problem of approximating the partition function has received much attention, both as a natural generalization of counting and because of its connections to sampling.¹ Recent progress in relating the complexity of approximating the partition function to phase transitions, which we now describe, has provided further impetus to this line of research.

The first such result was due to Weitz [50], who exploited the properties of the Gibbs measure of the hard core model on the infinite d -ary tree. It was well known that this model exhibits the following phase transition: there exists a *critical activity* $\lambda_c(d)$ such that the total variation distance between the marginal probability distributions induced at the root of the tree by any two fixings of the independent set on all the vertices at distance ℓ from the root decays exponentially in ℓ when $\lambda < \lambda_c(d) := \frac{d^d}{(d-1)^{d+1}}$, but remains bounded away from 0 even as $\ell \rightarrow \infty$ when $\lambda > \lambda_c(d)$ (the former condition is also referred to as *correlation decay*, since the correlation between the configuration at the root of the tree and a fixed configuration at distance ℓ from the root decays exponentially in ℓ ; it is also called *spatial mixing*). Weitz showed that for all $\lambda < \lambda_c(d)$ (i.e., in the regime where correlation decay holds on the d -ary tree), 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 $d + 1$ (note that the condition

¹ For the large class of self-reducible problems, it can be shown that approximating the partition function is polynomial-time equivalent to approximate sampling from the Gibbs distribution [24].

on λ is only in terms of the d -ary tree, while the FPTAS applies to all graphs). This connection to phase transitions was further strengthened by Sly [45] (see also [12,46]), who showed that a *fully polynomial randomized approximation algorithm* (FPRAS) for the partition function of the hard core model with $\lambda > \lambda_c(d)$ on graphs of degree $d + 1$ would imply $\text{NP} = \text{RP}$.²

In addition to establishing a close connection between the complexity of a natural computational problem and an associated phase transition, Weitz's algorithm had the further interesting feature of not being based on Markov chain Monte Carlo (MCMC) methods; rather, it used a deterministic procedure based on proving that decay of correlations on the d -ary tree implies decay of correlations on all graphs of degree at most $d + 1$. To date, no MCMC algorithms are known for the approximation of the partition function of the hard core model on graphs of degree at most $d + 1$ which run in polynomial time for all $\lambda < \lambda_c(d)$.

Weitz's algorithm led to an exploration of his approach for other problems too. For example, in the case of the monomer-dimer model—unlike that of the hard core model—there does exist a *randomized* polynomial time algorithm (based on MCMC) for approximating the partition function which works for every $\gamma > 0$, without any bounds on the degree of the graph [23]. However, finding a deterministic algorithm for the problem remains open. Bayati et al. [7] made progress on this question by showing that Weitz's approach could be used to derive a deterministic algorithm that runs in polynomial time for bounded degree graphs, and is sub-exponential on general graphs.

The algorithms of both Weitz and Bayati et al. are therefore polynomial time only on bounded degree graphs, and in particular, for a given value of the parameter λ (or γ in the case of the monomer-dimer model) the running time of these algorithms on graphs of maximum degree $d + 1$ depends upon the rate of decay of correlations on the infinite d -ary tree. Further, these results are obtained by showing that decay of correlations on the d -ary tree implies a similar decay on all graphs of maximum degree $d + 1$.

There are two important shortcomings of such results. First, in statistical physics one is often interested in special classes of graphs such as regular lattices. One can reasonably expect that the rate of decay of correlations on such graphs should be better than that predicted by their maximum degree. Second, these results have no non-trivial consequences even in very special classes of sparse unbounded degree graphs, such as graphs drawn from the Erdős-Rényi model $\mathcal{G}(n, d/n)$ for constant d .

This state of affairs leads to the following natural question: is there a finer notion of degree that can be used in these results in place of the maximum degree? Progress in this direction was made recently for the case of the hard core model in [44], where it was shown that one can get decay of correlation results in terms of the *connective constant*, a natural and well-studied notion of average degree. The connective constant of a regular

² An FPTAS for a quantity \mathcal{A} is a deterministic algorithm which, given an accuracy parameter $\epsilon > 0$, outputs in time polynomial in $1/\epsilon$ and the rest of the input an estimate \hat{A} satisfying $(1 - \epsilon)\mathcal{A} < \hat{A} < (1 + \epsilon)\mathcal{A}$. A *randomized* version of such an algorithm is called a *fully polynomial time randomized approximation scheme* (FPRAS). Both of these are natural and standard notions of algorithmic approximation. See Sect. 2.1 for details.

lattice of degree $d + 1$ is typically substantially less than d ; and it is bounded even in the case of sparse random graphs such as those drawn from $\mathcal{G}(n, d/n)$, which have unbounded maximum degree. By analogy with the bounded degree case, one might hope to get correlation decay on graphs with connective constant at most Δ for all $\lambda < \lambda_c(\Delta)$. In [44], such a result was proven under the stronger condition $\lambda < \lambda_c(\Delta + 1)$. The latter bound is tight asymptotically as $\Delta \rightarrow \infty$ (because $\lambda_c(\Delta + 1)/\lambda_c(\Delta) \rightarrow 1$ as $\Delta \rightarrow \infty$), but is sub-optimal in the important case of small Δ .

1.2 Contributions

In this paper, we show that one can indeed replace the maximum degree by the connective constant in the results of both Weitz [50] and Bayati et al. [7]. In particular, we show that for both the hard core and the monomer-dimer models, decay of correlations on the d -ary tree determines the rate of decay of correlations—as well as the complexity of deterministically approximating the partition function—in all graphs of *connective constant* at most d , without any dependence on the maximum degree. The specific notion of decay of correlations that we establish is known in the literature as *strong spatial mixing* [18, 33, 34, 50], and stipulates that the correlation between the state of a vertex v and another set S of vertices at distance ℓ from v should decay exponentially in ℓ even when one is allowed to fix the state of vertices close to v to arbitrary values (see Sect. 2.2.3 for a precise definition). Prior to the role it played in the design of deterministic approximate counting algorithms in Weitz's work [50], strong spatial mixing was already a widely studied notion in computer science and mathematical physics for its utility in analyzing the mixing time of Markov chains [18, 33, 34], and hence an improved understanding of conditions under which it holds is of interest in its own right.

We now give an informal description of the connective constant [19, 32]; see Sect. 2.6 for precise definitions. Given a graph G and a vertex v in G , let $N(v, \ell)$ denote the number of self avoiding walks in G of length ℓ starting at v . A graph family \mathcal{F} is said to have connective constant Δ if for all graphs in \mathcal{F} , the number of self-avoiding walks of length at most ℓ for large ℓ grows as Δ^ℓ , i.e., if $\ell^{-1} \log \sum_{i=1}^{\ell} N(v, i) \sim \log \Delta$ (the definition can be applied to both finite and infinite graphs; see Sect. 2.6). Note that in the special case graphs of maximum degree $d + 1$, the connective constant is at most d . It can, however, be much lower than this crude bound: for any $\epsilon > 0$, the connective constant of graphs drawn from $G(n, d/n)$ is at most $d(1 + \epsilon)$ with high probability (w.h.p.) (see, e.g., [44]), even though their maximum degree is $\Omega\left(\frac{\log n}{\log \log n}\right)$ w.h.p.

Our first main result can now be stated as follows.

Theorem 1 (Main, hard core model) *Let \mathcal{G} be a family of finite graphs of connective constant at most Δ , and let λ be such that $\lambda < \lambda_c(\Delta)$. Then there is an FPTAS for the partition function of the hard core model with vertex activity λ 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} .*

Remark 1 In [44], the above result was proved under the stronger hypothesis $\lambda < \lambda_c(\Delta + 1)$. The above result therefore subsumes the main results of [44]. It is also

optimal in the following sense: there cannot be an FPRAS for graphs of connective constant at most Δ which works for $\lambda > \lambda_c(\Delta)$, unless $\text{NP} = \text{RP}$. This follows immediately from the hardness results for the partition function of the hard core model on bounded degree graphs [45,46] since graphs of degree at most $d + 1$ have connective constant at most d .

An immediate corollary of Theorem 1 is the following.

Corollary 1 *Let $\lambda < \lambda_c(d)$. Then, there is an algorithm for approximating the partition function of graphs drawn from $\mathcal{G}(n, d/n)$ up to a factor of $(1 \pm \epsilon)$ which, with high probability over the random choice of the graph, runs in time polynomial in n and $1/\epsilon$.*

Similar results for $\mathcal{G}(n, d/n)$ have appeared in the literature in the context of rapid mixing of Glauber dynamics for the ferromagnetic Ising model [38], and also for the hard core model [11,37]. Although the authors of [37] do not supply an explicit range of λ for which their rapid mixing results hold, an examination of their proofs suggests that necessarily $\lambda < O(1/d^2)$. Similarly, the results of [11] hold when $\lambda < 1/(2d)$. In contrast, our bound approaches (and is always better than) the conjectured optimal value e/d . Further, unlike ours, the results of [11,37,38] are restricted to $\mathcal{G}(n, d/n)$ and certain other classes of sparse graphs.

A second consequence of Theorem 1 is a further improvement upon the spatial mixing bounds obtained in [44] for various lattices, as shown in Table 1. For each lattice, the table shows the best known upper bound for the connective constant and the strong spatial mixing (SSM) bounds we obtain using these values in Theorem 1. In the table, a value α in the “ λ ” column means that SSM is shown to hold for the appropriate lattice whenever $\lambda \leq \alpha$. As expected, improvements over our previous results in [44] 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. [42] and Vera et al. [47] using numerically intensive methods tailored to this special case. However, as we noted in [44], any improvement in the bound on the connective constant would immediately yield an improvement in our SSM bound. Indeed, in Appendix A, we 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 the papers [42,47]. We note that this improvement would not be possible using only our earlier results in [44].

We also apply our techniques to the study of uniqueness of the Gibbs measure of the hard core model on general trees. Uniqueness is a weaker notion than spatial mixing, requiring correlations to decay to zero with distance but not necessarily at an exponential rate (see Sect. 6 for a formal definition). We relate the phenomenon of the uniqueness of the Gibbs measure of the hard core model on a general

Table 1 Strong spatial mixing bounds for various lattices

Lattice	Max. Degree	Previous SSM bound λ	Connective constant Δ	SSM bound in [44] λ	Our SSM bound λ
\mathbb{T}	6	0.762 [50]	4.251419	0.937 [2]	0.961
\mathbb{H}	3	4.0 [50]	1.847760	4.706 [9]	4.976
\mathbb{Z}^2	4	2.48 [42,47]	2.679193	2.007 [41]	2.082 (2.538*)
\mathbb{Z}^3	6	0.762 [50]	4.7387	0.816 [41]	0.822
\mathbb{Z}^4	8	0.490 [50]	6.8040	0.506 [41]	0.508
\mathbb{Z}^5	10	0.360 [50]	8.8602	0.367 [41]	0.367
\mathbb{Z}^6	12	0.285 [50]	10.8886	0.288 [49]	0.288

* See Appendix A for a description of how this improved bound is obtained
 \mathbb{Z}^D is the D -dimensional Cartesian lattice; \mathbb{T} and \mathbb{H} denote the triangular and honeycomb lattices respectively

tree to the *branching factor* of the tree, another natural notion of average arity that has appeared in the study of uniqueness of Gibbs measure for models such as the ferromagnetic Ising model [30]. The details of these results can be found in Sect. 6.

Our second main result concerns the monomer-dimer model.

Theorem 2 (Main, Monomer-dimer model) *Let \mathcal{G} be a family of finite graphs of connective constant at most Δ , 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 γ for all graphs in \mathcal{G} . More specifically, the running time of the FPTAS for producing an $(1 \pm \epsilon)$ factor approximation is $(n/\epsilon)^{O(\sqrt{\gamma\Delta} \log \Delta)}$.*

The previous best deterministic approximation algorithm for the partition function of the monomer-dimer model was due to Bayati et al. [7], and ran in time $(n/\epsilon)^{O(\sqrt{\gamma d} \log d)}$ for graphs of degree at most $d + 1$. Thus, our 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 γ) in contrast to the sub-exponential time algorithm obtained by Bayati et al. [7]. Using an observation of Kahn and Kim [25], Bayati et al. also pointed out that the \sqrt{d} factor in the exponent of their running time was 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 (see the remark at the end of Sect. 5 for a more detailed discussion of this point). As an aside, we also note that when no bounds on the connective constant are available our FPTAS degrades to a sub-exponential algorithm, as does the algorithm of Bayati et al. in the case of unbounded degree graphs.

1.3 Techniques

The analyses by Weitz [50] and Bayati et al. [7] both begin with the standard observation that obtaining an FPTAS for the marginal probabilities of the Gibbs distribution is sufficient in order to obtain an FPTAS for the partition function. The next non-trivial step is to show that this computation of marginal probabilities at a given vertex v in a graph G can be carried out on the tree $T_{SAW}(v, G)$ of *self-avoiding walks* in G starting at v . Transferring the problem to a tree allows one to write down a recurrence for the marginal probabilities of a node in the tree in terms of the marginal probabilities of its children. However, since the tree is of exponential size, one needs to truncate the tree at a small (logarithmic) depth in order to obtain a polynomial time algorithm. Such a truncation in turn introduces an “error” at the leaves. The challenge then is to show that this error contracts exponentially as the recurrence works its way up to the root.

The approach of [7,50] (and similar results in [27,28,43]) for establishing this last condition takes the following general form: one shows that at each step of the recurrence, the correlation decay condition implies that the error at the parent node is less than a constant factor (less than 1) times the maximum (ℓ_∞ norm) of the errors at the children of the node. Intuitively, this strategy loses information about the structure of the tree by explaining the error at the parent in terms of *only one* of its children, and hence it is not surprising that the results obtained from it are only in terms of a local parameter such as the maximum degree.

The main technical contribution of [44] was to show that, by analyzing the decay in terms of the ℓ_2 norm—rather than the ℓ_∞ norm—of the errors at the children, one can get past this limitation and obtain a result in terms of the connective constant. Nevertheless, as stated above, the results obtained in [44] did not hold over the best possible range of parameters. Our main innovation in the present paper is to analyze instead a norm *adapted to* the parameters of the model, rather than a fixed norm such as ℓ_2 or ℓ_∞ . Specifically, we show that optimal results can be obtained by analyzing the decay in terms of a carefully picked ℓ_q norm where q is chosen as a function of the connective constant and the model parameters (the fugacity λ in the hard core model and the edge activity γ in the monomer-dimer model). At a technical level, the use of these adaptive norms implies that we can no longer employ the relatively simpler convexity arguments used in [44] in order to bound the propagation of errors; characterizing the “worst case” error vectors now requires solving a more involved optimization problem, which is the main new technical challenge in this paper. In Sect. 3, we give a general framework for tackling this problem. Our model specific main results are then obtained as direct applications of this framework. We conjecture that our framework may find applications to other approximate counting problems as well.

1.4 Related work

MCMC based algorithms for approximating the partition function of the hard core model on graphs of bounded degree $d + 1$ were obtained under the condition $\lambda < 1/(d - 2)$ by Luby and Vigoda [29], and later under the weaker condition $\lambda < 2/(d - 1)$

by Dyer and Greenhill [10] and Vigoda [48]. Weitz [50] obtained an FPTAS under the much weaker condition $\lambda < \lambda_c(d)$ by establishing a tight connection between the algorithmic problem and the decay of correlations on the d -ary tree. This connection was further tightened by Sly [45] (see also Galanis et al. [12] and Sly and Sun [46]) who showed that approximating the partition function of the hard core model on $(d + 1)$ -regular graphs is NP-hard when $\lambda > \lambda_c(d)$. Weitz's algorithm was also one of the first deterministic algorithms for approximating partition functions (along with the contemporaneous work of Bandhopadhyay and Gamarnik [4]) which exploited decay of correlations directly—in contrast to earlier algorithms which were mostly based on MCMC techniques. To date, no MCMC based algorithms for the partition function of the hard core model are known to have as large a range of applicability as Weitz's algorithm.

Weitz's approach has also been used to study the correlation decay phenomenon on specific lattices. Restrepo et al. [42] supplemented Weitz's approach with sophisticated computational arguments tailored to the special case of \mathbb{Z}^2 to obtain strong spatial mixing on \mathbb{Z}^2 under the condition $\lambda < 2.38$. Using even more extensive numerical work, this bound was improved to $\lambda < 2.48$ by Vera et al. [47]. In contrast, a direct application of Weitz's results yields the result only under the condition $\lambda < \lambda_c(3) = 1.6875$. Sampling from the hard core model on special classes of unbounded degree graphs has also been considered in the literature. Mossel and Sly [37] gave an MCMC based algorithm for sampling from the hard core model on graphs drawn from $\mathcal{G}(n, d/n)$. However, their algorithm is only applicable in the range $\lambda < O(1/d^2)$. Efthymiou [11] gave an MCMC based sampler under the much weaker condition $\lambda < 1/(2d)$. Hayes and Vigoda [21] also considered the question of sampling from the hard core model on special classes of unbounded degree graphs. They showed that for regular graphs on n vertices of degree $d(n) = \Omega(\log n)$ and of girth greater than 6, the Glauber dynamics for the hard core model mixes rapidly for $\lambda < e/d(n)$. These results are incomparable to Theorem 1. The latter neither requires the graph to be regular, nor any lower bounds on its degree or girth, but it does require additional information about the graph in the form of its connective constant. However, when the connective constant is available, then irrespective of the maximum degree of the graph or its girth, the theorem affords an FPTAS for the partition function.

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 [30] showed 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 [36,38] proved results analogous to our Theorem 1. An important ingredient in the arguments in both [30] and [36,38] 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 [30], Pemantle and Steif [40] defined the notion of a *robust phase transition (RPT)* and related 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.

A preliminary version of this paper [44] investigated decay of correlations in the hard core model in graphs with possibly unbounded degree but bounded connective constant. There, it was shown that when $\lambda < \lambda_c(\Delta + 1)$, all graphs of connective constant at most Δ exhibit decay of correlations and also admit an FPTAS for the partition function. The observation that for any $\epsilon > 0$, the connective constant of graphs drawn from $\mathcal{G}(n, d/n)$ is at most $d(1 + \epsilon)$ with high probability was then used to derive a polynomial time sampler for the hard core model on $\mathcal{G}(n, d/n)$ all the way up to $\lambda < e/d$ (which was the conjectured asymptotic bound). Even for the case of bounded degree graphs, it was shown in [44] that estimates on the connective constant could be used to improve the range of applicability of Weitz’s results. This latter result was then used in conjunction with well known upper bounds on the connective constants of various regular lattices to improve upon the known strong spatial mixing bounds for those lattices (these included \mathbb{Z}^d for $d \geq 3$). However, in the special case of \mathbb{Z}^2 , the bounds in [44] fell short of those obtained by Restrepo et al. [42] and Vera et al. [47].

The results of the present paper for the hard core model strengthen and unify the two distinct results of [44] mentioned above, by replacing the maximum degree constraint in Weitz’s result completely by an exactly corresponding constraint on the connective constant: in particular, we show that graphs of connective constant Δ admit an FPTAS (and exhibit strong spatial mixing) whenever $\lambda < \lambda_c(\Delta)$. Thus, for example, our results extend the range of applicability of the above quoted results of [44] for $\mathcal{G}(n, d/n)$ to $\lambda < \lambda_c(d)$; since $\lambda_c(d) > e/d$ for all $d \geq 1$, this is a strict improvement on [44]. Regarding the question of improved strong spatial mixing bounds on specific lattices, our results show that connective constant computations are sufficient to improve upon the results of Restrepo et al. [42] and Vera et al. [47] even in the case of \mathbb{Z}^2 , without taking recourse to any further numerical or computational work.

In contrast to the case of the hard core model, where algorithms with the largest range of applicability are already deterministic, much less is known about the deterministic approximation of the partition function of the monomer-dimer model. Jerrum and Sinclair [23] gave an MCMC-based FPRAS for the monomer-dimer model which runs in polynomial time on all graphs (without any bounds on the maximum degree) for any fixed value of the edge activity λ . However, no deterministic algorithms with this range of applicability are known, even in the case of specific graph families such as $\mathcal{G}(n, d/n)$. So far, the best result in this direction is due to Bayati, Gamarnik, Katz, Nair and Tetali [7], who gave an algorithm which produces a $(1 \pm \epsilon)$ factor approximation for the monomer-dimer partition function in time $(n/\epsilon)^{\tilde{O}(\sqrt{\gamma d})}$ on graphs of maximum degree d . Their result therefore yields super-polynomial (though sub-exponential) algorithms in the case of graphs such as those drawn from $\mathcal{G}(n, d/n)$, which have unbounded maximum degree even for constant d . In contrast, the present paper shows that the same running time can in fact be obtained for graphs of *connective constant* d , irrespective of the maximum degree. The running times obtained by applying the

results of this paper to the case of $\mathcal{G}(n, d/n)$ are therefore polynomial (and not merely sub-exponential) when d is a fixed constant.

The connective constant itself and various notions related to it have been the subject of an extensive and active area of research. Starting with the classical papers by Hammersley and Morton [20], Hammersley and Broadbent [8] and Hammersley [19], several natural combinatorial questions concerning the number and other properties of self-avoiding walks in various lattices have been studied in depth; see the monograph of Madras and Slade [32] for a survey. Much work has been devoted especially to finding rigorous upper and lower bounds for the connective constant of various lattices [1, 2, 22, 26, 41], and heuristic techniques from physics have also been brought to bear upon this question. For example, Nienhuis [39] conjectured on the basis of heuristic arguments that the connective constant of the honeycomb lattice \mathbb{H} must be $\sqrt{2 + \sqrt{2}}$. In 2012, Nienhuis's conjecture was rigorously confirmed in a celebrated paper of Duminil-Copin and Smirnov [9].

2 Preliminaries

2.1 Approximation algorithms

Deterministic and randomized fully polynomial approximation schemes are standard notions of algorithmic approximation. While the definitions we give here for them are specialized to the case of quantities which are functions of graphs (since this is the setting that concerns us in this paper), there is no difficulty in generalizing them to other contexts.

Definition 1 (*Fully polynomial approximation schemes*) Let $\mathcal{A}(G)$ be a quantity that depends upon an input graph G . A *fully polynomial time approximation scheme* (FPTAS) for \mathcal{A} is a deterministic algorithm which, on input G and a rational accuracy parameter ϵ , outputs in time polynomial in $1/\epsilon$ and the representation length of G a quantity \hat{A} satisfying $(1 - \epsilon)\mathcal{A}(G) \leq \hat{A} \leq (1 + \epsilon)\mathcal{A}(G)$. A *fully polynomial time randomized approximation scheme* (FPRAS) for \mathcal{A} is a randomized algorithm that on input G , an accuracy parameter ϵ and an error parameter δ , outputs in time polynomial in $1/\epsilon$, $\log(1/\delta)$ and the representation length of G a quantity \hat{A} which, with probability at least $1 - \delta$ over the randomness of the algorithm satisfies $(1 - \epsilon)\mathcal{A}(G) \leq \hat{A} \leq (1 + \epsilon)\mathcal{A}(G)$.

2.2 Probabilities and likelihood ratios

In this section, we define some standard marginals of the monomer-dimer and hard core distributions. The importance of these quantities for our work comes from the standard “self-reducibility” arguments which show that obtaining an FPTAS for these marginals is sufficient in order to obtain an FPTAS for the partition function of these models (see Sect. 2.5 for these reductions).

2.2.1 Monomer-dimer model

Definition 2 (*Monomer probability*) Consider the Gibbs distribution of the monomer-dimer model with dimer activity γ on a finite graph $G = (V, E)$, and let v be a vertex in V . We define the *monomer probability* $p(v, G)$ as

$$p_v := \mathbb{P}[v \notin M],$$

which is the probability that v is unmatched (i.e., a *monomer*) in a matching M sampled from the Gibbs distribution.

Remark 2 The monomer-dimer model is often described in the literature in terms of a *monomer activity* λ instead of the *dimer activity* γ used here. In this formulation, the weight of a matching M is $\lambda^{u(M)}$, where $u(M)$ is the number of unmatched vertices (*monomers*) in M . The two formulations are equivalent: with dimer activity γ corresponding to monomer activity $\lambda = \frac{1}{\gamma^2}$.

2.2.2 Hard core model

In the case of the hard core model, we will need to define the appropriate marginals in a slightly more generalized setting. Given a graph $G = (V, E)$, a *boundary condition* will refer to a partially specified independent set in G ; formally, a boundary condition $\sigma = (S, I)$ is a subset $S \subseteq V$ along with an independent set I on S (boundary conditions may be seen as special instances of initial conditions defined below).

Definition 3 (*Occupation probability and occupation ratio*) Consider the hard core model with vertex activity $\lambda > 0$ on a finite graph G , and let v be a vertex in G . Given a boundary condition $\sigma = (S, I_S)$ on G , the *occupation probability* $p_v(\sigma, G)$ at the vertex v is the probability that v is included in an independent set I sampled according to the hard core distribution conditioned on the event that I restricted to S coincides with I_S . The *occupation ratio* $R_v(\sigma, G)$ is then defined as

$$R_v(\sigma, G) = \frac{p_v(\sigma, G)}{1 - p_v(\sigma, G)}.$$

2.2.3 Strong spatial mixing

We present the definition of strong spatial mixing for the special case of the hard core model; the definition in the case of the monomer dimer model is exactly analogous. Our definition here closely follows the version used by Weitz [50].

Definition 4 (*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 σ and τ on G which differ only at a distance of at least ℓ from v , we have

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

for some fixed constant $0 \leq c < 1$.

An important special condition of the definition is when the family \mathcal{F} consists of a single infinite graph (e.g., \mathbb{Z}^2 or another regular lattice). The constant c in the definition is often referred to as the *rate* of strong spatial mixing.

2.3 Truncated recurrences with initial conditions

As in the case of other correlation decay based algorithms (e.g., in [13,50]), we will need to analyze recurrences for marginals on rooted trees with various initial conditions. We therefore set up some notation for describing such recurrences. 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 5 (Cutset) Let T be any tree rooted at ρ . A *cutset* C is a set of vertices in T satisfying the following two conditions:

- (1) Any path from ρ 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_ℓ of all vertices at distance ℓ from ρ in T .

Remark 3 In Sect. 6, we will need to use cutsets on rooted locally finite infinite trees. In this case, the first condition in the definition changes to “any infinite path starting from the root ρ must pass through C ”.

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 6 (Initial condition) An *initial condition* $\sigma = (S, P)$ is a set S of vertices in T along with an assignment $P : S \rightarrow [0, b]$ 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 $f_d : [0, b]^d \rightarrow [0, b]$ 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 σ . Formally, we define $F_u(\sigma) = b_0$ when $u \notin S$ is a leaf, and

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

2.4 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 that can be obtained by extending the self-avoiding walk represented by u by exactly one step (note that the self-avoiding walk tree for a finite graph G is of finite size, since any self-avoiding walk in G is of length at most the number of vertices in G). The importance of the self-avoiding walk tree for computation stems from the beautiful results of Godsil [16] (for the monomer-dimer model) and Weitz [50] (for the hard core model), 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.

2.4.1 Monomer-dimer model

Theorem 3 (Godsil [16]) *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 ρ , and let ρ_i , $1 \leq i \leq d$ be the children of ρ . Denoting by p_i the monomer probability $p_{\rho_i}(T_{\rho_i})$ at the root of the subtree T_{ρ_i} , one can then show that (see, e.g., [25])

$$p_\rho(T) = f_{d,\gamma}(p_1, p_2, \dots, p_d) := \frac{1}{1 + \gamma \sum_{i=1}^d p_i}. \tag{2}$$

Remark 4 In what follows, we will often suppress the dependence of $f_{d,\gamma}$ on γ 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 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_ℓ comprising all vertices at distance ℓ from ρ . Since $f_{d,\gamma}$ is monotonically decreasing in each of its arguments, we have

$$\begin{aligned} F_\rho(\mathbf{0}_\ell) &\leq p_\rho(T) \leq F_\rho(\mathbf{1}_\ell) && \text{when } \ell \text{ is even, and} \\ F_\rho(\mathbf{0}_\ell) &\geq p_\rho(T) \geq F_\rho(\mathbf{1}_\ell) && \text{when } \ell \text{ is odd.} \end{aligned} \tag{3}$$

Here, the initial condition $\mathbf{0}_\ell$ (respectively, $\mathbf{1}_\ell$) assigns the value 0 (respectively, 1) to every vertex in S_ℓ . 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 ℓ in order to establish that truncated versions of the recurrence converge to the true answer $p_\rho(T)$ exponentially fast in the “truncation length” ℓ .

2.4.2 Hard core model

Weitz [50] 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 4 (Weitz [50]) *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 condition³ \mathcal{W} on $T_{SAW}(v, G)$ such that for any boundary condition σ on G , we have*

$$R_v(\sigma, G) = R_v(\mathcal{W} \cup \sigma, T_{SAW}(v, G)), \tag{4}$$

where (1) the boundary condition σ on the right hand side denotes the natural translation of the boundary condition σ 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 σ (that is, σ overrides \mathcal{W} on vertices on which σ 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”.

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 that (see, e.g., [50]) that

$$R_\rho(\sigma, T) = f_{d,\lambda}(R_1, R_2, \dots, R_d) := \lambda \prod_{i=1}^d \frac{1}{1 + R_i}. \tag{5}$$

We now see that in terms of our notation for tree recurrences, the computation of $R_\rho(\sigma, T)$ corresponds to computing $F_\rho(\lambda_L \cup \sigma)$ (with the boundary value b_0 for leaves set to λ), where the initial condition $\lambda_L \cup \sigma$ assigns the value λ to all vertices in the set L of leaves, and then applies the boundary condition σ (possibly overriding previously assigned values). Note that the boundary condition σ assigns $R_v = \infty$ for vertices v which are set to occupied by σ , 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 the one which assigns $R_u = 0$ to the parent u of v . Thus, we may

³ For a description of how this boundary condition is computed, we refer the reader to the third paragraph of Appendix A.

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_ℓ of vertices at distance ℓ from the root ρ , and use the fact that $f_{d,\lambda}$ is monotonically decreasing in each of its arguments to see that

$$\begin{aligned} F_\rho(\mathbf{0}_\ell \cup \sigma) &\leq R_\rho(\sigma, T) \leq F_\rho(\lambda_\ell \cup \sigma) && \text{when } \ell \text{ is even, and} \\ F_\rho(\mathbf{0}_\ell \cup \sigma) &\geq R_\rho(\sigma, T) \geq F_\rho(\lambda_\ell \cup \sigma) && \text{when } \ell \text{ is odd.} \end{aligned} \tag{6}$$

Here, the initial condition $\mathbf{0}_\ell \cup \sigma$ (respectively, $\lambda_\ell \cup \sigma$) assigns the value 0 (respectively, λ) to every vertex in S_ℓ , after which the boundary condition σ is applied, possibly overriding the earlier assignments (from the previous discussion, we can assume that the effect of σ is limited to 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 ℓ 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” ℓ .

2.5 From probabilities to the partition function

In this section, we review some standard facts on how approximation algorithms for the marginal probabilities translate into approximation algorithms for the partition function (see, e.g, [13,50]). We provide the calculations here for the case of the monomer-dimer model, and refer to Weitz [50] for similar calculations for the hard core model.

Let v_1, v_2, \dots, 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

$$\begin{aligned} Z(G) &= \prod_{i=1}^n \frac{Z(G - \{v_1, \dots, v_{i-1}\})}{Z(G - \{v_1, \dots, v_i\})} \\ &= \prod_{i=1}^n \frac{1}{p_{v_i}(G - \{v_1, \dots, v_{i-1}\})}. \end{aligned} \tag{7}$$

Suppose, we have an FPTAS for the probabilities p_ρ 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 $t(n, 2n/\epsilon)$ to compute an approximation \hat{p}_i to the $p_{v_i}(G - \{v_1, \dots, v_{i-1}\})$. We then have for each i

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

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 $O(n \cdot t(n, 2n/\epsilon))$, 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.

2.6 The connective constant

We now recall the definition of the connective constant of a graph. Given a vertex v in a locally finite graph, we denote by $N(v, l)$ the number of self-avoiding walks of length l in the graph which start at v . The connective constant for infinite graphs is then defined as follows.

Definition 7 (*Connective constant: infinite graphs* [32]) Let $G = (V, E)$ be a locally finite infinite graph. The *connective constant* $\Delta(G)$ of G is $\sup_{v \in V} \limsup_{\ell \rightarrow \infty} N(v, \ell)^{1/\ell}$.

Remark 5 The supremum over v in the definition is clearly not required for vertex-transitive graphs such as Cartesian lattices. Further, in such graphs the lim sup can be replaced by a limit [32].

The definition was extended in [44] to families of finite graphs parametrized by size. As observed there, such a parametrization is natural for algorithmic applications.

Definition 8 (*Connective constant: finite graphs* [44]) Let \mathcal{F} be a family of finite graphs. The connective constant of \mathcal{F} is at most Δ 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 $\ell \geq a \log |V|$.

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. 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 (see [44]) that for any fixed $\epsilon > 0$, the connective constant of such a graph is at most $d(1 + \epsilon)$ w.h.p.

Remark 6 Note that the connective constant has a natural interpretation as the “average arity” of the SAW tree, since vertices in $T_{SAW}(v, G)$ at distance ℓ from the root are in bijection with self-avoiding walks of length ℓ starting at v .

3 Decay of correlations on the SAW tree

In this section, we lay the groundwork for proving decay of correlations results for the tree recurrences F_{ρ} defined in Eq. (1) for both the hard core and monomer-dimer models: such a result basically affirms that truncating the recurrence at a small depth

ℓ is sufficient in order to approximate F_ρ with good accuracy. Our proof will use the *message approach* [28,42,43], which proceeds by defining an appropriate function ϕ of the marginals being computed and then showing a decay of correlation result for this function.

Definition 9 (*Message* [28,42,43]) Given a positive real number b , a *message* is a strictly increasing and continuously differentiable function $\phi : (0, b] \rightarrow \mathbb{R}$, with the property that the derivative of ϕ is bounded away from 0 on its domain. A message ϕ is guaranteed to admit a continuously differentiable inverse, which we will denote by ψ .

In the rest of this section, we will work in the abstract framework described in Sect. 2.3, to illustrate how the message approach can be used to get strengthened decay of correlation estimates as compared to those obtained from direct analyses of one step of the recurrence. We will then instantiate our framework with appropriately chosen messages for the monomer-dimer and the hard core models in Sects. 4 and 5.

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 $f_d : [0, b]^d \rightarrow [0, b]$ 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 the hard core and the monomer-dimer model).

Condition 5 (*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 $\mathbf{x} \in \mathbb{R}^d$ is a permutation of $\mathbf{y} \in \mathbb{R}^d$, then $f_d(\mathbf{x}) = f_d(\mathbf{y})$.
- (2) If all but the first k co-ordinates of $\mathbf{x} \in \mathbb{R}^d$ are 0, then $f_d(\mathbf{x}) = f_k(x_1, x_2, x_3, \dots, x_k)$.

Given the message ϕ (and its inverse ψ), we further define f_d^ϕ by

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

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

Lemma 1 (*Mean value theorem*) Consider two vectors \mathbf{x} and \mathbf{y} in $\phi([0, B])^d$. Then there exists a vector $\mathbf{z} \in [0, \infty)^d$ such that

$$\left| f_d^\phi(\mathbf{x}) - f_d^\phi(\mathbf{y}) \right| \leq \Phi(f_d(\mathbf{z})) \sum_{i=1}^d \frac{|y_i - x_i|}{\Phi(z_i)} \left| \frac{\partial f_d}{\partial z_i} \right|, \tag{8}$$

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

The first step of our approach is similar to that taken in the papers [27,28,42,43] in that we will use an appropriate message—along with the estimate in Lemma 1—to argue that the “distance” between two input message vectors \mathbf{x} and \mathbf{y} at the children of a vertex shrinks by a constant factor at each step of the recurrence. Previous works [27, 28,42,43] showed such a decay on some version of the ℓ_∞ norm of the “error” vector $\mathbf{x} - \mathbf{y}$: this was achieved by bounding the appropriate dual ℓ_1 norm of the gradient of the recurrence. 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 ℓ_q norm for some $q < \infty$.

In line with the above plan, we will attempt to bound the right hand side of eq. (8) in terms of $\|\mathbf{x} - \mathbf{y}\|_q$ for an appropriate value of $q < \infty$ by maximizing the sum while keeping $f_d(\mathbf{z})$ fixed. In the special case $q = 2$, it is in fact possible to carry out this maximization using relatively simple concavity arguments. This was the approach taken in [44], but the restriction $q = 2$ leads to sub-optimal results. Here we get past this limitation by using a more flexible optimization than that used in [44]. 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 10 Given a consistent family of recurrences $\{f_d\}_{d \geq 1}$, a message ϕ (with $\Phi := \phi'$) is said to be *symmetrizable with exponent a* with respect to the family if it satisfies the following two conditions:

- (1) $\lim_{x_i \rightarrow 0^+} \frac{1}{\Phi(x_i)} \left| \frac{\partial f_d}{\partial x_i} \right| = 0$ for all $d \geq 1$, and for any fixed values of the $x_j, j \neq i$.
- (2) Let \mathcal{D} be the domain of the recurrence family, so that as discussed above, it is of the form $[0, b]$ for a positive real b . For every positive integer d and every real $B > 0$ for which the program

$$\begin{aligned} \max \quad & \sum_{i=1}^d \left(\frac{1}{\Phi(x_i)} \left| \frac{\partial f_d}{\partial x_i} \right| \right)^a, \quad \text{where} \\ & f_d(\mathbf{x}) = B \\ & x_i \in \mathcal{D}, \quad 1 \leq i \leq d \end{aligned}$$

is feasible, it also has a solution \mathbf{x} in which all the non-zero entries of \mathbf{x} are equal. Here, we use condition 1 to continuously extend the definition of the d functions $\frac{1}{\Phi(x_i)} \left| \frac{\partial f_d}{\partial x_i} \right|$ so that that they evaluate to 0 when $x_i = 0$.

For symmetrizable messages, we will be able to bound the quantity $|f_d^\phi(\mathbf{x}) - f_d^\phi(\mathbf{y})|$ in terms of $\|\mathbf{x} - \mathbf{y}\|_q$, 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 conditions used in [44]. Our bounds will be stated in terms of the following notion of decay.

Notation. 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 11 (*Decay factor α*) Let ϕ be a message with derivative Φ , 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:

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

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

The *decay factor α* is then defined as

$$\alpha = \alpha_{\phi,q} := \sup_{d \geq 1} \xi_{\phi,q}(d). \tag{9}$$

Armed with the above definitions, we are now ready to prove Lemma 2, 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 ϕ symmetrizable with exponent a for which the decay factor α is small; Lemma 3 below then shows how the decay factor comes into play in proving exponential decay of correlations over the tree.

Lemma 2 *Let ϕ be a message with derivative Φ , and let a and q be positive reals such that $\frac{1}{a} + \frac{1}{q} = 1$. If ϕ is symmetrizable with exponent a , then for any two vectors \mathbf{x}, \mathbf{y} in $\phi([0, b])^d$, there exists an integer $k \leq d$ such that*

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

Proof We apply Lemma 1. Assuming \mathbf{z} is as defined in that lemma, we have by Hölder’s inequality

$$\begin{aligned} \left| f_d^\phi(\mathbf{x}) - f_d^\phi(\mathbf{y}) \right| &\leq \Phi(f_d(\mathbf{z})) \sum_{i=1}^d \frac{|y_i - x_i|}{\Phi(z_i)} \left| \frac{\partial f_d}{\partial z_i} \right| \\ &\leq \Phi(f_d(\mathbf{z})) \left(\sum_{i=1}^d \left(\frac{1}{\Phi(z_i)} \left| \frac{\partial f_d}{\partial z_i} \right| \right)^a \right)^{1/a} \|\mathbf{x} - \mathbf{y}\|_q. \end{aligned}$$

Since ϕ is symmetrizable with exponent a , we can replace \mathbf{z} in the above inequality with a vector $\tilde{\mathbf{z}}$ 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 $\tilde{\mathbf{z}}$. Using the consistency condition, we then get

$$\begin{aligned} \left| f_d^\phi(\mathbf{x}) - f_d^\phi(\mathbf{y}) \right| &\leq \Phi(f_k(\tilde{\mathbf{z}})) \left(\sum_{i=1}^k \left(\frac{1}{k\Phi(\tilde{z})} |f'_k(\tilde{z})| \right)^a \right)^{1/a} \|\mathbf{x} - \mathbf{y}\|_q \\ &= \frac{1}{k^{1-1/a}} \frac{\Phi(f_k(\tilde{\mathbf{z}})) |f'_k(\tilde{z})|}{\Phi(\tilde{z})} \|\mathbf{x} - \mathbf{y}\|_q. \end{aligned}$$

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

Given a message ϕ satisfying the conditions of Lemma 2, 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 ρ under an initial condition σ . The lemma quantifies the dependence of $F_\rho(\sigma)$ on initial conditions σ which are fixed everywhere except at some cutset C , in terms of the distance of C from ρ .

Lemma 3 *Let T be a finite tree rooted at ρ . 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 σ and τ on $T_{\leq C'}$ which differ only on C' (so that they agree, in particular, on the default values they assign to the leaves of $T_{\leq C'}$ that are not in 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 ϕ 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 α is as defined in Eq. (9), 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 Appendix B.

4 A message for the hard core model

We now instantiate the approach outlined in Sect. 3 to prove Theorem 1 for the hard core model. Our message is the same as that used in [28]; we choose

$$\phi(x) := \sinh^{-1}(\sqrt{x}), \text{ so that } \Phi(x) := \phi'(x) = \frac{1}{2\sqrt{x(1+x)}}. \tag{10}$$

Notice that ϕ is a strictly increasing, continuously differentiable function on $(0, \infty)$, and also satisfies the technical condition that the derivative Φ be bounded away from zero on any finite interval, as required in the definition of a message. Our improvements over the results in [44] depend upon proving the following fact about the message ϕ .

Lemma 4 *For any $a \geq 2$, the message ϕ as defined in Eq. (10) 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 Appendix C.2. The crucial advantage of Lemma 4 is the flexibility it allows in the choice of exponent a .

The decay factor α in Lemma 3 which governs the rate of decay in the tree recurrences depends upon the choice of the exponent, and as we show later in this section, it is possible to obtain an optimal decay rate by choosing an appropriate exponent satisfying Lemma 4.

We now outline how the flexibility in the choice of the exponent afforded by Lemma 4 helps us obtain an optimal decay rate. Given the vertex activity λ , we first define the quantity $\Delta_c := \Delta_c(\lambda)$ as the unique solution in t of $\lambda_c(t) = \lambda$ (the existence and uniqueness of Δ_c follow from the easily verified fact that λ_c is a strictly decreasing function and maps the interval $(1, \infty)$ onto $(0, \infty)$). To get sufficient stepwise decay when Lemma 3 is applied, we need to show that for all $d > 0$, $\xi_{\phi,q}(d) \leq \frac{1}{\Delta_c}$. On the other hand, irrespective of the choice of q (and hence a), it is easy to show that $\xi_{\phi,q}(\Delta_c) = \frac{1}{\Delta_c}$. Thus, in order to fulfill our requirements for the decay rate, we need to choose a value of q such that $\xi_{\phi,q}(d)$ is maximized at $d = \Delta_c$. Now, it turns out that although symmetrizability is easier to establish when $q \geq 2$ (i.e., when $a \leq 2$), $\xi_{\phi,q}(d)$ is a strictly increasing function of d for such q . Hence, in order to produce a maximum at $d = \Delta_c$, we need to choose $q \leq 2$ (equivalently, $a \geq 2$). In fact, as seen in [44], even $q = a = 2$ is not sufficient for an optimal result, and we actually need $q < 2$ to fulfill the requirements on the decay rate. The hypotheses of Lemma 4 then require symmetrizability to be established for this more challenging range of values of a and q . In contrast, the simpler concavity arguments used in [44] restricted the choice of the exponent to $a = 2$, and hence led to a suboptimal decay rate.

Our analysis of the function $\xi_{\phi,q}(d)$ now begins with the following simple lemma, the special case $q = 2$ of which was proven in [44]. The lemma merely shows how to perform one of the maximizations needed in the definition of the decay factor α . In what follows, we drop the subscript ϕ for simplicity of notation.

Lemma 5 *Consider the hard core model with any fixed vertex activity $\lambda > 0$. For any $q \geq 1$ and with ϕ as defined in Eq. (10), we have $\xi_q(d) = \mathcal{E}_q(d, \tilde{x}_\lambda(d))$, where $\tilde{x}_\lambda(d)$ is the unique solution to*

$$d\tilde{x}_\lambda(d) = 1 + f_{d,\lambda}(\tilde{x}_\lambda(d)). \tag{11}$$

Proof Plugging in Φ from Eq. (10) in the definition of \mathcal{E} , we get

$$\mathcal{E}_q(d, x) = d^{q-1} \left(\frac{x}{1+x} \frac{f_{d,\lambda}(x)}{1+f_{d,\lambda}(x)} \right)^{q/2}.$$

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

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

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 \mathcal{E}_q . The fact that such a zero exists follows by noting that the partial derivative is positive at $x = 0$ and

negative as $x \rightarrow \infty$. Thus, $\mathcal{E}_q(d, x)$ is maximized at $\tilde{x}_\lambda(d)$ as defined above, and hence $\xi_q(d) = \mathcal{E}_q(d, \tilde{x}_\lambda(d))$, as claimed.

We now choose a and q as follows:

$$\frac{1}{q} = 1 - \frac{\Delta_c - 1}{2} \log \left(1 + \frac{1}{\Delta_c - 1} \right); \quad \frac{1}{a} = 1 - \frac{1}{q}. \tag{12}$$

Note that since $\log(1 + y) \leq y$ for all $y \geq 0$, we get that $q \leq 2$ (and hence $a \geq 2$) by using $y = \frac{1}{\Delta_c - 1}$, and noting that $\Delta_c > 1$. As noted above, $\xi_{\phi,q}(\Delta_c) = 1/\Delta_c$ holds for all q . The above value of q is derived by maximizing $\xi_{\phi,q}(d)$ over d for an arbitrary q , and then stipulating that this maximum occurs at $d = \Delta_c$. The details of this computation justifying this choice of q appear in Appendix C.1. Here, we record this property of our choice of q in the following lemma. To simplify notation and to emphasize dependence of q upon λ , we define the function $v_\lambda(d)$ as follows:

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

Lemma 6 *Fix $\lambda > 0$ and let $\Delta_c(\lambda) > 1$ be the unique solution to $\lambda_c(t) = \lambda$. The function $v_\lambda : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ is maximized at $d = \Delta_c := \Delta_c(\lambda)$. Further,*

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

The proof of the above lemma is somewhat technical, and is deferred to Appendix C.1. The lemma shows that when $\lambda < \lambda_c(\Delta)$, the decay factor $\alpha < \frac{1}{\Delta}$. We now proceed with the proof of Theorem 1. This only requires some standard arguments, with the only new ingredient being the improved estimate on decay factor proved in Lemma 6.

Proof of Theorem 1 Let \mathcal{G} be any family of finite or infinite graphs with connective constant Δ . We prove the result for any fixed λ such that $\lambda < \lambda_c(\Delta)$. For such λ , we have $\Delta_c(\lambda) > \Delta$ (since λ_c is a decreasing function). Using Lemma 6 we then see that there is an $\epsilon > 0$ such that $v_\lambda(d)\Delta \leq 1 - \epsilon$ 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{G} , v any vertex in G , and consider any boundary conditions σ and τ on G which differ only at a distance of at least ℓ from v . We consider the Weitz self-avoiding walk tree $T_{SAW}(v, G)$ rooted at v (as defined in Sect. 2.6). As before, we denote again by σ (respectively, τ) the translation of the boundary condition σ (respectively, τ) 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_ℓ denote the cutset in $T_{SAW}(v, G)$ consisting of all vertices at distance ℓ from v . Since G has connective constant at most Δ , it follows that for ℓ large enough, we have $|C_\ell| \leq \Delta^\ell(1 - \epsilon/2)^{-\ell}$. Further, in the notation of Lemma 3, $v_\lambda(d)\Delta = 1 - \epsilon$ implies that the decay factor α [defined in Eq.

(9)] is at most $(1 - \epsilon)/\Delta$. We now apply Lemma 3. We first observe that given our message ϕ , 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

$$\begin{aligned} |R_v(\sigma, G) - R_v(\tau, G)|^q &= |R_v(\mathcal{W} \cup \sigma, T_{SAW}(v, G)) - R_v(\mathcal{W} \cup \tau, T_{SAW}(v, G))|^q \\ &\leq c_0 \sum_{u \in C_\ell} \left(\frac{1 - \epsilon}{\Delta}\right)^\ell \\ &\leq c_0 \left(\frac{1 - \epsilon}{1 - \epsilon/2}\right)^\ell, \quad \text{using } |C_\ell| \leq \Delta^\ell (1 - \epsilon/2)^{-\ell}, \end{aligned}$$

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

We now consider the case when \mathcal{G} is a family of finite graphs, and G is a graph from \mathcal{G} of n vertices. Since the connective constant of the family is Δ , 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_ℓ is again chosen to be the set of all vertices at distance ℓ from v in $T_{SAW}(v, G)$, so that $|C_\ell| \leq c\Delta^\ell$. As before, we then have for $\ell > a \log n$,

$$\begin{aligned} |R_v(\sigma, G) - R_v(\tau, G)|^q &= |R_v(\mathcal{W} \cup \sigma, T_{SAW}(v, G)) - R_v(\mathcal{W} \cup \tau, T_{SAW}(v, G))|^q \\ &\leq c_0 \sum_{u \in C_\ell} \left(\frac{1 - \epsilon}{\Delta}\right)^\ell \\ &\leq c \cdot c_0 (1 - \epsilon)^\ell, \quad \text{using } |C_\ell| \leq c\Delta^\ell, \end{aligned} \tag{13}$$

which establishes the requisite strong spatial mixing bound.

In order to prove the algorithmic part, we first recall an observation of Weitz [50] 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 μ 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 μ 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 μ 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 σ . In this

case, we simply return 0. Otherwise, we expand $T_{SAW}(v, G)$ up to depth ℓ 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 Δ .

We now consider two extreme boundary conditions σ_+ and σ_- on C_ℓ : in σ_+ (respectively, σ_-) all vertices in C_ℓ that are not already fixed by σ 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. (13) 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 $O(\Delta^\ell)$, which is polynomial in n and $1/\mu$ as required.

We now prove Corollary 1.

Proof of Corollary 1 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 ℓ starting at v . Suppose $\ell \geq a \log n$, where a is a constant depending upon the parameters ϵ, β 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 ℓ with $a \log n \leq \ell \leq n$ and over all vertices v , we see that the connective constant Δ 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 1 are satisfied. This completes the proof.

5 A message for the monomer-dimer model

In this section, we apply the general framework of Sect. 3 to the monomer-dimer model. 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:

$$\phi(x) := \frac{1}{2} \log \left(\frac{x}{2-x} \right), \quad \text{so that } \Phi(x) := \phi'(x) = \frac{1}{x(2-x)}. \tag{14}$$

Note that ϕ is strictly increasing and continuously differentiable on the interval $(0, 1]$, and its derivative is bounded away from 0 on that interval. Thus, ϕ 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 3, we first study the symmetrizability of ϕ in the following technical lemma.

Lemma 7 Fix $r \in (1, 2]$. The message ϕ as defined in Eq. (14) 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 Appendix D.1. 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 ξ used in the definition of the decay factor. For ease of notation, we drop the subscript ϕ from the notation for ξ .

Lemma 8 Consider the monomer-dimer model with edge activity γ , and let ϕ be the message chosen in (14). For any $q > 1$, we have $\xi_q(d) = \mathcal{E}_q(d, \tilde{p}_\gamma(d))$, where $\tilde{p}_\gamma(d)$ satisfies $\mathcal{E}_q^{(0,1)}(d, \tilde{p}_\gamma(d)) = 0$ and is given by

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

Proof Plugging in Φ from eq. (14) in the definition of \mathcal{E} , we get

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

$$f_{d,\gamma}(x) = \frac{1}{1+\gamma dx}.$$

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

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

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 \mathcal{E}_q . By solving the quadratic, we see that $\tilde{p}_\gamma(d)$ as defined above is such a solution. Thus, $\mathcal{E}_q(d, x)$ is maximized at $\tilde{p}_\gamma(d)$ as defined above, and hence $\xi_q(d) = \mathcal{E}_q(d, \tilde{p}_\gamma(d))$.

Given the edge activity γ and an upper bound Δ 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

$$\frac{1}{r} = 1 - \frac{1}{\sqrt{1 + 4\gamma D}}; \quad \frac{1}{q} = 1 - \frac{1}{r} = \frac{1}{\sqrt{1 + 4\gamma D}}. \tag{15}$$

Note that the choice of D implies that $1 < r \leq 2$, so that ϕ 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

$$v_\gamma(d) := \xi_q(d), \quad \text{where } q \text{ is as chosen in Eq. (15).}$$

Lemma 9 Fix $\gamma > 0$ and $D > 3/4\gamma$, and let q be as chosen in (15). Then the function $v_\gamma : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ is maximized at $d = D$. Further, the decay factor α is given by

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

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

$$\begin{aligned} v'_\gamma(d) &= \mathcal{E}^{(1,0)}(d, \tilde{p}) + \mathcal{E}^{(0,1)}(d, \tilde{p}) \frac{d\tilde{p}}{dd} \\ &= \mathcal{E}^{(1,0)}(d, \tilde{p}), \text{ since } \mathcal{E}^{(0,1)}(d, \tilde{p}) = 0 \text{ by definition of } \tilde{p} \\ &= \frac{\mathcal{E}(d, \tilde{p})}{d(1 + 2\gamma d \tilde{p})} [q - 1 - 2\gamma d \tilde{p}] = \frac{\mathcal{E}(d, \tilde{p})}{d(1 + 2\gamma d \tilde{p})} \left[\sqrt{1 + 4\gamma D} - \sqrt{1 + 4\gamma d} \right], \end{aligned} \tag{16}$$

where we in the last line we substitute the values $\tilde{p}_\gamma(d) = (\sqrt{1 + 4\gamma d} - 1)/(2\gamma d)$ from Lemma 8 and $q = \sqrt{1 + 4\gamma D}$ from eq. (15). Now, we note that in eq. (16), 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 $v'_\gamma(d)$ has a unique zero at $d = D$ for $d \geq 0$, and this zero is a global maximum of v_γ .

We are now ready to prove our main result for the monomer-dimer model, Theorem 2 from the introduction. Given Lemmas 3 and 9, only some standard arguments are needed to finish the proof.

Proof of Theorem 2 Let \mathcal{F} be any family of finite graphs with connective constant at most Δ . Given the vertex activity γ of the monomer-dimer model, we choose $D = \max(\Delta, 3/(4\gamma))$. Using Lemma 9, we then see that the decay factor α appearing

in Lemma 3 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 Sect. 2.2, it is sufficient to construct an FPTAS for $p_v(G)$ in order to derive an FPTAS for the partition function.

Consider the self-avoiding walk tree $T_{SAW}(v, G)$ rooted at v (as defined in Sect. 3). From Godsil’s theorem (Theorem 3), we know that $p_v(G) = p_v(T_{SAW}(v, G))$. Let C_ℓ denote the cutset in $T_{SAW}(v, G)$ consisting of all vertices at distance ℓ from v . Since \mathcal{F} has connective constant at most Δ , 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 3 with q as defined in eq. (15). 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 .⁴ Now, defining $c_0 := (M/L)^q$, we have

$$\begin{aligned} & |F_v(\mathbf{0}_\ell) - F_v(\mathbf{1}_\ell)|^q \\ & \leq c_0 \sum_{u \in C_\ell} \alpha^\ell \leq c \cdot c_0 \cdot (\alpha\Delta)^\ell, \text{ using } |C_\ell| \leq c\Delta^\ell, \\ & \leq c \cdot c_0 \cdot \left(1 - \frac{2}{1 + \sqrt{1 + 4\gamma D}} \right)^{q\ell}, \text{ using } D \geq \Delta \text{ after substituting for } \alpha. \end{aligned} \tag{17}$$

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

$$|F_v(\mathbf{0}_\ell) - F_v(\mathbf{1}_\ell)| \leq \frac{1}{2} c^{1/\sqrt{1+4\gamma D}} \cdot \log(1 + 2\gamma n) \cdot \left(1 - \frac{2}{1 + \sqrt{1 + 4\gamma D}} \right)^\ell. \tag{18}$$

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 Sect. 2.3, $p_v(G)$ 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. (18) is at most $\epsilon/(1 + \gamma n)$. Denoting by β the quantity in the parenthesis on the right hand side of

⁴ Since the degree of every vertex v in the graph is n , every boundary condition σ satisfies $1 \geq F_v(\sigma) \geq \frac{1}{1+\gamma n}$. Substituting these bounds in the definition of M in Lemma 3 yields the claimed bound.

eq. (18), 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 ℓ , 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 ℓ . 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 Sect. 2.5, 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.

Remark 7 Note that eq. (18) 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 Δ , 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. [7] 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 Δ). This was due to the fact that their analysis essentially amounted to proving that eq. (18) hold for the special case of graphs of maximum degree $\Delta + 1$. Using an observation of Kahn and Kim [25] 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. (18) is optimal for graphs of connective constant Δ (in fact, the rate of strong spatial mixing on the d -ary tree is exactly $1 - \frac{2}{1 + \sqrt{1 + 4\gamma d}}$).

6 Branching factor and uniqueness of the Gibbs measure on general trees

We close with an application of our results to finding thresholds for the uniqueness of the Gibbs measure of the hard core model on locally finite infinite trees. Our bounds will be stated in terms of the *branching factor*, which has been shown to be the appropriate parameter for establishing phase transition thresholds for symmetric models such as the ferromagnetic Ising, Potts and Heisenberg models [30,40]. We begin with a general definition of the notion of uniqueness of Gibbs measure (see, for example, the survey article of Mossel [35]). Let T be a locally finite infinite tree rooted at ρ , and let C be a cutset in T . Consider the hard core model with vertex activity $\lambda > 0$ on T . We define the *discrepancy* $\delta(C)$ of C as follows. Let σ and τ be boundary conditions in

T which fix the state of the vertices on C , but not of any vertex in $T_{<C}$. Then, $\delta(C)$ is the maximum over all such σ and τ of the quantity $R_\rho(\sigma, T) - R_\rho(\tau, T)$.

Definition 12 (*Uniqueness of Gibbs measure*) The hard core model with vertex activity $\lambda > 0$ is said to exhibit *uniqueness of Gibbs measure* on T if there exists a sequence of cutsets $(B_i)_{i=1}^\infty$ such that $\lim_{i \rightarrow \infty} d(\rho, B_i) \rightarrow \infty$ and such that $\lim_{i \rightarrow \infty} \delta(B_i) = 0$.

Remark 8 Our definition of *uniqueness* here is similar in form to those used by Lyons [30] and Pemantle and Steif [40]. Notice, however, that the recurrence for the hard core model implies that the discrepancy is “monotonic” in the sense that if cutsets C and D are such that $C < D$ (i.e., every vertex in D is the descendant of some vertex in C) then $\delta(C) > \delta(D)$. This ensures that the choice of the sequence $(B_i)_{i=1}^\infty$ in the definition above is immaterial. For example, uniqueness is defined by Mossel [35] in terms of the cutsets C_ℓ consisting of vertices at distance exactly ℓ from the root. However, the above observation shows that for the hard core model, Mossel’s definition is equivalent to the one presented here.

We now define the notion of the branching factor of an infinite tree.

Definition 13 (*Branching factor* [30,31,40]) Let T be an infinite tree. The branching factor $\text{br}(T)$ is defined as follows:

$$\text{br}(T) := \inf \left\{ b > 0 \mid \inf_C \sum_{v \in C} b^{-|v|} = 0 \right\},$$

where the second infimum is taken over all cutsets C .

To clarify this definition, we consider some examples. If T is a d -ary tree, then $\text{br}(T) = d$. Further, by taking the second infimum over the cutsets C_ℓ of vertices at distance ℓ from the root, it is easy to see that the branching factor is never more than the connective constant. Further, Lyons [31] observes that in the case of spherically symmetric trees, one can define the branching factor as $\liminf_{\ell \rightarrow \infty} N(\rho, \ell)^\ell$.

We are now ready to state and prove our results on the uniqueness of the hard core model on general trees.

Theorem 6 *Let T be an infinite tree rooted at ρ with branching factor b . The hard core model with vertex activity $\lambda > 0$ exhibits uniqueness of Gibbs measure on T if $\lambda < \lambda_c(b)$.*

Proof As before, we apply Lemma 3 specialized to the message in eq. (10) with the values of a and q as chosen in eq. (12), where as before Δ_c is the unique positive real number satisfying $\lambda = \lambda_c(\Delta_c)$. Since $\lambda < \lambda_c(b)$, we then see from Lemma 6 that the decay factor $\alpha < \frac{1}{b}$. Hence, there is an $\epsilon > 0$ such that $\alpha = \frac{1}{b(1+\epsilon)}$ for some $\epsilon > 0$. Applying Lemma 3 to an arbitrary cutset C as in the proof of Theorem 1, we then get

$$\delta(C)^q \leq M_0 \sum_{v \in C} [(1 + \epsilon)b]^{-|v|}, \tag{19}$$

where M_0 is a constant. Since $b(1 + \epsilon) > \text{br}(T)$, the definition of $\text{br}(T)$ implies that we can find a sequence $(B_i)_{i=1}^{\infty}$ of cutsets such that

$$\lim_{i \rightarrow \infty} \sum_{v \in B_i} [(1 + \epsilon)b]^{-|v|} = 0.$$

Further, such a sequence must satisfy $\lim_{i \rightarrow \infty} d(\rho, B_i) = \infty$, since otherwise the limit above would be positive. Combining with eq. (19), this shows that $\lim_{i \rightarrow \infty} \delta(B_i) = 0$, which completes the proof.

7 Concluding remarks

Our results show that the connective constant is a natural notion of “average degree” of a graph that captures the spatial mixing property of the hard core and monomer-dimer models on the graph. We can then ask if such a correspondence holds also for other spin systems. More precisely, we start with results that establish decay of correlations in a graph when the parameters of the spin system lie in a region [28, 43, 50], say $R(d)$, determined only by the maximum degree d of the graph, and ask whether we can claim that the maximum degree d in those results can be replaced by $\Delta + 1$, where Δ is the connective constant of the graph (Theorem 1, for example, achieves precisely this goal for the hard core model).

It turns out that the only other spin system for which we know an exact analogue of Theorem 1 is the zero field Ising model (we omit the proof, but refer to Lyons [30], who proved a similar result for the uniqueness of Gibbs measure of the zero field Ising model on arbitrary trees). It is also the case that for the anti-ferromagnetic Ising model with field, one cannot hope to prove an exact analogue of Theorem 1. To see this, we consider the anti-ferromagnetic Ising model on a d -ary tree with vertex activity $\lambda > 1$ and edge activity $\beta < \frac{d-1}{d+1}$ (our notation for the anti-ferromagnetic Ising model is borrowed from [17] and [43]). For such β , it is known that there exists a critical activity $\lambda_c(\beta, d) > 1$ such that the model exhibits strong spatial mixing on the d -ary tree if and only if $|\log \lambda| > \lambda_c(\beta, d)$ (see [15, p. 254] and [43, Theorems 1 and 3]). Now consider a modification of the d -ary tree in which a new child is added separately to each vertex: it is clear that the connective constant of the resulting tree is still d , although the maximum degree is now $d + 2$. From the definition of the tree recurrences for the anti-ferromagnetic Ising model, it follows that the model with parameters (λ, β) on the new tree is equivalent to the same model on the original d -ary tree with parameters (λ', β) , where $\lambda' = \lambda \frac{\beta\lambda+1}{\beta+\lambda} < \lambda$ (when $\lambda > 1$). By taking the original λ sufficiently close to (but larger than) $\lambda_c(\beta, d)$, we can arrange that $1 < \lambda' < \lambda_c(\beta, d)$, which implies that the modified tree does not exhibit strong spatial mixing for some $\lambda > \lambda_c(\beta, d)$ despite having connective constant at most d . Note that the same argument does not apply to the hard core model because lowering the vertex activity of the hard core model is actually better for establishing strong spatial mixing (since we have strong spatial mixing on the d -ary tree whenever the vertex activity is smaller than a critical activity that depends upon the arity d of the tree). We note, however, that it is possible to obtain a weaker result for the anti-ferromagnetic Ising model which is analogous

to Theorem 1.3 in [44], in the sense that its hypotheses require an upper bound on the maximum degree as well. The proof of such a result mirrors the arguments in [44] exactly, and hence is omitted.

Pursuant to the above observations, we are thus led to ask if there is a natural notion of “average degree”—different from the connective constant—using which one can show decay of correlations for the anti-ferromagnetic Ising model with field on graphs of unbounded degree. More precisely, we would want such a notion to be powerful enough to yield an analogue of Corollary 1 for the anti-ferromagnetic Ising model with field on random sparse graphs.

The problem of identifying conditions under which a given spin system exhibits uniqueness of Gibbs measure on a specific lattices also has received much attention in the statistical physics literature. For example, it was predicted by Gaunt and Fisher [14] and by Baxter, Enting and Tsang [6] that the hard core model on \mathbb{Z}^2 should exhibit uniqueness of Gibbs measure as long as $\lambda < 3.79$. However, as we pointed out above, the condition $\lambda < 1.6875$ obtained by Weitz [50] was the best known rigorous upper bound until the work of Restrepo et al. [42] and Vera et al. [47], and these latter papers also employed Weitz’s idea of analyzing the self-avoiding walk (SAW) tree. Although the analysis in this paper is not as strictly specialized to the case of \mathbb{Z}^2 as were those of Restrepo et al. and Vera et al., it still follows Weitz’s paradigm of analyzing a SAW tree.

A natural question therefore arises: can the method of analyzing decay of correlations on the SAW tree establish uniqueness of Gibbs measure of the hard core model on \mathbb{Z}^2 under the condition $\lambda < 3.79$ predicted in [14] and [6]? One reason to suspect that this might not be the case is that the SAW tree approach typically establishes not only that uniqueness of the Gibbs measure holds on the lattice but also that *strong* spatial mixing holds on the SAW tree—the latter condition has algorithmic implications and is potentially stronger than the uniqueness of the Gibbs measure. It is in keeping with this suspicion that the best bounds we obtain for \mathbb{Z}^2 are still quite far from the conjectured bound of $\lambda < 3.79$. This contrast becomes even more pronounced when we consider the triangular lattice \mathbb{T} : Baxter [5] solved the hard core model on \mathbb{T} exactly (parts of Baxter’s solution were later made rigorous by Andrews [3]) using combinatorial tools, and showed that the uniqueness of Gibbs measure holds when $\lambda < 11.09$. On the other hand, the SAW tree approach—using published estimates of the connective constant of \mathbb{T} as a black box—can only show so far that strong spatial mixing holds under the condition $\lambda < 0.961$ (see Table 1). However, it must be pointed out that it may be possible to significantly improve this bound by analyzing the connective constant of the Weitz SAW tree of \mathbb{T} , as done in Appendix A for \mathbb{Z}^2 .

Acknowledgements We thank Elchanan Mossel, Allan Sly, Eric Vigoda and Dror Weitz for helpful discussions. We also thank the anonymous referees for detailed helpful comments.

A Description of numerical results

In this section, we describe the derivation of the numerical bounds in Table 1. As in [44], all of the bounds are direct applications of Theorem 1 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 [32, 49]. In contrast, for the honeycomb lattice, Duminil-Copin and Smirnov [9] rigorously established the connective constant to be \mathbb{H} is $\sqrt{2 + \sqrt{2}}$ in a recent breakthrough, and this is the bound we use for that lattice. In order to apply Theorem 1 for a given lattice of connective constant at most Δ , we simply need to compute $\lambda_c(\Delta) = \frac{\Delta^\Delta}{(\Delta-1)^{(\Delta+1)}}$, and the monotonicity of λ_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 the introduction, 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. [42] and Vera et al. [47] using sophisticated 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 [44], the proof of Theorem 1 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 [50]). 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 $v_0, v_1, \dots, v_l, v_0$ 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. [42].

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 viewed 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* [32]—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 ℓ 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* [42]. We therefore get $N(v, \ell) \leq \mathbf{e}_1^T M^\ell \mathbf{1}$, where $\mathbf{1}$ denotes the all 1’s vector, and \mathbf{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 γ . Using Gelfand’s formula (which states that $\gamma = \lim_{\ell \rightarrow \infty} \|M^\ell\|^{1/\ell}$, for any fixed matrix norm) with the ℓ_∞ norm to get the last equality, we see that

$$\limsup_{\ell \rightarrow \infty} N(v, \ell)^{1/\ell} \leq \limsup_{\ell \rightarrow \infty} (\mathbf{e}_1^T M^\ell \mathbf{1})^{1/\ell} \leq \limsup_{\ell \rightarrow \infty} \|M^\ell\|_\infty^{1/\ell} = \gamma.$$

Hence, the largest real eigenvalue γ 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 Δ , and applying Theorem 1 as described above, we get the bound 2.529 for λ in the notation of the table, which is better than the bounds obtained by Restrepo et al. [42] and Vera et al. [47]. 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 1 yields the bound 2.538 for λ , as quoted in Table 1.

B Proofs omitted from Sect. 3

We include here a proof of Lemma 1 for the convenience of the reader.

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

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

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

$$\begin{aligned} \left| f_{d,\lambda}^\phi(\mathbf{x}) - f_{d,\lambda}^\phi(\mathbf{y}) \right| &= |H'(s)| = \left| \left\langle \nabla f_{\lambda,d}^\phi(s\mathbf{x} + (1-s)\mathbf{y}), \mathbf{x} - \mathbf{y} \right\rangle \right| \\ &= \Phi(f_{d,\lambda}(\mathbf{z})) \left| \sum_{i=1}^d \frac{x_i - y_i}{\Phi(z_i)} \frac{\partial f_{d,\lambda}}{\partial z_i} \right|, \quad \text{using the chain rule} \\ &\leq \Phi(f_{d,\lambda}(\mathbf{z})) \sum_{i=1}^d \frac{|y_i - x_i|}{\Phi(z_i)} \left| \frac{\partial f_{d,\lambda}}{\partial z_i} \right|, \quad \text{as claimed.} \end{aligned}$$

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 $\mathbf{R} = \mathbf{z}$.

We now give the proof of the Lemma 3. The proof is syntactically identical to the proof of a similar lemma in [44], and the only difference (which is of course crucial for our purposes) is the use of the more specialized Lemma 2 in the inductive step.

Proof of Lemma 3 Recall that given a vertex v in $T_{\leq C}$, 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 σ 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_ρ , we will now show that for any vertex v in T_ρ which is at a distance $|v|$ from ρ , and has arity d_v , one has

$$|\phi(F_v(\sigma)) - \phi(F_v(\tau))|^q \leq M^q \sum_{u \in C_v} \alpha^{|u| - |v|}. \tag{20}$$

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. (20), the claim of the lemma follows from the above observation.

We now proceed to prove eq. (20). 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, \dots, v_{d_v} be the children of v , which satisfy eq. (20) by induction. In the remainder of the proof, we suppress the dependence of ξ on ϕ and q . Applying Lemma 2 followed by the induction hypothesis, we then have, for some positive integer $k \leq d_v$

$$\begin{aligned} & |\phi(R_v(\sigma)) - \phi(R_v(\tau))|^q \\ & \leq \xi(k) \sum_{i=1}^{d_v} |\phi(R_{v_i}(\sigma)) - \phi(R_{v_i}(\tau))|^q, \quad \text{using Lemma 2} \\ & \leq M^q \xi(k) \sum_{i=1}^{d_v} \sum_{u \in C_{v_i}} \alpha^{|u| - |v_i|}, \quad \text{using the induction hypothesis} \\ & \leq M^q \sum_{u \in C_v} \alpha^{|u| - |v|}, \quad \text{using } \xi(k) \leq \alpha \text{ and } |v_i| = |v| + 1. \end{aligned}$$

This completes the induction.

C Proofs omitted from Sect. 4

C.1 Maximum of ν_λ and implications for strong spatial mixing

In this section, we prove Lemma 6. We begin by giving a brief overview of how the value of q given in eq. (12) is chosen. As discussed in the paragraph following the equation, the idea is to start with the fact that $\xi_{\phi,q}(\Delta_c) = \frac{1}{\Delta_c}$ independent of the value of q , and then to choose q so that $\xi_{\phi,q}(d)$ is maximized at $d = \Delta_c$. Differentiation of $\xi_{\phi,q}(d)$ with respect to d for fixed λ and q leads to the expression in eq. (22), where $\tilde{x} = \tilde{x}_\lambda(d)$ is defined as the unique positive solution of the equation $d\tilde{x} = 1 + f_d(\tilde{x})$. We then require that, as a necessary condition for $\xi_{\phi,q}(d)$ to be maximized at $d = \Delta_c$, this derivative should vanish at this value of d . Using the fact that $\tilde{x}_\lambda(\Delta_c) = \frac{1}{\Delta_c - 1}$, we then see that q must be equal to the value given in eq. (12) for this requirement to be satisfied. The rest of the proof of the lemma then shows that when q is so chosen, the function $\nu_\lambda(d) := \xi_{\phi,q}(d)$ is indeed maximized at $d = \Delta_c$.

Proof of Lemma 6 We first prove that given λ , $\tilde{x}_\lambda(d)$ is a decreasing function of d . For ease of notation, we suppress the dependence of $\tilde{x}_\lambda(d)$ on d and λ . From Lemma 5, 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

$$\tilde{x}' = -\frac{(1 + \tilde{x}) [f_d(\tilde{x}) \log(1 + \tilde{x}) + \tilde{x}]}{d(1 + d)\tilde{x}}. \tag{21}$$

Since $\tilde{x} \geq \frac{1}{d} \geq 0$, this shows that \tilde{x} is a decreasing function of d .

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

$$\begin{aligned} v'_\lambda(d) &= \mathcal{E}^{(1,0)}(d, \tilde{x}) + \mathcal{E}^{(0,1)}(d, \tilde{x}) \frac{d\tilde{x}}{dd} \\ &= \mathcal{E}^{(1,0)}(d, \tilde{x}), \text{ since } \mathcal{E}^{(0,1)}(d, \tilde{x}) = 0 \text{ by definition of } \tilde{x} \\ &= \mathcal{E}(d, \tilde{x}) \left[\frac{q - 1}{d} - \frac{q \log(1 + \tilde{x})}{2(1 + f_{d,\lambda}(\tilde{x}))} \right] \\ &= \frac{q\mathcal{E}(d, \tilde{x})}{d} \left[1 - \frac{1}{q} - \frac{\log(1 + \tilde{x})}{2\tilde{x}} \right]. \end{aligned} \tag{22}$$

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 a strictly increasing function of \tilde{x} , and hence a strictly decreasing function of d . Since $\mathcal{E}(d, \tilde{x})$ is positive, this implies that there can be at most one positive zero of $v'_\lambda(d)$, and if such a zero exists, it is the unique maximum of $v_\lambda(d)$.

We now complete the proof by showing that $v'_\lambda(d) = 0$ 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 Φ was chosen in [28] in part to make sure that at the critical activity, the fixed point is also the maximizer of (an analogue of) $\mathcal{E}(d, \cdot)$.

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

$$\begin{aligned} v_\lambda(\Delta_c) &= \frac{q\mathcal{E}\left(\Delta_c, \frac{1}{\Delta_c-1}\right)}{2\Delta_c} \left[(\Delta_c - 1) \log\left(1 + \frac{1}{\Delta_c-1}\right) \right. \\ &\quad \left. - (\Delta_c - 1) \log\left(1 + \frac{1}{\Delta_c-1}\right) \right] = 0, \end{aligned}$$

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 v_λ , we have

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

which completes the proof.

C. 2 Symmetrizability of the message

In this section, we prove Lemma 4. We start with the following technical lemma.

Lemma 10 *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 10 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 .

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

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 $(x_1, 1/A]$. We thus see that h is strictly decreasing (and negative) in $(1, x_1)$ and strictly 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 strictly 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 $x = 1/A$.

We now prove Lemma 4.

Proof of Lemma 4 We begin by verifying the first condition in the definition of symmetrizability. For each $1 \leq i \leq d$, we have

$$\lim_{x_i \rightarrow 0^+} \frac{f_d(\mathbf{x})}{(1 + x_i)\Phi(x_i)} = \lim_{x_i \rightarrow 0^+} 2f_d(\mathbf{x})\sqrt{\frac{x_i}{1 + x_i}} = 0,$$

where we use the fact that $f_d(\mathbf{x}) > 0$ for non-negative \mathbf{x} . We now recall the program used in the definition of symmetrizability, with the definitions of Φ and f_d substituted, and with $r = a/2$:

$$\begin{aligned} \max \quad & \sum_{i=1}^d \left(\frac{x_i}{1 + x_i}\right)^r, \quad \text{where} \\ & \lambda \prod_{i=1}^d \frac{1}{1 + x_i} = B \\ & x_i \geq 0, \quad 1 \leq i \leq d \end{aligned} \tag{23}$$

Note that eq. (23) 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 \mathbf{y} be such a solution. Suppose without loss of generality that the first k co-ordinates of \mathbf{y} are non-zero while the rest are 0. We claim that $y_i = y_j \neq 0$ for all $1 \leq i \leq j \leq k$ and $y_i = 0$ for $i > k$.

To show this, we first define another vector 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. (23) is equivalent to

$$\prod_{i=1}^d s_i = B/\lambda. \tag{24}$$

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 10 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. (23, 24) 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 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 y is a maximum, and completes the proof.

D Proofs omitted from Sect. 5

D. 1 Symmetrizability of the message

In this section, we prove Lemma 7. As in the case of the hard core model, we begin with an auxiliary technical lemma.

Lemma 11 *Let r and a satisfy $1 < r \leq 2$ 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 $0 \leq s_1, s_2 \leq 1$ be such that $(s_1 + s_2)/2$ is constrained to be some fixed constant $a \leq 1$. 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 $a \leq 1/2$).

Proof of Lemma 11 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 γ , which is given by

$$\gamma'''(x) = -4r(r - 1)(1 - x)(1 - (1 - x)^2)^{r-2} \left[\frac{3 - (2r - 1)(1 - x)^2}{1 - (1 - x)^2} \right]. \tag{25}$$

Our first claim is that γ''' is strictly increasing in the interval $[0, 1]$ when $1 < r \leq 2$. In the case when $r = 2$, the last two factors in eq. (25) 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 strictly increasing and non-negative requires the condition that $r < 2$). Thus, because of the negative sign, γ''' 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 $A = 1 - a \neq a$, 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 $\gamma'(x) = 2rx^{r-1}(2 - x)^{r-1}(1 - x)$, and then observe that

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

We now start with the observation that $g'''(x) = \gamma'''(a + x) - \gamma'''(a - x)$, so that because of the strict monotonicity of γ''' 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 interval $(0, A]$ 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 $\{0, a\}$.

We now prove Lemma 7.

Proof of Lemma 7 We begin by verifying the first condition in the definition of symmetrizability:

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

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

$$\begin{aligned} \max \quad & \gamma^r f_{d,\gamma}(\mathbf{p})^{2r} \sum_{i=1}^d p_i^r (2 - p_i)^r, \quad \text{where} \\ & \frac{1}{1 + \gamma \sum_{i=1}^d p_i} = B \\ & 0 \leq p_i \leq 1, \quad 1 \leq i \leq d \end{aligned}$$

Since we are only interested in the values of \mathbf{p} solving the program, we can simplify the program as follows:

$$\begin{aligned} \max \quad & \sum_{i=1}^d p_i^r (2 - p_i)^r, \quad \text{where} \\ & \sum_{i=1}^d p_i = B' := \frac{1 - B}{\gamma B} \\ & 0 \leq p_i \leq 1, \quad 1 \leq i \leq d \end{aligned}$$

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 \mathbf{y} be such a solution. Suppose without loss of generality that the first k co-ordinates of \mathbf{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 11 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 $\gamma(y_i) + \gamma(y_j) = \gamma_i^r (2 - y_i)^r + \gamma_j^r (2 - y_j)^r$:

- (1) Making $y_i = y_j$, or
- (2) Making $y_i = 0$ (so that $y_j = 2a$). This case is possible only when $2a \leq 1$.

Thus, if \mathbf{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 \mathbf{y} is a maximum, and completes the proof.

References

1. Alm, S.E.: Upper bounds for the connective constant of self-avoiding walks. *Comb. Probab. Comput.* **2**(02), 115–136 (1993). doi:[10.1017/S0963548300000547](https://doi.org/10.1017/S0963548300000547)
2. Alm, S.E.: 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 (2005). doi:[10.1088/0305-4470/38/10/001](https://doi.org/10.1088/0305-4470/38/10/001)
3. Andrews, G.E.: The hard-hexagon model and Rogers–Ramanujan type identities. *Proc. Nat. Acad. Sci.* **78**(9), 5290–5292 (1981). <http://www.pnas.org/content/78/9/5290>. PMID: 16593082
4. Bandyopadhyay, A., Gamarnik, D.: Counting without sampling: asymptotics of the log-partition function for certain statistical physics models random structures and algorithms. *Random Struct. Algorithms* **33**(4), 452–479 (2008)
5. Baxter, R.J.: Hard hexagons: exact solution. *J. Phys. A: Math. Gen.* **13**(3), L61–L70 (1980). doi:[10.1088/0305-4470/13/3/007](https://doi.org/10.1088/0305-4470/13/3/007). <http://iopscience.iop.org/0305-4470/13/3/007>
6. Baxter, R.J., Enting, I.G., Tsang, S.K.: Hard-square lattice gas. *J. Stat. Phys.* **22**(4), 465–489 (1980). doi:[10.1007/BF01012867](https://doi.org/10.1007/BF01012867). <http://link.springer.com/article/10.1007/BF01012867>
7. Bayati, M., Gamarnik, D., Katz, D., Nair, C., Tetali, P.: Simple deterministic approximation algorithms for counting matchings. In: *Proc. 39th ACM Symp. Theory Comput.*, pp. 122–127. ACM (2007). doi:[10.1145/1250790.1250809](https://doi.org/10.1145/1250790.1250809)
8. Broadbent, S.R., Hammersley, J.M.: Percolation processes I. Crystals and mazes. *Math. Proc. Camb. Philos. Soc.* **53**(03), 629–641 (1957). doi:[10.1017/S0305004100032680](https://doi.org/10.1017/S0305004100032680)
9. Dumitriu-Copin, H., Smirnov, S.: The connective constant of the honeycomb lattice equals $\sqrt{2 + \sqrt{2}}$. *Ann. Math.* **175**(3), 1653–1665 (2012). doi:[10.4007/annals.2012.175.3.14](https://doi.org/10.4007/annals.2012.175.3.14)
10. Dyer, M., Greenhill, C.: On Markov chains for independent sets. *J. Algorithms* **35**(1), 17–49 (2000)
11. Efthymiou, C.: MCMC sampling colourings and independent sets of $\mathcal{G}(\setminus, \lceil/\setminus)$ near uniqueness threshold. In: *Proc. 25th ACM-SIAM Symp. Discret. Algorithms*, pp. 305–316. SIAM (2014). Full version available at [arXiv:1304.6666](https://arxiv.org/abs/1304.6666)
12. Galanis, A., Ge, Q., ŠtEFANKOVIČ, D., Vigoda, E., Yang, L.: Improved inapproximability results for counting independent sets in the hard-core model. *Random Struct. Algorithms* **45**(1), 78–110 (2014). doi:[10.1002/rsa.20479](https://doi.org/10.1002/rsa.20479)
13. Gamarnik, D., Katz, D.: Correlation decay and deterministic FPTAS for counting list-colorings of a graph. In: *Proc. 18th ACM-SIAM Symp. Discret. Algorithms*, pp. 1245–1254. SIAM (2007). <http://dl.acm.org/citation.cfm?id=1283383.1283517>
14. Gaunt, D.S., Fisher, M.E.: Hard-sphere lattice gases. I. Plane-square lattice. *J. Chem. Phys.* **43**(8), 2840–2863 (1965). doi:[10.1063/1.1697217](https://doi.org/10.1063/1.1697217). <http://scitation.aip.org/content/aip/journal/jcp/43/8/10.1063/1.1697217>
15. Georgii, H.O.: *Gibbs Measures and Phase Transitions*. De Gruyter Studies in Mathematics, Walter de Gruyter Inc, (1988)
16. Godsil, C.D.: Matchings and walks in graphs. *J. Graph Th.* **5**(3), 285–297 (1981). <http://onlinelibrary.wiley.com/doi/10.1002/jgt.3190050310/abstract>
17. Goldberg, L.A., Jerrum, M., Paterson, M.: The computational complexity of two-state spin systems. *Random Struct. Algorithms* **23**, 133–154 (2003)
18. Goldberg, L.A., Martin, R., Paterson, M.: Strong spatial mixing with fewer colors for lattice graphs. *SIAM J. Comput.* **35**(2), 486–517 (2005). doi:[10.1137/S0097539704445470](https://doi.org/10.1137/S0097539704445470). <http://link.aip.org/link/?SMJ/35/486/1>
19. Hammersley, J.M.: Percolation processes II. The connective constant. *Math. Proc. Camb. Philos. Soc.* **53**(03), 642–645 (1957). doi:[10.1017/S0305004100032692](https://doi.org/10.1017/S0305004100032692)
20. Hammersley, J.M., Morton, K.W.: Poor man’s Monte Carlo. *J. Royal Stat. Soc. B* **16**(1), 23–38 (1954). doi:[10.2307/2984008](https://doi.org/10.2307/2984008)
21. Hayes, T.P., Vigoda, E.: Coupling with the stationary distribution and improved sampling for colorings and independent sets. *Ann. Appl. Probab.* **16**(3), 1297–1318 (2006)

22. Jensen, I.: Enumeration of self-avoiding walks on the square lattice. *J. Phys. A* **37**(21), 5503 (2004). doi:[10.1088/0305-4470/37/21/002](https://doi.org/10.1088/0305-4470/37/21/002)
23. Jerrum, M., Sinclair, A.: Approximating the permanent. *SIAM J. Comput.* **18**(6), 1149–1178 (1989). doi:[10.1137/0218077](https://doi.org/10.1137/0218077). <http://epubs.siam.org/doi/abs/10.1137/0218077>
24. Jerrum, M., Valiant, L.G., Vazirani, V.V.: Random generation of combinatorial structures from a uniform distribution. *Theor. Comput. Sci.* **43**, 169–188 (1986)
25. Kahn, J., Kim, J.H.: Random matchings in regular graphs. *Combinatorica* **18**(2), 201–226 (1998). doi:[10.1007/PL00009817](https://doi.org/10.1007/PL00009817). <http://link.springer.com/article/10.1007/PL00009817>
26. Kesten, H.: On the number of self-avoiding walks. II. *J. Math. Phys.* **5**(8), 1128–1137 (1964). doi:[10.1063/1.1704216](https://doi.org/10.1063/1.1704216)
27. Li, L., Lu, P., Yin, Y.: Approximate counting via correlation decay in spin systems. In: Proc. 23rd ACM-SIAM Symp. Discret. Algorithms, pp. 922–940. SIAM (2012)
28. Li, L., Lu, P., Yin, Y.: Correlation decay up to uniqueness in spin systems. In: Proc. 24th ACM-SIAM Symp. Discret. Algorithms, pp. 67–84. SIAM (2013)
29. Luby, M., Vigoda, E.: Approximately counting up to four. In: Proc. 29th ACM Symp. Theory. Comput., pp. 682–687. ACM (1997). doi:[10.1145/258533.258663](https://doi.org/10.1145/258533.258663)
30. Lyons, R.: The Ising model and percolation on trees and tree-like graphs. *Commun. Math. Phys.* **125**(2), 337–353 (1989)
31. Lyons, R.: Random walks and percolation on trees. *Ann. Probab.* **18**(3), 931–958 (1990). doi:[10.1214/aop/1176990730](https://doi.org/10.1214/aop/1176990730)
32. Madras, N., Slade, G.: *The Self-Avoiding Walk*. Birkhäuser (1996)
33. Martinelli, F., Olivieri, E.: Approach to equilibrium of Glauber dynamics in the one phase region I. The attractive case. *Comm. Math. Phys.* **161**(3), 447–486 (1994). doi:[10.1007/BF02101929](https://doi.org/10.1007/BF02101929). <http://link.springer.com/article/10.1007/BF02101929>
34. Martinelli, F., Olivieri, E.: Approach to equilibrium of Glauber dynamics in the one phase region II. The general case. *Comm. Math. Phys.* **161**(3), 487–514 (1994). doi:[10.1007/BF02101930](https://doi.org/10.1007/BF02101930). <http://link.springer.com/article/10.1007/BF02101930>
35. Mossel, E.: Survey: Information flow on trees. In: *Graphs, Morphisms and Statistical Physics*, DIMACS Series in Discrete Mathematics and Theoretical Computer Science, vol. 63, pp. 155–170. American Mathematical Society (2004)
36. Mossel, E., Sly, A.: Rapid mixing of Gibbs sampling on graphs that are sparse on average. *Random Struct. Algorithms* **35**(2), 250–270 (2009). doi:[10.1002/rsa.20276](https://doi.org/10.1002/rsa.20276)
37. Mossel, E., Sly, A.: Gibbs rapidly samples colorings of $\mathcal{G}(\setminus, \lceil/\setminus)$. *Probab. Theory Relat. Fields* **148**(1–2), 37–69 (2010). doi:[10.1007/s00440-009-0222-x](https://doi.org/10.1007/s00440-009-0222-x)
38. Mossel, E., Sly, A.: Exact thresholds for Ising-Gibbs samplers on general graphs. *Ann. Probab.* **41**(1), 294–328 (2013). doi:[10.1214/11-AOP737](https://doi.org/10.1214/11-AOP737)
39. Nienhuis, B.: Exact critical point and critical exponents of $O(n)$ models in two dimensions. *Phys. Rev. Lett.* **49**(15), 1062–1065 (1982). doi:[10.1103/PhysRevLett.49.1062](https://doi.org/10.1103/PhysRevLett.49.1062)
40. Pemantle, R., Steif, J.E.: Robust phase transitions for Heisenberg and other models on general trees. *Ann. Probab.* **27**(2), 876–912 (1999)
41. Pönitz, A., Tittmann, P.: Improved upper bounds for self-avoiding walks in \mathbb{Z}^d . *Electron. J. Comb.*, **7**, Research Paper 21 (2000)
42. Restrepo, R., Shin, J., Tetali, P., Vigoda, E., Yang, L.: Improved mixing condition on the grid for counting and sampling independent sets. *Probab. Theory Relat. Fields* **156**(1–2), 75–99 (2013). Extended abstract in Proc. IEEE Symp. Found. Comput. Sci., 2011
43. Sinclair, A., Srivastava, P., Thurley, M.: Approximation algorithms for two-state anti-ferromagnetic spin systems on bounded degree graphs. *J. Stat. Phys.* **155**(4), 666–686 (2014)
44. Sinclair, A., Srivastava, P., Yin, Y.: Spatial mixing and approximation algorithms for graphs with bounded connective constant. In: Proc. 54th IEEE Symp. Found. Comput. Sci., pp. 300–309. IEEE Computer Society (2013). Full version available at [arXiv:1308.1762v1](https://arxiv.org/abs/1308.1762v1)
45. Sly, A.: Computational transition at the uniqueness threshold. In: Proc. 51st IEEE Symp. Found. Comput. Sci., pp. 287–296. IEEE Computer Society (2010)
46. Sly, A., Sun, N.: Counting in two-spin models on d -regular graphs. *Ann. Probab.* **42**(6), 2383–2416 (2014). doi:[10.1214/13-AOP888](https://doi.org/10.1214/13-AOP888). <http://projecteuclid.org/euclid.aop/1412083628>
47. Vera, J.C., Vigoda, E., Yang, L.: Improved bounds on the phase transition for the hard-core model in 2-dimensions. In: *Approximation, Randomization, and Combinatorial Optimization. Algorithms and*

- Techniques, Lecture Notes in Computer Science, vol 8096, pp. 699–713. Springer Berlin Heidelberg (2013). doi:[10.1007/978-3-642-40328-6_48](https://doi.org/10.1007/978-3-642-40328-6_48)
48. Vigoda, E.: A note on the Glauber dynamics for sampling independent sets. *Electron. J. Combin.* **8**(1), R8 (2001). <http://www.combinatorics.org/ojs/index.php/eljc/article/view/v8i1r8>
 49. Weisstein, E.W.: Self-avoiding walk connective constant. From MathWorld—a Wolfram Web Resource. <http://mathworld.wolfram.com/Self-AvoidingWalkConnectiveConstant.html>
 50. Weitz, D.: Counting independent sets up to the tree threshold. In: *Proc. 38th ACM Symp. Theory Comput.*, pp. 140–149. ACM (2006). doi:[10.1145/1132516.1132538](https://doi.org/10.1145/1132516.1132538)