On the Mixing Time of Glauber Dynamics for the Hard-Core and Related Models on $G(n, d/n)$

Charilaos Efthymiou
Computer Science, University of Warwick, Coventry, UK

Weiming Feng
School of Informatics, University of Edinburgh, Edinburgh, UK

Abstract

We study the single-site Glauber dynamics for the fugacity λ, Hard-Core model on the random graph $G(n, d/n)$. We show that for the typical instances of the random graph $G(n, d/n)$ and for fugacity $\lambda < \frac{d^d}{(d-1)^{d+1}}$, the mixing time of Glauber dynamics is $n^{1+O(1/\log\log n)}$.

Our result improves on the recent elegant algorithm in [Bezáková, Galanis, Goldberg and Štefankovič; ICALP'22]. The algorithm there is an MCMC-based sampling algorithm, but it is not the Glauber dynamics. Our algorithm here is *simpler*, as we use the classic Glauber dynamics. Furthermore, the bounds on mixing time we prove are smaller than those in Bezáková et al. paper, hence our algorithm is also *faster*.

The main challenge in our proof is handling vertices with unbounded degrees. We provide stronger results with regard the spectral independence via branching values and show that the our Gibbs distributions satisfy the approximate tensorisation of the entropy. We conjecture that the bounds we have here are optimal for $G(n, d/n)$.

As corollary of our analysis for the Hard-Core model, we also get bounds on the mixing time of the Glauber dynamics for the Monomer-Dimer model on $G(n, d/n)$. The bounds we get for this model are slightly better than those we have for the Hard-Core model

2012 ACM Subject Classification Theory of computation → Design and analysis of algorithms; Theory of computation → Randomness, geometry and discrete structures; Mathematics of computing → Discrete mathematics

Keywords and phrases spin-system, spin-glass, sparse random (hyper)graph, approximate sampling, efficient algorithm

Digital Object Identifier 10.4230/LIPIcs.ICALP.2023.54

Category Track A: Algorithms, Complexity and Games

Related Version *Full Version*: `https://arxiv.org/abs/2302.06172`

Funding *Charilaos Efthymiou*: EPSRC New Investigator Award (grant no. EP/V050842/1) and Centre of Discrete Mathematics and Applications (DIMAP), The University of Warwick.
Weiming Feng: European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 947778).

Acknowledgements Weiming Feng would like to thank Heng Guo for the helpful discussions.

1 Introduction

The Hard-Core model and the related problem of the geometry of independent sets on the sparse random graph $G(n, d/n)$ is a fundamental area of study in discrete mathematics [17, 11], in computer science they are studied in the context of the random *Constraint Satisfaction Problems* [10, 20], while in statistical physics they are studied as instances of *disordered systems*. Using the so-called *Cavity method* [25, 2], physicists make some impressive

predictions about the independent sets of $G(n, d/n)$, such as higher order replica symmetry breaking etc. Physicists' predictions are (typically) mathematically non-rigorous. Most of these predictions about independent sets still remain open as basic natural objects in the study such as the partition function, or the free energy are extremely challenging to analyse.

The Hard-Core model with fugacity $\lambda > 0$, is a distribution over the *independent sets* of an underlying graph G such that every independent set σ is assigned probability measure $\mu(\sigma)$ which is proportional to $\lambda^{|\sigma|}$, where $|\sigma|$ is the cardinality of σ. Here, we consider the case where the underlying graph is a typical instance of the sparse random graph $G(n, d/n)$. This is the random graph on n vertices, while each edge appears independently with probability $p = d/n$. The quantity $d > 0$ corresponds to the *expected degree*. For us here the expected degree is a bounded constant, i.e., we have $d = \Theta(1)$, hence the graph is sparse.

Our focus is on approximate sampling from the aforementioned distribution using *Glauber dynamics*. This is a classic, very popular, algorithm for approximate sampling. The popularity of this process, mainly, is due to its simplicity and the strong approximation guarantees that provides. The efficiency of Glauber dynamics for sampling is studied by means of the *mixing time*.

Recently, there has been an "explosion" of results about the mixing time of Glauber dynamics for *worst-case* instances the problem, e.g. [1, 9, 8, 12]. Combined with the earlier hardness results in [29, 30, 19] one could claim that for worst-case instances the behaviour of Glauber dynamics for the Hard-Core model, but also the related approximate sampling-counting problem, is well understood. Specifically, for the graphs of maximum degree Δ, Glauber dynamics exhibits $O(n \log n)$ mixing time for any fugacity $\lambda < (\Delta-1)^{\Delta-1}/(\Delta-2)^{\Delta}$, while the hardness results support that this region of λ is best possible.

The aforementioned upper bound on λ coincides with the *critical point* for the uniqueness/non-uniqueness phase transition of the Hard-Core model on the infinite Δ-regular tree [24]. At this point in the discussion, perhaps, it is important to note the dependency of the critical point on the *maximum degree*. This is the point where the situation with the random graph $G(n, d/n)$ differentiates from the worst case one.

For $G(n, d/n)$ and for the range of the expected degree d we consider here, typically, almost all of the vertices in the graph, e.g., say 99%, are of degree very close to d. On the other hand, the maximum degree of $G(n, d/n)$ is as large as $\Theta(\frac{\log n}{\log \log n})$, i.e., it is *unbounded*. In light of this observation, it is natural to expect that the Glauber dynamics on the Hard-Core model mixes fast for values of the fugacity that depend on the *expected degree*, rather the maximum degree. Note that, this implies to use Glauber dynamics to sample from the Hard-Core model with fugacity λ taking *much larger* values than what the worst-case bound implies.

For $d > 1$, let $\lambda_c(d) = \frac{d^d}{(d-1)^{(d+1)}}$. One of the main result in our paper is as follows: we show that for any $d > 1$ and for typical instances of $G(n, d/n)$, the Glauber dynamics on the Hard-Core with any fugacity $\lambda < \lambda_c(d)$, exhibits mixing time which is $n^{1+\frac{C}{\log \log n}} = n^{1+o(1)}$, for some absolute constant $C > 0$ which depends only on λ and d.

It is our *conjecture* that the bound on the mixing time for the hard-core is tight. Furthermore, following intuitions from [10], as well as from *statistical physics* predictions in [2], it is our *conjecture* that the bound $\lambda_c(d)$ on the fugacity λ is also tight, in the following sense: for $\lambda > \lambda_c(d)$ it is not precluded that there is a region where efficient approximate sampling is possible, however, the approximation guarantees are *weaker* than those we have here.

Our result improves on the elegant sampling algorithm that was proposed recently in [3] for the same distribution, i.e., the Hard-Core model on $G(n, d/n)$. That algorithm, similarly to the one we consider here, relies on the Markov Chain Monte Carlo method. The authors

use Spectral Independence [1, 9] to show that the underlying Markov chain exhibits mixing time which is $O\left(n^{1+\theta}\right)$ for any $\lambda < \lambda_c(d)$ and arbitrary small consant $\theta > 0$. The idea that underlies the algorithm in [3] is reminiscent of the *variable marking* technique that was introduced in [26] for approximate counting with the Lovász Local Lemma, and was further exploited in [14, 16, 21, 18]. Here, we use a different, more straightforward, approach and analyse directly the Glauber dynamics.

Note that both algorithms, i.e., here and in [3], allow for the same range for the fugacity λ. On the other hand, the algorithm we study here is the (much simpler) Glauber dynamics, while the running time guarantees we obtain here are asymptotically better.

Previous works in the area, i.e., even before [3], in order to prove their results and avoid the use of maximum degree, have been focusing on various parameters of $G(n, d/n)$ such as the expected degree [13], or the connective constant [28]. Which, as it turns out are not that different with each other. Here, we utilise the notion of *branching value*, which is somehow related to the previous ones.

The notion of the branching value as well as its use for establishing Spectral Independence was introduced in [3]. Unfortunately, the result there were not sufficiently strong to imply rapid mixing of Glauber dynamics. Their analytic tools for Spectral Independence (and others) seems to not be able to handle all that well vertices with unbounded degree. Here we derive stronger results for Spectral independence than those in [3] in the sense that they are more *general* and more *accurate*. Specifically, in our analysis we are able to accommodate vertices of *all degrees*, while we use a more elaborate matrix norm to establish spectral independence, reminiscent of those introduced in [12]. Furthermore, we utilise results from [8] that allow us deal with the unbounded degrees of the graph in order to establish our rapid mixing results.

2 Results

Consider the fixed graph $G = (V, E)$ on n vertices. Given the parameter $\lambda > 0$, which we call *fugacity*, we define the *Hard-Core* model $\mu = \mu_{G,\lambda}$ to be a distribution on the independent sets of the graph G, Specifically, every independent set σ is assigned probability measure $\mu(\sigma)$ defined by

$$\mu(\sigma) \propto \lambda^{|\sigma|} \ , \tag{1}$$

where $|\sigma|$ is equal to the size of the independent set σ.

We use $\{\pm 1\}^V$ to encode the configurations of the Hard-Core model, i.e., the independent sets of G. Particularly, the assignment $+1$ implies that the vertex is in the independent set, while -1 implies the opposite. We often use physics' terminology where the vertices with assignment $+1$ are called "occupied", whereas the vertices with -1 are "unoccupied".

We use the discrete time, (single site) *Glauber dynamics* to approximately sample from the aforementioned distributions. Glauber dynamics is a Markov chain with state space the support of the distribution μ. Typically, we assume that the chain starts from an arbitrary configuration $X_0 \in \{\pm 1\}^V$. For $t \geq 0$, the transition from the state X_t to X_{t+1} is according to the following steps:

1. Choose uniformly at random a vertex v.
2. For every vertex w different than v, set $X_{t+1}(w) = X_t(w)$.
3. Set $X_{t+1}(v)$ according to the marginal of μ at v, conditional on the neighbours of v having the configuration specified by X_{t+1}.

It is standard that when a Markov chain satisfies a set of technical conditions called *ergodicity*, then it converges to a unique stationary distribution. For the cases we consider here, Glauber dynamics is trivially ergodic, while the stationary distribution is the corresponding Hard-Core model μ.

Let P be the transition matrix of an ergodic Markov chain $\{X_t\}$ with a finite state space Ω and equilibrium distribution μ. For $t \geq 0$ and $\sigma \in \Omega$, let $P^t(\sigma, \cdot)$ denote the distribution of X_t when the initial state of the chain satisfies $X_0 = \sigma$. The *mixing time* of the Markov chain $\{X_t\}_{t\geq 0}$ is defined by

$$T_{\text{mix}} = \max_{\sigma\in\Omega} \min\left\{t > 0 \mid \|P^t(\sigma,\cdot) - \mu\|_{\text{TV}} \leq \frac{1}{2\text{e}}\right\} \ .$$

Our focus is on the mixing time of Glauber dynamics for the Hard-Core model for the case where the underlying graph is a typical instance of $G(n, d/n)$, where the expected degree $d > 0$ is a assumed to be a fixed number.

2.1 Mixing Time for Hard-Core Model

For $z > 1$, we let the function $\lambda_c(z) = \frac{z^z}{(z-1)^{(z+1)}}$. It is a well-known result from [24] that the uniqueness region of the Hard-Core model on the k-ary tree, where $k \geq 2$, holds for any λ such that

$$\lambda < \lambda_c(k) \ .$$

The following theorem is the main result of this work.

▶ **Theorem 1.** *For fixed $d > 1$ and any $\lambda < \lambda_c(d)$, there is a constant $C > 0$ such that the following is true:*

Let $\mu_{\boldsymbol{G}}$ be the Hard-Core model with fugacity λ on the graph $\boldsymbol{G} \sim G(n, d/n)$. With probability $1 - o(1)$ over the instances of $\boldsymbol{G}$, Glauber dynamics on $\mu_{\boldsymbol{G}}$ exhibits mixing time

$$T_{\text{mix}} \leq n^{\left(1+\frac{C}{\log\log n}\right)} \ .$$

2.2 Extensions to Monomer-Dimer Model

Utilising the techniques we develop in order to prove Theorem 1, we get mixing time bounds for the Glauber dynamics on the Monomer-Dimer model on $G(n, d/n)$.

Given a fixed graph $G = (V, E)$ and a parameter $\lambda > 0$, which we call *edge weight*, we define the Monomer-Dimer model $\mu = \mu_{G,\lambda}$ to be a distribution on the *matchings* of the graph G such that every matching σ is assigned probability measure $\mu(\sigma)$ defined by

$$\mu(\sigma) \propto \lambda^{|\sigma|} \ , \tag{2}$$

where $|\sigma|$ is equal to the number of edges in the matching σ.

Note that the Hard-Core model considers configurations on the vertices of G, while the Monomer-Dimer model considers configurations on the edges. Similarly to the independent sets, we use $\{\pm 1\}^E$ to encode the matchings of G. Specifically, the assignment $+1$ on the edge e implies that the edge is in matching, while -1 implies the opposite.

For the Monomer-Dimer model the definition of Glauber dynamics $\{X_t\}_{t\geq 0}$ extends in the natural way. That is, assume that the chain starts from an arbitrary configuration $X_0 \in]\pm 1^E$. For $t \geq 0$, the transition from the state X_t to X_{t+1} is according to the following steps:

1. Choose uniformly at random an edge e.
2. For every edge f different than e, set $X_{t+1}(f) = X_t(f)$.
3. Set $X_{t+1}(e)$ according to the marginal of μ at e, conditional on the neighbours of e having the configuration specified by X_{t+1}.

We consider the case of the Monomer-Dimer distribution where the underlying graph is an instance of $G(n, d/n)$. We prove the following result.

▶ **Theorem 2.** *For fixed $d > 1$ and any $\lambda > 0$, there is a constant $C > 0$ such that the following is true:*

Let $\mu_{\boldsymbol{G}}$ be the Monomer-Dimer model with edge weight λ on the graph $\boldsymbol{G} \sim G(n, d/n)$. With probability $1 - o(1)$ over the instances of $\boldsymbol{G}$, Glauber dynamics on $\mu_{\boldsymbol{G}}$ exhibits mixing time

$$T_{\text{mix}} \leq n^{\left(1+C\sqrt{\frac{\log\log n}{\log n}}\right)} .$$

The proof of Theorem 2 can be found in the full version of this paper.

For the Monomer-Dimer model on general graphs, the best-known result is the $\tilde{O}(n^2 m)$ mixing time of the Jerrum-Sinclair chain [23], where $m = |E|$ is the number of edges. For graphs with bounded maximum degree $\Delta = O(1)$, the spectral independence technique proved the $O(n \log n)$ mixing time of Glauber dynamics [9]. However, this result cannot be applied directly to the random graph $G(n, d/n)$, because the maximum degree of a random graph is typically unbounded. For the Monomer-Dimer model on $G(n, d/n)$, [3] gave a sampling algorithm with running time $n^{1+\theta}$, where $\theta > 0$ is an arbitrarily small constant, and [22] also proved the $n^{2+o(1)}$ mixing time of Glauber dynamics in a special case $\lambda = 1$. Our result in Theorem 2 proves the $n^{1+o(1)}$ mixing time of Glauber dynamics, which improves all the previous results for the Monomer-Dimer model on the random graph $G(n, d/n)$ with constant λ. It is an open problem to improve the mixing time in Theorem 2. Moreover, for general graphs, the tight mixing time of Glauber dynamics for the Monomer-Dimer model is also a challenging open problem.

We remark that for the Monomer-Dimer model, we actually proved the $n^{1+o(1)}$ mixing time of Glauber dynamics on *all* graphs satisfying $\Delta \log^2 \Delta = o(\log^2 n)$. See the full version of this paper for a more general result.

This version of the paper focuses on the Hard-Core model, i.e., proving Theorem 1. The proofs for the Monomer-Dimer model is in the full version.

Notation

Suppose that we are given a Gibbs distribution μ on the graph $G = (V, E)$. We denote with Ω the support of μ.

Suppose that Ω is a set of configurations at the vertices of G. Then, for any $\Lambda \subseteq V$ and any $\tau \in \{\pm 1\}^\Lambda$, we let $\mu^{\Lambda,\tau}$ (or μ^τ if Λ is clear from the context) denote the distribution μ conditional on that the configuration at Λ is τ. Alternatively, we use the notation $\mu(\cdot \mid (\Lambda, \tau))$ for the same conditional distribution. We let $\Omega^\tau \subseteq \Omega$ be the support of $\mu^{\Lambda,\tau}$. We call τ *feasible* if Ω^τ is nonempty.

For any subset $S \subseteq V$, let μ_S denote the marginal of μ at S, while let Ω_S denote the support of μ_S. In a natural way, we define the conditional marginal. That is, for $\Lambda \subseteq V \setminus S$ and $\sigma \in \{\pm 1\}^\Lambda$, we let $\mu_S^{\Lambda,\sigma}$ (or μ_S^σ if Λ is clear from the context) denote the marginal at S conditional on the configuration at Λ being σ. Alternatively we use $\mu_S(\cdot \mid (\Lambda, \sigma))$ for μ_S^σ. We let Ω_S^σ denote the support of μ_S^σ.

All the above notation for configurations on the vertices of G can be extended naturally for configurations on the edges of the graph G. We omit presenting it, because it is very similar to the above.

2.3 Hard-Core Model – Entropy Tensorisation for Rapid Mixing

We prove Theorem 1 by exploiting the notion of *approximate tensorisation of the entropy*.

Let μ be a distribution with support $\Omega \subseteq \{\pm 1\}^V$. For any function $f : \Omega \to \mathbb{R}_{\geq 0}$, we let $\mu(f) = \sum_{x\in\Omega} \mu(x)f(x)$, i.e., $\mu(f)$ is the expected value of f with respect to μ. Define the entropy of f with respect to μ by

$$\mathrm{Ent}_\mu(f) = \mu\left(f \log \tfrac{f}{\mu(f)}\right) \ ,$$

where we use the convention that $0 \log 0 = 0$.

Let $\tau \in \Omega_{V\setminus S}$ for some $S \subset V$. Define the function $f_\tau : \Omega_S^\tau \to \mathbb{R}_{\geq 0}$ by having $f_\tau(\sigma) = f(\tau \cup \sigma)$ for all $\sigma \in \Omega_S^\tau$ [1]. Let $\mathrm{Ent}_S^\tau(f_\tau)$ denote the entropy of f_τ with respect to the conditional distribution μ_S^τ. Furthermore, we let

$$\mu(\mathrm{Ent}_S(f)) = \sum_{\tau\in\Omega_{V\setminus S}} \mu_{V\setminus S}(\tau)\mathrm{Ent}_S^\tau(f_\tau) \ ,$$

i.e., $\mu(\mathrm{Ent}_S(f))$ is the average of the entropy $\mathrm{Ent}_S^\tau(f_\tau)$ with respect to the measure $\mu_{V\setminus S}(\cdot)$. When $S = \{v\}$, i.e., the set S is a singleton, we abbreviate $\mu(\mathrm{Ent}_{\{v\}}(f))$ to $\mu(\mathrm{Ent}_v(f))$.

▶ **Definition 3** (Approximate Tensorisation of Entropy). *A distribution μ with support $\Omega \subseteq \{\pm 1\}^V$ satisfies the approximate tensorisation of entropy with constant $C > 0$ if for all $f : \Omega \to \mathbb{R}_{\geq 0}$ we have that*

$$\mathrm{Ent}_\mu(f) \leq C \cdot \sum_{v\in V} \mu\left(\mathrm{Ent}_v(f)\right) \ .$$

On can establish bounds on the mixing time of Glauber dynamics by means of the approximate tensorisation of entropy of the equilibrium distribution μ. Specifically, if μ satisfies the approximate tensorisation of entropy with constant C, then after every transition of Glauber dynamics, the Kullback–Leibler divergence[2] between the current distribution and the stationary distribution decays by a factor which is at least $(1 - C/n)$, where $n = |V|$ is the number of variables.

As far as the mixing time of Glauber dynamics is concerned, if a distribution μ satisfies the approximate tensorisation of entropy with parameter C then we have following well known relation (e.g. see [9, Fact 3.5]),

$$T_{\mathrm{mix}} \leq \left\lceil Cn\left(\log\log\frac{1}{\mu_{\min}} + \log(2) + 2\right)\right\rceil, \qquad \text{where } \mu_{\min} = \min_{x\in\Omega}\mu(x) \ . \tag{3}$$

In light of the above, Theorem 1 follows as a corollary from the following result.

---

[1] With a slight abuse of notation we use $\tau \cup \sigma$ to indicate the configuration what agrees with τ at S and with σ at $V \setminus S$.

[2] For discrete probability distributions P and Q on a discrete space $\mathcal{X}$, the Kullback–Leibler divergence is defined by $D_{\mathrm{KL}}(P||Q) = \sum_{x\in\mathcal{X}} P(x) \log \frac{P(x)}{Q(x)}$.

▶ **Theorem 4** (Hard-Core Model Tensorisation). *For any fixed $d > 1$ and any $\lambda < \lambda_c(d)$, there is a constant $A > 0$ that depends only on d and λ such that the following is true:*

Let $\mu_{\boldsymbol{G}}$ be the Hard-Core model with fugacity λ on the graph $\boldsymbol{G} \sim G(n, d/n)$. With probability $1 - o(1)$ over the instances of $\boldsymbol{G}$, $\mu_{\boldsymbol{G}}$ satisfies the approximate tensorisation of entropy with parameter $n^{A/\log\log n}$.

Proof of Theorem 1. Theorem 1 follows from Theorem 4 and (3).

Specifically, plugging the result from Theorem 4 into (3) we get the following: with probability $1 - o(1)$ over the instances of $\boldsymbol{G}$ we have that

$$\begin{aligned} T_{\text{mix}} &\leq n^{1+\frac{A}{\log\log n}} \left(\log\log \frac{1}{\mu_{\min}} + \log(2) + 2 \right) \\ &\leq n^{1+\frac{A}{\log\log n}} \left(\log\log \left(1 + \lambda + \lambda^{-1}\right)^n + \log(2) + 2 \right) \\ &= n^{1+\frac{A}{\log\log n}} \left(\log n + \log\log(1 + \lambda + \lambda^{-1}) \right) \leq n^{1+\frac{2A}{\log\log n}} \quad . \end{aligned}$$

For the second derivation, we note that for the Hard-Core distribution $\mu = \mu_{\boldsymbol{G}}$, we have that $\mu_{\min}$ is at least $\min\{1, \lambda^n\}/(1+\lambda)^n$, which implies that $\mu_{\min} \geq (1 + \lambda + \lambda^{-1})^{-n}$.

Note that Theorem 1 follows from the above, by setting $C = 2A$. ◀

3 Our Approach & Contributions

In this section we describe our approach towards establishing our results. Our focus is on the Hard-Core model.

3.1 Tensorisation and Block-Factorisation of Entropy

We establish the tensorisation of the entropy, described in Theorem 4, by exploiting the recently introduced notion of *block factorisation of entropy* in [5]. Specifically, we build on the framework introduced in [9] to relate the tensorisation and the block factorisation of the entropy.

The framework in [9] relies on the assumption that the maximum degree of the underlying graph is bounded. Otherwise, the results it implies are not strong. In our setting here, a vanilla application of this approach would not be sufficient to give the desirable bounds on the tensorisation constant due to the fact that the typical instances of $G(n, d/n)$ have unbounded maximum degree. To this end, we employ techniques from [8].

Given the graph $G = (V, E)$, and the integer $\ell \geq 0$, we let $\binom{V}{\ell}$ denote all subsets $S \subseteq V$ with $|S| = \ell$.

▶ **Definition 5** (ℓ-block Factorisation of Entropy). *Let μ be a distribution over $\{\pm 1\}^V$ and $1 \leq \ell \leq |V| = n$ be an integer. The distribution μ satisfies the ℓ block factorisation of entropy with parameter C if for all $f : \Omega \to \mathbb{R}_{\geq 0}$ we have that*

$$\text{Ent}_\mu(f) \leq \frac{C}{\binom{n}{\ell}} \sum_{S \in \binom{V}{\ell}} \mu\left(\text{Ent}_S(f)\right) \quad . \tag{4}$$

The notion of the ℓ block factorisation of entropy generalises that of the approximate tensorisation of entropy. Specifically, a distribution that satisfies the $\ell = 1$ block factorisation of entropy with parameter C, also satisfies the approximate tensorisation of entropy with parameter C/n.

As far as the Hard-Core model on $G(n, d/n)$ is concerned, we show the following theorem via the spectral independence technique, which is one of the main technical results in our paper.

▶ **Theorem 6.** *For fixed $d > 1$ and any $0 < \lambda < \lambda_c(d)$, consider $\boldsymbol{G} \sim G(n, d/n)$ and let $\mu_{\boldsymbol{G}}$ be the Hard-Core model on $\boldsymbol{G}$ with fugacity λ. With probability $1 - o(1)$ over the instances of $\boldsymbol{G}$ the following is true: There is a constant $K = K(d, \lambda) > 0$, such that for*

$$\frac{1}{\alpha} = K \frac{\log n}{\log \log n} \ ,$$

for any $1/\alpha \leq \ell < n$, $\mu_{\boldsymbol{G}}$ satisfies the ℓ-block factorisation of entropy with parameter $C = (\frac{en}{\ell})^{1+1/\alpha}$.

Let us have a high level overview of how we use the ℓ-block factorisation and particularly Theorem 6 to establish our entropy tensorisation result in Theorem 4.

Note that Theorem 6 essentially implies the following: Suppose that $G = (V, E)$ is a *typical instance* of $G(n, d/n)$. Then, the Hard-Core model μ on G, with fugacity $\lambda < \lambda_c(d)$, is such that for any $f : \Omega \to \mathbb{R}_{>0}$ we have

$$\mathrm{Ent}_{\mu}(f) \leq \left(\frac{\mathrm{e}}{\theta}\right)^{1+1/\alpha} \frac{1}{\binom{n}{\ell}} \sum_{S \in \binom{V}{\ell}} \mu\left(\mathrm{Ent}_S(f)\right) \ , \tag{5}$$

where $\ell = \lceil \theta n \rceil$ and $\theta \in (0, 1)$ is a constant satisfying $\lceil \theta n \rceil \geq 1/\alpha = \Omega(\log n / \log \log n)$.

Let $G[S]$ be the subgraph of G that is induced by the vertices in the set S. On the RHS of (5), the entropy is evaluated with respect to conditional distributions μ_S^τ, which is the Hard-Core model on the subgraph $G[S]$ given the boundary condition τ on $V \setminus S$.

We let $C(S)$ denote the set of connected components in $G[S]$. With a slight abuse of notation, we use $U \in C(S)$ to denote the set of vertices in the component U, as well. It is not hard to see that the Hard-Core model μ_S^τ, for $\tau \in \Omega_{V \setminus S}$, factorises as a product distribution over Gibbs marginals at the components $U \in C(S)$, i.e.,

$$\mu_S^\tau = \bigotimes_{U \in C(S)} \mu_U^\tau \ .$$

We use the following result for the factorisation of entropy on product distributions [6, 4, 9].

▶ **Lemma 7** ([9, Lemma 4.1])**.** *For any $S \subseteq V$, any $\tau \in \Omega_{V \setminus S}$, any $f : \Omega_S^\tau \to \mathbb{R}_{\geq 0}$,*

$$\mathrm{Ent}_S^\tau(f) \leq \sum_{U \in C(S)} \mu_S^\tau[\mathrm{Ent}_U(f)] \ .$$

Combining Lemma 7 and (5) we get that

$$\mathrm{Ent}_{\mu}(f) \leq \left(\frac{\mathrm{e}}{\theta}\right)^{1+1/\alpha} \mathbb{E}_{\boldsymbol{S} \sim \binom{V}{\ell}} \left[\sum_{U \in C(\boldsymbol{S})} \mu\left(\mathrm{Ent}_U(f)\right)\right] \ , \tag{6}$$

where $\boldsymbol{S} \sim \binom{V}{\ell}$ denotes that $\boldsymbol{S}$ is a uniformly random element from $\binom{V}{\ell}$.

The above step allows us to reduce the proof of approximate tensorisation to that of the components in $C(\boldsymbol{S})$. We choose the parameter $\ell = \lceil \theta n \rceil$ so that the connected components in $C(\boldsymbol{S})$ are typically small.

In light of the above, Theorem 4 follows by establishing two results: The first one is to derive a bound on the constant of the approximate tensorisation of entropy for the components of size k in $C(\boldsymbol{S})$, for each $k > 0$. The second result is to derive tail bounds on the size of the components in $C(\boldsymbol{S})$ for $\boldsymbol{S} \sim \binom{V}{\ell}$. Since the components are small with high probability, the following crude bound on the approximate tensorisation of entropy is enough for our analysis.

▶ **Lemma 8.** *For any fixed $d > 0$, for any $\lambda < \lambda_c(d)$, consider $\boldsymbol{G} \sim G(n, d/n)$. With probability $1 - o(1)$ over the instances of $\boldsymbol{G}$, the following is true:*

For any $k \geq 1$ and $H \subseteq V$ such that $|H| = k$, the Hard-Core model μ_H on $\boldsymbol{G}[H]$ with fugacity λ satisfies the approximate tensorisation of entropy with constant

$$\mathrm{AT}(k) \leq \min\left\{2k^2\left(1+\lambda+1/\lambda\right)^{2k+2},\ 3\log\left(1+\lambda+1/\lambda\right)\cdot\left((1+\lambda)k\right)^{2+2\eta}\right\}\ , \tag{7}$$

where $\eta = B(\log n)^{1/r}$, while $B = B(d,\lambda)$ and $r = r(d) \in (1,2)$ are constants that depend on d, λ.

As far as size of the components in $C(\boldsymbol{S})$ is concerned, we use the following result from [3].

▶ **Lemma 9** ([3])**.** *Let $d > 1$ be a constant. There is a constant $L = L(d)$ such that the following holds with probability at least $1 - o(1)$ over the $\boldsymbol{G} \sim G(n, d/n)$. Let $\boldsymbol{S} \sim \binom{V}{\ell}$, while let $C_v \subseteq \boldsymbol{S}$ be the set of vertices that are in the same component as vertex v in $\boldsymbol{G}[\boldsymbol{S}]$. For any integer $k \geq \log n$, it holds that*

$$\Pr[|C_v| = k] \leq (2\mathrm{e})^{\mathrm{e}Lk}\left(\frac{2\ell}{n}\right)^k \leq (2\mathrm{e})^{\mathrm{e}Lk}\,(2\theta)^k\ .$$

Theorem 4 follows by combining Theorem 6, with Lemmas 9 and 8. For a full proof of Theorem 4, see Section 5.

3.2 Spectral Independence with Branching Values

An important component in our proof of Theorem 6 is to establish Spectral Independence bounds for the Hard-Core model on typical instances of $G(n, d/n)$.

For worst-case graph instances (i.e., non random), typically, we establish Spectral Independence for a region of the parameters of the Gibbs distribution which is expressed in terms of the maximum degree Δ of the underlying graph G. As far as $G(n, d/n)$ is concerned, the maximum degree does not seem to be the appropriate graph parameter to consider for this problem.

Here, we utilise the notion of branching value. The notion of the branching value as well as its use for establishing Spectral Independence was introduced in [3]. Unfortunately, the result there were not sufficiently strong to imply rapid mixing of Glauber dynamics. Here we derive stronger results for Spectral independence than those in [3] in the sense that they are more *general* and more *accurate*. Specifically, in our analysis we are able to accommodate vertices of *all degrees*, while we use a more elaborate matrix norm to establish spectral independence, reminiscent of those introduced in [12]. Furthermore, we utilise results from [8] that allow us deal with the unbounded degrees of the graph in order to establish our rapid mixing results.

Before getting to further details in our discussion, let us first introduce some basic notions. We start with the *pairwise influence matrix* $\mathcal{I}_G^{\Lambda,\tau}$ and the related notion of Spectral Independence. These notions were first introduced in [1]. In this paper, we use the absolute version introduced in [15].

Consider a *fixed* graph $G = (V, E)$. Assume that we are given a Gibbs distribution μ on the configuration space $\{\pm 1\}^V$. We define the pairwise influence matrix $\mathcal{I}_G^{\Lambda,\tau}$ as follows: for a set of vertices $\Lambda \subset V$ and a configuration τ at Λ, the matrix $\mathcal{I}_G^{\Lambda,\tau}$ is indexed by the vertices in $V \setminus \Lambda$, while for any two vertices, different with each other $u, w \in V \setminus \Lambda$, if w can take both values ± 1 given τ, we have that

$$\mathcal{I}_G^{\Lambda,\tau}(w,u) = ||\mu_u(\cdot \mid (\Lambda,\tau),(\{w\},+)) - \mu_u(\cdot \mid (\Lambda,\tau),(\{w\},-))||_{\text{TV}} \ ; \tag{8}$$

if w can only take one value in ± 1 given τ, we have $\mathcal{I}_G^{\Lambda,\tau}(w,u) = 0$. Also, we have that $\mathcal{I}_G^{\Lambda,\tau}(w,w) = 0$ for all $w \in V \setminus \Lambda$. That is, the diagonal of $\mathcal{I}_G^{\Lambda,\tau}$ is always zero.

Recall that, above, $\mu_u(\cdot \mid (\Lambda,\tau),(\{w\},1))$ is the Gibbs marginal that vertex u, conditional that the configuration at Λ is τ and the configuration at w is 1. We have the analogous for $\mu_u(\cdot \mid (\Lambda,\tau),(\{w\},-1))$.

▶ **Definition 10** (Spectral Independence). *For a real number $\eta > 0$, the Gibbs distribution μ_G on $G = (V,E)$ is η-spectrally independent, if for every $0 \le k \le |V| - 2$, $\Lambda \subseteq V$ of size k and $\tau \in \{\pm 1\}^\Lambda$ the spectral radius of $\mathcal{I}_G^{\Lambda,\tau}$ satisfies that $\rho(\mathcal{I}_G^{\Lambda,\tau}) \le \eta$.*

We bound the spectral radius of $\mathcal{I}_G^{\Lambda,\tau}$ by means of matrix norms. Specifically, we use the following norm of $\mathcal{I}_G^{\Lambda,\tau}$

$$\left|\left| D^{-1} \cdot \mathcal{I}_G^{\Lambda,\tau} \cdot D \right|\right|_\infty , \tag{9}$$

where D is the diagonal matrix indexed by the vertices in $V \setminus \Lambda$ such that

$$D(u,u) = \begin{cases} \deg_G(v)^{1/\chi} & \text{if } \deg_G(v) \ge 1 \\ 1 & \text{if } \deg_G(v) = 0 \ , \end{cases} \tag{10}$$

where the parameter χ is being specified later.

Let $G = (V,E)$ be a fixed graph. For any vertex $v \in V$ and integer $\ell \ge 0$, we use $N_{v,\ell}$ to denote the number of simple paths with $\ell + 1$ vertices that start from v in graph G. By definition, we have that $N_{v,0} = 1$.

▶ **Definition 11** (d-branching value). *Let $d \ge 1$ be a real number and $G = (V,E)$ be a graph. For any vertex $v \in V$, the d-branching value S_v is defined by $\sum_{\ell \ge 0} N_{v,\ell}/d^\ell$.*

We establish spectral independence results that utilise the notion of d-branching value that was introduced in [3]. The following theorem is an example of the Spectral Independence results we derive here. In our proof, we actually use the stronger result in Theorem 19 . This analysis of spectral independence is of independent interest.

▶ **Theorem 12.** *Let $d > 1$ be a real number and $G = (V,E)$ be a graph. Let μ_G be the Hard-Core model with fugacity $\lambda < \lambda_c(d)$. For any $\alpha > 0$ such that the d-branching value $S_v \le \alpha$ for all $v \in V$ the following is true: μ_G is η-spectrally independent for*

$$\eta \le C_0 \cdot \alpha^{1/r} \ ,$$

where $C_0 = C_0(d,\lambda)$ and $r = r(d) \in (1,2)$ are constants.

There are a couple of interesting point about Theorem 12 to make. The first one is that the bound on η does not have any dependence on the degrees of the graph G. This is because we utilise the matrix norm $\|D^{-1} \cdot \mathcal{I}_G^{\Lambda,\tau} \cdot D\|_\infty$ instead of $\|\mathcal{I}_G^{\Lambda,\tau}\|_\infty$ that is typically used to establish the bound on the spectral independence. Furthermore, note that Theorem 12 is *not* necessarily about $G(n,d/n)$, i.e., it applies to an arbitrary graph. As a matter of fact in order to use the above result for $G(n,d/n)$ we need to establish bounds on its branching value. To this end, we use the following result from [3] so that we can take $\alpha = \log n$ in Theorem 12.

▶ **Lemma 13** ([3, Lemma 9]). *Let $d \geq 1$. For any fixed $d' > d$, with probability $1 - o(1)$ over $\boldsymbol{G} \sim G(n, d/n)$, the d'-branching factor of every vertex in G is at most $\log n$.*

It is worth mentioning that Lemma 13, here, is a *weaker* version of Lemma 9 in [3], i.e., we do not really need the the full strength of the result there.

In light of the above results, an interesting open problem is to turn the branching-value based spectral independence result in Theorem 12 into rapid mixing bound one for Glauber dynamics on a general graphs with bounded branching value. Note that this is not possible with the techniques we develop here.

Concluding this short introductory section about Spectral Independence, let us remark that for our results we work with the so-called *Complete Spectral Independence* for the Hard-Core model, introduced in [7, 8]. This is more general a notion compared to the (standard) Spectral Independence. For further discussion see Section 4.2.

4 Entropy Factorisation from Stability and Spectral Independence

In this section we establish the ℓ-block factorisation of entropy for the Hard-Core model on $G(n, d/n)$ as it is described in Theorem 6. To this end, we employ techniques from [8]. This means that we study the Hard-Core model on $G(n, d/n)$ in terms of the stability of ratios of the marginals and the so-called Complete Spectral Independence.

4.1 Ratios of Gibbs Marginals & Stability

Consider the *fixed* graph $G = (V, E)$ and a Gibbs distribution μ on this graph. For a vertex $w \in V$, the region $K \subseteq V \setminus \{w\}$ and $\tau \in \{\pm 1\}^K$, we consider the *ratio of marginals* at w denoted as $R^{K,\tau}(w)$ such that

$$R_G^{K,\tau}(w) = \frac{\mu_w(+1 \mid K, \tau)}{\mu_w(-1 \mid K, \tau)} \ . \tag{11}$$

Recall that $\mu_w(\cdot \mid K, \tau)$ denotes the marginal of the Gibbs distribution $\mu(\cdot \mid K, \tau)$ at vertex w. Also, note that the above allows for $R^{K,\tau}(w) = \infty$, e.g., when $\mu_w(-1 \mid K, \tau) = 0$ and $\mu_w(+1 \mid K, \tau) \neq 0$.

▶ **Definition 14** (Marginal stability). *Let $\zeta > 0$ be a real number. The Gibbs distribution μ_G on $G = (V, E)$ is called ζ-marginally stable if for any $\Lambda \subset V$, any $w \in V \setminus \Lambda$, for any configuration τ at Λ and any $S \subseteq \Lambda$ we have that*

$$R_G^{\Lambda,\tau}(w) \leq \zeta \qquad \text{and} \qquad R_G^{\Lambda,\tau}(w) \leq \zeta \cdot R_G^{S,\tau_S}(w) \ . \tag{12}$$

As far as the stability of the Hard-Core marginals at $G(n, d/n)$ is concerned, we prove the following result.

▶ **Theorem 15** (Stability Hard-Core Model). *For any fixed $d > 0$, for any $\lambda < \lambda_c(d)$, consider $\boldsymbol{G} \sim G(n, d/n)$ and let $\mu_{\boldsymbol{G}}$ be the Hard-Core model on $\boldsymbol{G}$ with fugacity λ. With probability $1 - o(1)$ over the instances $\boldsymbol{G}$, $\mu_{\boldsymbol{G}}$ is $2(1+\lambda)^{\frac{2\log n}{\log\log n}}$-marginally stable.*

Proof. Let $\zeta = 2(1+\lambda)^{2\frac{\log n}{\log\log n}}$. Also, let $N(w)$ be the set of the neighbours of w.

For any $\Lambda \subseteq V$ and any $\tau \in \{\pm\}^\Lambda$, we have that $\mu_w(+1 \mid \Lambda, \tau) \leq \frac{\lambda}{1+\lambda}$. One can see that the equality holds if $N(w) \subseteq \Lambda$ and for every $u \in N(w)$ we have that $\tau(w) = -1$. Noting that $R_{\boldsymbol{G}}^{\Lambda,\tau}(w)$ is increasing in the value of the Gibbs marginal $\mu_w(+1 \mid \Lambda, \tau)$, it is immediate that

$$\Pr\left[R_{\boldsymbol{G}}^{\Lambda,\tau}(w) \leq \lambda < \zeta \quad \forall \Lambda \subseteq V,\ \forall w \in V \setminus \Lambda\right] = 1 \ . \tag{13}$$

It remains to show that

$$\Pr\left[R_{\boldsymbol{G}}^{\Lambda,\tau}(w) \leq \zeta \cdot R_{\boldsymbol{G}}^{S,\tau_S}(w) \quad \forall \Lambda \subset V,\ \forall S \subset \Lambda,\ \forall w \in V \setminus \Lambda\right] = 1 - o(1) \ . \tag{14}$$

In light of (13), (14) follows by showing that

$$\Pr\left[R_{\boldsymbol{G}}^{S,\tau_S}(2) > 2\lambda\,(1+\lambda)^{-2\frac{\log n}{\log\log n}} \quad \forall \Lambda \subset V,\ \forall S \subset \Lambda,\ \forall w \in V \setminus \Lambda\right] = 1 - o(1) \ . \tag{15}$$

If there is $u \in N(w)$ such that $\tau(u) = +1$, then $R_{\boldsymbol{G}}^{\Lambda,\tau}(w) = 0$ and (14) holds trivially since $R_{\boldsymbol{G}}^{S,\tau_S}(w) \geq 0$. We focus on the case that all vertices $u \in N(w) \cap \Lambda$ satisfy $\tau(u) = -1$.

Let $\mathcal{E}$ be the event that none of the vertices in $N(w)$ is occupied, while let γ_S be the probability of the event $\mathcal{E}$ under the Gibbs distribution $\mu(\cdot \mid S, \tau_S)$. It is standard to show that

$$R_{\boldsymbol{G}}^{S,\tau_S}(w) = \frac{\frac{\lambda}{1+\lambda}\gamma_S}{1 - \frac{\lambda}{1+\lambda}\gamma_S} \ .$$

Noting that the function $f(x) = \frac{x}{1-x}$ is increasing in $x \in (0,1)$, while $\gamma_S \geq (\frac{1}{1+\lambda})^{\deg_{\boldsymbol{G}}(w)}$, we have that

$$R_{\boldsymbol{G}}^{S,\tau_S}(w) \geq \frac{\frac{\lambda}{1+\lambda}(\frac{1}{1+\lambda})^{\deg_{\boldsymbol{G}}(w)}}{1 - \frac{\lambda}{1+\lambda}(\frac{1}{1+\lambda})^{\deg_{\boldsymbol{G}}(w)}} = \frac{\lambda}{(1+\lambda)^{\deg_{\boldsymbol{G}}(w)+1} - \lambda} \ .$$

From the above, it is immediate to get (15). Specifically, it follows from the above inequality and a standard bound on the maximum degree of random graph which implies that for any fixed number $\epsilon > 0$, the maximum degree in $\boldsymbol{G}$ is less than $(1+\epsilon)\frac{\log n}{\log\log n}$ with probability $1 - o(1)$.

This concludes the proof of Theorem 15. ◀

4.2 (Complete) Spectral Independence

The notions of the pairwise influence matrix $\mathcal{I}_G^{\Lambda,\tau}$ and the Spectral Independence, as we introduce them in Section 3.2, are typically used to establish bounds on the spectral gap for Glauber dynamics and hence derive bounds on the mixing time of the chain.

The authors in [9], make a further use of Spectral Independence to obtain the approximate tensorisation of entropy. Unfortunately, a vanilla application of their technique is not sufficient to prove our tensorisation results, mainly, because of the unbounded degrees we typically have in $G(n, d/n)$.

In this work, we exploit ideas from [9] together with the related notion of the *Complete Spectral Independence*, in order to establish our factorisation results for the entropy in Theorem 6. Specifically, we utilise the connection between complete spectral independence and the ℓ block factorisation of entropy that was established in [8] (see further details in the following section).

Since the notions of the pairwise influence matrix $\mathcal{I}_G^{\Lambda,\tau}$ and the Spectral Independence are so important, let us recall them once more, even though they have already been defined in Section 3.2. Consider a *fixed* graph $G = (V, E)$. Assume that we are given a Gibbs distribution μ on the configuration space $\{\pm 1\}^V$.

We define the pairwise influence matrix $\mathcal{I}_G^{\Lambda,\tau}$ as follows: for a set of vertices $\Lambda \subset V$ and a configuration τ at Λ, the matrix $\mathcal{I}_G^{\Lambda,\tau}$ is indexed by the vertices in $V \setminus \Lambda$, while for any two vertices $v, w \in V \setminus \Lambda$, different with each other, if w can take both values ± 1 given τ, we have that

$$\mathcal{I}_G^{\Lambda,\tau}(w,u) = ||\mu_u(\cdot \mid (\Lambda,\tau),(\{w\},+)) - \mu_u(\cdot \mid (\Lambda,\tau),(\{w\},-))||_{\text{TV}} \ ; \tag{16}$$

if w can only take one value in ± 1 given τ, we have $\mathcal{I}_G^{\Lambda,\tau}(w,u) = 0$. Also, we have that $\mathcal{I}_G^{\Lambda,\tau}(w,w) = 0$ for all $w \in V \setminus \Lambda$. That is, the diagonal of $\mathcal{I}_G^{\Lambda,\tau}$ is always zero.

Recall that, above, $\mu_u(\cdot \mid (\Lambda,\tau),(\{w\},1))$ is the Gibbs marginal that vertex u, conditional that the configuration at Λ is τ and the configuration at w is 1. We have the analogous for $\mu_u(\cdot \mid (\Lambda,\tau),(\{w\},-1))$.

▶ **Definition 16** (Spectral Independence). *For a real number $\eta > 0$, the Gibbs distribution μ_G on $G = (V,E)$ is η-spectrally independent, if for every $0 \le k \le |V| - 2$, $\Lambda \subseteq V$ of size k and $\tau \in \{\pm 1\}^\Lambda$ the spectral radius of $\mathcal{I}_G^{\Lambda,\tau}$ satisfies that $\rho(\mathcal{I}_G^{\Lambda,\tau}) \le \eta$.*

We proceed to introduce the Complete Spectral Independence. First, consider the notion of the Magnetising operation.

▶ **Definition 17** (Magnetising operation). *Let μ_G be a Gibbs distribution on the graph $G = (V,E)$. For any local fields $\vec{\phi} \in \mathbb{R}^V_{>0}$, the magnetised distribution $\vec{\phi} * \mu$ satisfies*

$$\forall \sigma \in \{\pm 1\}^V, \quad (\vec{\phi} * \mu)(\sigma) \propto \mu(\sigma) \prod_{v \in V : \sigma_v = +1} \phi_v \ .$$

We denote $\vec{\phi} * \mu$ by $\phi * \mu$ if $\vec{\phi}$ is a constant vector with value ϕ.

Suppose that μ is the Hard-Core model on G with fugacity λ. It is immediate that the magnetisied distribution $\vec{\phi} * \mu$ can be viewed as the *non-homogenious* Hard-Core model such that each vertex v has its own fugacity $\lambda_v = \lambda \cdot \phi_v$.

▶ **Definition 18** (Complete Spectral Independence). *For two reals $\eta > 0$ and $s > 0$, the Gibbs distribution μ_G on $G = (V,E)$ is (η, s)-completely spectrally independent, if the magnetised distribution $\vec{\phi} * \mu$ is η-spectrally independent for all $\vec{\phi} \in (0, 1+s]^V$.*

As far as the Hard-Core model on the random graph $G(n, d/n)$ is concerned, we prove the following result.

▶ **Theorem 19.** *For any fixed $d > 1$ and $\lambda < \lambda_c(d)$, there exist bounded constants $r = r(d,\lambda) \in (1,2)$, $B = B(d,\lambda) > 0$ and $s = s(d,\lambda) > 0$ such that the following holds:*

Consider $\boldsymbol{G} \sim G(n, d/n)$ and let $\mu_{\boldsymbol{G}}$ be the Hard-Core model on $\boldsymbol{G}$ with fugacity λ. With probability $1 - o(1)$ over the instances of $\boldsymbol{G}$, $\mu_{\boldsymbol{G}}$ is $(B \cdot (\log n)^{1/r}, s)$-completely spectrally independent.

The proof of Theorem 19 appears in the full version of this paper, where we first relate the influence matrix on the graph to the influence matrix on the self-avoiding walk tree [31, 27] and then use the potential function in [28] to analysis the *weighted* total influence on the self-avoiding walk tree.

4.3 Entropy Block Factorisation - Proof of Theorem 6

The following theorem, from [8], allows us to derive a bound on the ℓ- block factorisation parameter of the entropy by using the result in Theorem 15 for the stability of Gibbs marginals and the result in Theorem 19 for Complete Spectral Independence.

▶ **Theorem 20** ([8, Lemma 2.3]). *Let $\eta > 0, \xi > 0$ and $\zeta > 0$ be parameters. Let μ_G be a Gibbs distribution on $G = (V, E)$. If μ_G is (η, ξ)-completely spectrally independent and ζ-marginally stable, then for any $1/\alpha \leq \ell < n$, μ_G satisfies the ℓ block factorisation of entropy with parameter $C = (\frac{en}{\ell})^{1+1/\alpha}$, where*

$$\alpha = \min\left\{\frac{1}{2\eta}, \frac{\log(1+\xi)}{\log(1+\xi)+\log 2\zeta}\right\} .$$

Proof of Theorem 6. From Theorem 19 we have the following: with probability $1 - o(1)$ over the instances of $\boldsymbol{G}$ we have that $\mu_{\boldsymbol{G}}$ is (η, s)-completely spectrally independent where $s = s(d, \lambda)$ is *constant*, while

$$\eta = B \cdot (\log n)^{1/r} = o\left(\frac{\log n}{\log\log n}\right),$$

where $B = B(d, \lambda)$ and $r = r(d, \lambda) \in (1, 2)$ are constants specified in the statement of Theorem 19. The second equality above follows by noting that $1/r < 1$, bounded away from 1.

Furthermore, from Theorem 15 we have the following: With probability $1 - o(1)$ over the instances of $\boldsymbol{G}$, the distribution $\mu_{\boldsymbol{G}}$ is ζ-marginally stable, where

$$\zeta \leq 2(1+\lambda)^{2\frac{\log n}{\log\log n}} .$$

In light of all the above, the theorem follows by plugging the above values into Theorem 20. ◀

5 Approximate Tensorisation of Entropy

In this section we prove our results related to the approximate tensorisation of the entropy. These are Theorem 4 and Lemma 8.

5.1 Proof of Theorem 4

In this section we give the full proof of Theorem 4. Recall the high level description of the steps we follow towards this endeavour in Section 3.1.

Proof of Theorem 4. From Theorem 6 we have the following: For $d > 1$ and $\lambda < \lambda_c(d)$, consider $\boldsymbol{G} \sim G(n, d/n)$, while let $\mu = \mu_{\boldsymbol{G}}$ be the Hard-Core model on $\boldsymbol{G}$ with fugacity λ. Let the number $\theta = \theta(d, \lambda)$ in the interval $(0, 1)$ be a parameter whose value is going to be specified later. Then, with probability $1 - o(1)$ over the instances of $\boldsymbol{G}$, for $\ell = \lceil \theta n \rceil$ and for any $f : \Omega \to \mathbb{R}_{>0}$ we have that

$$\mathrm{Ent}_\mu(f) \leq \left(\frac{\mathrm{e}}{\theta}\right)^{1+1/\alpha} \frac{1}{\binom{n}{\ell}} \sum_{S \in \binom{V}{\ell}} \mu\left(\mathrm{Ent}_S(f)\right). \tag{17}$$

Recall that $C(S)$ denotes the set of connected components in $\boldsymbol{G}[S]$, the subgraph that is induced by vertices in S. With a slight abuse of notation, we use $U \in C(S)$ to denote the set of vertices in the component U. By the conditional independence property of the Gibbs distribution and Lemma 7, we have

$$\begin{aligned}
\mathrm{Ent}_\mu(f) &\le \left(\frac{\mathrm{e}}{\theta}\right)^{1+1/\alpha} \frac{1}{\binom{n}{\ell}} \sum_{S\in\binom{V}{\ell}} \sum_{U\in C(S)} \mu\left(\mathrm{Ent}_U(f)\right) \\
\text{(by Lemma 8)} \quad &\le \left(\frac{\mathrm{e}}{\theta}\right)^{1+1/\alpha} \frac{1}{\binom{n}{\ell}} \sum_{S\in\binom{V}{\ell}} \sum_{U\in C(S)} \mathrm{AT}(|U|) \sum_{v\in U} \mu[\mathrm{Ent}_v(f)] \\
&\le \left(\frac{\mathrm{e}}{\theta}\right)^{1+1/\alpha} \sum_{v\in V} \mu[\mathrm{Ent}_v(f)] \sum_{k\ge 1} \mathrm{AT}(k)\Pr[|C_v|=k] \ , \qquad (18)
\end{aligned}$$

where C_v is the connected component in $\boldsymbol{G}[S]$, where S is sampled from $\binom{V}{\ell}$ uniformly at random. In order to bound the innermost summation on the R.H.S. of (18) we distinguish two cases for k. For $1 \le k \le \log n$, we use the trivial bound $\Pr[|C_v| = k] \le 1$, while Lemma 8 implies that

$$\begin{aligned}
\sum_{k=1}^{\log n} \mathrm{AT}(k)\Pr[|C_v|=k] \le \sum_{k=1}^{\log n} \mathrm{AT}(k) &= \sum_{k=1}^{\log n} 3\log\left(1+\lambda+\lambda^{-1}\right)\cdot((1+\lambda)k)^{2+2\eta} \\
&\le 3\log\left(1+\lambda+\lambda^{-1}\right)\cdot\log n\cdot((1+\lambda)\log n)^{2+2\eta} \\
&\le 3\log\left(1+\lambda+\lambda^{-1}\right)\cdot((1+\lambda)\log n)^{3+2\eta} \ ,
\end{aligned}$$

where $\eta = B(\log n)^{1/r}$, for constants $B = B(d,\lambda)$ and $r = r(d) \in (1,2)$. Elementary calculations imply that

$$\sum_{k=1}^{\log n} \mathrm{AT}(k)\Pr[|C_v|=k] \le 3\log\left(1+\lambda+\lambda^{-1}\right)\cdot((1+\lambda)\log n)^{3+2\eta} \le n^x \ , \qquad (19)$$

for $x = o\left(\frac{1}{\log\log n}\right)$.

For $k \ge \log n$, we use the bound in Lemma 9 for $\Pr[|C_v| = k]$, while from Lemma 8 we have

$$\sum_{k\ge\log n} \mathrm{AT}(k)\Pr[|C_v|=k] \le 2k^2\left(1+\lambda+\lambda^{-1}\right)^{2k+2}(2\mathrm{e})^{\mathrm{e}Lk}(2\theta)^k \ ,$$

where $L = L(d)$ is the parameter in Lemma 9. We choose sufficiently small $\theta = \theta(d,\lambda)$ such that

$$\forall k \ge 1, \quad 2k^2\left(1+\lambda+\lambda^{-1}\right)^{2k+2}(2\mathrm{e})^{\mathrm{e}Lk}(2\theta)^k \le (1/2)^k \ .$$

This implies that

$$\sum_{k\ge\log n} \mathrm{AT}(k)\Pr[|C_v|=k] \le \sum_{k\ge\log n}\left(\frac{1}{2}\right)^k \le 1 \ . \qquad (20)$$

Plugging (19), (20) into (18), we get the following: With probability $1-o(1)$ over the instances of $\boldsymbol{G}$ we have that

$$\mathrm{Ent}_\mu(f) \le \left(\frac{\mathrm{e}}{\theta}\right)^{1+1/\alpha}\left(n^{\left(\frac{1}{\log\log n}\right)}+1\right)\sum_{v\in V}\mu[\mathrm{Ent}_v(f)] \ .$$

Since, by Theorem 6 we have that $\frac{1}{\alpha} = K(\frac{\log n}{\log\log n})$, for a constant $K = K(d,\lambda)$, and $\theta = \theta(d,\lambda)$ is also a constant, the above inequality can be written as follows: there is a constant $A = A(d,\lambda)$ such that

$$\mathrm{Ent}_\mu(f) \leq n^{\left(\frac{A}{\log\log n}\right)} \sum_{v\in V} \mu[\mathrm{Ent}_v(f)] \ .$$

The above concludes the proof of Theorem 4. ◀

References

1 Nima Anari, Kuikui Liu, and Shayan Oveis Gharan. Spectral independence in high-dimensional expanders and applications to the hardcore model. In *FOCS*, pages 1319–1330, 2020.

2 Jean Barbier, Florent Krzakala, Lenka Zdeborová, and Pan Zhang. The hard-core model on random graphs revisited. *Journal of Physics: Conference Series*, 473(1):012021, December 2013.

3 Ivona Bezáková, Andreas Galanis, Leslie Ann Goldberg, and Daniel Štefankovič. Fast sampling via spectral independence beyond bounded-degree graphs. In *ICALP*, volume 229 of *LIPIcs*, pages 21:1–21:16. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2022.

4 Pietro Caputo, Georg Menz, and Prasad Tetali. Approximate tensorization of entropy at high temperature. *Ann. Fac. Sci. Toulouse Math. (6)*, 24(4):691–716, 2015.

5 Pietro Caputo and Daniel Parisi. Block factorization of the relative entropy via spatial mixing. *arXiv preprint*, 2020. `arXiv:2004.10574`.

6 Filippo Cesi. Quasi-factorization of the entropy and logarithmic Sobolev inequalities for Gibbs random fields. *Probab. Theory Related Fields*, 120(4):569–584, 2001.

7 Xiaoyu Chen, Weiming Feng, Yitong Yin, and Xinyuan Zhang. Rapid mixing of glauber dynamics via spectral independence for all degrees. In *FOCS*, pages 137–148. IEEE, 2021.

8 Xiaoyu Chen, Weiming Feng, Yitong Yin, and Xinyuan Zhang. Optimal mixing for two-state anti-ferromagnetic spin systems. In *FOCS*, pages 588–599. IEEE, 2022.

9 Zongchen Chen, Kuikui Liu, and Eric Vigoda. Optimal mixing of Glauber dynamics: Entropy factorization via high-dimensional expansion. In *STOC*, 2021. arXiv:2011.02075. `arXiv:2011.02075`.

10 Amin Coja-Oghlan and Charilaos Efthymiou. On independent sets in random graphs. *Random Struct. Algorithms*, 47(3):436–486, 2015.

11 Varsha Dani and Cristopher Moore. Independent sets in random graphs from the weighted second moment method. In *RANDOM*, volume 6845 of *Lecture Notes in Computer Science*, pages 472–482. Springer, 2011.

12 Charilaos Efthymiou. Spectral independence beyond uniqueness using the topological method. *CoRR*, abs/2211.03753, 2022. `arXiv:2211.03753`.

13 Charilaos Efthymiou, Thomas P. Hayes, Daniel Stefankovic, and Eric Vigoda. Sampling random colorings of sparse random graphs. In *SODA*, pages 1759–1771, 2018.

14 Weiming Feng, Heng Guo, Yitong Yin, and Chihao Zhang. Fast sampling and counting k-SAT solutions in the local lemma regime. *J. ACM*, 68(6):40:1–40:42, 2021.

15 Weiming Feng, Heng Guo, Yitong Yin, and Chihao Zhang. Rapid mixing from spectral independence beyond the boolean domain. In *SODA*, pages 1558–1577, 2021.

16 Weiming Feng, Kun He, and Yitong Yin. Sampling constraint satisfaction solutions in the local lemma regime. In *STOC*, pages 1565–1578. ACM, 2021.

17 Alan M. Frieze. On the independence number of random graphs. *Discret. Math.*, 81(2):171–175, 1990.

18 Andreas Galanis, Leslie Ann Goldberg, Heng Guo, and Kuan Yang. Counting solutions to random CNF formulas. *SIAM J. Comput.*, 50(6):1701–1738, 2021.

19 Andreas Galanis, Daniel Štefankovič, and Eric Vigoda. Inapproximability of the partition function for the antiferromagnetic Ising and hard-core models. *Combinatorics, Probability and Computing*, 25(04):500–559, 2016.

20 David Gamarnik and Madhu Sudan. Limits of local algorithms over sparse random graphs. In *ITCS*, pages 369–376. ACM, 2014.

21 Vishesh Jain, Huy Tuan Pham, and Thuy-Duong Vuong. On the sampling Lovász local lemma for atomic constraint satisfaction problems. *arXiv preprint*, 2021. `arXiv:2102.08342`.

22 Vishesh Jain, Huy Tuan Pham, and Thuy Duong Vuong. Spectral independence, coupling with the stationary distribution, and the spectral gap of the Glauber dynamics. *arXiv preprint*, 2021. `arXiv:2105.01201`.

23 Mark Jerrum and Alistair Sinclair. Approximating the permanent. *SIAM J. Comput.*, 18(6):1149–1178, 1989.

24 F. P. Kelly. Stochastic models of computer communication systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, 47:379–395, 1985.

25 Florent Krzakala, Andrea Montanari, Federico Ricci-Tersenghi, Guilhem Semerjian, and Lenka Zdeborová. Gibbs states and the set of solutions of random constraint satisfaction problems. *Proc. Natl. Acad. Sci. USA*, 104(25):10318–10323, 2007.

26 Ankur Moitra. Approximate counting, the Lovász local lemma, and inference in graphical models. *J. ACM*, 66(2):10:1–10:25, 2019.

27 Jesús Salas and Alan D Sokal. Absence of phase transition for antiferromagnetic Potts models via the Dobrushin uniqueness theorem. *J. Stat. Phys.*, 86(3):551–579, 1997.

28 Alistair Sinclair, Piyush Srivastava, Daniel Štefankovič, and Yitong Yin. Spatial mixing and the connective constant: optimal bounds. *Probab. Theory Related Fields*, 168(1-2):153–197, 2017.

29 Allan Sly. Computational transition at the uniqueness threshold. In *FOCS*, pages 287–296, 2010.

30 Allan Sly and Nike Sun. Counting in two-spin models on d-regular graphs. *Ann. Probab.*, 42(6):2383–2416, 2014.

31 Dror Weitz. Counting independent sets up to the tree threshold. In *STOC*, pages 140–149, 2006.