2016 IEEE 57th Annual Symposium on Foundations of Computer Science

Convergence of MCMC and Loopy BP in the Tree Uniqueness Region for the Hard-Core Model

Charilaos Efthymiou*, Thomas P. Hayes†, Daniel Štefankovič‡, Eric Vigoda § and Yitong Yin¶
*Mathematics Institute, Goethe University, Frankfurt am Main, 60325, Germany
Email: efthymiou@gmail.com,
† Department Computer Science, University of New Mexico, Albuquerque, NM 87131, USA
Email: hayes@cs.unm.edu
‡Department of Computer Science, University of Rochester, Rochester, NY 14627, USA
Email: stefanko@cs.rochester.edu
§ School of Computer Science, Georgia Institute of Technology, Atlanta, GA 30332, USA
Email: ericvigoda@gmail.com
¶State Key Laboratory for Novel Software Technology. Nanjing University, Nanjing, China
Email: yinyt@nju.edu.cn

***Abstract*—We study the hard-core (gas) model defined on independent sets of an input graph where the independent sets are weighted by a parameter (aka fugacity) $\lambda > 0$. For constant Δ, previous work of Weitz (2006) established an FPTAS for the partition function for graphs of maximum degree Δ when $\lambda < \lambda_c(\Delta)$. Sly (2010) showed that there is no FPRAS, unless NP=RP, when $\lambda > \lambda_c(\Delta)$. The threshold $\lambda_c(\Delta)$ is the critical point for the statistical physics phase transition for uniqueness/non-uniqueness on the infinite Δ-regular tree. The running time of Weitz's algorithm is exponential in $\log \Delta$. Here we present an FPRAS for the partition function whose running time is $O^*(n^2)$. We analyze the simple single-site Markov chain known as the Glauber dynamics for sampling from the associated Gibbs distribution. We prove there exists a constant Δ_0 such that for all graphs with maximum degree $\Delta \geq \Delta_0$ and girth ≥ 7 (i.e., no cycles of length ≤ 6), the mixing time of the Glauber dynamics is $O(n \log n)$ when $\lambda < \lambda_c(\Delta)$. Our work complements that of Weitz which applies for small constant Δ whereas our work applies for all Δ at least a sufficiently large constant Δ_0 (this includes Δ depending on $n = |V|$).**

Our proof utilizes loopy BP (belief propagation) which is a widely-used algorithm for inference in graphical models. A novel aspect of our work is using the principal eigenvector for the BP operator to design a distance function which contracts in expectation for pairs of states that behave like the BP fixed point. We also prove that the Glauber dynamics behaves locally like loopy BP. As a byproduct we obtain that the Glauber dynamics, after a short burn-in period, converges close to the BP fixed point, and this implies that the fixed point of loopy BP is a close approximation to the Gibbs distribution. Using these connections we establish that loopy BP quickly converges to the Gibbs distribution when the girth ≥ 6 and $\lambda < \lambda_c(\Delta)$.

C. Efthymiou is supported by DFG grant EF 103/11, E. Vigoda is supported by NSF grants 1555579, 1617306, and 1563838, D. Štefankovič is supported by NSF grant CCF-1318374, Y. Yin is supported by NSFC grants 61272081 and 61321491.

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

I. Introduction

Background

The hard-core gas model is a natural combinatorial problem that has played an important role in the design of new approximate counting algorithms and for understanding computational connections to statistical physics phase transitions. For a graph $G = (V, E)$ and a fugacity $\lambda > 0$, the hard-core model is defined on the set Ω of independent sets of G where $\sigma \in \Omega$ has weight $w(\sigma) = \lambda^{|\sigma|}$. The equilibrium state of the system is described by the Gibbs distribution μ in which an independent set σ has probability $\mu(\sigma) = w(\sigma)/Z$. The partition function $Z = \sum_{\sigma \in \Omega} w(\sigma)$.

We study the closely related problems of efficiently approximating the partition function and approximate sampling from the Gibbs distribution. These problems are important for Bayesian inference in graphical models where the Gibbs distribution corresponds to the posterior or likelihood distributions. Common approaches used in practice are Markov Chain Monte Carlo (MCMC) algorithms and message passing algorithms, such as loopy BP (belief propagation), and one of the aims of this paper is to prove fast convergence of these algorithms.

Exact computation of the partition function is #P-complete [37], even for restricted input classes [10], hence the focus is on designing an efficient approximation scheme, either a deterministic FPTAS or randomized FPRAS. The existence of an FPRAS for the partition function is polynomial-time inter-reducible to approximate sampling from the Gibbs distribution.

A beautiful connection has been established: there is a computational phase transition on graphs of maximum degree Δ that coincides with the statistical physics phase transition on Δ-regular trees. The critical point for both of these phase transitions is $\lambda_c(\Delta) := (\Delta - 1)^{\Delta - 1}/(\Delta - 2)^{\Delta}$.

0272-5428/16 $31.00 © 2016 IEEE
DOI 10.1109/FOCS.2016.80

In statistical physics, $\lambda_c(\Delta)$ is the critical point for the uniqueness/non-uniqueness phase transition on the infinite Δ-regular tree $\mathbb{T}_\Delta$ [18] (roughly speaking, this is the phase transition for the decay versus persistence of the influence of the leaves on the root). For some basic intuition about the value of this critical point, note its asymptotics $\lambda_c(\Delta) \sim e/(\Delta-2)$ and the following basic property: $\lambda_c(\Delta) > 1$ for $\Delta \leq 5$ and $\lambda_c(\Delta) < 1$ for $\Delta \geq 6$.

Weitz [41] showed, for all constant Δ, an FPTAS for the partition function for all graphs of maximum degree Δ when $\lambda < \lambda_c(\Delta)$. To properly contrast the performance of our algorithm with Weitz's algorithm let us state his result more precisely: for all $\delta > 0$, there exists constant $C = C(\delta)$, for all Δ, all $G = (V, E)$ with maximum degree Δ, all $\lambda < (1-\delta)\lambda_c(\Delta)$, all $\epsilon > 0$, there is a deterministic algorithm to approximate Z within a factor $(1 \pm \epsilon)$ with running time $O\left((n/\epsilon)^{C \log \Delta}\right)$. An important limitation of Weitz's result is the exponential dependence on $\log \Delta$ in the running time. Hence it is polynomial-time only for constant Δ, and even in this case the running time is unsatisfying.

On the other side, Sly [33] (extended in [7], [8], [34], [9]) has established that, unless $NP = RP$, for all $\Delta \geq 3$, there exists $\gamma > 0$, for all $\lambda > \lambda_c(\Delta)$, there is no polynomial-time algorithm for triangle-free Δ-regular graphs to approximate the partition function within a factor $2^{\gamma n}$.

Weitz's algorithm was extremely influential: many works have built upon his algorithmic approach to establish efficient algorithms for a variety of problems (e.g., [28], [31], [19], [20], [32], [38], [21], [30], [22]). One of its key conceptual contributions was showing how decay of correlations properties on a Δ-regular tree are connected to the existence of an efficient algorithm for graphs of maximum degree Δ. We believe our paper enhances this insight by connecting these same decay of correlations properties on a Δ-regular tree to the analysis of widely-used Markov Chain Monte Carlo (MCMC) and message passing algorithms.

Main Results

As mentioned briefly earlier on, there are two widely-used approaches for the associated approximate counting/sampling problems, namely MCMC and message passing approaches. A popular MCMC algorithm is the simple single-site update Markov chain known as the Glauber dynamics. The Glauber dynamics is a Markov chain (X_t) on Ω whose transitions $X_t \to X_{t+1}$ are defined by the following process:

1) Choose v uniformly at random from V.
2) If $N(v) \cap X_t = \emptyset$ then let

$$X_{t+1} = \begin{cases} X_t \cup \{v\} & \text{with probability } \lambda/(1+\lambda) \\ X_t \setminus \{v\} & \text{with probability } 1/(1+\lambda) \end{cases}$$

3) If $N(v) \cap X_t \neq \emptyset$ then let $X_{t+1} = X_t$.

The mixing time T_{mix} is the number of steps to guarantee that the chain is within a specified (total) variation distance of the stationary distribution. In other words, for $\epsilon > 0$,

$$T_{\text{mix}}(\epsilon) = \min\{t : \text{ for all } X_0, d_{\mathsf{TV}}(X_t, \mu) \leq \epsilon\},$$

where $d_{\mathsf{TV}}()$ is the variation distance. We use $T_{\text{mix}} = T_{\text{mix}}(1/4)$ to refer to the mixing time for $\epsilon = 1/4$.

It is natural to conjecture that the Glauber dynamics has mixing time $O(n \log n)$ for all $\lambda < \lambda_c(\Delta)$. Indeed, Weitz's work implies rapid mixing for $\lambda < \lambda_c(\Delta)$ for amenable graphs. On the other hand Mossel et al. in [25] show slow mixing when $\lambda > \lambda_c(\Delta)$ on random regular bipartite graphs. The previously best known results for MCMC algorithms are far from reaching the critical point. It was known that the mixing time of the Glauber dynamics (and other simple, local Markov chains) is $O(n \log n)$ when $\lambda < 2/(\Delta-2)$ for any graph with maximum degree Δ [5], [23], [39]. In addition, [14] analyzed Δ-regular graphs with $\Delta = \Omega(\log n)$ and presented a polynomial-time simulated annealing algorithm when $\lambda < \lambda_c(\Delta)$.

Here we prove $O(n \log n)$ mixing time up to the critical point when the maximum degree is at least a sufficiently large constant Δ_0, and there are no cycles of length ≤ 6 (i.e., girth ≥ 7).

Theorem 1. *For all $\delta > 0$, there exists $\Delta_0 = \Delta_0(\delta)$ and $C = C(\delta)$, for all graphs $G = (V, E)$ of maximum degree $\Delta \geq \Delta_0$ and girth ≥ 7, all $\lambda < (1-\delta)\lambda_c(\Delta)$, all $\epsilon > 0$, the mixing time of the Glauber dynamics satisfies:*

$$T_{\text{mix}}(\epsilon) \leq Cn \log(n/\epsilon).$$

Note that Δ and λ can be a function of $n = |V|$. The above sampling result yields (via [35], [16]) an FPRAS for estimating the partition function Z with running time $O^*(n^2)$ where $O^*()$ hides multiplicative $\log n$ factors. The algorithm of Weitz [41] is polynomial-time for small constant Δ, in contrast our algorithm is polynomial-time for all $\Delta > \Delta_0$ for a sufficiently large constant Δ_0.

A family of graphs of particular interest are random Δ-regular graphs and random Δ-regular *bipartite* graphs. These graphs do not satisfy the girth requirements of Theorem 1 but they have few short cycles. Hence, as one would expect the above result extends to these graphs.

Theorem 2. *For all $\delta > 0$, there exists $\Delta_0 = \Delta_0(\delta)$ and $C = C(\delta)$, for all $\Delta \geq \Delta_0$, all $\lambda < (1-\delta)\lambda_c(\Delta)$, all $\epsilon > 0$, with probability $1 - o(1)$ over the choice of an n-vertex graph G chosen uniformly at random from the set of all Δ-regular (bipartite) graphs, the mixing time of the Glauber dynamics on G satisfies:*

$$T_{\text{mix}}(\epsilon) \leq Cn \log(n/\epsilon).$$

Theorem 2 complements the work in [25] which shows slow mixing for random Δ-regular bipartite graphs when $\lambda > \lambda_c(\Delta)$.

The other widely used approach is BP (belief propagation) based algorithms. BP, introduced by Pearl [27], is a simple recursive scheme designed on trees to correctly compute the marginal distribution for each vertex to be occupied/unoccupied. In particular, consider a rooted tree $T = (V, E)$ where for $v \in V$ its parent is denoted as p and its children are $N(v)$. Let

$$q(v) = \mathbf{Pr}_{\mu}\left[v \text{ is occupied} \mid p \text{ is unoccupied}\right]$$

denote the probability in the Gibbs distribution that v is occupied conditional on its parent p being unoccupied. It is convenient to work with ratios of the marginals, and hence let $R_{v\to p(v)} = q(v)/(1-q(v))$ denote the ratio of the occupied to unoccupied marginal probabilities. Because T is a tree then it is not difficult to show that this ratio satisfies the following recurrence:

$$R_{v\to p(v)} = \lambda \prod_{w\in N(v)\setminus\{p(v)\}} \frac{1}{1+R_{w\to v}}.$$

This recurrence explains the terminology of BP that $R_{w\to v}$ is a "message" from w to its parent v. Given the messages to v from all of its children then v can send its message to its parent. Finally the root r (with a parent p always fixed to be unoccupied and thus removed) can compute the marginal probability that it is occupied by: $q(r) = R_{r\to p}/(1+R_{r\to p})$.

The above formulation defines (the sum-product version of) BP a simple, natural algorithm which works efficiently and correctly for trees. For general graphs *loopy BP* implements the above approach, even though there are now cycles and so the algorithm no longer is guaranteed to work correctly. For a graph $G = (V, E)$, for $v \in V$ let $N(v)$ denote the set of all neighbors of v. For each $p \in N(v)$ and time $t \geq 0$ we define a message

$$R^t_{v\to p} = \lambda \prod_{w\in N(v)\setminus\{p\}} \frac{1}{1+R^{t-1}_{w\to v}}.$$

The corresponding estimate of the marginal can be computed from the messages by:

$$q^t(v,p) = \frac{R^t_{v\to p}}{1+R^t_{v\to p}}. \tag{1}$$

Loopy BP is a popular algorithm for estimating marginal probabilities in general graphical models (e.g., see [26]), but there are few results on when loopy BP converges to the Gibbs distribution (e.g., Weiss [40] analyzed graphs with one cycle, and [36], [15], [17] presented various sufficient conditions, see also [2], [29] for analysis of BP variants). We have an approach for analyzing loopy BP and in this project we will prove that loopy BP works well in a broad range of parameters. Its behavior relates to phase transitions in the underlying model, we detail our approach and expected results after formally presenting phase transitions.

We prove that, on any graph with girth ≥ 6 and maximum degree $\Delta \geq \Delta_0$ where Δ_0 is a sufficiently large constant, loopy BP quickly converges to the (marginals of) Gibbs distribution μ. More precisely, $O(1)$ iterations of loopy BP suffices, note each iteration of BP takes $O(n+m)$ time where $n = |V|$ and $m = |E|$.

Theorem 3. *For all $\delta, \epsilon > 0$, there exists $\Delta_0 = \Delta_0(\delta,\epsilon)$ and $C = C(\delta,\epsilon)$, for all graphs $G = (V, E)$ of maximum degree $\Delta \geq \Delta_0$ and girth ≥ 6, all $\lambda < (1-\delta)\lambda_c(\Delta)$, the following holds: for $t \geq C$, for all $v \in V$, $p \in N(v)$,*

$$\left|\frac{q^t(v,p)}{\mu(v \textit{ is occupied} \mid p \textit{ is unoccupied})} - 1\right| \leq \epsilon$$

where $\mu(\cdot)$ is the Gibbs distribution.

Contributions

Our main conceptual contribution is formally connecting the behavior of BP and the Glauber dynamics. We will analyze the Glauber dynamics using path coupling [1]. In path coupling we need to analyze a pair of *neighboring configurations*, in our setting this is a pair of independent sets X_t, Y_t which differ at exactly one vertex v. The key is to construct a one-step coupling $(X_t, Y_t) \to (X_{t+1}, Y_{t+1})$ and introduce a distance function $\Phi : \Omega \times \Omega \to \mathbf{R}_{\geq 0}$ which "contracts" meaning that the following *path coupling condition* holds for some $\gamma > 0$:

$$\mathbb{E}\left[\Phi(X_{t+1}, Y_{t+1}) \mid X_t, Y_t\right] \leq (1-\gamma)\Phi(X_t, Y_t).$$

We use a simple maximal one-step coupling and hence in our setting the path coupling condition simplifies to:

$$(1-\gamma)\Phi(X_t, Y_t) \geq \sum_{z\in N(v)} \frac{\lambda}{1+\lambda} \mathbf{1}\left\{z \text{ is unblocked in } X_t\right\}\Phi(z),$$

where *unblocked* means that $N(z) \cap X_t = \emptyset$, i.e., all neighbors of z are unoccupied, and we have assumed there are no triangles so as to ignore the possibility that X_t and Y_t differ on the neighborhood of z.

The distance function Φ must satisfy a few basic conditions such as being a path metric, and if $X \neq Y$ then $\Phi(X, Y) \geq 1$ (so that by Markov's inequality $\mathbf{Pr}\left[X_t \neq Y_t\right] \leq \mathbb{E}\left[\Phi(X_t, Y_t)\right]$). A standard choice for the distance function is the Hamming distance. In our setting the Hamming distance does not suffice and our primary challenge is determining a suitable distance function.

We cannot construct a suitable distance function which satisfies the path coupling condition for arbitrary neighboring pairs X_t, Y_t. But, a key insight is that we can show the existence of a suitable Φ when the local neighborhood of the disagreement v behaves like the BP fixpoint. Our construction of this Φ is quite intriguing.

In our proofs it is useful to consider the (unrooted) BP recurrences corresponding to the probability that a vertex is unblocked. This corresponds to the following function F :

$[0,1]^V \to [0,1]^V$ which is defined as follows, for any $\omega \in [0,1]^V$ and $z \in V$:

$$F(\omega)(z) = \prod_{y \in N(z)} \frac{1}{1+\lambda\omega(y)}. \tag{2}$$

Also, for some integer $i \geq 0$, let $F^i(\omega) : [0,1]^V \to [0,1]^V$ be the i-iterate of F. This recurrence is closely related to the standard BP operator $R()$ and hence under the hypotheses of our main results, we have that $F()$ has a unique fixed point ω^*, and for any ω, all $z \in V$, $\lim_{i\to\infty} F^i(z) = \omega^*(z)$.

To construct the distance function Φ we start with the Jacobian of this BP operator $F()$. By a suitable matrix diagonalization we obtain the path coupling condition. Since $F()$ converges to a fixed point, and, in fact, it contracts at every level with respect to an appropriately defined potential function, we then know that the Jacobian of the BP operator $F()$ evaluated at its fixed point ω^* has spectral radius < 1 and hence the same holds for the path coupling condition for pairs of states that are BP fixed points. This yields a function Φ that satisfies the following system of inequalities

$$\Phi(v) > \sum_{z \in N(v)} \frac{\lambda\omega^*(z)}{1+\lambda\omega^*(z)}\Phi(z). \tag{3}$$

However for the path coupling condition a stronger version of the above is necessary. More specifically, the sum on the r.h.s. should be appropriately bounded away from $\Phi(v)$, i.e. we need to have

$$(1-\gamma)\Phi(v) > \sum_{z \in N(v)} \frac{\lambda\omega^*(z)}{1+\lambda\omega^*(z)}\Phi(z).$$

Additionally, Φ should be a distance metric, e.g. $\Phi > 0$. It turns out that we use further properties of the distance function Φ, hence we need to explicitly derive a Φ.

There are previous works [12], [13] which utilize the spectral radius of the adjacency matrix of the input graph G to design a suitable distance function for path coupling. In contrast, we use insights from the analysis of the BP operator to derive a suitable distance function. We believe this is a richer connection that can potentially lead to stronger results since it directly relates to convergence properties on the tree. Our approach has the potential to apply for a more general class of spin systems, we comment on this in more detail in the conclusions.

The above argument only implies that we have contraction in the path coupling condition for pairs of configurations which are BP fixed points. A priori we don't even know if the BP fixed points on the tree correspond to the Gibbs distribution on the input graph. We prove that the Glauber dynamics (approximately) satisfies a recurrence that is close to the BP recurrence; this builds upon ideas of Hayes [11] for colorings. This argument requires that there are no cycles of length ≤ 6 for the Glauber dynamics (and no cycles of length ≤ 5 for the direct analysis of the Gibbs distribution). Some local sparsity condition is necessary since if there are many short cycles then the Gibbs distribution no longer behaves similarly to a tree and hence loopy BP may be a poor estimator.

As a consequence of the above relation between BP and the Glauber dynamics, we establish that from an arbitrary initial configuration X_0, after a short burn-in period of $T = O(n \log \Delta)$ steps of the Glauber dynamics the configuration X_T is a close approximation to the BP fixed point. In particular, for any vertex v, the number of unblocked neighbors of v in X_T is $\approx \sum_{z \in N(v)} \omega^*(z)$ with high probability. As is standard for concentration results, our proof of this result necessitates that Δ is at least a sufficiently large constant. Finally we adapt ideas of [4] to utilize these burn-in properties and establish rapid mixing of the Glauber dynamics.

Outline of Paper

The full proofs of our results are quite lengthy and so we defer many to the full version of our work which appears online in [6].

In the following section we state results about the convergence of the BP recurrences. We then present in Section III our theorem showing the existence of a suitable distance function for path coupling for pairs of states at the BP fixed point. Section IV sketches the proofs for our local uniformity results that after a burn-in period the Glauber dynamics behaves locally similar to the BP recurrences. Finally, in Section V we outline the proof of Theorem 1 of rapid mixing for the Glauber dynamics. The extension to random regular (bipartite) graphs as stated in Theorem 2 is proven in Section F of the full version. Theorem 3 about the efficiency of loopy BP is proven in Section B of the full version, the key technical results in the proof are sketched in Section IV.

II. BP Convergence

Here we state several useful results about the convergence of BP to a unique fixed point, and stepwise contraction of BP to the fixed point. The lemmas presented in this section are proved in Section A of the full version.

Our first lemma (which is proved using ideas from [28], [20], [31]) says that the recurrence for $F()$ defined in (2) has a unique fixed point.

Lemma 4. *For all $\delta > 0$, there exists $\Delta_0 = \Delta_0(\delta)$, for all $G = (V, E)$ of maximum degree $\Delta \geq \Delta_0$, all $\lambda < (1-\delta)\lambda_c(\Delta)$, the function F has a unique fixed point ω^*.*

A critical result for our approach is that the recurrences $F()$ have stepwise contraction to the fixed point ω^*. To obtain contraction we use the following potential function Ψ. Let the function $\Psi : [0,1] \to \mathbb{R}_{\geq 0}$ be as follows,

$$\Psi(x) = (\sqrt{\lambda})^{-1}\text{arcsinh}\left(\sqrt{\lambda \cdot x}\right). \tag{4}$$

Our main motivation for introducing Ψ is as a normalizing potential function that we use to define the following distance metric, D, on functions $\omega \in [0,1]^V$:

$$D(\omega_1, \omega_2) = \max_{z \in V} |\Psi(\omega_1(z)) - \Psi(\omega_2(z))|.$$

We will also need a variant, $D_{v,R}$, of this metric whose value only depends on the restriction of the function to a ball of radius ℓ around vertex v. For any $v \in V$, integer $\ell \geq 0$, let $B(v,\ell)$ be the set of vertices within distance $\leq \ell$ of v. Moreover, for functions $\omega_1, \omega_2 \in [0,1]^V$, we define:

$$D_{v,\ell}(\omega_1, \omega_2) = \max_{z \in B(v,\ell)} |\Psi(\omega_1(z)) - \Psi(\omega_2(z))|. \quad (5)$$

We can now state the following convergence result for the recurrences, which establishes stepwise contraction.

Lemma 5. *For all $\delta > 0$, there exists $\Delta_0 = \Delta_0(\delta)$, for all $G = (V,E)$ of maximum degree $\Delta \geq \Delta_0$, all $\lambda < (1-\delta)\lambda_c(\Delta)$, for any $\omega \in [0,1]^V$, $v \in V$ and $\ell \geq 1$, we have:*

$$D_{v,\ell-1}(F(\omega), \omega^*) \leq (1-\delta/6) D_{v,\ell}(\omega, \omega^*).$$

where ω^* is the fixed point of F.

III. Path Coupling Distance Function

We now prove that there exists a suitable distance function Φ for which the path coupling condition holds for configurations that correspond to the fixed points of $F()$.

Theorem 6. *For all $\delta > 0$, there exists $\Delta_0 = \Delta_0(\delta)$, for all $G = (V,E)$ of maximum degree $\Delta \geq \Delta_0$, all $\lambda < (1-\delta)\lambda_c(\Delta)$, there exists $\Phi : V \to \mathbb{R}_{\geq 0}$ such that for every $v \in V$,*

$$1 \leq \Phi(v) \leq 12, \quad (6)$$

and

$$(1-\delta/6)\Phi(v) \geq \sum_{u \in N(v)} \frac{\lambda \omega^*(u)}{1+\lambda\omega^*(u)} \Phi(u), \quad (7)$$

where ω^* is the fixed point of F defined in (2)*.*

Proof: We will prove here that the convergence of BP provides the existence of a distance function Φ satisfying (7). We defer the technical proof of (6) to Section A of the full version.

The Jacobian J of the BP operator F is given by

$$J(v,u) = \left| \frac{\partial F(\omega)(v)}{\partial \omega(u)} \right| = \begin{cases} \frac{\lambda F(\omega)(v)}{1+\lambda\omega(u)} & \text{if } u \in N_v \\ 0 & \text{otherwise} \end{cases}$$

Let $J^* = J|_{\omega=\omega^*}$ denote the Jacobian at the fixed point $\omega = \omega^*$. Let D be the diagonal matrix with $D(v,v) = \omega^*(v)$ and let $\hat{J} = D^{-1} J^* D$.

The path coupling condition (7) is in fact

$$\hat{J}\Phi \leq (1-\delta/6)\Phi. \quad (8)$$

The fact that ω^* is a Jacobian attractive fixpoint implies the existence of a nonnegative Φ with $\hat{J}\Phi < \Phi$. Thus, the theorem would follow immediately if the spectral radius of $\hat{J}$ is $\rho(\hat{J}) \leq 1-\delta/6$ and $\hat{J}$ has a principal eigenvector with each entry from the bounded range $[1,12]$. However, explicitly calculating this principal eigenvector can be challenging on general graphs.

The convergence of BP which is established in Lemmas 4, 5, with respect to the potential function Ψ, guides us to an explicit construction of Φ such that $\hat{J}\Phi < \Phi$. Indeed, let $\Psi'(x) = \frac{1}{2\sqrt{x(1+\lambda x)}}$ denote the derivative of the potential function Ψ. It will follow from the proof of Lemma 5 that:

$$\sum_{u \in N(v)} J^*(v,u) \frac{\Psi'(\omega^*(v))}{\Psi'(\omega^*(u))} \leq 1-\delta/6.$$

This inequality is due to the contraction of the BP system at the fixed point with respect to the potential function Ψ. It is equivalent to the following:

$$\sum_{u \in N(v)} \frac{\hat{J}(v,u)}{\omega^*(u)\Psi'(\omega^*(u))} \leq \frac{1-\delta/6}{\omega^*(v)\Psi'(\omega^*(v))}.$$

Then, (8) is trivially satisfied by choosing Φ such that $\Phi(v) = \frac{1}{2\omega^*(v)\Psi'(\omega^*(v))} = \sqrt{\frac{1+\lambda\omega^*(v)}{\omega^*(v)}}$. In turn we get the path coupling condition (7). The verification of (6) is in Section A of the full version.

■

IV. Local Uniformity for the Glauber Dynamics

We will prove that the Glauber dynamics, after a sufficient burn-in, behaves with high probability locally similar to the BP fixed points. In this section we will formally state some of these "local uniformity" results and sketch the main ideas in their proof. The proofs are quite technical and deferred to Section D of the full version.

For an independent set σ, for $v \in V$, and $p \in N(v)$ let

$$\mathbf{U}_{v,p}(\sigma) = \mathbf{1}\{\sigma \cap (N(v) \setminus \{p\}) = \emptyset\} \quad (9)$$

be the indicator of whether the children of v leave v unblocked.

We now state our main local uniformity results. We first establish that the Gibbs distribution behaves as in the BP fixpoint, when the girth ≥ 6. We will prove that for any vertex v, the number of unblocked neighbors of v is $\approx \sum_{z \in N(v)} \omega^*(z)$ with high probability. Hence, for $v \in V$ let

$$\mathbf{S}_X(v) = \sum_{z \in N(v)} \mathbf{U}_{z,v}(X),$$

denote the number of unblocked neighbors of v in configuration X.

Theorem 7. *For all $\delta, \epsilon > 0$, there exists $\Delta_0 = \Delta_0(\delta,\epsilon)$ and $C = C(\delta,\epsilon)$, for all graphs $G = (V,E)$ of maximum*

degree $\Delta \geq \Delta_0$ and girth ≥ 6, all $\lambda < (1-\delta)\lambda_c(\Delta)$, for all $v \in V$, it holds that:

$$\mathbf{Pr}_{X\sim\mu}\left[\left|\mathbf{S}_X(v) - \sum_{z\in N(v)} \omega^*(z)\right| \leq \epsilon\Delta\right] \geq 1-\exp\left(-\Delta/C\right),$$

where ω^* is the fixpoint from Lemma 4.

Theorem 7 will be the key ingredient in the proof of Theorem 3 (to be precise, the upcoming Lemma 9 is the key element in the proofs of Theorems 3 and 7).

For our rapid mixing result (Theorem 2) we need an analogous local uniformity result for the Glauber dynamics. This will require the slightly higher girth requirement ≥ 7 since the grandchildren of a vertex v no longer have a certain conditionally independence and we need the additional girth requirement to derive an approximate version of the conditional independence (this is discussed in more detail in Section C.3 of the full version).

The path coupling proof weights the vertices according to Φ. Hence, in place of $\mathbf{S}$ we need the following weighted version $\mathbf{W}$. For $v \in V$ and $\Phi : V \to \mathbb{R}_{\geq 0}$ as defined in Theorem 6 let

$$\mathbf{W}_\sigma(v) = \sum_{z\in N(v)} \mathbf{U}_{z,v}(\sigma)\ \Phi(z). \tag{10}$$

We then prove that the Glauber dynamics, after sufficient burn-in, also behaves as in the BP fixpoint with a slightly higher girth requirement ≥ 7. (For path coupling we only need an upper bound on the number of unblocked neighbors, hence we state and prove this simpler form.)

Theorem 8. *For all $\delta, \epsilon > 0$, let $\Delta_0 = \Delta_0(\delta,\epsilon), C = C(\delta,\epsilon)$, for all graphs $G = (V,E)$ of maximum degree $\Delta \geq \Delta_0$ and girth ≥ 7, all $\lambda < (1-\delta)\lambda_c(\Delta)$, let (X_t) be the Glauber dynamics on the hard-core model. For all $v \in V$, it holds that*

$$\mathbf{Pr}\left[(\forall t \in \mathcal{I}) \quad \mathbf{W}_{X_t}(v) < \sum_{z\in N(v)} \omega^*(z)\Phi(z) + \epsilon\Delta\right] \geq 1-\exp\left(-\Delta/C\right), \tag{11}$$

where the time interval $\mathcal{I} = [Cn\log\Delta, n\exp(\Delta/C)]$.

A. Proof sketch for local uniformity results

Here we sketch the simpler proof of Theorem 7 of the local uniformity results for the Gibbs distribution. This will illustrate the main conceptual ideas in the proof for the Gibbs distribution, and we will indicate the extra challenge for the analysis of the Glauber dynamics in the proof of Theorem 8. The full proofs for Theorems 7 and 8 are in Section D of the full version.

Consider a graph $G = (V,E)$. For a vertex v and an independent set σ, consider the following quantity:

$$\mathbf{R}(\sigma, v) = \prod_{z\in N(v)} \left(1 - \frac{\lambda}{1+\lambda}\mathbf{U}_{z,v}(\sigma)\right), \tag{12}$$

where $\mathbf{U}_{z.v}(\sigma)$ is defined in (9) (it is the indicator that the children of z leave it unblocked). The important aspect of this quantity $\mathbf{R}$ is the following qualitative interpretation. Let Y be distributed as in the Gibbs measure w.r.t. G. For triangle-free G we have

$$\mathbf{R}(\sigma, v) = \mathbf{Pr}\left[v \text{ is unblocked} \mid v \notin Y,\ Y(S_2(v)) = \sigma(S_2(v))\right],$$

where $S_2(z)$ are those vertices distance 2 from z and by "$z \notin \sigma$" we mean that z is not occupied. Moreover, conditional on the configuration at z and $S_2(z)$ the neighbors of z are independent in the Gibbs distribution and hence:

$$\mathbf{R}(\sigma, v) = \prod_{z\in N(v)} \mathbf{Pr}\left[z \notin Y \mid v \notin Y,\ Y(S_2(v)) = \sigma(S_2(v))\right]. \tag{13}$$

In the special case where the underlying graph is a *tree* we can extend (13) to the following recursive equations: Let X be distributed as in μ. We have that

$$\mathbf{R}(X, v) = \prod_{z\in N(v)} \left(1 - \frac{\lambda}{1+\lambda}\mathbf{R}(X, z)\right) + O(1/\Delta), \tag{14}$$

For our purpose it turns out that $\mathbf{R}(X,\cdot)$ is an approximate version of $F()$ defined in (2). The error term $O(1/\Delta)$ in (14) is negligible. For understanding $\mathbf{R}(X,\cdot)$ qualitatively, this error term can be completely ignored.

Consider the (BP system of) equations in (14), which is exact on trees. Nothing prevents us from applying (14) on the graph G and get the loopy version of the equations. Now, (14) does not necessarily compute the probability for v to be unblocked. However, we show the following interesting result regarding the quantity $\mathbf{S}_X(v)$, for every $v \in V$. With probability $\geq 1 - \exp(-\Omega(\Delta))$, it holds that

$$\left|\mathbf{S}_X(v) - \sum_{z\in N(v)} \mathbf{R}(X, z)\right| \leq \epsilon\Delta. \tag{15}$$

That is, we can approximate $\mathbf{S}_X(v)$ by using quantities that arise from the loopy BP equations. Still, getting a handle on $\mathbf{R}(X,z)$ in (15) is a non-trivial task. To this end, we show that $X \sim \mu$ satisfies (14) in the following approximate sense:

Lemma 9. *For all $\gamma, \delta > 0$, there exists $\Delta_0, C > 0$, for all graphs $G = (V,E)$ of maximum degree $\Delta \geq \Delta_0$ and girth ≥ 6 all $\lambda < (1-\delta)\lambda_c(\Delta)$ for all $v \in V$ the following is true:*

Let X be distributed as in μ. Then with probability $\geq 1 - \exp(-\Delta/C)$ it holds that

$$\left|\mathbf{R}(X, v) - \prod_{z\in N(v)} \left(1 - \frac{\lambda}{1+\lambda}\mathbf{R}(X, z)\right)\right| < \gamma. \tag{16}$$

We will argue (via (16)) that $\mathbf{R}()$ is an approximate version of $F()$ and then we can apply Lemma 5 to deduce

convergence (close) to the fixpoint ω^*. Consequently, we will prove that for every $v \in V$, with probability at least $1 - \exp(-\Omega(\Delta))$, it holds that

$$|\mathbf{R}(X, v) - \omega^*(v)| \leq \epsilon. \tag{17}$$

(See Lemma 16 in Section B.2 of the full version for a formal statement.) Combining (17) and (15) will finish the proof of Theorem 7. For the detailed proof of Theorem 7 see Section D.1 in the full version.

B. Approximate recurrence - Proof of Lemma 9

Here we prove Lemma 9 which shows that $\mathbf{R}$ satisfies an approximate recurrence similar to loopy BP, this is the main result in the proof of Theorem 7. Before beginning the proof we illustrate the necessity of the girth assumption.

Recall that for triangle-free graphs we have conditional independence in (13) for the neighbors of vertex z. In (15) we need to consider $\sum_{z \in N(v)} \mathbf{R}(X, z)$. To get independence on the grandchildren of v we need to condition on $S_3(v)$, this will require girth ≥ 6, see (18) below.

Proof of Lemma 9: Consider X distributed as in μ. Given some vertex $v \in V$, let $\mathcal{F}$ be the σ-algebra generated by the configuration of v and the vertices at distance ≥ 3 from v.

Note that $\lambda_c(\Delta) \sim e/\Delta$. So, for $\lambda < \lambda_c(\Delta)$ and $\Delta > \Delta_0$ we have $\lambda = O(1/\Delta)$.

Note that $\mathbf{S}_X(v)$ is a function of the configuration at $S_2(v)$. Conditional on $\mathcal{F}$, for any $z, z' \in N(v)$ the configurations at $N(z)\backslash\{v\}$ and $N(z')\backslash\{v\}$ are independent with each other. That is, conditional on $\mathcal{F}$, the quantity $\mathbf{S}_X(v)$ is a sum of $|N(v)|$ many independent random variables in $\{0, 1\}$. Then, applying Azuma's inequality (the Lipschitz constant is 1) we get that

$$\mathbf{Pr}\left[|\mathbb{E}\left[\mathbf{S}_X(v) \mid \mathcal{F}\right] - \mathbf{S}_X(v)| \leq \beta\Delta\right] \geq 1 - 2\exp\left(-\beta^2\Delta/2\right), \tag{18}$$

for any $\beta > 0$.

For $x \in \mathbb{R}_{\geq 0}$, let $f(x) = \exp\left(-\frac{\lambda}{1+\lambda}x\right)$. Since $\lambda \leq e/\Delta$ for $\Delta \geq \Delta_0$, then for $|\gamma| \leq (3e)^{-1}$ it holds that $f(x+\gamma\Delta) \leq 10\gamma$. Using these observations and (18) we get the following: for $0 < \beta < (3e)^{-1}$ it holds that

$$\begin{aligned}\mathbf{Pr}\left[|f(\mathbf{S}_X(v)) - f(\mathbb{E}\left[\mathbf{S}_X(v) \mid \mathcal{F}\right])| \leq 10\beta\right] &\geq 1 - 2\exp\left(-\beta^2\Delta/2\right).\end{aligned} \tag{19}$$

Recalling the definition of $\mathbf{R}(X, v)$, we have that

$$\begin{aligned}\mathbf{R}(X, v) &= \prod_{z \in N(v)}\left(1 - \frac{\lambda}{1+\lambda}\mathbf{U}_{z,v}(X)\right) \\ &= \exp\left(-\frac{\lambda}{1+\lambda}\sum_{z \in N(v)}\mathbf{U}_{z,v}(X) + O\left(1/\Delta\right)\right) \\ &= f(\mathbf{S}_X(v)) + O\left(1/\Delta\right),\end{aligned} \tag{20}$$

where the second equality we use the fact that $\lambda = O(1/\Delta)$ and that for $|x| < 1$ we have $1 + x = \exp(x + O(x^2))$; the last equality follows by noting that $f(\mathbf{S}_X(v)) \leq 1$.

We are now going to show that for every $z \in N(v)$ it holds that

$$|\mathbb{E}\left[\mathbf{U}_{z,v}(X) \mid \mathcal{F}\right] - \mathbf{R}(X, z)| \leq 2\lambda. \tag{21}$$

Before showing that (21) is indeed correct, let us show how we use it to get the lemma.

We have that

$$\begin{aligned}&f(\mathbb{E}\left[\mathbf{S}_X(v) \mid \mathcal{F}\right]) \\ &\quad= \exp\left(-\frac{\lambda}{1+\lambda}\sum_{z \in N(v)}\mathbb{E}\left[\mathbf{U}_{z,v}(X_t) \mid \mathcal{F}\right]\right) \\ &\quad= \exp\left(-\frac{\lambda}{1+\lambda}\sum_{z \in N(v)}\mathbf{R}(X, z)\right) + O(1/\Delta),\end{aligned} \tag{22}$$

where in the first derivation we use linearity of expectation and in the second derivation we use (21) and the fact that $\lambda = O(1/\Delta)$.

The lemma follows by plugging (22) and (20) into (19) and taking sufficiently large Δ.

It remains to show (21). We first get an appropriate upper bound for $\mathbb{E}\left[\mathbf{U}_{z,v}(X) \mid \mathcal{F}\right]$. Using the fact that $\mathbf{U}_{z,w}(X) \leq 1$ and $\mathbf{Pr}\left[z \in X|\mathcal{F}\right] \leq \lambda$ we have that

$$\begin{aligned}&\mathbb{E}\left[\mathbf{U}_{z,v}(X) \mid \mathcal{F}\right] \\ &\quad= \mathbb{E}\left[\mathbf{U}_{z,v}(X) \mid \mathcal{F}, z \in X\ \right] \cdot \mathbf{Pr}\left[z \in X|\mathcal{F}\right] \\ &\qquad+\mathbb{E}\left[\mathbf{U}_{z,v}(X) \mid \mathcal{F}, z \notin X\right] \cdot \mathbf{Pr}\left[z \notin X\ |\mathcal{F}\right] \\ &\quad\leq \mathbf{Pr}\left[z \in X|\mathcal{F}\right] + \mathbb{E}\left[\mathbf{U}_{z,v}(X) \mid \mathcal{F}, z \notin X\right] \\ &\quad\leq \lambda + \mathbb{E}\left[\mathbf{U}_{z,v}(X) \mid \mathcal{F}, z \notin X\right] \\ &\quad= \lambda + \prod_{u \in N(z)\backslash\{v\}}\left(1 - \frac{\lambda}{1+\lambda}\mathbf{U}_{u,z}(X)\right) \qquad (23) \\ &\quad\leq 2\lambda + \prod_{u \in N(z)}\left(1 - \frac{\lambda}{1+\lambda}\mathbf{U}_{u,z}(X)\right) \\ &\quad= 2\lambda + \mathbf{R}(X, z), \qquad (24)\end{aligned}$$

where (23) uses the fact that given $\mathcal{F}$ the values of $\mathbf{U}_{u,z}(X)$, for $u \in N(z) \setminus \{v\}$ are fully determined. Similarly, we get the lower bound:

$$\begin{aligned}\mathbb{E}\left[\mathbf{U}_{z,v}(X) \mid \mathcal{F}\right] &= \mathbb{E}\left[\mathbf{U}_{z,v}(X) \mid \mathcal{F}, z \in X\right] \cdot \mathbf{Pr}\left[z \in X|\mathcal{F}\right] \\ &\quad+\mathbb{E}\left[\mathbf{U}_{z,v}(X) \mid \mathcal{F}, z \notin X\right] \cdot \mathbf{Pr}\left[z \notin X|\mathcal{F}\right] \\ &\geq (1 - 2\lambda)\mathbb{E}\left[\mathbf{U}_{z,v}(X) \mid \mathcal{F}, z \notin X\right] \\ &\geq (1 - 2\lambda)\prod_{u \in N(z)\backslash\{w\}}\left(1 - \frac{\lambda}{1+\lambda}\mathbf{U}_{u,z}(X)\right) \\ &\geq (1 - 2\lambda)\prod_{u \in N(z)}\left(1 - \frac{\lambda}{1+\lambda}\mathbf{U}_{u,z}(X)\right) \\ &= (1 - 2\lambda)\mathbf{R}(X, z) \\ &\geq \mathbf{R}(X, z) - 2\lambda,\end{aligned} \tag{25}$$

where in the last inequality we use the fact that $\mathbf{R}(X, z) \leq 1$.

From (24) and (25) we have proven (21), which completes the proof of the lemma. ■

V. Sketch of Rapid Mixing Proof

Theorem 8 tells us that after a burn-in period the Glauber dynamics locally behaves like the BP fixpoints ω^* with high probability (whp). (In this discussion, we use the term whp to refer to events that occur with probability $\geq 1-\exp(-\Omega(\Delta))$.) Meanwhile Theorem 6 says that there is an appropriate distance function Φ for which path coupling has contraction for pairs of states that behave as in ω^*. The snag in simply combining this pair of results and deducing rapid mixing is that when Δ is constant then there is still a constant fraction of the graph that does not behave like ω^*, and our disagreements in our coupling proof may be biased towards this set. We follow the approach in [4] to overcome this obstacle and complete the proof of Theorem 1. We give a brief sketch of the approach, the details are contained in Section E of the full version.

The burn-in period for Theorem 8 to apply is $O(n \log \Delta)$ steps from the worst-case initial configuration X_0. In fact, for a "typical" initial configuration only $O(n)$ steps are required as we only need to update $\geq 1 - \epsilon$ fraction of the neighbors of every vertex in the local neighborhood of the specified vertex v. The "bad" initial configurations are ones where almost all of the neighbors of v (or many of its grandchildren) are occupied. We call such configurations "heavy" (see Section C.2 of the full version for details). We first prove that after $O(n \log \Delta)$ steps a chain is not-heavy in the local neighborhood of v, and this property persists whp (see Lemma 22 in the full version). Then, only $O(n)$ steps are required for the burn-in period (see Theorem 27 in Section D of the full version).

Our argument has two stages. We start with a pair of chains X_0, Y_0 that differ at a single vertex v. In the first stage we burn-in for $T_b = O(n \log \Delta)$ steps. After this burn-in period, we have the following properties whp: every vertex in the local neighborhood of v is not-heavy, the number of disagreements is $\leq \mathrm{poly}(\Delta)$, and the disagreements are all in the local neighborhood of v (see Lemma 31, parts 2 and 4, in Section E of the full version).

In the second stage we have sets of epochs of length $T = O(n)$ steps. For the pair of chains X_{T_b}, Y_{T_b} we apply path coupling again. Now we consider a pair of chains that differ at one vertex z which is not heavy. We look again at the local neighborhood of z (in this case, that means all vertices within distance $\leq \sqrt{\Delta}$ of z). After T steps, whp every vertex in the local neighborhood has the local uniformity properties and the disagreements are contained in this local neighborhood. Then we have contraction in the path coupling condition (by applying Theorem 6), and hence after $O(n)$ further steps the expected Hamming distance is small (see Lemma 32 in the full version). Combining a sequence of these $O(n)$ length epochs we get that the original pair has is likely to have coupled and we can deduce rapid mixing.

VI. Conclusions

The work of Weitz [41] was a notable accomplishment in the field of approximate counting/sampling. However a limitation of his approach is that the running time depends exponentially on $\log \Delta$. It is widely believed that the Glauber dynamics has mixing time $O(n \log n)$ for all G of maximum degree Δ when $\lambda < \lambda_c(\Delta)$. However, until now there was little theoretical work to support this conjecture. We give the first such results which analyze the widely used algorithmic approaches of MCMC and loopy BP.

One appealing feature of our work is that it directly ties together with Weitz's approach: Weitz uses decay of correlations on trees to truncate his self-avoiding walk tree, whereas we use decay of correlations to deduce a contracting metric for the path coupling analysis, at least when the chains are at the BP fixed point. We believe this technique of utilizing the principal eigenvector for the BP operator for the path coupling metric will apply to a general class of spin systems, such as 2-spin antiferromagnetic spin systems (Weitz's algorithm was extended to this class [20]).

We hope that in the future more refined analysis of the local uniformity properties will lead to relaxed girth assumptions. However dealing with very short cycles, such as triangles, will require a new approach since loopy BP no longer seems to be a good estimator of the Gibbs distribution for certain examples.

References

[1] R. Bubley and M. Dyer. Path coupling, Dobrushin uniqueness, and approximate counting. In *Proceedings of the 38th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, 223–231, 1997.

[2] V. Chandrasekaran, M. Chertkov, D. Gamarnik, D. Shah, and J. Shin. Counting Independent Sets using the Bethe Approximation. *SIAM J. Discrete Math.*, 25(2):1012-1034, 2011.

[3] D. Dubhashi and D. Ranjan. Balls and bins: a study in negative dependence. *Random Structures and Algorithms*, 13:99-124, 1998.

[4] M. E. Dyer, A. M. Frieze, T. P. Hayes, and E. Vigoda. Randomly Coloring Constant Degree Graphs. *Random Structures and Algorithms*, 43(2):181-200, 2013.

[5] M. Dyer and C. Greenhill. On Markov chains for independent sets. *J. Algorithms* 35(1):17–49, 2000.

[6] C. Efthymiou, T. P. Hayes, D. Štefankovič, E. Vigoda and Y. Yin Convergence of MCMC and Loopy BP in the Tree Uniqueness Region for the Hard-Core Model. *Preprint 2016*, available from arxiv at http://arxiv.org/abs/1604.01422

[7] A. Galanis, Q. Ge, D. Štefankovič, E. Vigoda, and L. Yang. Improved inapproximability results for counting independent sets in the hard-core model. *Random Structures and Algorithms*, 45(1):78-110, 2014.

[8] A. Galanis, D. Štefankovič, and E. Vigoda. Inapproximability of the partition function for the antiferromagnetic Ising and hard-core models. *Combinatorics, Probability & Computing* 25(4): 500-559, 2016.

[9] A. Galanis, D. Štefankovič, and E. Vigoda. Inapproximability for Antiferromagnetic Spin Systems in the Tree Non-Uniqueness Region. *Journal of the ACM*, 62(6):50, 2015.

[10] C. Greenhill. The complexity of counting colourings and independent sets in sparse graphs and hypergraphs. *Computational Complexity*, 9(1):52–72, 2000.

[11] T. P. Hayes. Local uniformity properties for Glauber dynamics on graph colorings. *Random Structures and Algorithms*, 43(2):139-180, 2013.

[12] T. P. Hayes. A simple condition implying rapid mixing of single-site dynamics on spin systems. In *47th Annual IEEE Symposium on Foundations of Computer Science* (FOCS), pages 39-46, 2006.

[13] T. P. Hayes, J. C. Vera, and E. Vigoda. Randomly coloring planar graphs with fewer colors than the maximum degree. *Random Struct. Algorithms* 47(4): 731-759, 2015

[14] T. Hayes and E. Vigoda. Coupling with the Stationary Distribution and Improved Sampling for Colorings and Independent Sets. *Annals of Applied Probability*, 16(4):1297-1318, 2006.

[15] T. Heskes. On the uniqueness of loopy belief propagation fixed points. *Neural Computation*, 16(11):2379-2413, 2004.

[16] M. Huber. Approximation algorithms for the normalizing constant of Gibbs distributions. *Annals of Applied Probability*, 25(2):974-985, 2015.

[17] A. T. Ihler, J. W. Fisher III, and A. S. Willsky. Loopy Belief Propagation: Convergence and Effects of Message Errors. *Journal of Machine Learning Research*, 6:905-936, 2005.

[18] F. P. Kelly. Stochastic Models of Computer Communication Systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, 47(3):379-395, 1985.

[19] L. Li, P. Lu, and Y. Yin. Approximate counting via correlation decay in spin systems. In *Proceedings of the 23rd Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 922-940, 2012.

[20] L. Li, P. Lu, and Y. Yin. Correlation Decay up to Uniqueness in Spin Systems. In *Proceedings of the 24th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 67-84, 2013.

[21] C. Lin, J. Liu, and P. Lu. A Simple FPTAS for Counting Edge Covers. In *Proceedings of the 25th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 341-348, 2014.

[22] J. Liu and P. Lu. FPTAS for #BIS with Degree Bounds on One Side. In *Proceedings of the 47th Annual ACM Symposium on Theory of Computing (STOC)*, pages 549-556, 2015.

[23] M. Luby and E. Vigoda. Fast Convergence of the Glauber Dynamics for Sampling Independent Sets. *Random Structures and Algorithms*, 15(3-4):229-241, 1999.

[24] M. Mitzenmacher and E. Upfal. Probability and computing: Randomized algorithms and probabilistic analysis, Cambridge University Press, 2005.

[25] E. Mossel, D. Weitz, and N. Wormald. On the hardness of sampling independent sets beyond the tree threshold. *Probability Theory and Related Fields*, 143(3-4):401-439, 2009.

[26] K. P. Murphy, Y. Weiss, and M. I. Jordan. Loopy belief propagation for approximate inference: an empirical study. In *Proceedings of the 15th Conference on Uncertainty in Artificial Intelligence (UAI)*, pages 467-475, 1999.

[27] J. Pearl. Reverend Bayes on inference engines: A distributed hierarchical approach. In *Proceedings of the 2nd National Conference on Artificial Intelligence* (AAAI), pages 133-136, 1982.

[28] R. Restrepo, J. Shin, P. Tetali, E. Vigoda, and L. Yang. Improved Mixing Condition on the Grid for Counting and Sampling Independent Sets. *Probability Theory and Related Fields*, 156(1-2):75-99, 2013.

[29] J. Shin. The Complexity of Approximating a Bethe Equilibrium. *IEEE Transactions on Information Theory*, 60(7): 3959-3969, 2014.

[30] A. Sinclair, P. Srivastava, D. Štefankovič, and Y. Yin. Spatial mixing and the connective constant: Optimal bounds. To appear in *Probability Theory & Related Fields.*

[31] A. Sinclair, P. Srivastava, and M. Thurley. Approximation algorithms for two-state anti-ferromagnetic spin systems on bounded degree graphs. *J. Stat. Phys.* 155 (4), pp. 666?686, 2014.

[32] A. Sinclair, P. Srivastava, and Y. Yin. Spatial Mixing and Approximation Algorithms for Graphs with Bounded Connective Constant. In *Proceedings of the 54th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 300-309, 2013.

[33] A. Sly. Computational Transition at the Uniqueness Threshold. In *Proceedings of the 51st Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 287-296, 2010.

[34] A. Sly and N. Sun. The computational hardness of counting in two-spin models on d-regular graphs. *The Annals of Probability*, 42(6):2383-2416, 2014.

[35] D. Štefankovič, S. Vempala, and E. Vigoda. Adaptive Simulated Annealing: A Near-optimal Connection between Sampling and Counting. *Journal of the ACM*, 56(3):1-36, 2009.

[36] S. Tatikonda and M. Jordan. Loopy belief propagation and Gibbs measures. In *Proceedings of the 18th Conference on Uncertainty in Artificial Intelligence* (UAI), pages 493-500, 2002.

[37] L. G. Valiant. The Complexity of Enumeration and Reliability Problems. *SIAM Journal on Computing*, 8(3):410-421, 1979.

[38] J. C. Vera, E. Vigoda, and L. Yang. Improved Bounds on the Phase Transition for the Hard-Core Model in 2-Dimensions. *SIAM Journal on Discrete Mathematics*, 29(4):1895-1915, 2015.

[39] E. Vigoda. A Note on the Glauber Dynamics for Sampling Independent Sets. *Electronic Journal of Combinatorics*, 8(1): Research paper 8, 2001.

[40] Y. Weiss. Correctness of Local Probability Propagation in Graphical Models with Loops. *Neural Computation*, 12(1):1-41, 2000.

[41] D. Weitz. Counting independent sets up to the tree threshold. In *Proceedings of the 38th Annual ACM Symposium on Theory of Computing (STOC)*, pages 140-149, 2006.

[42] N. C. Wormald. The asymptotic distribution of short cycles in random regular graphs. *J. Combinatorial Theory (Series B)*, 31:168-182, 1981.