Rapid Mixing via Coupling Independence for Spin Systems with Unbounded Degree

Xiaoyu Chen* Weiming Feng†

Abstract

We develop a new framework to prove the mixing or relaxation time for the Glauber dynamics on spin systems with unbounded degree. It works for general spin systems including both 2-spin and multi-spin systems. As applications for this approach:

- We prove the optimal $O(n)$ relaxation time for the Glauber dynamics of random q-list-coloring on an n-vertices triangle-tree graph with maximum degree Δ such that $q/\Delta > \alpha^\star$, where $\alpha^\star \approx 1.763$ is the unique positive solution of the equation $\alpha = \exp(1/\alpha)$. This improves the $n^{1+o(1)}$ relaxation time for Glauber dynamics obtained by the previous work of Jain, Pham, and Vuong (2022). Besides, our framework can also give a near-linear time sampling algorithm under the same condition.
- We prove the optimal $O(n)$ relaxation time and near-optimal $\widetilde{O}(n)$ mixing time for the Glauber dynamics on hardcore models with parameter λ in *balanced* bipartite graphs such that $\lambda < \lambda_c(\Delta_L)$ for the max degree Δ_L in left part and the max degree Δ_R of right part satisfies $\Delta_R = O(\Delta_L)$. This improves the previous result by Chen, Liu, and Yin (2023).

At the heart of our proof is the notion of *coupling independence* which allows us to consider multiple vertices as a huge single vertex with exponentially large domain and do a "coarse-grained" local-to-global argument on spin systems. The technique works for general (multi) spin systems and helps us obtain some new comparison results for Glauber dynamics.

1 Introduction

The *spin system* is a fundamental probabilistic graphical model. It is defined on a graph $G = (V, E)$, where every vertex is a random variable and every edge models the local interactions. Each variable takes a value from a discrete domain $[q] = \{1, 2, \ldots, q\}$. Each vertex has a vector $b \in \mathbb{R}^q_{\geq 0}$ called the *external field* and each edge has a symmetric matrix $A \in \mathbb{R}^{q\times q}_{\geq 0}$ called the *interaction matrix*. The spin system defines a Gibbs distribution over $[q]^V$ such that for any configuration $\sigma \in [q]^V$,

$$\mu(\sigma) \propto \prod_{v\in V} b(\sigma_v) \prod_{e=\{u,v\}\in E} A(\sigma_u, \sigma_v).$$

The spin system covers many important distributions including the uniform distribution of graph colorings, the Ising model, the hardcore gas model in Physics, and a broad class of undirected graphical models in machine learning [MM09].

*State Key Laboratory for Novel Software Technology, New Cornerstone Science Laboratory, Nanjing University, China. E-mail: chenxiaoyu233@smail.nju.edu.cn

†Institute for Theoretical Studies, ETH Zürich, Switzerland. Email: weiming.feng@eth-its.ethz.ch. Research supported by Dr. Max Rössler, the Walter Haefner Foundation and the ETH Zürich Foundation.

Sampling from the Gibbs distribution is a central algorithmic task for spin systems. The *Glauber dynamics* is a fundamental Markov chain Monte Carlo (MCMC) method for sampling from high-dimensional distributions. Given a distribution μ over $[q]^V$, it starts from an arbitrary $X \in \Omega(\mu)$, where $\Omega(\mu) \subseteq [q]^V$ is the support of μ. In each step, it updates the current state X as follows:

- pick a variable $v \in V$ uniformly at random;
- resample the value of X_v from the conditional distribution $\mu_v(\cdot \mid X_{V\setminus v})$.

It is well-known that if the state space $\Omega(\mu)$ is connected through the moves of Glauber dynamics, then the distribution μ is the unique stationary distribution for the Glauber dynamics.

In this paper, we study the convergence rate of the Glauber dynamics. Let $(X_t)_{t\geq 0}$ denote the random sequence generated by the Glauber dynamics. Many notions capture the convergence rate. The most standard one is the *mixing time*, which is defined by

$$T_{\mathrm{mix}}^{\mathrm{GD}}(\mu, \varepsilon) := \max_{X_0 \in \Omega(\mu)} \min\left\{t > 0 \mid \mathcal{D}_{\mathrm{TV}}\left(X_t \parallel \mu\right) \leq \varepsilon\right\}, \tag{1}$$

where $\mathcal{D}_{\mathrm{TV}}\left(X_t \parallel \mu\right)$ denote the standard *total variation distance* between μ and the distribution of X_t. In words, if the Glauber dynamics starts from the worst initial state X_0, the mixing time is the minimum number t such that the total variation distance between X_t and μ is below a sufficiently small constant. Another widely used notion is the *relaxation time*. Let $P : \Omega(\mu) \times \Omega(\mu) \to [0, 1]$ denote the transition matrix of the Glauber dynamics. A standard fact says that P only has non-negative real eigenvalues $1 = \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_{|\Omega|} \geq 0$ [DGU14]. The gap $\lambda_1 - \lambda_2 = 1 - \lambda_2$ is called the *spectral gap* of Glauber dynamics. The relaxation time is defined by

$$T_{\mathrm{rel}}^{\mathrm{GD}}(\mu) := \frac{1}{1 - \lambda_2}.$$

It is well known that $T_{\mathrm{mix}}^{\mathrm{GD}}(\mu, \varepsilon) = O(T_{\mathrm{rel}}^{\mathrm{GD}}(\mu) \log \frac{1}{\varepsilon \mu_{\min}})$, where $\mu_{\min} = \min_{\sigma \in \Omega(\mu)} \mu(\sigma)$.

Recently, in a series of works [ALOV19, AL20, ALO20] studied Glauber dynamics using high dimensional expanders. An important notion called *spectral independence* was developed during this process. Anari, Liu, and Oveis Gharan [ALO20] first introduced spectral independence for Boolean distributions. The follow-up works [CGŠV21, FGYZ21] then generalized it to non-Boolean distributions. For example, for a Boolean distribution μ over $\{-1, +1\}^{[n]}$, the *influence matrix* $\Psi \in \mathbb{R}_{\geq 0}^{n \times n}$ is defined by $\Psi(u, v) := \mathbf{Pr}_{X\sim\mu}\left[X_v = + \mid X_u = +\right] - \mathbf{Pr}_{X\sim\mu}\left[X_v = + \mid X_u = -\right]$. A distribution μ is C-spectrally independent if the maximum eigenvalue of Ψ is at most C. If every conditional distribution of μ is C-spectrally independent, then by the local-to-global argument [AL20], both relaxation and mixing time of Glauber dynamics are bounded by $n^{O(C)}$, where n is the number of variables. Given this polynomial bound $n^{O(C)}$, many works tried to obtain an improved or even the optimal mixing/relaxation time for Glauber dynamics, especially when μ is a Gibbs distribution defined by spin systems. Chen, Liu and Vigoda [CLV21] proved that for spin systems on bounded degree graphs, the spectral independence implies both $O(n \log n)$ optimal mixing time and $O(n)$ optimal relaxation time.

The next question is how to deal with spin systems on *unbounded degree* graphs. Many works [JPV22, CFYZ21, AJK+22, CE22, CFYZ22] focused on this question. Significant progress was made, especially for 2-spin systems ($q = 2$). [JPV22] first studied coloring and weighted independent sets (hardcore model) in high-girth graphs and proved the near-optimal $n^{1+o(1)}$ relaxation time. [CFYZ21] introduced a stronger variant of spectral independence called *complete spectral independence*, and proved the optimal $O(n)$ relaxation time for anti-ferromagnetic 2-spin systems in the uniqueness regime. To obtain the optimal mixing time, [AJK+22] made the first step and defined

a new notion called *entropic independence*. After a line of works [AJK+22, CE22, CFYZ22], the optimal $O(n \log n)$ mixing time was established for a broad class of 2-spin systems.

Most of the previous techniques [CFYZ21, AJK+22, CE22, CFYZ22] for unbounded degree graphs are restricted to the 2-spin systems. We consider the following question in this paper.

How to prove the optimal mixing/relaxation time
for Glauber dynamics on (multi) spin systems with unbounded degree?

To the best of our knowledge, the only previous result beyond 2-spin systems is the $n^{1+o(1)}$ relaxation time for graph coloring [JPV22]. However, [JPV22] relies on the coupling analysis for colorings in [HV06], which makes it difficult to be generalized to other spin systems.

In this work, we develop a new framework for proving mixing/relaxation time for the Glauber dynamics on general spin systems including both 2-spin and multi-spin systems. Our new framework is based on a stronger variant of the spectral independence known as the *coupling independence*, which is already used implicitly or explicitly in many previous works [Liu21, BCC+21, CZ23, CG24, CLMM23, Jer24]. A spin system μ on $[q]^V$ is C-coupling independent if for any $v \in V$ and $a, b \in [q]$, there is a coupling (X, Y) where $X \sim \mu^{v \leftarrow a}$ and $Y \sim \mu^{v \leftarrow b}$ such that

$$\mathbb{E}\left[d_H(X, Y)\right] \leq C.$$

Here, $d_H(X, Y) = |\{v \in V \mid X_v \neq Y_v\}|$ denotes the hamming distance between X and Y and $\mu^{v \leftarrow a}$ the distribution induced by μ conditional on v taking the value a.

Given a spin system on a graph G with Gibbs distribution μ, we show that if μ and all the conditional distributions induced by μ satisfy the coupling independence and the maximum degree of G is greater than a large constant, then the following comparison results hold for Glauber dynamics.

- **Relaxation time comparison.** The relaxation time satisfies $T_{\text{rel}}^{\text{GD}}(\mu) = O(T_{\text{rel}}^{\text{GD}}(\mu^\star))$, where $\mu^\star$ is a conditional distribution obtained from μ by fixing the values on a subset $\Lambda \subseteq V$ of variables. The set Λ is chosen intentionally such that the induced subgraph $G[V \setminus \Lambda]$ on other vertices has smaller maximum degree. For many spin systems, the distribution $\mu^\star$ is in an "easy regime" so that the mixing/relaxation time for $\mu^\star$ is easy to analyze. We can bound the relaxation time for μ via this comparison result (see Theorem 9).
- **Mixing time comparison.** If μ is a monotone spin system (Definition 14) and the Glauber dynamics starts from a specific initial configuration, then the mixing time satisfies $T_{\text{mix}}^{\text{GD}}(\mu, \varepsilon) = \widetilde{O}(T_{\text{mix}}^{\text{GD}}(\mu^\star, \frac{1}{4e}))$, where $\widetilde{O}$ hides a polylog(n/ε) factor (See Theorem 15).

We obtain the relaxation/mixing time bounds via the above comparison results. In the relaxation time comparison result, the constant factor in $O(\cdot)$ is independent of the degree of the graph. In applications, the distribution $\mu^\star$ is in an "easy regime", we can use some standard technique to show $T_{\text{rel}}^{\text{GD}}(\mu^\star) = O(n)$. The comparison result gives the optimal $T_{\text{rel}}^{\text{GD}}(\mu) = O(n)$ relaxation time. Similarly, in the applications of monotone systems, we can obtain the near-optimal $\widetilde{O}(n)$ mixing time for general graphs. Our comparison results only hold for graphs with large maximum degrees. It does not cause any issue in applications, because coupling independence implies spectral independence, and for graphs with bounded maximum degree, [CLV21] already established the optimal relaxation/mixing time.

Our proving techniques can also give a near-linear time (in input size) sampling algorithm (see Theorem 13). Furthermore, we introduce a general technique to establish coupling independence for 2-spin systems (Theorem 16). Specifically, many spectral independence results for 2-spin systems are proved by analyzing the decay of correlation in the self-avoiding walk tree [CLV20, CLY23]. We show that all of such proofs can be translated to a proof of coupling independence.

Organization of the paper In Section 1.1, we first exhibit some concrete applications. In Section 2, we give our technical results and an overview of proof techniques. Section 3 is for preliminaries. In Section 4, we prove the relaxation time comparison result. In Section 5, we give a near-liner time sampling algorithm. In Section 6, we prove the mixing time comparison result for monotone systems. In Section 7, we give a general technique for establishing coupling independence. Section 8 and Section 9 are for proofs of applications.

1.1 Applications

Let $G = (V, E)$ be a graph with maximum degree Δ and $[q] = \{1, 2, \dots, q\}$ a set of colors. Given a set of color lists $L_v \subseteq [q], v \in V$, a proper list-coloring $X \in [q]^V$ assigns a color $X_v \in L_v$ to each vertex $v \in V$ such that adjacent vertices receive different colors. In a special case when $L_v = [q]$ for all $V \in V$, the list coloring becomes the standard graph q-coloring. We use μ to denote the uniform distribution over all proper list-colorings in G. For the list coloring, in each step, the Glauber dynamics picks a random vertex v and update its color to a random available color. There is a long line of works studying the mixing and relaxation time of Glauber dynamics e.g. [Jer95, Vig00, CDM+19].

In the era of spectral independence, the proper list-coloring has been re-studied by a series of works [CGŠV21, FGYZ21, CLMM23]. Though the technique varies, all these works ended up establishing some coupling independence results for the proper list-coloring. For list colorings on triangle-free graphs. Let $\alpha^\star \approx 1.763$ denote the unique positive solution to the equation $\alpha = \exp(1/\alpha)$. When $|L_v| > (\alpha^\star + \delta)\Delta$, the $O_\delta(1)$-coupling independence can be established by techniques in [CGŠV21, FGYZ21]. Our framework gives the optimal relaxation time of Glauber dynamics even if the maximum degree of G is unbounded.

Theorem 1 (Coloring: Relaxation Time)**.** *Let $\delta > 0$ be a constant. For any triangle-free graph $G = (V, E)$ and color lists $(L_v)_{v \in V}$, if $|L_v| \geq (\alpha^\star + \delta)\Delta$ for all $v \in V$, where $\Delta \geq 3$ is the maximum degree of G, then relaxation time of Glauber dynamics is $O_\delta(n)$, where n is the number of vertices in G.*

Under the condition of Theorem 1, the relaxation time of the Glauber dynamics has been studied by many previous works. Combining the spectral independence technique [AL20, ALO20] with the correlation decay analysis [GMP05, GKM13], two independent works [CGŠV21, FGYZ21] proved the polynomial relaxation time $n^{O(1/\delta)}$ of Glauber dynamics. For graphs with bounded maximum degree $\Delta = O(1)$, Chen, Liu and Vigoda [CLV21] established the $O_{\Delta,\delta}(n)$ relaxation time, where $O_{\Delta,\delta}(\cdot)$ hides a constant factor like $\Delta^{O(\Delta^2/\delta)}$. For general graphs with possibly unbounded maximum degree, Jain, Pham and Vuong [JPV22] proved the first almost linear relaxation time $O_\delta(ne^{(\log\log n)^2}) = O_\delta(n^{1+o(1)})$. Their proof combined the techniques in [CLV21] with the coupling analysis in [HV06]. Compared to previous results, Theorem 1 gives the optimal linear relaxation time for general graphs.

We prove Theorem 1 by first verifying the coupling independence condition (Definition 7) and then applying our comparison result (Theorem 9). Theorem 1 requires $|L_v| \geq (\alpha^\star + \delta)\Delta$ because the current best coupling independence result requires this number of colors but our comparison result does not require such a strong condition. It is conjectured that $O_\delta(1)$-coupling independence should hold for proper list-coloring in general graphs when $|L_v| \geq (1+\delta)\Delta + O(1)$.

Conjecture 2 (Folklore)**.** Let $\delta > 0$ be a constant. For any graph $G = (V, E)$ with maximum degree Δ and color lists $(L_v)_{v \in V}$ such that $|L_v| \geq (1+\delta)\Delta + O(1)$ for all $v \in V$, the the uniform distribution μ over all the proper list-colorings of G is $O_\delta(1)$-coupling independent.

Our comparison framework can prove optimal relaxation time for Glauber dynamics on proper list-colorings of graphs (with potentially unbounded degree) as long as Conjecture 2 holds.

Proposition 3. *If Conjecture 2 holds with $\delta > 0$, then for any list coloring instance in Conjecture 2, the relaxation time of Glauber dynamics is $O_\delta(n)$.*

The standard relation between relaxation time and mixing time implies that the Glauber dynamics mixes in time $O_\delta(n^2 \log q)$, which yields a sampling algorithm for the uniform distribution μ of graph colorings in time $O_\delta(\Delta n^2 \log q)$ because each step of Glauber dynamics can be simulated in time $O(\Delta)$. However, in terms of sampling algorithm, our technique would directly give an algorithm (not the Glauber dynamics) in time $\widetilde{O}_\delta(\Delta n)$. Since the input graph G contains Δn edges, the running time is linear-near in the input size.

Theorem 4 (Coloring: Algorithm)**.** *Let $\delta > 0$ be a constant. There exists an algorithm such that given any $\varepsilon > 0$, any triangle-free graph $G = (V, E)$ and color lists $(L_v)_{v \in V}$, if $|L_v| \geq (\alpha^\star + \delta)\Delta$ for all $v \in V$, where $\Delta \geq 3$ is the maximum degree of G, it returns a random sample X satisfying $\mathcal{D}_{\mathrm{TV}}(X \parallel \mu) \leq \varepsilon$ in time $\Delta n (\log \frac{n}{\varepsilon})^{C(\delta)}$, where $C(\delta)$ is a constant depending only on δ.*

The next example is the hardcore model. Let $G = (V, E)$ be a graph. Let $\lambda > 0$ be the fugacity. The hardcore model defines a distribution μ over all independent sets $S \subseteq V$ in G such that $\mu(S) \propto \lambda^{|S|}$. Let $\Delta \geq 3$ denote the maximum degree of graph G. There is a critical threshold for the tree uniqueness phase transition [Kel85]

$$\lambda_c(\Delta) := \frac{(\Delta - 1)^{(\Delta - 1)}}{(\Delta - 2)^\Delta}.$$

such that if $\lambda \leq \lambda_c(\Delta)$ the correction between two vertices decays in their distance; if $\lambda > \lambda_c(\Delta)$, the long-range correlation exists. A computational phase transition occurs at the same threshold. If $\lambda < \lambda_c(\Delta)$, polynomial time sampling algorithm exists [Wei06]; if $\lambda > \lambda_c(\Delta)$, the sampling problem is hard unless **NP** = **RP** [Sly10]. The mixing and relaxation time of the Glauber dynamics for hardcore model were also extensively studied [LV99, HV06, EHŠ+19]. Recent works analyzed Glauber dynamics via spectral independence [ALO20]. The optimal $O_\delta(n \log n)$ mixing time and the optimal $O_\delta(n)$ relaxation time were established when $\lambda \leq (1 - \delta)\lambda_c(\Delta)$ for general graphs [CLV21, CE22, CFYZ22].

However, for the hardcore model on bipartite graphs, the picture is not very clear. Consider the hardcore model in a bipartite graph $G = (V = V_L \uplus V_R, E)$. Let Δ_L denote the maximum degree in the left part. Assume $3 \leq \Delta_L$. It is recently known that the uniqueness threshold for the hardcore model on the bipartite graph can be refined to $\lambda_c(\Delta_L) \geq \lambda_c(\Delta)$ where $\Delta \geq \Delta_L$ is the maximum degree of the bipartite graph [LL15, CLY23]. The Glauber dynamics is also proved to have polynomial mixing time when $\lambda < \lambda_c(\Delta_L)$ [CLY23].

On the other side, when $\lambda > \lambda_c(\Delta_L)$, the lower bound in [Sly10] does not hold for bipartite graphs and the problem is #BIS-hard [CGG+16], where #BIS is the problem of counting the independent sets in bipartite graphs. A line of works (e.g. [JKP20, CP20, LLLM22, CGŠV22, JPP23]) studied various sampling algorithms in the low-temperature (large λ) regime.

Within the critical threshold $\lambda < \lambda_c(\Delta_L)$, we consider "balanced" bipartite graphs. Let Δ_R be the maximum degree in V_R. We say a bipartite graph is θ-balanced if $\Delta_R \leq \theta \Delta_L$.

Theorem 5 (Bipartite Hardcore: Relaxation Time)**.** *Let $\delta \in (0, 1)$ and $\theta > 1$ be two constants. For any hardcore model on a θ-balanced bipartite graph G with fugacity λ, if $\lambda \leq (1 - \delta)\lambda_c(\Delta_L)$, then the relaxation time of Glauber dynamics is $O_{\delta,\theta}(n)$, where n is the number of vertices in G.*

For the mixing time, again, the standard relation gives $O_{\delta,\theta}(n^2)$ mixing time of the Glauber dynamics. However, since the bipartite graph hardcore is a monotone system, our technique also implies the $\widetilde{O}_{\delta,\theta}(n)$ *mixing time* of Glauber dynamics starting from the independent set containing all vertices in the left part: $X_0 = V_L$. Formally, for any $S \in \Omega(\mu)$,

$$T^{\mathrm{GD}}_{\mathrm{mix}}(\mu, \varepsilon \mid S) = \min\left\{t > 0 \mid \mathcal{D}_{\mathrm{TV}}\left(X_t \parallel \mu\right) \le \varepsilon \wedge X_0 = S\right\}.$$

Theorem 6 (Bipartite Hardcore: Mixing Time)**.** *Let $\delta \in (0,1)$ and $\theta > 1$ be two constants. For any hardcore model on a θ-balanced bipartite graph G with fugacity λ, if $\lambda \le (1-\delta)\lambda_c(\Delta_L)$, then the mixing time of Glauber dynamic starting from V_L satisfies $T^{\mathrm{GD}}_{\mathrm{mix}}(\mu, \varepsilon \mid V_L) = n(\log \frac{n}{\varepsilon})^{C(\delta,\theta)}$, where $C(\delta,\theta)$ is a constant depending only on δ and θ.*

The previous work [CLY23] established the $(\frac{\Delta_L \log n}{\lambda})^{O(1/\delta)} n^2$ relaxation time for the Glauber dynamics, which, by the standard relation, implies the $(\frac{\Delta_L \log n}{\lambda})^{O(1/\delta)} n^3 \log \frac{1+\lambda}{\lambda}$ mixing time. The previous result holds for general bipartite graphs as long as $\lambda \le (1-\delta)\lambda_c(\Delta_L)$. For balanced bipartite graphs, we obtained the optimal relaxation time $O_{\delta,\theta}(n)$ and the near-optimal mixing time $\widetilde{O}_{\delta,\theta}(n)$, which significantly improved the dependency to n and Δ_L compared to the previous result. For example, in the critical case when $\lambda_c = (1-\delta)\lambda_c(\Delta_L) = \Theta(1/\Delta_L)$, previous result gives $\Delta_L^{O(1/\delta)} n^2 \cdot \mathrm{polylog}(n)$ relaxation time and $\Delta_L^{O(1/\delta)} n^3 \cdot \mathrm{polylog}(n)$ mixing time but our result gives $O(n)$ relaxation time and $n \cdot \mathrm{polylog}(n)$ mixing time. However, our result works only on balanced bipartite graphs. The result in [CLY23] is still state-of-the-art for general bipartite graphs.

Finally, we point out that our technique could also recover many previous $O(n)$ relaxation time results for anti-ferromagnetic 2-spin systems in [CFYZ21]. See Remark 10 for one example.

2 Technical Results and Proof Overview

2.1 Coupling Independence

In this section, we give our general results for spin systems. Let $G = (V, E)$ be a graph. Let $[q] = \{1, 2, \ldots, q\}$ be a set of $q \ge 2$ spins. For each vertex $v \in V$, let vector $b_v \in \mathbb{R}^q_{\ge 0}$ be the *external field* at vertex v. For each edge $e \in E$, let symmetric matrix $A_e \in \mathbb{R}^{q\times q}_{\ge 0}$ be the *interaction matrix* at edge e. A spin system defines a *Gibbs distribution* μ over $[q]^V$ such that,

$$\forall \sigma \in [q]^V \quad \mu(\sigma) \propto w(\sigma) := \prod_{u \in V} b_u(\sigma_u) \prod_{e=\{v,w\} \in E} A_e(\sigma_v, \sigma_w).$$

We often use $\Omega(\mu) \subseteq [q]^V$ to denote the support of the Gibbs distribution μ.

Let $\Lambda \subseteq V$ be a subset of vertices. Given any *pinning* $\tau \in [q]^{V\setminus\Lambda}$, we define a conditional distribution μ^τ by for any configuration $\sigma \in [q]^V$,

$$\mu^\tau(\sigma) \propto w^\tau(\sigma) := \mathbf{1}[\sigma_\Lambda = \tau] \cdot \prod_{u \in \Lambda} b_u(\sigma_u) \prod_{\substack{e=\{v,w\} \in E \\ v,w \in \Lambda}} A_e(\sigma_v, \sigma_w) \prod_{\substack{e=\{v,w\} \in E \\ v \in \Lambda \wedge w \notin \Lambda}} A_e(\sigma_v, \tau_w). \tag{2}$$

In words, μ^τ is a Gibbs distribution obtained for μ by removing all edges $e \subseteq V \setminus \Lambda$ and putting a constraint that every vertex in $v \in V \setminus \Lambda$ must take the value τ_v. In particular, if τ is feasible (e.g. τ belongs the support of the marginal distribution $\mu_{V\setminus\Lambda}$), then μ^τ is exactly the conditional distribution induced by μ given the condition τ. For all spin systems considered in this paper, it holds that $\sum_\sigma w^\tau(\sigma) > 0$ for all τ. The distribution in (2) is well-defined. Furthermore, for any subset S, we use μ^τ_S to denote the marginal distribution on S projected from μ^τ.

The following conditions plays a key role in the proof of our main results. Let $\rho : V \to \mathbb{N}_{>0}$ be a function that maps every vertex $v \in V$ to a positive integer. We call the function ρ the Hamming weight function. For any two (possibly partial) configurations $\sigma, \tau \in [q]^\Lambda$, where $\Lambda \subseteq V$, define their weighted Hamming distance with respect to ρ by

$$H_\rho(\sigma, \tau) := \sum_{v \in \Lambda : \sigma(v) \neq \tau(v)} \rho(v). \tag{3}$$

Definition 7 (Coupling Independence)**.** Let $C \geq 1$ be a constant. A distribution μ over $[q]^V$ is said to be C-coupling independent (C-CI) if there exists Hamming weight function $\rho : V \to \mathbb{N}_{>0}$ such that the following holds. For any pinning $\sigma_1, \sigma_2 \in [q]^S$, where $S \subseteq V$ and σ_1, σ_2 disagree only at one vertex $v_0 \in S$, there exists a coupling (X, Y), where $X \sim \mu^{\sigma_1}$ and $Y \sim \mu^{\sigma_2}$, such that

$$\frac{\mathbb{E}\left[H_\rho(X, Y)\right]}{\rho(v_0)} \leq C.$$

The expectation $\mathbb{E}\left[H_\rho(X, Y)\right]$ is a kind of Wasserstein distance between μ^{σ_1} and μ^{σ_2}. The notion of coupling independence was introduced explicitly in [CZ23] to study the spectral independence property. For example, for Boolean distributions[1] ($q = 2$), given any pinning $\tau \in \{-, +\}^{V \setminus \Lambda}$, the $|\Lambda| \times |\Lambda|$ influence matrix [ALO20] is defined by

$$\Psi_\mu^\tau(v, u) := \mu_u^{\tau \wedge v^+}(+) - \mu_u^{\tau \wedge v^-}(+), \tag{4}$$

where $u, v \in \Lambda$ and $\tau \wedge v^\pm$ denotes the pinning τ together with v taking the value $\pm$. A distribution μ is C-spectrally independent if the maximum eigenvalue of Ψ_μ^τ is at most C for any pinning τ. It is not hard to show that C-coupling independence implies C-spectral independence. Hence, recent works [CZ23, CLMM23, CG24] utilized coupling independence to establish the spectral independence for various spin systems.

2.2 Compare Markov Chains via Coupling Independence

In this work, we find more applications for coupling independence beyond establishing spectral independence. We build some comparison results of Markov chains via coupling independence. As a by-product result, we also show that the coupling independence gives fast sampling algorithms.

Let μ be a Gibbs distribution over $[q]^V$ on graph $G = (V, E)$. For any $\Lambda \subseteq V$, we use $G[\Lambda]$ to denote the induced subgraph of G on vertex set Λ.

Definition 8 (Relaxation Time with Pinning)**.** Let μ be a Gibbs distribution on graph $G = (V, E)$ with maximum degree Δ. Let $\eta \in [0, 1]$. Let $D(\eta)$ denote all subsets $\Lambda \subseteq V$ such that the maximum degree of $G[\Lambda]$ is at most $\eta\Delta$. Define

$$T_{\text{rel}}^{(\eta)}(\mu) := \max\left\{ T_{\text{rel}}^{\text{GD}}(\mu^\tau) \mid \Lambda \in D(\eta) \wedge \tau \in [q]^{V \setminus \Lambda} \right\}.$$

In the above definition, μ^τ is a distribution on $[q]^V$. In every step, the Glauber dynamics picks $v \in V$ uniformly at random then resamples the value on v. If $v \notin \Lambda$, the value of v after resampling is always τ_v. Indeed, μ^τ is essentially the same as μ_Λ^τ. But, considering μ^τ would help us simplify some results and proofs. The following is our main comparison result.

[1] The influence matrix and spectral independence are also defined for general distributions with $q \geq 2$. See [CGŠV21] for the detailed definition.

Theorem 9 (Relaxation Time Comparison). *Let $M \geq 1$ and $0 < \eta \leq \frac{1}{2\lceil M \rceil}$ be two constants. There exists $\Delta_0 = \Omega(\frac{M^2}{\eta^2} \log \frac{M}{\eta})$ such that for any Gibbs distribution μ on graph G with the maximum degree $\Delta \geq \Delta_0$, if μ satisfies M-coupling independence, then the relaxation time of Glauber dynamics on μ satisfies*

$$T_{\text{rel}}^{\text{GD}}(\mu) \leq 2^{O(M/\eta)} \cdot T_{\text{rel}}^{(\eta)}(\mu).$$

The theorem is proved in Section 4. See Section 2.4 for a proof overview.

The above theorem is a comparison result between two kinds of relaxation times. Consider the case when parameters of μ are close to the critical threshold so that the relaxation time $T_{\text{rel}}^{\text{GD}}(\mu)$ is hard to analyze. By choosing a sufficiently small η, suppose for any $\Lambda \in D(\eta)$ and any $\tau \in [q]^{V\setminus\Lambda}$, the conditional distribution μ^τ falls into an easy regime. The relaxation time $T_{\text{rel}}^{(\eta)}(\mu)$ is easy to analyze. Theorem 9 boosts the relaxation time from an easy regime to the hard regime if μ satisfies the coupling independence and the maximum degree Δ is greater than a constant Δ_0.

When applying Theorem 9 to a specific spin system with Gibbs distribution μ, we first need to show that the μ satisfies the coupling independence property. Next, we choose a small constant η to guarantee that $T_{\text{rel}}^{(\eta)}(\mu)$ is easy to analysis. Now, the constant parameter Δ_0 in Theorem 9 is fixed. If the maximum degree $\Delta \leq \Delta_0 = O(1)$ is bounded, then since the coupling independence implies the spectral independence, the previous work [CLV21] already established the optimal relaxation time for μ. If the maximum degree $\Delta \geq \Delta_0$, we can apply our boosting result to bound the relaxation time. We show how to prove Theorem 1 and Proposition 3 via Theorem 9.

Proof Sketch of Theorem 1 Given a triangle-free graph $G = (V, E)$ and color lists $L_v \subseteq [q]$ with $|L_v| \geq (\alpha^\star + \delta)\Delta$ for all $v \in V$, let μ denote the uniform distribution over all proper list-colorings. By going through the analysis in [FGYZ21], we can prove that μ satisfies $O(1/\delta)$-coupling independence. Let η be a parameter to be fixed later. For any $\Lambda \subseteq D(\eta)$, any pinning $\tau \in [q]^{V\setminus\Lambda}$, the distribution μ^τ is essentially the same as the distribution μ_Λ^τ because the coloring outside Λ is fixed by τ. By self-reducibility, μ_Λ^τ is a list coloring on $G' = G[\Lambda]$ with color list $L'_v = L_v \setminus \{\tau_u \mid u \notin \Lambda \wedge \{u, v\} \in E\}$. Let $\deg'(v)$ and $\deg(v)$ denote the degree of v in G' and G respectively. The new instance satisfies

$$\forall v \in \Lambda, \quad |L'_v| \geq |L_v| - (\deg(v) - \deg'(v)) \implies \frac{|L'_v|}{\Delta'} \geq \frac{|L_v| - \deg(v)}{\Delta'},$$

where Δ' denote the maximum degree of G'. By the definition of $D(\eta)$, $\deg'(v) \leq \Delta' \leq \eta\Delta$. We have $|L_v| - \deg(v) > (\alpha^\star - 1)\Delta \geq \frac{\alpha^\star - 1}{\eta}\Delta'$. if we set the parameter $\eta \leq \frac{1}{10}$, then

$$\forall v \in \Lambda, \quad |L'_v| \geq 5\Delta'. \tag{5}$$

In this easy regime, one can use path coupling [BD97] to show $T_{\text{rel}}^{(\eta)}(\mu) = O(n)$. To apply Theorem 9, we pick a small $\eta = O(\delta)$ and $\eta < \frac{1}{10}$. If $\Delta \geq \Delta_0 = \Theta(\frac{1}{\delta^4} \log \frac{1}{\delta})$, then

$$T_{\text{rel}}^{(\eta)}(\mu) = 2^{O(1/\delta^2)} n = O_\delta(n).$$

On the other hand, if $\Delta \leq \Delta_0 = \Theta(\frac{1}{\delta^4} \log \frac{1}{\delta})$, then the maximum degree is bounded, we can use the result in [CLV21] to obtain the relaxation time $T_{\text{rel}}^{(\eta)}(\mu) = O_\delta(n)$ in the same order. This gives the proof sketch of Theorem 1. The only missing component is how to establish the coupling independence, which can be found in Section 9.

Proof of Proposition 3 To obtain (5), we only need to use the fact that $\alpha^\star > 1$. If we replace $\alpha^\star$ with $(1+\delta)$, then we can set $\eta \leq \frac{\delta}{5}$ and (5) still holds. The same analysis proves Proposition 3.

Remark 10 (Hardcore Model in Uniqueness Regime)**.** Theorem 9 could also rediscover some previous results. For example, for the hardcore model on a graph $G = (V, E)$ with fugacity $\lambda \leq (1-\delta)\Delta$, [CFYZ21] proved the optimal $O_\delta(n)$ relaxation time. For a fixed λ, the hardcore model falls into an easy regime if we can reduce the maximum degree of the graph by a constant factor. The hardcore model in the uniqueness regime satisfies $O(1/\delta)$-coupling independence (which can be proved by Theorem 16 in this paper). Using a similar argument as that for list coloring, one can rediscover the optimal $O_\delta(n)$ relaxation time using Theorem 9.

We remark that the relaxation time result for the bipartite graph hardcore model (Theorem 5) is not a direct consequence from Theorem 9. We need to tweak the proof of Theorem 9 to prove Theorem 5. The proof of Theorem 5 is in Section 8. See Section 2.4 for a proof overview.

Remark 11 (Compare Theorem 9 to the Technique in [CFYZ21])**.** Another comparison result about relaxation time was given in [CFYZ21]. The previous result considers general Boolean distribution (not necessarily Gibbs distribution) μ over $\{-,+\}^V$. Given a vector $\lambda = (\lambda_v)_{v\in V}$, $(\lambda * \mu)$ denotes the distribution such that for any $\sigma \in \{-,+\}^V$, $(\lambda * \mu) \propto \mu(\sigma)\prod_{v\in V:\sigma(v)=+}\lambda_v$. The result says if $(\lambda * \mu)$ is spectrally independent for all $\lambda \in (0,1]^V$, then one can compare $T^{\mathrm{GD}}_{\mathrm{rel}}(\mu)$ to $T^{\mathrm{GD}}_{\mathrm{rel}}(\lambda_\theta * \mu)$, where λ_θ is the vector with constant value $0 < \theta < 1$. When applying results to Gibbs distributions, here are some differences between Theorem 9 and the previous result.

- Theorem 9 works for general domain $[q]$ but previous result works only for Boolean domain;
- The condition is incomparable. Theorem 9 requires coupling independence for μ and a degree lower bound for the underlying graph but the previous result requires spectral independence for a family of distributions;
- The easy regime is incomparable. The easy regime in Theorem 9 is the conditional distributions on a small degree subgraph but the easy regime in the previous result is $\lambda_\theta * \mu$;

For many spin systems, one can use Theorem 9 to establish the optimal $O(n)$ relaxation for Glauber dynamics, where n is the number of variables in the spin system. By the standard relation between mixing and relaxation time, the mixing time of Glauber dynamics can usually be bound by $O(n^2)$. Each transition of Glauber dynamics can be simulated in time $O(\Delta)$. Hence, one can obtain a sampling algorithm in time $O(\Delta n^2)$. Alternatively, we can give a faster sampling algorithm in time $\widetilde{O}(\Delta n)$ if the easy regime has linear-near mixing time.

Definition 12 (Mixing Time with Pinning)**.** Let μ be a Gibbs distribution on graph $G = (V, E)$ with maximum degree Δ. Let $\eta \in [0,1]$. Let $D(\eta)$ denote all subsets $\Lambda \subseteq V$ such that the maximum degree of $G[\Lambda]$ is at most $\eta\Delta$. Define

$$T^{(\eta)}_{\mathrm{mix}}(\mu) := \max\left\{T^{\mathrm{GD}}_{\mathrm{mix}}\left(\mu^\tau, \frac{1}{4e}\right) \mid \Lambda \in D(\eta) \wedge \tau \in [q]^{V\setminus\Lambda}\right\}.$$

In words, for any pinning τ on $V \setminus \Lambda$ with $\Lambda \in D(\eta)$, $T^{(\eta)}_{\mathrm{mix}}(\mu)$ is an upper bound for the mixing time T of Glauber dynamics for μ^τ such that starting from the worst initial X_0, the total variation distance between X_T and μ^τ is at most $\frac{1}{4e}$.

Theorem 13 (Fast Sampling Algorithm)**.** *Let $M \geq 1$ and $0 < \eta \leq \frac{1}{2\lceil M\rceil}$ be two constants. There exists an algorithm such that given any $\varepsilon \in (0,1)$ and any Gibbs distribution μ on graph G with the maximum*

degree $\Delta \geq \Delta_0 = (\frac{10M}{\eta})^2 \log \frac{10M}{\eta}$*, if* μ *satisfies* M*-coupling independence such that the weighted hamming distance* ρ *satisfies* $\frac{\rho_{\max}}{\rho_{\min}} = \mathrm{poly}(n)$*, then it returns a random sample* X *satisfying* $\mathcal{D}_{\mathrm{TV}}(X \parallel \mu) \leq \varepsilon$ *in time*

$$\Delta T^{(\eta)}_{\mathrm{mix}}(\mu) \left(\log \frac{n}{\varepsilon}\right)^{O(M/\eta)},$$

where n *is the number of vertices in* G *and we use* $\rho_{\max} = \max_{v \in V} \rho(v)$ *and* $\rho_{\min} = \min_{v \in V} \rho(v)$*.*

The theorem is proved in Section 5. See Section 2.4 for a proof overview.

In the above theorem, suppose $\frac{\rho_{\max}}{\rho_{\min}} = O(n^d)$ for some universal constant d, then the running time in above theorem should be $T^{\mathrm{GD}}_{\mathrm{mix}}(\mu, \eta)\Delta(\log \frac{n}{\varepsilon})^{C(M/\eta+d)}$ for some universal constant C. We then hide the constants C and d by $O(\cdot)$ in Theorem 13.

Theorem 4 can be obtained from Theorem 13. Consider the list coloring on a triangle free graphs $G = (V, E)$ with $|L_v| \geq (\alpha^\star + 1)\Delta$. The uniform distribution μ satisfies $O(1/\delta)$-coupling-independence with standard Hamming weight $\rho(v) = 1$ for all $v \in V$. Take $\eta = O(1/\delta)$ be a small constant with $\eta < \frac{1}{10}$. By (5), a simple path coupling [BD97] shows that $T^{\mathrm{GD}}_{\mathrm{mix}}(\mu, \eta) = O(n \log n)$. Hence, if the maximum degree of Δ is greater than Δ_0, we run the algorithm in Theorem 13 and the running time is $\Delta n \cdot \mathrm{polylog}(\frac{n}{\varepsilon})$. Otherwise, the maximum degree is bounded and the result in [CLV21] gives the $O_\delta(n \log n)$ mixing time of the Glauber dynamics, then we can simulate Glauber dynamics to obtain a sampling algorithm. The proof of Theorem 4 is in Section 9.

The algorithm in Theorem 13 is *not* the Glauber dynamics. Roughly speaking, the algorithm uses some strategy to pick vertex and uses the Glauber update to resample the value of the picked vertex. However, for *monotone spin systems*, we can compare this algorithm to Glauber dynamics via censoring inequality [PW13] and then we can bound the mixing time of Glauber dynamics.

Let μ over $[q]^V$ be the Gibbs distribution. Define a partial order $\leq$ for $[q]^V$ as follows. For each $v \in V$, pick a *total order* $\leq_v$ on $[q]$. For any two $X, Y \in [q]^V$,

$$X \leq Y \quad \Longleftrightarrow \quad \forall v \in V, \quad X_v \leq_v Y_v. \tag{6}$$

For two distributions μ and ν over $[q]^V$, we say μ *is stochastic dominated by* ν (i.e., $\mu \preceq \nu$) if there is a coupling $\mathcal{C}$ between μ, ν such that $\mathbf{Pr}_{(X,Y)\sim\mathcal{C}}[X \leq Y] = 1$. Let P be the transition matrix of the Glauber dynamics on μ, which can be written as

$$P = \frac{1}{n} \sum_{v \in V} P_v, \tag{7}$$

where P_v performs *updates* at the $v \in V$ such that $P_v(X, Y) = \mu^{X_{V\setminus v}}(Y)$, for all $X, Y \in [q]^V$.

Definition 14. [LPW17, Chapter 22] We say μ is a *monotone spin system* if for every $v \in V$, P_v is ordering persists, which means for any $X, Y \in [q]^V$ with $X \leq Y$, it holds that $P_v(X, \cdot) \preceq P_v(Y, \cdot)$.

By the definition of the partial ordering $\leq$ in $[q]^V$, there is a unique maximum configuration for the ordering. Denote this state as X^+. Recall $T^{\mathrm{GD}}_{\mathrm{mix}}(\mu, \varepsilon \mid X^+)$ denotes the mixing time of Glauber dynamics starting from X^+.

Theorem 15 (Mixing Time Comparison)**.** *Let* $M \geq 1$ *and* $0 < \eta \leq \frac{1}{2\lceil M \rceil}$ *be two constants. For any monotone spin system* μ *on graph* G *with the maximum degree* $\Delta = \Omega(\frac{M^2}{\eta^2} \log \frac{M}{\eta})$*, if* μ *satisfies* M*-coupling-independence such that the Hamming weight* ρ *satisfies* $\frac{\rho_{\max}}{\rho_{\min}} = \mathrm{poly}(n)$*, then the mixing time of Glauber dynamics starting from the maximum configuration satisfies*

$$T^{\mathrm{GD}}_{\mathrm{mix}}(\mu, \varepsilon \mid X^+) \leq \left(\log \frac{n}{\varepsilon}\right)^{O(M/\eta)} \cdot T^{(\eta)}_{\mathrm{mix}}(\mu),$$

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

The theorem is proved in Section 6. See Section 2.4 for a proof overview.

The above theorem is of independent interest. Suppose μ is a monotone system with coupling independence property. The parameters of μ are in the critical regime and the underlying graph has an unbounded maximum degree. If we can choose a proper constant η such that $T_{\text{mix}}^{(\eta)}(\mu) = O(n \log n)$, then the theorem gives a linear-optimal $\widetilde{O}(n)$ mixing time of Glauber dynamics for μ starting from the maximum configuration.

To obtain the linear-near mixing time for μ, some previous works [CFYZ22, CE22, AJK+22] developed comparison techniques for the modified log-Sobolev (MLS) constants. Roughly speaking, if one can lower bound the MLS constant $\text{mls}(\mu)$ of the Glauber dynamics for μ, then one can obtain the optimal $O(n \log n)$ mixing time. Previous works compared $\text{mls}(\mu)$ to $\text{mls}(\mu')$, where μ' is a distribution in the easy regime, and such comparison requires μ to satisfy certain entropic independence [AJK+22] condition. In general, it is not easy to verify the entropic independence condition and analyze $\text{mls}(\mu')$ even if μ' is in an easy regime. Theorem 15 only requires the coupling independence condition and directly compares the mixing time. However, Theorem 15 requires monotone systems, and the final mixing result is restricted.

We remark that although the hardcore model in bipartite graphs is a monotone system, Theorem 6 is not a direct consequence from Theorem 15. We need to tweak the proof of Theorem 15 to prove Theorem 6. The proof of Theorem 6 is in Section 8. See Section 2.4 for a proof overview.

2.3 Establish Coupling Independence

The next question is how to establish the coupling independence condition for spin systems. Previously, spectral independence was known for many spin systems. The coupling independence was often a by-product result when proving spectral independence. Hence, it is known for some specific spin systems such as subgraph world [CZ23], b-matching [CG24] and coloring in high girth graphs [CLMM23].

In this paper, we give a tool to turn many existing spectral independence results into coupling independence results. A large family of spin systems is 2-spin systems. Let $G = (V, E)$ be a graph with maximum degree $\Delta \geq 3$. Let $0 \leq \beta \leq \gamma$ be the edge interactions such that $\gamma > 0$. Let $\lambda > 0$ be the external field. Let μ be the Gibbs distribution on G with parameters β, γ, λ such that for any $\sigma \in \{-, +\}$, $\mu(\sigma) \propto \lambda^{n_+(\sigma)} \beta^{m_+(\sigma)} \gamma^{m_-(\sigma)}$, where $n_+(\sigma)$ is the number of vertices v with $\sigma_v = +$ and $m_\pm(\sigma)$ is the number of edges $\{u, v\}$ with $\sigma_u = \sigma_v = \pm$. The 2-spin system is said to be ferromagnetic if $\beta\gamma > 1$ and anti-ferromagnetic if $\beta\gamma < 1$.

Anari, Liu, and Oveis Gharan [ALO20] analyzed the spectral independence for the hardcore model. Chen, Liu, and Vigoda [CLV20] extended the analysis to general 2-spin systems. Recall the influence matrix Ψ_μ^τ is defined in (4). The maximum eigenvalue $\text{Eig}_{\max}(\Psi_\mu^\tau)$ can be upper bound by the total influence from one vertex

$$\text{Eig}_{\max}(\Psi_\mu^\tau) \leq \max_v \sum_{u \in V} |\Psi_\mu(v, u)|. \tag{8}$$

The RHS is called the *total influence bound*. For 2-spin systems, the analysis is performed on the Self-Avoiding-Walk (SAW) tree [Wei06]. Roughly speaking, fix a vertex v, the SAW tree T_v enumerates all the SAWs in graph G starting from v. By defining a proper 2-spin system on T_v, one can use the total influence from the root in T_v to upper bound the total influence from v in G, and thus establish the spectral independence for Gibbs distribution μ. In [CLV20], a weighted version of (8) is studied to deal with general 2-spin systems. We give the following result for coupling independence.

Theorem 16 (Informal version of Lemma 39)**.** *For 2-spin systems, the (weighted) total influence bound in the Self-Avoid-Walk tree implies coupling independence.*

As a consequence, all the spectral independence results for 2-spin systems in [CLV20] can be turned into coupling independence results in black-box. For the hardcore model in bipartite graphs(Theorem 5 and Theorem 6), we can also use the above result to transform the total influence bound in [CLY23] into coupling independence result.

Theorem 16 is proved by constructing a recursive coupling in Section 7. Fix a vertex v in G. We build a coupling (X, Y) between μ^{v^+} and μ^{v^-} and show the discrepancy between X and Y are bounded by the total influence in the SAW tree T_v. Suppose v has d neighbors $u_1, u_2, \ldots, u_d$. We split v into d copies $v_1, v_2, \ldots, v_d$ such that v_i only has one neighbor u_i. Define the pinning σ_i such that v_j for $j \geq i$ takes the value $+$ and v_j for $j < i$ takes the value $-$. Then $\mu^{v^+} = \mu^{\sigma_0}$ and $\mu^{v^-} = \mu^{\sigma_d}$. We couple each adjacent $\mu^{\sigma_{i-1}}$ and μ^{σ_i}, then merge them into a coupling between two endpoints μ^{σ_0} and μ^{σ_d}. For each adjacent pair, the only difference between σ_i and σ_{i-1} is the pinning at v_i. Hence, we first couple the only neighbor u_i of v_i then construct the coupling recursively if the coupling at u_i fails. This recursive processing essentially enumerates all SAWs from v. We can relate the coupling with the SAW tree to prove the theorem.

For multi-spin systems such as list-coloring, we can mimic the recursive coupling for 2-spin systems. Since the previous spectral independence results for list-coloring were also obtained via the SAW tree [FGYZ21, CGŠV21], a similar proof gives the coupling independence.

2.4 Proof Overview

We give a proof overview for the relaxation time comparison result in Theorem 9. Let $G = (V, E)$ be a graph with maximum degree Δ. Let ℓ and k be two constant integers such that $\ell < k$. Their specific values will be fixed later. We first partition all the vertices in V into k parts $U_1, U_2, \ldots, U_k$ such that for any vertex $v \in V$, each U_i does not have more than $\frac{\eta}{\ell}\Delta$ neighbors of v. In other words, let $\Gamma_v = \{u \mid (u, v) \in E\}$ denote the set of neighbors of v in graph G. For any $i \in [k]$, $|\Gamma_v \cap U_i| \leq \frac{\eta}{\ell}\Delta$. The existence of the partition is guaranteed by the Lovász local lemma. However, the local lemma requires the maximum degree Δ to be sufficiently large. That is why we require a lower bound for Δ in our technical results. We also remark that in our proof, the degree lower bound is used solely to ensure the existence of the partition. A similar partition appeared in the previous work [JSS21].

The input Gibbs distribution μ over $[q]^V$ is a joint distribution of n variables $(X_v)_{v \in V}$, where each variable takes its value from $[q]$. Now, we can view μ as a joint distribution of k variables $Y = (Y_i)_{i \in [k]}$ such that each variable $Y_i = X_{U_i}$ takes its value from a huge domain $[q]^{U_i}$. We define the $k \leftrightarrow (k - \ell)$ down-up walk on Y. Given $Y = (Y_1, Y_2, \ldots, Y_k)$, the Markov chain does as follows

- Down-walk: Sample a set $S \in \binom{[k]}{\ell}$ of ℓ indices uniformly at random and then remove the configuration on the set S: $Y \to Y_{[k]\setminus S}$;
- Up-walk: Resample Y_S from μ conditional on $Y_{[k]\setminus S}$ and then go back to a full configuration $Y_{[k]\setminus S} \to Y_{[k]\setminus S} \cup Y_S$.

A full configuration $Y = (Y_1, Y_2, \ldots, Y_k)$ is on the level k. In the down-walk, we sample a random subset of indices $S \subseteq [k]$ with size ℓ. By dropping the configuration Y_S, we move from a full configuration at level k to a partial configuration at level $k - \ell$. In the up-walk, we resample Y_S and go back to the level k. The process can also be viewed as a kind of block dynamics for configuration $X \in [q]^V$. In every step, we pick a random subset $U_S = \cup_{i \in S} U_i \subseteq V$ of variables and resample $X(U_S)$ conditional on $X(V \setminus U_S)$.

We use local-to-global technique [AL20, ALO20] to analyze the spectral gap of the $k \leftrightarrow (k - \ell)$ down-up walk for Y. The local-to-global technique suggests to analyze the relaxation time of $k \leftrightarrow 1$

down-up walk[2]. In the down walk, we pick a random S of size $|S| = k-1$ and drop Y_S. In the up-walk, we resample Y_S and go back to level k. We use coupling independence to analyze this $k \leftrightarrow 1$ down-up walk via path coupling. For simplicity, suppose μ satisfies C-coupling independence with standard Hamming distance ($\rho(v) = 1$ for all $v \in V$). We can view this $k \leftrightarrow 1$ down-up walk on Y as a block dynamics on $X \in [q]^V$, where it updates a block U_S in every step. Given two $X \in [q]^V$ and $X' \in [q]^V$ that disagree only at one vertex $v \in V$, say $v \in U_1$, we couple the transition of $k \leftrightarrow 1$ down-up walk. Let two $k \leftrightarrow 1$ down-up walks (starting from X and X', respectively) select the same random subset $S \subseteq [k]$ such that $|S| = k-1$.

- If $1 \in S$, which happens with probability $\frac{k-1}{k}$, then since $v \in U_1$ the value of v is removed in the down-walk, and thus X and X' can be coupled perfectly after the transition.
- If $1 \notin S$, which happens with probability $\frac{1}{k}$, then since $v \in U_1$, the disagreement at v may percolate to other blocks in the up-walk step. We use the coupling in the M-coupling independence to couple the up-walk so that the expected Hamming distance between X and X' after the transition is at most M.

Hence, the expected Hamming distance between X and X' after transition is at most $\frac{M}{k}$. If $k > M$, the path coupling gives the $O(\log n)$ mixing time and $O(1)$ relaxation time for this down-up walk. To apply the local-to-global technique, we also need to fix a configuration Y_Λ, where $\Lambda \subseteq [k]$ and $|\Lambda| = t \le \ell$, and consider the $(k-t) \leftrightarrow 1$ down-up walk for $Y_{[k]\setminus\Lambda}$. The same path coupling works if $k - t > M$. By choosing k and ℓ such that $k - \ell > M$ and using the local-to-global technique, we can show that the $k \leftrightarrow (k-\ell)$ down-up walk for Y has $O(1)$ relaxation time.

We then compare the $k \leftrightarrow (k-\ell)$ down-up for $Y = (Y_1, Y_2, \dots, Y_k)$ to the Glauber dynamics for $X \in [q]^V$. Recall that $k \leftrightarrow (k-\ell)$ down-up walk is a kind of block dynamics for X. In every step, the block dynamics updates a subset $U_S = \cup_{i \in S} U_i$ with $|S| = \ell$. The update step is to resample $X(U_S)$ conditional on $X(V \setminus U_S)$. This step samples from the conditional Gibbs distribution $\mu_{U_S}^{X(V\setminus U_S)}$ on subgraph $G[U_S]$. By the construction of the partition, the maximum degree of $G[U_S]$ is at most $\eta\Delta$ so that we have the relaxation time bound $T_{\mathrm{rel}}^{(\eta)}(\mu)$ for Glauber dynamics on $\mu_{U_S}^{X_{V\setminus U_S}}$. Let $T_{\mathrm{rel}}^{\text{down-up}}$ denote the relaxation time of $k \leftrightarrow (k-\ell)$ down-up walk. By some standard comparison argument between block dynamics and the Glauber dynamics, we can prove Theorem 9 by showing that

$$T_{\mathrm{rel}}^{\mathrm{GD}}(\mu) \le T_{\mathrm{rel}}^{\text{down-up}} \times T_{\mathrm{rel}}^{(\eta)}(\mu) = O(1) \times T_{\mathrm{rel}}^{(\eta)}(\mu).$$

Next, we briefly explain how to get the near-linear time sampling algorithm in Theorem 13 and the mixing time in Theorem 15. Note that for $k \leftrightarrow 1$ down-up walk, the path coupling actually gives the $O(\log n)$ mixing time. For one update step, it selects a subset $S \subseteq [k]$ with $|S| = k-1$. Let i denote the missing index, i.e. $S \cup \{i\} = [k]$. The update step resamples Y_S conditional on Y_i. We can simulate this transition step using $(k-1) \leftrightarrow 1$ down-up walk for the conditional distribution on Y_S. This down-up walk also has the $O(\log n)$ mixing time. We do this recursively until we need to sample from a conditional distribution on $Y_{S'}$ with $|S'| = \ell$. Note that the maximum degree of the graph $G[U_{S'}]$ is at most $\eta\Delta$. Now, we simulate the Glauber dynamics for $T_{\mathrm{mix}}^{(\eta)}(\mu)$ steps to sample from the conditional distribution. Hence, the total number of Glauber steps is $(\log n)^{O(\ell)} T_{\mathrm{mix}}^{(\eta)}(\mu)$. For monotone systems, we can compare this algorithm to Glauber dynamics via censoring inequality.

[2] In [AL20, ALO20], the local walk is essentially defined as the $1 \leftrightarrow k$ up-down walk. Every state is Y_i for $i \in [k]$. In the up-walk, it extends Y_i to a full configuration Y. In the down-walk step, it samples a random index $j \in [k]$ and updates Y to Y_j. It is well-known that $1 \leftrightarrow k$ up-down walk and $k \leftrightarrow 1$ up-down walk has the same relaxation time.

The results for list-coloring are consequences of general technical results. However, we need to tweak the analysis to prove the results for hardcore model in bipartite graphs (Theorem 5 and Theorem 6). The reason is that for hardcore model on $G = (V_L \cup V_R, E)$, we only know $\lambda < \lambda_c(\Delta_L)$ but we cannot control the degree Δ_R in the right part V_R. Our technique can only prove the coupling independence for μ_L, which is the marginal distribution on V_L projected from μ. To prove the relaxation time and mixing time results, we first partition V_L into disjoint part $U_1, U_2, \ldots, U_k$ such that for any vertex $v \in V_R$, v has no more than $\frac{\eta}{\ell}\Delta$ neighbors in each U_i. Again, the existence of the partition is guaranteed by the local lemma. Let $X \sim \mu_L$ be a partial configuration on V_L. We can define $Y = (Y_1, Y_2, \ldots, Y_k)$, where $Y_i = X_{U_i}$. By a similar proof, we show that the $k \leftrightarrow (k-\ell)$ down-up walk for Y is rapid mixing. We consider a global Markov chain over $\{-,+\}^{V_L \cup V_R}$ defined as follows. Let $\overline{X} \in \{-,+\}^{V_L \cup V_R}$ be a full configuration.

- Drop the right part to obtain $X \leftarrow \overline{X}(V_L)$;
- Update X using the $k \leftrightarrow (k-\ell)$ down-up walk for $Y = (Y_1, \ldots, Y_k)$, where $Y_i = X(U_i)$;
- Sample $X(V_R) \sim \mu^X_{V_R}$ and let $\overline{X} \leftarrow X \cup X(V_R)$.

We first compare this chain to the $k \leftrightarrow (k-\ell)$ down-up walk and then compare the Glauber dynamics for μ to this Markov chain. This gives the relaxation time of Glauber dynamics. For the mixing time, we can first obtain a near-linear time sampling algorithm for μ_L, since hardcore model in bipartite graph is a monotone system, we then compare the algorithm to the Glauber dynamics for μ via censoring inequality.

3 Preliminaries

3.1 ϕ-Divergences and ϕ-Entropies

Let $\phi : D \to \mathbb{R}$ be a convex function with domain $D \subseteq R$. Let μ be a distribution over a finite set Ω. For any random variable $f : \Omega \to D$, the ϕ-entropy of f with respect to μ is defined as

$$\mathbf{Ent}^{\phi}_{\mu}[f] := \mathbb{E}_{\mu}[\phi(f)] - \phi(\mathbb{E}_{\mu}[f]). \tag{9}$$

Note that $\mathbf{Ent}^{\phi}_{\mu}[f] \geq 0$ follows directly from the Jensen's inequality since ϕ is a convex function. In particular, when $D \supseteq R_{\geq 0}$, the notion of ϕ-entropy can be used to measure the distance between distributions. Let ν and μ be distributions on Ω such that ν is absolutely continuous with respect to μ, then the ϕ-divergence $\mathcal{D}_{\phi}(\nu \parallel \mu)$ between ν and μ is defined as

$$\mathcal{D}_{\phi}(\nu \parallel \mu) := \mathbf{Ent}^{\phi}_{\mu}\left[\frac{\nu}{\mu}\right]. \tag{10}$$

In practice, typical choice of the function ϕ are

- $\mathrm{TV}(x) = \frac{1}{2}|x-1|$: this defines the *TV-distance* $\mathcal{D}_{\mathrm{TV}}(\nu \parallel \mu) = \frac{1}{2}\sum_{x\in\Omega}|\nu(x) - \mu(x)|$;
- $\chi^2(x) = x^2$: this defines the *χ^2-divergence* $\mathcal{D}_{\chi^2}(\nu \parallel \mu) = (\sum_{x\in\Omega}\nu(x)\cdot\frac{\nu(x)}{\mu(x)}) - 1$;
- $\mathrm{KL}(x) = x\log x$: this defines the *KL-divergence* $\mathcal{D}_{\mathrm{KL}}(\nu \parallel \mu) = \sum_{x\in\Omega}\nu(x)\log\frac{\nu(x)}{\mu(x)}$.

By conventions, the χ^2-entropy is usually called variance (with the notation $\mathbf{Var}_\mu[f]$) and the KL-entropy is usually called entropy (with the notation $\mathbf{Ent}_\mu[f]$):

$$\mathbf{Var}_\mu[f] = \mathbb{E}_\mu[f^2] - \mathbb{E}_\mu[f]^2 \tag{11}$$
$$\mathbf{Ent}_\mu[f] = \mathbb{E}_\mu[f \log f] - \mathbb{E}_\mu[f] \log \mathbb{E}_\mu[f]. \tag{12}$$

3.2 Markov Chain Background

Definition 17. Let Ω and Ω' be two finite sets. A *Markov kernel* P from Ω to Ω' assigns to every element $x \in \Omega$ a distribution $P(x, \cdot)$ on Ω'. In particular, P could be seen as a matrix in $\mathbb{R}_{\geq 0}^{\Omega \times \Omega'}$.

When $\Omega = \Omega'$, the Markov kernel P becomes the *transition matrix* of some Markov chain $(X_t)_{t \geq 0}$. We use P to refer to this Markov chain if it is clear in the context. The Markov chain P is

- *irreducible*, if for any $x, y \in \Omega$, there is a $t > 0$ such that $P^t(x, y) > 0$;
- *aperiodic*, if for any $x \in \Omega$, it holds that $\gcd\{x \mid P^t(x, x) > 0\} = 1$.

A distribution μ on Ω is called the stationary distribution of P if $\mu P = \mu$. If a Markov chain is both irreducible and aperiodic, then it has a unique stationary distribution.

Definition 18 (time-reversal)**.** Let Ω and Ω' be two finite sets. The *time-reversal* P^* of a Markov kernel P from Ω to Ω' with respect to μ is defined by the following *detailed balanced equation*:

$$\forall x \in \Omega, y \in \Omega', \quad \mu(x)P(x, y) = \mu^*(y)P^*(y, x), \tag{13}$$

where $\mu^* = \mu P$ is the corresponding distribution on Ω'.

In particular, let P be a Markov chain on Ω. Let P^* be its time-reversal with respect to μ. Then P is called *reversible* with respect to μ if $P = P^*$. This implies μ is a stationary distribution of P.

Let P be a Markov chain on Ω with the unique stationary distribution μ. The mixing time of Glauber dynamics is defined in (1). It can be defined similarly for a general Markov chain P.

In this paper, we are particularly interested in the block dynamics defined as follows.

Block dynamics Let μ be a distribution over $[q]^V$, not necessarily Gibbs distribution. Let $\mathcal{B} = \{B_1, B_2, \ldots, B_\ell\}$ a set of blocks, where $B_i \subseteq V$. Define the following block dynamics. Given any $X \in \Omega(\mu)$, the block dynamics does as follows

- down-walk D: sample $i \in [\ell]$ uniformly at random and let $X \to X_{V \setminus B_i}$;
- up-walk U: sample $X_{B_i} \sim \mu_{B_i}^{X_{V \setminus B_i}}$ and extend $X_{V \setminus B_i} \to X_{B_i} \cup X_{V \setminus B_i}$.

Here, we decompose the block dynamics into two steps, the down-walk D and the up-walk U. Let $\Omega = \Omega(\mu)$ be the set of full configurations. Let $\Omega^* = \{X_{V \setminus B_i} \mid X \in \Omega \wedge 1 \leq i \leq \ell\}$ be a set of partial configurations. The down-walk $D : \Omega \times \Omega^* \to \mathbb{R}_{\geq 0}$ goes from a full configuration to a random partial configuration. And the up-walk $U : \Omega^* \times \Omega \to \mathbb{R}_{\geq 0}$ goes from a partial configuration to a random full configuration. We view both D and U as transition matrices. Moreover, $U = D^*$ is the time-reversal of D with respect to μ. Let $P = DU$ be the composition of D and U. P is the transition matrix of the block dynamics and P is time-reversible with respect to μ. In particular, if each B_i only contains a single variable, then the block dynamics above is exactly the Glauber dynamics.

In order to investigate the relaxation time of the block dynamics. We primarily use the notion of approximate tensorization [MSW03, CMT15] and block factorization of variance [CP21]. Recall

μ is a distribution over $[q]^V$. We use $\Omega(\mu)$ to denote the support of μ. Let $f : \Omega(\mu) \to \mathbb{R}$ be a function. And $X \sim \mu$ be a random sample. We will use $\mathbb{E}_\mu[f]$ to denote $\mathbb{E}[f(X)]$. Recall that the variance of f is the χ^2-entropy of f defined as $\mathbf{Var}_\mu[f] = \mathbf{Var}[f(X)] = \mathbb{E}\left[f^2(X)\right] - \mathbb{E}[f(X)]^2$.

For any subset $S \subseteq V$, define

$$\mathbf{Var}_S[f] := \mathbf{Var}\left[f(X) \mid X_{V\setminus S}\right]$$
$$\text{and} \quad \mu[\mathbf{Var}_S[f]] := \mathbb{E}\left[\mathbf{Var}\left[f(X) \mid X_{V\setminus S}\right]\right]$$

as a shorthand. We note that $\mathbf{Var}_S[f]$ is a random variable $\Omega(\mu) \to \mathbb{R}$. Given a sample $x \in \Omega(\mu)$, the value of this random variable is given by

$$\mathbf{Var}_S^x[f] := \mathbf{Var}\left[f(X) \mid X_{V\setminus S}\right](x) = \mathbf{Var}\left[f(X) \mid X_{V\setminus S} = x_{V\setminus S}\right].$$

Definition 19. Let $\mathcal{B} = \{B_1, B_2, \ldots, B_\ell\}$ a set of blocks, where $B_i \subseteq V$. We say μ satisfies *$\mathcal{B}$-block factorization of variance* with parameter C if the following inequality holds for every $f : \Omega(\mu) \to \mathbb{R}$:

$$\mathbf{Var}_\mu[f] \le \frac{C}{\ell} \sum_{B\in\mathcal{B}} \mu[\mathbf{Var}_B[f]].$$

In particular, if every B_i is a single variable $v \in V$ and $\ell = |V|$, then we say μ satisfies *C-approximate tensorization of variance*.

The block factorization of variance is closely related to the contraction of the down-walk D on the χ^2-divergence. We formally define it as follow.

Definition 20. We say the down-walk of the block dynamics has *δ-contraction for χ^2-divergence* if for any distribution ν over $\Omega(\mu)$ such that ν is absolutely continuous with respect to μ,

$$\mathcal{D}_{\chi^2}(\nu D \parallel \mu D) \le (1-\delta) D_{\chi^2}(\nu \parallel \mu).$$

It is well-known that the relaxation time, block factorization and the contraction of the down-walk are equivalent. We summarize this into the following proposition.

Proposition 21. *The following three properties are equivalent for block dynamics:*

1. *the relaxation time of block dynamics is T;*
2. *μ satisfies $\mathcal{B}$-block factorization of variance with parameter T;*
3. *the down-walk has $1/T$-contraction for χ^2-divergence.*

For the equivalence of Item 1 and Item 2, we refer the readers to [CMT15, Cap23]. And for the equivalence of Item 2 and Item 3, we refer the readers to [CLV21, Lemma 2.7].

3.3 Coupling and Path Coupling

Let (X, Y) be two jointly distributed random variables such that $X \sim \nu$ and $Y \sim \mu$. Then, (X, Y) is called a *coupling* of ν and μ. The TV-distance between two distribution ν and μ over Ω can be bounded by coupling. The following result is standard and is called the coupling lemma. We refer the readers to [LPW17, Proposition 4.7] for a proof.

Lemma 22 (coupling lemma)**.** *Let ν and μ be two probability distributions on Ω. Then*

$$\mathcal{D}_{\mathrm{TV}}(\nu \parallel \mu) = \inf\{\mathbf{Pr}[X \neq Y] \mid (X, Y) \textit{ is a coupling of } \nu \textit{ and } \mu\}.$$

There exists an optimal coupling of ν and μ that achieves the infimum.

Coupling can be used to bound both mixing time and relaxation time of a Markov chain. Let $\rho : V \to \mathbb{N}_{>0}$ be the Hamming distance weight in Definition 7. Let $\Omega \subseteq [q]^V$. Let P be the transition matrix of a Markov chain on Ω. Suppose there is a $\delta \in (0,1)$ that for each $x, y \in \Omega$, there is a coupling (X_1, Y_1) of $P(x,\cdot)$ and $P(y,\cdot)$ satisfying

$$\mathbb{E}\left[H_\rho(X_1, Y_1)\right] \leq (1-\delta) H_\rho(x,y), \tag{14}$$

where H_ρ is the weighted Hamming distance in (3). For convenience, let $\rho_{\max} := \max_{v\in V} \rho_v$ and $\rho_{\min} := \min_{v\in V} \rho_v$. Then it is standard to have the following results.

Lemma 23. *If (14) holds, then* $T_{\mathrm{mix}}(P,\varepsilon) \leq \frac{1}{\delta}\left(\log \frac{\rho_{\max}}{\rho_{\min}} + \log \frac{1}{\varepsilon}\right)$.

Lemma 24. *If (14) holds and the matrix P is positive semi-definite, then the relaxation time is bounded by*

$$T_{\mathrm{rel}}(P) \leq 1/\delta.$$

We refer the reader to [BD97] for a proof of Lemma 23 and [LPW17, Theorem 13.1] for a proof of Lemma 24. In particular, all the Markov chains considered in this paper are down-up walks (i,e. block dynamics). And it is well-known that down-up walks are positive semi-definite [AL20, Section 2.4]. Hence, when (14) holds, we can directly use Lemma 24 to bound the relaxation time of the down-up walk and block dynamics. Finally, the path coupling is a tool to construct the coupling of the Markov chain.

Lemma 25 ([BD97])**.** *Let $P : [q]^V \times [q]^V \to \mathrm{R}_{\geq 0}$ be a Markov chain. If for any $x, y \in [q]^V$ such that x and y disagree only at a single vertex v, there exists a coupling $(x,y) \to (X_1, Y_1)$ of P such that*

$$\mathbb{E}\left[\rho(X_1, Y_1)\right] \leq (1-\delta)\rho(x,y) = (1-\delta)\rho(v_0),$$

then for any $x, y \in [q]^V$, there exists a coupling satisfying (14).

With path coupling, one only needs to construct coupling for adjacent x, y rather than all x, y. The path coupling requires that P is defined over $[q]^V \times [q]^V$ rather than $\Omega \times \Omega$, where $\Omega \subseteq [q]^V$. In this paper, we only consider block dynamics. It is easy to extend the block dynamics to $[q]^V \times [q]^V$ using the conditional distribution defined in (2).

3.4 Lovász Local Lemma and Algorithmic Local Lemma

The Lovasz local lemma [EL75] is used to prove the existence of a combinatorial object. In this paper, we will focus on its symmetric and algorithmic version [MT10].

Let $\mathcal{E}_1, \cdots, \mathcal{E}_n$ be a set of bad events in some probability space. We want to show that there is a sample in the probability space that is not included in any bad events. A *dependency graph* for $\mathcal{E}_1, \cdots, \mathcal{E}_n$ is a graph $G = ([n], E)$ such that for each $i \in [n]$, event $\mathcal{E}_i$ is mutually independent of the events $\{\mathcal{E}_j \mid (i,j) \notin E\}$.

Lemma 26 (Lovász local lemma, [EL75, MT10])**.** *Let $p = \max_i \mathbf{Pr}[\mathcal{E}_i]$ and Δ be the degree of dependency graph for $\mathcal{E}_1, \cdots, \mathcal{E}_n$. If $\mathrm{e}p(\Delta+1) \leq 1$, then it holds that*

$$\mathbf{Pr}\left[\bigcap_{i=1}^{n} \overline{\mathcal{E}_i}\right] > \left(1 - \frac{1}{\Delta+1}\right)^n,$$

where we use $\overline{\mathcal{E}_i}$ to denote the negation of $\mathcal{E}_i$.

In particular, if all the events $\mathcal{E}_1, \cdots, \mathcal{E}_n$ are determined by a set of mutually independent random variables $X_1, \cdots X_m$, then there is a Las Vegas algorithm that runs in time $\frac{np}{1-p}$ in expectation. When it halts, it will output an assignment $\boldsymbol{x} = (x_1, \cdots, x_m)$ of X_is such that event $\mathcal{E}_j$ does not holds under $\boldsymbol{x}$ for all j.

4 Proof of Relaxation Time Comparison Result

In this section, we will prove Theorem 9. Let μ be a Gibbs distribution on graph $G = (V, E)$. The proof works whenever the graph G has a good partition. We give it a formal definition as follow.

Definition 27 ((ξ, k)-degree partition)**.** Let $k \geq 1$ be an integer and $\xi > 0$. A graph $G = (V, E)$ is said to have a (ξ, k)-degree partition if there exists a partition $V = U_1 \uplus U_2 \uplus \ldots \uplus U_k$ such that

$$\forall 1 \leq i \leq k, \forall v \in V, \quad |\Gamma_v \cap U_i| \leq \frac{(1+\xi)\Delta}{k}.$$

We note that a similar but simpler definition appeared in [JSS21], where they partition the graph into 2 parts with certain degree constrains.

The following result states that we can do comparison between relaxation time of Glauber dynamics given a good partition and M-coupling-independence.

Theorem 28. *Let μ be a Gibbs distribution on $G = (V, E)$ such that:*

- *μ satisfies M-coupling-independence for some integer $M \geq 1$;*
- *G has a (ξ, k)-degree partition such that $k \geq 2M$.*

Then, the relaxation time of Glauber dynamics on μ satisfies

$$T_{\mathrm{rel}}^{\mathrm{GD}}(\mu) \leq 2^{k-2M} \cdot T_{\mathrm{rel}}^{(\eta)}(\mu), \quad \text{such that } \eta = 2(1+\xi)M/k.$$

Note that Theorem 28 only need the existence of such partition. It does not need the explicit construction of the partition. When the maximum degree Δ of G is sufficiently large, the existence of (ζ, k)-partition could be proved by the Lovász local lemma.

Proposition 29. *Let $k \geq 1$ be an integer and $\xi \in (0, 1)$. For any graph $G = (V, E)$ with maximum degree $\Delta \geq \Delta_0(k, \xi) = \Omega(\frac{k^2}{\xi^2} \log k)$ has a (ξ, k)-degree partition.*

We defer the proof of Proposition 29 to Section 4.3. Now, we are ready to prove Theorem 9.

Proof of Theorem 9. For simplicity, we just fix $\xi = 1$. Note that M-coupling-independence directly implies $\lceil M \rceil$-coupling-independence. We take $k = \lceil 4 \lceil M \rceil / \eta \rceil$, which implies $k > 2 \lceil M \rceil$ automatically. By Proposition 29, when $\Delta \geq \Delta_0(k, 1) = \Omega(k^2 \log k) = \Omega(\frac{M^2}{\eta^2} \log \frac{M}{\eta})$, graph G has a $(1, k)$-partition. We make a summary of what we have as follow:

- μ is $\lceil M \rceil$-coupling-independence;
- G has $(1, k)$-degree partition and $k \geq 2 \lceil M \rceil$.

Then, by Theorem 28, we have

$$T_{\mathrm{rel}}^{\mathrm{GD}}(\mu) \leq 2^{k-2\lceil M \rceil} \cdot T_{\mathrm{rel}}^{(4\lceil M \rceil / \lceil 4 \lceil M \rceil / \eta \rceil)}(\mu) \leq 2^{O(M/\eta)} \cdot T_{\mathrm{rel}}^{(\eta)}(\mu),$$

where in the last inequality, we use the fact that $k = O(M/\eta)$ and $4 \lceil M \rceil \leq \lceil 4 \lceil M \rceil / \eta \rceil \cdot \eta$. □

Now, we only left to prove Theorem 28. Let μ be a Gibbs distribution on graph $G = (V, E)$. Suppose G has a partition $V = U_1 \uplus U_2 \uplus \ldots \uplus U_k$. We note that for $1 \leq i \leq k$, it is possible that U_i is empty. Let $1 \leq \ell \leq k$ be an integer. Consider the following $k \leftrightarrow \ell$ down-up walk defined on the partition. It starts from an arbitrary $X \in \Omega(\mu)$ and it does as follows in each step:

- pick a subset $R \subseteq [k]$ with $|R| = \ell$ uniformly at random;
- resample $X_{V\setminus U_R} \sim \mu_{V\setminus U_R}(\cdot \mid X_R)$, where $U_R = \cup_{i\in R} U_i$.

In the down walk, it picks a subset $R \subseteq [k]$ of size ℓ and remove the configuration at $V \setminus U_R$. In the up walk, it resamples the configuration on $V \setminus U_R$ conditional on the configuration on U_R. Let $T_{\mathrm{rel}}(k, \ell)$ denote the relaxation time of the above $k \leftrightarrow \ell$ down-up walk. This down-up walk is a *block dynamics* for the Gibbs distribution μ. In each step, it picks random block U_R and update the values on other variables conditional on U_R.

The following result bound the relaxation time of the $k \leftrightarrow (k - 2M)$ down-up walk.

Lemma 30. *Let $M \geq 1$ and $k \geq 2M$ be two integers. For any Gibbs distribution μ on graph G, if μ satisfies M-coupling independence, then for any partition $U_1, \cdots U_k$ of G, the relaxation time of $k \leftrightarrow (k - 2M)$ down-up walk satisfies $T_{\mathrm{rel}}(k, k - 2M) \leq 2^{k-2M}$.*

Lemma 30 is proved in Section 4.1. It follows from the standard local-to-global paradigm. However, unlike the usual situation, the "local" part here is actually non-local in the sense that it may have exponential sized *influence matrix* or *correlation matrix*. We surpass this obstacle by working with *coupling independence* instead of standard *spectral independence* condition in [ALO20].

Then, according to standard comparison argument between the block dynamics and the Glauber dynamics, we can bound the relaxation time of the Glauber dynamics by the relaxation time of the block dynamics. Taking into account that the block dynamics actually runs on a (ξ, k)-partition, we have the following result which is proved in Section 4.2.

Lemma 31. *For any Gibbs distribution μ on graph G, if G has a (ξ, k)-partition $U_1, \cdots, U_k$. Then for any $0 \leq \ell \leq k - 1$, the relaxation time of Glauber dynamics is bounded by*

$$T^{\mathrm{GD}}_{\mathrm{rel}}(\mu) \leq T_{\mathrm{rel}}(k, \ell) \cdot T^{(\eta(\ell))}_{\mathrm{rel}}(\mu), \quad \textit{where } \eta(\ell) = (1 + \xi) \cdot \frac{k - \ell}{k},$$

and $T_{\mathrm{rel}}(k, \ell)$ denotes the relaxation time of the $k \leftrightarrow \ell$ down-up walk on the partition $U_1, \cdots, U_k$.

Now, we are ready to prove Theorem 28.

Proof of Theorem 28. Pick $\ell = k - 2M$ and combine Lemma 30 and Lemma 31. □

4.1 Relaxation Time of Block Dynamics via Coupling Independence

In this section, we prove Lemma 30. To do so, we introduce some auxiliary Markov chains to help us analyze the relaxation time of $k \leftrightarrow (k - 2M)$ down up walk. Let $R \subseteq [k]$ with size $|R| = r$. For any $\tau \in [q]^{U_R}$, we define the $(k - r) \leftrightarrow 1$ down up walk on $\mu^{\tau}_{V\setminus U_R}$. The chain starts from arbitrary $Y \in \Omega(\mu^{\tau}_{V\setminus U_R})$. In each step, it does as follows

- pick an index $j \in [k] \setminus R$ uniformly at random and let $\Lambda = V \setminus R \setminus U_j$;
- resample $Y_\Lambda \sim \mu_\Lambda(\cdot \mid Y_{U_j}, \tau)$.

In words, in this Markov chain, the configuration on R is fixed by τ and the configuration on $V \setminus U_R$ is free. Hence, we have $k - r$ levels in total. In the down walk, it picks one index j and remove the configurations on all $V \setminus U_R$ except U_j, and thus the chain goes to level 1. In the up walk, it goes back to the level s by sampling a random configuration on Λ conditional on Y_{U_j} and τ. Again, let $T^{\tau}_{\mathrm{rel}}(k - r, 1)$ denote the relaxation time of $(k - r) \leftrightarrow 1$ down-up walk conditioning on τ.

Proposition 32 ([AL20]). *Let $0 \le \ell \le k-1$. If there exists $\gamma_0, \gamma_1, \dots, \gamma_{\ell-1}$ such that for any $0 \le r \le \ell - 1$, any $R \subseteq [k]$ with $|R| = r$, any $\tau \in [q]^{U_R}$, $T^{\tau}_{\mathrm{rel}}(k-r, 1) \le \gamma_r$, then the $k \leftrightarrow \ell$ down-up walk satisfies*

$$T_{\mathrm{rel}}(k, \ell) \le \prod_{i=0}^{\ell-1} \gamma_i.$$

The proof of Proposition 32 follows from standard local to global argument in [AL20]. We give a simple proof in Appendix A for the completeness.

Finally, the relaxation time bound for $(k-r) \leftrightarrow 1$ down-up walk could be established from coupling independence via path coupling.

Lemma 33. *If μ satisfies M-coupling-independent, then for any $0 \le r \le k - 2M$, any $R \subseteq [k]$ with $|R| = r$, and any pinning $\tau \in [q]^{U_R}$, it holds that*

$$T^{\tau}_{\mathrm{rel}}(k-r, 1) \le \gamma_r = 2.$$

Proof. We use path coupling to analyze the $(k-r) \leftrightarrow 1$ down-up walk. For the simplicity of notation, we denote $\overline{U_R} = V \setminus U_R$. Fix $R \subseteq [k]$ and $\tau \in [q]^{U_R}$. Note that $\mu^{\tau}_{\overline{U_R}}$ is a distribution over $[q]^{\overline{U_R}}$. One may assume $\overline{U_R}$ is not empty. Otherwise, the lemma is trivial.

Fix two configurations X and Y in $[q]^{\overline{U_R}}$ (not necessarily feasible in $\mu^{\tau}_{\overline{U_R}}$) that differs only at single vertex v_0. We couple the transition $(X, Y) \to (X', Y')$ of $(k-r) \leftrightarrow 1$ down-up walk as follows

- sample the same $j \in [k] \setminus R$, construct the same $\Lambda = \overline{U_{R \cup \{j\}}}$;
- if $v_0 \notin U_j$, let $X'_{U_j} = Y'_{U_j} = X_{U_j} = Y_{U_j}$ and perfectly couple X'_Λ and Y'_Λ;
- if $v_0 \in U_j$, use the coupling in Definition 7 to sample (X'', Y'') where $X'' \sim \mu^{\tau \wedge X_{U_j}}$ and $Y'' \sim \mu^{\tau \wedge Y_{U_j}}$ and let $X' = X''_{\overline{U_R}}$ and $Y' = Y''_{\overline{U_R}}$.

We can apply the coupling in Definition 7 because μ is M-CI. Specifically, we apply the coupling with pinnings $\tau \wedge X_{U_j}$ and $\tau \wedge Y_{U_j}$. Both X'', Y'' returned by the coupling are full configurations on V. We take partial configurations on U to get X' and Y'. We have $H_\rho(X', Y') = H_\rho(X'', Y'')$ because $X''_{U_R} = Y''_{U_R} = \tau$. It is straightforward to verify both $X \to X'$ and $Y \to Y'$ follows the transition rule of $(k-r) \leftrightarrow 1$ down-up walk. We have the following result

$$\mathbb{E}\left[H_\rho(X', Y') \mid X, Y\right] \le \frac{M}{k-r} \cdot \rho(v_0) = \frac{M}{k-r} H_\rho(X, Y) \le \frac{1}{2} H_\rho(X, Y),$$

where the last inequality holds because $k - r \ge 2M$.

By path coupling argument, for any Z_1 and Z_2 in $[q]^U$ (Z_1 and Z_2 may differ at multi vertices), there exists a coupling $(Z_1, Z_2) \to (Z'_1, Z'_2)$ such that $\mathbb{E}\left[H_\rho(Z'_1, Z'_2) \mid Z_1, Z_2\right] \le (1 - \frac{1}{2}) H_\rho(Z_1, Z_2)$. Hence, the step-wise decay coupling implies the relaxation time $T^{\tau}_{\mathrm{rel}}(s, 1) \le 2$. □

Now, we are ready to prove Lemma 30. Combining Proposition 32 and Lemma 33,

$$T_{\mathrm{rel}}(k, k-2M) \le \prod_{i=0}^{k-2M-1} \gamma_i = 2^{k-2M}.$$

4.2 Compare Glauber Dynamics to Block Dynamics (Proof of Lemma 31)

In this section, we will prove Lemma 31. Let $0 \le \ell < k$ be the integer in theorem. Recall that G has a (ξ, k)-degree partition $V = U_1 \uplus \ldots \uplus U_k$. According to Proposition 21, the relaxation time of the $k \to \ell$ down up walk implies the block factorization of variance as follow: for any function $f : \Omega(\mu) \to \mathbb{R}$, it holds that

$$\mathbf{Var}_\mu[f] \le \frac{T_{\mathrm{rel}}(k,\ell)}{\binom{k}{k-\ell}} \sum_{S\subseteq[k]:|S|=k-\ell} \mu[\mathbf{Var}_{U_S}[f]] = \frac{T_{\mathrm{rel}}(k,\ell)}{\binom{k}{k-\ell}} \sum_{S\subseteq[k]:|S|=k-\ell} \sum_{\tau\in[q]^{V\setminus U_S}} \mu_{V\setminus U_S}(\tau)\mathbf{Var}_{\mu^\tau}[f].$$

In distribution μ^τ, consider the graph $G[U_S]$. By Definition 27, the maximum degree of the subgraph is at most $(1+\xi)\frac{k-\ell}{k}\Delta \le \eta(\ell)\Delta$. Let n be the number of vertices in V. We can apply Definition 8 on μ^τ to obtain

$$\mathbf{Var}_{\mu^\tau}[f] \le T_{\mathrm{rel}}^{(\eta(\ell))}(\mu)\frac{1}{n}\sum_{v\in V}\mu^\tau[\mathbf{Var}_v[f]] = T_{\mathrm{rel}}^{(\eta(\ell))}(\mu)\frac{1}{n}\sum_{v\in U_S}\mu^\tau[\mathbf{Var}_v[f]],$$

where the last equation holds because the values on $V \setminus U_i$ are fixed and so the variance is 0. Combining the above two inequalities together implies

$$\begin{aligned}\mathbf{Var}_{\mu^\tau}[f] &\le T_{\mathrm{rel}}^{(\eta(\ell))}(\mu)\frac{T_{\mathrm{rel}}(k,\ell)}{\binom{k}{k-\ell}}\cdot\frac{1}{n}\sum_{S\subseteq[k]:|S|=k-\ell}\sum_{v\in U_S}\mu[\mathbf{Var}_v[f]]\\ &= T_{\mathrm{rel}}^{(\eta(\ell))}(\mu)\frac{(k-\ell)T_{\mathrm{rel}}(k,\ell)}{k}\cdot\frac{1}{n}\sum_{v\in V}\mu[\mathbf{Var}_v[f]] \le T_{\mathrm{rel}}^{(\eta(\ell))}(\mu)T_{\mathrm{rel}}(k,\ell)\cdot\frac{1}{n}\sum_{v\in V}\mu[\mathbf{Var}_v[f]].\end{aligned}$$

This proves $T_{\mathrm{rel}}^{\mathrm{GD}} \le T_{\mathrm{rel}}^{(\eta(\ell))}(\mu)\cdot T_{\mathrm{rel}}(k,\ell)$.

4.3 Partition the Graph via Local Lemma (Proof of Proposition 29)

Let $G = (V, E)$ be a graph. For any vertex $v \in V$, we use Γ_v to denote the neighborhood of v.

We use the Lovász local lemma to prove the existence. For each vertex v, sample an index $i \in [k]$ uniformly and independently and let v join the set U_i. For each $v \in V$, define bad event B_v as there exists $i \in [k]$ such that $|\Gamma_v \cap U_i| > \frac{(1+\xi)\Delta}{k}$. Suppose the degree of v is $1 \le d \le \Delta$. Fix $i \in [k]$. For each $j \in [d]$, we use $X_j \in \{0,1\}$ to indicate whether the j-th neighbor of v belongs to U_i. Then

$$\mathbf{Pr}\left[|\Gamma_v\cap U_i| > \frac{(1+\xi)\Delta}{k}\right] \le \mathbf{Pr}\left[\sum_{i=1}^{d} X_j \ge \frac{d}{k}+\frac{\xi\Delta}{k}\right] \le \exp\left(-\frac{2\xi^2\Delta^2}{dk^2}\right) \le \exp\left(-\frac{2\xi^2\Delta}{k^2}\right),$$

where the tail bound follows from Hoeffding's inequality. By a union bound,

$$\mathbf{Pr}[B_v] \le k\exp\left(-\frac{2\xi^2\Delta}{k^2}\right).$$

Finally, B_u and B_v are dependent with each other only if $\mathrm{dist}_G(u,v) \le 2$. The maximum degree of dependency graph is at most $\Delta^2 - 1$. The Lovász local lemma says the partition in the lemma exists if $e\cdot\Delta^2\cdot k e^{-2\xi^2\Delta/k^2} < 1$, which is true if $\Delta = \Omega(\frac{k^2}{\xi^2}\log k)$. This finishes the proof of Proposition 29.

We remark that to prove the results about Glauber dynamics, we only use this partition in the analysis and we do not need to explicitly construct the partition.

5 A Fast Sampling Algorithm

In this section, we prove Theorem 13. We will use the same setting as Section 4. Assume G has a $(1,k)$-degree partition $V = U_1 \uplus U_2 \uplus \cdots \uplus U_k$. By Lemma 26 and Section 4.3, not only such partition exists, but we can construct it explicitly in linear time in expectation. We can further modify it to obtain a construction algorithm that runs in worst case linear time. Let T be the running time of the Las Vegas algorithm. By definition, T is a random variable with $\mathbb{E}[T] = O(n)$. By the Markov inequality, if we run the Las Vegas algorithm until time $2\mathbb{E}[T]$, then it halts with probability $\mathbf{Pr}[T \leq 2\mathbb{E}[T]] = 1 - \mathbf{Pr}[T \geq 2\mathbb{E}[T]] \geq 1/2$. Hence, we can run $\log_2 \frac{2}{\varepsilon}$ copies of Las Vegas algorithm until time $2\mathbb{E}[T]$. If there is at least one copy halts, then we use that partition to continue. If none of them halts, then we output an arbitrary sample. The latter case will contribute to the TV-distance but it only happens with probability $(1/2)^{\log_2 \frac{2}{\varepsilon}} = \varepsilon/2$. Hence, in the rest part of this section, we can assume the partition $U_1, \cdots, U_k$ is already provided for simplicity.

We choose $k = \left\lceil \frac{4\lceil M \rceil}{\eta} \right\rceil$ as we did in the proof of Theorem 9 in Section 4, where M, η are parameters used in Theorem 13. For convenience, for $R \subseteq [k]$, we will use U_R to denote $\cup_{i \in R} U_i$ and $\Lambda_R = V \setminus U_R$. Then, for any $R \subseteq [k]$ with $X_{U_R} \in \Omega(\mu_{U_R})$, we will consider the $(k - |R|) \leftrightarrow 1$ down-up walk on the conditional distribution $\mu_{\Lambda_R}(\cdot \mid X_{U_R})$. To simplify the notation, we will use the following notation.

Definition 34. We use the notion $\mu[X_{U_R}]$ to denote the distribution $\mu_{\Lambda_R}(\cdot \mid X_{U_R}) = \mu_{V \setminus U_R}(\cdot \mid X_{U_R})$.

Recall that it starts from an arbitrary $Y = Y_{\Lambda_R} \in \Omega(\mu[X_{U_R}])$, in each step,

- pick $i \in [k] \setminus R$ uniformly at random;
- resample $Y_{\Lambda_{R \cup \{i\}}} \sim \mu[X_{U_R} \uplus Y_{U_i}]$.

In the rest of this section, we will denote this down-up walk as $P_{X(U_R)}$.

Now, in order to have a fast sampling algorithm for the distribution μ, we try to implement $P_\varnothing$ on this partition. Note that after picking the index $i \in [k]$, the resample phase of $P_\varnothing$ is non-trivial. Suppose i is fixed, now the problem is reduced to sampling from $\mu[X_{U_i}]$. We could then further decompose this task by running $P_{X(U_i)}$.

We use the above scheme recursively until the problem is reduced to sampling from $\mu[X_{U_R}]$ such that $|R| = k - 2M$. Then, in order to sample from $\mu[X_{U_R}]$, we simply run the Glauber dynamics for plenty of steps.

Formally, we will use the algorithm $\mathsf{SimDownUp}(X, U_R)$ in Algorithm 1 to simulate the down-up walk $P_{X(U_R)}$ or the Glauber dynamics (depends on $|R|$) for plenty of steps. We will determine the parameters T_0 and T_1 in Algorithm 1 later. This should be able to generate a random sample within small TV-distance from $\mu[X_{U_R}]$.

Now, we are ready to prove Theorem 13.

Proof of Theorem 13. In order to prove Theorem 13, Let $X \in \Omega(\mu)$ be an arbitrary state. We only need to verify that $\mathsf{SimDownUp}(X, \varnothing)$ will eventually returns a sample Y such that $\mathcal{D}_{\mathrm{TV}}(Y \parallel \mu) \leq \varepsilon$ after the time claimed in Theorem 13. First, we fix T_0 and T_1 as follow

$$T_0 = \left\lceil T^{\mathrm{GD}}_{\mathrm{mix}}(\mu, \eta) \cdot C \cdot \log\left(\frac{n}{\varepsilon}\right) \right\rceil,$$
$$T_1 = \left\lceil C \cdot \log\left(\frac{n}{\varepsilon}\right) \right\rceil,$$

where $C = O(M/\eta)$. Without loss of generality, we may assume that n is sufficiently large such that $2 \leq 2C \leq \log n$ to simplify the formulations. Since otherwise, n is a constant and we can run

Algorithm 1:

```
SimDownUp(X, U_R) begin
    Let Y = X be the initial state;
    if |M| ≥ k − 2M then
        update Y_{Λ_R} by running Glauber dynamics on μ[X_{U_R}] for T_0 steps;
    else
        for t = 1, ··· , T_1 do
            pick i ∈ [k] \ R uniformly at random;
            update Y_{Λ_{R∪{i}}} = SimDownUp(Y, X_{U_R} ⊎ Y_{U_i});
        end
    end
    return Y_{Λ_R};
end
```

a brute force algorithm to generate samples. Then, $T_1 \le 1 + C \log \frac{n}{\varepsilon} \le \left(\log \frac{n}{\varepsilon}\right)^2$ and the running time claimed in Theorem 13 follows directly from our setting for T_0 and T_1.

Now, we are only left to show that the sample returned by $\mathsf{SimDownUp}(\varnothing)$ has the desired accuracy. To achieve this, we will show that for every $R \subseteq [k]$ such that $0 \le |R| \le k - 2M$ and for all $X \in \Omega(\mu)$, it holds that

$$\mathcal{D}_{\mathrm{TV}}\left(\mathsf{SimDownUp}(X, U_R) \parallel \mu[X_{U_R}]\right) \le \left(\frac{\varepsilon}{n}\right)^{C/4} \frac{\rho_{\max}}{\rho_{\min}} \sum_{i=0}^{k-2M-|R|} T_1^i =: F(|R|). \tag{15}$$

Then, by $T_1 \ge 1$, $k = \lceil 4 \lceil M \rceil / \eta \rceil \le n$, and the fact that $\rho_{\max}/\rho_{\min} = \mathrm{poly}(n)$, we note that

$$F(0) \le \left(\frac{\varepsilon}{n}\right)^{C/4} \mathrm{poly}(n) T_1^k \le \left(\frac{\varepsilon}{n}\right)^{C/4} \mathrm{poly}(n) \left(\log\left(\frac{n}{\varepsilon}\right)\right)^{2k},$$

where in the last inequality, we use the assumption $2 \le 2C \le \log n$ make the formulation simple. Now, it is direct to see that there exists a C of the order $O(M/\eta)$ such that $F(0) \le \varepsilon$. This proves Theorem 13.

Now, we only left to prove (15). We will prove (15) via induction on $|R|$.

Base case $|R| = k - 2M$. In this case, $\mathsf{SimDownUp}(X, U_R)$ degenerates to the Glauber dynamics on $\mu[X_{U_R}]$ that runs T_0 steps. By our assumption on T_0, this gives

$$\mathcal{D}_{\mathrm{TV}}\left(\mathsf{SimDownUp}(X, U_R) \parallel \mu[X_{U_R}]\right) \le \left(\frac{\varepsilon}{n}\right)^{C} \le F(k - 2M).$$

Inductive case $0 \le |R| < k - 2M$. Fix $0 \le r < k - 2M$. Suppose (15) holds when $r < |R| \le k - 2M$, we will show that it also holds for $|R| = r$. Recall that we use Λ_R to denote $V \setminus U_R$. Suppose X is the initial state used in Algorithm 1. We run $P_{X(U_R)}$ from $X(\Lambda_R)$ for T_1 steps and we denote these steps as $X(\Lambda_R) = Z_0, Z_1, \cdots, Z_{T_1}$. Similarly, we denote each steps of $\mathsf{SimDownUp}(X, U_R)$ as $X(\Lambda_R) = Y_0, Y_1, \cdots, Y_{T_1}$. We note that Z_is and Y_is here are partial configurations on $\Omega(\mu_{\Lambda_R})$. According to this definition we know $Z_{T_1} \sim P^{T_1}_{X(U_R)}(X(\Lambda_R), \cdot)$ and Y_{T_1} is returned from $\mathsf{SimDownUp}(X, U_R)$. By the triangle inequality,

$$\mathcal{D}_{\mathrm{TV}}\left(\mathsf{SimDownUp}(X, U_R) \parallel \mu[X_{U_R}]\right) \le \mathcal{D}_{\mathrm{TV}}\left(Z_{T_1} \parallel Y_{T_1}\right) + \mathcal{D}_{\mathrm{TV}}\left(Z_{T_1} \parallel \mu[X_{U_R}]\right). \tag{16}$$

We note that the TV-distance between Z_{T_1} and $\mu[X_{U_R}]$ can be bounded via the decay of the Markovian coupling, which is already done in the proof of Lemma 33. This means

$$\mathcal{D}_{\mathrm{TV}}\left(Z_{T_1} \parallel \mu[X_{U_R}]\right) \leq \left(\frac{1}{2}\right)^{T_1} \cdot \frac{\rho_{\max}}{\rho_{\min}} \cdot n \leq \left(\frac{\varepsilon}{n}\right)^{C/4} \cdot \frac{\rho_{\max}}{\rho_{\min}}, \tag{17}$$

where, in the last equation, we use the fact that $C \geq 4$ to simplify the formula.

In order to bound $\mathcal{D}_{\mathrm{TV}}\left(Z_{T_1} \parallel Y_{T_1}\right)$, we use the standard way to construct a coupling between Z_{T_1} and Y_{T_1}. Note that $\mathsf{SimDownUp}(X, U_R)$ and $P_{X(U_R)}$ only differs at the resample stage. Recall that for $R \subseteq [k]$, we use Λ_R to denote $V \setminus U_R$. For $t = 1, \cdots T_1$, We build the coupling between Z_t and Y_t recursively as follow:

1. sample the same index $i \in [k] \setminus R$ and let $Z_t(U_i) = Z_{t-1}(U_i)$ and $Y_t(U_i) = Y_{t-1}(U_i)$;
2. in the resampling stage:
   - if $Z_{t-1} = Y_{t-1}$, then $Z_t(\Lambda_{R\cup\{i\}})$ and $Y_t(\Lambda_{R\cup\{i\}})$ are sampled from the optimal coupling between $\mu[X(U_R) \uplus Y_{t-1}(U_i)]$ and $\mathsf{SimDownUp}(Y_{t-1} \uplus X(U_R), X(U_R) \uplus Y_{t-1}(U_i))$;
   - otherwise, if $Z_{t-1} \neq Y_{t-1}$, then $Z_t(\Lambda_{R\cup\{i\}})$ and $Y_t(\Lambda_{R\cup\{i\}})$ are sampled from $\mu[X(U_R) \uplus Y_{t-1}(U_i)]$ and $\mathsf{SimDownUp}(Y_{t-1} \uplus X(U_R), X(U_R) \uplus Y_{t-1}(U_i))$ independently.

For simplicity, we use $\nu[X(U_R) \uplus Y_{t-1}(U_i)]$ to denote the distribution generated by $\mathsf{SimDownUp}(Y_{t-1} \uplus X(U_R), X(U_R) \uplus Y_{t-1}(U_i))$. Hence, according to the coupling lemma (Lemma 22), we have

$$\begin{aligned}
\mathcal{D}_{\mathrm{TV}}\left(Z_{T_1} \parallel Y_{T_1}\right) &\leq \mathbf{Pr}\left[Z_{T_1} \neq Y_{T_1}\right] \leq \mathbf{Pr}\left[\exists t, \text{s.t. } t \text{ is the first time that } Z_t \neq Y_t\right] \\
\text{(union bound)} \quad &\leq \sum_{t=1}^{T_1} \mathbf{Pr}\left[Z_t \neq Y_t \text{ and } \forall_{j<t} Z_j = Y_j\right] \leq \sum_{t=1}^{T_1} \mathbf{Pr}\left[Z_t \neq Y_t \mid Z_{t-1} = Y_{t-1}\right] \\
&= \sum_{t=1}^{T_1} \mathbb{E}_{i, Y_{t-1}}\left[\mathcal{D}_{\mathrm{TV}}\left(\mu[X(U_R) \uplus Y_{t-1}(U_i)] \parallel \nu[X(U_R) \uplus Y_{t-1}(U_i)]\right)\right] \\
\text{(I.H.)} \quad &\leq T_1 \cdot F(|R|+1).
\end{aligned} \tag{18}$$

Combining (16), (17), and (18), we have

$$\mathcal{D}_{\mathrm{TV}}\left(\mathsf{SimDownUp}(X, U_R) \parallel \mu[X(U_R)]\right) \leq F(|R|+1) \cdot T_1 + \left(\frac{\varepsilon}{n}\right)^{C/4} \cdot \frac{\rho_{\max}}{\rho_{\min}}.$$

Plugging in the definition of F in (15), we finish the proof of (15) and the proof of Theorem 13. □

6 Mixing Time Comparison for Monotone Systems

In this section, we prove Theorem 15. For monotone spin systems, the famous censoring inequality states that censoring updates never decreases the distance to stationary. We refer the readers to [PW13] and the textbook [LPW17] for details. We refer Definition 14 for the definition of monotone spin systems. Then the censoring inequality on monotone spin systems is stated as follow.

Lemma 35 ([PW13])**.** *Let μ be a monotone spin system. Let $v_1, v_2, \cdots, v_m$ be a random sequence of vertices. Let π be the distribution resulting from updates at $v_1, \cdots, v_m$, starting from the maximum state. Let ν be the distribution resulting from updates at a random subsequence $v_{i_1}, \cdots, v_{i_k}$, also started from the maximum state. The randomness for the updates is independent from the randomness of the sequence $v_1, v_2, \cdots, v_m$ and the subsequence $v_{i_1}, \cdots, v_{i_k}$. Then $\pi \preceq \nu$, and*

$$\mathcal{D}_{\mathrm{TV}}\left(\pi \parallel \mu\right) \leq \mathcal{D}_{\mathrm{TV}}\left(\nu \parallel \mu\right).$$

For monotone spin systems, the censoring inequality allows us to compare the Glauber dynamics with the simulate algorithm SimDownUp in Algorithm 1. However, to finish the comparison, we need a slightly stronger partition. We define it formally as follow.

Definition 36. Let $U_1, \cdots, U_k$ be a (ζ, k)-partition of V as defined in Definition 27. If it further holds that $|U_i| \geq \frac{n}{2k}$, then we say $U_1 \cdots, U_k$ is a *balanced* (ζ, k)-partition.

The following result shows that such partition exists.

Proposition 37. *Let $k \geq 1$ be an integer and $\xi \in (0,1)$. Then, any n-vertices graph $G = (V, E)$ such that $n = \Omega(k \log k)$ and the maximum degree $\Delta = \Omega(\frac{k^2}{\xi^2} \log k)$ has a balanced (ξ, k)-degree partition.*

Proof. For each vertex v, sample an index $i_v \in [k]$ uniformly at random and let v join the set U_{i_v}. According to the proof of Proposition 29 in Section 4.3 and Lemma 26, when $\Delta = \Omega(\frac{k^2}{\zeta^2} \log k)$,

$$\mathbf{Pr}\left[(U_i)_{i\in[k]} \text{ is a } (\zeta, k)\text{-partition}\right] \geq \left(1 - \frac{1}{\Delta + 1}\right)^n.$$

Then for any $t \in [k]$, let $Y_v = \mathbf{1}[i_v = t]$ to indicate that if vertex v is added to U_t. Let $Y = \sum_v Y_v$ and we have $\mathbb{E}[Y] = n/k$ directly. Then by the Chernoff bound, it holds that

$$\mathbf{Pr}\left[|U_t| < \frac{n}{2k}\right] \leq \mathbf{Pr}\left[Y \leq \mathbb{E}[Y]/2\right] \leq \exp\left(-\mathbb{E}[Y]/8\right) = \exp\left(-\frac{n}{8k}\right).$$

Then according to a union bound

$$\begin{aligned}
&\mathbf{Pr}\left[(U_i)_{i\in[k]} \text{ is a balanced } (\zeta, k)\text{-partition}\right] \\
&\qquad \geq 1 - \mathbf{Pr}\left[(U_i)_{i\in[k]} \text{ is not a } (\zeta, k)\text{-partition}\right] - \sum_{t\in[k]} \mathbf{Pr}\left[|U_t| < \frac{n}{2k}\right] \\
&\qquad \geq \left(1 - \frac{1}{\Delta + 1}\right)^n - k \exp\left(-\frac{n}{8k}\right),
\end{aligned}$$

which is positive when $n = \Omega(k \log k)$ and $\Delta = \Omega(k)$. In summary, when $\Delta = \Omega(\frac{k^2}{\zeta^2} \log k)$ and $n = \Omega(k \log k)$, then such partition exists. □

Proof of Theorem 15. In this prove, we will basicly use the same setting as the proof of Theorem 13 in Section 5. However, to apply the censoring inequality, we need to do the following changes:

- In Section 5, we are using the $(1, k)$-partition for the proof. Here, we use the stronger balanced $(1, k)$-partition. The existence of such partition is guaranteed by Proposition 37. Since $k = O(M/\eta)$ is only a constant depending on M and η, we can assume $n = \Omega(k \log k)$ without loss of generality to meet the requirement in Proposition 37.
- Also note that in the proof of Theorem 13 in Section 5, the initial state $X \in \Omega(\mu)$ for the algorithm $\mathsf{SimDownUp}(X, \varnothing)$ is picked arbitrarily from $\Omega(\mu)$. Here, in order to use censoring inequality, we have to use the maximum state $X_0 \in \Omega(\mu)$ as initial state.

Let $T_{\mathsf{sim}} = T^{\mathrm{GD}}_{\mathrm{mix}}(\mu, \eta) \left(\log \frac{2n}{\varepsilon}\right)^{O(M/\eta)}$ be the number of updates performed by Glauber dynamics required by Theorem 13. We note that the extra Δ in Theorem 13 is used to implement one step of the Glauber dynamics.

In this manner, the algorithm $\mathsf{SimDownUp}(X_0, \varnothing)$ can be seen as repeatedly applying Glauber dynamics updates P_v (see (7)) at some randomly chosen vertex v to the current states initiating from the maximum over the whole state space. Suppose the update sequence of this process is

$$\begin{array}{lll} v_{1,1}, & \cdots, & v_{1,T_0}, \\ v_{2,1}, & \cdots, & v_{2,T_0}, \\ \cdots, & \cdots, & \cdots, \\ v_{N,1}, & \cdots, & v_{N,T_0}, \end{array}$$

where T_0 is the parameter defined in Algorithm 1 and we use N to denote T_{sim}/T_0. We note that for each $1 \le i \le N$, the vertices $\{v_{i,j}\}_{1 \le j \le T_0}$ are picked independently from a subset $\Lambda_{R_i} = V \setminus U_{R_i}$ ($R_i \subseteq [k]$) uniformly at random. Note that R_i is also picked from $[k]$ according to some distribution.

We run the Glauber dynamics P from the maximal state for CT_{sim} steps where C is some integer parameter that we will determine later. The update sequence of the Glauber dynamics is

$$\begin{array}{lll} u_{1,1}, & \cdots, & u_{1,CT_0}, \\ u_{2,1}, & \cdots, & u_{2,CT_0}, \\ \cdots, & \cdots, & \cdots, \\ u_{N,T_0}, & \cdots, & u_{N,CT_0}, \end{array}$$

where each $u_{i,j}$ is picked from V uniformly at random. We consider an indicator $\mathbb{I}_{i,j} = \mathbf{1}[u_{i,j} \in \Lambda_{R_i}]$. Note that conditioning on $\mathbb{I}_{i,j} = 1$, it holds that $u_{i,j}$ is distributed uniformly over Λ_{R_i}.

Now, we consider a censored Glauber dynamics P_1 starting from the maximum state. An update $u_{i,j}$ in P is censored P_1 if $\mathbb{I}_{i,j} = 0$. After that, we further censor P_1 to P_2 to make sure that there are at most T_0 updates for each row i, $1 \le i \le N$.

For fixed $\mathbb{I}_{\cdot,\cdot}$, let the resulting distribution of P be $\pi_{\mathbb{I}}$ and the resulting distribution of P_2 be $\nu_{\mathbb{I}}$. By the censoring inequality in Lemma 35

$$\mathcal{D}_{\mathrm{TV}}\left(\pi_{\mathbb{I}} \parallel \mu\right) \le \mathcal{D}_{\mathrm{TV}}\left(\nu_{\mathbb{I}} \parallel \mu\right).$$

Let ρ be the resulting distribution of the algorithm $\mathsf{SimDownUp}(X_0, \varnothing)$. By the triangle inequality:

$$\mathcal{D}_{\mathrm{TV}}\left(\pi_{\mathbb{I}} \parallel \mu\right) \le \mathcal{D}_{\mathrm{TV}}\left(\nu_{\mathbb{I}} \parallel \rho\right) + \mathcal{D}_{\mathrm{TV}}\left(\rho \parallel \mu\right).$$

Taking expectation w.r.t. $\mathbb{I}_{\cdot,\cdot}$ at both side, we arrive at

$$\mathbb{E}_{\mathbb{I}}\left[\mathcal{D}_{\mathrm{TV}}\left(\pi_{\mathbb{I}} \parallel \mu\right)\right] \le \mathbb{E}_{\mathbb{I}}\left[\mathcal{D}_{\mathrm{TV}}\left(\nu_{\mathbb{I}} \parallel \rho\right)\right] + \mathcal{D}_{\mathrm{TV}}\left(\rho \parallel \mu\right). \tag{19}$$

By a standard coupling argument, it holds that $\mathcal{D}_{\mathrm{TV}}\left(\pi \parallel \mu\right) \le \mathbb{E}_{\mathbb{I}}\left[\mathcal{D}_{\mathrm{TV}}\left(\pi_{\mathbb{I}} \parallel \mu\right)\right]$. It sufficient for us to bound the RHS of (19). According to our choice of T_{sim} and Theorem 13, we know that $\mathcal{D}_{\mathrm{TV}}\left(\rho \parallel \mu\right) \le \varepsilon/2$. Now, it is sufficient for us to show that $\mathbb{E}_{\mathbb{I}}\left[\mathcal{D}_{\mathrm{TV}}\left(\nu_{\mathbb{I}} \parallel \rho\right)\right] \le \varepsilon/2$. Note that when $\mathbb{I}_{\cdot,\cdot}$ is picked that P_2 has exactly T_0 updates in each row, then P_2 and $\mathsf{SimDownUp}$ becomes the same random process, and hence $\mathcal{D}_{\mathrm{TV}}\left(\nu_{\mathbb{I}} \parallel \rho\right) = 0$. Hence we have

$$\begin{aligned} \mathbb{E}_{\mathbb{I}}\left[\mathcal{D}_{\mathrm{TV}}\left(\nu_{\mathbb{I}} \parallel \rho\right)\right] &\le 1 - \mathbf{Pr}_{\mathbb{I}}\left[P_2 \text{ has exactly } T_0 \text{ updates in each row}\right] \\ &= \mathbf{Pr}_{\mathbb{I}}\left[\exists i, P_1 \text{ has} < T_0 \text{ updates in the } i\text{-th row}\right] \\ \text{(union bound)} \quad &\le \sum_i \mathbf{Pr}_{\mathbb{I}}\left[P_1 \text{ has} < T_0 \text{ updates in the } i\text{-th row}\right]. \end{aligned} \tag{20}$$

For the i-th row, according to the definition of the balanced $(1,k)$-partition $(U_i)_{i \in [k]}$ in Definition 36, we note that each $\mathbb{I}_{i,j} = 1$ with probability $|\Lambda_{R_i}| / |V| \ge \frac{1}{2k}$. This means $(\mathbb{I}_{i,j})_{1 \le j \le CT_0}$ are i.i.d

random 0/1 random variables with mean $\mathbb{E}\left[\mathbb{I}_{i,j}\right] \geq \frac{1}{2k}$. Let $X = \sum_{1\leq j\leq CT_0} \mathbb{I}_{i,j}$, we know $\mathbb{E}[X] \geq \frac{CT_0}{2k}$. According to the Chernoff bound

$$\begin{aligned}\mathbf{Pr}[X < T_0] = \mathbf{Pr}\left[X < \frac{CT_0}{2k}\cdot\frac{2k}{C}\right] &\leq \mathbf{Pr}\left[X \leq \mathbb{E}[X]\cdot\frac{2k}{C}\right] \\ &\leq \exp\left(-\frac{\mathbb{E}[X](1-2k/C)^2}{2}\right) \leq \exp\left(-\frac{CT_0}{10k}\right) \end{aligned} \tag{21}$$

where we use the assumption $C \geq 10k$ to simplify the formulation.

Combining (20) and (21), it holds that

$$\mathbb{E}_{\mathbb{I}}\left[\mathcal{D}_{\mathrm{TV}}\left(\nu_{\mathbb{I}} \parallel \rho\right)\right] \leq N\cdot\exp\left(-\frac{CT_0}{10k}\right).$$

It is then directly to notice that in order to make $\mathbb{E}_{\mathbb{I}}\left[\mathcal{D}_{\mathrm{TV}}\left(\nu_{\mathbb{I}} \parallel \rho\right)\right] \leq \varepsilon/2$, we need

$$C \geq \frac{10k}{T_0}\log\frac{2N}{\varepsilon}.$$

Recall the parameters related to SimDownUp that we have $k = O(M/\eta)$ and

$$\begin{aligned} T_0 &= T_{\mathrm{mix}}^{\mathrm{GD}}\cdot O(M/\eta)\cdot\log\left(\frac{2n}{\varepsilon}\right), \\ N \leq T_{\mathsf{sim}} &= T_{\mathrm{mix}}^{\mathrm{GD}}(\mu,\eta)\left(\log\frac{2n}{\varepsilon}\right)^{O(M/\eta)}. \end{aligned}$$

This means we can assume n to be sufficiently large and pick $C = O(M/\eta)$. We finish the proof by noticing that the number of updates performed by the Glauber dynamics is bounded by $CT_{\mathsf{sim}} = T_{\mathrm{mix}}^{\mathrm{GD}}(\mu,\eta)\left(\log\frac{2n}{\varepsilon}\right)^{O(M/\eta)}$. □

7 Establish Coupling Independence via Self-Avoiding-Walk Tree

Let μ be the Gibbs distribution of a 2-spin system with parameters β,γ,λ on a connected graph $G = (V,E)$. Assume there is an arbitrary ordering of vertices in G. We fix an arbitrary total order $<$ on V. For any vertex $r \in V$, the self-avoiding-walk (SAW) tree [Wei06, LLY13] $T_{\mathrm{SAW}}(G,r)$ is a tree with pinnings on some leaves. The tree enumerates all self-avoiding-walks $v_0, v_1, \ldots, v_\ell$ in G starting from $r = v_0$ such that

- all vertices $v_0, v_1, \ldots, v_{\ell-1}$ are distinct;
- either the degree of v_ℓ is 1 in G or v_ℓ is a cycle-closing vertex (i.e. $v_i = v_\ell$ for some $i < \ell$);
- for every cycle-closing vertex v_ℓ in a SAW $v_0, v_1, \ldots, v_\ell$ with $v_\ell = v_i$ for some $i < \ell$, the value of v_ℓ is fixed as $-$ if $v_{i+1} > v_{\ell-1}$ and the value of v_ℓ is fixed as $+$ if $v_{i+1} < v_{\ell-1}$.

In the definition of the SAW tree, each vertex $v \in V$ in graph G may have multiple copies in $T_{\mathrm{SAW}}(G,r)$. We say a copy is free if its value is not fixed. One can define a two-spin system with pinning on T_{SAW} using the same parameters β,γ,λ. We use π to denote the Gibbs distribution on $T_{\mathrm{SAW}}(G,r)$.

We can extend a pinning in G to a pinning in $T_{\mathrm{SAW}}(G,r)$. Let $\Lambda \subseteq V$ and $r \notin \Lambda$. Given any pinning $\tau \in \{-,+\}^\Lambda$, for any vertex $v \in \Lambda$, we find all copies $\hat{v}$ of v in $T_{\mathrm{SAW}}(G,r)$ such that $\hat{v}$ is free,

fix the value of $\hat{v}$ as τ_v, and remove all descendants of $\hat{v}$. Again, the pinning only appears in leaves of this new tree. We slightly abuse the notation to denote the new SAW tree by $T^{\tau}_{\text{SAW}}(G,r)$ and the Gibbs distribution in this SAW tree by π^{τ}. We also call the SAW tree $T^{\tau}_{\text{SAW}}(G,r)$ as $T_{\text{SAW}}(G,r)$ with pinning τ.

The SAW tree also admits an inductive definition (e.g. see [CLV20]), which will be used in our analysis. Suppose r has d neighbors $u_1 < u_2 < \ldots < u_d$. We split r into $r_1, r_2, \ldots, r_d$ and connect r_i to u_i to obtain a graph $\overline{G}$. Let U_i denote the pinning that fixes the value of r_j with $j < i$ to be $-$ and the value for r_j with $j > i$ to be $+$. To construct the SAW tree $T^{\tau}_{\text{SAW}}(G,r)$, one can first construct all $T_i = T^{\tau \wedge U_i}_{\text{SAW}}(\overline{G} - r_i, u_i)$ and then connect r to the root of each T_i.

Recall the definition of influence matrix Ψ^{τ}_{π} in (4). We list some standard properties of SAW tree, which are widely-used for establishing the spectral independence for 2-spin systems [ALO20, CLV20, CFYZ21].

Proposition 38 ([Wei06, ALO20])**.** *For any SAW tree $T = T^{\tau}_{SAW}(G,r)$ with Gibbs distribution π,*

- *$\pi_r = \mu^{\tau}_r$, where μ is the Gibbs distribution on graph G;*
- *for any free copy $\hat{u}$ of u, the degree of $\hat{u}$ in T is the same as the degree of u in G;*
- *for any u, v in T, any w in the path between u and v, $\Psi^{\tau}_{\pi}(u,v) = \Psi^{\tau}_{\pi}(u,w)\Psi^{\tau}_{\pi}(w,v)$.*

We give the following result that relates the coupling independence to the influences in SAW tree. Many previous works proved the ℓ_∞ spectral independence for the SAW trees, which is a sufficient condition for the spectral independence of the original Gibbs distribution on G. With this lemma, we can transform them into coupling independence results in a black-box manner. The proof of the lemma is similar to the recursive coupling introduced in [GMP05].

Lemma 39. *For any $\Lambda \subseteq V$, any $\tau \in \{-,+\}^{V\setminus\Lambda}$, any $v \in \Lambda$, there exists a coupling (X,Y) of $\mu^{\tau\wedge v^-}_{\Lambda}$ and $\mu^{\tau\wedge v^+}_{\Lambda}$, where $\tau \wedge v^c$ is the condition τ together with v taking $c \in \{-,+\}$, such that*

$$\forall u \in \Lambda \setminus \{v\}, \quad \mathbf{Pr}\left[X_u \neq Y_u\right] \leq \sum_{\hat{u}:\hat{u} \text{ is a copy of } u \text{ in } T^{\tau}_{\text{SAW}}(G,v)} |\Psi^{\tau}_{\pi}(v,\hat{u})|, \tag{22}$$

where Ψ^{τ}_{π} is the influence matrix for Gibbs distribution π^{τ} in $T_{\text{SAW}}(G,v)$.

Proof. Fix parameters β, γ, λ of the 2-spin system. For any Gibbs distribution μ on a graph $G = (V,E)$, any pinning $\tau \in [q]^{V\setminus\Lambda}$, where $\Lambda \subseteq V$, let $|\Lambda|$ denote the number of free variables for μ^{τ}. For any $k \geq 1$, we use a induction proof to show that any conditional Gibbs distribution μ^{τ} with k free variables satisfies (22) for any free variable v, where μ can be defined on an arbitrary finite graph as long as μ^{τ} has k free variables. The base case $k = 1$ is trivial.

Suppose the induction hypothesis holds for all $k' < k$. We prove it for conditional Gibbs distribution μ^{τ} with $k = |\Lambda|$ free variables, where μ is defined on $G = (V,E)$ and $\Lambda \subseteq V$. Let $u_1 < u_2 < \cdots < u_d$ be the neighbors of v in G. Then, we construct a graph $\overline{G}$ from G by splitting v into $\{v_1, v_2, \cdots, v_d\}$. Using the same parameters β, γ, λ, we can define a Gibbs distribution $\overline{\mu}$ on $\overline{G}$. Define the pinning σ_i such that all v_j for $j \leq i$ take value - and all v_j for $j > i$ take the value $+$. To simplify the notation, we denote $\nu = \mu^{\tau}_{\Lambda}$ and $\nu_i = \overline{\mu}^{\tau\wedge\sigma_i}_{\Lambda\setminus v}$. It is easy to see $\nu^{v^+}_{\Lambda\setminus v} = \nu_0$ and $\nu^{v^-}_{\Lambda\setminus v} = \nu_d$. To couple ν^{v^+} and ν^{v^-}, we only need to couple ν_0 and ν_d.

The construction of the coupling (X_0, X_d) between ν_0 and ν_d is achieved by path coupling method [BD97]. Specifically, we construct couplings (X_{i-1}, X_i) between ν_{i-1} and ν_i for all $1 \leq i \leq d$. We first sample (X_0, X_1) from the first coupling, then conditional on the value X_{i-1}, we sample X_i from the i-th coupling. Finally, we get a coupling (X_0, X_d) between ν_0 and ν_d.

Consider the i-th coupling (X_{i-1}, X_i). The only difference between ν_{i-1} and ν_i is the pinning at vertex v_i. The only neighbor of v_i is u_i. Each distribution ν_i has $|\Lambda \setminus v| = k - 1$ free variables, so we can use the I.H. on distribution ν_i. The coupling does as follows.

- Couple $X_{i-1}(u_i)$ and $X_i(u_i)$ via the optimal coupling between marginal distributions.
- If $X_{i-1}(u_i) = X_i(u_i)$, then $X_{i-1} = X_i$ can be coupled perfectly. Because, by conditional independence, given the same value c on u_i, the conditional distribution on $\Lambda \setminus v$ induced from ν_i and ν_{i-1} are the same, i.e. $\nu_{i-1,\Lambda\setminus v}^{u_i^c} = \nu_{i,\Lambda\setminus v}^{u_i^c}$ for any $c \in \{-,+\}$; If $X_{i-1}(u_i) \neq X_i(u_i)$, by I.H., there is a coupling between $\nu_{i,\Lambda\setminus v}^{u_i^-}$ and $\nu_{i,\Lambda\setminus v}^{u_i^+}$ satisfying (22), and we use this coupling to couple all variables $\Lambda \setminus v$[3] in X_{i-1} and X_i.

By the construction of coupling, we have for any vertex $u \in \Lambda$ and $u \neq v$,

$$\mathbf{Pr}\left[X_{i-1}(u) \neq X_i(u)\right] = \mathbf{Pr}\left[X_{i-1}(u_i) \neq X_i(u_i)\right] \cdot \mathbf{Pr}\left[X_{i-1}(u) \neq X_i(u) \mid X_{i-1}(u_i) \neq X_i(u_i)\right].$$

The first probability on the RHS is the total variation distance between two marginals. Consider SAW tree $T_{\mathrm{SAW}}^{\tau}(G, v)$. The root v has d child $u_1, u_2, \ldots, u_d$. We use T_i to denote the subtree rooted at u_i. By the inductive definition of SAW tree, the tree $T_i + v$ with the value of v being $-$ (resp. $+$) is exactly $T_{\mathrm{SAW}}(\overline{G}, u_i)$ with pinning $\sigma_i \wedge \tau$ (resp. $\sigma_{i-1} \wedge \tau$). Since SAW tree preserves the marginal distribution at the root u_i (the first property in Proposition 38),

$$\mathbf{Pr}\left[X_{i-1}(u) \neq X_i(u)\right] = \mathcal{D}_{\mathrm{TV}}\left(\nu_{i-1,u_i} \parallel \nu_{i,u_i}\right) = |\Psi_{\pi}^{\tau}(v, u_i)|.$$

Consider the second step of coupling. Recall that $\nu_i = \overline{\mu}_{\Lambda\setminus v}^{\tau\wedge\sigma_i}$. Consider the SAW tree $T_{\mathrm{SAW}}(\overline{G}, u_i)$ with pinning $\tau \wedge \sigma_i$. Let $\pi_i^{\tau\wedge\sigma_i}$ denote its Gibbs distribution. Let $\Psi_{\pi_i}^{\tau\wedge\sigma_i}$ denote the influence matrix for $\pi_i^{\tau\wedge\sigma_i}$. By I.H, for any $u \in \Lambda \setminus v$,

$$\mathbf{Pr}\left[X_{i-1}(u) \neq X_i(u) \mid X_{i-1}(u_i) \neq X_i(u_i)\right] \leq \sum_{\hat{u}\text{: copies of } u \text{ in } T_{\mathrm{SAW}}^{\tau\wedge\sigma_i}(\overline{G}, u_i)} |\Psi_{\pi_i}^{\tau\wedge\sigma_i}(u_i, \hat{u})|.$$

Finally, we need to relate $T_{\mathrm{SAW}}^{\tau\wedge\sigma_i}(\overline{G}, u_i)$ to $T_{\mathrm{SAW}}^{\tau}(G, v)$. Recall T_i is the i-th subtree of the root v in $T_{\mathrm{SAW}}^{\tau}(G, v)$. Again, by the induction definition of the SAW tree, $T_i + v$ with pinning $+$ on v is exactly $T_{\mathrm{SAW}}^{\tau\wedge\sigma_i}(\overline{G}, u_i)$. By the conditional independence property, given the value of u_i, every variable in subtree T_i is independent from all variables outside T_i. Hence, for any $\hat{u} \in T_i$, we have

$$\Psi_{\pi_i}^{\tau\wedge\sigma_i}(u_i, \hat{u}) = \Psi_{\pi}^{\tau}(u_i, \hat{u}).$$

Combining all of above analysis together and using a union bound for path coupling, we have

$$\begin{aligned}
\mathbf{Pr}\left[X_0(u) \neq X_d(u)\right] &\leq \sum_{i=1}^{d} \mathbf{Pr}\left[X_{i-1}(u) \neq X_i(u)\right] \leq \sum_{i=1}^{d} |\Psi_{\pi}^{\tau}(v, u_i)| \sum_{\hat{u}\text{: copies of } u \text{ in } T_i} |\Psi_{\pi}^{\tau}(u_i, \hat{u})| \\
\text{(by last property in Proposition 38)} \quad &= \sum_{\hat{u}\text{: copies of } u \text{ in } T_{\mathrm{SAW}}^{\tau}(G, v)} |\Psi_{\pi}^{\tau}(v, \hat{u})|.
\end{aligned}$$

This finishes the induction step of the proof. □

[3] We remark that $u_i \in \Lambda \setminus v$ and u_i must take $+$ (and $-$) in $\nu_i^{u_i^+}$ (and $\nu_i^{u_i^-}$).

8 Hardcore Model in Bipartite Graphs

We prove Theorem 5 and Theorem 6 in this section. First, let us recall the parameters. Let $\delta > 0, \theta > 1$ be two constants. Let $G = (V_L, V_R, E)$ be a bipartite graph. Let $\Delta_L \geq 3$ and Δ_R denote the degree in the left and right parts respectively. Assume $\lambda \leq (1-\delta)\Delta_L$ and $\Delta_R \leq \theta\Delta_L$. Let $V = V_L \uplus V_R$. Let μ over $\{-,+\}^V$ denote the hardcore distribution in graph G with fugacity λ, where for any $X \sim \mu$ corresponds the the independent set $\{v \in V \mid X_v = +\}$ in graph G.

The proof can be outlined as follows. We use μ_L to denote the marginal distribution on L projected from μ. Note that μ_L is not a Gibbs distribution on graph G.

8.1 Coupling Independence for Marginal Distributions

Lemma 40. *The marginal μ_L satisfies $O(1/\delta)$-coupling-independence with $\rho(v) = 1$ for all $v \in V$.*

The coupling independence of μ_L can be obtained from Lemma 39 and the SAW-tree analysis in [CLY23]. Fix a pinning $\Lambda \subseteq V_L$ and $\tau \in \{0,1\}^{V_L \setminus \Lambda}$. For any vertex $v \in \Lambda$, we need to show a good coupling exists for $\mu_L^{\tau \wedge v^+}$ and $\mu_L^{\tau \wedge v^-}$. Similar to Section 7, starting from a vertex v, we can define the SAW tree $T_{\text{SAW}}(G, v)$ and extend the pinning τ to the SAW tree to obtain a hardcore distribution π^τ in SAW tree. We use L_k to denote the set of vertices in level k, where the root v is in the level $k = 0$. Hence, all copies of vertices in V_L are in level L_k for even k. When $\lambda \leq (1-\delta)\lambda_c(\Delta_L)$, the following result is proved in [CLY23, see Theorem 2, Lemma 63, Lemma 64 and inequality (84)].

Lemma 41 ([CLY23])**.** *For any $k \geq 0$, $\sum_{w \in L_{2k}} |\Psi_\pi^\tau(v,w)| \leq \frac{\Delta_L}{\Delta_L - 1}(1+\lambda)^{\Delta_L}(1-\frac{\delta}{10})^k$.*

By Lemma 39, we can easily transform the influence bound in [CLY23] into a coupling independence result. We now prove Lemma 40.

Proof of Lemma 40. Note that τ in a pinning on $V_L \setminus \Lambda$. Vertices in $V_R \uplus \Lambda$ is free given τ. There exists a coupling (X, Y) between $\mu_{\Lambda \uplus V_R}^{\tau \wedge v^+}$ and $\mu_{\Lambda \uplus V_R}^{\tau \wedge v^-}$ satisfying the condition in Lemma 39. We can project both X into $\Lambda \subseteq V_L$ to obtain a coupling (X_Λ, Y_Λ) between $\mu_\Lambda^{\tau \wedge v^+}$ and $\mu_\Lambda^{\tau \wedge v^-}$. The expected Hamming distance $|X_\Lambda \oplus Y_\Lambda| = |\{u \in \Lambda \mid X_u \neq Y_u\}|$ can be bounded by

$$\mathbb{E}\left[|X_\Lambda \oplus Y_\Lambda|\right] \leq 1 + \sum_{k \geq 1} \sum_{w \in L_{2k}} |\Psi_\pi^\tau(v,w)| \leq \frac{\Delta_L}{\Delta_L - 1}(1+\lambda)^{\Delta_L} \sum_{k \geq 0} (1 - \frac{\delta}{10})^k,$$

where the last inequality is from Lemma 41. Finally, note that $\frac{\Delta_L}{\Delta_L - 1} \leq 1.5$ and $(1+\lambda)^{\Delta_L} = (1 + O(\frac{1}{\Delta_L}))^{\Delta_L} = O(1)$. The expected Hamming distance is at most $O(1/\delta)$. Since this bound holds for any pinning τ, we proved the $O(1/\delta)$-coupling-independence for the marginal distribution μ_L. □

8.2 Graph Partition for Bipartite Graphs

We partition all the vertices in the left part V_L into k disjoint parts $U_1, U_2, \ldots, U_k$ such that for

$$\forall v \in V_R, \forall i \in [k], \quad |\Gamma_v \cap U_i| \leq \Delta_L, \tag{23}$$

where $\Gamma_v \subseteq V_L$ denote neighbors of $v \in V_R$ in graph G. Since the degree of $v \in V_R$ is $\theta\Delta_L$, we roughly need to partition V_L into $k = \Omega(\theta)$ parts, which can be achieved via local lemma.

Proposition 42. *For any $k \geq \lceil 2\theta \rceil$, if $\Delta_L = \Omega(\theta \log(k\theta))$, then there exists a partition $V_L = U_1 \uplus U_2 \uplus \ldots \uplus U_k$ satisfying (23).*

Proof. We use the Lovász local lemma (Lemma 26) to prove the existence. We first set the parameter $k \geq \lceil 2\theta \rceil$. For each vertex $v \in V_L$, sample an index $i \in [k]$ uniformly and independently and let v join the set U_i. For each $u \in V_R$, define bad event B_u as there exists $i \in [k]$ such that $|\Gamma_u \cap U_i| > \Delta_L$. Suppose the degree of v is $1 \leq d \leq \Delta_R \leq \theta\Delta_L$. We have $\frac{d}{k} + \frac{\Delta_L}{2} \leq \Delta_L$. Fix $i \in [k]$. For each $j \in [d]$, we use $X_j \in \{0,1\}$ to indicate whether the j-th neighbor of v belongs to U_i. Then

$$\mathbf{Pr}\left[|\Gamma_u \cap U_i| > \Delta_L\right] \leq \mathbf{Pr}\left[\sum_{i=1}^{d} X_j \geq \frac{d}{k} + \frac{\Delta_L}{2}\right] \leq \exp\left(-\frac{\Delta_L^2}{2d}\right) \leq \exp\left(-\frac{\Delta_L}{2\theta}\right),$$

where the tail bound follows from Hoeffding's inequality. By a union bound,

$$\mathbf{Pr}\left[B_u\right] \leq k \exp\left(-\frac{\Delta_L}{2\theta}\right).$$

Finally, B_u and B_v are dependent with each other only if $\mathrm{dist}_G(u,v) \leq 2$. The maximum degree of dependency graph is at most $\theta^2\Delta_L^2$. The Lovász local lemma says the partition in the lemma exists if $e \cdot \theta^2\Delta_L^2 \cdot k \exp\left(-\frac{\Delta_L}{2\theta}\right) < 1$, which is true if $\Delta_L = \Omega(\theta \log \theta k)$. □

8.3 Relaxation Time and Mixing Time Bounds

Let $M = O(1/\delta)$ be the coupling-independence parameter. Without loss of generality, we assume M is an integer at least 1. If not, we can around M up to an integer. We set parameter $k = \max\{\lceil 2\theta \rceil, 10M\}$. Let $\Delta_0 = \Delta_0(\theta, \delta) = \Theta(\theta \log(k\theta)) = O(\theta \log \frac{\theta}{\delta}) = O_{\delta,\theta}(1)$ be the threshold for Δ_L in Proposition 42 such that the good k-partition exists if $\Delta_L \geq \Delta_0$.

Now, we consider two cases: $\Delta_L < \Delta_0 = O(1)$ and $\Delta_L \geq \Delta_0$. The first one is the easier case. We claim the following result.

Proposition 43. *Let $\delta \in (0,1)$ and $\theta > 1$ be two constants. For any hardcore model on a θ-balanced bipartite graph G with fugacity λ, if $\Delta_L < \Delta_0(\theta, \delta)$ and $\lambda \leq (1-\delta)\lambda_c(\Delta_L)$, then the relaxation time and mixing time of Glauber dynamics is $O_{\delta,\theta}(n)$ and $O_{\delta,\theta}(n \log n)$, where n is the number of vertices in G.*

Proposition 43 can be proved by applying the technique in [CLV21] to the marginal distribution μ_L and then compare some block dynamics on μ_L with the Glauber dynamics on μ. The proof in [CLV21] works for spin systems with bounded maximum degree. Although μ_L is not a spin system, the conditional independence results still hold on the power graph $G^2[V_L]$ and the proof technique can be applied. We give the proof of Proposition 43 in Appendix B.

With Proposition 43, we only need to consider the large degree case $\Delta \geq \Delta_0$, where the partition in Proposition 42 exists. Let $U_1, U_2, \ldots, U_k$ denote the partition of V_L. Similar to Section 4.1, we study the $k \leftrightarrow (k-\ell)$ down-up walk on the partition. The only difference is that we now consider the partition on V_L, so the down-up walk is defined on the marginal distribution μ_L. Specifically, in each step, given the current configuration $X \in \{-,+\}^{V_L}$, the down-up walk does as follows

- pick a subset $S \subseteq [k]$ with $|S| = \ell$ uniformly at random;
- resample $X_{U_S} \sim \mu_{L,U_S}(\cdot \mid X_{V_L \setminus U_S}) = \mu_{U_S}(\cdot \mid X_{V_L \setminus U_S})$, where $U_S = \cup_{i \in S} U_i$.

Since μ_L is the marginal distribution on V_L projected from μ, $\mu_{L,U_S}(\cdot \mid X_{V_L \setminus U_S})$ is the same as the conditional distribution $\mu_{U_S}(\cdot \mid X_{V_L \setminus U_S})$ induced from μ.

Let $S \subseteq [k]$ with size $|S| = s$. Let $U_{[k]\setminus S} = \uplus_{i \in [k]\setminus S} U_i$. For any $\tau \in [q]^{U_{[k]\setminus S}}$, we can similarly define the $s \leftrightarrow 1$ down-up walk for distribution μ_L as that in Section 4.1. The only difference between here

and Section 4.1 is that here we define down-up walks on μ_L, which is not a Gibbs distribution of a spin system but the down-up walks in Section 4.1 are defined for Gibbs distributions. However, both the local-to-global argument in Proposition 32 and the path argument in Lemma 33 work for general distributions. Hence, the same proof in Section 4.1 implies the following result.

Lemma 44. *The relaxation time of $k \leftrightarrow (k-2M)$ down-up walk for μ_L on partition $U_1, U_2, \ldots, U_k$ satisfies $T_{\mathrm{rel}}(k, k-2M) \leq 2^{k-2M}$.*

Lemma 44 shows that $k \leftrightarrow (k-2M)$ down-up walk for μ_L has a constant relaxation $O_{\theta,\delta}(1)$. Now, we relate this $k \leftrightarrow (k-2M)$ down-up walk for μ_L to the following block dynamics $\mathcal{B}$ on μ. The block dynamics maintains a hardcore configuration $X \in \{-,+\}^V$ on the whole graph G. In each step,

- pick a subset $S \subseteq [k]$ with $|S| = 2M$ uniformly at random;
- resample $X_{U_S \cup V_R} \sim \mu_{U_S \cup V_R}(\cdot \mid X_{V \setminus U_S})$.

The difference between the above block dynamics on μ and down-up walk on μ_L is that in the above dynamics, we always resample the configuration on V_R in every step. We prove the following result by comparing $\mathcal{B}$ to the $k \leftrightarrow (k-2M)$ down-up walk.

Lemma 45. *The relaxation time for block dynamics $\mathcal{B}$ is at most 2^{k-2M}.*

Proof. We decompose the $k \leftrightarrow (k-2M)$ down-up walk into two steps. Given the current configuration $X \in \{-,+\}^{V_L}$, the down operator D sample a uniformly random subset $S \subseteq [k]$ with $|S| = 2M$ and map X to $X_{V_L \setminus U_S}$, and the up operator complete the partial configuration $X_{V_L \setminus U_S}$ into a configuration $X \in \{-,+\}^{V_L}$ by sampling from the conditional distribution. Recall the relaxation time is equivalent to the contraction of the down-walk (Proposition 21). Lemma 44 implies that for any distribution π_L over $\{-,+\}^{V_L}$, it holds that

$$D_{\chi^2}(\pi_L D \,\|\, \mu_L D) \leq (1-C) D_{\chi^2}(\pi_L \,\|\, \mu_L), \tag{24}$$

where $0 < C = 2^{-k+2M} < 1$. Similarly, we can define a down operator $D_{\mathcal{B}}$ for the block dynamics $\mathcal{B}$. Given a full configuration $Y \in \{-,+\}^V$, let P_L be the projection operator that maps Y to Y_{V_L}. Then, the down operator $D_{\mathcal{B}} = P_L D$, where D is the down operator for the $(k-2M)$ down-up walk. For any distribution $\pi \in \{0,1\}^V$, let $\pi_L = \pi P_L$, we have

$$\begin{aligned}
D_{\chi^2}(\pi D_{\mathcal{B}} \,\|\, \mu D_{\mathcal{B}}) &= D_{\chi^2}(\pi P_L D \,\|\, \mu P_L D) \\
&= D_{\chi^2}(\pi_L D \,\|\, \mu_L D) \\
\text{(by (24))} \quad &\leq (1-C) D_{\chi^2}(\pi_L \,\|\, \mu_L) \\
&= (1-C) D_{\chi^2}(\pi P_L \,\|\, \mu P_L) \\
\text{(by data-processing inequality)} \quad &\leq (1-C) D_{\chi^2}(\pi \,\|\, \mu). \qquad (25)
\end{aligned}$$

The above inequality implies the relaxation time of $\mathcal{B}$. □

Finally, we can prove Theorem 5 by comparing $\mathcal{B}$ to the Glauber dynamics.

Proof of Theorem 5. Let $C = 2^{k-2M}$. Since $k = O(\theta + \frac{1}{\delta})$, we have $C = 2^{O(\theta+1/\delta)}$ is a constant. By

Lemma 45, we have the following block factorization of variance.

$$\forall f : \Omega(\mu) \to \mathbb{R}, \quad \mathbf{Var}_\mu[f] \le C\binom{k}{2M}^{-1} \sum_{S \in \binom{k}{2M}} \mu[\mathbf{Var}_{U_S \cup V_R}[f]]$$
$$= C\binom{k}{2M}^{-1} \sum_{S \in \binom{k}{2M}} \sum_{\tau \in \{-,+\}^{V_L \setminus U_S}} \mu_{V_L \setminus U_S}(\tau) \mathbf{Var}_{\mu^\tau}[f].$$

Given any pinning $\tau \in \{-,+\}^{V_L \setminus U_S}$, μ^τ is a hardcore model in $G[U_S \cup V_R]$ with boundary condition τ. By Proposition 42, the maximum degree of $G[U_S \cup V_R]$ is at most Δ_L. Since $\lambda \le (1-\delta)\lambda_c(\Delta_L)$, we know that the Glauber dynamics in μ^τ has the relaxation time $2^{O(1/\delta)}n$ [CFYZ21]. We remark that we can use [CFYZ21] because in the graph $G[U_S \cup V_R]$, the degree of *every* vertex is at most Δ_L. We cannot directly apply the previous result on G because we only have the degree bound for the left part of G. As a consequence,

$$\mathbf{Var}_{\mu^\tau}[f] \le 2^{O(1/\delta)} \sum_{u \in V_R \cup U_S} \mu^\tau[\mathbf{Var}_u[f]].$$

Putting everything together, we have

$$\mathbf{Var}_\mu[f] \le C \cdot 2^{O(1/\delta)} \binom{k}{2M}^{-1} \sum_{S \in \binom{k}{2M}} \sum_{\tau \in \{-,+\}^{V_L \setminus U_S}} \mu_{V_L \setminus U_S}(\tau) \sum_{u \in V_R \cup U_S} \mu^\tau[\mathbf{Var}_u[f]]$$
$$= C \cdot 2^{O(1/\delta)} \binom{k}{2M}^{-1} \sum_{S \in \binom{k}{2M}} \sum_{u \in V_R \cup U_S} \mu[\mathbf{Var}_u[f]].$$

In the above summation, every vertex $u \in V_L$ is counted for $\binom{k-1}{2M-1}$ times and every vertex $v \in V_R$ is counted for $\binom{k}{2M}$. Since the variance is non-negative, we have the following bound

$$\mathbf{Var}_\mu[f] \le C \cdot 2^{O(1/\delta)} \sum_{u \in V_L \cup V_R} \mu[\mathbf{Var}_u[f]].$$

This proves the $C \cdot 2^{O(1/\delta)} \cdot n = O_{\delta,\theta}(n)$ relaxation time for Glauber dynamics. □

In the rest part of this section, we will prove Theorem 6. The proof follows the same high-level plan as the proof of Theorem 15. First, we need to strengthen the partition used in Proposition 42 to its balanced version to apply the censoring inequality. By combining the proof for Proposition 42 and Proposition 37, one could prove the following result.

Proposition 46. *For any $k \ge \lceil 2\theta \rceil$, if $\Delta_L = \Omega(\theta \log(k\theta))$ and $|V_L| = \Omega(k \log k)$, then there exists a partition $V_L = U_1 \uplus U_2 \uplus \ldots \uplus U_k$ satisfying (23) and for every $i \in [k]$, $|U_i| \ge \frac{|V_L|}{2k}$.*

Technically, we need one more observation for the original proof to make it work for the bipartite hardcore model. Here, we run the same algorithm SimDownUp on the distribution μ. Recall $U_1, U_2, \cdots, U_k$ is a partition of V_L. So, the difference here is that $U_1, \cdots U_k$ are just disjoint sets rather than a partition of $V = V_L \uplus V_R$.

Actually, it is straightforward to notice that the same proof for Theorem 15 and Theorem 13 still works in this scenario provided that the down-up walk or Glauber dynamics on the conditional distribution $\mu[X_{U_R}]$ is fast mixing (recall that $\mu[X_{U_R}]$ is the notion we defined in Definition 34).

When $|R| = k - 2M$, the distribution $\mu[X_{U_R}]$ is a hardcore model on $G[U_S \cup V_R]$ with boundary condition X_{U_R}, where we let $U_S = V_L \setminus U_R$. By Proposition 42, the maximum degree of $G[U_S \cup V_R]$ is at most Δ_L. Since $\lambda \le (1-\delta)\lambda_c(\Delta_L)$, the Glauber dynamics in $\mu[X_{U_R}]$ is fast mixing.

Lemma 47 ([CFYZ22, CE22]). *When $\lambda \le (1-\delta)\lambda_c(\Delta)$, the Glauber dynamics for the hardcore model on any n-vertices graph with maximum degree Δ has mixing time $O_\delta(n \log n)$.*

When $|R| < k - 2M$, the mixing of the $(k - |R|) \leftrightarrow 1$-down-up walk on $\mu[X_{U_R}] = \mu_{U_S \cup V_R}(\cdot \mid X_{U_R})$ is guaranteed by the mixing of the $(k - |R|) \leftrightarrow 1$-down-up walk on $\mu_{U_S}(\cdot \mid X_{U_R})$. This is already observed in [CLY23] and we include this result to prove Theorem 6. Note that the original proof in [CLY23] is only for Glauber dynamics. But is is straight forward to generalize the proof for the down-up walks (i.e., block dynamics).

Lemma 48 ([CLY23, Lemma 102]). *Let P_1 be the $(k - |R|) \leftrightarrow 1$-down-up walk on $\mu[X_{U_R}]$ and let P_2 be the $(k - |R|) \leftrightarrow 1$-down-up walk on $\mu_{U_S}(\cdot \mid X_{U_R})$, then we have for every $x \in \Omega(\mu[X_{U_R}])$ every $t \ge 1$,*

$$\mathcal{D}_{\mathrm{TV}}\left(P_1^t(x,\cdot) \parallel \mu[X_{U_R}]\right) \le \mathcal{D}_{\mathrm{TV}}\left(P_2^t(y,\cdot) \parallel \mu_{U_S}(\cdot \mid X_{U_R})\right),$$

where $y = x_{U_S}$ is the projection of x onto U_S.

Proof of Theorem 6. The proof of Theorem 6 is finished by go through the same proof as Theorem 13 and Theorem 15 with the mixing time of Glauber dynamics and down-up walks required by the proof being guaranteed by Lemma 47 and Lemma 48. □

9 List Coloring

Theorem 1 is a consequence of Theorem 9 and Theorem 4 is a consequence of Theorem 13. The proofs of Theorem 1 and Theorem 4 are given in Section 2.2. The only missing part is the following coupling independence result.

Lemma 49 ([FGYZ21]). *Let $\delta \in (0, 0.5)$ be a constant and $G = (V, E)$ be a triangle-free graph with maximum degree $\Delta \ge 3$. Let $\mathcal{L} = (\mathcal{L}_v)_{v \in V}$ be the color list such that*

$$\forall v \in V, \quad |\mathcal{L}_v| - \deg_G(v) \ge (\alpha^\star + \delta - 1)\Delta, \tag{26}$$

where $\alpha^\star \approx 1.763$ is the unique solution to the equation $\alpha = \exp(1/\alpha)$. Let $\mu = \mu_{G,\mathcal{L}}$ be the uniform distribution over all the $\mathcal{L}$-list-colorings of G. Then, it holds that μ is $O(1/\delta)$-coupling independence.

In the above lemma, we assume $\delta < 0.5$. If $\delta \ge 0.5$, then $|\mathcal{L}_v| - \deg_G(v) > 1.1\Delta$ and optimal relaxation and mixing time can be obtained from path coupling.

In [FGYZ21], the spectral independence result is proved for list-coloring. The above lemma cannot be obtained by applying Lemma 39, because list-coloring is not two spin systems. However, the proof technique in [FGYZ21] also yields the coupling independence result.

Given a list coloring instance $G = (V, E), \mathcal{L}$, a pinning $\tau \in \otimes_{v \in V \setminus \Lambda} \mathcal{L}_v$ and a vertex $v \in \Lambda$, the proof in [FGYZ21] builds the following coupling between $\mu_\Lambda^{\tau \wedge v^a}$ and $\mu_\Lambda^{\tau \wedge v^b}$, where a, b are two different colors in L_v and $\mu_\Lambda^{\tau \wedge v^a}$ is the marginal distribution on v conditional on τ and v taking the color a. By self-reducibility, we can turn the pinning τ to a new list-coloring instance in graph $G \leftarrow G[\Lambda]$. For any $u \in \Lambda$, $\mathcal{L}_u \leftarrow \mathcal{L}_u \setminus \{\tau_w \mid w \notin \Lambda \wedge \{u, w\} \in E\}$. The proof of [FGYZ21] constructs the following coupling, which is essentially the recursive coupling in [GMP05]. Let $u_1, u_2, \ldots, u_d$ denote the neighbors of v in G. We split v into d vertices $v_1, v_2, \ldots, v_d$ and connect v_i with u_i. Denote the graph as G_v. Let π denote the uniform distribution of list-colorings in G_v, where each color list of v_i is the same as the list of v. Define the pinning σ_i on $\{v_1, v_2, \ldots, v_d\}$ as $\sigma_i(v_j) = b$ for $j \le i$ and $\sigma_i(v_j) = a$ for $j > i$. We only need to couple $\pi^{\sigma_0} = \mu_\Lambda^{\tau \wedge v^a}$ and $\pi^{\sigma_d} = \mu_\Lambda^{\tau \wedge v^b}$. To do this, we couple each adjacent pair $\pi^{\sigma_{i-1}}$ and π^{σ_i} and merge them via path coupling. To couple $X \sim \pi^{\sigma_{i-1}}$ and $Y \sim \pi^{\sigma_i}$, the analysis in [FGYZ21] essentially consider the following coupling. Note that σ_i and σ_{i-1} differ only at v_i and the only neighbor of v_i is u_i.

- Couple $X(u_i)$ and $Y(u_i)$ via the optimal coupling of marginal distributions at u_i;
- If $X(u_i) = Y(u_i)$, then couple all other vertices perfectly;
- If $X(u_i) \neq Y(u_i)$, then use the above process recursively (splitting u_i and applying path coupling) to couple $\pi^{\sigma_{i-1}(-i)\wedge X(u_i)}$ and $\pi^{\sigma_i(-i)\wedge Y(u_i)}$, where $\sigma_i(-i) = \{\sigma_i(v_j) \mid j \neq i\}$ and we can remove v_i because conditional on u_i, v_i has no effect on other vertices.

The above process gives a coupling (X, Y) between $\mu_\Lambda^{\tau\wedge v^a}$ and $\mu_\Lambda^{\tau\wedge v^b}$. One can extend it to a coupling between $\mu^{\tau\wedge v^a}$ and $\mu^{\tau\wedge v^b}$ by letting $X_{V\setminus\Lambda} = Y_{V\setminus\Lambda} = \tau$. The analysis in [FGYZ21] bound the discrepancy of the above recursively coupling such that

$$\mathbb{E}\left[H_\rho(X,Y)\right] \leq \frac{9}{2\delta} + 1[X_v \neq Y_v] = \frac{9}{2\delta} + 1 = O\left(\frac{1}{\delta}\right),$$

where $H_\rho(X, Y)$ is the standard Hamming distance ($\rho(u) = 1$ for all $u \in V$) between X and Y.

Acknowledgment We would like to thank Heng Guo for bringing the paper [JSS21] to our attention and for the helpful discussions at the beginning of this project. We thank Eric Vigoda for suggesting a better statement of technical results. We also thank Charlie Carlson, Yitong Yin, and Xinyuan Zhang for their helpful discussions.

References

[AJK+22] Nima Anari, Vishesh Jain, Frederic Koehler, Huy Tuan Pham, and Thuy-Duong Vuong. Entropic independence: optimal mixing of down-up random walks. In *STOC*, pages 1418–1430, 2022.

[AL20] Vedat Levi Alev and Lap Chi Lau. Improved analysis of higher order random walks and applications. In *STOC*, pages 1198–1211, 2020.

[ALO20] 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.

[ALOV19] Nima Anari, Kuikui Liu, Shayan Oveis Gharan, and Cynthia Vinzant. Log-concave polynomials II: high-dimensional walks and an FPRAS for counting bases of a matroid. In *STOC*, pages 1–12, 2019.

[BCC+21] Antonio Blanca, Pietro Caputo, Zongchen Chen, Daniel Parisi, Daniel Štefankovič, and Eric Vigoda. On mixing of markov chains: Coupling, spectral independence, and entropy factorization. *arXiv preprint arXiv:2103.07459*, 2021.

[BCKL13] Christian Borgs, Jennifer Chayes, Jeff Kahn, and László Lovász. Left and right convergence of graphs with bounded degree. *Random Structures Algorithms*, 42(1):1–28, 2013.

[BD97] Russ Bubley and Martin Dyer. Path coupling: A technique for proving rapid mixing in markov chains. In *FOCS*, pages 223–231, 1997.

[Cap23] Pietro Caputo. Lecture notes on Entropy and Markov Chains. *Preprint, available from: http://www.mat.uniroma3.it/users/caputo/entropy.pdf*, 2023.

[CDM+19] Sitan Chen, Michelle Delcourt, Ankur Moitra, Guillem Perarnau, and Luke Postle. Improved bounds for randomly sampling colorings via linear programming. In *SODA*, pages 2216–2234, 2019.

[CE22] Yuansi Chen and Ronen Eldan. Localization schemes: A framework for proving mixing bounds for markov chains. In *FOCS*, pages 110–122, 2022.

[Ces01] Filippo Cesi. Quasi-factorization of the entropy and logarithmic Sobolev inequalities for Gibbs random fields. *Probab. Theory Related Fields*, 120(4):569–584, 2001.

[CFYZ21] 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, 2021.

[CFYZ22] Xiaoyu Chen, Weiming Feng, Yitong Yin, and Xinyuan Zhang. Optimal mixing for two-state anti-ferromagnetic spin systems. In *FOCS*, pages 588–599, 2022.

[CG24] Zongchen Chen and Yuzhou Gu. Fast sampling of b-matchings and b-edge covers. In *SODA*, pages 4972–4987, 2024.

[CGG+16] Jin-Yi Cai, Andreas Galanis, Leslie Ann Goldberg, Heng Guo, Mark Jerrum, Daniel Štefankovič, and Eric Vigoda. #BIS-hardness for 2-spin systems on bipartite bounded degree graphs in the tree non-uniqueness region. *J. Comput. System Sci.*, 82(5):690–711, 2016.

[CGŠV21] Zongchen Chen, Andreas Galanis, Daniel Štefankovič, and Eric Vigoda. Rapid mixing for colorings via spectral independence. In *SODA*, pages 1548–1557, 2021.

[CGŠV22] Zongchen Chen, Andreas Galanis, Daniel Štefankovič, and Eric Vigoda. Sampling colorings and independent sets of random regular bipartite graphs in the non-uniqueness region. In *SODA*, pages 2198–2207, 2022.

[CLMM23] Zongchen Chen, Kuikui Liu, Nitya Mani, and Ankur Moitra. Strong spatial mixing for colorings on trees and its algorithmic applications. In *FOCS*, pages 810–845, 2023.

[CLV20] Zongchen Chen, Kuikui Liu, and Eric Vigoda. Rapid mixing of Glauber dynamics up to uniqueness via contraction. In *FOCS*, pages 1307–1318, 2020.

[CLV21] Zongchen Chen, Kuikui Liu, and Eric Vigoda. Optimal mixing of Glauber dynamics: entropy factorization via high-dimensional expansion. In *STOC*, pages 1537–1550, 2021.

[CLY23] Xiaoyu Chen, Jingcheng Liu, and Yitong Yin. Uniqueness and rapid mixing in the bipartite hardcore model. In *FOCS*, pages 1991–2005. IEEE, 2023.

[CMT15] 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.

[CP20] Sarah Cannon and Will Perkins. Counting independent sets in unbalanced bipartite graphs. In *SODA*, pages 1456–1466, 2020.

[CP21] Pietro Caputo and Daniel Parisi. Block factorization of the relative entropy via spatial mixing. *Comm. Math. Phys.*, 388(2):793–818, 2021.

[CZ23] Xiaoyu Chen and Xinyuan Zhang. A near-linear time sampler for the Ising model with external field. In *SODA*, pages 4478–4503, 2023.

[DGU14] Martin Dyer, Catherine Greenhill, and Mario Ullrich. Structure and eigenvalues of

heat-bath Markov chains. *Linear Algebra Appl.*, 454:57–71, 2014.

[EHŠ+19] Charilaos Efthymiou, Thomas P. Hayes, Daniel Štefankovič, Eric Vigoda, and Yitong Yin. Convergence of MCMC and loopy BP in the tree uniqueness region for the hard-core model. *SIAM J. Comput.*, 48(2):581–643, 2019.

[EL75] Paul Erdős and László Lovász. Problems and results on 3-chromatic hypergraphs and some related questions. *Infinite and finite sets, volume 10 of Colloquia Mathematica Societatis János Bolyai*, pages 609–628, 1975.

[FGYZ21] Weiming Feng, Heng Guo, Yitong Yin, and Chihao Zhang. Rapid mixing from spectral independence beyond the boolean domain. In *SODA*, pages 1558–1577, 2021.

[GKM13] David Gamarnik, Dmitriy Katz, and Sidhant Misra. Strong spatial mixing of list coloring of graphs. *Random Structures Algorithms*, 2013.

[GMP05] Leslie Ann Goldberg, Russell A. Martin, and Mike Paterson. Strong spatial mixing with fewer colors for lattice graphs. *SIAM J. Comput.*, 35(2):486–517, 2005.

[HV06] Thomas P. Hayes and Eric Vigoda. Coupling with the stationary distribution and improved sampling for colorings and independent sets. *Ann. Appl. Probab.*, 16(3):1297–1318, 2006.

[Jer95] Mark Jerrum. A very simple algorithm for estimating the number of k-colorings of a low-degree graph. *Random Struct. Algorithms*, 7(2):157–165, 1995.

[Jer24] Mark Jerrum. Glauber dynamics for the hard-core model on bounded-degree H-free graphs. *CoRR*, abs/2404.07615, 2024.

[JKP20] Matthew Jenssen, Peter Keevash, and Will Perkins. Algorithms for #BIS-hard problems on expander graphs. *SIAM J. Comput.*, 49(4):681–710, 2020.

[JPP23] Matthew Jenssen, Will Perkins, and Aditya Potukuchi. Approximately counting independent sets in bipartite graphs via graph containers. *Random Structures Algorithms*, 63(1):215–241, 2023.

[JPV22] Vishesh Jain, Huy Tuan Pham, and Thuy Duong Vuong. Spectral independence, coupling with the stationary distribution, and the spectral gap of the Glauber dynamics. *Inf. Process. Lett.*, 177:106268, 2022.

[JSS21] Vishesh Jain, Ashwin Sah, and Mehtaab Sawhney. Perfectly sampling $k \geq (8/3 + o(1))\Delta$-colorings in graphs. In *STOC*, pages 1589–1600, 2021.

[Kel85] Frank P Kelly. Stochastic models of computer communication systems. *Journal of the Royal Statistical Society: Series B (Methodological)*, 47(3):379–395, 1985.

[Liu21] Kuikui Liu. From coupling to spectral independence and blackbox comparison with the down-up walk. *arXiv preprint arXiv:2103.11609*, 2021.

[LL15] Jingcheng Liu and Pinyan Lu. FPTAS for #BIS with degree bounds on one side. In *STOC*, pages 549–556, 2015.

[LLLM22] Chao Liao, Jiabao Lin, Pinyan Lu, and Zhenyu Mao. An FPTAS for the hardcore model on random regular bipartite graphs. *Theoret. Comput. Sci.*, 929:174–190, 2022.

[LLY13] Liang Li, Pinyan Lu, and Yitong Yin. Correlation decay up to uniqueness in spin systems. In *SODA*, pages 67–84, 2013. arXiv:1111.7064.

[LPW17] David A. Levin, Yuval Peres, and Elizabeth L. Wilmer. *Markov chains and mixing times*. American Mathematical Society, Providence, RI, 2017.

[LV99] Michael Luby and Eric Vigoda. Fast convergence of the glauber dynamics for sampling independent sets. *Random Structures Algorithms*, 15(3-4):229–241, 1999.

[MM09] Marc Mezard and Andrea Montanari. *Information, physics, and computation*. Oxford University Press, 2009.

[MSW03] Fabio Martinelli, Alistair Sinclair, and Dror Weitz. The ising model on trees: Boundary conditions and mixing time. In *FOCS*, pages 628–639, 2003.

[MT10] Robin A Moser and Gábor Tardos. A constructive proof of the general Lovász local lemma. *Journal of the ACM (JACM)*, 57(2):11, 2010.

[PW13] Yuval Peres and Peter Winkler. Can extra updates delay mixing? *Comm. Math. Phys.*, 323(3):1007–1016, 2013.

[Sly10] Allan Sly. Computational transition at the uniqueness threshold. In *FOCS*, pages 287–296, 2010.

[Vig00] Eric Vigoda. Improved bounds for sampling colorings. *J. Math. Phys.*, 41(3):1555–1569, 2000.

[Wei06] Dror Weitz. Counting independent sets up to the tree threshold. In *STOC*, pages 140–149, 2006.

A Local to Global Proof

Proof of Proposition 32. Let $f : \Omega(\mu) \to \mathbb{R}$ be a function. Let $\mathbf{Var}_\mu[f]$ denote the variance of f with respect to μ. For any subset $S \subseteq V$, let $\mu[\mathbf{Var}_S[f]]$ denote the average of the variance $\mathbf{Var}_{\mu^\sigma}[f]$ where $\sigma \sim \mu_{V\setminus S}$. For any set U, we use $R \sim \binom{U}{i}$ to denote sample a uniform subset of $R \subseteq U$ with $|R| = i$. Note that when $\ell = 0$, then Proposition 32 becomes trivial. Hence, without loss of generality, we may assume $\ell \geq 1$. By the definition of relaxation time $T^{\varnothing}_{\mathrm{rel}}(k,1)$,

$$\mathbf{Var}_\mu[f] \leq T^{\varnothing}_{\mathrm{rel}}(k,1)\mathbb{E}_{R\sim\binom{[k]}{1}}\left[\mu[\mathbf{Var}_{\overline{U_R}}[f]]\right] \leq \gamma_0\mathbb{E}_{R\sim\binom{[k]}{1}}\left[\mu[\mathbf{Var}_{\overline{U_R}}[f]]\right],$$

where for convenience, we use $\overline{U_R}$ to denote the set $V \setminus R$. In the above inequality, one needs to deal with variance $\mathbf{Var}_{\mu^\tau}[f]$, where $\tau \sim \mu_{U_R}$. To bound it, we consider the $(k-1) \leftrightarrow 1$ down-up walk on distribution $\pi = \mu^\tau_{U_R}$ to obtain

$$\forall g \in \Omega(\pi), \quad \mathbf{Var}_\pi[g] \leq T^{\tau}_{\mathrm{rel}}(k-1,1)\mathbb{E}_{R'\sim\binom{[k]\setminus R}{1}}\left[\pi[\mathbf{Var}_{\overline{U_{R'\cup R}}}[g]]\right].$$

Since the configuration on U_R is fixed as τ in μ^τ, the above inequality implies

$$\mathbf{Var}_{\mu^\tau}[f] \leq T^{\tau}_{\mathrm{rel}}(k-1,1)\mathbb{E}_{R'\sim\binom{[k]\setminus R}{1}}\left[\mu^\tau[\mathbf{Var}_{\overline{U_{R'\cup R}}}[f]]\right] \leq \gamma_1\mathbb{E}_{R'\sim\binom{[k]\setminus R}{1}}\left[\mu^\tau[\mathbf{Var}_{\overline{U_{R'\cup R}}}[f]]\right].$$

Combining the above inequalities, we have

$$\mathbf{Var}_\mu[f] \leq \gamma_k\gamma_{k-1}\mathbb{E}_{R\sim\binom{[k]}{1}}\left[\mathbb{E}_{R'\sim\binom{[k]\setminus R}{1}}\left[\mu[\mathbf{Var}_{\overline{U_{R\cup R'}}}[f]]\right]\right] = \gamma_k\gamma_{k-1}\mathbb{E}_{R\sim\binom{[k]}{2}}\left[\mu[\mathbf{Var}_{\overline{U_R}}[f]]\right].$$

Using the above argument iteratively, we have

$$\begin{aligned}\mathbf{Var}_{\mu}[f] &\leq \gamma_0\gamma_1\mathbb{E}_{R\sim\binom{[k]}{2}}\left[\mu[\mathbf{Var}_{\overline{U_R}}[f]]\right]\\ &\leq \gamma_0\gamma_1\gamma_2\mathbb{E}_{R\sim\binom{[k]}{3}}\left[\mu[\mathbf{Var}_{\overline{U_R}}[f]]\right]\\ &\leq \dots\\ &\leq \left(\prod_{i=0}^{\ell-1}\gamma_i\right)\mathbb{E}_{R\sim\binom{[k]}{\ell}}\left[\mu[\mathbf{Var}_{U_B}[f]]\right].\end{aligned}$$

This proves the proposition. □

B Bipartite Graph Hardcore: Bounded Degree Case

In this section, we give a brief proof of Proposition 43. The proof applies techniques in [CLV21] to this problem. We need to do some modifications because we only have coupling independence on the left part.

The maximum degree Δ_R in the right part is bounded by $\theta\Delta_L = O_{\theta,\delta}(1)$. If $\lambda < \frac{1}{2\theta\Delta_L}$, then the whole hardcore model satisfies $\lambda < \frac{1}{2\Delta}$, where $\Delta = \max\{\Delta_L, \Delta_R\}$. The proposition follows from standard path coupling [BD97].

Now, we assume $\frac{1}{2\theta\Delta_L} \leq \lambda \leq (1-\delta)\lambda_c(\Delta_L)$. As a consequence, $\lambda = \Theta_{\theta,\delta}(\frac{1}{\Delta})$. For the distribution μ on the entire graph, for any pinning $\sigma \in \{-,+\}^{\Lambda}$ with $\Lambda \subseteq V_L \cup V_R$ and $v \notin \Lambda$,

$$\min\left(\mu_v^{\sigma}(+), \mu_v^{\sigma}(-)\right) \geq b(\theta,\delta) = \Omega_{\theta,\delta}(1). \tag{27}$$

This property is called b-marginal boundedness of μ in [CLV21]. It is easy to see the b-marginal boundedness of μ implies the b-marginal boundedness of μ_L. Since the coupling independence implies the spectral independence, by Lemma 40, μ_L is $O(1/\delta)$-spectrally independent. Define

$$\alpha := \left(\frac{b^2}{100e(\Delta_L+\Delta_R)}\right)^{\Delta_L} = \Theta_{\delta,\theta}(1). \tag{28}$$

By [CLV21, Lemma 2.5], the distribution μ_L satisfies the following factorization of entropy

$$\forall f : \{-,+\}^{V_L} \to \mathbb{R}_{\geq 0}, \quad \mathbf{Ent}_{\mu_L}[f] \leq \frac{C(\theta,\delta)}{\binom{n_L}{\ell}} \sum_{S\in\binom{V_L}{\ell}} \mu_L[\mathbf{Ent}_S[f]], \tag{29}$$

where $n_L = |V_L|$ and $\ell = \lceil \alpha n_L \rceil$. The constant C depends only on the marginal boundedness parameter b, spectral independence parameter $O(1/\delta)$, and the parameter α. Since all of them depend on δ, θ, it holds that $C = C(\delta,\theta)$ is a constant. [CLV21, Lemma 2.5] holds for $n_L \geq n_0(\theta,\delta)$. However, if $n_L \leq n_0(\theta,\delta)$, then $n_L, \Delta_L, \Delta_R, \lambda = \Theta_{\theta,\delta}(1)$ are all constants, we can take $C(\theta,\delta)$ sufficiently large to make the above inequality hold.

The inequality mentioned above is referred to as block factorization of entropy. In a similar manner to the proof of Lemma 45, we consider two Markov chains. The first one is the $n_L \leftrightarrow n_L - \ell$ down-up walk for μ_L. Given any configuration $X \in \{-,+\}^{V_L}$, it updates X as follows.

- Down-Walk D: sample $S \in \binom{V_L}{\ell}$ uniformly at random and update X to $X_{V_L\setminus S}$;
- Up-Walk U: extend $X_{V_L\setminus S}$ to a configuration on V_L by sampling $X_S \sim \mu_{L,S}^{X_{V_L\setminus S}}$.

The second Markov chain is the block dynamics $\mathcal{B}$ for μ. Given any configuration $X \in \{-,+\}^{V_L \cup V_R}$ of the entire bipartite graph $G = (V_L \cup V_R, E)$, it updates X as follows.

- Down-Walk $D_{\mathcal{B}}$: sample $S \in \binom{V_L}{\ell}$ uniformly at random and update X to $X_{V_L \setminus S}$;
- Up-Walk $U_{\mathcal{B}}$: extend $X_{V_L \setminus S}$ to a configuration on $V_L \cup V_R$ by sampling $X_{S \cup V_R} \sim \mu^{X_{V_L \setminus S}}_{S \cup V_R}$.

The factorization in (29) implies the down-walk D contracts the KL-divergence by a factor of $1 - \frac{1}{C(\theta,\delta)}$ [CLV21, Lemma 2.7]. Using the same analysis as that for (25) but replacing χ^2-divergence with KL-divergence, we have for any distribution ν over $\{-,+\}^{V_L \cup V_R}$,

$$\mathcal{D}_{\mathrm{KL}}\left(\nu D_{\mathcal{B}} \parallel \mu D_{\mathcal{B}}\right) \leq \left(1 - \frac{1}{C(\theta,\delta)}\right) \mathcal{D}_{\mathrm{KL}}\left(\nu \parallel \mu\right).$$

The above contraction of KL-divergence implies the following factorization of entropy

$$\forall f : \Omega(\mu) \to \mathbb{R}_{\geq 0}, \quad \mathbf{Ent}_{\mu}\left[f\right] \leq \frac{C(\theta,\delta)}{\binom{n_L}{\ell}} \sum_{S \in \binom{V_L}{\ell}} \mu[\mathbf{Ent}_{S \cup V_R}\left[f\right]].$$

Consider a subset $S \subseteq V_L$. Let $\mathcal{C}(S \cup V_R)$ denote the set of connected components graph $G[S \cup V_R]$. Given any $\tau \in \{-,+\}^{V_L \setminus S}$, in the conditional distribution μ^{τ}, all components in $\mathcal{C}(S \cup V_R)$ are mutually independent. By [Ces01, CMT15] (also see [CLV21, Lemma 4.1]),

$$\mu[\mathbf{Ent}_{S \cup V_R}\left[f\right]] \leq \sum_{U \in \mathcal{C}(S \cup V_R)} \mu[\mathbf{Ent}_{U}\left[f\right]]. \tag{30}$$

For any U, any $\xi \in \{-,+\}^{(V_L \cup V_R) \setminus U}$, by [CLV21, Lemma 4.2],

$$\mathbf{Ent}_{\mu^{\xi}}\left[f\right] \leq \frac{3|U|^2 \log(1/b)}{2b^{2|U|+2}} \sum_{v \in U} \mu^{\xi}[\mathbf{Ent}_{v}\left[f\right]], \tag{31}$$

where b is defined in (27). Finally, we prove the following lemma for bounding the size of components, which is similar to [CLV21, Lemma 4.3].

Lemma 50. *If we sample $S \in \binom{V_L}{\ell}$ uniformly at random, where $\ell = \lceil \alpha n_L \rceil$, then for any vertex $v \in V_L \cup V_R$,*

$$\mathbf{Pr}\left[|S_v| \geq k\right] \leq (e(\Delta_L + \Delta_R))^k (2\alpha)^{(k-1)/\Delta_L},$$

where S_v is the component in $G[S \cup V_R]$ containing v and $S_v = \varnothing$ if $v \notin S \cup V_R$.

Proof. Fix a vertex $v \in V_L \cup V_R$. The maximum degree Δ of the graph is $\max\{\Delta_L, \Delta_R\}$. The number of $U \subseteq V_L \cup V_R$ such $v \in U$, $G[U]$ is connected and $|U| = k$ is at most $(e\Delta)^k$ [BCKL13]. Note that $G[U]$ has at least $k-1$ edges because it is connected. Then, $G[U]$ has at least $\ell_k = \lceil (k-1)/\Delta_L \rceil$ vertices from V_L, because every edge has one vertex in V_L. By a union bound,

$$\mathbf{Pr}\left[|S_v| \geq k\right] \leq (e\Delta)^k \frac{\binom{n_L - \ell_k}{\ell - \ell_k}}{\binom{n_L}{\ell}} \leq (e\Delta)^k \frac{\ell}{n_L} \cdot \frac{\ell - 1}{n_L - 1} \cdot \ldots \cdot \frac{\ell - \ell_k + 1}{n_L - \ell_k + 1} \leq (e\Delta)^k (2\alpha)^{\ell_k}.$$

The lemma holds because $\Delta \leq \Delta_L + \Delta_R$ and $\alpha < 1/2$. □

Hence, we can bound the entropy of f as follows

$$\begin{aligned}
\mathbf{Ent}_\mu[f] &\le \frac{C(\theta,\delta)}{\binom{n_L}{\ell}} \sum_{S\in\binom{V_L}{\ell}} \mu[\mathbf{Ent}_{S\cup V_R}[f]] \\
\text{(by (30))} \quad &\le \frac{C(\theta,\delta)}{\binom{n_L}{\ell}} \sum_{S\in\binom{V_L}{\ell}} \sum_{U\in C(S\cup V_R)} \mu[\mathbf{Ent}_U[f]] \\
\text{(by (31))} \quad &\le \frac{C(\theta,\delta)}{\binom{n_L}{\ell}} \sum_{S\in\binom{V_L}{\ell}} \sum_{U\in C(S\cup V_R)} \frac{3|U|^2\log(1/b)}{2b^{2|U|+2}} \sum_{v\in U} \mu[\mathbf{Ent}_v[f]] \\
&\le \frac{3C(\theta,\delta)\log(1/b)}{2b^2} \sum_{v\in V_L\cup V_R} \mu[\mathbf{Ent}_v[f]] \sum_{k\ge 0} \mathbf{Pr}_{S\sim\binom{n_L}{\ell}}[|S_v|=k]\,\frac{3k^2}{b^{2k}} \\
\text{(by Lemma 50)} \quad &\le \frac{3C(\theta,\delta)\log(1/b)}{2b^2} \sum_{v\in V_L\cup V_R} \mu[\mathbf{Ent}_v[f]] \sum_{k\ge 0} (e(\Delta_L+\Delta_R))^k (2\alpha)^{(k-1)/\Delta_L} \frac{3k^2}{b^{2k}}.
\end{aligned}$$

Note that both C, b, Δ_L, Δ_R are constants depending on δ and θ. By the definition of α in (28),

$$(e(\Delta_L+\Delta_R))^{k-1}(2\alpha)^{(k-1)/\Delta_L} = \left(\frac{b^2}{50}\right)^{k-1}.$$

We have

$$\mathbf{Ent}_\mu[f] = O_{\theta,\delta}(1) \sum_{v\in V_L\cup V_R} \mu[\mathbf{Ent}_v[f]] \sum_{k\ge 0} \left(\frac{1}{50}\right)^{k-1} k^2 = O_{\theta,\delta}(1) \sum_{v\in V_L\cup V_R} \mu[\mathbf{Ent}_v[f]].$$

The above inequality is the approximate tensorization of entropy for μ, which implies $O_{\theta,\delta}(n)$ relaxation time and $O_{\theta,\delta}(n\log n)$ mixing time for Glauber dynamics [CLV21, Fact 3.5], where $n = |V_L \cup V_R|$.