A NEAR-LINEAR TIME SAMPLER FOR THE ISING MODEL WITH EXTERNAL FIELD

XIAOYU CHEN AND XINYUAN ZHANG

ABSTRACT. We give a near-linear time sampler for the Gibbs distribution of the ferromagnetic Ising models with edge activities $\beta > 1$ and external fields $\lambda < 1$ (or symmetrically, $\lambda > 1$) on general graphs with bounded or unbounded maximum degree.

Our algorithm is based on the field dynamics given in [CFYZ21b]. We prove the correctness and efficiency of our algorithm by establishing spectral independence of distribution of the random cluster model and the rapid mixing of Glauber dynamics on the random cluster model in a low-temperature regime, which may be of independent interest.

1. INTRODUCTION

The Ising model [Isi25] introduced by Ising and Lenz is an extensively studied statistical physics model which leads to many inspiring discoveries in physics, discrete probability, machine learning, and theoretical computer science. Let $G = (V, E)$ be an undirected graph with n vertices and m edges, $\beta \in (1, +\infty)^E$ be the *edge activities*, and $\lambda \in [0, 1]^V$ be the *external fields*. The *Gibbs distribution* $\mu_{\beta,\lambda}^{\mathrm{Ising}}$ over 2^V of the *ferromagnetic Ising model* is defined by

$$\forall S \subseteq V, \quad \mu_{\beta,\lambda}^{\mathrm{Ising}}(S) := \frac{1}{Z_{\beta,\lambda}^{\mathrm{Ising}}} \prod_{e \in m(S)} \beta_e \prod_{v \in S} \lambda_v,$$

where $m(S) := \{e \in E \mid e \cap S = e \text{ or } e \cap S = \varnothing\}$ denotes the set of "monochromatic" edges, and $Z_{\beta,\lambda}^{\mathrm{Ising}} := \sum_{S \subseteq V} \prod_{e \in m(S)} \beta_e \prod_{v \in S} \lambda_v$ is known as the partition function. A major problem is to sample from the Gibbs distribution of the ferromagnetic Ising model.

One of the most well-known approaches is the Markov Chain Monte Carlo (MCMC) method. The Glauber dynamics, also known as the Gibbs sampler, is an example of this method. There are numerous researches establishing the rapid mixing results of Glauber dynamics [MS13, CLV20, CLV21a, CFYZ21b, AJK+21a, AJK+21b, CE22] when $\beta < \beta_c(\Delta)$, where Δ is the maximum degree of a graph G and $\beta_c(\Delta) := \Delta/(\Delta - 2)$ is the critical threshold. However, when $\beta > \beta_c(\Delta)$, it is known that there exist graphs such that the Glauber dynamics is exponentially slow in the size of the graph [GM07].

Even though the Glauber dynamics fails to be efficient, there still exist fast algorithms to sample from the Gibbs distribution of the ferromagnetic Ising model. The random cluster model [FK72, For72a, For72b] and subgraph-world model are two statistical mechanics models that are closely related to the Ising model. Leveraging the connection between the partition function of the Ising model and the subgraph-world model [NM53], [JS93] showed that the 1/2-lazy Metropolis chain on the subgraph-world model converges rapidly by using the technique of canonical path [JS89]. The following works [GJ18, FGW22a] established a similar mixing time of Glauber dynamics on the random cluster model via multicommodity flow based on the canonical path in [JS93] and the coupling in [GJ09]. All these results can be translated into fast Ising samplers that run in time $\widetilde{O}_{\beta,\lambda}(m^3)$ when $\beta > 1$ and $\lambda < 1$. Furthermore, these samplers also work when $\lambda = 1$, where the

(Xiaoyu Chen, Xinyuan Zhang) STATE KEY LABORATORY FOR NOVEL SOFTWARE TECHNOLOGY, NANJING UNIVERSITY, 163 XIANLIN AVENUE, NANJING, JIANGSU PROVINCE, CHINA. E-mails: chenxiaoyu233@smail.nju.edu.cn, zhangxy@smail.nju.edu.cn

running time degenerates to $\widetilde{O}_\beta(n^4 m^3)$. It is also worth mentioning that for the case $\lambda = 1$, there is a specific Markov chain called worm process on the Prokof'ev-Svistunov measure of the Ising model studied in statistical physics, which is proved to be rapid mixing [CGHT16].

With bounded degree assumption, [CLV21b] proved the optimal mixing time of Glauber dynamics on the subgraph-world model via spectral independence, implying a fast Ising sampler that runs in $\Delta^{O_{\beta,\lambda}(\Delta)} \cdot \widetilde{O}(m)$ when $\beta > 1$ and $\lambda < 1$.

Apart from MCMC based method, algorithms based on zero-freeness property and polymer model also achieve polynomial running time with restrictions. By the celebrated Lee-Yang circle theorem [LY52] and the polynomial interpolation algorithm framework [Bar16, PR17, PR19], there is an algorithm for sampling the ferromagnetic Ising model that runs in time $n^{O_\lambda(\log \Delta)}$ for $\beta > 1$ and $\lambda < 1$ [LSS19], where Δ is the maximum degree of the graph. While algorithms based on the polymer model usually require graph G to be an α-expander for some constant $\alpha > 0$ and parameter $\beta = \Omega(\log(\Delta)/\alpha)$. An algorithm of this type with running time $n^{O_{\beta,\alpha}(\log \Delta)}$ was given in [JKP20], and follow-up works [CGG+21, BCP22a] improved the running time to $\widetilde{O}_{\beta,\alpha}(n)$.

Besides, there are many other fast samplers for the Ising model on special families of graphs [MSW03, GŠV19, BCŠV21, BCP+22b, GS22], such as lattice graph and torus graph.

In summary, no algorithms in previous studies run faster than cubic time without any assumption on graphs or parameters. It is natural to ask the following question:

Are there faster algorithms to sample from Gibbs distribution of the Ising model in the general case?

In this paper, we answer this question in the affirmative.

Theorem 1.1. *Let $\delta_\beta, \delta_\lambda \in (0, 1)$ be constants, and μ be the Gibbs distribution of the ferromagnetic Ising model specified by graph $G = (V, E)$, parameters $\beta \in [1 + \delta_\beta, +\infty)^E$ and $\lambda \in [0, 1 - \delta_\lambda]^V$. There exists an algorithm that samples X satisfying $d_{TV}(X, \mu) \leq \varepsilon$ for any given parameter $\varepsilon > 0$ within running time*

$$m \cdot (\log n)^{O_{\delta_\beta, \delta_\lambda}(1)},$$

where m is the number of edges and n is the number of vertices.

Remark 1.2. *Let μ be the Gibbs distribution of the Ising model with parameters $\beta \in \mathbb{R}_{>1}^E$ and $\lambda \in \mathbb{R}_{>1}^V$, and define $\overline{\mu}$ by letting $\overline{\mu}(S) = \mu(V \setminus S)$ for each $S \subseteq V$. Note that $\overline{\mu}$ is the Gibbs distribution of the Ising model with parameters β and $\lambda^{-1} \in (0, 1)^V$. Therefore, we can sample from $\overline{\mu}$ via the sampler in Theorem 1.1, which implies a sampler for μ.*

Compared to previous works, our algorithm could handle general instances while it only takes a near-linear running time when parameters are bounded away from $\mathbf{1}$. We give a detailed comparison between Theorem 1.1 and previous results in Table 1.

As in [GJ18, FGW22a], we leverage the Edwards-Sokal coupling [ES88] (see Proposition 2.2), which reduces the task of sampling from the distribution of the ferromagnetic Ising model to the random cluster model (see Section 2.3.2 for formal definition).

Theorem 1.3. *Let $\delta_p, \delta_\lambda \in (0, 1)$ be constants and μ be the distribution of a random cluster model specified by graph $G = (V, E)$, parameters $p \in [\delta_p, 1)^E$ and $\lambda \in [0, 1 - \delta_\lambda]^V$. There is an algorithm that samples X satisfying $d_{TV}(X, \mu) \leq \varepsilon$ for any given parameter $\varepsilon > 0$ within running time*

$$m \cdot (\log n)^{O_{\delta_p, \delta_\lambda}(1)},$$

where m is the number of edges and n is the number of vertices.

The proof of Theorem 1.1 assuming Theorem 1.3 is deferred to Appendix A.1.

In the recent progress on high-dimensional expansion and the analysis of Glauber dynamics, a new Markov chain called *field dynamics* has played an important role [CFYZ21b, AJK+21b, CFYZ22, CE22]. The field dynamics was originally used to obtain a boosted optimal spectral gap or modified

results	running time	requirements
[JS93]	$\widetilde{O}_{\beta,\lambda}(m^3)$	$\lambda < 1$
	$\widetilde{O}_{\beta}(n^4 m^3)$	$\lambda = 1$
[LSS19]	$n^{O_\lambda(\log \Delta)}$	$\lambda < 1$
[CLV21b]	$\Delta^{O_{\beta,\lambda}(\Delta)} \cdot \widetilde{O}(m)$	bounded degree
[CGG$^+$21]	$\widetilde{O}_{\beta,\alpha}(n)$	α-expander, $\lambda = 1$
		$\beta = \Omega(\log(\Delta)/\alpha)$
This work (Theorem 1.1)	$\widetilde{O}_{\beta,\lambda}(m)$	$\lambda < 1$

TABLE 1. Comparison of running times and requirements for Ising samplers

log-Sobolev constant from a sub-critical regime. However, it turns out that the field dynamics could also be used to design fast sampler. A recent work [AJK$^+$21b] used field dynamics with interleaved systematic scans to build a fast sampler for the hardcore model. In our result, Theorem 1.3 is another example in which field dynamics is used as an algorithmic tool to design fast sampler. The proof of Theorem 1.3 is outlined in Section 3.

The key ingredients in Theorem 1.3 are to establish spectral independence of the random cluster model for graphs with both bounded and unbounded maximum degree, and to prove the mixing results for the random cluster model in a low-temperature regime. The previous work [CLV21b] established an $O_{\Delta,p,\eta}(1)$ bound for the spectral independence of the subgraph-world model (see Section 2.3.1) by using the analytical property offered by the zero-freeness region, which leads to a $\Delta^{O_{\beta,\lambda}(\Delta)}$ factor in the running time of the sampler for the Ising model or the random cluster model. In this work, we remove the dependency on Δ and prove an $O_\eta(1)$ spectral independence for the subgraph-world model by using a novel coupling based argument (see lemma 4.4). Unlike the previous analysis in [CLV21b], this coupling based argument enables us to lift the spectral independence bound from the subgraph-world model to the random cluster model by using the nature coupling between these models. Therefore, we are able to prove the first $O_\lambda(1)$ bound for the spectral independence of the random cluster model (see lemma 4.1). Finally, in section 5, we use the coupling with stationary argument to show that the Glauber dynamics on the random cluster model mixes rapidly in a low-temperature regime (i.e. when p_{min} is close to 1). With these ingredients, we develop a near-linear time sampler via the field dynamics for the random cluster model on general graphs with both bounded and unbounded maximum degree.

We remark that the recently updated version [FGW22b] of [FGW22a] proved the optimal mixing time of the Glauber dynamics of the random cluster model when the fields are bounded away from 1 and the maximum degree of graphs is bounded by a universal constant. Their proof is also based on the high-dimensional expander.

1.1. **Open problems.** In this paper, we developed a near-linear time sampler for Ising models with parameters $\beta > 1$ and $\lambda < 1$ (or symmetrically, $\lambda > 1$). It still leaves several open problems.

- Develop a good sampler for the ferromagnetic Ising model when field $\lambda = 1$. Our algorithm fails due to the exponential reliance on the gap of λ and 1, which stems from the analysis based on high dimensional expander technique. Therefore, It is still tempting to surpass Jerrum and Sinclair's algorithm [JS93] in this case.
- Better analysis on the Glauber dynamics and the Swendsen-Wang dynamics of the random cluster model. These simple yet powerful dynamics are of great interests in the study of

random cluster model [GJ99, GJ18, GŠV19, BCŠV21, FGW22a]. Though, the current mixing time bounds for these dynamics on general graphs are still far from optimal. We hope our techniques and results could be an inspiration for works in this field.

2. Preliminaries

2.1. **Notation.** Let μ be a distribution over 2^U for some ground set U, and τ, Λ be subsets of U. $\mathcal{P}_{\tau,\Lambda}$ denote the set of configurations $S \subseteq U$ that agree with τ on Λ, i.e.

$$\mathcal{P}_{\tau,\Lambda} = \{S \subseteq U \mid S \cap \Lambda = \tau \cap \Lambda\}.$$

The distribution conditional on event $\mathcal{E} \subseteq 2^U$ is defined by[1]

$$\forall S \subseteq U, \quad \mu(S \mid \mathcal{E}) = \begin{cases} \frac{\mu(S)}{\mu(\mathcal{E})} & \text{if } S \in \mathcal{E}, \\ 0 & \text{otherwise.} \end{cases}$$

For simplicity, we denote by $\mu(\cdot \mid i)$ (resp. $\mu(\cdot \mid \bar{i})$) the distribution $\mu\left(\cdot \mid \mathcal{P}_{\{i\},\{i\}}\right)$ (resp. $\mu\left(\cdot \mid \mathcal{P}_{\varnothing,\{i\}}\right)$) projected on $U \setminus \{i\}$ for some $i \in U$. Furthermore, we denote by $\mu(i)$ (resp. $\mu(\bar{i})$) be the probability $\mathbf{Pr}_{S\sim\mu}[i \in S]$ (resp. $\mathbf{Pr}_{S\sim\mu}[i \notin S]$).

Let $\mathbf{p} \in [0,1]^U$. The distribution $\mu = \bigotimes_{i\in U} \mathrm{Ber}(p_i)$ is defined by

$$\forall S \subseteq E, \quad \mu(S) = \prod_{i\in S} p_i \prod_{i\notin S}(1 - p_i).$$

We will write $\mathbf{1}$ be the constant vector with value 1, and $\mathbf{1}_u = (1_{i=u})_{i\in U}$ for some $u \in U$. Lastly, let $X, Y \subseteq U$ be two subsets of the ground set U, we use $X \oplus Y := (X \setminus Y) \cup (Y \setminus X)$ to denote the symmetric difference between X and Y.

2.2. **Markov chains, entropy and mixing time.**

2.2.1. *Basic definitions.* Let $(X_t)_{t\in\mathbb{N}}$ be a Markov chain over a finite state space Ω with transition matrix $P = (p_{x,y})_{x,y\in\Omega} \in \mathbb{R}_{\geq 0}^{\Omega\times\Omega}$. $(X_t)_{t\in\mathbb{N}}$ is *irreducible*, if for any $x, y \in \Omega$, there exists $t > 0$, such that $P^t(x,y) > 0$. $(X_t)_{t\in\mathbb{N}}$ is *aperiodic*, if for any $x \in \Omega$, $\gcd\left\{t \in \mathbb{N}_{>0} \mid P^t(x,x) > 0\right\} = 1$. A distribution μ over Ω is a *stationary distribution* of $(X_t)_{t\in\mathbb{N}}$, if $\mu = \mu P$. The fundamental theorem of Markov chain says that a Markov chain $(X_t)_{t\in\mathbb{N}}$ has a unique stationary distribution, if the Markov chain is irreducible and aperiodic. A distribution μ over Ω is *reversible* with respect to $(X_t)_{t\in\mathbb{N}}$, if μ satisfies the *detailed balance condition*, i.e. $\mu(x)P(x,y) = \mu(y)P(y,x)$ holds for all $x, y \in \Omega$. It is known that μ is the stationary distribution of $(X_t)_{t\in\mathbb{N}}$ if μ is reversible with respect to the Markov chain.

Let μ, ν be two distributions over the finite state space Ω, the *total variation distance* is defined as

$$d_{\mathrm{TV}}(\mu, \nu) = \max_{S\subseteq\Omega} |\mu(S) - \nu(S)| = \frac{1}{2}\sum_{\sigma\in\Omega} |\nu(\sigma) - \mu(\sigma)|.$$

Suppose μ is the stationary distribution of a Markov chain $(X_t)_{t\in\mathbb{N}}$ with transition matrix P. The *mixing time* initialized from configuration X_0 is defined by

$$T_{\mathrm{mix}}(\varepsilon, X_0) = \min\{t \in \mathbb{N} \mid d_{\mathrm{TV}}(P^t(X_0, \cdot), \mu) < \varepsilon\}.$$

The *mixing time* is defined by $T_{\mathrm{mix}}(\varepsilon) = \max_{X_0\in\Omega} T_{\mathrm{mix}}(\varepsilon, X_0)$.

---

[1]Event $\mathcal{E}$ must satisfy $\mu(\mathcal{E}) = \sum_{S\in\mathcal{E}} \mu(S) > 0$.

2.2.2. *Glauber dynamics.* Let μ be a distribution over 2^U on a finite ground set U. One of the most famous single-site dynamics is the *Glauber dynamics* (a.k.a. Gibbs sampler). In each step, the Glauber dynamics updates configuration $X \in 2^U$ according to the following rules:

- pick $i \in U$ uniformly at random;
- update X according to distribution $\mu\left(\cdot \mid \mathcal{P}_{X,U\setminus\{i\}}\right)$.

It can be verified that μ is reversible with respect to Glauber dynamics.

2.2.3. *Entropy decay and mixing time.* The relation between the functional inequalities and the mixing time of the Markov chain has been studied in literature [DSC96, BT06]. We now introduce the decay of the relative entropy and its implication on the mixing time of the Markov chain.

Let μ, ν be distributions over finite state set Ω and ν is absolutely continuous with respect to μ. The *relative entropy* (known as *Kullback-Leibler divergence*) between ν and μ is defined as

$$D_{KL}(\nu \parallel \mu) = \sum_{\sigma \in \Omega} \nu(\sigma) \log\left(\frac{\nu(\sigma)}{\mu(\sigma)}\right),$$

with convention $0 \cdot \infty = 0$.

Let $(X_t)_{t \in \mathbb{N}}$ be a Markov chain with transition matrix P and stationary distribution μ. The relative entropy *decays with rate* α, if for any distribution ν absolutely continuous with respect to μ,

$$D_{KL}(\nu P \parallel \mu P) \leqslant (1 - \alpha) D_{KL}(\nu \parallel \mu).$$

It is known (see, for example, [BCP+22b, Lemma 2.4]) that the mixing time $T_{\mathrm{mix}}(\varepsilon, X_0)$ satisfies

$$(1) \qquad T_{\mathrm{mix}}(\varepsilon, X_0) \leqslant \alpha^{-1}\left(\log \log \frac{1}{\mu(X_0)} + \log \frac{1}{2\varepsilon^2}\right).$$

2.3. **Subgraph-world model, random cluster model, and relation of models.**

2.3.1. *Generalized subgraph-world model.* Let $G = (V, E)$ be an undirected graph, $\mathbf{p} \in [0,1]^E, \boldsymbol{\eta} \in [0,1]^V$ and $\sigma \in \{0,1\}^V$. The weight of a configuration $S \subseteq E$ in the generalized subgraph-world model is defined by:

$$w_{E,\mathbf{p},\boldsymbol{\eta},\sigma}^{\mathrm{GSW}}(S) := \prod_{e \in S} p_e \prod_{f \in E \setminus S} (1 - p_f) \prod_{\substack{v \in V \\ |S \cap E_v| \equiv \sigma_v (\mathrm{mod}\ 2)}} \eta_v,$$

where E_v denotes the set of edges that is incident to v. For ease of notation, we may use $\mathbf{p}^S, (1-\mathbf{p})^{E\setminus S}$ to denote $\prod_{e \in S} p_e$ and $\prod_{f \in E\setminus S}(1 - p_f)$, respectively. The distribution $\mu_{E,\mathbf{p},\boldsymbol{\eta},\sigma}^{\mathrm{GSW}}$ on 2^E is

$$\forall S \subseteq E, \quad \mu_{E,\mathbf{p},\boldsymbol{\eta},\sigma}^{\mathrm{GSW}}(S) := \frac{w_{E,\mathbf{p},\boldsymbol{\eta},\sigma}^{\mathrm{GSW}}(S)}{Z_{E,\mathbf{p},\boldsymbol{\eta},\sigma}^{\mathrm{GSW}}},$$

where $Z_{E,\mathbf{p},\boldsymbol{\eta},\sigma}^{\mathrm{GSW}} := \sum_{S \subseteq E} w_{E,\mathbf{p},\boldsymbol{\eta},\sigma}^{\mathrm{GSW}}(S)$ is the partition function of this system [2]. We remark that when $\mathbf{p}' \in [0,1]^{E'}$, where $E' \supseteq E$ is a superset of E, the distribution $\mu_{E,\mathbf{p}',\boldsymbol{\eta},\sigma}^{\mathrm{GSW}}$ is defined by

$$\mu_{E,\mathbf{p}',\boldsymbol{\eta},\sigma}^{\mathrm{GSW}} = \mu_{E,\mathbf{p}'|_E,\boldsymbol{\eta},\sigma}^{\mathrm{GSW}},$$

where $\mathbf{p}'|_E$ is the vector obtained by restricting $\mathbf{p}'$ to E. When $\sigma = \mathbf{1}$, our definition matches the definition of the subgraph-world model [JS93, FGW22a]. In this case, we may denote the distribution and partition function by $\mu_{E,\mathbf{p},\boldsymbol{\eta}}^{\mathrm{SW}}$ and $Z_{E,\mathbf{p},\boldsymbol{\eta}}^{\mathrm{SW}}$ instead.

---

[2] $Z_{E,\mathbf{p},\boldsymbol{\eta},\sigma}^{\mathrm{GSW}}$ may equal to zero. In this case, the system is invalid and we will not consider such case.

2.3.2. *Random cluster model.* Let $G = (V, E)$ be an undirected graph, $\mathbf{p} \in [0,1]^E, \boldsymbol{\lambda} \in [0,1]^V$ be parameters. The weight of a configuration $S \subseteq E$ in the random cluster model is defined by:

$$w_{E,\mathbf{p},\boldsymbol{\lambda}}^{\mathrm{RC}}(S) := \prod_{e \in S} p_e \prod_{f \in E \setminus S} (1 - p_e) \prod_{C \in \kappa(V,S)} \left(1 + \prod_{j \in C} \lambda_j \right),$$

where we use $\kappa(V, S)$ to denote the set of connected components of graph (V, S). The distribution $\mu_{E,\mathbf{p},\boldsymbol{\lambda}}^{\mathrm{RC}}$ is defined by

$$\forall S \subseteq E, \quad \mu_{E,\mathbf{p},\boldsymbol{\lambda}}^{\mathrm{RC}}(S) := \frac{w_{E,\mathbf{p},\boldsymbol{\lambda}}^{\mathrm{RC}}(S)}{Z_{E,\mathbf{p},\boldsymbol{\lambda}}^{\mathrm{RC}}},$$

where $Z_{E,\mathbf{p},\boldsymbol{\lambda}}^{\mathrm{RC}} := \sum_{S \subseteq E} w_{E,\mathbf{p},\boldsymbol{\lambda}}^{\mathrm{RC}}(S)$ is the partition function of this system. When $\boldsymbol{\lambda} = \mathbf{1}$, our definition matches the classical definition of the random cluster model with $q = 2$ [FK72].

2.3.3. *Relation between models.* A well-known holographic transformation [JS93, Val08] connects the partition functions of the ferromagnetic Ising model, the subgraph-world model and the random cluster model.

Proposition 2.1 ([FGW22a, Lemma 2.1]). *Let $G = (V, E)$ be a graph, $\boldsymbol{\beta} \in (1, +\infty)^E$ and $\boldsymbol{\lambda} \in [0,1]^V$ be parameters, then*

$$Z_{\boldsymbol{\beta},\boldsymbol{\lambda}}^{\mathrm{Ising}} = \left(\prod_{e \in E} \beta_e \right) Z_{E,\mathbf{p},\boldsymbol{\lambda}}^{\mathrm{RC}} = \left(\prod_{e \in E} \beta_e \right) \left(\prod_{v \in V} (1 + \lambda_v) \right) Z_{E,\mathbf{p}/2,\boldsymbol{\eta}}^{\mathrm{SW}},$$

where $\mathbf{p} = 1 - \boldsymbol{\beta}^{-1} = \left(1 - \beta_e^{-1} \right)_{e \in E}$ and $\boldsymbol{\eta} = \left(\frac{1-\lambda_v}{1+\lambda_v} \right)_{v \in V}$.

The standard Edwards-Sokal coupling connects the Gibbs distribution of the ferromagnetic Ising model and the distribution of the random cluster model.

Proposition 2.2 ([FGW22a, Proposition 2.3]). *Let $G = (V, E)$ be a graph, $\boldsymbol{\beta} \in (1, +\infty)^E$ and $\boldsymbol{\lambda} \in [0,1]^V$ be parameters. Furthermore, let $\mathbf{p} = 1 - \boldsymbol{\beta}^{-1} = (1 - \beta_e)_{e \in E}$. Construct Y according to the following rules.*

- *initialize $Y = \varnothing$ and sample $X \sim \mu_{E,\mathbf{p},\boldsymbol{\lambda}}^{\mathrm{RC}}$;*
- *for each $C \in \kappa(V, X)$, add C to Y with probability $\frac{\prod_{i \in C} \lambda_i}{1 + \prod_{i \in C} \lambda_i}$.*

Then, it holds that $Y \sim \mu_{\boldsymbol{\beta},\boldsymbol{\lambda}}^{\mathrm{Ising}}$.

3. Proof outline

In this section, we outline the proof of Theorem 1.3. We first summarize previous results of the field dynamics in Section 3.1, and then introduce the field dynamics simulator, prove its correctness as an approximate sampler, and analyze its running time in Section 3.2.

3.1. **Field dynamics.** The *field dynamics* is an adaptive select-update dynamics, first introduced in [CFYZ21b]. Intuitively, the field dynamics serves as a reduction from a critical instance to a sub-critical instance, i.e., instance that are much easier to handle.

Let μ be a distribution over 2^U on ground set U and $\boldsymbol{\lambda} \in \mathbb{R}_{>0}^U$. Denote by $\boldsymbol{\lambda} * \mu$ the distribution over 2^U satisfying

$$\forall S \subseteq U, \quad (\boldsymbol{\lambda} * \mu)(S) \propto \boldsymbol{\lambda}^S \mu(S),$$

where $\boldsymbol{\lambda}^S = \prod_{i \in S} \lambda_i$. In particular, if $\boldsymbol{\lambda}$ is a constant vector with $\lambda_u = \lambda$, we may write $\lambda * \mu$ instead.

The field dynamics P_θ^{FD} with parameter $\theta \in (0, 1)$ in each round updates a configuration $X \in 2^U$ according to the following rules:

- sample $S' \sim \bigotimes_{u \in U} \mathrm{Ber}(\theta)$ and let $S = S' \cup X$;
- update X according to distribution $(\theta^{-1} * \mu) \left(\cdot \mid \mathcal{P}_{X, U \setminus S} \right)$,

where we recall the event $\mathcal{P}_{X,Y} = \{T \subseteq U \mid T \cap Y = X \cap Y\}$.

It has been showed in [CFYZ21b] that the field dynamics P_θ^{FD} has the stationary distribution μ. We note that the only non-trivial step in the field dynamics is to sample from a new distribution $(\theta^{-1} * \mu) \left(\cdot \mid \mathcal{P}_{X, U \setminus S} \right)$. A key intuition of the field dynamics is that $(\theta^{-1} * \mu) \left(\cdot \mid \mathcal{P}_{X, U \setminus S} \right)$ might be relatively easy to sample from when we choose a good parameter θ. Hence, when the field dynamics itself is rapid mixing, it actually reduces the task of sampling from μ to an easier task of sampling from $(\theta^{-1} * \mu) \left(\cdot \mid \mathcal{P}_{X, U \setminus S} \right)$. Our algorithm for the random cluster model is based on this idea where we use a Glauber dynamics to generate approximate samples from $(\theta^{-1} * \mu) \left(\cdot \mid \mathcal{P}_{X, U \setminus S} \right)$ so as to sample from the original distribution (see algorithm 1 and algorithm 2 for the details). Moreover, We also establish the rapid mixing of field dynamics via the spectral independence.

In recent years, there is a long line of works [ALO20, CLV21a, CFYZ21b, AJK+21a, BCC+21, AJK+21b, CE22, CFYZ22] establishing the relation between the mixing time of select-update dynamics and the spectral independence. We first introduce the notion of spectral independence.

Definition 3.1 (influence matrix, [ALO20])**.** Let μ be a distribution over 2^U on ground set U of size n. The influence matrix Ψ_μ^{inf} is an n by n matrix defined as

$$\forall i, j \in U, \quad \Psi_\mu^{\mathrm{inf}}(i,j) = \begin{cases} \mu(i \mid j) - \mu(i \mid \bar{j}) & \text{if } i \neq j \text{ and } 0 < \mathbf{Pr}_{S \sim \mu}[j \in S] < 1, \\ 0 & \text{otherwise.} \end{cases}$$

Definition 3.2 (spectral independence in infinite norm)**.** Let μ be a distribution over 2^U on ground set U. The distribution μ is C-spectrally independent, if

$$\left\| \Psi_\mu^{\mathrm{inf}} \right\|_\infty \leqslant C.$$

Furthermore, μ is C-spectrally independent under all pinnings, if for any $\tau, \Lambda \subseteq U$ with $\mu(\mathcal{P}_{\tau,\Lambda}) > 0$, $\mu(\cdot \mid \mathcal{P}_{\tau,\Lambda})$ projected on $U \setminus \Lambda$ is C-spectrally independent.

We remark the notion of spectral independence in Definition 3.2 is stronger than that in [ALO20], where the distribution μ is C-spectrally independent, if $\lambda_{\max}(\Psi_\mu^{\mathrm{inf}}) \leqslant C$.

In recent progress of high-dimensional expansion and the analysis of Glauber dynamics on antiferromagnetic two-state spin systems, the following entropy decay result for field dynamics is established [AJK+21b, CFYZ21a, CE22, CFYZ22].

Lemma 3.3 (entropy decay of field dynamics)**.** *Let μ be a distribution over 2^U on ground set U. If $\lambda * \mu$ is C-spectrally independent under all pinnings for all $\lambda \in \mathbb{R}_{>0}^U$, then for any $\theta \in (0,1)$ and distribution ν absolutely continuous with respect to μ, let $\kappa = (\theta/e)^{C+3}$, it holds that*

$$D_{\mathrm{KL}}(\nu P_\theta^{\mathrm{FD}} \parallel \mu P_\theta^{\mathrm{FD}}) \leqslant (1-\kappa) D_{\mathrm{KL}}(\nu \parallel \mu).$$

For completeness, we include the proof of Lemma 3.3 in Appendix A.3. By (1), this entropy decay result implies a tight bound on the mixing time of field dynamics.

3.2. **Field dynamics simulator.** We are now ready to introduce the field dynamics simulator for random cluster model. Parameters θ, T^{FD}, and T^{GD} are to be determined.

Algorithm 1: field dynamics simulator for random cluster model

---

input : Graph $G = (V, E)$, parameters $\mathbf{p} \in (0,1)^E$, $\boldsymbol{\lambda} \in [0,1)^V$ and $\varepsilon \in (0,1)$

output: a random configuration $X \subseteq E$ satisfying $d_{TV}(X, \mu^{RC}_{E,\mathbf{p},\boldsymbol{\lambda}}) < \varepsilon$

1 Initialize $X = E$;

2 **for** $t \leftarrow 1$ **to** T^{FD} **do**

3   draw $S' \sim \bigotimes_{e \in E} \mathrm{Ber}(\theta)$ and let $S = S' \cup X$;

4   update $X \leftarrow \mathrm{Resample}(G' = (V, S), \mathbf{p}^\star, \boldsymbol{\lambda}, (2T^{FD})^{-1} \varepsilon)$, where $p_e^\star = \frac{p_e}{\theta(1-p_e)+p_e}, \forall e \in S$.

5 **return** X

Algorithm 2: Resample(G, $\mathbf{p}$, $\boldsymbol{\lambda}$, ε)

---

input : Graph $G = (V, E)$, parameters $\mathbf{p} \in (0,1)^E$, $\boldsymbol{\lambda} \in [0,1)^V$, and $\varepsilon \in (0,1)$

output: a random configuration $X \subseteq E$ satisfying $d_{TV}(X, \mu^{RC}_{E,\mathbf{p},\boldsymbol{\lambda}}) < \varepsilon$

1 Initialize $X = E$;

2 **for** $t \leftarrow 1$ **to** T^{GD} **do**

3   draw $e \in E$ uniformly at random;

4   update X according to $\mu^{RC}_{E,\mathbf{p},\boldsymbol{\lambda}} \left(\cdot \mid \mathcal{P}_{X, E \setminus \{e\}} \right)$.

5 **return** X

Remark 3.4. In Algorithm 1, since $S = S' \cup X$, it holds that $X \cap (E \setminus S) = \varnothing$. This means all the elements in $E \setminus S$ are in the "out" state in X. So, it is straight forward to verify that

$$\left(\theta^{-1} * \mu^{RC}_{E,\mathbf{p},\boldsymbol{\lambda}} \right) \left(\cdot \mid \mathcal{P}_{X, E \setminus S} \right) = \mu^{RC}_{S, \mathbf{p}^\star, \boldsymbol{\lambda}}(\cdot),$$

where $p_e^\star = \frac{p_e}{p_e + \theta(1-p_e)}$ for all $e \in S$. Hence, when $|V| > N_0$, Algorithm 1 is exactly the field dynamics assuming Algorithm 2 being a perfect sampler. Though in our implementation, Algorithm 2 returns approximate samples and causes biases.

Let $\mu^{RC}_{E,\mathbf{p},\boldsymbol{\lambda}}$ be the distribution of the random cluster model specified by graph $G = (V, E)$, parameters $\mathbf{p} \in (0,1)^E$ and $\boldsymbol{\lambda} \in [0,1)^V$. Furthermore, let

$$p_{\min} = \min_{e \in E} p_e \text{ and } \lambda_{\max} = \max_{v \in V} \lambda_v.$$

We first state the mixing time results for both field dynamics and Glauber dynamics, and then prove Theorem 1.3 with these results.

Lemma 3.5. *The mixing time of the field dynamics initialized from E satisfies*

$$\forall \varepsilon \in (0,1), \quad T(\varepsilon, E) \leqslant \left(\frac{e}{\theta} \right)^{5(1-\lambda_{\max})^{-2}} \left(2 \log n + \log \log \frac{2}{p_{\min}} + \log \frac{1}{2\varepsilon^2} \right).$$

This mixing time result is a corollary of Lemma 3.3 and the spectral independence of $\mu^{RC}_{E,\mathbf{p},\boldsymbol{\lambda}}$.

Lemma 3.6. $\mu^{RC}_{E,\mathbf{p},\boldsymbol{\lambda}}$ *is $2(1 - \lambda_{\max})^{-2}$-spectrally independent under all pinnings.*

The proof of Lemma 3.5 assuming Lemma 3.6 is deferred to Appendix A.2.

Remark 3.7. Establishing spectral independence is a challenging task. A series of works[CLV20, BCC+21, Liu21, CLV21b, ALO22] establish spectral independence via different approaches, including correlation decay, path coupling, stability of polynomial, and the trickle-down phenomenon. An $O_{\Delta,\mathbf{p},\boldsymbol{\eta}}(1)$ spectral independence result for the subgraph-world model was established

in [CLV22]. This result has the dependency on Δ and does not imply a spectral independence bound for the random cluster model. Our method is quite different from the previous approaches for bounding the spectral independence. In contrast to previous works, we establish a better spectral independence result for the subgraph-world model that is independent of Δ by a sophisticated coupling procedure. This coupling procedure also enables us to lift the spectral independence result from the subgraph-world model to the random cluster model. As far as we know, lemma 3.6 gives the first spectral independence bound for the random cluster model. The details will be described in Section 4.

Lemma 3.8. *If it holds that*

$$(1 - p_{\min}) \log n \leqslant \min \left\{ 10^{-7}, \frac{1 - \lambda_{\max}}{27} \right\},$$

then the mixing time of the Glauber dynamics satisfies

$$\forall \varepsilon \in (0, 1), \quad T_{\mathrm{mix}}(\varepsilon) \leqslant 25 m \log m \log \frac{1}{\varepsilon} + 1.$$

The proof of Lemma 3.8 involves a coupling with stationary argument, which will be presented in Section 5. We are now ready to prove Theorem 1.3.

Proof of Theorem 1.3. We choose parameters in Algorithm 1 and Algorithm 2 as follows:

parameter	value
θ	$\dfrac{\min\left\{ 10^{-7}, \frac{1-\lambda_{\max}}{27} \right\} p_{\min}}{\log n}$
T^{FD}	$\left\lceil \left(\frac{e}{\theta}\right)^{5(1-\lambda_{\max})^{-2}} \left(2 \log n + \log \log \frac{2}{p_{\min}} + \log \frac{2}{\varepsilon^2} \right) \right\rceil$
T^{GD}	$\left\lceil 25 m \log m \log \frac{2 T^{\mathrm{FD}}}{\varepsilon} \right\rceil + 1$

To prove Theorem 1.3, it suffices to show that:

 (1) The sample X returned from Algorithm 1 satisfies $d_{\mathrm{TV}}(X, \mu_{E,\mathbf{p},\boldsymbol{\lambda}}^{\mathrm{RC}}) < \varepsilon$.

 (2) The overall running time can be achieved in $m \left(\varepsilon^{-1} \cdot \delta_p^{-1} \cdot \log n \right)^{O(\delta_\lambda^{-2})}$.

First, we prove requirement (1). For convenience, let $\widehat{P_\theta^{\mathrm{FD}}}$ denote the transition matrix of the simulation algorithm (Algorithm 1) and $\mu = \mu_{E,\mathbf{p},\boldsymbol{\lambda}}^{\mathrm{RC}}$. Let $(X_t)_{0 \leqslant t \leqslant T^{\mathrm{FD}}}$ (respectively, $(Y_t)_{0 \leqslant t \leqslant T^{\mathrm{FD}}}$) be the Markov chain starting from configuration $X_0 = E$ (respectively, $Y_0 = E$) with transition matrix $\widehat{P_\theta^{\mathrm{FD}}}$ (respectively, P_θ^{FD}). It holds that

$$d_{\mathrm{TV}}\left(X_{T^{\mathrm{FD}}}, \mu\right) \leqslant d_{\mathrm{TV}}\left(X_{T^{\mathrm{FD}}}, Y_{T^{\mathrm{FD}}}\right) + d_{\mathrm{TV}}\left(Y_{T^{\mathrm{FD}}}, \mu\right) \leqslant \mathbf{Pr}\left[X_{T^{\mathrm{FD}}} \neq Y_{T^{\mathrm{FD}}}\right] + d_{\mathrm{TV}}\left(Y_{T^{\mathrm{FD}}}, \mu\right).$$

Note that $X_{T^{\mathrm{FD}}} \neq Y_{T^{\mathrm{FD}}}$ is equivalent to that there exists $1 \leqslant i \leqslant T^{\mathrm{FD}}$ such that $X_i \neq Y_i$ but $X_j = Y_j$ for all $j < i$. Hence,

$$\mathbf{Pr}\left[X_{T^{\mathrm{FD}}} \neq Y_{T^{\mathrm{FD}}}\right] = \sum_{i=1}^{T^{\mathrm{FD}}} \mathbf{Pr}\left[X_i \neq Y_i \text{ and } \forall j < i, X_j = Y_j\right] \leqslant \sum_{i=1}^{T^{\mathrm{FD}}} \mathbf{Pr}\left[X_i \neq Y_i \mid X_{i-1} = Y_{i-1}\right].$$

For any $1 \leqslant i \leqslant T^{\mathrm{FD}}$, consider the following coupling of X_i and Y_i condition on $X_{i-1} = Y_{i-1}$:

 (1) Generate set $S' \sim \bigotimes_{u \in U} \mathrm{Ber}(\theta)$ and $S = S' \cup X_{i-1}$, i.e., the first step of the field dynamics;

 (2) Generate X_i and Y_i according to the optimal coupling of distribution $\mu_{S,\mathbf{p}^\star,\boldsymbol{\lambda}}^{\mathrm{RC}}$ and $\widehat{\mu_{S,\mathbf{p}^\star,\boldsymbol{\lambda}'}^{\mathrm{RC}}}$ where $\widehat{\mu_{S,\mathbf{p}^\star,\boldsymbol{\lambda}}^{\mathrm{RC}}}$ is the distribution generated by $\mathrm{Resample}((V, S), \mathbf{p}^\star, \boldsymbol{\lambda}, (2T^{\mathrm{FD}})^{-1}\varepsilon)$.

9

Hence, we have

$$\mathbf{Pr}\left[X_i \neq Y_i \mid X_{i-1} = Y_{i-1}\right] \leq \max_{S \subseteq E} d_{\mathrm{TV}}\left(\widetilde{\mu_{S,\mathbf{p}^\star,\lambda}^{\mathrm{RC}}}, \mu_{S,\mathbf{p}^\star,\lambda}^{\mathrm{RC}}\right) =: \varepsilon'.$$

Note that $(1 - p_{\min}^\star) \log n \leq \frac{\theta \log n}{p_{\min}} \leq \min\left\{10^{-7}, \frac{1-\lambda_{\max}}{27}\right\}$. By Lemma 3.8 and our choice of T^{GD}, $\varepsilon' \leq \frac{\varepsilon}{2T^{\mathrm{FD}}}$. Therefore, by Lemma 3.5 and our choice of T^{FD}, it holds that

$$d_{\mathrm{TV}}\left(X_{T^{\mathrm{FD}}}, \mu\right) \leq T^{\mathrm{FD}} \varepsilon' + \varepsilon/2 \leq \varepsilon.$$

This proves the requirement (1).

Lastly, we verify requirement (2). In each iteration of the field dynamics in Algorithm 1, we need $O(m)$ time to generate the set S. This consumes mT^{FD} time. Besides, the algorithm needs T^{FD} iterations of T^{GD}-steps Glauber dynamics starting from configuration $X = S$ on distribution $\mu_{S,\mathbf{p}^\star,\lambda}^{\mathrm{RC}}$ for some $S \subseteq E$. We claim that each step of the Glauber dynamics could be implemented in $\mathrm{polylog}(n)$ time, then the total running time is

$$(2) \qquad\qquad mT^{\mathrm{FD}} + T^{\mathrm{FD}} \cdot T^{\mathrm{GD}} \cdot \mathrm{polylog}(n).$$

By our choice of T^{FD} and T^{GD}, it holds that

$$T^{\mathrm{FD}} = \left((1 - \lambda_{\max})^{-1} \cdot p_{\min}^{-1} \cdot \log n\right)^{O((1-\lambda_{\max})^{-2})} \text{ and } T^{\mathrm{GD}} \leq 25m \log m \log \frac{2T^{\mathrm{FD}}}{\varepsilon} + 2.$$

Together with $\lambda_{\max} \leq 1 - \delta_\lambda$ and $p_{\min} \geq \delta_\beta$, the total running time (2) could be bounded by

$$m \left(\delta_\lambda^{-1} \cdot \delta_p^{-1} \cdot \log n\right)^{O(\delta_\lambda^{-2})}.$$

We only left to show that each step of the Glauber dynamics could be implemented in $\mathrm{polylog}(n)$ time. Suppose the current configuration is X, the Glauber dynamics will first use $O(\log n)$ time to draw a random edge $e = (u, v) \in E$. Let $C_u = C_u(X)$ and $C_v = C_v(X)$ be the connected components in graph $(V, X \setminus \{e\})$, containing u and v, respectively. Then the probability $p_{X,e}$ that X will be updated by $X \cup \{e\}$ is

$$(3) \qquad\qquad p_{X,e} = \begin{cases} p & \text{if } C_u = C_v, \\ \frac{1+\lambda^{C_u \cup C_v}}{1+\lambda^{C_u \cup C_v}+(1-p_e^\star)(\lambda^{C_u}+\lambda^{C_v})} & \text{otherwise,} \end{cases}$$

where $\lambda^R = \prod_{i \in R} \lambda_i$ for $R \subseteq V$. In order to calculate $p_{X,e}$ fast, we need a data structure that supports the following operations:

- update $X \leftarrow X \cup \{e\}$ for any given $e \in E$;
- update $X \leftarrow X \setminus \{e\}$ for any given $e \in E$;
- query if $C_u(X) = C_v(X)$ for any given $u, v \in V$;
- query $\lambda^{C_u(X)}$ for any given $u \in V$.

These updates and queries can all be handled in $O(\log^2 n)$ amortized time by the data structure in [WN13, Section 3]. This concludes the proof of requirement (2) and Theorem 1.3. □

4. Spectral independence of random cluster model

In this section, we are going to prove Lemma 3.6. We prove Lemma 3.6 via the following lemma.

Lemma 4.1. *Let μ be the distribution of the random cluster model specified by graph $G = (V, E)$, parameters $\mathbf{p} \in [0, 1]^E$ and $\lambda \in [0, 1)^V$. Then, μ is $2(1 - \lambda_{\max})^{-2}$-spectrally independent.*

Proof of Lemma 3.6. For any $\tau, \Lambda \subseteq E$, define $\tilde{\mathbf{p}} \in [0,1]^E$ by

$$\forall e \in E, \quad \tilde{p}_e = \begin{cases} 0 & \text{if } e \in \Lambda \setminus \tau, \\ 1 & \text{if } e \in \Lambda \cap \tau, \\ p_e & \text{if } e \in E \setminus \Lambda. \end{cases}$$

Note that $\mu^{RC}_{E,\tilde{\mathbf{p}},\lambda}$ is exactly $\mu(\cdot \mid \mathcal{P}_{\tau,\Lambda})$. This concludes the proof of Lemma 3.6. $\qquad \square$

In order to prove Lemma 4.1, we introduce a simple coupling criteria for spectral independence.

Definition 4.2 (coupling independence)**.** A distribution μ over 2^E on ground set E is C-coupling independent, if for all $i \in E$, there exists a coupling (X, Y) of distribution $\mu(\cdot \mid i)$ and $\mu(\cdot \mid \bar{i})$, that

$$\mathbb{E}[|X \oplus Y|] \leq C.$$

Furthermore, a distribution is C-coupling independent under all pinnings, if for any $\Lambda, \tau \subseteq U$ with $\mu(\mathcal{P}_{\tau,\Lambda}) > 0$, $\mu(\cdot \mid \mathcal{P}_{\tau,\Lambda})$ projected on $U \setminus \Lambda$ is C-coupling independent.

Proposition 4.3. *If a distribution μ over 2^E is C-coupling independent, then μ is C-spectrally independent.*

Proof. Fix $i \in E$. Let (X, Y) be a coupling of $\mu(\cdot \mid i)$ and $\mu(\cdot \mid \bar{i})$ such that $\mathbb{E}[|X \oplus Y|] \leq C$, then

$$\sum_{j \in E \setminus \{i\}} \left| \mu(j \mid i) - \mu(j \mid \bar{i}) \right| \leq \sum_{j \in E} \mathbb{E}\left[\mathbf{1}[X_j \neq Y_j]\right] = \mathbb{E}[|X \oplus Y|] \leq C,$$

where the first inequality holds by standard coupling lemma. Therefore,

$$\left\| \Psi^{\inf}_\mu \right\|_\infty = \max_{i \in E} \sum_{j \in E \setminus \{i\}} \left| \mu(j \mid i) - \mu(j \mid \bar{i}) \right| \leq C. \qquad \square$$

Now, to prove Lemma 4.1, we first claim the coupling independence for the distribution of subgraph-world model.

Lemma 4.4. *Let ν be the distribution of subgraph-world model specified by graph $G = (V, E)$, and vectors $\mathbf{p} \in [0, \frac{1}{2}]^E$, $\boldsymbol{\eta} \in (0,1]^V$. It holds that ν is $\frac{1}{2\eta^2_{\min}}$-coupling independent.*

Then, we show that once we have a coupling of distributions of generalized subgraph-world model, we could "lift" it to the random cluster model.

Lemma 4.5. *Let $G = (V, E)$ be an undirected graph, $\mathbf{p} \in [0,1]^E$ and $\boldsymbol{\lambda} \in [0,1)^V$ be parameters. Let μ be the distribution of a random cluster model specified by graph G, parameters $\mathbf{p}$ and $\boldsymbol{\lambda}$. Let ν be the distribution of a subgraph-world model specified by graph G, parameters $\frac{\mathbf{p}}{2} = \left(\frac{p_e}{2}\right)_{e \in E}$ and $\boldsymbol{\eta} = \left(\frac{1-\lambda_v}{1+\lambda_v}\right)_{v \in V}$.*

If ν is C-coupling independent, then μ is also C-coupling independent.

Lemma 4.1 is proved by combining Proposition 4.3, Lemma 4.4, and Lemma 4.5.
The proof of Lemma 4.4 and Lemma 4.5 are given in Section 4.1 and Section 4.2 respectively.

4.1. Coupling independence of generalized subgraph-world model.

In this section, we prove Lemma 4.4. For convenience, for $\sigma \in \mathbb{R}^V$, we use $\sigma^{u \leftarrow c}$ to denote a vector σ with σ_u being changed to value c. Meanwhile, for $\sigma, \tau \in \{0,1\}^V$, we use $\sigma \oplus \tau$ to denote the bitwise exclusive or of σ and τ. We now prove a generalized version of Lemma 4.4.

Lemma 4.6. *Let $G = (V, E)$ be an undirected graph, $\sigma \in \{0,1\}^V$ be parity constraints on vertices, $\mathbf{p} \in \left[0, \frac{1}{2}\right]^E$ and $\boldsymbol{\eta} \in (0,1]^V$ be parameters. For any $u \in V$, there is a coupling (X, Y) between $\mu^{GSW}_{E,\mathbf{p},\boldsymbol{\eta},\sigma}$ and $\mu^{GSW}_{E,\mathbf{p},\boldsymbol{\eta},\sigma \oplus \mathbf{1}_u}$ such that $\mathbb{E}[|X \oplus Y|] \leq \frac{1}{4\eta^2_{\min}}$.*

FIGURE 1. This is an illustration of Algorithm 3. Here, vertices v_i is colored red if $u = v_i$, blue if $\sigma_{v_i} = 1$ and green if $\sigma_{v_i} = 0$. Moreover, edge that has not been revealed is represented as dotted line, edge in both X and Y is colored with black, edge in exactly one of X and Y is colored with red, and edge that has been revealed but not in either X or Y is removed.

We now prove Lemma 4.4. The proof follows from a standard coupling argument.

Proof of Lemma 4.4. Fix $e \in (u, v) \in E$ and let $\nu = \mu_{E, \mathbf{p}, \boldsymbol{\eta}, \sigma}^{\text{GSW}}$. By definition,

$$\nu(\cdot \mid \overline{e}) = \mu_{E \setminus \{e\}, \mathbf{p}, \boldsymbol{\eta}, \mathbf{1}}^{\text{GSW}} \quad \text{and} \quad \nu(\cdot \mid e) = \mu_{E \setminus \{e\}, \mathbf{p}, \boldsymbol{\eta}, \mathbf{1} \oplus \mathbf{1}_u \oplus \mathbf{1}_v}^{\text{GSW}},$$

Consider an intermediate distribution $\widetilde{\nu} := \mu_{E \setminus \{e\}, \mathbf{p}, \boldsymbol{\eta}, \mathbf{1} \oplus \mathbf{1}_u}^{\text{GSW}}$. By Lemma 4.6, there are couplings $\mathcal{C}_1$ of $\nu(\cdot \mid \overline{e})$ and $\widetilde{\nu}$ as well as $\mathcal{C}_2$ of $\widetilde{\nu}$ and $\nu(\cdot \mid e)$ satisfying

$$\forall i \in \{1, 2\}, \quad \mathbb{E}_{(X, Y) \sim \mathcal{C}_i} [|X \oplus Y|] \leqslant \frac{1}{4\eta_{\min}^2}.$$

Using $\mathcal{C}_1$ and $\mathcal{C}_2$, we could construct a coupling (X, Y) of $\nu(\cdot \mid \overline{e})$ and $\nu(\cdot \mid e)$ by: (1) sampling $X \sim \nu(\cdot \mid \overline{e})$; (2) sampling Z proportional to $\mathcal{C}_1(X, \cdot)$; (3) sampling Y proportional to $\mathcal{C}_2(Z, \cdot)$. It could be verified from the definition of $\mathcal{C}_1$ and $\mathcal{C}_2$ that X, Y have correct marginals. Again, by Lemma 4.6, it holds that

$$\mathbb{E}[|X \oplus Y|] \leqslant \mathbb{E}[|X \oplus Z| + |Z \oplus Y|] \leqslant \mathbb{E}[|X \oplus Z|] + \mathbb{E}[|Z \oplus Y|] \leqslant \frac{1}{2\eta_{\min}^2},$$

where the last inequality follows from the fact that $(X, Z) \sim \mathcal{C}_1$ and $(Z, Y) \sim \mathcal{C}_2$. $\square$

The rest part of this section is dedicated to the proof of Lemma 4.6. We construct the coupling (X, Y) using the procedure $\text{Couple}(G, \mathbf{p}, \boldsymbol{\eta}, \sigma, u, U)$ in Algorithm 3, where U denotes the set of visited vertices, and is initialized to $\varnothing$. Figure 1 is an illustration of Algorithm 3.

In order to prove Lemma 4.6, it suffices to prove the following properties.

(1) Each recursive call in Algorithm 3 is valid.

12

(2) $\mathbb{E}\left[|X \oplus Y|\right] \leqslant \frac{1}{4\eta_{\min}^2}$;

(3) $(X, Y) = \text{Couple}(u, E, \boldsymbol{\eta}, \sigma, \varnothing)$ is indeed a coupling of $\mu_{E,\mathbf{p},\boldsymbol{\eta},\sigma}^{\text{GSW}}$ and $\mu_{E,\mathbf{p},\boldsymbol{\eta},\sigma\oplus\mathbf{1}_u}^{\text{GSW}}$, that is

$$X \sim \mu_{E,\mathbf{p},\boldsymbol{\eta},\sigma}^{\text{GSW}} \text{ and } Y \sim \mu_{E,\mathbf{p},\boldsymbol{\eta},\sigma\oplus\mathbf{1}_u}^{\text{GSW}}.$$

Algorithm 3: $\text{Couple}(G, \mathbf{p}, \boldsymbol{\eta}, \sigma, u, U)$

---

input : graph $G = (V, E)$, parameters $\mathbf{p} \in [0,1]^E$, $\boldsymbol{\eta} \in [0,1]^V$, $\sigma \in \{0,1\}^V$, vertex $u \in V$ and set of visited vertices U satisfying

    (1) $\eta_u = 0$ if and only if $u \in U$;
    (2) $Z_{E,\mathbf{p},\boldsymbol{\eta},\sigma}^{\text{GSW}}, Z_{E,\mathbf{p},\boldsymbol{\eta},\sigma\oplus\mathbf{1}_u}^{\text{GSW}} > 0$.

output: a pair of random configuration $(X, Y) \in 2^E \times 2^E$.

1 if $u \notin U$ **then**
2   update $U \leftarrow U \cup \{u\}$;
3   let $\mathcal{A} = \{S \subseteq E \mid |S \cap E_u| \equiv 0 \mod 2\}$;
4   let $R = \left(\sum_{S\in\mathcal{A}} \mu_{E,\mathbf{p},\boldsymbol{\eta}^{u\leftarrow 1},\sigma}^{\text{GSW}}(S)\right)\left(\sum_{S\notin\mathcal{A}} \mu_{E,\mathbf{p},\boldsymbol{\eta}^{u\leftarrow 1},\sigma}^{\text{GSW}}(S)\right)^{-1}$, $q_0 = \frac{\eta_u R}{\eta_u R + 1}$ and $q_1 = \frac{R}{R+\eta_u}$;
5   draw $r \sim \text{Uniform}(0,1)$;
6   update $\eta_u \leftarrow 0$;
7   **if** $r \geqslant q_1$ **then**
8   sample $C \sim \mu_{E,\mathbf{p},\boldsymbol{\eta},\sigma^{u\leftarrow 0}}^{\text{GSW}}$;
9   **return** $(X, Y) = (C, C)$
10   **if** $r \leqslant q_0$ **then**
11   sample $C \sim \mu_{E,\mathbf{p},\boldsymbol{\eta},\sigma^{u\leftarrow 1}}^{\text{GSW}}$;
12   **return** $(X, Y) = (C, C)$

13 pick an arbitrary $e = (u, v) \in E_u$;
14 let ν, π be the distributions of $\mu_{E,\mathbf{p},\boldsymbol{\eta},\sigma}^{\text{GSW}}$ and $\mu_{E,\mathbf{p},\boldsymbol{\eta},\sigma\oplus\mathbf{1}_u}^{\text{GSW}}$ projected on e respectively;
15 sample (X_1, Y_1) from an optimal coupling of ν and π;
16 if $X_1 = \{e\}$ **then**
17   update $\sigma \leftarrow \sigma \oplus \mathbf{1}_u \oplus \mathbf{1}_v$;
18 if $X_1 = Y_1$ **then**
19 $(X_2, Y_2) \leftarrow \text{Couple}((V, E \setminus \{e\}), \mathbf{p}, \boldsymbol{\eta}, \sigma, u, U)$;
20   **return** $(X, Y) = (X_1 \cup X_2, Y_1 \cup Y_2)$
21 else
22 $(X_2, Y_2) \leftarrow \text{Couple}((V, E \setminus \{e\}), \mathbf{p}, \boldsymbol{\eta}, \sigma, v, U)$;
23   **return** $(X, Y) = (X_1 \cup X_2, Y_1 \cup Y_2)$

The first property can be verified easily. We now prove Property (2) with following observations.

Proposition 4.7. *Let* U *be the set of visited vertices upon termination. For any* $k \geqslant 1$,

$$\mathbf{Pr}\left[|U| \geqslant k\right] \leqslant \left(\frac{1-\eta_{\min}}{1+\eta_{\min}}\right)^{k-1}.$$

13

Proof. Note that $|U| \geqslant k$ implies that first $k-1$ random numbers $r_1, r_2, \ldots, r_{k-1}$ drawn in Line 5 all lie in their corresponding segments. Therefore,

$$\mathbf{Pr}\left[|U| \geqslant k\right] \leqslant \left(\max_{R>0, u \in V}\left\{\frac{R}{R+\eta_u} - \frac{\eta_u R}{\eta_u R + 1}\right\}\right)^{k-1} \leqslant \left(\frac{1-\eta_{\min}}{1+\eta_{\min}}\right)^{k-1}. \qquad \square$$

Proposition 4.8. *Let U be the set of visited vertices upon termination and (X, Y) be the returned pair of configurations in Algorithm 3. For each $e = (u, v) \in X \oplus Y$, both $u, v \in U$.*

Proof. This directly follows from the coupling procedure. $\qquad \square$

Now, we are ready to prove Property (2).

Proof of Property (2). Let U be the set of visited vertices upon termination and (X, Y) be the returned configuration. By Proposition 4.7 and Proposition 4.8,

$$\mathbb{E}\left[|X \oplus Y|\right] \leqslant \sum_{k=1}^{|V|}\binom{k}{2}\mathbf{Pr}\left[|U|=k\right] \leqslant \sum_{k=1}^{+\infty} k\mathbf{Pr}\left[|U| \geqslant k+1\right] \leqslant \frac{1}{4\eta_{\min}^2}. \qquad \square$$

Now, we only left to prove Property (3). To begin with, we need the following propositions.

Proposition 4.9. $\mu_{E,\mathbf{p},\boldsymbol{\eta},\sigma^{u \leftarrow 0}}^{\text{GSW}}(\mathcal{A}) = q_0$ *and* $\mu_{E,\mathbf{p},\boldsymbol{\eta},\sigma^{u \leftarrow 1}}^{\text{GSW}}(\mathcal{A}) = q_1$.

Proof. Without loss of generality, we only prove the first part. It holds that

$$\mu_{E,\mathbf{p},\boldsymbol{\eta},\sigma^{u \leftarrow 0}}^{\text{GSW}}(\mathcal{A}) = \left(\sum_{S \in \mathcal{A}} w_{E,\mathbf{p},\boldsymbol{\eta},\sigma^{u \leftarrow 0}}^{\text{GSW}}(S)\right)\left(\sum_{S \in \mathcal{A}} w_{E,\mathbf{p},\boldsymbol{\eta},\sigma^{u \leftarrow 0}}^{\text{GSW}}(S) + \sum_{S \notin \mathcal{A}} w_{E,\mathbf{p},\boldsymbol{\eta},\sigma^{u \leftarrow 0}}^{\text{GSW}}(S)\right)^{-1}$$

$$= \left(\eta_u \sum_{S \in \mathcal{A}} w_{E,\mathbf{p},\boldsymbol{\eta}^{u \leftarrow 1},\sigma}^{\text{GSW}}(S)\right)\left(\eta_u \sum_{S \in \mathcal{A}} w_{E,\mathbf{p},\boldsymbol{\eta}^{u \leftarrow 1},\sigma}^{\text{GSW}}(S) + \sum_{S \notin \mathcal{A}} w_{E,\mathbf{p},\boldsymbol{\eta}^{u \leftarrow 1},\sigma}^{\text{GSW}}(S)\right)^{-1}$$

$$= \frac{\eta_u R}{\eta_u R + 1} = q_0.$$

$$\square$$

Fact 4.10. $\mu_{E,\mathbf{p},\boldsymbol{\eta}^{u \leftarrow 0},\sigma^{u \leftarrow 0}}^{\text{GSW}} = \mu_{E,\mathbf{p},\boldsymbol{\eta},\sigma}^{\text{GSW}}(\cdot \mid \overline{\mathcal{A}})$ *and* $\mu_{E,\mathbf{p},\boldsymbol{\eta}^{u \leftarrow 0},\sigma^{u \leftarrow 1}}^{\text{GSW}} = \mu_{E,\mathbf{p},\boldsymbol{\eta},\sigma}^{\text{GSW}}(\cdot \mid \mathcal{A})$ [3].

Now, we are ready to prove Property (3).

Proof of Property (3). It suffices to prove that, for any valid input $((V, E), \mathbf{p}, \boldsymbol{\eta}, \sigma, u, U)$, a pair of configurations (X, Y) drawn in procedure Couple$((V, E), \mathbf{p}, \boldsymbol{\eta}, \sigma, u, U)$ satisfies

$$X \sim \mu_{E,\mathbf{p},\boldsymbol{\eta},\sigma}^{\text{GSW}} \text{ and } Y \sim \mu_{E,\mathbf{p},\boldsymbol{\eta},\sigma \oplus \mathbf{1}_u}^{\text{GSW}}.$$

Without loss of generality, we only prove $X \sim \mu_{E,\mathbf{p},\boldsymbol{\eta},\sigma}^{\text{GSW}}$. We prove by induction on $m = |E|$.

The base case $m = 0$ is trivial. Suppose Property (3) holds for all E with $|E| = m' < m$. We will show that it also holds when $|E| = m$. We will considering two cases: (1) $u \in U$; (2) $u \notin U$.

When $u \in U$, Algorithm 3 will

- select an arbitrary $e = (u, v) \in E_u = \{f \in E \mid f = (u, y) \text{ for some } y\}$;
- sample $X_1 \sim \nu$, which is the distribution $\mu_{E,\mathbf{p},\boldsymbol{\eta},\sigma}^{\text{GSW}}$ projected on e;

---

[3] We only consider the case where distributions are well-defined, i.e. $Z_{E,\mathbf{p},\boldsymbol{\eta}^{u \leftarrow c},\sigma^{u \leftarrow c}}^{\text{GSW}} > 0$ for $c = 0, 1$.

- sample C_1 via procedure $\mathrm{Couple}(u, E \setminus \{e\}, \boldsymbol{\eta}, \sigma^\star, U)$ or $\mathrm{Couple}(u, E \setminus \{e\}, \boldsymbol{\eta}, \sigma^\star, U)$, where

$$\sigma^\star = \begin{cases} \sigma & , X_1 = \varnothing; \\ \sigma \oplus \mathbf{1}_u \oplus \mathbf{1}_v & , X_1 = \{e\} \end{cases}$$

By induction hypothesis, $C_1 \sim \mu^{\mathrm{GSW}}_{E \setminus \{e\}, \boldsymbol{p}, \boldsymbol{\eta}, \sigma^\star}$. By the definition of $\sigma^\star$, it holds that

$$\mu^{\mathrm{GSW}}_{E \setminus \{e\}, \boldsymbol{p}, \boldsymbol{\eta}, \sigma^\star} = \begin{cases} \mu^{\mathrm{GSW}}_{E, \boldsymbol{p}, \boldsymbol{\eta}, \sigma}(\cdot \mid e) & , X_1 = \{e\}, \\ \mu^{\mathrm{GSW}}_{E, \boldsymbol{p}, \boldsymbol{\eta}, \sigma}(\cdot \mid \overline{e}) & , X_1 = \varnothing. \end{cases}$$

Hence, $X = C_1 \cup X_1 \sim \mu^{\mathrm{GSW}}_{E, \boldsymbol{p}, \boldsymbol{\eta}, \sigma}$.

When $u \notin U$, let q_0, q_1 be defined in Line 4 in Algorithm 3. Algorithm 3 behaves as follows:

- with probability $1 - q_1$, sample X from distribution $\mu^{\mathrm{GSW}}_{E, \boldsymbol{p}, \boldsymbol{\eta}^{u \leftarrow 0}, \sigma^{u \leftarrow 0}}$;

- with probability q_0, sample X from distribution $\mu^{\mathrm{GSW}}_{E, \boldsymbol{p}, \boldsymbol{\eta}^{u \leftarrow 0}, \sigma^{u \leftarrow 1}}$;

- with probability remaining probability $q_1 - q_0$, sample X from distribution $\mu^{\mathrm{GSW}}_{E, \boldsymbol{p}, \boldsymbol{\eta}^{u \leftarrow 0}, \sigma}$.

Here, the last term follows from the analysis of previous case. Hence for $S \subseteq E$, it holds that

$$\mathbf{Pr}\left[X = S\right] = q_0 \mu^{\mathrm{GSW}}_{E, \boldsymbol{p}, \boldsymbol{\eta}^{u \leftarrow 0}, \sigma^{u \leftarrow 1}}(S) + (q_1 - q_0) \mu^{\mathrm{GSW}}_{E, \boldsymbol{p}, \boldsymbol{\eta}^{u \leftarrow 0}, \sigma}(S) + (1 - q_1) \mu^{\mathrm{GSW}}_{E, \boldsymbol{p}, \boldsymbol{\eta}^{u \leftarrow 0}, \sigma^{u \leftarrow 0}}(S).$$

When $\sigma_u = 0$, it holds that

$$\mathbf{Pr}\left[X = S\right] = q_0 \mu^{\mathrm{GSW}}_{E, \boldsymbol{p}, \boldsymbol{\eta}^{u \leftarrow 0}, \sigma^{u \leftarrow 1}}(S) + (1 - q_0) \mu^{\mathrm{GSW}}_{E, \boldsymbol{p}, \boldsymbol{\eta}^{u \leftarrow 0}, \sigma^{u \leftarrow 0}}(S)$$

$$= \mu^{\mathrm{GSW}}_{E, \boldsymbol{p}, \boldsymbol{\eta}, \sigma^{u \leftarrow 0}}(\mathcal{A}) \mu^{\mathrm{GSW}}_{E, \boldsymbol{p}, \boldsymbol{\eta}, \sigma}(S \mid \mathcal{A}) + \mu^{\mathrm{GSW}}_{E, \boldsymbol{p}, \boldsymbol{\eta}, \sigma^{u \leftarrow 0}}(\overline{\mathcal{A}}) \mu^{\mathrm{GSW}}_{E, \boldsymbol{p}, \boldsymbol{\eta}, \sigma}(S \mid \overline{\mathcal{A}})$$

(by $\sigma_u = 0$) $\quad = \mu^{\mathrm{GSW}}_{E, \boldsymbol{p}, \boldsymbol{\eta}, \sigma}(S),$

where the second equation holds by Proposition 4.9, Fact 4.10.

Similarly, when $\sigma_u = 1$, it holds that

$$\mathbf{Pr}\left[X = S\right] = q_1 \mu^{\mathrm{GSW}}_{E, \boldsymbol{p}, \boldsymbol{\eta}^{u \leftarrow 0}, \sigma^{u \leftarrow 1}}(S) + (1 - q_1) \mu^{\mathrm{GSW}}_{E, \boldsymbol{p}, \boldsymbol{\eta}^{u \leftarrow 0}, \sigma^{u \leftarrow 0}}(S)$$

$$= \mu^{\mathrm{GSW}}_{E, \boldsymbol{p}, \boldsymbol{\eta}, \sigma^{u \leftarrow 1}}(\mathcal{A}) \mu^{\mathrm{GSW}}_{E, \boldsymbol{p}, \boldsymbol{\eta}, \sigma}(S \mid \mathcal{A}) + \mu^{\mathrm{GSW}}_{E, \boldsymbol{p}, \boldsymbol{\eta}, \sigma^{u \leftarrow 1}}(\overline{\mathcal{A}}) \mu^{\mathrm{GSW}}_{E, \boldsymbol{p}, \boldsymbol{\eta}, \sigma}(S \mid \overline{\mathcal{A}})$$

(by $\sigma_u = 1$) $\quad = \mu^{\mathrm{GSW}}_{E, \boldsymbol{p}, \boldsymbol{\eta}, \sigma}(S),$

where the second equation holds by Proposition 4.9, Fact 4.10.

Combining these two cases, we have $X \sim \mu^{\mathrm{GSW}}_{E, \boldsymbol{p}, \boldsymbol{\eta}, \sigma}$ and hence complete the proof. $\quad\square$

4.2. **Lifting coupling independence.** In this section, we will prove Lemma 4.5. Let μ be the distribution of random cluster model specified by graph $G = (V, E)$, and parameters $\boldsymbol{p} \in [0, 1]^E$, $\boldsymbol{\lambda} \in [0, 1)^V$. Furthermore, let ν be the distribution of the subgraph-world model model specified by the same graph G, and parameters $\frac{\boldsymbol{p}}{2} = \left(\frac{p_e}{2}\right)_{e \in E}$ and $\eta = \left(\frac{1 - \lambda_v}{1 + \lambda_v}\right)_{v \in V}$. A natural coupling between μ and ν is observed by previous works [GJ18, FGW22a], which is summarized below.

Lemma 4.11 ([FGW22a, Lemma 3.3]). *Let* $q := (p_i/(2 - p_i))_{i \in E}$. *Suppose* $Z \sim \bigotimes_{i \in E} \mathrm{Ber}(q_i)$, $X \sim \nu$, *and* $Y = X \cup Z$ *then it holds that* $Y \sim \mu$. *Equivalently, for every* $Y \subseteq E$, *we have*

$$(4) \qquad \mu(Y) = \sum_{X \subseteq Y} \nu(X) \prod_{h \in Y \setminus X} q_h \prod_{f \in E \setminus Y} (1 - q_f).$$

Now, fix $e \in E$, let (X_0, X_1) be a coupling of $\nu(\cdot \mid \overline{e})$ and $\nu(\cdot \mid e)$. Let $t_e := \frac{q_e \nu(\overline{e})}{q_e \nu(\overline{e}) + \nu(e)}$ be a real number. We will construct (Y_0, Y_1) as a coupling of $\mu(\cdot \mid \overline{e})$ and $\mu(\cdot \mid e)$ as follow:

- sample $Z \sim \bigotimes_{f \in E} \mathrm{Ber}(q_f)$;
- with probability t_e, let $Y_0 = (X_0 \cup Z) \setminus \{e\}$ and $Y_1 = (X_0 \cup Z) \cup \{e\}$;
- with probability $1 - t_e$, let $Y_0 = (X_0 \cup Z) \setminus \{e\}$ and $Y_1 = X_1 \cup Z$.

Now, Lemma 4.5 could be simply proved by

$$\mathbb{E}[|Y_0 \oplus Y_1|] = t_e + (1 - t_e) \cdot \mathbb{E}[|(X_0 \cup Z) \setminus \{e\} \oplus X_1 \cup Z|]$$
$$\leqslant t_e + (1 - t_e) \cdot \mathbb{E}[|X_0 \oplus X_1|] \leqslant C,$$

where in the last inequality, we use the fact that $C \geqslant 1$, which holds by definition.

Now, we only need to verify that Y_0, Y_1 follow the correct distribution as we claimed. For convenience, for $S \subseteq E$, we use q^S to denote $\prod_{i \in S} q_i$. Then for any $Y \subseteq E \setminus \{e\}$,

$$\mathbf{Pr}[Y_0 = Y] = \sum_{T \subseteq Y} \nu(T \mid \overline{e}) q^{Y \setminus T} (1 - q)^{E \setminus Y \setminus \{e\}}$$
$$= \frac{\sum_{T \subseteq Y} \nu(T) q^{Y \setminus T} (1 - q)^{E \setminus Y}}{\nu(\overline{e})(1 - q_e)} \overset{(\star)}{=} \frac{\mu(Y)}{\mu(\overline{e})} = \mu(Y \mid \overline{e}),$$

where in $(\star)$ we use (4) and the fact that $\mu(\overline{e}) = \nu(\overline{e})(1 - q_e)$, which could be implied from Lemma 4.11. A similar calculation shows that for any $\{e\} \subseteq Y \subseteq E$,

$$\mathbf{Pr}[Y_1 = Y] = t_e \sum_{T \subseteq Y \setminus \{e\}} \nu(T \mid \overline{e}) q^{Y \setminus \{e\} \setminus T}(1 - q)^{E \setminus Y} + (1 - t_e) \sum_{\{e\} \subseteq T \subseteq Y} \nu(T \mid e) q^{Y \setminus T}(1 - q)^{E \setminus Y}$$
$$= \left(t_e \sum_{T \subseteq Y \setminus \{e\}} \frac{\nu(T)}{\nu(\overline{e})} q^{Y \setminus \{e\} \setminus T} + (1 - t_e) \sum_{\{e\} \subseteq T \subseteq Y} \frac{\nu(T)}{\nu(e)} q^{Y \setminus T} \right) (1 - q)^{E \setminus Y}$$
$$\overset{(+)}{=} \frac{\left(\sum_{T \subseteq Y \setminus \{e\}} \nu(T) q^{Y \setminus T} + \sum_{\{e\} \subseteq T \subseteq Y} \nu(T) q^{Y \setminus T} \right) (1 - q)^{E \setminus Y}}{q_e \nu(\overline{e}) + \nu(e)}$$
$$= \frac{\sum_{T \subseteq Y} \nu(T) q^{Y \setminus T} (1 - q)^{E \setminus Y}}{q_e \nu(\overline{e}) + \nu(e)} \overset{(\star)}{=} \frac{\mu(Y)}{\mu(e)} = \mu(Y \mid e),$$

where $(+)$ holds by the fact that $t_e = \frac{q_e \nu(\overline{e})}{q_e \nu(\overline{e}) + \nu(e)}$ and $(\star)$ holds by Lemma 4.11.

5. Rapid mixing of Glauber dynamics in good regime

Let μ be the distribution of the random cluster model specified by graph $G = (V, E)$, parameters $p \in (0, 1)^E$ and $\lambda \in (0, 1)^V$. Let P_μ^{GD} be the Glauber dynamics with stationary distribution μ.

We restate Lemma 3.8 for convenience.

Lemma 5.1. *Let* $\varepsilon \in \left(0, \frac{1}{4}\right)$ *be a real number. If*

$$(5) \qquad K := (1 - p_{\min}) \log n \leqslant \min\left\{ 10^{-7}, \frac{1 - \lambda_{\max}}{27} \right\},$$

then the mixing time of Glauber dynamics satisfies

$$T_{\mathrm{mix}}(\varepsilon) \leqslant 25 m \log m \log(1/\varepsilon) + 1.$$

Furthermore, if

$$K \leqslant \min\left\{ 10^{-5} \exp\left(-\frac{\log(8/\varepsilon)}{\log n} \right), \frac{1 - \lambda_{\max}}{27} \right\},$$

then the mixing time of Glauber dynamics initialized from E *satisfies*

$$T_{\mathrm{mix}}(\varepsilon, E) \leqslant 2m(\log m + \log(2/\varepsilon)).$$

We remark that a trade-off exists between the mixing time and the condition on K. This trade-off arises from a subtle distinction in the proof of the mixing time. To prove the first part, we employ the burn-in method and a "doubling" argument, whereas the second part does not rely on these techniques. We will use the first part of Lemma 5.1 in the subsequent analysis, and the proof of the second part is deferred to Appendix B.

In the rest of this section, we will always assume that K is picked as in (5). Define $\mathcal{C}$ by

$$\mathcal{C} := \{S \subseteq V \mid |S| \leqslant n/2 \text{ and } |E(S, V \setminus S)| \geqslant |S| \log n\},$$

where $E(S, V \setminus S)$ is the set of edges between S and $V \setminus S$. Furthermore, define the good event $\mathcal{G}$ by

$$\mathcal{G} := \{X \subseteq E \mid \forall S \in \mathcal{C}, |X \cap E(S, V \setminus S)| > 0\}.$$

Intuitively, the good event $\mathcal{G}$ is the set of configurations $X \subseteq E$, such that there exists no connected components C in graph (V, X) with large $|E(C, V \setminus C)|$.

Let $(X_t)_{t \geqslant 0}$ and $(Y_t)_{t \geqslant 0}$ be Glauber dynamics starting from different configurations X_0 and Y_0. We will show that after a burn-in phase of length $O(m \log m)$, the good event $\mathcal{G}$ happens in high probability. Furthermore, when $X_t, Y_t \in \mathcal{G}$, there exists a coupling so that the hamming distance of X_t and Y_t contracts with rate $1 - \frac{1}{2m}$, where $m = |E|$.

Lemma 5.2. *If $t \geqslant \theta m \log m$ for some $\theta > 0$, it holds that*

$$\mathbf{Pr}\left[X_t \notin \mathcal{G}\right] \leqslant n^{\log(27K)} + m^{1-\theta} \text{ and } \mathbf{Pr}\left[Y_t \notin \mathcal{G}\right] \leqslant n^{\log(27K)} + m^{1-\theta}.$$

Lemma 5.3. *When $K \leqslant \frac{1 - \lambda_{\max}}{27}$, there exists a coupling of $(X_t)_{t \geqslant 0}$ and $(Y_t)_{t \geqslant 0}$ so that when $X_t, Y_t \in \mathcal{G}$,*

$$\mathbb{E}\left[|X_{t+1} \oplus Y_{t+1}| \mid X_t, Y_t\right] \leqslant \left(1 - \frac{1}{2m}\right) |X_t \oplus Y_t|.$$

The proof of Lemma 5.2 and Lemma 5.3 is deferred to Section 5.2 and Section 5.1 respectively. To prove Lemma 5.1 via Lemma 5.2 and Lemma 5.3, we need the coupling with stationary lemma.

Lemma 5.4 ([HV06, Theorem 3.1]). *Let $(X_t)_{t \geqslant 0}$ and $(Y_t)_{t \geqslant 0}$ be coupled Markov chains. Let $\varepsilon \in (0, 1)$, $T_0, T_1 \in \mathbb{N}$ be parameters satisfying $T_0 < T_1$. Furthermore, $\mathcal{E}_t$ denote the event*

$$\mathbb{E}\left[|X_{t+1} \oplus Y_{t+1}| \mid X_t, Y_t\right] \leqslant \left(1 - \frac{1}{2m}\right) |X_t \oplus Y_t|.$$

Suppose $\mathbf{Pr}\left[\overline{\mathcal{E}_t}\right] \leqslant \delta$ for all $T_0 \leqslant t < T_1$, then

$$\mathbf{Pr}\left[X_{T_1} \neq Y_{T_1}\right] \leqslant \left(1 - \frac{1}{2m}\right)^{T_1 - T_0} m + \delta \cdot 2m^2.$$

We are now ready to prove Lemma 5.1.

Proof for the first part of Lemma 5.1. Without loss of generality, we will only prove the case where $\varepsilon = \frac{1}{4}$. This can be extended to general case by a doubling argument [LPW17]. Moreover, we assume $n \geqslant 2$ and $m \geqslant 2$. Since $n \leqslant 1$ implies that $m = 0$ and when $m \leqslant 1$, Glauber dynamics mixes in one step.

Let $T_0 = 9m \log m$, $T_1 = 15m \log m$ and X_0, Y_0 be arbitrary starting configurations. We consider the Glauber dynamics $(X_t)_{t \geqslant 0}$ and $(Y_t)_{t \geqslant 0}$ starting from X_0 and Y_0 respectively. By Lemma 5.2 and Lemma 5.3, there exists a coupling of $(X_t)_{t \geqslant 0}$ and $(Y_t)_{t \geqslant 0}$ such that for all $T_0 \leqslant t < T_1$, the event $\mathcal{E}_t$ happens with probability no more than $2n^{\log(27K)} + 2m^{-8}$, where $\mathcal{E}_t$ denotes the event $\mathbb{E}\left[|X_{t+1} \oplus Y_{t+1}| \mid X_t, Y_t\right] \leqslant \left(1 - \frac{1}{2m}\right) |X_t \oplus Y_t|$. By $K \leqslant 10^{-7}$, it holds that

$$\left(2n^{\log(27K)} + 2m^{-8}\right) \cdot 2m^2 \leqslant 4n^{4+\log(27K)} + 4m^{-6} \leqslant 4n^{-6} + 4m^{-6} \leqslant \frac{1}{8},$$

17

where the last inequality follows from our assumption $n, m \geqslant 2$. Therefore, by Lemma 5.4,

$$\mathbf{Pr}\left[X_{T_1} \neq Y_{T_1}\right] \leqslant \left(1 - \frac{1}{2m}\right)^{T_1 - T_0} m + \frac{1}{8} \leqslant \frac{1}{4}.$$

Finally, by the geometric convergence of Markov chain, it holds that

$$T_{\mathrm{mix}}(\varepsilon) \leqslant 15 m \log m \log_2(1/\varepsilon) \leqslant 25 m \log m \log(1/\varepsilon) + 1,$$

where the $+1$ is used to handle the cases where $n \leqslant 1$ or $m \leqslant 1$. $\qquad\square$

5.1. **Contraction in $\mathcal{G}$ (proof of Lemma 5.3).** For any $S, T \in \mathcal{G}$, let $e_1, e_2, \ldots, e_x$ be edges in $S \setminus T$, and $f_1, f_2, \ldots, f_y$ be edges in $T \setminus S$. We design the following path of configurations $P = (P_0, P_1, \ldots, P_{x+y})$ from S to T.

$$\forall 0 \leqslant i \leqslant s + t, \quad P_i = \begin{cases} S & \text{if } i = 0, \\ P_{i-1} \cup \{f_i\} & \text{if } 1 \leqslant i \leqslant y, \\ P_{i-1} \setminus \{e_{i-y}\} & \text{if } i > y. \end{cases}$$

Note that $|P_{i-1} \oplus P_i| = 1$ for all $1 \leqslant i \leqslant x + y$, $P_i \in \mathcal{G}$ for all $0 \leqslant i \leqslant x + y$, and the length of this path is exactly $|S \oplus T|$. Hence, by the standard path coupling argument [BD97], it suffices to prove that, for every $X_t, Y_t \in \mathcal{G}$ satisfying $|X_t \oplus Y_t| = 1$,

$$(6) \qquad\qquad \mathbb{E}\left[|X_{t+1} \oplus Y_{t+1}| \mid X_t, Y_t\right] \leqslant 1 - \frac{1}{2m}.$$

Without loss of generality, we assume that $Y_t = X_t \cup \{e\}$, where $e = (u, v) \in E$. Using the one step optimal coupling of the Glauber dynamics, it holds that

$$\mathbb{E}\left[|X_{t+1} \oplus Y_{t+1}| \mid X_t, Y_t\right] = 1 - \frac{1}{m} + \frac{1}{m} \sum_{f \in E \setminus \{e\}} d(X_t, Y_t, f),$$

where $d(X_t, Y_t, f) = \left| \frac{\mu(X_t \cup \{f\})}{\mu(X_t \cup \{f\}) + \mu(X_t \setminus \{f\})} - \frac{\mu(Y_t \cup \{f\})}{\mu(Y_t \cup \{f\}) + \mu(Y_t \setminus \{f\})} \right|$.

We first discuss the value of $d(X_t, Y_t, f)$. Fix edge $f = (x, y) \neq e$, let

$$S_{\overline{f}, \overline{e}} := X_t \setminus \{f\}, S_{f, \overline{e}} := X_t \cup \{f\}, S_{\overline{f}, e} := Y_t \setminus \{f\}, S_{f, e} := Y_t \cup \{f\}.$$

It can be verified that

$$\frac{\mu(X_t \cup \{f\})}{\mu(X_t \cup \{f\}) + \mu(X_t \setminus \{f\})} = \begin{cases} p_f, & \text{if } C_{x, \overline{f}, \overline{e}} = C_{y, \overline{f}, \overline{e}}, \\ \frac{1 + \lambda^{C_{x, \overline{f}, \overline{e}} \cup C_{y, \overline{f}, \overline{e}}}}{1 + \lambda^{C_{x, \overline{f}, \overline{e}} \cup C_{y, \overline{f}, \overline{e}}} + (1 - p_f)\left(\lambda^{C_{x, \overline{f}, \overline{e}}} + \lambda^{C_{y, \overline{f}, \overline{e}}}\right)} & \text{otherwise,} \end{cases}$$

$$\frac{\mu(Y_t \cup \{f\})}{\mu(Y_t \cup \{f\}) + \mu(Y_t \setminus \{f\})} = \begin{cases} p_f & \text{if } C_{x, \overline{f}, e} = C_{y, \overline{f}, e}, \\ \frac{1 + \lambda^{C_{x, \overline{f}, e} \cup C_{y, \overline{f}, e}}}{1 + \lambda^{C_{x, \overline{f}, e} \cup C_{y, \overline{f}, e}} + (1 - p_f)\left(\lambda^{C_{x, \overline{f}, e}} + \lambda^{C_{y, \overline{f}, e}}\right)} & \text{otherwise,} \end{cases}$$

where $C_{r, i, j}$ is the connected component containing r in graph $(V, S_{i,j})$ where $i \in \{f, \overline{f}\}$ and $j \in \{e, \overline{e}\}$, and $\lambda^S = \prod_{r \in S} \lambda_r$. For general subset $R \subseteq E$, we also use $C_r(R)$ to denote the connected component containing r in graph (V, R). We consider the following cases as illustrated in Figure 2.

(1) $C_{u, \overline{f}, \overline{e}} = C_{v, \overline{f}, \overline{e}}$;
(2) $C_{u, \overline{f}, \overline{e}} \neq C_{v, \overline{f}, \overline{e}}$ and $C_{u, f, \overline{e}} = C_{v, f, \overline{e}}$;
(3) $C_{u, \overline{f}, \overline{e}} \neq C_{v, \overline{f}, \overline{e}}$ and $C_{u, f, \overline{e}} \neq C_{v, f, \overline{e}}$.

18

FIGURE 2. Illustration of three cases. Each ellipse is a connected component in graph $(V, S_{\bar{f},\bar{e}})$.

5.1.1. *Case* (1). In this case, $(V, S_{\bar{f},\bar{e}})$ and $(V, S_{\bar{f},e})$ have the same structure of connected components. Hence we have, $C_{x,\bar{f},\bar{e}} = C_{x,\bar{f},e}$ and $C_{y,\bar{f},\bar{e}} = C_{y,\bar{f},e}$, which implies $d(X_t, Y_t, f) = 0$.

5.1.2. *Case* (2). In this case, both edges $e = (u,v)$ and $f = (x,y)$ connect $C_{u,\bar{f},\bar{e}}$ and $C_{v,\bar{f},\bar{e}}$ in graph $(V, S_{\bar{f},\bar{e}})$. Thus, $C_{x,\bar{f},\bar{e}} \neq C_{y,\bar{f},\bar{e}}$ and $C_{x,\bar{f},e} = C_{y,\bar{f},e}$, which implies

$$d(X_t, Y_t, f) = \frac{p_f(1-p_f)(\lambda^{C_{u,\bar{f},\bar{e}}} + \lambda^{C_{v,\bar{f},\bar{e}}})}{1 + (1-p_f)(\lambda^{C_{u,\bar{f},\bar{e}}} + \lambda^{C_{v,\bar{f},\bar{e}}}) + \lambda^{C_{u,\bar{f},\bar{e}} \cup C_{v,\bar{f},\bar{e}}}}$$

$$(7) \qquad\qquad \leqslant (1-p_{\min})(\lambda_{\max}^{\left|C_{u,\bar{f},\bar{e}}\right|} + \lambda_{\max}^{\left|C_{v,\bar{f},\bar{e}}\right|}).$$

We further consider two sub-cases: (a) $f \in X_t$; (b) $f \notin X_t$.

Case (2a). As $f \in X_t$, it holds that $C_u(X_t) = C_v(X_t)$ and there exists a path $P = (x_0, x_1, \ldots, x_\ell)$ that connects u and v in the graph (G, X_t) for some $\ell > 0$. Note that only edges on path P may satisfy the requirement of Case (2) i.e. $C_{u,\bar{f},\bar{e}} \neq C_{v,\bar{f},\bar{e}}$. Therefore, by (7), the sum of $d(X_t, Y_t, f)$ in this case can be bounded by

$$(8) \qquad \sum_{f \in P} (1-p_{\min})(\lambda_{\max}^{\left|C_{u,\bar{f},\bar{e}}\right|} + \lambda_{\max}^{\left|C_{v,\bar{f},\bar{e}}\right|}) \overset{(\star)}{\leqslant} (1-p_{\min}) \sum_{i=1}^{\ell} (\lambda_{\max}^{i} + \lambda_{\max}^{\ell+1-i}) \leqslant \frac{2K}{1-\lambda_{\max}},$$

where $(\star)$ follows from the fact that $C_{u,\bar{f},\bar{e}}$ and $C_{v,\bar{f},\bar{e}}$ contains at least i and $\ell+1-i$ vertices respectively, where $f = (x_{i-1}, x_i)$.

Case (2b). As $f \notin X_t$, it holds that $X_t = S_{\bar{f},e}$, and only edges in $E(C_{u,\bar{f},\bar{e}}, C_{v,\bar{f},\bar{e}})$ may satisfy the requirement of Case (2) (i.e. f connects two components), where we recall that $E(C_{u,\bar{f},\bar{e}}, C_{v,\bar{f},\bar{e}})$ denotes the set of edges between $C_{u,\bar{f},\bar{e}}$ and $C_{v,\bar{f},\bar{e}}$. Therefore, the sum of $d_{X_t,Y_t,f}$ in this case can be bounded by

$$\left|E(C_{u,\bar{f},\bar{e}}, C_{v,\bar{f},\bar{e}})\right| \cdot (1-p_{\min})(\lambda_{\max}^{\left|C_{u,\bar{f},\bar{e}}\right|} + \lambda_{\max}^{\left|C_{v,\bar{f},\bar{e}}\right|})$$

$$\overset{(\star)}{\leqslant} \log n \cdot \min\left\{\left|C_{u,\bar{f},\bar{e}}\right|, \left|C_{v,\bar{f},\bar{e}}\right|\right\} \cdot (1-p_{\min})(\lambda_{\max}^{\left|C_{u,\bar{f},\bar{e}}\right|} + \lambda_{\max}^{\left|C_{v,\bar{f},\bar{e}}\right|})$$

$$\leqslant 2\log n \cdot (1-p_{\min}) \cdot \max_{z>0} z\, \lambda_{\max}^{z}$$

$$(9) \qquad\qquad \leqslant \frac{2K}{1-\lambda_{\max}},$$

where $(\star)$ follows from fact that $X_t \in \mathcal{G}$.

19

5.1.3. *Case* (3). Note that $d(X_t, Y_t, f) = 0$ when

(10) $$C_{u,\overline{f},\overline{e}} = C_{u,f,\overline{e}} \text{ and } C_{v,\overline{f},\overline{e}} = C_{v,f,\overline{e}},$$

i.e. both x and y are not in $C_{u,\overline{f},\overline{e}} \cup C_{v,\overline{f},\overline{e}}$ where $f = (x, y)$. Furthermore, two constraints in (10) cannot be violated at the same time, as $f = (x, y)$ does not connect $C_{u,\overline{f},\overline{e}}$ and $C_{v,\overline{f},\overline{e}}$. Therefore, it suffices to consider the case where $C_{u,\overline{f},\overline{e}} \neq C_{u,f,\overline{e}}$ and $C_{v,\overline{f},\overline{e}} = C_{v,f,\overline{e}}$.

Let $A = C_{u,\overline{f},\overline{e}}, B = C_{v,\overline{f},\overline{e}}, C = C_{u,f,\overline{e}} \setminus A$. A similar calculation yields

$$d(X_t, Y_t, f) = \left| \frac{p_f\left(1 + \lambda^{A \cup B}\right)}{1 + \lambda^{A \cup B} + (1 - p_f)\left(\lambda^A + \lambda^B\right)} - \frac{p_f\left(1 + \lambda^{A \cup B \cup C}\right)}{1 + \lambda^{A \cup B \cup C} + (1 - p_f)\left(\lambda^{A \cup C} + \lambda^B\right)} \right|$$

(11) $$\leqslant (1 - p_{\min})\lambda_{\max}^{|A|}.$$

We further consider two sub-cases: (a) $f \in X_t$; (b) $f \notin X_t$.

Case (3a). When $f \in X_t$, in order to make contribution, the deletion of f split $C_{u,f,\overline{e}}$ into $A = C_{u,\overline{f},\overline{e}}$ and $C = C_{u,f,\overline{e}} \setminus A$. In this case, f must be a bridge in graph $C_{u,f,\overline{e}}$. Pick an arbitrary spanning tree T rooted at u in $C_{u,f,\overline{e}}$. Obviously, f must be an edge in T. Let ℓ be the size of T and $\{f_1, f_2, \ldots, f_\ell\}$ be the edges in T sorted by the size of $C_{u,\overline{f_i},\overline{e}}$ in decreasing order. Note that f_j must be in component $C_{u,\overline{f_i},\overline{e}}$ for all $1 \leqslant i < j \leqslant \ell$, since $|C_{u,\overline{f_i},\overline{e}}| \geqslant |C_{u,\overline{f_j},\overline{e}}|$ means f_i cannot be an ancestor of f_j. Hence, $|C_{u,\overline{f_i},\overline{e}}| \geqslant \ell - i + 1$. Together with (11), we bound the sum of $d(X_t, Y_t, f)$ in this case by

(12) $$\sum_{i=1}^{\ell} (1 - p_{\min})\lambda_{\max}^i \leqslant \frac{\lambda_{\max}(1 - p_{\min})}{1 - \lambda_{\max}} \leqslant \frac{K}{1 - \lambda_{\max}}.$$

Case (3b). When $f \notin X_t$, f must connects $C_{u,\overline{f},\overline{e}}$ and other component in $(V, S_{\overline{f},\overline{e}})$ except $C_{v,\overline{f},\overline{e}}$. Therefore, the sum of $d(X_t, Y_t, f)$ in this case is bounded by (similar calculation as in (9))

(13) $$\left| E(C_{u,\overline{f},\overline{e}}, V \setminus C_{u,\overline{f},\overline{e}}) \right| \cdot (1 - p_{\min})\lambda_{\max}^{|C_u|} \leqslant \frac{K}{1 - \lambda_{\max}},$$

where we use (11) and the fact that $S_{\overline{f},\overline{e}} = X_t \in \mathcal{G}$.

5.1.4. *Wrapping up.* Recall that by our assumption of K in (5.3), it holds that $K \leqslant (1 - \lambda_{\max})/27$. Together with previous analysis, we bound $\sum_{f \in E \setminus \{e\}} d(X_t, Y_t, f)$ by

(14) $$\sum_{f \in E \setminus \{e\}} d(X_t, Y_t, f) \leqslant (8) + (9) + 2 \times (12) + 2 \times (13) \leqslant \frac{8K}{1 - \lambda_{\max}} \leqslant \frac{1}{2}.$$

Therefore, it holds that

$$\mathbb{E}\left[|X_{t+1} \oplus Y_{t+1}| \mid X_t, Y_t\right] \leqslant 1 - \frac{1}{m}\left(1 - \frac{1}{2}\right) = 1 - \frac{1}{2m},$$

which concludes the proof of Lemma 5.3.

5.2. **Bad event happens with small probability (proof of Lemma 5.2).** Without loss of generality, we will only bound the probability $\mathbf{Pr}\left[X_t \notin \mathcal{G}\right]$. For simplicity of notation, we denote *update sequence* $\mathcal{L}_t = (e_1, e_2, \ldots, e_t)$ the chosen edges of the Glauber dynamics in the first t rounds, and denote $\mathcal{S}(\mathcal{L}_t) = \{e_1, e_2, \ldots, e_t\}$ be the set of edges of an update sequence $\mathcal{L}_t$.

Fix $t \geqslant \theta m \log m$ for some $\theta > 0$. We claim that $\mathcal{S}(\mathcal{L}_t) = E$ happens with high probability. Furthermore, condition on any updating sequence $\mathcal{L}_t$ with $\mathcal{S}(\mathcal{L}_t) = E$, the distribution X_t stochastically dominates the product distribution $\nu = \bigotimes_{e \in E} \text{Ber}(1 - 3K)$, where $K = (1 - p_{\min})\log n$.

20

Lemma 5.5. *If* $t \geqslant \theta m \log m$ *for some* $\theta > 0$, *event* $S(\mathcal{L}_t) \neq E$ *occurs with probability at most* $m^{1-\theta}$.

Lemma 5.6. *For every* $S \subseteq E$ *and* $e \in E$, *it holds that*

$$\frac{\mu(S \cup \{e\})}{\mu(S \cup \{e\}) + \mu(S \setminus \{e\})} \geqslant 1 - 3K.$$

Lemma 5.7. *For any updating sequence* $\mathcal{L}_t$ *satisfying* $S(\mathcal{L}_t) = E$ *the distribution* X_t *condition on* $\mathcal{L}_t$ *stochastically dominates* ν, *i.e., there exists a coupling* $C = (S, T)$ *of* $(X_t \mid \mathcal{L}_t)$ *and* ν *satisfying* $T \subseteq S$.

We first prove Lemma 5.2 with Lemma 5.5 and Lemma 5.7.

Proof of Lemma 5.2. For any $t \geqslant \theta m \log m$, it holds that

$$\mathbf{Pr}\left[X_t \notin \mathcal{G}\right] = \sum_{\substack{\mathcal{L}_t \in E^t \\ S(\mathcal{L}_t) = E}} \mathbf{Pr}\left[X_t \notin \mathcal{G} \mid \mathcal{L}_t\right] \mathbf{Pr}\left[\mathcal{L}_t\right] + \sum_{\substack{\mathcal{L}_t \in E^t \\ S(\mathcal{L}_t) \neq E}} \mathbf{Pr}\left[X_t \notin \mathcal{G} \mid \mathcal{L}_t\right] \mathbf{Pr}\left[\mathcal{L}_t\right]$$

$$\leqslant \max_{\substack{\mathcal{L}_t \in E^t \\ S(\mathcal{L}_t) = E}} \mathbf{Pr}\left[X_t \notin \mathcal{G} \mid \mathcal{L}_t\right] + \mathbf{Pr}\left[S(\mathcal{L}_t) \neq E\right]$$

Fix an update sequence $\mathcal{L}_t$ satisfying $S(\mathcal{L}_t) = E$. By Lemma 5.7, the definition of good event $\mathcal{G}$ and the definition of ν,

$$\mathbf{Pr}\left[X_t \notin \mathcal{G} \mid \mathcal{L}_t\right] \leqslant \sum_{S \in \mathcal{C}} (3K)^{|E(S, V \setminus S)|} \leqslant \sum_{S \in \mathcal{C}} (3K)^{|S| \log n} \leqslant \sum_{j=1}^{+\infty} n^j n^{j \log(3k)} \leqslant n^{\log(27K)}.$$

Furthermore, by Lemma 5.5, the probability of $S(\mathcal{L}_t) \neq E$ is upper bounded by $m^{1-\theta}$. Therefore,

$$\mathbf{Pr}\left[X_t \notin \mathcal{G}\right] \leqslant n^{\log(27K)} + m^{1-\theta}. \qquad \square$$

In the rest of this section, we are dedicated to proving Lemma 5.5 and Lemma 5.7.

Proof of Lemma 5.5. Fix $t \geqslant \theta m \log m$. For each edge $e \in E$, the probability that $e \notin \mathcal{L}_t$ is at most

$$\mathbf{Pr}\left[e \notin \mathcal{L}_t\right] \leqslant \left(1 - \frac{1}{m}\right)^t \leqslant \left(1 - \frac{1}{m}\right)^{\theta m \log m} \leqslant m^{-\theta}.$$

By union bound,

$$\mathbf{Pr}\left[S(\mathcal{L}_t) \neq E\right] \leqslant \sum_{e \in E} \mathbf{Pr}\left[e \notin \mathcal{L}_t\right] \leqslant m^{1-\theta}. \qquad \square$$

Proof of Lemma 5.7 assuming lemma 5.6. Instead of proving the original statement, we will prove that for any update sequence $\mathcal{L}_t \in E$, there exists a coupling (S, T) of $(X_t \mid \mathcal{L}_t)$ and ν satisfying that $(T \cap R) \subseteq (S \cap R)$, where $R = S(\mathcal{L}_t)$. We remark that this is stronger compared to the original statement.

We prove by induction on t. Base case $t = 0$ follows from $R = \varnothing$. Now we assume our assumption holds for any $t' \leqslant t$. Fix an update sequence $\mathcal{L}_{t+1} = (e_1, e_2, \ldots, e_{t+1})$. Let $\mathcal{L}_t = (e_1, e_2, \ldots, e_t)$ be the update sequence. By induction hypothesis, there exists a coupling (S, T) of $(X_t \mid \mathcal{L}_t)$ and ν satisfying $(T \cap R) \subseteq (S \cap R)$, where $R = S(\mathcal{L}_t)$. We design the following coupling (S', T') of $(X_t \mid \mathcal{L}_{t+1})$ and ν based on (S, T):

- let $q_0 = \frac{\mu(S \cup \{e_{t+1}\})}{\mu(S \cup \{e_{t+1}\}) + \mu(S \setminus \{e_{t+1}\})}$ and $q_1 = 1 - 3K$;
- draw $r \sim \text{Uniform}(0, 1)$;
    - if $r < q_0$, set $S' = S \cup \{e_{t+1}\}$; Otherwise, set $S' = S \setminus \{e_{t+1}\}$;
    - if $r < q_1$, set $T' = T \cup \{e_{t+1}\}$; Otherwise, set $T' = T \setminus \{e_{t+1}\}$.

21

Note that q_0 is the probability of adding e_{t+1} to S in the update step of Glauber dynamics. It can be verified that (S', T') is indeed a coupling of $(X_{t+1} \mid \mathcal{L}_{t+1})$ and ν. Moreover, if $q_0 \geqslant q_1$, it holds that $(T' \cap R') \subseteq (S' \cap R')$, where $R' = \mathcal{S}(\mathcal{L}_{t+1})$. Hence, it remains to show $q_0 \geqslant q_1 = 1 - 3K$, which follows from lemma 5.6. This concludes the proof of Lemma 5.7. $\qquad\square$

Proof of lemma 5.6. For convenience, let $e = (u, v)$. We consider the following cases.

   (1) u, v are in the same connected component in graph $(V, S \setminus \{e\})$. In this case,

$$\frac{\mu(S \cup \{e\})}{\mu(S \cup \{e\}) + \mu(S \setminus \{e\})} = p_e \geqslant 1 - K/\log n \geqslant 1 - 3K.$$

   (2) u, v are in different connected components in graph $(V, S \setminus \{e\})$. Let C_u and C_v be the connected components that u and v are in respectively. In this case,

$$\frac{\mu(S \cup \{e\})}{\mu(S \cup \{e\}) + \mu(S \setminus \{e\})} = \frac{p_e(1 + \lambda^{C_u \cup C_v})}{1 + \lambda^{C_u \cup C_v} + (1 - p_e)\left(\lambda^{C_u} + \lambda^{C_v}\right)} \geqslant \frac{p_e}{3 - 2p_e}.$$

Recall that $p_e \geqslant 1 - \frac{K}{\log n} \geqslant 1 - K$. Therefore, $\frac{\mu(S \cup \{e\})}{\mu(S \cup \{e\}) + \mu(S \setminus \{e\})} \geqslant \frac{1-K}{1+2K} \geqslant 1 - 3K.$ $\qquad\square$

6. Acknowledgement

References

[AJK+21a] Nima Anari, Vishesh Jain, Frederic Koehler, Huy Tuan Pham, and Thuy-Duong Vuong. Entropic independence I: Modified log-sobolev inequalities for fractionally log-concave distributions and high-temperature ising models. *arXiv preprint arXiv:2106.04105*, 2021.

[AJK+21b] Nima Anari, Vishesh Jain, Frederic Koehler, Huy Tuan Pham, and Thuy-Duong Vuong. Entropic independence II: Optimal sampling and concentration via restricted modified log-sobolev inequalities. *arXiv preprint arXiv:2111.03247*, 2021.

[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.

[ALO22] Dorna Abdolazimi, Kuikui Liu, and Shayan Oveis Gharan. A matrix trickle-down theorem on simplicial complexes and applications to sampling colorings. In *FOCS*, pages 161–172, 2022.

[Bar16] Alexander Barvinok. *Combinatorics and complexity of partition functions*, volume 30 of *Algorithms and Combinatorics*. Springer, Cham, 2016.

[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.

[BCP22a] Antonio Blanca, Sarah Cannon, and Will Perkins. Fast and perfect sampling of subgraphs and polymer systems. *arXiv preprint arXiv:2202.05907*, 2022.

[BCP+22b] Antonio Blanca, Pietro Caputo, Daniel Parisi, Alistair Sinclair, and Eric Vigoda. Entropy decay in the Swendsen-Wang dynamics on $\mathbb{Z}^d$. *Ann. Appl. Probab.*, 32(2):1018–1057, 2022.

[BCŠV21] Antonio Blanca, Zongchen Chen, Daniel Štefankovič, and Eric Vigoda. *The Swendsen-Wang dynamics on trees*, volume 207 of *LIPIcs. Leibniz Int. Proc. Inform.* Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2021.

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

[BT06] Sergey G. Bobkov and Prasad Tetali. Modified logarithmic Sobolev inequalities in discrete settings. *J. Theoret. Probab.*, 19(2):289–336, 2006.

[CE22] Yuansi Chen and Ronen Eldan. Localization schemes: A framework for proving mixing bounds for markov chains. *arXiv preprint arXiv:2203.04163*, 2022.

[CFYZ21a] Xiaoyu Chen, Weiming Feng, Yitong Yin, and Xinyuan Zhang. Optimal mixing time for the ising model in the uniqueness regime. *arXiv preprint arXiv:2111.03034*, 2021.

[CFYZ21b] 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. *arXiv preprint arXiv:2203.07771*, 2022.

[CGG⁺21] Zongchen Chen, Andreas Galanis, Leslie A. Goldberg, Will Perkins, James Stewart, and Eric Vigoda. Fast algorithms at low temperatures via Markov chains. *Random Structures Algorithms*, 58(2):294–321, 2021.

[CGHT16] Andrea Collevecchio, Timothy M. Garoni, Timothy Hyndman, and Daniel Tokarev. The worm process for the Ising model is rapidly mixing. *J. Stat. Phys.*, 164(5):1082–1102, 2016.

[CLV20] Zongchen Chen, Kuikui Liu, and Eric Vigoda. Rapid mixing of Glauber dynamics up to uniqueness via contraction. In *FOCS*, pages 1307–1318, 2020. arXiv:2004.09083.

[CLV21a] Zongchen Chen, Kuikui Liu, and Eric Vigoda. Optimal mixing of Glauber dynamics: Entropy factorization via high-dimensional expansion. In *STOC*, 2021. arXiv:2011.02075.

[CLV21b] Zongchen Chen, Kuikui Liu, and Eric Vigoda. Spectral independence via stability and applications to holant-type problems. *arXiv preprint arXiv:2106.03366*, 2021. (To appear in *FOCS'21*).

[CLV22] Zongchen Chen, Kuikui Liu, and Eric Vigoda. Spectral independence via stability and applications to holant-type problems. In *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 149–160. IEEE, 2022.

[DSC96] P. Diaconis and L. Saloff-Coste. Logarithmic Sobolev inequalities for finite Markov chains. *Ann. Appl. Probab.*, 6(3):695–750, 1996.

[ES88] Robert G. Edwards and Alan D. Sokal. Generalization of the Fortuin-Kasteleyn-Swendsen-Wang representation and Monte Carlo algorithm. *Phys. Rev. D (3)*, 38(6):2009–2012, 1988.

[FGW22a] Weiming Feng, Heng Guo, and Jiaheng Wang. Sampling from the ferromagnetic ising model with external fields. *arXiv preprint arXiv:2205.01985 [v1]*, 2022.

[FGW22b] Weiming Feng, Heng Guo, and Jiaheng Wang. Sampling from the ferromagnetic ising model with external fields. *arXiv preprint arXiv:2205.01985 [v2]*, 2022.

[FK72] Cornelius Marius Fortuin and Piet W Kasteleyn. On the random-cluster model. I. Introduction and relation to other models. *Physica*, 57:536–564, 1972.

[For72a] Cornelius Marius Fortuin. On the random-cluster model. II. The percolation model. *Physica*, 58:393–418, 1972.

[For72b] Cornelius Marius Fortuin. On the random-cluster model. III. The simple random-cluster model. *Physica*, 59:545–570, 1972.

[GJ99] Vivek K Gore and Mark R Jerrum. The swendsen–wang process does not always mix rapidly. *Journal of Statistical Physics*, 97(1):67–86, 1999.

[GJ09] Geoffrey Grimmett and Svante Janson. Random even graphs. *The Electronic Journal of Combinatorics [electronic only]*, 16(1):Research–Paper, 2009.

[GJ18] Heng Guo and Mark Jerrum. Random cluster dynamics for the Ising model is rapidly mixing. *The Annals of Applied Probability*, 28(2):1292–1313, 2018.

[GM07] Antoine Gerschenfeld and Andrea Montanari. Reconstruction for models on random graphs. In *48th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2007), October 20-23, 2007, Providence, RI, USA, Proceedings*, pages 194–204. IEEE Computer Society, 2007.

[GS22] Reza Gheissari and Alistair Sinclair. Low-temperature ising dynamics with random initializations. In Stefano Leonardi and Anupam Gupta, editors, *STOC '22: 54th Annual ACM SIGACT Symposium on Theory of Computing, Rome, Italy, June 20 - 24, 2022*, pages 1445–1458. ACM, 2022.

[GŠV19] Andreas Galanis, Daniel Štefankovič, and Eric Vigoda. Swendsen-Wang algorithm on the mean-field Potts model. *Random Structures Algorithms*, 54(1):82–147, 2019.

[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. (conference version in *SODA*'05).

[Isi25] Ernst Ising. Beitrag zur theorie des ferromagnetismus. *Zeitschrift für Physik*, 31(1):253–258, 1925.

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

[JS89] Mark Jerrum and Alistair Sinclair. Approximating the permanent. *SIAM J. Comput.*, 18(6):1149–1178, 1989.

[JS93] Mark Jerrum and Alistair Sinclair. Polynomial-time approximation algorithms for the Ising model. *SIAM Journal on Computing*, 22(5):1087–1116, 1993.

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

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

[LSS19] Jingcheng Liu, Alistair Sinclair, and Piyush Srivastava. The Ising partition function: zeros and deterministic approximation. *J. Stat. Phys.*, 174(2):287–315, 2019.

[LY52] T. D. Lee and C. N. Yang. Statistical theory of equations of state and phase transitions. II. Lattice gas and Ising model. *Phys. Rev. (2)*, 87:410–419, 1952.

[MS13] Elchanan Mossel and Allan Sly. Exact thresholds for Ising–Gibbs samplers on general graphs. *The Annals of Probability*, 41(1):294–328, 2013.

[MSW03] Fabio Martinelli, Alistair Sinclair, and Dror Weitz. The ising model on trees: Boundary conditions and mixing time. In *44th Symposium on Foundations of Computer Science (FOCS 2003), 11-14 October 2003, Cambridge, MA, USA, Proceedings*, pages 628–639. IEEE Computer Society, 2003.

[NM53] Gordon F. Newell and Elliott W. Montroll. On the theory of the Ising model of ferromagnetism. *Rev. Modern Physics*, 25:353–389, 1953.

[PR17] Viresh Patel and Guus Regts. Deterministic polynomial-time approximation algorithms for partition functions and graph polynomials. *SIAM J. Comput.*, 46(6):1893–1919, 2017.

[PR19] Han Peters and Guus Regts. On a conjecture of Sokal concerning roots of the independence polynomial. *Michigan Math. J.*, 68(1):33–55, 2019.

[Val08] Leslie G Valiant. Holographic algorithms. *SIAM Journal on Computing*, 37(5):1565–1594, 2008.

[WN13] Christian Wulff-Nilsen. Faster deterministic fully-dynamic graph connectivity. In *Proceedings of the twenty-fourth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1757–1769. SIAM, 2013.

Appendix A. Missing proofs

A.1. **Proof of Theorem 1.1.** Let $\mathcal{A}(G = (V, E), \mathbf{p}, \boldsymbol{\lambda}, \varepsilon)$ be the approximate sampler in Theorem 1.3 for distribution of random cluster model specified by graph G and parameters $\mathbf{p}$ and $\boldsymbol{\lambda}$ within total variation distance ε. The sampler $\mathcal{B}(G = (V, E), \boldsymbol{\beta}, \boldsymbol{\lambda}, \varepsilon)$ for Gibbs distribution of Ising model is defined as follows.

- initialize $X \leftarrow \varnothing$ and let $\mathbf{p} = \left(1 - \beta_i^{-1}\right)_{i \in V}$;
- draw S from $\mathcal{A}(G, \mathbf{p}, \boldsymbol{\lambda}, \varepsilon)$;
- for each $C \in \kappa(V, X)$, add C to X with probability $\frac{\lambda^C}{1 + \lambda^C}$;
- return X.

By a standard coupling argument and Proposition 2.2, X drawn in $\mathcal{B}(G, \boldsymbol{\beta}, \boldsymbol{\lambda}, \varepsilon)$ satisfies

$$d_{TV}(X, \mu_{\boldsymbol{\beta}, \boldsymbol{\lambda}}^{Ising}) \leqslant \varepsilon.$$

This concludes the proof of Theorem 1.1.

A.2. **Proof of Lemma 3.5.** By Lemma 3.6, Lemma 3.3, and (1), it suffices to show

$$\mu_{E, \mathbf{p}, \boldsymbol{\lambda}}^{RC}(E) \geqslant \left(\frac{p_{\min}}{2}\right)^{n^2}.$$

By Proposition 2.1, $Z_{E, \mathbf{p}, \boldsymbol{\lambda}}^{RC} \cdot \boldsymbol{\beta}^E = Z_{\boldsymbol{\beta}, \boldsymbol{\lambda}}^{Ising} \leqslant \boldsymbol{\beta}^E \cdot 2^n$, which implies $Z_{E, \mathbf{p}, \boldsymbol{\lambda}}^{RC} \leqslant 2^n$. Therefore,

$$\mu_{E, \mathbf{p}, \boldsymbol{\lambda}}^{RC}(E) \geqslant \frac{\mathbf{p}^E \left(1 + \boldsymbol{\lambda}^V\right)}{2^n} \geqslant \left(\frac{p_{\min}}{2}\right)^{2^n}.$$

A.3. **Proof of Lemma 3.3.** We give a proof of Lemma 3.3 in this section for completeness. Recall that in Section 3.1, we consider a distribution $\overline{\mu}$ over 2^U on the ground set U. In each round, the field dynamics $\overline{P_\theta^{FD}}$ for $\overline{\mu}$ with parameter $\theta \in (0, 1)$ updates a configuration $X \in 2^U$ as

- sample $S' \sim \bigotimes_{u \in U} \mathrm{Ber}(\theta)$ and let $S = S' \cup X$;
- update X according to distribution $(\theta^{-1} * \overline{\mu})(\cdot \mid \mathcal{P}_{X, U \setminus S})$.

We want to prove that when $\boldsymbol{\lambda} * \overline{\mu}$ is C-spectrally independent under all pinnings for all $\boldsymbol{\lambda} \in \mathbb{R}_{>0}^U$, then for $\theta \in (0, 1)$, and any distribution $\overline{\nu}$ that is absolutely continous with respect to $\overline{\mu}$, we have

(15) $$D_{KL}(\overline{\nu}\, \overline{P_\theta^{FD}} \parallel \overline{\mu}\, \overline{P_\theta^{FD}}) \leqslant (1 - \kappa) D_{KL}(\overline{\nu} \parallel \overline{\mu}),$$

where κ is defined as $\kappa = (\theta/e)^{C+3}$.

Note that the field dynamics we use here is not in its standard form in previous works [CFYZ21b]. In order to be more compatible with previous results, we first transform the setting we use to a more standard version. Specifically, let μ and ν be the distribution over 2^U defined as

$$\forall S \subseteq E, \quad \mu(S) := \overline{\mu}(E \setminus S), \text{ and } \nu(S) := \overline{\nu}(E \setminus S).$$

It is standard to check that ν is absolutely continous with respect to μ and $\boldsymbol{\lambda} * \mu$ is C-spectrally independent under all pinnings for all $\boldsymbol{\lambda} \in \mathbb{R}_{>0}^U$. In each round, the field dynamics P_θ^{FD} with parameter $\theta \in (0, 1)$ falls into its standard form that updates a configuration $X \in 2^U$ as

- sample $S' \sim \bigotimes_{u \in U} \mathrm{Ber}(\theta)$ and let $S = S' \cup (E \setminus X)$;

- update X according to the distribution $(\theta * \mu)(\cdot \mid \mathcal{P}_{X, U \setminus S})$.

Note that P_θ^{FD} is exact $\overline{P_\theta^{FD}}$ except we exchange the role of "in" and "out" for each element in U, and it could be easily checked that

$$\forall X, Y \subseteq E, \quad P_\theta^{FD}(X, Y) = \overline{P_\theta^{FD}}(E \setminus X, E \setminus Y),$$

which implies that

$$\forall X \subseteq E, \quad \mu P^{FD}(X) = \overline{\mu}\, \overline{P_\theta^{FD}}(E \setminus X) \text{ and } \nu P^{FD}(X) = \overline{\nu}\, \overline{P_\theta^{FD}}(E \setminus X).$$

Hence, in order to prove Equation (15), it is sufficient for us to prove

$$(16) \qquad D_{KL}(\nu P_\theta^{FD} \parallel \mu P_\theta^{FD}) \leqslant (1 - \kappa) D_{KL}(\nu \parallel \mu),$$

for $\kappa = (\theta/e)^{C+3}$, under the assumption that $\lambda * \mu$ is C-spectrally independent under all pinnings for all $\lambda \in \mathbb{R}_{>0}^U$ and ν is absolutely continuous with respect to μ.

For convenience, we just inherit the notation that is used in [CFYZ21a]. Suppose μ is a distribution over 2^U over the ground set U, it could also be interpreted as a distribution over $\{-1, +1\}^U$. Let $\theta \in (0, 1)$ and $\pi := \theta * \mu$. Formally, for every $\sigma \in \{-1, +1\}^U$, we have

$$\pi(\sigma) := \frac{\mu(\sigma)\theta^{\|\sigma\|_+}}{Z_\pi},$$

where $Z_\pi := \sum_{\sigma \in \{-1, +1\}^U} \mu(\sigma)\theta^{\|\sigma\|_+}$ and $\|\sigma\|_+$ denotes the number of $+1$ in the vector σ. For a function $f : \{-1, +1\}^U \to \mathbb{R}_{\geqslant 0}$ and a distribution μ over $\{-1, +1\}^U$, the entropy is defined as $\mathbf{Ent}_\mu[f] := \mathbb{E}_\mu[f \log f] - \mathbb{E}_\mu[f] \log \mathbb{E}_\mu[f]$.

Lemma A.1 ([CFYZ21a, Lemma 2.3]). *Let $\theta \in (0, 1)$ be a real number. Let μ be a distribution over $\{-1, +1\}^U$ and $\eta > 0$. If μ is C-spectrally independent under all pinnings for all $\lambda \in \mathbb{R}_{>0}^V$, then μ satisfies the following inequality for any function $f : \{-1, +1\}^U \to \mathbb{R}_{\geqslant 0}$ with $\kappa = \left(\frac{\theta}{e}\right)^{C+3}$,*

$$(17) \qquad \mathbf{Ent}_\mu[f] \leqslant \kappa^{-1} \cdot \frac{Z_\pi}{\theta^{|U|}} \sum_{R \subseteq U} (1 - \theta)^{|R|}\theta^{|U|-|R|} \cdot \pi_R(\mathbf{1}_R) \cdot \mathbf{Ent}_{\pi^{\mathbf{1}_R}}[f],$$

where $\mathbf{1}_R$ is the all-1 vector on R.

Remark A.2. In [CFYZ21a], (17) is called θ-magnetized block factorization of entropy.

Note that the field dynamics could be decomposed into two components that mimics the so-called down-up walk. Let $\Omega := \{\sigma \in \{-1, +1\}^U\}$ and $\omega := \{\mathbf{1}_R \mid R \subseteq U\}$ that is

$$P_\theta^{FD} = P^\downarrow P^\uparrow,$$

where $P^\downarrow \in \mathbb{R}_{\geqslant 0}^{\Omega \times \omega}$ and $P^\uparrow \in \mathbb{R}_{\geqslant 0}^{\omega \times \Omega}$ are defined as follow: $\forall \sigma \in \Omega, \mathbf{1}_R \in \omega$ we have

$$P^\downarrow(\sigma, \mathbf{1}_R) := \mathbf{1}[R \subseteq \sigma^{-1}(+1)](1 - \theta)^{|R|}\theta^{\|\theta\|_+ - |R|}$$

$$\text{and} \quad P^\uparrow(\mathbf{1}_R, \sigma) := \mathbf{1}[R \subseteq \sigma^{-1}(+1)]\pi^{\mathbf{1}_R}(\sigma),$$

where $\sigma^{-1}(+1) := \{u \in U \mid \sigma_u = +1\}$ denotes the set of $+1$-spin elements in U according to σ. Moreover, let $\mu_0 := \mu P^\downarrow$, these two operator have the following adjoint property: $\forall \sigma \in \Omega, \mathbf{1}_R \in \omega$,

$$\mu(\sigma)P^\downarrow(\sigma, \mathbf{1}_R) = \mu_0(\mathbf{1}_R)P^\uparrow(\mathbf{1}_R, \sigma).$$

Without loss of generality, if we assume $R \subseteq \sigma^{-1}(+1)$, then we have

$$
\begin{aligned}
\mu_0(\mathbf{1}_R) P^\uparrow(\mathbf{1}_R, \sigma) &= \sum_{\tau : \tau_R = \mathbf{1}_R} (1-\theta)^{|R|} \theta^{\|\tau\|_+ - |R|} \mu(\tau) \cdot \pi^{\mathbf{1}_R}(\sigma) \\
&= \sum_{\tau : \tau_R = \mathbf{1}_R} (1-\theta)^{|R|} \theta^{\|\tau\|_+ - |R|} \cdot \frac{\mu(\tau)\pi(\sigma)}{\pi_R(\mathbf{1}_R)} \\
&= \sum_{\tau : \tau_R = \mathbf{1}_R} (1-\theta)^{|R|} \theta^{\|\sigma\|_+ - |R|} \cdot \frac{\pi(\tau)\mu(\sigma)}{\pi_R(\mathbf{1}_R)} \\
&= \mu(\sigma)(1-\theta)^{|R|} \theta^{\|\sigma\|_+ - |R|} \sum_{\tau : \tau_R = \mathbf{1}_R} \frac{\pi(\tau)}{\pi_R(\mathbf{1}_R)} \\
&= \mu(\sigma) \cdot (1-\theta)^{|R|} \theta^{\|\sigma\|_+ - |R|} = \mu(\sigma) P^\downarrow(\sigma, \mathbf{1}_R).
\end{aligned}
$$

Let ν be a distribution over $\{-1, +1\}^U$, let $f = \frac{\nu}{\mu}$. By standard result, adjoint property gives us

$$
\text{(18)} \qquad\qquad \frac{\nu P^\downarrow}{\mu P^\downarrow} = P^\uparrow f.
$$

Now, we are ready to prove Lemma 3.3. We claim that when $f = \frac{\nu}{\mu}$, (17) is equivalent to

$$
\text{(19)} \qquad D_{KL}(\nu \parallel \mu) \leqslant \kappa^{-1} \left(D_{KL}(\nu \parallel \mu) - D_{KL}(\nu P^\downarrow \parallel \mu P^\downarrow) \right),
$$

which is equivalent to

$$
D_{KL}(\nu P^\downarrow \parallel \mu P^\downarrow) \leqslant (1-\kappa) D_{KL}(\nu \parallel \mu).
$$

Then, (16) (and Lemma 3.3) follows directly from the data processing inequality as

$$
D_{KL}(\nu P_\theta^{FD} \parallel \mu P_\theta^{FD}) = D_{KL}(\nu P^\downarrow P^\uparrow \parallel \mu P^\downarrow P^\uparrow) \leqslant D_{KL}(\nu P^\downarrow \parallel \mu P^\downarrow) \leqslant (1-\kappa) D_{KL}(\nu \parallel \mu).
$$

Now, we are going to prove (17) $\Leftrightarrow$ (19) by a brute force calculation. First, note that

$$
\text{(20)} \qquad
\begin{aligned}
&\frac{Z_\pi}{\theta^{|U|}} \sum_{R \subseteq U} (1-\theta)^{|R|} \theta^{|U|-|R|} \cdot \pi_R(\mathbf{1}_R) \cdot \mathbf{Ent}_{\pi^{\mathbf{1}_R}}[f] \\
&= \frac{Z_\pi}{\theta^{|U|}} \sum_{R \subseteq U} (1-\theta)^{|R|} \theta^{|U|-|R|} \pi_R(\mathbf{1}_R) \mathbb{E}_{\pi^{\mathbf{1}_R}}[f \log f] \quad -
\end{aligned}
$$

$$
\text{(21)} \qquad \frac{Z_\pi}{\theta^{|U|}} \sum_{R \subseteq U} (1-\theta)^{|R|} \theta^{|U|-|R|} \pi_R(\mathbf{1}_R) \mathbb{E}_{\pi^{\mathbf{1}_R}}[f] \log \mathbb{E}_{\pi^{\mathbf{1}_R}}[f].
$$

We will show that

$$
\text{(20)} = D_{KL}(\nu \parallel \mu) = \mathbf{Ent}_\mu[f] \quad \text{and} \quad \text{(21)} = D_{KL}(\nu P^\downarrow \parallel \mu P^\downarrow).
$$

We start from (20) and note that

$$
(20) = \frac{Z_\pi}{\theta^{|U|}} \sum_{R \subseteq U} (1-\theta)^{|R|} \theta^{|U|-|R|} \sum_{\sigma : \sigma_R = \mathbf{1}_R} \pi(\sigma) f(\sigma) \log f(\sigma)
$$

$$
= \frac{Z_\pi}{\theta^{|U|}} \sum_{\sigma \in \Omega} \pi(\sigma) \sum_{R \subseteq \sigma^{-1}(+1)} (1-\theta)^{|R|} \theta^{|U|-|R|} \cdot f(\sigma) \log f(\sigma)
$$

$$
= \sum_{\sigma \in \Omega} \mu(\sigma) \sum_{R \subseteq \sigma^{-1}(+1)} (1-\theta)^{|R|} \theta^{\|\sigma\|_+ - |R|} \cdot f(\sigma) \log f(\sigma)
$$

$$
= \sum_{\sigma \in \Omega} \mu(\sigma) \cdot f(\sigma) \log f(\sigma) \sum_{R \subseteq \sigma^{-1}(+1)} (1-\theta)^{|R|} \theta^{\|\sigma\|_+ - |R|}.
$$

$$
= \sum_{\sigma \in \Omega} \mu(\sigma) \cdot f(\sigma) \log f(\sigma) = \mathbf{Ent}_\mu [f].
$$

Now, we only left to prove $(21) = D_{KL}(\nu P^\downarrow \parallel \mu P^\downarrow)$. By the definition of KL-divergence, we have

$$
D_{KL}(\nu P^\downarrow \parallel \mu P^\downarrow) = \sum_{R \subseteq U} \mu P^\downarrow(\mathbf{1}_R) \cdot \frac{\nu P^\downarrow(\mathbf{1}_R)}{\mu P^\downarrow(\mathbf{1}_R)} \log \frac{\nu P^\downarrow(\mathbf{1}_R)}{\mu P^\downarrow(\mathbf{1}_R)}
$$

$$
\text{(by (18))} \quad = \sum_{R \subseteq U} \mu P^\downarrow(\mathbf{1}_R) \cdot P^\uparrow f(\mathbf{1}_R) \log P^\uparrow f(\mathbf{1}_R)
$$

$$
\stackrel{(\star)}{=} \sum_{R \subseteq U} \sum_{\sigma : \sigma_R = \mathbf{1}_R} \mu(\sigma)(1-\theta)^{|R|} \theta^{\|\sigma\|_+ - |R|} \cdot \mathbb{E}_{\pi^{\mathbf{1}_R}} [f] \log \mathbb{E}_{\pi^{\mathbf{1}_R}} [f]
$$

$$
= \frac{Z_\pi}{\theta^{|U|}} \sum_{R \subseteq U} \sum_{\sigma : \sigma_R = \mathbf{1}_R} \pi(\sigma)(1-\theta)^{|R|} \theta^{|U|-|R|} \cdot \mathbb{E}_{\pi^{\mathbf{1}_R}} [f] \log \mathbb{E}_{\pi^{\mathbf{1}_R}} [f]
$$

$$
= \frac{Z_\pi}{\theta^{|U|}} \sum_{R \subseteq U} (1-\theta)^{|R|} \theta^{|U|-|R|} \left(\sum_{\sigma : \sigma_R = \mathbf{1}_R} \pi(\sigma) \right) \cdot \mathbb{E}_{\pi^{\mathbf{1}_R}} [f] \log \mathbb{E}_{\pi^{\mathbf{1}_R}} [f]
$$

$$
= \frac{Z_\pi}{\theta^{|U|}} \sum_{R \subseteq U} (1-\theta)^{|R|} \theta^{|U|-|R|} \pi_R(\mathbf{1}_R) \cdot \mathbb{E}_{\pi^{\mathbf{1}_R}} [f] \log \mathbb{E}_{\pi^{\mathbf{1}_R}} [f] = (21),
$$

where in $(\star)$ we use the fact that

$$
\mu P^\downarrow(\mathbf{1}_R) = \sum_{\sigma : \sigma_R = \mathbf{1}_R} \mu(\sigma)(1-\theta)^{|R|} \theta^{\|\sigma\|_+ - |R|} \quad \text{and} \quad P^\uparrow f(\mathbf{1}_R) = \mathbb{E}_{\pi^{\mathbf{1}_R}} [f].
$$

Appendix B. Proof for the second part of Lemma 5.1

Let $(X_t)_{t \geqslant 0}$ and $(Y_t)_{t \geqslant 0}$ be Glauber dynamics starting from E and the stationary distribution μ respectively. Just as in the proof of the first part, we present the following lemma. The proof of Lemma B.1 follows a similar fashion to that of Lemma 5.2, and we shall omit the proof for brevity.

Lemma B.1. *For any* $t \geqslant 0$, *it holds that*

$$
\mathbf{Pr}[X_t \notin \mathcal{G}] \leqslant n^{\log(27K)} \text{ and } \mathbf{Pr}[Y_t \notin \mathcal{G}] \leqslant n^{\log(27K)}.
$$

Proof for the second part of Lemma 5.1. Let $T = 2m(\log m + \log(2/\varepsilon))$ and $\mathcal{E}_t$ be the event that $X_t \notin \mathcal{G}$ or $Y_t \notin \mathcal{G}$ happens. By Lemma B.1 and union bound,

$$
\forall 0 \leqslant t \leqslant T-1, \quad \mathbf{Pr}\left[\overline{\mathcal{E}_t}\right] \leqslant 2n^{\log(27K)}.
$$

When $K \leqslant 10^{-5} \exp\left(-\frac{\log(8/\varepsilon)}{\log n}\right)$, it holds that

$$2n^{\log(27K)} \cdot 2m^2 \leqslant 4n^{4+\log(27K)} \leqslant \frac{\varepsilon}{2}.$$

Together with Lemma 5.3 and Lemma 5.4, we have

$$\mathbf{Pr}\left[X_T \neq Y_T\right] \leqslant \left(1 - \frac{1}{2m}\right)^T + \frac{\varepsilon}{2} \leqslant \varepsilon. \qquad \square$$