RAPID MIXING IN POSITIVELY WEIGHTED RESTRICTED BOLTZMANN MACHINES

WEIMING FENG, HENG GUO, MINJI YANG

ABSTRACT. We show polylogarithmic mixing time bounds for the alternating-scan sampler for positively weighted restricted Boltzmann machines. This is done via analysing the same chain and the Glauber dynamics for ferromagnetic two-spin systems, where we obtain new mixing time bounds up to the critical thresholds.

1. INTRODUCTION

The restricted Boltzmann machine (RBM) [AHS85, Smo86] is a popular model to represent many different types of data [HOT06, SMH07, MH10]. Its simple two-layer structure also makes it useful as a basic building block for deep belief networks [HOT06]. The development of RBMs is recognised as a main contribution for Geoffrey E. Hinton's Nobel prize in physics in 2024 [nob24]. As it would distract from the main focus of our paper, we do not attempt to give a comprehensive overview of RBMs here.

The training of RBMs relies on estimating the gradient, which is often done via the MCMC method. One of the most popular Markov chains here is the alternating-scan sampler [Hin02], which updates the two layers of the variables alternately conditioned on the other layer. The mixing time of this sampler (namely the time it takes to converge to its stationary distribution) is very important in learning RBMs, as emphasised in Hinton's practical guide [Hin12].

Despite RBMs' popularity, rigorous mixing time bounds of the alternating-scan sampler are rather sparse. The only available results require either bounded interaction strengths [Tos16] or special structures [KQWW26]. The lack of good bounds is perhaps for a good reason. Via an equivalent formulation of anti-ferromagnetic two-spin systems, when parameters cross the critical threshold, the mixing time in negatively weighted RBMs is exponentially large, and in fact, in this case sampling and approximate counting are **NP**-hard [Sly10, SS14, GŠV16]. On the other hand, the contrastive divergence method [Hin12] in practise typically runs the alternating-scan sampler for a constant number of rounds. In this paper, we show a polylogarithmic mixing time bound for the alternating-scan sampler on positively weighted RBMs, bypassing the bounded interaction strengths requirement and complementing the hardness for the negative weight case.

Next we introduce our main result more precisely. A *Boltzmann machine* [AHS85] with a set V of variables of size n is specified by an n-by-n symmetric interaction matrix $W = \{w_{uv}\}_{u,v\in V}$ and variable weights $\theta = \{\theta_v\}_{v\in V}$. A configuration $\sigma : V \to \{0,1\}$ is associated with the Hamiltonian or the energy function:

$$E(\sigma) := \sum_{u,v\in V} w_{uv}\sigma_u\sigma_v + \sum_{v\in V} \theta_v\sigma_v. \tag{1}$$

Without loss of generality we may assume that the diagonal entries of W are all 0. The Gibbs distribution μ is defined as $\mu(\sigma) = \frac{e^{E(\sigma)}}{Z}$, where $Z := \sum_{\sigma\in\{0,1\}^V} e^{E(\sigma)}$ is the normalizing constant, namely the partition function.

A *restricted Boltzmann machine* (RBM) [Smo86] is one where the variables can be partitioned into two parts $V = V_0 \uplus V_1$ (the visible and the hidden layers) such that $w_{uv} = 0$ whenever $u \in V_0$ and $v \in V_1$.

(Heng Guo) SCHOOL OF INFORMATICS, UNIVERSITY OF EDINBURGH, INFORMATICS FORUM, UK
(Weiming Feng and Minji Yang) SCHOOL OF COMPUTING AND DATA SCIENCE, THE UNIVERSITY OF HONG KONG, HK
E-mail address: wfeng@hku.hk, hguo@inf.ed.ac.uk, ymjessen02@connect.hku.hk.

We may also view an RBM as over a bipartite graph where the edge set E represents nonzero interaction weights.

A popular algorithm to sample from RBMs is the aforementioned alternating-scan sampler [Hin12], which is a systematic scan variant of the Gibbs sampler where we scan the two partitions in order. Starting from an arbitrary configuration $X \in \{0,1\}^n$. For any $t \geqslant 1$, in the t-th step, it updates the current configuration X as follows: pick the part V_i with index $i = (t \mod 2)$ and resample the configuration on V_i conditional on the current configuration of the other part V_{1-i}. More formally, at step t,

(1) pick the part V_i with index $i = (t \mod 2)$;
(2) resample $X_{V_i} \sim \mu_{V_i}^{X(V_{1-i})}$, where $\mu_{V_i}^{X(V_{1-i})}$ is the marginal distribution of all variables in the part V_i induced by μ conditioned on the configuration $X(V_{1-i})$ on the other part V_{1-i};

The mixing time of a Markov chain is defined as the number of steps until the configuration X is close to the stationary distribution μ in total variation distance. Formally, let P be the transition matrix of the Markov chain. Then, the mixing time is defined as

$$\forall \epsilon > 0, \quad t_{\text{mix}}(\epsilon) = \max_{X_0 \in \{0,1\}^V} \min \left\{ t \geqslant 0 : D_{\text{TV}}\left(P^t(X_0, \cdot), \mu\right) < \epsilon \right\}, \tag{2}$$

where $D_{\text{TV}}(\nu, \mu) = \frac{1}{2}\sum_{x \in \{0,1\}^V} |\nu(x) - \mu(x)|$ denotes the total variation distance and X_0 is called the starting configuration or state.

Now we can state our main result.

Theorem 1. *Let $c > 0$ be an arbitrary constant. For any restricted Boltzmann machine (W, θ) with n variables, if for all u, v, either $w_{uv} \geqslant c$ or $w_{uv} = 0$, and for all $v \in V$, $\theta_v \geqslant 0$, then the alternating-scan sampler over the Gibbs distribution μ of the RBM has mixing time at most $O((\log n)^C \log \frac{1}{\epsilon})$, where $C = C(c) > 0$ is a constant depending on c.*

We note that the lower bound $c > 0$ is to avoid cases where, for example, some $w_{uv} = 1/n$. Certain technical conditions we rely on would break in such a case. We believe that this is an artifact of our proof, and the theorem should hold with $c = 0$. On the other hand, the main strength of Theorem 1 is that we do not need to assume any upper bound on w_{uv}'s.

Previously, Tosh [Tos16] showed that the alternating-scan sampler mixes in logarithmic time when $\|W\|_1 \|W^t\|_1 < 4$ via a one-step coupling, where $\|\cdot\|_1$ denotes the 1-norm of matrices. Kwon, Qin, Wang, and Wei [KQWW26] considered the setting where $W_{uv} = c/n$ for any $u \in V_0$ and $v \in V_1$ for some c. They obtained logarithmic mixing time as long as $c > -5.87$ via a drift and contraction coupling technique. In contrast, Theorem 1 does not have any upper bound on the interaction strength or assumption on the structure, and the proof technique is a significant departure from these two results.

Alternatively, rigorous efficient algorithms for positively weighted RBMs can be obtained via an equivalent formulation of ferromagnetic two-spin systems [GJP03, LLZ14, GL18, GLL20]. Theorem 1 is also proved via this connection, so we will explain it next.

1.1. Ferromagnetic two-spin systems.

Boltzmann machines are a special case of the so-called two-spin systems. Let $G = (V, E)$ be a graph. For each edge $e \in E$, let $\beta_e, \gamma_e > 0$ be the edge activity at e. For each vertex $v \in V$, let $\lambda_v \leqslant 1$ be the external field at v. A two-spin system $(G, (\beta_e, \gamma_e)_{e \in E}, (\lambda_v)_{v \in V})$ defines a Gibbs distribution μ over $\Omega = \{0,1\}^V$ such that

$$\forall \sigma \in \{0,1\}^V, \quad \mu(\sigma) \propto \prod_{v \in V : \sigma_v = 0} \lambda_v \prod_{\{u,v\} \in E : \sigma_u = \sigma_v = 0} \beta_e \prod_{\{u,v\} \in E : \sigma_u = \sigma_v = 1} \gamma_e.$$

A two-spin system is said to be ferromagnetic if $\beta_e \gamma_e \geqslant 1$ for all $e \in E$.

Positively weighted Boltzmann machines with parameters (W, θ) can be viewed as ferromagnetic two-spin systems over the complete graph via the following reparameterisation:

$$\forall v \in V,\ \lambda_v = \exp(-\theta_v) \text{ and } \forall u, v \in V,\ \beta_{uv} = 1, \gamma_{uv} = \exp(w_{uv}).$$

We may also remove edges with zero weights. This way, restricted Boltzmann machines become ferromagnetic two-spin systems defined over bipartite graphs.

We mainly consider families of ferromagnetic two-spin systems given as follows.

Definition 2 ((β, γ, λ)-ferromagnetic two-spin systems)**.** Let $\beta \leqslant 1 < \gamma$, $\beta\gamma > 1$ and $\lambda > 0$ be three constants. A ferromagnetic two-spin system $(G, (\beta_e, \gamma_e)_{e \in E}, (\lambda_v)_{v \in V})$ is said to be a (β, γ, λ)-ferromagnetic two-spin system if $\lambda_v < \lambda$ for all $v \in V$ and $\beta_e \leqslant \beta \leqslant 1 < \gamma \leqslant \gamma_e$, $\beta\gamma \geqslant \beta_e\gamma_e > 1$ for all $e \in E$.

Restricted Boltzmann machines in Theorem 1 are special cases of ferromagnetic two-spin systems in Definition 2 over bipartite graphs with $\beta = 1$, $\gamma = \exp(c)$, and $\lambda = 1 + \epsilon$ for an arbitrarily small $\epsilon > 0$. It is important that here β, γ, and λ are all constants, and we do not need to assume any of the β_e, γ_e, or λ_v to be constants. Theorem 1 is in fact implied by the following more general result for the mixing time of the alternating-scan sampler on ferromagnetic two-spin systems in bipartite graphs.

Theorem 3. *Let $\beta \leqslant 1 < \gamma$, $\beta\gamma > 1$ and $\lambda < \lambda_0(\beta, \gamma) := \sqrt{\gamma/\beta}$ be three constants. For any (β, γ, λ)-ferromagnetic two-spin system over a bipartite graph with n vertices, the alternating-scan sampler on the Gibbs distribution has mixing time at most $O((\log n)^C \log \frac{1}{\epsilon})$, where $C = C(\beta, \gamma, \lambda) > 0$ is a constant depending on (β, γ, λ).*

In addition to the alternating-scan sampler, we also analyse Glauber dynamics, which is another fundamental Markov chain to sample from Gibbs distributions. Starting from an arbitrary configuration $X \in \{0, 1\}^V$, in each step, the Glauber dynamics updates the current configuration X as follows:

- pick a vertex v uniformly at random from V;
- resample $X_v \sim \mu_v^{X(V \setminus \{v\})}$, where $\mu_v^{X(V \setminus \{v\})}$ is the marginal distribution on v induced by μ conditioned on the configuration $X(V \setminus \{v\})$ on other variables $V \setminus \{v\}$;

We show that under the same conditions as in Theorem 3, Glauber dynamics mixes in near-linear time.

Theorem 4. *Let $\beta, \gamma, \lambda > 0$ be three constants such that $\beta \leqslant 1 < \gamma$, $\beta\gamma > 1$, and $\lambda < \lambda_0 := \sqrt{\gamma/\beta}$. For any (β, γ, λ)-ferromagnetic two-spin system with n vertices, the Glauber dynamics on the Gibbs distribution has mixing time at most $n(\log n)^C \log \frac{1}{\epsilon}$, where $C = C(\beta, \gamma, \lambda) > 0$ is a constant depending on (β, γ, λ).*

The same threshold $\lambda_0 = \sqrt{\gamma/\beta}$ also appeared in [GJP03], where the authors showed that for ferromagnetic two-spin systems with uniform parameters $\lambda_v = \lambda$, $\beta_e = \beta$, and $\gamma_e = \gamma$, there exists a polynomial-time sampling algorithm if $\beta\gamma > 1$ and $\lambda \leqslant \lambda_0$. The condition for a polynomial-time sampling algorithm was later shown to be $\lambda \leqslant \gamma/\beta$ by [LLZ14]. Both algorithms are obtained by reducing the problem of sampling from ferromagnetic two-spin systems to that of sampling from a *ferromagnetic Ising model* with consistent external fields. Specifically, the resulting Ising model is a two-spin system $(G, (\beta_e, \gamma_e)_{e \in E}, (\lambda_v)_{v \in V})$ such that every edge has interaction strength $\beta_e = \gamma_e = \sqrt{\beta\gamma} > 1$ and every vertex has external field $\lambda_v \leqslant 1$. Jerrum and Sinclair [JS93] gave the first polynomial-time sampling algorithm to this Ising model. After the reduction in [LLZ14], there is a constant gap between the external field λ and 1. In this case, the best sampling algorithm runs in near-linear time as well [CZ23], via yet another connection [ES88, GJ18, FGW23] to the *random cluster model* [FK72].

Our results in Theorem 3 and Theorem 4 are the first near-optimal mixing results for the alternating-scan sampler and Glauber dynamics on ferromagnetic two-spin systems with $\lambda < \lambda_0$, whereas all previous algorithms rely on a reduction to sampling from other models. From a technical perspective, our approach is completely different from the reduction technique used in [GJP03, LLZ14]. We develop a unified framework for analyzing the mixing of a family of *heat-bath* and *systematic scan* dynamics on ferromagnetic two-spin systems, which covers the alternating-scan sampler and Glauber dynamics as special cases. We give a proof overview in Section 2.

Another advantage of the direct mixing time bound in Theorem 4 is that, unlike previous results, it allows us to extend our mixing time analysis beyond λ_0 to a larger threshold

$$\lambda_c(\beta,\gamma) := (\gamma/\beta)^{\frac{\sqrt{\beta\gamma}}{\sqrt{\beta\gamma}-1}} \geqslant \lambda_0(\beta,\gamma),$$

which was previously identified as the potentially critical threshold for ferromagnetic 2-spin systems [GL18].[1] In fact, Guo and Lu [GL18] designed efficient sampling and approximate counting algorithms for ferromagnetic 2-spin systems below this threshold via correlation decay [Wei06]. However, their algorithms run in time $O(n^C)$ where C is a large constant depending on (β,γ,λ). Later, Guo, Liu, and Lu [GLL20] designed another algorithm based on the zeros of polynomials method [Bar16, PR17], which works for all β,γ such that $\beta\gamma > 1$ but with a lower threshold for λ[2] and requires bounded degree graphs. In any case, it has a similar $O(n^C)$ running time. Our next result improves the exponent in the running times for sampling and approximate counting to absolute constants. For sampling, our time bound is $\widetilde{O}(n^2)$.

Theorem 5. *Let $\beta,\gamma,\lambda > 0$ be three constants such that $\beta \leqslant 1 < \gamma$, $\beta\gamma > 1$, and $\lambda < \lambda_c := (\gamma/\beta)^{\frac{\sqrt{\beta\gamma}}{\sqrt{\beta\gamma}-1}}$. There exists a constant $C = C(\beta,\gamma,\lambda) > 0$ such that for any (β,γ,λ)-ferromagnetic two-spin system with n vertices, the Glauber dynamics on the Gibbs distribution μ has mixing time*

- *at most $n^2 \cdot (\log n)^C \cdot \log \frac{1}{\epsilon}$ starting from the all-1 configuration;*
- *at most $n^3 \cdot (\log n)^C \cdot \log \frac{1}{\epsilon}$ starting from an arbitrary configuration.*

Remark. In the proof of Theorem 5, we show that the spectral gap of the Glauber dynamics is $\mathsf{gap} = \widetilde{\Omega}(n^{-1})$, which is optimal. Since $\mu(\mathbf{1}) = 2^{-\Omega(n)}$ for the all-1 configuration, this yields the upper bound $O\left(\frac{1}{\mathsf{gap}} \log \frac{1}{\epsilon\,\mu(\mathbf{1})}\right)$ on the mixing time from the all-1 starting configuration. For the mixing time from an arbitrary starting configuration, the standard approach is to use the bound $O\left(\frac{1}{\mathsf{gap}} \log \frac{1}{\epsilon\,\mu_{\min}}\right)$, where $\mu_{\min} = \min_{x\in\{0,1\}^V} \mu(x)$. However, for spin systems in Definition 2, the parameters λ_v and β_e may be arbitrarily small, while γ_e may be arbitrarily large, resulting in a potentially arbitrarily small $\mu_{\min}$. We resolve this issue by showing that the Glauber dynamics quickly reaches a warm-start configuration with high probability, and then bounding the mixing time from such a warm-start configuration (see Lemma 55). We remark that even if all parameters $\lambda_v, \beta_e, \gamma_e$ are assumed to be constants, $\mu_{\min}$ can still be as small as $\exp(-O(n^2))$. The reason is that the graph can be very dense and contain an $\Omega(n^2)$ number of edges.

Our result is also the first polynomial mixing time bound for Glauber dynamics on ferromagnetic two-spin systems with $\lambda < \lambda_c$ in general graphs. All previous polynomial mixing time results work only on *bounded degree* graphs. Let Δ denote the maximum degree of the graph. Chen, Liu, and Vigoda first proved $n^{e^{O(\Delta)}}$ mixing time bound for Glauber dynamics [CLV23b], and later they improved the bound to $e^{e^{O(\Delta)}} n \log n$ [CLV23a]. In contrast, our Theorem 5 does not depend on Δ.

Theorem 5 is proved by combining Theorem 4 with a mixing time boosting technique developed by Chen, Feng, Yin, and Zhang [CFYZ21]. Roughly speaking, by verifying a certain spectral independence condition [ALO24] for ferromagnetic two-spin systems when $\lambda < \lambda_c$, we can reduce the analysis to the case $\lambda < \lambda_0$, which is handled by Theorem 4. It is important that Theorem 4 provides a direct mixing time bound rather than a reduction based sampling algorithm as in [GJP03, LLZ14]; otherwise, the mixing time boosting technique would not be applicable. The detailed proof is given in Section 9.3.

As mentioned before, we believe that the lower bound $c > 0$ requirement can be removed in Theorem 1, but some new ideas are required to handle the case where, for example, some $w_{uv} = 1/n$. Another interesting open problem is to prove a near-optimal $\widetilde{O}(n)$ mixing time bound for ferromagnetic

[1] Roughly speaking, up to an integral gap, systems above this threshold are #**BIS**-hard [LLZ14], where #**BIS** is conjectured to be computationally hard [DGGJ04].

[2] Their threshold is roughly $\sqrt{\lambda_c}$.

two-spin systems when $\lambda < \lambda_c$. Due to technical obstacles (see Section 2), we cannot directly extend the analysis of Theorem 4 to this regime. A possible alternative is to use the refined mixing-time boosting techniques developed in [CFYZ22, CE22, FY26]. However, this approach requires a stronger *entropic independence* [AJK+22] condition, which is not known to hold for the class of ferromagnetic two-spin systems studied here. More broadly, our proof framework applies to general ferromagnetic two-spin systems, for which there is still a big gap between the known algorithmic [GL18, GLL20, SS21] and the hardness threshold [LLZ14], especially when $\beta, \gamma > 1$. In that case, worst-case correlation decay results, such as those in [GL18], no longer hold. We believe that our "typical-case" SSM (more detail in Section 2) is the first step on the right direction.

2. Proof overview

We give a proof overview for the mixing time of Glauber dynamics on ferromagnetic two-spin systems. For the simplicity of the overview, consider a ferromagnetic two-spin system μ defined on a graph $G = (V, E)$ with unified parameters, where $\lambda_v = \lambda$ for all $v \in V$ and $\beta_e = \beta, \gamma_e = \gamma$ for all $e \in E$ for constants λ, β, γ. We outline the proof of $n \cdot \mathrm{polylog}(n)$ mixing time bound in Theorem 4 when $\lambda < \lambda_0 = \sqrt{\gamma/\beta}$. Other results can be proved as follows.

- The proof technique of Theorem 4 can be generalized to the alternating-scan sampler in Theorem 3.
- The mixing result in Theorem 5 when $\lambda < \lambda_c$ can be proved by combining the mixing result in Theorem 4 with the existing results in [CFYZ21, FY26].

2.1. **All-to-one influence bound.** Let μ over $\{0,1\}^V$ be a Gibbs distribution defined on variable set V. For any two variables $u, v \in V$, the influence from u on v is defined as

$$\Psi(u, v) := \Pr[X_v = 1 \mid X_u = 1] - \Pr[X_v = 1 \mid X_u = 0].$$

Anari, Liu, and Oveis Gharan [ALO24] showed that if the maximum eigenvalue of the influence matrix Ψ is bounded by a constant, then the Glauber dynamics mixes in polynomial time. The maximum eigenvalue of the influence matrix is bounded by the *all-to-one influence* $\max_{v \in V} \sum_{u \in V} \Psi(u, v)$. We show in Theorem 19 that if $\lambda < \lambda_c$, then the all-to-one influence is $O(1)$. The proof is inspired by the analysis in [ALO24], where they analysed the all-to-one influence of the hardcore model in the uniqueness regime. Here, we need to deal with the ferromagnetic spin system in general graph with possibly unbounded degree. We use the correlation decay technique developed in [GL18] to prove the bound.

The all-to-one influence only gives an $n^{O(C)}$ mixing time bound, where the influence bound C can be a very large constant. However, this is still useful in getting the local mixing bounds we need later. To obtain our $n \cdot \mathrm{polylog}(n)$ mixing result in general graphs, we use a local mixing to global mixing argument based on the *aggregate strong spatial mixing* (*ASSM*) property.

2.2. **Mixing from typical-case ASSM.** A ferromagnetic two-spin system is a monotone system. Mossel and Sly [MS13] showed that the mixing of Glauber dynamics on monotone systems can be proved via the ASSM property. Let $v \in V$ and $S_v \subseteq V$ a subset of vertices containing v. Let ∂S_v be the outer *boundary* of S_v, which is the set of vertices not in S_v but adjacent to S_v. Define the *influence* of u on v by

$$\widehat{a}_u := \max_{\sigma \in \{0,1\}^{\partial S_v}} D_{\mathrm{TV}}\left(\mu_v^{\sigma^{u \leftarrow 0}}, \mu_v^{\sigma^{u \leftarrow 1}}\right), \tag{3}$$

where $\sigma^{u \leftarrow c}$ denotes the configuration on ∂S_v obtained from σ by changing the value of u to c. Mossel and Sly showed that if the ASSM property $\sum_{u \in \partial S_v} \widehat{a}_u \leqslant \frac{1}{20}$ holds and the mixing time of Glauber dynamics on the conditional distribution $\mu_{S_v}^{\sigma}$ is at most T_{local} for any $\sigma \in \{0,1\}^{\partial S_v}$, then the mixing time of Glauber dynamics on μ is at most $O(T_{\mathrm{local}} \cdot n \log n \cdot \max_{v \in V} \log |S_v \cup \partial S_v|)$. Their result works for ferromagnetic two-spin systems on graphs with *bounded degrees.* For the Ising model in the uniqueness regime, the ASSM property can be verified if the region S_v is a ball centered at v with radius $\ell_0 = O(1)$ [MS13]. Since the degrees are bounded, $|S_v \cup \partial S_v|$ is a constant, implying that $T_{\text{local}} = O(1)$. The overall mixing time of the Glauber dynamics on μ is $O(n \log n)$.

However, what we consider are general graphs with possibly unbounded degrees. Consider a star centered at v. If we choose $S_v = \{v\}$, then ASSM does not necessarily hold. If we choose S_v as a ball centered at v with radius 1, the resulting S_v is the whole V, and bounding local mixing T_{local} is the same as bounding the mixing time of Glauber dynamics on μ. To resolve these issues, we introduce a weaker version of the ASSM property. For each vertex $v \in V$, we algorithmically choose a region S_v and also define a set of good boundary configurations $\Omega_{\partial S_v} \subseteq \{0,1\}^{\partial S_v}$. The specific choice of S_v and $\Omega_{\partial S_v}$ will be given in later. We define a new influence bound a_u as

$$a_u := \max_{\sigma \in \Omega_{\partial S_v}} D_{TV}\left(\mu_v^{\sigma^{u \leftarrow 0}}, \mu_v^{\sigma^{u \leftarrow 1}}\right),$$

Compared with (3), the new influence considers only "typical" boundary conditions, namely those from $\Omega_{\partial S_v}$, on ∂S_v. Using the monotone coupling technique, we show that the mixing time of Glauber dynamics on μ is at most $O(T_{\text{burn-in}} + T_{\text{local}} \cdot n \log n \cdot \max_{v \in V} \log |S_v \cup \partial S_v|) = O(T_{\text{burn-in}} + T_{\text{local}} \cdot n \log^2 n)$ as long as the following conditions holds for two parameters $T_{\text{burn-in}}$ and T_{local}.

- For the Glauber dynamics $(X_t)_{t \geqslant 0}$ on μ, starting from an arbitrary $X_0 \in \{0,1\}^V$, for any $t \geqslant T_{\text{burn-in}}$, any $v \in V$, with probability at least $1 - \frac{1}{\text{poly}(n)}$, it holds that $X_t(\partial S_v) \in \Omega_{\partial S_v}$.
- ASSM holds for typical boundary conditions: $\sum_{u \in \partial S_v} a_u \leqslant \frac{1}{20}$ for all $v \in V$.
- For any vertex $v \in V$ and any $\sigma \in \{0,1\}^{\partial S_v}$, the Glauber dynamics on $\mu_{S_v}^{\sigma}$ has mixing time T_{local}.

Compared with the result of Mossel and Sly, the key advantage is that we only require the ASSM property to hold under a typical-boundary condition after burn-in, while they require the ASSM property to hold for worst-case boundary conditions. For the star graph centered at v, we can simply take $S_v = \{v\}$ and let $\Omega_{\partial S_v}$ contains all configurations on neighbors of v such that at least a constant fraction of them are assigned 1. Note that the parameter setting is $\beta \leqslant 1 < \gamma$. If $\Omega(n)$ neighbors of v are in state 1, then since $\gamma > 1$, the value on v is almost fixed to be 1, so we can bound the sum of the influences. However, in the original definition of Mossel and Sly, the maximum influence for w to v is achieved when all other vertices are 0. In this case, if $\beta = 1$, then each $a_u = \Omega(1)$, so the total influence is $\Omega(n)$.

2.3. **Typical-case ASSM for ferro spin systems.** To carry out the ideas in the previous section, we need to carefully choose the region S_v and the set of good boundary configurations $\Omega_{\partial S_v}$ so that all the above conditions hold, which is the most technical part of the proof. We will guarantee that $|S_v| = \text{polylog}(n)$. Then, for the local mixing bound T_{local}, the conditionally distribution is defined on $N = \text{polylog}(n)$ vertices. Using the all-to-one influence bound and the result in [ALO24], we have $T_{\text{local}} \leqslant N^{O(1)} = \text{polylog}(n)$.

We next give a detailed construction of the region S_v. To illustrate the idea, let us first focus on a special case when the graph G is a tree. We run a DFS starting from the root v. Suppose the DFS procedure visits a vertex w. We first add w into the region S_v. Next, let $u_0 = v, u_1, \ldots, u_k = w$ be the path from v to w in the tree. For each u_i, let d_i denote the number of children of u_i in the tree rooted at v.

- If $\sum_{i=1}^{k} d_i < D_1 = O(\log \log n)$, we recursively do the DFS on all children of w;
- If $\sum_{i=1}^{k} d_i \geqslant D_1$, we will *not* recursively do the DFS on any child of w. Instead, if the number of children of w is less than $D_2 = (\log n)^3$, we add all these children into the region S_v and terminate the exploration in this branch. Otherwise, we stop at w.

Overall, the DFS procedure will construct a region S_v, where the induced subgraph $T_{S_v} = G[S_v]$ is a subtree rooted at v. For any vertex $w \in S_v$, in the subtree $G[S_v]$, we can upper bound the degree sum of all vertices on the path from v to w. Using this property, we can show that $|S_v| = \text{polylog}(n)$.

Let ∂S_v be the outer boundary of S_v. Define $\Omega_{\partial S_v}$ as the set of all boundary configurations $\sigma \in \{0,1\}^{\partial S_v}$ such that for any vertex $w \in S_v$ with K neighbors in ∂S_v, if $K \geqslant D_2/3$, then at least $K/\log n = \Omega((\log n)^2)$ neighbors are assigned 1 in σ. In other words, if w has many neighbors in ∂S_v, then a significant proportion of them are assigned 1. Since $\beta \leqslant 1$, when Glauber dynamics updates a vertex u, with a constant probability, the value on u is updated to 1. After running Glauber dynamics for $T_{\text{burn-in}} = O(n \log n)$ steps, a simple coupon collector and Chernoff bound argument shows that with high probability, the configuration on ∂S_v is in $\Omega_{\partial S_v}$.

We next bound the sum $\sum_{u \in \partial S_v} a_u$. Fix a vertex $u \in \partial S_v$. We first explain why a single influence a_u is small. Then we give some high level ideas on how to bound the sum of the influences. To analyze the influence a_u, we need to consider a spin system with pinnings defined on the induced subgraph $G[S_v \cup \partial S_v]$. Using the self-reducibility property of ferromagnetic two-spin systems, we can remove the pinning and analyze a spin system on $T' = G[S_v \cup \{u\}]$ with some effective external fields on the inner boundary of S_v. Then, a_u is the one-to-one influence from u to v in T'. Guo and Lu [GL18] showed the following *computationally efficient* correlation decay result. Let $v_0 = v, v_1, \dots, v_k = u$ be the path from v to u in the tree T'. Let d_i' be the number of children of v_i in T'. Then

$$a_u \leqslant C_1 \exp\left(-\sum_{i=1}^{k-1} d_i'/C_2\right),$$

for some sufficiently large constants $C_1, C_2 > 0$. Let $d_1, d_2, \dots, d_k$ be the number of children of v_i in the tree G rooted at v. By the definition of $T' = G[S_v \cup \{u\}]$ and the construction of S_v, we have $d_i' = d_i$ for $1 \leqslant i \leqslant k-2$ and $d_{k-1}' = 1$. Depending on how v_{k-1} is added to S_v, there are two cases.

- The vertex v_{k-1} is added to S_v because the DFS stops at the vertex v_{k-2} and v_{k-2} has less than D_2 children. However, stopping at v_{k-2} means that $\sum_{i=1}^{k-2} d_i \geqslant D_1$. Thus $\sum_{i=1}^{k-1} d_i' \geqslant \sum_{i=1}^{k-2} d_i \geqslant D_1 = \Omega(\log \log n)$ and then a_u is small.
- The vertex v_{k-1} is added to S_v because the DFS stops at the vertex v_{k-1} and v_{k-1} has at least D_2 children. Now, although $\sum_{i=1}^{k-1} d_i \geqslant D_1$, we have no lower bound on $\sum_{i=1}^{k-1} d_i'$ because d_{k-1}' can be much smaller than d_{k-1}. However, in this case v_{k-1} has many neighbors in ∂S_v because $d_{k-1} \geqslant D_2$. By the definition of $\Omega_{\partial S_v}$, many neighbors of v_{k-1} are assigned 1. Since the spin system is ferromagnetic, the value on v_{k-1} is almost fixed to be 1. The vertex v_{k-1} *blocks* the influence from u to v and a_u is small.

To bound the sum of influences $\sum_{u \in \partial S_v} a_u$, we decompose the sum as $\sum_{k \geqslant 1} \sum_{u \in L_k(v) \cap \partial S_v} a_u := \sum_{k \geqslant 1} \text{Inf}(k)$, where $L_k(v)$ denotes the set of vertices at level k in the tree G rooted at v and $\text{Inf}(k)$ is the sum of influences at level k. We then use correlation decay analysis to bound $\text{Inf}(k)$ for each level k. Compared to the all-to-one influence bound, which is proved using a similar methodology, a new challenge is the presence of the boundary conditions in a_u's. For two vertices u and u' in ∂S_v at the same level k, the boundary conditions to achieve a_u and $a_{u'}$ may be very different and some of the disagreements may be very close to the root v. This makes the correlation decay analysis difficult to carry out.

To resolve this issue, we showed that, roughly speaking, if $\lambda < \lambda_0$ (for the definition of λ_0, recall Theorem 3), then we can assume that

$$\forall w \in L_{<k}(v) \cap \partial S_v, \quad \sigma(w) = \tau(w), \text{ where } L_{<k}(v) := \cup_{j=1}^{k-1} L_j(v), \tag{4}$$

where σ and τ are two boundary conditions that achieve a_u and $a_{u'}$. Hence, when analysing $\text{Inf}(k)$, we can assume all pinnings above level k are consistent for all $u \in L_k(v) \cap \partial S_v$. The disagreements only appear after level k. Details of this argument are in Lemma 42. With its help we then can apply the correlation decay analysis to establish ASSM. We remark that (4) is the only place where we need to use the stronger condition $\lambda < \lambda_0$ in instead of $\lambda < \lambda_c$. If one can verify the typical-case ASSM property when $\lambda < \lambda_c$, then the above analysis framework gives an improved $\tilde{O}(n)$ mixing time to Theorem 5.

So far, all the discussion above assumes the graph G itself is a tree. For a general graph G, the set S_v can be constructed as follows. We first construct a *self-avoiding walk* (SAW) tree T_{SAW} of the graph G rooted at v (a tree enumerating all self-avoiding walks from v in graph G). Then, using the same construction as in the tree case, we construct the region S_v^T for the SAW tree T_{SAW} and then map all vertices in S_v^T back to the original graph G to obtain S_v. Details of this construction are in Section 6. Let ∂S_v be the outer boundary of S_v in G. The good boundary condition $\sigma \in \Omega_{\partial S_v}$ is defined similarly as above: if a vertex $w \in S_v$ has many neighbors in ∂S_v, then many of them are assigned 1 in σ. To prove the ASSM property in G, we reduce the task to analyzing influences in the self-avoiding walk tree T_{SAW}. Using the ideas above, we show that every path in the SAW tree contributes a good decay of correlation, so that typical-case ASSM holds in general graphs.

3. Preliminaries

3.1. **Markov chain and mixing time.** Let X_t be a Markov chain on a state space Ω with transition matrix P. We call a Markov chain *irreducible* if for any two states $x, y \in \Omega$, there exists a positive integer t such that $P^t(x, y) > 0$, *aperiodic* if for any $x \in \Omega$, $\gcd\{t \geqslant 1 : P^t(x, x) > 0\} = 1$, and *reversible* with respect to a distribution μ if $\mu(x)P(x, y) = \mu(y)P(y, x)$ for all $x, y \in \Omega$. An irreducible, aperiodic, and reversible Markov chain has a unique stationary distribution μ. The mixing time is defined as

$$t^P_{\mathrm{mix}}(\epsilon) = \max_{x\in\Omega} \min\{t \geqslant 0 : D_{TV}\left(P^t(x, \cdot), \mu\right) < \epsilon\}.$$

We often consider the mixing time when $\epsilon = 1/(4e)$ because of the following general bound

$$\forall \epsilon > 0, \quad t^P_{\mathrm{mix}}(\epsilon) \leqslant t^P_{\mathrm{mix}}\left(\frac{1}{4e}\right) \log \frac{1}{\epsilon}. \tag{5}$$

Let μ be a distribution over $\Omega = \{0, 1\}^V$. Let P be the Glauber dynamics on μ. Then, the transition matrix P is positive semi-definite with real non-negative eigenvalues $1 = \lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_{|\Omega|} \geqslant 0$. The spectral gap of P is defined as $\gamma_{GD} = 1 - \lambda_2$. For any distribution ν over Ω, it is well known that

$$D_{\chi^2}(\nu P \| \mu P) \leqslant (1 - \gamma_{GD})^t \cdot D_{\chi^2}(\nu \| \mu),$$

where $D_{\chi^2}(\nu\|\mu) = \sum_{x\in\Omega} \frac{(\nu(x)-\mu(x))^2}{\mu(x)}$ is the chi-squared divergence between ν and μ. The following relationship between the total variation distance and the chi-squared divergence holds:

$$D_{TV}(\nu, \mu) \leqslant \sqrt{D_{\chi^2}(\nu\|\mu)}.$$

As a consequence, for the Glauber dynamics on μ starting from an arbitrary configuration $X_0 = \sigma$,

$$D_{TV}(X_t, \mu) \leqslant \sqrt{D_{\chi^2}(X_t\|\mu)} \leqslant \sqrt{(1 - \gamma_{GD})^t \cdot D_{\chi^2}(X_0\|\mu)} \leqslant \sqrt{(1 - \gamma_{GD})^t \cdot \frac{1}{\mu(\sigma)}}.$$

where X_t is the distribution of the Glauber dynamics on μ after t steps starting from X_0. Hence, the mixing time of the Glauber dynamics on μ is at most

$$t^P_{\mathrm{mix}}(\epsilon) \leqslant \frac{1}{\gamma_{GD}} \log \frac{1}{\epsilon^2 \mu(\sigma)}. \tag{6}$$

The ratio $\frac{1}{\gamma_{GD}}$ is called the relaxation time of the Glauber dynamics on μ. For the other direction,

$$\forall \epsilon > 0, \quad t^P_{\mathrm{mix}}(\epsilon) \geqslant \left(\frac{1}{\gamma_{GD}} - 1\right) \log \frac{1}{2\epsilon}. \tag{7}$$

Next, consider a Gibbs distribution μ defined on a bipartite graph $G = (V_0, V_1, E)$. Let Q be the alternating-scan sampler on μ. Formally, let P_0 denote the transition matrix of updating the configuration on V_0 conditional on the current configuration of the other part V_1, and let P_1 denote the transition matrix of updating the configuration on V_1 conditional on the current configuration of the other part V_0. Then, the transition matrix Q of the alternating-scan sampler is defined as

$$Q = P_1 P_0.$$

When μ is the Gibbs distribution of a restricted Boltzmann machine, the Markov chain Q is irreducible, aperiodic, and has the unique stationary distribution μ. However, Q may not be reversible with respect to μ. Let the multiplicative reversiblization be $R(Q) = QQ^*$, where Q^* is defined by

$$Q^*(\sigma, \tau) = \frac{\mu(\tau)}{\mu(\sigma)} Q(\tau, \sigma).$$

Then $R(Q)$ is reversible with respect to μ. Furthermore, all eigenvalues of $R(Q)$ are real and non-negative [Fil91]. The relaxation time of the alternating-scan sampler is defined by

$$T_{\mathrm{rel}}(Q) = \frac{1}{1 - \sqrt{1 - \gamma(R(Q))}},$$

where $\gamma(R(Q)) = 1 - \lambda_2(R(Q))$, and $\lambda_2(R(Q))$ is the second largest eigenvalue of $R(Q)$.

Proposition 6 ([GKZ18, Theorem 1]). *For a RBM on a bipartite graph with Gibbs distribution* μ*,*

$$T_{rel}(Q) \leqslant \frac{2}{\gamma_{GD}},$$

where γ_{GD} *is the spectral gap of the Glauber dynamics on* μ*.*

Remark. Theorem 1 in [GKZ18] considers the spectral gap of the lazy version of the Glauber dynamics on μ, which is $\frac{1}{2}I + \frac{1}{2}P$, where I is the identity matrix. Hence, we add a factor of 2 in the above proposition.

The mixing time of the alternating-scan sampler on μ can be bounded by the following proposition.

Proposition 7 ([GKZ18, Theorem 3]). *For the alternating-scan sampler* Q *on an RBM, starting from a configuration* $\sigma \in \{0,1\}^V$*, after running* Q *for* $T_{rel}(Q) \log \frac{4e^2}{\epsilon^2 \mu(\sigma)}$ *steps, the total variation distance between the resulting distribution and the stationary distribution is at most* ϵ*.*

Remark. The mixing time upper bound stated in [GKZ18, Theorem 3] is $T_{\mathrm{rel}}(Q) \log \frac{4e^2}{\mu_{\min}}$, where $\mu_{\min} := \min_{\sigma \in \{0,1\}^V} \mu(\sigma)$ and they define the mixing time by setting $\epsilon = 1/(2e)$. To get Proposition 7, generalising from $1/(2e)$ to an arbitrary ϵ is straightforward, and the proof of [Fil91, Theorem 2.1], which the proof in [GKZ18] is based on, already deals with $\mu(\sigma)$ instead of $\mu_{\min}$.

3.2. **Self-reducibility.** Let $G = (V, E)$ be a graph. Let μ be the Gibbs distribution of a ferromagnetic two-spin system on G with parameters $(\beta_e, \gamma_e)_{e \in E}, (\lambda_v)_{v \in V}$.

Fix a subset $\Lambda \subseteq V$. Let $\sigma \in \{0,1\}^{V \setminus \Lambda}$ be a configuration on $V \setminus \Lambda$. We use μ^σ to denote the distribution of $X \sim \mu$ conditional on $X(\Lambda) = \sigma$. The pinning σ induces a conditional distribution μ^σ_Λ on Λ given σ. Note that μ^σ_Λ is a Gibbs distribution of a ferromagnetic two-spin system on $G[\Lambda]$ with edge activities $(\beta_e, \gamma_e)_{e \in G[\Lambda]}$. For all vertices $v \in \Lambda$, the vertex activity at v is updated to $\lambda'_v = \lambda_v \prod_{e \in E_c} \beta_e \prod_{e \in E_1} \frac{1}{\gamma_e} \leqslant \lambda_v$, where E_c is the set of edges $\{v, u\}$ for $u \in V \setminus \Lambda$ and $\sigma_u = c$ for $c \in \{0,1\}$.

Observation 8 (Self-reducibility under pinning). *Let* $\beta \leqslant 1 < \gamma$*,* $\beta\gamma > 1$ *and* $\lambda < \lambda_c(\beta, \gamma)$*. For any* (β, γ, λ)*-ferromagnetic two-spin system with Gibbs distribution* μ *in* $G = (V, E)$*, for any pinning* σ *on a subset* $\Lambda \subseteq V$*,* μ^σ_Λ *is also the Gibbs distribution of a* (β, γ, λ)*-ferromagnetic two-spin system on* $G[\Lambda]$*.*

3.3. **Self-avoiding walk tree.** Let $G = (V, E)$ be a graph. Assume that there is a total ordering of all vertices V in G. The self-avoiding walk (SAW) tree is defined as follows.

Definition 9 (SAW tree [Wei06]). Let $G = (V, E)$ be a graph. For any vertex $v \in V$, the SAW tree $T_{\mathrm{SAW}}(G, v)$ rooted at v enumerates all SAWs from v such that every path $v_0 - v_1 - \cdots - v_\ell$ from root to leaf satisfies that either it is a SAW that ends at v_ℓ (namely the degree $\deg_G(v_\ell)$ of v_ℓ is 1) or it is a SAW that ends at a cycle-closing vertex v_ℓ ($v_0 - v_1 - \cdots - v_{\ell-1}$ is a SAW and $v_\ell = v_i$ for some $0 \leqslant i \leqslant \ell - 2$).

In addition, we also need to consider SAW trees when a boundary is present. Let $S \subseteq V$ be a set of boundary vertices. The SAW tree $T = T_{SAW}(G, v, S)$ rooted at v with boundary S is same as $T = T_{SAW}(G, v)$ defined in Definition 9, except that any SAW stops immediately after reaching a boundary vertex $u \in S$, in which case u is the last vertex in that SAW. Thus, $T_{SAW}(G, v)$ is the same as $T_{SAW}(G, v, \emptyset)$.

The following two observations are straightforward to verify from the definition.

Observation 10. *For any non-leaf vertex u in $T_{SAW}(G, v, S)$, the degree of u in T is the same as the degree of its preimage $f(u)$ in G.*

Observation 11. *Any leaf u in $T_{SAW}(G, v, S)$ falls into three disjoint types: (1) u is a copy of some vertex in the boundary S; (2) u is a cycle-closing vertex; (3) u has degree one in G and is not a copy of any vertex in S. As a corollary, any cycle-closing vertex u cannot be a copy of any vertex in S.*

Consider a spin system on graph G with parameters $(\beta_e, \gamma_e)_{e \in E}, (\lambda_v)_{v \in V}$ and the Gibbs distribution μ. Fix a vertex v and a pinning $\sigma \in \{0,1\}^S$ over boundary S. To analyse the conditional marginal distribution μ_v^σ, we need to use the following construction of SAW trees with pinnings.

Definition 12 (SAW tree with pinning)**.** Let $\sigma \in \{0,1\}^S$ be a partial pinning on S, where $S \subseteq V$ is a set of boundary vertices. The SAW tree $T_{SAW}(G, v, \sigma)$ rooted at v with pinning σ is constructed as follows.

(1) Construct the SAW tree $T = T_{SAW}(G, v, S)$ with boundary S.
(2) For any leaf vertex in T that is a copy of some $u \in S$, pin its value to be $\sigma(u)$.
(3) For any cycle-closing leaf vertex v_ℓ in T, say $v_\ell = v_i$ for some $0 \leqslant i \leqslant \ell - 2$ in the SAW, we pin the value of v_ℓ to be 0 if $v_{i+1} > v_\ell$ and pin the value of v_ℓ to be 1 if $v_{i+1} < v_\ell$ according to the total order of V.

By Observation 11, if some leaf vertex u in $T_{SAW}(G, v, S)$ gets pinned in the second step of Definition 12, then the pinning on u will not be changed in the third step because u cannot be a cycle-closing vertex.

Let $T = T_{SAW}(G, v, \sigma)$. Denote $T = (V_T, E_T)$, where V_T is all vertices in T and E_T are all edges in T. By Definition 9, some leaf vertices of T are cycle-closing vertices and we define

$$\Gamma := \{w \in V_T : w \text{ is a cycle-closing leaf vertex of } T\}. \tag{8}$$

We remark that $\Gamma \subseteq V_T$ is determined by the tuple (G, v, S) and all vertices in Γ are leaf vertices of T. We use ρ_Γ to denote the pinning on all cycle-closing leaf vertices of T.

For a vertex w in graph G, it may have multiple copies in T. We use $\mathrm{copy}(w)$ to denote the set of all copies of w in T. Define the set of all copies of vertices in S as

$$\bar{S} := \bigcup_{w \in S} \mathrm{copy}(w). \tag{9}$$

By the construction of T, $\bar{S}$ is a subset of leaf vertices in T. We use $\sigma_{\bar{S}}$ to denote the pinning on all vertices in $\bar{S}$. Note that $\sigma_{\bar{S}}$ is determined by the pinning $\sigma \in \{0,1\}^S$.

Every vertex in T is a copy of some vertex in G and every edge in T is a copy of some edge in G. We can naturally define a Gibbs distribution on T by inheriting the parameters of the two-spin systems on G. Denote the Gibbs distribution on T as π. Let $\pi^{\bar{\sigma}}$ be the Gibbs distribution on T with pinning $\bar{\sigma} = \rho_\Gamma \cup \sigma_{\bar{S}}$. The main point of all these constructions is the following well-known result by Weitz [Wei06].

Proposition 13 ([Wei06])**.** *For the root vertex v, two marginal distributions μ_v^σ and $\pi_v^{\bar{\sigma}}$ are identical.*

3.4. **Tree recursion and potential function.** Consider a SAW tree T rooted at v with pinning $\bar{\sigma}$ on a subset of leaf vertices. For each vertex $w \in T$, let T_w be the sub-tree of T rooted at w. Consider the spin system induced by the sub-tree T_w on the vertices in T_w. Let $p_w(0)$ and $p_w(1)$ be the marginal probabilities of w being 0 and 1 in the Gibbs distribution induced by the sub-tree T_w respectively. Define

$$R_w = \frac{p_w(0)}{p_w(1)}. \tag{10}$$

If the value of w is pinned to be 0, then $p_w(1) = 0$ and $R_w = \infty$. This happens only at leaves.

Let u be a vertex in T. Let $u_1, u_2, \ldots, u_d$ be the children of u. The tree recursion function $F_u : [0, \infty]^d \to \mathbb{R}$ at the vertex u is defined as

$$F_u(x_1, x_2, \ldots, x_d) = \lambda_u \prod_{i=1}^{d} \frac{\beta_{u,u_i} x_i + 1}{x_i + \gamma_{u,u_i}}. \tag{11}$$

Weitz [Wei06] shows a well-known recursion relation

$$R_u = F_u(R_{u_1}, \ldots, R_{u_d}).$$

Let $\beta \leqslant 1 < \gamma$, $\beta\gamma > 1$ and $\lambda > 0$ be three parameters. Now, let us consider a (β, γ, λ)-ferromagnetic two-spin system on graph G in Definition 2. Guo and Lu [GL18] used a potential function method to analyze the recursion function. By (11), the image space of F_u is within $[0, \lambda)$. Let $\Phi : [0, \lambda) \to \mathbb{R}$ be a differentiable and increasing potential function. Instead of analyzing the recursion of R_w, they analyze the recursion of $\Phi(R_w)$. The tree recursion in (11) at vertex u with potential function Φ is

$$F_u^{\Phi}(y_1, y_2, \ldots, y_d) = (\Phi \circ F_u \circ \Phi^{-1})(y_1, y_2, \ldots, y_d),$$

where $y = \Phi(x)$, $y_i = \Phi(x_i)$ and all $x_i \in [0, \lambda)$.

The potential function used by Guo and Lu [GL18] for ferromagnetic two-spin systems is given implicitly via its derivative $\phi(x) = \Phi'(x)$, which is

$$\phi(x) := \min\left\{ \frac{1}{x \log \frac{\lambda}{x}}, \frac{1}{t} \right\}, \text{ where } t = t(\beta, \gamma, \lambda) > 0 \text{ is a constant.} \tag{12}$$

The following observation is easy to prove using the definition of $\phi(x)$.

Observation 14. *There exist constants $C_{\max} > 0$ and $C_{\min} > 0$ such that*

$$\forall x \in [0, \lambda), \quad C_{\min} \leqslant \phi(x) \leqslant C_{\max}.$$

The specific definition of the constant t can be found in [GL18]. The potential function is then

$$\Phi(x) = \int_0^x \phi(s)\, ds. \tag{13}$$

The potential function $\Phi(x)$ satisfies the following property.

Lemma 15 ([GL18])**.** *Let $\beta \leqslant 1 < \gamma$, $\beta\gamma > 1$ and $\lambda < \lambda_c(\beta, \gamma) := (\gamma/\beta)^{\frac{\sqrt{\beta\gamma}}{\sqrt{\beta\gamma}-1}}$ be three parameters. Consider the recursion function F_u in (11) with $\lambda_u < \lambda$ and for any edge $e = \{u, u_i\}$, $\beta_e \leqslant \beta \leqslant 1 < \gamma \leqslant \gamma_e$, $\beta\gamma \geqslant \beta_e\gamma_e > 1$. Then, there exist a constant $0 < \alpha = \alpha(\beta, \gamma, \lambda) < 1$ such that for all $x_1, \ldots, x_d \in (0, \lambda)$,*

$$C_{\phi,d}(\boldsymbol{x}) := \phi(F_u(\boldsymbol{x})) \sum_{i=1}^{d} \left| \frac{\partial F_u}{\partial x_i}(\boldsymbol{x}) \right| \frac{1}{\phi(x_i)} \leqslant 1 - \alpha.$$

In [GL18], Lemma 15 is proved for uniform parameters, namely, the same β, γ for all edges and the same λ for all vertices. For non-uniform parameters $(\lambda_v)_{v \in V}$ and $(\beta_e, \gamma_e)_{e \in E}$, a proof is given in Appendix A.

In addition, we also have the following trivial bound for each term in the sum $C_{\phi,d}(\boldsymbol{x})$.

Lemma 16. *Let $\beta \leqslant 1 < \gamma$, $\beta\gamma > 1$ and $\lambda > 0$ be three parameters. Consider the recursion function F_u in (11) with $\lambda_u < \lambda$ and for any edge $e = \{u, u_i\}$, $\beta_e \leqslant \beta \leqslant 1 < \gamma \leqslant \gamma_e$, $\beta\gamma \geqslant \beta_e\gamma_e > 1$. For any $1 \leqslant i \leqslant d$,*

$$\forall x_1, x_2, \ldots, x_d \in (0, \lambda), \quad \phi(F_u(\boldsymbol{x})) \left| \frac{\partial F_u}{\partial x_i}(\boldsymbol{x}) \right| \frac{1}{\phi(x_i)} \leqslant C_{trl} \cdot \lambda_u \left(\frac{\beta\lambda + 1}{\lambda + \gamma} \right)^{d-1} = \lambda_u \exp(-\Omega(d)),$$

where C_{trl} is a constant depending on β, γ, λ.

Proof. By Observation 14, we have $\phi(F_u(\mathbf{x})) \leqslant C_{\max}$ and $\frac{1}{\phi(x_i)} \leqslant \frac{1}{C_{\min}}$. Further,

$$\left|\frac{\partial F_u}{\partial x_i}(\mathbf{x})\right| = \lambda_u \frac{\beta_{u,u_i}\gamma_{u,u_i} - 1}{(x_i + \gamma_{u,u_i})^2} \prod_{1 \leqslant j \leqslant d : j \neq i} \frac{\beta_{u,u_j} x_j + 1}{x_j + \gamma_{u,u_j}} \leqslant \lambda_u \frac{\beta\gamma - 1}{\gamma^2}\left(\frac{\beta\lambda + 1}{\lambda + \gamma}\right)^{d-1}.$$

The lemma holds by taking the constant $C_{\mathrm{trl}} = \frac{C_{\max}}{C_{\min}} \frac{\beta\gamma - 1}{\gamma^2}$. □

Using the potential function and the above property, Guo and Lu [GL18] showed the following strong spatial mixing (SSM) result for (β, γ, λ)-ferromagnetic two-spin systems.

Lemma 17 ([GL18])**.** *Let $\beta \leqslant 1 < \gamma$, $\beta\gamma > 1$ and $\lambda < \lambda_c(\beta, \gamma) := (\gamma/\beta)^{\frac{\sqrt{\beta\gamma}}{\sqrt{\beta\gamma}-1}}$ be three parameters. Consider the Gibbs distribution μ of a (β, γ, λ)-ferromagnetic two-spin system on graph $G = (V, E)$. There exist constants $A = A(\beta, \gamma, \lambda) > 0$ and $0 < B = B(\beta, \gamma, \lambda) < 1$ such that for any two configurations σ and τ in a subset $\Lambda \subseteq V$, where σ and τ differ only at subset $D \subseteq \Lambda$, then for any vertex $v \notin \Lambda$,*

$$\left|\frac{\mu_v^\sigma(0)}{\mu_v^\sigma(1)} - \frac{\mu_v^\tau(0)}{\mu_v^\tau(1)}\right| \leqslant A(1 - B)^\ell,$$

where $\ell = \min_{u \in D} d(u, v)$ is the distance from v to the closest vertex in D.

4. All-to-one influence bound

We start by establishing the all-to-one influence bound. The analysis here is also useful later to establish ASSM in Section 8.

Definition 18 (All-to-one influence)**.** Let μ be a distribution over $\{0, 1\}^V$. We say that μ has C_{inf}-bounded all-to-one influence if, for every vertex $v \in V$,

$$\sum_{u \in V \setminus \{v\}} \left|\Pr_{X \sim \mu}[X(v) = 0 \mid X(u) = 0] - \Pr_{X \sim \mu}[X(v) = 0 \mid X(u) = 1]\right| \leqslant C_{\mathrm{inf}}.$$

Theorem 19. *Let $\beta \leqslant 1 < \gamma$, $\beta\gamma > 1$, and $\lambda < \lambda_c(\beta, \gamma) := (\gamma/\beta)^{\frac{\sqrt{\beta\gamma}}{\sqrt{\beta\gamma}-1}}$. Let μ be the Gibbs distribution for a (β, γ, λ)-ferromagnetic two-spin system on $G = (V, E)$. It has C_{inf}-bounded all-to-one influence, where $C_{\mathrm{inf}} = C_{\mathrm{inf}}(\beta, \gamma, \lambda) > 0$ is a constant depending only on β, γ, λ.*

To prove this theorem, consider the SAW tree $T = T_{\mathrm{SAW}}(G, v, \emptyset)$ rooted at v. The cycle-closing leaves of T have fixed pinned values. We use the self-reducibility property in Observation 8 to remove all cycle-closing leaves from the SAW tree and update the external fields at their neighbours. Thus, without loss of generality, we can assume there is no pinning on T. Let π denote the Gibbs distribution on $T = (V_T, E_T)$, where the parameters are inherited from μ. Fix a vertex $w \in V$. Let $S = \mathrm{copy}(w)$ be the set of all copies of w in T. By Proposition 13, $\mu_v^{w \leftarrow c}$ is identical to $\pi_v^{S \leftarrow c}$ for $c \in \{0, 1\}$, where $S \leftarrow c$ is the pinning on S such that all $x \in S$ are pinned to be c.

For any vertex $u \in V_T$, let R_u be the marginal ratio at u defined in (10). The ratio R_u can be computed recursively using the tree recursion function F_u in (11) in a bottom-up manner. From this perspective, T can also be viewed as a computation tree for the ratio R_u.

Definition 20 (Pinning on the computation tree)**.** Let $u \in V_T$ and S be a subset of vertices in the subtree of u, where $u \notin S$. Let $\sigma : S \to [0, \infty]$ be a pinning on S (of ratios). For each $x \in S$, we remove all the descendants of x and fix the value $R_x = \sigma(x)$. Then, all pinnings are on the leaves of the subtree rooted at u. For all other leaf vertices x', we set $R_{x'} = \lambda_{x'}$ as the definition of $R_{x'}$ in (10). We use R_u^σ to denote the marginal ratio at u computed via tree recursion in a bottom-up manner.

We also use the notation R_u^σ even if σ contains pinning outside the subtree of u. In this case, $R_u^\sigma = R_u^{\bar{\sigma}}$, where $\bar{\sigma}$ is the pinning obtained from σ by removing the pinning outside the subtree of u.

By definition, it is straightforward to verify that $R_v^{S\leftarrow\infty} = \frac{\mu_v^{w\leftarrow 0}(0)}{\mu_v^{w\leftarrow 0}(1)}$ and $R_v^{S\leftarrow 0} = \frac{\mu_v^{w\leftarrow 1}(0)}{\mu_v^{w\leftarrow 1}(1)}$, where S is the set of all copies of w in T. Note that for the computation tree, pinnings are with respect to the ratio R instead of the state, although it is easy to translate between the two. To emphasize that the pinning is on all copies of w, we denote

$$R_v^{w^0} = R_v^{S\leftarrow\infty} \quad \text{and} \quad R_v^{w^1} = R_v^{S\leftarrow 0}.$$

The following lemma is straightforward.

Lemma 21. *The influence of* w *on* v *can be bounded by*

$$D_{TV}\left(\mu_v^{w\leftarrow 0}, \mu_v^{w\leftarrow 1}\right) \leqslant \left| R_v^{w^0} - R_v^{w^1} \right|,$$

Proof. By Proposition 13, $\mu_v^{w\leftarrow c}$ coincides with $\pi_v^{S\leftarrow c}$ for $c \in \{0,1\}$, where $S = \text{copy}(w)$ and $R_v^{w^0} = R_v^{S\leftarrow\infty}$, $R_v^{w^1} = R_v^{S\leftarrow 0}$ as in the notation above. So $D_{TV}\left(\mu_v^{w\leftarrow 0}, \mu_v^{w\leftarrow 1}\right) = D_{TV}\left(\pi_v^{S\leftarrow\infty}, \pi_v^{S\leftarrow 0}\right)$. The marginals at v are Bernoulli: $\pi_v^{S\leftarrow\infty}(1) = 1/(1+R_v^{w^0})$ and $\pi_v^{S\leftarrow 0}(1) = 1/(1+R_v^{w^1})$. Thus

$$D_{TV}\left(\pi_v^{S\leftarrow\infty}, \pi_v^{S\leftarrow 0}\right) = \left| \frac{1}{1+R_v^{w^0}} - \frac{1}{1+R_v^{w^1}} \right| = \frac{\left|R_v^{w^0} - R_v^{w^1}\right|}{(1+R_v^{w^0})(1+R_v^{w^1})} \leqslant \left|R_v^{w^0} - R_v^{w^1}\right|. \qquad \square$$

In $R_v^{w^0}$ and $R_v^{w^1}$, a set of vertices is pinned to 0 or 1. Next, we decompose the influence into the sum of influences contributed by individual vertices in this set. We define the following notion of influence from one vertex in the computation tree. A similar definition and analysis for the hardcore model appears in [ALO24], but we need a more careful definition for ferromagnetic two-spin systems. Define the set of vertices at level k by

$$\forall k \in \mathbb{N}, \quad L_k(u) = \{v \in V_T : d(v,u) = k\},$$

where $d(v,u)$ is the distance from v to u in the SAW tree T. A vertex u' is called a sibling of u if u' has the same parent as u.

Definition 22 (Influence from one vertex in the computation tree)**.** Let $u \in L_k(v)$ be a vertex in the computational tree T at level k. Define the influence of u on v as

$$I_v^u = \sup_{\sigma\in\mathcal{S}} \left| R_v^{\sigma\wedge u\leftarrow\infty} - R_v^{\sigma\wedge u\leftarrow 0} \right|,$$

where $\mathcal{S}$ contains all pinnings $\sigma : L_k(v)\backslash\{u\} \to [0,\infty]$ satisfying that for all siblings u' of u, $\sigma(u') \in (0,\lambda)$.

Compared to the definition in [ALO24], our definition explicitly constrains the siblings of u. We next prove the following influence bound using the technique in [ALO24].

Lemma 23. *The influence satisfies*

$$|R_v^{w^0} - R_v^{w^1}| \leqslant 2 \sum_{u\in \mathit{copy}(w)} I_v^u.$$

Proof. Let $u_1, \ldots, u_m$ be the vertices in $\text{copy}(w)$ in the increasing order of the distance to root v, which means $d(v,u_i) \leqslant d(v,u_j)$ for $1 \leqslant i < j \leqslant m$. Let $S_i := \{u_i, \cdots, u_m\}$ for $1 \leqslant i \leqslant m$. For j from 0 to m, we inductively show that:

$$|R_v^{w^0} - R_v^{w^1}| \leqslant 2\sum_{i=1}^{j} I_v^{u_i} + |R_v^{S_{j+1}\leftarrow\infty} - R_v^{S_{j+1}\leftarrow 0}|,$$

where $S_{m+1} = \emptyset$ and $|R_v^{S_{m+1}\leftarrow\infty} - R_v^{S_{m+1}\leftarrow 0}| = 0$. When $j = 0$, the inequality holds trivially. Assume that the inequality holds for j for some $0 \leqslant j < m$. We next show that the inequality also holds for $j+1$. By the triangle inequality, we have

$$|R_v^{S_{j+1}\leftarrow\infty} - R_v^{S_{j+1}\leftarrow 0}| \leqslant |R_v^{S_{j+1}\leftarrow\infty} - R_v^{S_{j+2}\leftarrow\infty}| + |R_v^{S_{j+2}\leftarrow\infty} - R_v^{S_{j+2}\leftarrow 0}| + |R_v^{S_{j+2}\leftarrow 0} - R_v^{S_{j+1}\leftarrow 0}|.$$

To verify the $j+1$ case, using the induction hypothesis on j, it suffices to show that the first and third terms are each bounded by $I_v^{u_{j+1}}$. We only prove this for the first term, since the third term is analogous. By the monotonicity of the recursion function,

$$|R_v^{S_{j+1}\leftarrow\infty} - R_v^{S_{j+2}\leftarrow\infty}| \leqslant |R_v^{S_{j+2}\leftarrow\infty\wedge u_{j+1}\leftarrow\infty} - R_v^{S_{j+2}\leftarrow\infty\wedge u_{j+1}\leftarrow 0}|.$$

Because $d(v, u_{j+1}) \leqslant d(v, u_i)$ for all $i > j+1$, all pinnings on S_{j+2} induce pinnings on $L_k(v) \setminus \{u_{j+1}\}$, where k is the level of u_{j+1} in T. Moreover, all siblings of u_{j+1} are not in $\mathrm{copy}(w)$, and thus they are unpinned. By the definition of the tree recursion, when we compute the tree recursion from bottom to top, all siblings u' of u_{j+1} obtain a value in $(0, \lambda)$, which is the induced pinning on u'. For all other vertices $u'' \in L_k(v) \setminus \{u_{j+1}\}$ that is not a sibling of u_{j+1}, the ratio on u'' computed via the tree recursion can be any value in $[0, \infty]$. Therefore, by the definition of $I_v^{u_{j+1}}$ in Definition 22, we have

$$|R_v^{S_{j+1}\leftarrow\infty} - R_v^{S_{j+2}\leftarrow\infty}| \leqslant I_v^{u_{j+1}}.$$

The same argument gives $|R_v^{S_{j+2}\leftarrow 0} - R_v^{S_{j+1}\leftarrow 0}| \leqslant I_v^{u_{j+1}}$. This proves the $j+1$ case and hence the lemma. □

Using Lemma 23 and Lemma 21, we have the following bound

$$\sum_{w\in V\setminus\{v\}} D_{\mathrm{TV}}\left(\mu_v^{w\leftarrow 0}, \mu_v^{w\leftarrow 1}\right) \leqslant 2 \sum_{w\in V\setminus\{v\}} \sum_{u\in\mathrm{copy}(w)} I_v^u = 2\sum_{k\geqslant 1}\sum_{w\in L_k(v)} I_v^w. \tag{14}$$

Next, fix an integer $k \geqslant 1$. We bound the sum of influences over all vertices in $L_k(v)$. We also work with the potential function Φ defined in (12). Fix a vertex $w \in L_k(v)$. Let σ^w be a pinning on $L_k(v) \setminus \{w\}$ that attains (or is arbitrarily close to) the supremum in the definition of I_v^w. We emphasize that σ^w depends on w. Instead of directly bounding I_v^w, we bound the potential difference $|\Phi(R_v^{\sigma^w\wedge w\leftarrow\infty}) - \Phi(R_v^{\sigma^w\wedge w\leftarrow 0})|$. We use the following general relation.

Lemma 24. *For any two $x^0, x^1 \in (0, \lambda)$, we have*

$$\frac{1}{C_{\max}}\left|\Phi(x^0) - \Phi(x^1)\right| \leqslant \left|x^0 - x^1\right| \leqslant \frac{1}{C_{\min}}\left|\Phi(x^0) - \Phi(x^1)\right|,$$

where $C_{\max}$ and $C_{\min}$ are constants defined in Observation 14.

Proof. For the potential Φ with derivative $\phi = \Phi'$ from (12), the mean value theorem gives

$$|\Phi(x^0) - \Phi(x^1)| = \phi(\eta)\,|x^0 - x^1|$$

for some η between x^0 and x^1. By Observation 14, for any $z \in (0, \lambda)$, we have $\phi(z) \geqslant C_{\min}$ and $\phi(z) \leqslant C_{\max}$. The lemma can be proved using the following equation:

$$|x^0 - x^1| = \frac{|\Phi(x^0) - \Phi(x^1)|}{\phi(\eta)}$$ □

Now, our task is reduced to bound the difference of the potential $\Phi(R_v^{\sigma^w\wedge w\leftarrow\infty}) - \Phi(R_v^{\sigma^w\wedge w\leftarrow 0})$. In Section 4.1, we give some general influence decay results. In Section 4.2, we apply these general results to prove the influence bound.

4.1. **General influence decay results.** Next, we present general results for proving the influence bound, which will also be used later to prove aggregate strong spatial mixing. Consider a ferromagnetic two-spin system $\mathcal{S}$ on a tree T, rooted at v. For each vertex $w \in L_k(v)$, let σ^w be a pinning on $L_k(v) \setminus \{w\}$. Different vertices w may correspond to different pinnings σ^w. Define the potential-based influence from w to the root v as

$$K_v^w = \left| \Phi(R_v^{\sigma^w \wedge w \leftarrow \infty}) - \Phi(R_v^{\sigma^w \wedge w \leftarrow 0}) \right|. \tag{15}$$

More generally, for any vertex u on the path between w and v, define the influence K_u^w of w on u by

$$K_u^w = \left| \Phi(R_u^{\sigma^w \wedge w \leftarrow \infty}) - \Phi(R_u^{\sigma^w \wedge w \leftarrow 0}) \right|, \tag{16}$$

where $R_u^{\sigma^w \wedge w \leftarrow \infty}$ and $R_u^{\sigma^w \wedge w \leftarrow 0}$ are ratios computed by tree recursion in $\mathcal{S}$, with σ^w restricted to the subtree rooted at u.

The following two general influence decay results hold.

Lemma 25. *Suppose $\mathcal{S}$ in T is a (β, γ, λ)-ferromagnetic two-spin system with $\beta \leqslant 1 < \gamma$ and $\beta\gamma > 1$. Let $u \in L_\ell(v)$ be a vertex at level ℓ, where $0 \leqslant \ell \leqslant k-2$. Let $u_1, u_2, \ldots, u_d$ be the children of u. Then*

$$\sum_{w \in L_{k-\ell}(u)} K_u^w \leqslant C_{trl} \cdot \lambda_u d \left(\frac{\beta\lambda + 1}{\lambda + \gamma} \right)^{d-1} \cdot \max_{1 \leqslant i \leqslant d} \sum_{w \in L_{k-\ell-1}(u_i)} K_{u_i}^w$$
$$= \lambda_u d \exp(-\Omega(d)) \max_{1 \leqslant i \leqslant d} \sum_{w \in L_{k-\ell-1}(u_i)} K_{u_i}^w,$$

where $C_{trl} = C_{trl}(\beta, \gamma, \lambda)$ is the constant in Lemma 16 and $L_j(u)$ denotes the set of vertices at level j in the subtree rooted at u.

Lemma 26. *Suppose $\mathcal{S}$ in T is a (β, γ, λ)-ferromagnetic two-spin system with $\beta \leqslant 1 < \gamma$, $\beta\gamma > 1$, and $\lambda < \lambda_c(\beta, \gamma) = (\frac{\gamma}{\beta})^{\frac{\sqrt{\beta\gamma}}{\sqrt{\beta\gamma}-1}}$. There exist constants $\ell_0 = \ell_0(\beta, \gamma, \lambda)$ and $0 < \delta = \delta(\beta, \gamma, \lambda) < 1$ such that if $k > \ell_0$, then for any $0 \leqslant \ell \leqslant k - \ell_0$, for any vertex $u \in L_\ell(v)$ with children $u_1, \cdots, u_d$, it holds that*

$$\sum_{w \in L_{k-\ell}(u)} K_u^w \leqslant (1-\delta) \max_{1 \leqslant i \leqslant d} \sum_{w \in L_{k-\ell-1}(u_i)} K_{u_i}^w.$$

These two lemmas can be proved by combining the techniques developed in [GL18, ALO24]. Compared to the proof in [ALO24] for the hardcore model, our proof needs to carefully analyze the potential function Φ and use the decay results in Lemma 15 and Lemma 16 to control the influence decay.

Proof of Lemma 25. We have $L_{k-\ell}(u) = \bigcup_{i=1}^d L_{k-\ell-1}(u_i)$ (disjoint). Fix a $w \in L_{k-\ell-1}(u_i)$, where w lies in the subtree of u_i. For each $j \neq i$, define the marginal ratio z_j^w at u_j as $z_j^w = R_{u_j}^{\sigma^w}$. For the subtree rooted at u_i, define two ratios $z_i^{w,0}$ and $z_i^{w,\infty}$ as $z_i^{w,0} = R_{u_i}^{\sigma^w \wedge w \leftarrow 0}$ and $z_i^{w,\infty} = R_{u_i}^{\sigma^w \wedge w \leftarrow \infty}$. Then, two ratios $R_u^{\sigma^w \wedge w \leftarrow 0}$ and $R_u^{\sigma^w \wedge w \leftarrow \infty}$ can be written as

$$R_u^{\sigma^w \wedge w \leftarrow 0} = F_u(z_1^w, \ldots, z_{i-1}^w, z_i^{w,0}, z_{i+1}^w, \ldots, z_d^w)$$
$$R_u^{\sigma^w \wedge w \leftarrow \infty} = F_u(z_1^w, \ldots, z_{i-1}^w, z_i^{w,\infty}, z_{i+1}^w, \ldots, z_d^w).$$

Let $y_j^w = \Phi(z_j^w)$ for $j \neq i$, $y_i^{w,0} = \Phi(z_i^{w,0})$, $y_i^{w,\infty} = \Phi(z_i^{w,\infty})$. The potential recursion is

$$\Phi(R_u^{\sigma^w \wedge w \leftarrow 0}) = (\Phi \circ F_u \circ \Phi^{-1})(y_1^w, \ldots, y_{i-1}^w, y_i^{w,0}, y_{i+1}^w, \ldots, y_d^w)$$
$$\Phi(R_u^{\sigma^w \wedge w \leftarrow \infty}) = (\Phi \circ F_u \circ \Phi^{-1})(y_1^w, \ldots, y_{i-1}^w, y_i^{w,\infty}, y_{i+1}^w, \ldots, y_d^w).$$

By definition, $K_u^w = \left|\Phi(R_u^{\sigma^w \wedge w \leftarrow 0}) - \Phi(R_u^{\sigma^w \wedge w \leftarrow \infty})\right|$. Applying the mean value theorem to the map $y_i \mapsto (\Phi \circ F_u \circ \Phi^{-1})(y_1^w, \dots, y_i^w, \dots, y_d^w)$ (with y_j^w for $j \neq i$ fixed), there exists $\tilde{y}_i^w$ between $y_i^{w,0}$ and $y_i^{w,\infty}$ such that

$$K_u^w = \left|\frac{\partial(\Phi \circ F_u \circ \Phi^{-1})}{\partial y_i}(y_1^w, \dots, \tilde{y}_i^w, \dots, y_d^w)\right| \cdot \left|y_i^{w,0} - y_i^{w,\infty}\right|.$$

Let $\tilde{z}_i^w = \Phi^{-1}(\tilde{y}_i^w)$; then $\tilde{z}_i^w$ lies between $z_i^{w,0}$ and $z_i^{w,\infty}$. Compute the partial derivative $\frac{\partial(\Phi \circ F_u \circ \Phi^{-1})}{\partial y_i}$ by the chain rule. With $\boldsymbol{z}^w = (z_1^w, \dots, z_{i-1}^w, \tilde{z}_i^w, z_{i+1}^w, \dots, z_d^w)$ we have

$$K_u^w = \frac{\phi(F_u(\boldsymbol{z}^w))}{\phi(\tilde{z}_i^w)}\left|\frac{\partial F_u}{\partial z_i}(\boldsymbol{z}^w)\right|\left|y_i^{w,0} - y_i^{w,\infty}\right| \leqslant \frac{\phi(F_u(\boldsymbol{z}^w))}{\phi(\tilde{z}_i^w)}\left|\frac{\partial F_u}{\partial z_i}(\boldsymbol{z}^w)\right| \cdot K_{u_i}^w, \tag{17}$$

where the last equation holds because $\left|y_i^{w,0} - y_i^{w,\infty}\right| = \left|\Phi(z_i^{w,0}) - \Phi(z_i^{w,\infty})\right| = K_{u_i}^w$. Summing over $w \in L_{k-\ell}(u)$, we have

$$\sum_{w \in L_{k-\ell}(u)} K_u^w \leqslant \sum_{i=1}^{d} \sum_{w \in L_{k-\ell-1}(u_i)} \frac{\phi(F_u(\boldsymbol{z}^w))}{\phi(\tilde{z}_i^w)}\left|\frac{\partial F_u}{\partial z_i}(\boldsymbol{z}^w)\right| \cdot K_{u_i}^w. \tag{18}$$

By the assumption of the lemma, $\ell \leqslant k-2$. Hence, all z_j^w for $j \neq i$ and $\tilde{z}_i^w$ are in the range $(0, \lambda)$. Using Lemma 16, we have the following bound

$$\begin{aligned}\sum_{w \in L_{k-\ell}(u)} K_u^w &\leqslant \sum_{i=1}^{d} C_{\mathrm{trl}} \cdot \lambda_u \left(\frac{\beta\lambda+1}{\lambda+\gamma}\right)^{d-1} \sum_{w \in L_{k-\ell-1}(u_i)} K_{u_i}^w \\ &\leqslant C_{\mathrm{trl}} \cdot \lambda_u d \left(\frac{\beta\lambda+1}{\lambda+\gamma}\right)^{d-1} \cdot \max_{1 \leqslant i \leqslant d} \sum_{w \in L_{k-\ell-1}(u_i)} K_{u_i}^w.\end{aligned}$$

□

Proof of Lemma 26. We start from (18). For each i, the coefficient of $K_{u_i}^w$ depends on w through $\boldsymbol{z}^w$ (every z_j^w for depends on w). If we could use a single $\boldsymbol{z} = (z_1, \dots, z_d)$ for all w, then Lemma 15 would give $\phi(F_u(\boldsymbol{z}))\sum_{i=1}^{d}\left|\frac{\partial F_u}{\partial z_i}(\boldsymbol{z})\right|\frac{1}{\phi(z_i)} < 1-\alpha$, so we can use the lemma to bound K_u^w. But here $\boldsymbol{z}^w$ depends on w. Using the technique in [ALO24], we resolve this when $\ell \leqslant k - \ell_0$ using SSM in Lemma 17.

Define the pinning τ on $L_k(v)$ such that τ fixes all vertices in $L_k(v)$ to be 0. Define

$$z_i = R_{u_i}^{\tau} \text{ and } \boldsymbol{z} = (z_1, \dots, z_d).$$

For $w \in L_{k-\ell-1}(u_i)$, the distance from w to u_i is $k - \ell - 1 \geqslant \ell_0 - 1$. By Lemma 17, $\|\boldsymbol{z}^w - \boldsymbol{z}\|_\infty \leqslant \eta$ with $\eta = A\exp(-B(\ell_0 - 1))$. Furthermore, using Lemma 17 at vertex u, $|F_u(\boldsymbol{z}^w) - F_u(\boldsymbol{z})| \leqslant \eta$. Define

$$C(\boldsymbol{a}) := \frac{\phi(F_u(\boldsymbol{a}))}{\phi(a_i)}\left|\frac{\partial F_u}{\partial z_i}(\boldsymbol{a})\right|, \qquad \text{so that} \qquad \frac{C(\boldsymbol{z}^w)}{C(\boldsymbol{z})} = \frac{\phi(F_u(\boldsymbol{z}^w))}{\phi(F_u(\boldsymbol{z}))} \cdot \frac{\phi(z_i)}{\phi(\tilde{z}_i^w)} \cdot \frac{\left|\frac{\partial F_u}{\partial z_i}(\boldsymbol{z}^w)\right|}{\left|\frac{\partial F_u}{\partial z_i}(\boldsymbol{z})\right|}. \tag{19}$$

To analyze the above ratio, we need to use the following lemma.

Lemma 27. *Recall* $\phi(x) = \min\{\frac{1}{t}, \frac{1}{x \log \frac{\lambda}{x}}\}$, *where* $t = t(\beta, \gamma, \lambda)$ *is the constant. For any two numbers* $a, b \in (0, \lambda)$ *with* $|a - b| \leqslant \eta$, *it holds that* $\frac{\phi(a)}{\phi(b)} \leqslant 1 + O_{\beta,\gamma,\lambda}(\eta)$.

Proof. Note that $x \log \frac{\lambda}{x} \leqslant \frac{\lambda}{e}$ for all $x \in (0, \lambda)$. Also note that if $t \geqslant \frac{\lambda}{e}$, then $\frac{1}{x \log \frac{\lambda}{x}} \geqslant \frac{1}{t}$ for all $x \in (0, \lambda)$. In this case, $\phi(x) = 1/t$ is a constant and the lemma holds trivially.

Let us assume $t < \frac{\lambda}{e}$. Then, there are two roots to $x \log \frac{\lambda}{x} = t$ in $(0, \lambda)$, denoted by $x_1 < x_2$. We have

$$\phi(x) = \begin{cases} \frac{1}{t} & \text{if } x \in (0, x_1], \\ \frac{1}{x \log \frac{\lambda}{x}} & \text{if } x \in (x_1, x_2), \\ \frac{1}{t} & \text{if } x \in [x_2, \lambda). \end{cases}$$

Since t is a constant depends on β, γ, λ, we have x_1 and x_2 are also constants depending on β, γ, λ. For $x \in (x_1, x_2)$, the derivative $|\phi'(x)|$ is bounded by a constant c depending only on β, γ, λ. Hence, the ratio can be bounded by

$$\frac{\phi(a)}{\phi(b)} \leqslant 1 + \frac{|\phi(a) - \phi(b)|}{\phi(b)} \leqslant 1 + \frac{c|a-b|}{C_{\min}} = 1 + O_{\beta,\gamma,\lambda}(\eta),$$

where $C_{\min} = C_{\min}(\beta, \gamma, \lambda)$ is the constant in Observation 14. □

Using Lemma 27, we can bound the first two terms in (19) as

$$\frac{\phi(F_u(z^w))}{\phi(F_u(z))} \cdot \frac{\phi(z_i)}{\phi(\tilde{z}_i^w)} = \left(1 + O_{\beta,\gamma,\lambda}(\eta)\right)^2.$$

Now, for the last term, recall that $\beta_i = \beta_{u,u_i}$ and $\gamma_i = \gamma_{u,u_i}$, we can write the ratio as

$$\frac{\left|\frac{\partial F_u}{\partial z_i}(z^w)\right|}{\left|\frac{\partial F_u}{\partial z_i}(z)\right|} = \frac{F_u(z^w)}{F_u(z)} \cdot \frac{(\beta_i z_i + 1)(z_i + \gamma_i)}{(\beta_i \tilde{z}_i^w + 1)(\tilde{z}_i^w + \gamma_i)}.$$

Let $\beta_j = \beta_{u,u_j}$ and $\gamma_j = \gamma_{u,u_j}$ for all $j \in [d]$. For two numbers $a, b \in (0, \lambda)$ and $|a - b| \leqslant \eta$, we have

$$\left(\frac{\beta_j a + 1}{a + \gamma_j}\right) \Big/ \left(\frac{\beta_j b + 1}{b + \gamma_j}\right) \leqslant 1 + \frac{(\beta_j \gamma_j - 1)|a-b|}{(a + \gamma_j)(\beta_j b + 1)} \leqslant 1 + O_{\beta,\gamma}(|a-b|) \leqslant 1 + O_{\beta,\gamma}(\eta);$$

$$\begin{aligned} \frac{(\beta_i a + 1)(a + \gamma_i)}{(\beta_i b + 1)(b + \gamma_i)} &\leqslant 1 + \frac{(\beta_i(a+b) + \beta_i \gamma_i - 1)|a-b|}{(\beta_i b + 1)(b + \gamma_i)} \\ &\leqslant 1 + \frac{(2\lambda\beta + \beta\gamma - 1)|a-b|}{\gamma} \leqslant 1 + O_{\beta,\gamma,\lambda}(\eta). \end{aligned}$$

Using the above two bounds, the last term in (19) can be bounded as

$$\frac{\left|\frac{\partial F_u}{\partial z_i}(z^w)\right|}{\left|\frac{\partial F_u}{\partial z_i}(z)\right|} \leqslant \left(1 + O_{\beta,\gamma,\lambda}(\eta)\right)^{d+1}.$$

Finally, by putting all the bounds together, we have

$$\frac{C(z^w)}{C(z)} = \frac{\phi(F_u(z^w))}{\phi(F_u(z))} \cdot \frac{\phi(z_i)}{\phi(\tilde{z}_i^w)} \cdot \frac{\left|\frac{\partial F_u}{\partial z_i}(z^w)\right|}{\left|\frac{\partial F_u}{\partial z_i}(z)\right|} \leqslant \left(1 + O_{\beta,\gamma,\lambda}(\eta)\right)^{d+3}.$$

The sum of the influence in (18) now can be bounded by

$$\begin{aligned} \sum_{w \in L_{k-\ell}(u)} K_u^w &\leqslant \sum_{i=1}^{d} \sum_{w \in L_{k-\ell-1}(u_i)} \frac{\phi(F_u(z^w))}{\phi(\tilde{z}_i^w)} \left|\frac{\partial F_u}{\partial z_i}(z^w)\right| \cdot K_{u_i}^w \\ &\leqslant \left(1 + O_{\beta,\gamma,\lambda}(\eta)\right)^{d+3} \sum_{i=1}^{d} \frac{\phi(F_u(z))}{\phi(z_i)} \left|\frac{\partial F_u}{\partial z_i}(z)\right| \sum_{w \in L_{k-\ell-1}(u_i)} K_{u_i}^w \\ &\leqslant \left(1 + O_{\beta,\gamma,\lambda}(\eta)\right)^{d+3} \left(\sum_{i=1}^{d} \frac{\phi(F_u(z))}{\phi(z_i)} \left|\frac{\partial F_u}{\partial z_i}(z)\right|\right) \cdot \left(\max_{i \in [d]} \sum_{w \in L_{k-\ell-1}(u_i)} K_{u_i}^w\right). \end{aligned}$$

For the middle term in the above formula, using Lemma 16 and Lemma 15, we have

$$\sum_{i=1}^{d} \frac{\phi(F_u(z))}{\phi(z_i)} \left|\frac{\partial F_u}{\partial z_i}(z)\right| \leqslant \min\left\{1-\alpha, C_{\mathrm{trl}} \cdot d\lambda_u \left(\frac{\beta\lambda+1}{\lambda+\gamma}\right)^{d-1}\right\} = \min\{1-\alpha, C_1 \exp(-C_2 d)\},$$

where $\alpha = \alpha(\beta,\gamma,\lambda) < 1$ is the constant in Lemma 15 and $C_{\mathrm{trl}} = C_{\mathrm{trl}}(\beta,\gamma,\lambda)$ is the constant in Lemma 16. Note that $\frac{\beta\lambda+1}{\lambda+\gamma} < 1$ and $\lambda_u \leqslant \lambda$, so the second bound is upper bounded by $dC_1 \exp(-C_2 d)$ for some constants $C_1, C_2 > 0$ depending on β,γ,λ. We can choose sufficiently large constants $d_0 = d_0(\beta,\gamma,\lambda)$ and $\ell_0 = \ell_0(\beta,\gamma,\lambda)$ such that the following holds. If $d > d_0$, we use

$$(1+O_{\beta,\gamma,\lambda}(\eta))^{d+3} \cdot dC_1 \exp(-C_2 d) \leqslant dC_1(1+O_{\beta,\gamma,\lambda}(\eta))^3 \cdot \exp((-C_2 + O_{\beta,\gamma,\lambda}(\eta))d).$$

By choosing ℓ_0 large enough, we can make sure that $\eta = A\exp(-B(\ell_0-1))$ is sufficiently small so that $-C_2 + O_{\beta,\gamma,\lambda}(\eta) < -C_2/2$. Since $d \geqslant d_0$, by taking the constant d_0 sufficiently large, the whole term is bounded by $1-\alpha^2$. If $d \leqslant d_0$, then

$$(1+O_{\beta,\gamma,\lambda}(\eta))^{d+3} \cdot (1-\alpha) \leqslant (1+O_{\beta,\gamma,\lambda}(\eta))^{d_0+3} \cdot (1-\alpha) \leqslant 1-\alpha^2,$$

where the last inequality holds by choosing ℓ_0 large enough so that η is small enough and the $(1+O_{\beta,\gamma,\lambda}(\eta))^{d_0+3}$ term is at most $1+\alpha$. Combining the two cases, the lemma holds with $\delta = \alpha^2$. □

4.2. **Proof of the influence bound.** We are now ready to prove the influence bound. Using (14), we bound the sum of the influence level by level. Fix an integer $k \geqslant 0$, to bound the sum $\sum_{w \in L_k(u)} K_v^w$, we apply Lemma 25 and Lemma 26. Formally, we first truncate the tree T and only keep levels up to k to form a new tree T_k. By definition of I_v^w, for every w, it fixes the pinning on the k-th level of T_k. Hence, we can only consider the tree T_k when analysing the influence. Using Lemma 25 and Lemma 26 recursively, we finally reach a vertex u at level $k-1$ with children $u_1, u_2, \ldots, u_d$ such that

$$\sum_{w \in L_k(u)} K_v^w \leqslant (1-\delta)^{\max\{0,k-\ell_0+1\}} \cdot \left(C_{\mathrm{trl}} \cdot \lambda_u d \left(\frac{\beta\lambda+1}{\lambda+\gamma}\right)^{d-1}\right)^{\max\{0,\min\{\ell_0-2,k-2\}\}} \cdot \sum_{i=1}^{d} K_u^{u_i}.$$

Note that $C_{\mathrm{trl}} \cdot \lambda_u d \left(\frac{\beta\lambda+1}{\lambda+\gamma}\right)^{d-1} = d\exp(-\Omega(d)) = O_{\beta,\gamma,\lambda}(1)$ can be upper bounded by a constant, and that ℓ_0, δ are the constants in Lemma 26. Hence, we can write the above inequality as

$$\sum_{w \in L_k(u)} K_v^w = O_{\beta,\gamma,\lambda}(1) \cdot (1-\delta)^k \cdot \sum_{i=1}^{d} K_u^{u_i}.$$

Finally, we bound each $K_u^{u_i}$. By definition of the influence in Definition 22, we can write the influence as

$$K_u^{u_i} = \left|\Phi(R_u^{\sigma^i \wedge u_i \leftarrow \infty}) - \Phi(R_u^{\sigma^i \wedge u_i \leftarrow 0})\right|,$$

where σ^i is a pinning on all u_j with $j \neq i$ and $\sigma^i(u_j) \in (0,\lambda)$ for all $j \neq i$. A simple calculation shows

$$\|R^{\sigma^i \wedge u_i \leftarrow \infty} - R^{\sigma^i \wedge u_i \leftarrow 0}\| \leqslant \lambda \left(\frac{\beta\lambda+1}{\lambda+\gamma}\right)^{d-1} \cdot \left(\frac{\beta\gamma-1}{\gamma}\right) = \exp(-\Omega(d)).$$

Using Lemma 24, we have

$$\sum_{i=1}^{d} K_u^{u_i} \leqslant \sum_{i=1}^{d} O_{\beta,\gamma,\lambda}(1) \cdot \|R^{\sigma^i \wedge u_i \leftarrow \infty} - R^{\sigma^i \wedge u_i \leftarrow 0}\| \leqslant O_{\beta,\gamma,\lambda}(1) \cdot d \cdot \exp(-\Omega(d)) = O_{\beta,\gamma,\lambda}(1).$$

Finally, combining (14), Lemma 24, and the above bounds, the total influence is bounded by

$$\sum_{w \in V \setminus \{v\}} D_{\mathrm{TV}}\left(\mu_v^{w \leftarrow 0}, \mu_v^{w \leftarrow 1}\right) \leqslant 2 \sum_{k \geqslant 1} \sum_{w \in L_k(v)} I_v^w \leqslant O_{\beta,\gamma,\lambda}(1) \sum_{k \geqslant 1} \sum_{w \in L_k(v)} K_v^w$$

$$\leqslant O_{\beta,\gamma,\lambda}(1)\sum_{k\geqslant 1}(1-\delta)^k = O_{\beta,\gamma,\lambda}(1).$$

5. Mixing from typical-case aggregate strong spatial mixing

The ferromagnetic two-spin systems are monotone systems. To make this notion precise, recall that μ^σ denotes the distribution of $X \sim \mu$ conditional on $X(\Lambda) = \sigma$, where $\Lambda \subseteq V$ is a subset of vertices and $\sigma \in \{0,1\}^\Lambda$ is a configuration on Λ. Define a partial ordering $\preceq$ as follows. For any $\Lambda \subseteq V$, any two configurations $\sigma, \tau \in \{0,1\}^\Lambda$,

$$\sigma \preceq \tau \quad \Leftrightarrow \quad \sigma_v \leqslant \tau_v \quad \forall v \in \Lambda. \tag{20}$$

Definition 28 (Monotone spin systems)**.** A two-spin system is said to be monotone if for any $\Lambda \subseteq V$, any two configurations $\sigma, \tau \in \{0,1\}^\Lambda$, if $\sigma \preceq \tau$, then μ^σ is stochastically dominated by μ^τ, which means that there exists a coupling (X, Y) such that $X \sim \mu^\sigma$ and $Y \sim \mu^\tau$ and $\Pr[X \preceq Y] = 1$.

As a well-known fact, any Gibbs distribution of ferromagnetic two-spin system is a monotone spin system. We provide a proof for the sake of completeness in Appendix B.

Proposition 29. *Any Gibbs distribution of ferromagnetic two-spin system is a monotone spin system.*

We study the block dynamics on two-spin systems with Gibbs distribution μ. Let $\mathcal{B} = \{B_1, B_2, \ldots, B_r\}$ be a set of blocks, where each block $B_i \subseteq V$ and $\cup_{i=1}^r B_i = V$. We consider two kinds of block dynamics: heat-bath block dynamics and systematic scan block dynamics.

Starting from an initial configuration $X \in \Omega = \{0,1\}^V$, in each step, the heat-bath block dynamics updates the current configuration X as follows:

- pick a block B uniformly at random from $\mathcal{B}$;
- resample $X(B) \sim \mu_B^{X(V\setminus B)}$, where $\mu_B^{X(V\setminus B)}$ is the marginal distribution on B induced by μ conditioned on the configuration $X(V \setminus B)$ on other variables $V \setminus B$ outside of B.

The systematic scan block dynamics updates the current configuration X as follows: for each update step,

- scan all the blocks B_i for i from 1 to r in order, and resample the configuration on B_i conditional on the current configuration of other variables: $X(B_i) \sim \mu_{B_i}^{X(V\setminus B_i)}$.

For each block B_i, let P_{B_i} denote the transition matrix of updating the configuration on B_i conditional on the current configuration of other variables. The transition matrix of heat-bath block dynamics is then

$$P_{HB} = \frac{1}{r}\sum_{i=1}^{r} P_{B_i},$$

and the transition matrix of systematic scan block dynamics is

$$P_{Scan} = P_{B_r} \cdot P_{B_{r-1}} \cdots P_{B_1}.$$

The result in this section works for both the heat-bath block dynamics and the systematic scan block dynamics. In the rest of the proof in this section, we use the phrase "block dynamics" to refer to both the heat-bath block dynamics and the systematic scan block dynamics.

As before, the mixing time of block dynamics is defined as the number of steps until the configuration X is close to the stationary distribution μ in total variation distance. Formally, let $P : \Omega \times \Omega \to [0,1]$ be the transition matrix of the block dynamics. Then, the mixing time is defined as

$$\forall \epsilon > 0, \quad t_{mix}^{P}(\epsilon) = \max_{\sigma\in\Omega} \min\left\{t \geqslant 0 : D_{TV}\left(P^t(\sigma,\cdot), \mu\right) < \epsilon\right\}.$$

Monotone systems admit monotone grand couplings. The following standard result applies to P_{HB} and P_{Scan}. For the sake of completeness, we provide a proof in Appendix B.

Proposition 30 (Monotone grand coupling of block dynamics). *Let μ be a Gibbs distribution of a ferromagnetic two-spin system on graph $G = (V, E)$. Let P be a block dynamics on μ. Then, there exists a monotone coupling function $f : \Omega \times [0, 1] \to \Omega$ such that for any $\sigma \in \Omega$, real vector $r \in [0, 1]^{n+1}$ uniformly at random, $\sigma \to \tau$ where $\tau = f(\sigma, r)$ follows the law of P. Furthermore, for any $\sigma \preceq \sigma'$, it holds that*

$$\Pr_r[f(\sigma, r) \preceq f(\sigma', r)] = 1.$$

To analyse this grand coupling, due to the monotonicity, it suffices to consider two chains starting from all-one configuration $\mathbf{1}$ and all-zero configuration $\mathbf{0}$.

Definition 31. Let $(r_t)_{t\geqslant 1}$ be a sequence of independent uniformly random real vectors in $[0, 1]^{n+1}$. Let X_0^+ be the all-ones configuration and X_0^- be the all-zeros configuration. Define the monotone coupling $(X_t^+, X_t^-)_{t\geqslant 0}$ as for any $t \geqslant 1$, $X_t^+ = f(X_{t-1}^+, r_t)$ and $X_t^- = f(X_{t-1}^-, r_t)$, where $f(\cdot, \cdot)$ is the monotone coupling function in Proposition 30.

In addition, to facilitate the analysis later, define the following censored block dynamics.

Definition 32 (Censored block dynamics). Let μ be the Gibbs distribution of a ferromagnetic two-spin system on graph $G = (V, E)$. Let $P : \Omega \times \Omega \to [0, 1]$ be the transition matrix of a block dynamics on μ with a set of blocks $\mathcal{B} = \{B_1, B_2, \dots, B_r\}$. For any subset $S \subseteq V$, any pinning $\sigma \in \{0, 1\}^{V\setminus S}$, the censored block dynamics P_S on μ_S^σ is defined as follows.

- The Markov chain starts from an arbitrary $X \in \{0, 1\}^V$ with $X(V \setminus S) = \sigma$.

For the heat-bath block dynamics, in each step,

- sample $B \in \mathcal{B}$ uniformly at random, and resample the configuration on $B \cap S$ conditional on the current configuration of other variables: $X(B \cap S) \sim \mu_{B\cap S}^{X(V\setminus(B\cap S))}$.

For the systematic scan block dynamics, in each step,

- scan all the blocks B_i for i from 1 to r in order, and resample the configuration on $B_i \cap S$ conditional on the current configuration of other variables: $X(B_i \cap S) \sim \mu_{B_i\cap S}^{X(V\setminus(B_i\cap S))}$.

The censored block dynamics P_S^{censored} only updates the configuration on S while keeping the configuration on $V \setminus S$ fixed. Intuitively, updates outside of S are "censored". During the whole process, the configuration on $V \setminus S$ is fixed as σ. Let $(X_t)_{t\geqslant 0}$ be the Markov chain generated by P_S^{censored} on μ_S^σ. As before, the mixing time of censored block dynamics P_S^{censored} on μ_S^σ is

$$\forall \epsilon > 0, \quad t_{\text{mix}}^{P_S^{\text{censored}}, \mu_S^\sigma}(\epsilon) = \max_{X_0 : X_0(V\setminus S)=\sigma} \min \left\{ t \geqslant 0 : D_{TV}\left((P_S^{\text{censored}})^t(X_0, \cdot), \mu_S^\sigma \right) < \epsilon \right\}.$$

The key to our proof is the notion of good neighbourhood and boundary conditions, which facilitates typical-case ASSM. Let $S \subseteq V$ be a subset of vertices. The outer boundary ∂S of S is the set of vertices $v \in V \setminus S$ such that there exists an edge $\{u, v\} \in E$ with $u \in S$.

Definition 33. For any $v \in V$, we call a neighbourhood $S_v \ni v$ and a set of boundary conditions $\Omega_{\partial S_v} \subseteq \{0, 1\}^{\partial S_v}$ *good* with local mixing time T_{local} if the following three properties hold:

- **Closed under shortest paths.** For any $\sigma, \tau \in \Omega_{\partial S_v}$, there exists a path of good boundary configurations $\eta_0, \eta_1, \dots, \eta_t \in \Omega_{\partial S_v}$ such that $\eta_0 = \sigma$, $\eta_t = \tau$, and for any $1 \leqslant i \leqslant t$, η_i and η_{i+1} differ only at one vertex, where $t = |\{\sigma(u) \neq \tau(u) : u \in \partial S_v\}|$ is the Hamming distance between σ and τ.
- **ASSM under good boundary conditions.** For any $u \in \partial S_v$, define the influence of u on v as

$$a_u := \max_{\sigma \in \Omega_{\partial S_v}} D_{TV}\left(\mu_v^{\sigma^{u\leftarrow 0}}, \mu_v^{\sigma^{u\leftarrow 1}} \right), \tag{21}$$

where $\sigma^{u\leftarrow c}$ denotes the configuration on ∂S_v obtained from σ by changing the value of u to c. Then, the following aggregate strong spatial mixing (ASSM) property holds

$$\sum_{u\in\partial S_v} a_u \leqslant \frac{1}{20}. \tag{22}$$

- **Local mixing.** For any outside configuration $\sigma \in \{0,1\}^{V\setminus S_v}$, the censored block dynamics $P^{\text{censored}}_{S_v}$ on $\mu^{\sigma}_{S_v}$ has mixing time $t^{P^{\text{censored}}_{S_v},\mu^{\sigma}_{S_v}}_{\text{mix}}(\frac{1}{4e}) \leqslant T_{\text{local}}$.

Now we are ready to show the main theorem of this section.

Theorem 34. *Let μ be the Gibbs distribution of a ferromagnetic two-spin system on graph $G = (V, E)$. Let P be a block dynamics on μ with a set $\mathcal{B}$ of blocks. Let $T_{\text{local}} > 0$ and $T_{\text{burn-in}} > 0$ be two integers. Suppose for any $v \in V$, there exists $S_v \subseteq V$ and $\Omega_{\partial S_v} \subseteq \{0,1\}^{\partial S_v}$ such that*

- *$(S_v, \Omega_{\partial S_v})$ is good with local mixing time T_{local} as in Definition 33;*
- *the monotone coupling $(X^+_t, X^-_t)_{t\geqslant 0}$ of P in Definition 31 satisfies that for any $t \geqslant T_{\text{burn-in}}$,*

$$\Pr[X^+_t(\partial S_v) \notin \Omega_{\partial S_v} \vee X^-_t(\partial S_v) \notin \Omega_{\partial S_v}] \leqslant \frac{1}{n^3}, \tag{23}$$

where $n = |V|$ is the number of vertices.

Then the mixing time of block dynamics P is bounded by

$$t^{P}_{\text{mix}}\left(\frac{1}{4e}\right) = O\left(T_{\text{burn-in}} + T_{\text{local}} \cdot \max_{v\in V} \log|R_v| \cdot \log n\right), \quad \textit{where } R_v = S_v \cup \partial S_v. \tag{24}$$

In (24) we set $\epsilon = 1/(4e)$ for convenience later. It is standard to extend it to general $\epsilon > 0$. The proof of Theorem 34 follows similar lines as in [MS13].

Proof of Theorem 34. Let $(X^+_t, X^-_t)_{t\geqslant 0}$ be the monotone coupling of P in Definition 31. Define $T_{\text{phase}} := T_{\text{local}} \cdot \max_{v\in V} \log(20|R_v|)$. We show that for any integer $k \geqslant 1$, it holds that

$$\begin{aligned} &\max_{v\in V} \Pr\left[X^+_{T_{\text{burn-in}}+(k+1)\cdot T_{\text{phase}}}(v) \neq X^-_{T_{\text{burn-in}}+(k+1)\cdot T_{\text{phase}}}(v)\right] \\ &\leqslant \frac{1}{2}\max_{v\in V} \Pr\left[X^+_{T_{\text{burn-in}}+k\cdot T_{\text{phase}}}(v) \neq X^-_{T_{\text{burn-in}}+k\cdot T_{\text{phase}}}(v)\right] + \frac{1}{n^2}. \end{aligned} \tag{25}$$

Solving the recursion in (25), after $T := T_{\text{burn-in}} + O(T_{\text{phase}} \log n)$ steps,

$$\max_{v\in V} \Pr\left[X^+_T(v) \neq X^-_T(v)\right] \leqslant \left(\frac{1}{2}\right)^{O(\log n)} + \frac{2}{n^2} \leqslant \frac{3}{n^2}.$$

By a union bound over all $v \in V$, it holds that $\Pr[X^+_T \neq X^-_T] \leqslant \frac{3}{n} \leqslant \frac{1}{4e}$. This holds for two chains starting from the all-one configuration $\mathbf{1}$ and all-zero configuration $\mathbf{0}$. By monotonicity, namely Proposition 30, starting from an arbitrary pair of initial configurations, the two chains can be coupled successfully with probability at least $1 - \frac{1}{4e}$. Therefore, by the standard coupling argument, the mixing time bound in (24) is proved. Our task is reduced to verify the recursion in (25).

Fix an integer $k \geqslant 0$. Let $s = k \cdot T_{\text{phase}} + T_{\text{burn-in}}$. Fix a vertex $v \in V$ and the corresponding region $S_v \subseteq V$. We construct another two instances of Markov chains $(Y^+_j, Y^-_j)_{j\geqslant 0}$ by the following process:

- for $0 \leqslant j \leqslant s$, let $(Y^+_j, Y^-_j) = (X^+_j, X^-_j)$;
- for $j > s$, the two processes $Y^+_{j-1} \to Y^+_j$ and $Y^-_{j-1} \to Y^-_j$ both follow the transition rule of the censored block dynamics $P^{\text{censored}}_{S_v}$.

For two random variables X and Y over $\{0,1\}^V$, we say that the distribution of X is stochastically dominated by the distribution of Y, denoted by $X \preceq_D Y$, if there exists a coupling (X, Y) such that $X \preceq Y$ with probability 1, where the partial order $\preceq$ is defined in (20). The following result holds for the censored block dynamics $P^{\text{censored}}_{S_v}$. A similar result appeared in [BCV20, Theorem 7]. For the sake of completeness, we provide a brief proof in Appendix B.

Claim 35. *The following stochastic dominance relations hold:*

$$\forall j \geq 0, \quad Y^-_j \preceq_D X^-_j \preceq_D X^+_j \preceq_D Y^+_j.$$

Claim 35 states a stochastic dominance relation among four random variables $X^-_j, X^+_j, Y^-_j, Y^+_j$. The statement itself only involves the marginal distribution of four random variables. For instance, the distribution of Y^-_j is stochastic dominated by the distribution of X^-_j. The claim itself states nothing about the joint distribution of four random variables.

Recall that $R_v = S_v \cup \partial S_v$. Let $t = s + T_{\text{phase}}$. Since $(X^+_t, X^-_t)_{t \geq 0}$ forms a monotone coupling, we have $X^+_t(v) \geq X^-_t(v)$ with probability 1. To upper bound the probability of $X^+_t(v) \neq X^-_t(v)$, we only need to upper bound $\Pr[X^+_t(v) = 1] - \Pr[X^-_t(v) = 1]$. The stochastic dominance relations in Claim 35 shows

$$\Pr[X^-_t(v) = 1] \geq \Pr[Y^-_t(v) = 1] \text{ and } \Pr[X^+_t(v) = 1] \leq \Pr[Y^+_t(v) = 1].$$

Therefore, we have the following upper bound:

$$\Pr[X^+_t(v) \neq X^-_t(v)] = \Pr[X^+_t(v) = 1] - \Pr[X^-_t(v) = 1] \leq \Pr[Y^+_t(v) = 1] - \Pr[Y^-_t(v) = 1]. \tag{26}$$

For any two configurations $\sigma^+, \sigma^- \in \{0,1\}^{R_v}$, let $\mathcal{C}(\sigma^+, \sigma^-)$ be the event $X^+_s(R_v) = \sigma^+$ and $X^-_s(R_v) = \sigma^-$. We only consider σ^+, σ^- such that $\mathcal{C}(\sigma^+, \sigma^-)$ happens with a positive probability. For $t > s$, we will upper bound the difference between the probabilities of $Y^+_t(v) = 1$ and $Y^-_t(v) = 1$ conditioned on $\mathcal{C}(\sigma^+, \sigma^-)$. Let $\tau^+ = \sigma^+(\partial S_v)$ and $\tau^- = \sigma^-(\partial S_v)$ be the configurations on the boundary ∂S_v induced by σ^+ and σ^- respectively. We also define $\mathcal{C}(\tau^+, \tau^-)$ be the event $X^+_s(\partial S_v) = \tau^+$ and $X^-_s(\partial S_v) = \tau^-$. By the triangle inequality, we have

$$\begin{aligned}
&\left|\Pr[Y^+_t(v) = 1 \mid \mathcal{C}(\sigma^+, \sigma^-)] - \Pr[Y^-_t(v) = 1 \mid \mathcal{C}(\sigma^+, \sigma^-)]\right| \\
\leq &\left|\Pr[Y^+_t(v) = 1 \mid \mathcal{C}(\sigma^+, \sigma^-)] - \mu^{\tau^+}_v(1)\right| + \left|\mu^{\tau^+}_v(1) - \mu^{\tau^-}_v(1)\right| \\
&+ \left|\Pr[Y^-_t(v) = 1 \mid \mathcal{C}(\sigma^+, \sigma^-)] - \mu^{\tau^-}_v(1)\right|,
\end{aligned} \tag{27}$$

By the law of total probability and the triangle inequality, the probability of $Y^+_t(v) \neq Y^-_t(v)$ is at most

$$\begin{aligned}
&\Pr[Y^+_t(v) = 1] - \Pr[Y^-_t(v) = 1] \\
= &\sum_{(\sigma^+, \sigma^-): \sigma^+ \neq \sigma^-} \Pr[\mathcal{C}(\sigma^+, \sigma^-)] \cdot \left(\Pr[Y^+_t(v) = 1 \mid \mathcal{C}(\sigma^+, \sigma^-)] - \Pr[Y^-_t(v) = 1 \mid \mathcal{C}(\sigma^+, \sigma^-)]\right) \\
\leq &\sum_{(\sigma^+, \sigma^-): \sigma^+ \neq \sigma^-} \Pr[\mathcal{C}(\sigma^+, \sigma^-)] \cdot \left|\Pr[Y^+_t(v) = 1 \mid \mathcal{C}(\sigma^+, \sigma^-)] - \Pr[Y^-_t(v) = 1 \mid \mathcal{C}(\sigma^+, \sigma^-)]\right|
\end{aligned} \tag{28}$$

Note that the sum above enumerates only pairs of distinct feasible boundary configurations $\sigma^+, \sigma^- \in \{0,1\}^{R_v}$, namely $\sigma^+ \neq \sigma^-$. This is because, when $\sigma^+ = \sigma^-$, using the conditional independence property of spin systems, two Markov chains $Y^+_t(v)$ and $Y^-_t(v)$ are exactly the same stochastic processes (the same starting configuration and same transition matrix) inside S_v, and therefore $\Pr[Y^+_t(v) = 1 \mid \mathcal{C}(\sigma^+, \sigma^-)] =$

$\Pr[Y_t^-(v)=1 \mid \mathcal{C}(\sigma^+,\sigma^-)]$. Combining (27) and (28), we have

$$
\begin{aligned}
&\Pr[Y_t^+(v)=1]-\Pr[Y_t^-(v)=1]\\
\leqslant & \sum_{(\sigma^+,\sigma^-):\sigma^+\neq\sigma^-}\Pr[\mathcal{C}(\sigma^+,\sigma^-)]\left|\Pr[Y_t^+(v)=1 \mid \mathcal{C}(\sigma^+,\sigma^-)]-\mu_v^{\tau^+}(1)\right|\\
&+\sum_{(\sigma^+,\sigma^-):\sigma^+\neq\sigma^-}\Pr[\mathcal{C}(\sigma^+,\sigma^-)]\left|\mu_v^{\tau^+}(1)-\mu_v^{\tau^-}(1)\right|\\
&+\sum_{(\sigma^+,\sigma^-):\sigma^+\neq\sigma^-}\Pr[\mathcal{C}(\sigma^+,\sigma^-)]\left|\Pr[Y_t^-(v)=1 \mid \mathcal{C}(\sigma^+,\sigma^-)]-\mu_v^{\tau^-}(1)\right|.
\end{aligned}
\tag{29}
$$

Consider the first and the third terms in (29). Note that $Y_t^+(v)$ and $Y_t^-(v)$ both follow the censored transition matrix $P_{S_v}^{\text{censored}}$. The configuration outside S_v is fixed in the censored process, and the configuration inside S_v converges to the conditional marginal distribution $\mu_v^{\tau^+}$ and $\mu_v^{\tau^-}$ respectively. Therefore, by the local mixing property of Definition 33 and since $t-s=T_{\text{local}}\cdot\max_{v\in V}\log(20|R_v|)$, by (5),

$$
\forall \sigma^+,\sigma^-\in\{0,1\}^{R_v},\quad
\begin{aligned}
&\left|\Pr[Y_t^+(v)=1 \mid \mathcal{C}(\sigma^+,\sigma^-)]-\mu_v^{\tau^+}(1)\right|\leqslant\frac{1}{20|R_v|};\\
&\left|\Pr[Y_t^-(v)=1 \mid \mathcal{C}(\sigma^+,\sigma^-)]-\mu_v^{\tau^-}(1)\right|\leqslant\frac{1}{20|R_v|}.
\end{aligned}
$$

Therefore the first and the third terms in (29) can be bounded by

$$
\begin{aligned}
&\sum_{(\sigma^+,\sigma^-):\sigma^+\neq\sigma^-}\Pr[\mathcal{C}(\sigma^+,\sigma^-)]\left|\Pr[Y_t^+(v)=1 \mid \mathcal{C}(\sigma^+,\sigma^-)]-\mu_v^{\tau^+}(1)\right|\\
&\quad+\sum_{(\sigma^+,\sigma^-):\sigma^+\neq\sigma^-}\Pr[\mathcal{C}(\sigma^+,\sigma^-)]\left|\Pr[Y_t^-(v)=1 \mid \mathcal{C}(\sigma^+,\sigma^-)]-\mu_v^{\tau^-}(1)\right|\\
\leqslant&\sum_{(\sigma^+,\sigma^-):\sigma^+\neq\sigma^-}\frac{\Pr[\mathcal{C}(\sigma^+,\sigma^-)]}{20|R_v|}+\sum_{(\sigma^+,\sigma^-):\sigma^+\neq\sigma^-}\frac{\Pr[\mathcal{C}(\sigma^+,\sigma^-)]}{20|R_v|}\\
=&\frac{\Pr[X_s^+(R_v)\neq X_s^-(R_v)]}{10|R_v|}\leqslant\frac{\max_{u\in V}\Pr[X_s^+(u)\neq X_s^-(u)]}{10}.
\end{aligned}
\tag{30}
$$

To bound the second term in (29), we first have that

$$
\begin{aligned}
&\sum_{(\sigma^+,\sigma^-):\sigma^+\neq\sigma^-}\Pr[\mathcal{C}(\sigma^+,\sigma^-)]\left|\mu_v^{\tau^+}(1)-\mu_v^{\tau^-}(1)\right|\\
=&\sum_{(\tau^+,\tau^-)\in\{0,1\}^{\partial S_v}\times\{0,1\}^{\partial S_v}:\tau^+\neq\tau^-}\Pr[\mathcal{C}(\tau^+,\tau^-)]\left|\mu_v^{\tau^+}(1)-\mu_v^{\tau^-}(1)\right|,
\end{aligned}
$$

because whenever $\tau^+=\sigma^+(\partial S_v)=\sigma^-(\partial S_v)=\tau^-$, it holds that $\mu_v^{\tau^+}(1)=\mu_v^{\tau^-}(1)$. We then construct a path $\eta_0,\eta_1,\ldots,\eta_t\in\{0,1\}^{\partial S_v}$ such that $\eta_0=\tau^+$, $\eta_t=\tau^-$, and for any $1\leqslant i\leqslant t$, η_i and η_{i+1} differ only at one vertex, where $t=\{\tau^+(u)\neq\tau^-(u):u\in\partial S_v\}$ is the Hamming distance between τ^+ and τ^-. There are two cases depending on whether both τ^+ and τ^- are in $\Omega_{\partial S_v}$. If so, by the first property of Definition 33, we can further assume that $\eta_i\in\Omega_{\partial S_v}$ for all $0\leqslant i\leqslant t$. Then,

$$
\begin{aligned}
&\sum_{\tau^+\neq\tau^-}\Pr[\mathcal{C}(\tau^+,\tau^-)]\left|\mu_v^{\tau^+}(1)-\mu_v^{\tau^-}(1)\right|\leqslant\sum_{\tau^+\neq\tau^-}\Pr[\mathcal{C}(\tau^+,\tau^-)]\sum_{i=1}^{t}\left|\mu_v^{\eta_{i-1}}(1)-\mu_v^{\eta_i}(1)\right|\\
\leqslant&\sum_{\tau^-\neq\tau^+}\Pr[\mathcal{C}(\tau^+,\tau^-)]\sum_{u\in\partial S_v}\mathbb{1}\{\tau^+(u)\neq\tau^-(u)\}(\mathbb{1}\{\tau^+,\tau^-\in\Omega_{\partial S_v}\}a_u+\mathbb{1}\{\tau^+\text{ or }\tau^-\notin\Omega_{\partial S_v}\}\cdot 1),
\end{aligned}
$$

where in the last inequality, we split the two cases. If both τ^+ and τ^- are in $\Omega_{\partial S_v}$, then $\eta_i \in \Omega_{\partial S_v}$ for all $0 \leqslant i \leqslant t$. It implies that the difference between $\mu_v^{\eta_{i-1}}(1)$ and $\mu_v^{\eta_i}(1)$ is at most a_u, where u is the vertex that η_{i-1} and η_i differ on and a_u is defined in (21). Otherwise τ^+ or τ^- is not in $\Omega_{\partial S_v}$, then the difference between $\mu_v^{\eta_{i-1}}(1)$ and $\mu_v^{\eta_i}(1)$ is at most 1. Rearranging the terms, we have

$$\begin{aligned}
&\sum_{u\in\partial S_v} a_u \cdot \sum_{\tau^+\neq\tau^-} \Pr[\mathcal{C}(\tau^+,\tau^-)] \cdot \mathbb{1}\{\tau^+(u)\neq\tau^-(u)\} \cdot \mathbb{1}\{\tau^+,\tau^- \in \Omega_{\partial S_v}\} \\
&\quad + \sum_{u\in\partial S_v} \sum_{\tau^+\neq\tau^-} \Pr[\mathcal{C}(\tau^+,\tau^-)] \cdot \mathbb{1}\{\tau^+(u)\neq\tau^-(u)\} \cdot \mathbb{1}\{\tau^+ \text{ or } \tau^- \notin \Omega_{\partial S_v}\} \\
&\leqslant \sum_{u\in\partial S_v} a_u \Pr\left[X_s^+(u)\neq X_s^-(u)\right] + \sum_{u\in\partial S_v} \Pr\left[X_s^-(\partial S_v)\notin\Omega_{\partial S_v} \text{ or } X_s^+(\partial S_v)\notin\Omega_{\partial S_v}\right] \\
&\leqslant \max_{u\in V} \Pr[X_s^+(u)\neq X_s^-(u)] \sum_{u\in\partial S_v} a_u + |\partial S_v| \cdot \frac{1}{n^3},
\end{aligned}$$

where we used $\mathbb{1}\{\tau^+,\tau^- \in \Omega_{\partial S_v}\} \leqslant 1$ in the first inequality, and the condition in (23) in the second. Finally, the sum of a_u can be bounded by the ASSM property in Definition 33. As $|\partial S_v| \leqslant n$, the second term in (29) can be bounded by

$$\sum_{(\sigma^+,\sigma^-):\sigma^+\neq\sigma^-} \Pr[\mathcal{C}(\sigma^+,\sigma^-)] \left|\mu_v^{\tau^+}(1) - \mu_v^{\tau^-}(1)\right| \leqslant \frac{\max_{u\in V}\Pr[X_s^+(u)\neq X_s^-(u)]}{20} + \frac{1}{n^2}. \tag{31}$$

Combining (26), (29), (30), and (31), we have for all $v \in V$,

$$\Pr[X_t^+(v)\neq X_t^-(v)] \leqslant \frac{3\max_{u\in V}\Pr[X_s^+(u)\neq X_s^-(u)]}{20} + \frac{1}{n^2}.$$

Taking the maximum over $v \in V$ proves (25). ☐

Remark (Relaxing the local mixing condition). In Definition 33, the local mixing condition is assumed for an arbitrary outside configuration $\sigma \in \{0,1\}^{V\setminus S_v}$. For applications considered in this paper, we can verify this strong assumption of local mixing. However, the proof technique above works fine with a relaxed condition of local mixing, where we consider only $\sigma \in \{0,1\}^{V\setminus S_v}$ such that $\sigma(\partial S_v) \in \Omega_{\partial S_v}$ instead of an arbitrary outside configuration. The mixing result in Theorem 34 still holds under this relaxed local mixing condition.

6. Construct the good neighbourhood

In this section, we show how to construct the good neighbourhood. Let $G = (V,E)$ be a graph. For any $v \in V$ we construct the good neighbourhood $S_v \subseteq V$ such that $v \in S_v$. We first need some definitions. Recall Definition 9, the SAW tree. Let $\mathrm{cld}_T(u)$ be the set of children of u in a tree T. For a SAW tree $T = T_{\mathrm{SAW}}(G,v,\partial S_v)$ rooted at v, and for any vertex $u \in T$ that is a copy of some vertex in S_v, define

$$F_T(u) := |\{w \in \mathrm{cld}_T(u) : w \text{ is a copy of some vertex in } S_v \text{ and } w \text{ is not a cycle-closing vertex in } T\}|. \tag{32}$$

Lemma 36. *Let $G = (V,E)$ be a graph. Let $1 \leqslant D_1 \leqslant D_2$ be two integer parameters. For any vertex $v \in V$, there exists $S_v \subseteq V$ with $v \in S_v$ such that $|S_v| \leqslant \exp(D_1)\cdot D_2$ and the following property holds for the SAW tree $T = T_{\mathrm{SAW}}(G,v,\partial S_v)$. For any leaf vertex w in T such that w is a copy of some vertex in ∂S_v, at least one of the following two conditions holds:*

- *Let $v = u_1, u_2, \cdots, u_k, w$ be the path from the root v to w in T, where $k \geqslant 1$ is the distance between v and w in T. It holds that $\sum_{i=1}^{k-1} F_T(u_i) \geqslant D_1$;*
- *there exists an ancestor u of w such that the number of non-cycle-closing children of u is at least D_2.*

Proof of Lemma 36. Fix $v \in V$ and we construct the region S_v as follows. Consider the SAW tree $T_\emptyset = T_{\mathrm{SAW}}(G, v, \emptyset)$. By removing all cycle-closing vertices in $T_\emptyset$, we obtain a tree T'. We use a DFS starting from the root v to first construct a region Q_v as in Algorithm 1. (Algorithm 1 is the same as the procedure for trees described in Section 2.3.) In the algorithm, for each vertex $u \in T'$,

$$\mathrm{degsum}(u) := \sum_{w \in \mathrm{path}(v,u)} |\mathrm{cld}_{T'}(w)|,$$

where $\mathrm{path}(v, u)$ is the set of vertices on the path from v to u in T', including v and u.

Algorithm 1: Construction of the region Q_v

1 Initialize $Q_v = \emptyset$;
2 DFS(v);
3 **return** Q_v;
4 **Procedure** DFS(u)
5     $Q_v \leftarrow Q_v \cup \{u\}$;
6   **if** u *is a leaf in* T' **then**
7   **return**;
8   **else if** *degsum*$(u) \geqslant D_1$ **then**
9   **if** $|\mathit{cld}_{T'}(u)| < D_2$ **then**
10             $Q_v \leftarrow Q_v \cup \mathrm{cld}_{T'}(u)$;
11   **return**;
12   **else**
13   **for** *each child* w *of* u **do**
14   DFS(w);

After constructing Q_v by Algorithm 1, define

$$S_v := \{u \in G : \exists u' \in Q_v \text{ such that } u' \text{ is a copy of } u\}.$$

Let $T = T_{\mathrm{SAW}}(G, v, \partial S_v)$ be the SAW tree rooted at v with boundary ∂S_v. We first show that for each $w \in T$ that is a copy of some vertex in ∂S_v, at least one of the two conditions in the lemma holds.

Let $v = u_1, u_2, \cdots, u_k, w$ be the path from the root v to w in T, where k is the distance between v and w. If there exists an ancestor u of w such that the number of non-cycle-closing children of u in T is at least D_2, then the second condition holds. Otherwise, for any ancestor u_i of w, it must hold that the number of non-cycle-closing children of any u_i is less than D_2 in T. Recall the tree T' obtained from $T_\emptyset = T_{\mathrm{SAW}}(G, v, \emptyset)$ by removing all cycle-closing vertices. Since none of the $\{u_i\}_{i\in[k]}$ is cycle-closing, the path $v = u_1, u_2, \cdots, u_k$ must be present in T' as well. Since u_i is not a leaf vertex in T, it has the same set of children in T as in $T_\emptyset$. Hence, the non-cycle-closing children of u_i in T are exactly the children of u_i in T'. Therefore, $|\mathrm{cld}_{T'}(u_i)| < D_2$ for all $1 \leqslant i \leqslant k$. Consider the DFS procedure in T'. When we do the DFS along the path $u_1, u_2, \cdots, u_k$ in T', the DFS procedure must stop at some u_j for $1 \leqslant j \leqslant k-1$ because:

- the DFS procedure must have stopped at some u_j for $1 \leqslant j \leqslant k$. Otherwise, w is added to Q_v and then w cannot be a copy of some vertex in ∂S_v;
- furthermore, the DFS procedure cannot stop at u_k. Otherwise, since u_k is not a leaf vertex in T', u_k must satisfy the condition in Line 8. Note that $\mathrm{cld}_{T'}(u_k) < D_2$. Then, all children of u_k, including w, are added to the set Q_v. This contradicts the assumption that w is a copy of some vertex in ∂S_v.

Note that the DFS procedure can stop only when it reaches the condition in Line 8, because $u_1, u_2, \cdots, u_{k-1}$ are not leaves in T'. Furthermore, since $|\mathrm{cld}_{T'}(u_i)| < D_2$ for all $1 \leqslant i \leqslant k-1$, the DFS procedure can stop only after executing Line 10 at some u_j for $1 \leqslant j \leqslant k-1$. Therefore,

$$\sum_{i=1}^{k-1} F_T(u_i) \geqslant \sum_{i=1}^{j} F_T(u_i) = \mathrm{degsum}(u_j) \geqslant D_1,$$

where the equality follows from the fact that all children of u_i in T' are added to Q_v for $1 \leqslant i \leqslant j$ (for $i < j$, we run DFS on all children of u_i, and for $i = j$, since $\left|\mathrm{cld}_{T'}(u_j)\right| < D_2$, we add all children of u_j to Q_v directly). Hence, all children of u_i in T' are copies of some vertices in S_v, and none of them is cycle-closing by the definition of T'. Therefore, for each $1 \leqslant i \leqslant j$, we have $F_T(u_i) = |\mathrm{cld}_{T'}(u_i)|$, which gives the equality above. This implies that the first condition holds.

Finally, we bound the size of S_v. Since there is a surjection from Q_v to S_v, we have $|S_v| \leqslant |Q_v|$. Consider the following optimisation problem. Let $g(m)$ be the maximum number of vertices in a tree T_0 such that: for any leaf u in T_0,

$$\sum_{w \in \mathrm{path}(v,u), w \neq u} |\mathrm{cld}_{T_0}(w)| < m.$$

In other words, $g(m)$ denotes the size of the largest tree T_0 satisfying the condition above with parameter m. By definition, $g(1) = 1$. We claim that the following recursive relation holds:

$$g(m) = \max_{d \in [1, m-1]} \{1 + d \cdot g(m-d)\}.$$

Indeed, for any tree T_0 satisfying the requirement with parameter m, let d be the number of children of the root v in T_0. Then each subtree rooted at a child of v satisfies the same condition with parameter $m - d$, and hence each such subtree contains at most $g(m-d)$ vertices.

We prove $g(m) \leqslant \exp(m)$ by induction. The base case $g(1) = 1 \leqslant \exp(1)$ holds. Assume $g(m') \leqslant \exp(m')$ for all $m' < m$. Then

$$\begin{aligned} \forall d \in [1, m-1], \quad 1 + d \cdot g(m-d) &\leqslant 1 + d \cdot \exp(m-d) \\ &\leqslant 1 + (\exp(d) - 1)\exp(m-d) \\ &= \exp(m) + 1 - \exp(m-d) \leqslant \exp(m), \end{aligned}$$

where we use $\exp(d) \geqslant 1 + d$ for all $d \in \mathbb{R}$ in the second inequality.

Back to the size of $|Q_v|$. If we omit the children added to Q_v in Line 10, then the remaining DFS tree has at most $g(D_1) \leqslant \exp(D_1)$ vertices by the optimisation problem analysed above. Each time Line 10 is executed, at most D_2 children are added to Q_v, and these added vertices do not trigger further DFS calls. Since Line 10 can be executed at most once for each vertex in the remaining DFS tree, we obtain

$$|Q_v| \leqslant g(D_1) \cdot D_2 \leqslant \exp(D_1) \cdot D_2.$$

Therefore, $|S_v| \leqslant |Q_v| \leqslant \exp(D_1) \cdot D_2$. □

7. Reducing the ASSM property from graphs to SAW trees

In this section, we verify the conditions in Definition 33 for the neighbourhood constructed in Lemma 36. Fix a vertex $v \in V$ in the graph $G = (V, E)$. Let $S_v \subseteq V$ be the region constructed by Lemma 36 with the following parameters

$$D_1 := C_D \cdot \log\log n, \quad D_2 := (\log n)^3, \tag{33}$$

where $C_D = C_D(\beta, \gamma, \lambda)$ is a constant depending on β, γ, λ. The value of C_D will be determined in (45). Recall ∂S_v, the outer boundary of S_v. For any $u \in S_v$, define the boundary-neighbors of u as

$$N^G_{\partial S_v}(u) := \{w \in \partial S_v : (u, w) \in E\}. \tag{34}$$

Definition 37 (Good boundary condition)**.** We say a configuration $\sigma \in \{0,1\}^{\partial S_v}$ is *good* if for any $u \in S_v$ with $|N^G_{\partial S_v}(u)| > D_2/3$, it satisfies

$$|\{w \in N^G_{\partial S_v}(u) : \sigma(w) = 1\}| \geqslant |N^G_{\partial S_v}(u)|/(\log n) + 2.$$

Let $\Omega_{\partial S_v}$ denote the set of all good boundary conditions.

Good boundary conditions admit typical-case ASSM.

Lemma 38. *Let $\beta \leqslant 1 < \gamma$, $\beta\gamma > 1$ and $\lambda < \lambda_0(\beta,\gamma) := \sqrt{\gamma/\beta}$ be three constants. Let μ be the Gibbs distribution of a (β,γ,λ)-ferromagnetic two-spin system with parameters $(\beta_e,\gamma_e)_{e\in E},(\lambda_v)_{v\in V}$ on $G = (V,E)$ as in Definition 2. For any $u \in \partial S_v$, let a_u be the influence of u on v in the distribution μ, defined as in (21), where the boundary condition set $\Omega_{\partial S_v}$ is given by Definition 37. Then, $\sum_{u\in\partial S_v} a_u \leqslant \frac{1}{20}$.*

In this section, we carry out the first step in the proof of Lemma 38. We reduce the problem to verifying a similar ASSM statement on the SAW tree T instead of on the original graph G. Next, in Section 8, we prove the ASSM property on T. Lemma 38 follows from combining the two steps.

Let $\sigma \in \{0,1\}^{\partial S_v}$ be a good boundary condition. Let $T = T_{SAW}(G,v,\sigma) = (V_T,E_T)$ be the SAW tree with boundary ∂S_v defined in Definition 12. We first recall some notation and background on the SAW tree T. Let $\Gamma \subseteq V_T$ be the set of cycle-closing leaf vertices of T, and let ρ_Γ be the pinning on Γ. Let Λ be the set of all leaf vertices in T that are copies of vertices in ∂S_v. Let σ_Λ be the pinning on Λ inherited from σ. We use $\bar\sigma := \rho_\Gamma \cup \sigma_\Lambda$ to denote the total pinning on $\Gamma\cup\Lambda$. Note that all vertices in $\Gamma\cup\Lambda$ are leaves of T. Let π be the Gibbs distribution on T obtained by inheriting the parameters of μ on G. By Proposition 13, the marginals μ^σ_v and $\pi^{\bar\sigma}_v$ are identical.

We next prune the SAW tree T by removing all cycle-closing leaf vertices. Using the self-reducibility property in Observation 8, we can remove all cycle-closing leaf vertices from T and modify the external fields at their neighbors accordingly. From now on, we use $T = (V_T,E_T)$ to denote the pruned SAW tree and π to denote the Gibbs distribution on this pruned tree. Note that π is still a Gibbs distribution of a (β,γ,λ)-ferromagnetic two-spin system on $T = (V_T,E_T)$.

As in (22), we want to prove $\sum_{u\in\partial S_v} a_u \leqslant \frac{1}{20}$, where $a_u = \max_{\sigma\in\Omega_{\partial S_v}} D_{TV}\left(\mu_v^{\sigma^{u\leftarrow 0}},\mu_v^{\sigma^{u\leftarrow 1}}\right)$. Our goal is to reduce this to verifying a similar ASSM statement on T. For this purpose, we extend the definitions of boundary-neighbors and good boundary conditions from the graph G to the SAW tree T. For every vertex $w \in V_T \setminus \Lambda$, similar to (34), define

$$N^T_\Lambda(w) := \{u \in \Lambda : \{w,u\} \in E_T\}.$$

Intuitively, one can view Λ as the boundary of T. Then $N^T_\Lambda(w)$ is the set of boundary-neighbors of w in T. We next define a good boundary condition on T. Note that in the pruned SAW tree T, the pinning is defined only on Λ, because Γ has been removed. We introduce the following notion of a good boundary condition for the SAW tree T, analogous to Definition 37.

Definition 39 (Good boundary for the SAW tree)**.** We say a configuration $\tau \in \{0,1\}^\Lambda$ is a good boundary condition if for any $w \notin \Lambda$ with $|N^T_\Lambda(w)| > D_2/3$, it satisfies

$$|\{u \in N^T_\Lambda(w) : \tau(u) = 1\}| \geqslant |N^T_\Lambda(w)|/(\log n) + 1. \tag{35}$$

We use Ω_Λ to denote the set of all good boundary conditions on T.

Finally, for any vertex $w \in \Lambda$, define the influence of w on v in the distribution π by

$$b_w = \max_{\tau\in\Omega_\Lambda} D_{TV}\left(\pi_v^{\tau^{w\leftarrow 0}},\pi_v^{\tau^{w\leftarrow 1}}\right). \tag{36}$$

We show the following relationship between the influence bounds in G and T.

Lemma 40. *The influence bounds in* G *and* T *satisfy*

$$\sum_{u \in \partial S_v} a_u \leqslant \sum_{w \in \Lambda} b_w.$$

Proof. Since $\sum_{w\in\Lambda} b_w = \sum_{u\in\partial S_v}\sum_{w\in\mathrm{copy}(u)} b_w$, it suffices to show that for any $u \in \partial S_v$,

$$a_u \leqslant \sum_{w \in \mathrm{copy}(u)} b_w. \tag{37}$$

For a pinning $\sigma \in \Omega_{\partial S_v}$, the corresponding pinning on T is σ_Λ, and

$$D_{\mathrm{TV}}\left(\mu_v^{\sigma^{u\leftarrow 0}}, \mu_v^{\sigma^{u\leftarrow 1}}\right) = D_{\mathrm{TV}}\left(\pi_v^{\sigma_\Lambda^{\mathrm{copy}(u)\leftarrow 0}}, \pi_v^{\sigma_\Lambda^{\mathrm{copy}(u)\leftarrow 1}}\right),$$

where $\sigma_\Lambda^{\mathrm{copy}(u)\leftarrow c}$ is the pinning on T obtained from σ_Λ by changing the value of all copies of u to c.

List all copies of u in T as $\mathrm{copy}(u) = \{u_1, \cdots, u_k\}$. By the triangle inequality, we can write

$$D_{\mathrm{TV}}\left(\mu_v^{\sigma^{u\leftarrow 0}}, \mu_v^{\sigma^{u\leftarrow 1}}\right) \leqslant \sum_{i=1}^{k} D_{\mathrm{TV}}\left(\pi_v^{\sigma_{\Lambda,i-1}}, \pi_v^{\sigma_{\Lambda,i}}\right), \tag{38}$$

where, for any $i \geqslant 1$, $\sigma_{\Lambda,i}$ is obtained from σ_Λ by changing the values of $u_1, \cdots, u_i$ to 0 and the values of $u_{i+1}, \cdots, u_k$ to 1. Note that $\sigma_{\Lambda,i-1}$ and $\sigma_{\Lambda,i}$ differ only at the single vertex u_i.

Next, we show that $\sigma_{\Lambda,i} \in \Omega_\Lambda$ for all $i = 1, \cdots, k$. Consider any vertex $w \notin \Lambda$. The vertex w is a copy of some vertex $w' \in S_v$. By the construction of T in Definition 12, each vertex x in $N_\Lambda^T(w)$ corresponds bijectively to a vertex y in $N_{\partial S_v}^G(w')$, and x is a copy of y. Thus, for any $w \notin \Lambda$ with $|N_\Lambda^T(w)| > D_2/3$, we can find $w' \in S_v$ such that w is a copy of w' and $|N_{\partial S_v}^G(w')| = |N_\Lambda^T(w)| > D_2/3$. Moreover,

$$\begin{aligned}|\{x \in N_\Lambda^T(w) : \sigma_\Lambda(x) = 1\}| &= |\{y \in N_{\partial S_v}^G(w') : \sigma(y) = 1\}| \\ &\geqslant |N_{\partial S_v}^G(w')|/(\log n) + 2 \\ &= |N_\Lambda^T(w)|/(\log n) + 2,\end{aligned} \tag{39}$$

where the inequality holds because $\sigma \in \Omega_{\partial S_v}$ in Definition 37. For $\sigma_{\Lambda,i}$, the only difference from σ_Λ is that the values of some copies of u are changed. In the SAW tree, no two copies of u can be children of the same vertex, so $|\{x \in N_\Lambda^T(w) : \sigma_{\Lambda,i}(x) = 1\}| \geqslant |\{x \in N_\Lambda^T(w) : \sigma_\Lambda(x) = 1\}| - 1$. Hence, for any $\sigma \in \Omega_{\partial S_v}$, combining (35) and (39), we obtain $\sigma_{\Lambda,i} \in \Omega_\Lambda$. Since the definition of b_w in (36) ranges over all pinnings in Ω_Λ, we have

$$a_u = \max_{\sigma \in \Omega_{\partial S_v}} D_{\mathrm{TV}}\left(\mu_v^{\sigma^{u\leftarrow 0}}, \mu_v^{\sigma^{u\leftarrow 1}}\right) \leqslant \sum_{i=1}^{k} D_{\mathrm{TV}}\left(\pi_v^{\sigma_{\Lambda,i}}, \pi_v^{\sigma_{\Lambda,i-1}}\right) \leqslant \sum_{w \in \mathrm{copy}(u)} b_w.$$

Summing over all $u \in \partial S_v$ proves the lemma. □

8. ASSM on the SAW tree

We now prove the ASSM property on the SAW tree. Fix a vertex $v \in V$ and a region S_v. Given a good boundary condition $\sigma \in \Omega_{\partial S_v}$, we construct the SAW tree $T = T_{\mathrm{SAW}}(G, v, \sigma)$ and prune all cycle-closing vertices in T. Recall that Λ consists of all copies of vertices in ∂S_v. To prove Lemma 38, by Lemma 40, we need to show that

$$\sum_{w \in \Lambda} b_w = \sum_{w \in \Lambda} \max_{\tau \in \Omega_\Lambda} D_{\mathrm{TV}}\left(\pi_v^{\tau^{w\leftarrow 0}}, \pi_v^{\tau^{w\leftarrow 1}}\right) \leqslant \frac{1}{20},$$

where π is the Gibbs distribution on the SAW tree.

For each vertex $w \in \Lambda$, let τ^w be the pinning of Λ in Ω_Λ that maximizes the total variation distance $D_{TV}\left(\pi_v^{\tau^{w \leftarrow 0}}, \pi_v^{\tau^{w \leftarrow 1}}\right)$. We write a superscript w to emphasize that the pinning τ^w depends on w. In the analysis, we view the SAW tree T as a computation tree and use the tree recursion to compute the marginal ratio at the root v. For each vertex w, define the corresponding ratio pinning $\rho^w : \Lambda \setminus \{w\} \to [0, \infty]$ such that

$$\forall u \in \Lambda \setminus \{w\}, \quad \rho^w(u) = \begin{cases} \infty & \text{if } \tau^w(u) = 0; \\ 0 & \text{if } \tau^w(u) = 1. \end{cases} \tag{40}$$

Consider two ratios $R_v^{\rho^w \wedge w \leftarrow \infty}$ and $R_v^{\rho^w \wedge w \leftarrow 0}$ at v under the two pinnings $\rho^w \wedge w \leftarrow \infty$ and $\rho^w \wedge w \leftarrow 0$, respectively, where the ratio is computed via the tree recursion (see Definition 20). Using the same proof as in Lemma 21, it is straightforward to show that

$$\begin{aligned} b_w &= \left| \frac{1}{1 + R_v^{\rho^w \wedge w \leftarrow \infty}} - \frac{1}{1 + R_v^{\rho^w \wedge w \leftarrow 0}} \right| = \frac{\left| R_v^{\rho^w \wedge w \leftarrow \infty} - R_v^{\rho^w \wedge w \leftarrow 0} \right|}{(1 + R_v^{\rho^w \wedge w \leftarrow \infty})(1 + R_v^{\rho^w \wedge w \leftarrow 0})} \\ &\leqslant \left| R_v^{\rho^w \wedge w \leftarrow \infty} - R_v^{\rho^w \wedge w \leftarrow 0} \right|. \end{aligned}$$

Hence, it suffices to bound the difference $\left| R_v^{\rho^w \wedge w \leftarrow \infty} - R_v^{\rho^w \wedge w \leftarrow 0} \right|$. However, for different vertices w, the pinnings $\rho^w : \Lambda \setminus \{w\} \to [0, \infty]$ can be different. We show that we can modify each pinning ρ^w to a pinning σ^w such that σ^w is similar to $\sigma^{w'}$ whenever two vertices w and w' lie on the same level of the SAW tree T. Recall that $L_k(v)$ is the set of all descendants of v at distance k from v in the SAW tree T.

Definition 41. A pinning $\sigma^* : \Lambda \to \{0, \infty\}$ is defined as follows. For each non-leaf vertex u,

- if $|N_\Lambda^T(u)| \leqslant D_2/3$, we set $\sigma^*(w) = \infty$ for all $w \in \Lambda$ that are children of u;
- if $|N_\Lambda^T(u)| > D_2/3$, let $w_1, w_2, \ldots, w_d \in \Lambda$ be the children of u in the SAW tree T, where $d = |N_\Lambda^T(u)|$. Let $\gamma_i = \gamma_{u,w_i}$ and $\beta_i = \beta_{u,w_i}$. Suppose all w_i are sorted in increasing order of $\beta_i\gamma_i$ (breaking ties arbitrarily). For the first $\lfloor |N_\Lambda^T(u)|/(\log n) \rfloor$ children, we set $\sigma^*(w_i) = 0$. For the remaining children, we set $\sigma^*(w_i) = \infty$.

Intuitively, the pinning σ^* is a pinning in Ω_Λ that maximizes the ratio $R_v^{\sigma^*}$. To see this, since we consider a ferromagnetic two-spin system, setting $\sigma^*(w) = \infty$ for all w would maximize the ratio $R_v^{\sigma^*}$. However, by Definition 39, if $|N_\Lambda^T(u)| > D_2/3$, then there is a restriction on the pinning at children of u. Hence, in the above definition, we need to pay special attention when $|N_\Lambda^T(u)| > D_2/3$.

The following lemma plays a key role in the analysis. For any k, let $L_{<k}(v) = \cup_{0 \leqslant j < k} L_j(v)$.

Lemma 42. *Let $\beta \leqslant 1 < \gamma$ such that $\beta\gamma > 1$ and $\lambda < \lambda_0(\beta, \gamma) := \sqrt{\gamma/\beta}$ be three constants. Suppose π is the Gibbs distribution of a (β, γ, λ)-ferromagnetic two-spin system on T. Let $w \in \Lambda$ be a vertex. Suppose $w \in L_k(v)$ for some $k \in \mathbb{N}$. For any pinning ρ^w obtained from $\tau_w \in \Omega_\Lambda$ as in (40), there exists a pinning $\sigma^w : (L_k(v) \setminus \{w\}) \cup (\Lambda \cap L_{<k}(v)) \to [0, \infty]$ such that*

- *for all vertices $u \in \Lambda$ with $u \in L_{k'}(v)$ for $k' < k$, $\sigma^w(u) = \sigma^*(u)$;*
- *for all siblings u of the vertex w, $\sigma^w(u) = \rho^w(u)$ if $u \in \Lambda$ and $\sigma^w(u) \in (0, \lambda)$ if $u \notin \Lambda$.*

Then the following inequality holds:

$$\left| R_v^{\rho^w \wedge w \leftarrow \infty} - R_v^{\rho^w \wedge w \leftarrow 0} \right| \leqslant \left| R_v^{\sigma^w \wedge w \leftarrow \infty} - R_v^{\sigma^w \wedge w \leftarrow 0} \right|. \tag{41}$$

The proof of Lemma 42 is given in Section 8.2. Using this lemma, we can bound the sum of the influences at each level. For each integer $k \geqslant 1$, the sum of influences can be bounded as

$$\sum_{w \in L_k(v) \cap \Lambda} \left| R_v^{\rho^w \wedge w \leftarrow \infty} - R_v^{\rho^w \wedge w \leftarrow 0} \right| \leqslant \sum_{w \in L_k(v) \cap \Lambda} \left| R_v^{\sigma^w \wedge w \leftarrow \infty} - R_v^{\sigma^w \wedge w \leftarrow 0} \right| := \mathrm{Inf}(k).$$

By the definition of the pinning σ^w, for any $w \in L_k(v) \cap \Lambda$, the restriction of σ^w to $L_{<k}(v)$ is the same. The only difference lies in the pinning on vertices at level k. Hence, we reduce the task of proving aggregate strong spatial mixing to the problem analyzed in Section 4.1. We also remark that Lemma 42 is the only place where we use the stronger condition $\lambda < \lambda_0$.

Define the following ferromagnetic two-spin system for each level $k \geqslant 1$.

Definition 43. Let $k \geqslant 1$ be an integer. Define a ferromagnetic two-spin system $\mathcal{S}_k$ as follows.

- Truncate the SAW tree T to keep the first k levels. Let T_k be the truncated SAW tree. The vertices and edges in T_k inherit the parameters of the original ferromagnetic two-spin system on T.
- For each vertex $w \in L_{<k}(v) \cap \Lambda$, the value of w is fixed by the pinning σ^*. Using self-reducibility in Observation 8, we remove the leaf vertex w and modify the external field of its parent.

By Observation 8, $\mathcal{S}_k$ is a (β, γ, λ)-ferromagnetic two-spin system.

For each vertex $w \in L_k(v)$, we use σ_k^w to denote the pinning σ^w restricted on $L_k(v)$. Hence, σ_k^w is a pinning on all leaf vertices of the tree T_k except the vertex w. Let $R_{v,k}^{\sigma_k^w \wedge w \leftarrow \infty}$ and $R_{v,k}^{\sigma_k^w \wedge w \leftarrow 0}$ be the ratio computed via the tree recursion in the ferromagnetic two-spin system $\mathcal{S}_k$. By definition, we have

$$\text{Inf}(k) = \sum_{w \in L_k(v) \cap \Lambda} \left| R_{v,k}^{\sigma_k^w \wedge w \leftarrow \infty} - R_{v,k}^{\sigma_k^w \wedge w \leftarrow 0} \right|. \tag{42}$$

Recall that D_1 and D_2 are defined in (33). Let n denote the number of vertices in the original graph G. We have the following two lemmas.

Lemma 44. *If $k > (\log \log n)^3$, then $\mathit{Inf}(k) \leqslant C' \cdot (1-\delta)^k \cdot (\log n)^3$, where $\delta = \delta(\beta, \gamma, \lambda) < 1$ and $C' = C'(\beta, \gamma, \lambda) > 0$ are constants depending on β, γ, λ.*

Lemma 45. *If $1 \leqslant k \leqslant (\log \log n)^3$, then $\mathit{Inf}(k) < \frac{1}{\log n}$.*

Assuming Lemma 44 and Lemma 45 hold, we can bound the sum of the influence as follows.

$$\begin{aligned} \sum_{k \geqslant 1} \text{Inf}(k) &\leqslant \frac{(\log \log n)^3}{\log n} + \sum_{k > (\log \log n)^3} C' \cdot (1-\delta)^k \cdot (\log n)^3 \\ &\leqslant o(1) + \frac{C'(1-\delta)^{(\log \log n)^3}}{\delta} (\log n)^3 = o(1) < \frac{1}{20}. \end{aligned}$$

The last equality holds because when n is sufficiently large, we have $(1-\delta)^{(\log \log n)^3} \ll \frac{1}{(\log n)^3}$. The above analysis shows that $\sum_{w \in \Lambda} b_w \leqslant \frac{1}{20}$. Combining it with Lemma 40 proves Lemma 38.

8.1. **Analysis of the sum of the influence.** We prove Lemma 44 and Lemma 45 in this subsection. We consider the following setting. Fix an integer $k \geqslant 1$. Let $\mathcal{S}_k$ be the ferromagnetic two-spin system defined in Definition 43 in the tree T_k, where T_k is a tree with k levels rooted at v. Recall that T_k is constructed by the following procedure. First, let $T_{\partial S_v} = T_{\text{SAW}}(G, v, \partial S_v)$. After pruning all cycle-closing vertices in $T_{\partial S_v}$, we obtain a tree T. Finally, we truncate the tree T and keep levels $0, 1, \ldots, k$, and then prune all vertices in $\Lambda \cap L_{<k}(v)$. When pruning a vertex, we modify the external field of its parent using self-reduction.

Lemma 46. *Let $u \in L_{k'}(v)$ be a vertex at level k' of the tree T_k, where $k' \leqslant k-2$.*

- *The number of children $|cld_{T_k}(u)|$ of u in T_k satisfies $|cld_{T_k}(u)| = F_{T_{\partial S_v}}(u)$, where $F_{T_{\partial S_v}}$ is defined in (32) and $T_{\partial S_v} = T_{\text{SAW}}(G, v, \partial S_v)$.*
- *If the number of non-cycle-closing children of u in $T_{\partial S_v}$ is at least D_2, then either u has at least $D_2/2$ children in T_k or $\lambda_u \leqslant \lambda(1/\gamma)^{D_2/(5 \log n)}$, where λ_u is the external field of u in $\mathcal{S}_k$.*

Proof. By the construction of T_k, for vertex u, we have pruned all its cycle-closing children and children in Λ from the tree $T_{\partial S_v}$. The first property holds from the definition of $F_{T_{\partial S_v}}(u)$.

For the second property, if the number of non-cycle-closing children of u in $T_{\partial S_v}$ is at least D_2, then one of the following two conditions must hold:

- u has at least $D_2/2$ children in $T_{\partial S_v}$ that are copies of vertices in S_v. All of them remain in T_k. Hence, u has at least $D_2/2$ children in T_k.
- u has at least $D_2/2$ children in $T_{\partial S_v}$ that are copies of vertices in ∂S_v. Hence, u has at least $D_2/2$ children in T that belong to Λ. By the definition of σ^* in Definition 41, at least $\lfloor |N_\Lambda^T(u)|/(\log n) \rfloor$ children in $N_\Lambda^T(u)$ satisfy $\sigma^*(w_i) = 0$. Note that when we prune a vertex and modify the external field of its parent using self-reduction, we can only decrease the external field of the parent because $\beta_e \leqslant 1$ and $\gamma_e > 1$ for all edges e. Hence, the external field of u in T_k can be bounded by

$$\lambda_u \leqslant \lambda \cdot \left(\frac{\beta 0 + 1}{0 + \gamma}\right)^{\lfloor |N_\Lambda^T(u)|/\log n \rfloor} \leqslant \lambda \cdot \left(\frac{1}{\gamma}\right)^{\lfloor |N_\Lambda^T(u)|/\log n \rfloor}$$
$$\leqslant \lambda \cdot \left(\frac{1}{\gamma}\right)^{\lfloor D_2/(2\log n) \rfloor} \leqslant \lambda \cdot \left(\frac{1}{\gamma}\right)^{D_2/(5\log n)}.$$

Hence, the second property holds. □

For any vertex $w \in L_k(v) \cap \Lambda$, there is an associated pinning σ_k^w on $L_k(v) \setminus \{w\}$. By Lemma 42, the pinning σ_k^w satisfies the following condition.

Lemma 47. *Let $w \in L_k(v) \cap \Lambda$ be a vertex at level k of the tree T_k. Let u be the parent of w in T_k, where u is at level $k-1$. The following two properties hold for the pinning σ_k^w.*

- *For any sibling $w' \notin \Lambda$ of w, $\sigma_k^w(w') \in [0, \lambda)$.*
- *If u has more than $D_2/3$ children in Λ (i.e., $|N_\Lambda^{T_k}(u)| > D_2/3$), then at least $\lfloor |N_\Lambda^{T_k}(u)|/\log n \rfloor$ siblings w' of w satisfy $\sigma_k^w(w') = 0$.*

Proof. The first property follows directly from Lemma 42. For the second property, if $|N_\Lambda^{T_k}(u)| > D_2/3$, then in the pinning ρ^w from Lemma 42, at least $\lfloor |N_\Lambda^{T_k}(u)|/\log n \rfloor + 1$ children w' of u satisfy $\rho^w(w') = 0$. This is because ρ^w is obtained from a good pinning $\tau^w \in \Omega_\Lambda$; see (40). Note that in T_k and $T_{\partial S_v}$, the children of u in Λ are the same. By the definition of a good boundary pinning in Definition 39, at least $\lfloor |N_\Lambda^{T_k}(u)|/\log n \rfloor + 1$ children of u satisfy $\tau^w(w') = 1$, and thus $\rho^w(w') = 0$. Using Lemma 42, all siblings $w' \in \Lambda$ of w satisfy $\sigma_k^w(w') = \rho^w(w')$. Hence, at least $\lfloor |N_\Lambda^{T_k}(u)|/\log n \rfloor$ siblings $w' \in \Lambda$ of w satisfy $\sigma_k^w(w') = 0$. □

Recall that the influence we need to bound is

$$\mathrm{Inf}(k) = \sum_{w \in L_k(v) \cap \Lambda} \left| R_{v,k}^{\sigma_k^w \wedge w \leftarrow \infty} - R_{v,k}^{\sigma_k^w \wedge w \leftarrow 0} \right|,$$

where $R_{v,k}^{\cdot}$ is the ratio computed by tree recursion in T_k rooted at v. We will use the general results Lemmas 25 and 26 to bound the influence. To use these lemmas in a black-box way, for each $w \in L_k(v) \setminus \Lambda$, we also associate w with an arbitrary pinning σ_k^w on $L_k(v) \setminus \{w\}$ that satisfies the condition in Lemma 47. We can define an upper bound on the influence by

$$\overline{\mathrm{Inf}}(k) = \sum_{w \in L_k(v)} \left| R_{v,k}^{\sigma_k^w \wedge w \leftarrow \infty} - R_{v,k}^{\sigma_k^w \wedge w \leftarrow 0} \right| \geqslant \mathrm{Inf}(k).$$

In $\mathrm{Inf}(k)$, the influence is contributed by the vertices in $L_k(v) \cap \Lambda$. In $\overline{\mathrm{Inf}}(k)$, the influence is contributed by all vertices in $L_k(v)$. Similarly to (15) and (16), we can define the potential-based influence $K_{v,k}^w$ and $K_{u,k}^w$, where we add a subscript k to emphasise that the quantity is defined on the tree T_k.

8.1.1. *Proof of Lemma 44.* Suppose $k \geqslant (\log\log n)^3$. We use Lemma 26 $(k-\ell_0+1)$ times and then use Lemma 25 (ℓ_0-2) times, where ℓ_0 is from Lemma 26. We arrive at a vertex u at level $k-1$ with children $u_1, u_2, \ldots, u_d$ such that

$$\sum_{w \in L_k(v)} K_{v,k}^{w} \leqslant (1-\delta)^{k-\ell_0+1} \cdot \left(C_{\mathrm{trl}} \cdot \lambda_u d \left(\frac{\beta\lambda+1}{\lambda+\gamma} \right)^{d-1} \right)^{\ell_0-2} \cdot \sum_{i=1}^{d} K_{u,k}^{u_i}.$$

Note that $d(\frac{\beta\lambda+1}{\lambda+\gamma})^{d-1} = d\exp(-\Omega(d)) = O_{\beta,\gamma,\lambda}(1)$ and $\delta = \delta(\beta,\gamma,\lambda)$ is a constant. Hence, we have

$$\sum_{w \in L_k(v)} K_{v,k}^{w} \leqslant O_{\beta,\gamma,\lambda}(1) \cdot (1-\delta)^{k} \cdot \sum_{i=1}^{d} K_{u,k}^{u_i}. \tag{43}$$

We need the following lemma to bound the influence coming from the last level.

Lemma 48. *Let u be a vertex at level $k-1$ with children $u_1, u_2, \ldots, u_d$. Then*

$$\sum_{i=1}^{d} K_{u,k}^{u_i} \leqslant \begin{cases} \lambda C_{\max} \cdot (\log n)^3 & \text{if } d < D_2 = (\log n)^3; \\ \exp(-d/(C_0 \log n)) & \text{if } d \geqslant D_2 = (\log n)^3. \end{cases}$$

where $C_{\max}$ is the constant in Lemma 24, and $C_0 > 1$ is a sufficiently large constant depending on β, γ, λ.

Assuming Lemma 48 holds, the last-level influence is at most $O_{\beta,\gamma,\lambda}(1) \cdot (\log n)^3$. Combining with (43),

$$\sum_{w \in L_k(v)} K_{v,k}^{w} \leqslant O_{\beta,\gamma,\lambda}(1) \cdot (1-\delta)^{k} \cdot (\log n)^3.$$

Combining the above bound with Lemma 24 proves Lemma 44. We now prove Lemma 48.

Proof of Lemma 48. Consider the two possible cases of the parameter d. If $d < D_2 = (\log n)^3$, then by the definition of the tree recursion, the influence

$$\left| R_{u,k}^{\sigma_k^{u_i} \wedge u_i \leftarrow \infty} - R_{u,k}^{\sigma_k^{u_i} \wedge u_i \leftarrow 0} \right| \leqslant \lambda,$$

because the recursion function has the image space in $[0, \lambda)$. Note that

$$K_{u,k}^{u_i} = \left| \Phi(R_{u,k}^{\sigma_k^{u_i} \wedge u_i \leftarrow \infty}) - \Phi(R_{u,k}^{\sigma_k^{u_i} \wedge u_i \leftarrow 0}) \right|.$$

Using Lemma 24, we have $K_{u,k}^{u_i} \leqslant C_{\max}\lambda$. Summing up all u_i (at most $d < D_2$) gives the first bound.

Suppose $d \geqslant D_2 = (\log n)^3$. Then either u has at least $d/2 \geqslant D_2/2$ children in Λ or at least $d/2 \geqslant D_2/2$ children not in Λ. Suppose we are in the first case. By Lemma 47, at least $\lfloor |N_{\Lambda}^{T_k}(u)|/\log n \rfloor \geqslant d/(5\log n)$ siblings w of u_i satisfy $\sigma_k^{u_i}(w) = 0$. Note that $\frac{\beta x+1}{x+\gamma} \leqslant 1$ for all $x \geqslant 0$. Hence,

$$\left| R_{u,k}^{\sigma_k^{u_i} \wedge u_i \leftarrow \infty} - R_{u,k}^{\sigma_k^{u_i} \wedge u_i \leftarrow 0} \right| \leqslant \lambda \left(\frac{1}{\gamma} \right)^{d/(5\log n)} \left(\frac{\beta\gamma-1}{\gamma} \right) \leqslant \exp\left(-\frac{d}{C_1 \log n} \right),$$

for some constant $C_1 > 1$ large enough. Here, the first factor λ comes from the external field $\lambda_u \leqslant \lambda$ of u; the second factor $(\frac{1}{\gamma})^{d/(5\log n)}$ comes from the siblings w of u_i with $\sigma_k^{u_i}(w) = 0$; and the third factor $(\frac{\beta\gamma-1}{\gamma}) \geqslant |\beta_{u,u_i} - \frac{1}{\gamma_{u,u_i}}|$ comes from the different pinnings at u_i.

For the second case, u has at least $d/2 \geqslant D_2/2$ children not in Λ. By Lemma 47, the pinning values at these children are at most λ. Then

$$\left| R_{u,k}^{\sigma_k^{u_i} \wedge u_i \leftarrow \infty} - R_{u,k}^{\sigma_k^{u_i} \wedge u_i \leftarrow 0} \right| \leqslant \lambda \left(\frac{\beta\lambda+1}{\lambda+\gamma} \right)^{d/2} \left(\frac{\beta\gamma-1}{\gamma} \right) \leqslant \exp\left(-\frac{d}{C_1} \right),$$

where the last inequality holds for some constant $C_1 > 1$ large enough. Finally, summing over all u_i and using the bound in Lemma 24 gives

$$\sum_{i=1}^{d} K_{u,k}^{u_i} \leqslant C_{\max} d \exp\left(-\frac{d}{C_1 \log n}\right) \leqslant \exp\left(-\frac{d}{C_0 \log n}\right),$$

for some constant $C_0 > 1$ large enough. □

8.1.2. *Proof of Lemma 45.* Let $\ell_1 := \max\{-1, k - \ell_0\}$, where ℓ_0 is from Lemma 26. By applying Lemma 25 and Lemma 26, we go through a path from the root ν to a vertex u at level $k-1$ with children $u_1, u_2, \ldots, u_d$. Let the path be $\nu = \nu_0, \nu_1, \ldots, \nu_{k-1} = u$, and the number of children of ν_i is d_i, where $d_{k-1} = d$. We have

$$\begin{aligned}\sum_{w \in L_k(\nu)} K_{\nu,k}^{w} \leqslant &\prod_{i=0}^{\ell_1} \min\left\{\left(C_{\text{trl}} \cdot \lambda_{u_i} d_i \left(\frac{\beta\lambda+1}{\lambda+\gamma}\right)^{d_i}\right), 1-\delta\right\} \\ &\cdot \prod_{i=\ell_1+1}^{k-2} \left(C_{\text{trl}} \cdot \lambda_{u_i} d_i \left(\frac{\beta\lambda+1}{\lambda+\gamma}\right)^{d_i}\right) \cdot \sum_{i=1}^{d} K_{u,k}^{u_i}.\end{aligned} \tag{44}$$

For any $0 \leqslant i \leqslant k-2$, we have $d_i = F_{T_{\partial S_\nu}}(u)$, where $F_{T_{\partial S_\nu}}$ is defined in (32) and $T_{\partial S_\nu} = T_{\text{SAW}}(G, \nu, \partial S_\nu)$. If there exists $j \in [0, k-2]$ such that $d_j \geqslant D_2 = (\log n)^3$, then similar to the proof of Lemma 48, we have

$$C_{\text{trl}} \cdot \lambda_{u_j} d_j \left(\frac{\beta\lambda+1}{\lambda+\gamma}\right)^{d_j} \leqslant \left(-\frac{d_j}{C_2} + C_3\right),$$

for sufficiently large constants $C_2, C_3 > 0$. Note that here we choose C_2 and C_3 large so that the estimate above holds for any integer $d_j \geqslant 1$, although with sufficiently large n we could absorb C_3 into C_2. This is because this estimate will be used again later when we do not have the assumption that $d_j \geqslant D_2$. Next we have

$$\begin{aligned}\sum_{w \in L_k(\nu)} K_{\nu,k}^{w} &\leqslant \exp\left(-\frac{d_j}{C_2} + C_3\right) \prod_{i=0, i \neq j}^{\ell_1} \min\left\{\left(C_{\text{trl}} \cdot \lambda_{u_i} d_i \left(\frac{\beta\lambda+1}{\lambda+\gamma}\right)^{d_i - 1}\right), 1-\delta\right\} \\ &\quad \cdot \prod_{i=\ell_1+1, i \neq j}^{k-2} \left(C_{\text{trl}} \cdot \lambda_{u_i} d_i \left(\frac{\beta\lambda+1}{\lambda+\gamma}\right)^{d_i}\right) \cdot \sum_{i=1}^{d} K_{u,k}^{u_i} \\ &\leqslant \exp\left(-\frac{d_j}{C_2} + C_3\right) \cdot O_{\beta,\gamma,\lambda}(1) \cdot (\log n)^3 < \frac{1}{(\log n)^2},\end{aligned}$$

where the second inequality holds because every factor in the first product is at most $1 - \delta < 1$, the second product is no larger than C^{ℓ_0} for some constant $C > 0$ and $\ell_0 = O_{\beta,\gamma,\lambda}(1)$, and the term $\sum_{i=1}^{d} K_{u,k}^{u_i}$ is bounded by Lemma 48. The last inequality holds for large enough n as $\exp\left(-\frac{d_j}{C_2} + C_3\right) \leqslant \exp\left(-\frac{(\log n)^3}{C_2} + C_3\right) \leqslant \frac{1}{n}$. Since $\sum_{w \in L_k(\nu)} K_{\nu,k}^{w}$ is at most $\frac{1}{(\log n)^2}$, using Lemma 24, the sum of the influence without potential function is at most $O(\frac{1}{(\log n)^2}) < \frac{1}{\log n}$.

If $d = d_{k-1} \geqslant D_2 = (\log n)^3$, then by Lemma 48, we have $\sum_{i=1}^{d} K_{u,k}^{u_i} \leqslant \exp(-\frac{d}{C_0 \log n})$. Therefore,

$$\sum_{w \in L_k(\nu)} K_{\nu,k}^{w} \leqslant O_{\beta,\gamma,\lambda}(1) \sum_{i=1}^{d} K_{u,k}^{u_i} \leqslant O_{\beta,\gamma,\lambda}(1) \cdot \exp\left(-\frac{d}{C_0 \log n}\right) < \frac{1}{(\log n)^2},$$

where the first inequality holds because every factor in the first product in (44) is at most $1 - \delta < 1$, and the second product is no larger than C^{ℓ_0} for some constant $C > 0$ and $\ell_0 = O_{\beta,\gamma,\lambda}(1)$. The last inequality holds for sufficiently large n because $\exp\left(-\frac{d}{C_0 \log n}\right) \leqslant \exp\left(-\frac{(\log n)^3}{C_0 \log n}\right) = \exp\left(-\frac{(\log n)^2}{C_0}\right) \leqslant \frac{1}{n}$. Again, using Lemma 24, the sum of the influence without potential function is at most $O(\frac{1}{(\log n)^2}) < \frac{1}{\log n}$.

For the remaining case, we have $d_i < D_2$ for all $i \in [0, k-1]$. Hence, by Lemma 36, we have $\sum_{i=0}^{d-2} d_i \geqslant D_1$. Let C_4 be a large enough constant such that $(1-\delta)^{C_4/2} \leqslant \exp(-5)$ and C_5 be a large enough constant such that $\exp(-C_5/2C_2) \leqslant \exp(-5)$. We set

$$C_D := 2C_2C_3C_4 + C_5, \tag{45}$$

and recall that $D_1 = C_D \log\log n$. There are two subcases:

(1) $|\{d_i : d_i < 2C_2C_3 \wedge i \in [0, k-2]\}| \geqslant C_4 \cdot \log\log n$, we have $k \geqslant C_4 \cdot \log\log n$ and

$$\begin{aligned}\sum_{w \in L_k(v)} K_{v,k}^{w} \leqslant & (1-\delta)^{k-\ell_0+1} \cdot O_{\beta,\gamma,\lambda}(1) \cdot (\log n)^3 \\ \leqslant & (1-\delta)^{C_4 \cdot \log\log n/2} \cdot O_{\beta,\gamma,\lambda}(1) \cdot (\log n)^3 \\ \leqslant & \exp(-5 \cdot \log\log n) \cdot O_{\beta,\gamma,\lambda}(1) \cdot (\log n)^3 < \frac{1}{(\log n)^{1.5}};\end{aligned}$$

(2) $|\{d_i : d_i < 2C_2C_3 \wedge i \in [0, k-2]\}| < C_4 \cdot \log\log n$, then $\sum_{0 \leqslant i \leqslant k-2 : d_i \geqslant 2C_2C_3} d_i \geqslant D_1 - 2C_2C_3C_4 \cdot \log\log n = C_5 \log\log n$. We have $\exp\left(-\frac{x}{C_2} + C_3\right) \leqslant \exp\left(-\frac{x}{2C_2}\right)$ for $x \geqslant 2C_2C_3$. Hence,

$$\begin{aligned}\sum_{w \in L_k(v)} K_{v,k}^{w} \leqslant & \prod_{i=0}^{\ell_1} \min\left\{\left(C_{\mathrm{trl}} \cdot \lambda_{u_i} d_i \left(\frac{\beta\lambda+1}{\lambda+\gamma}\right)^{d_i}\right), 1-\delta\right\} \\ & \cdot \prod_{i=\ell_1+1}^{k-2} \left(C_{\mathrm{trl}} \cdot \lambda_{u_i} d_i \left(\frac{\beta\lambda+1}{\lambda+\gamma}\right)^{d_i}\right) \cdot \sum_{i=1}^{d} K_{u,k}^{u_i} \\ \leqslant & \prod_{0 \leqslant i \leqslant k-2 : d_i \geqslant 2C_2C_3} \exp\left(-\frac{d_i}{2C_2}\right) \cdot O_{\beta,\gamma,\lambda}(1) \cdot (\log n)^3 \\ = & \exp\left(-\frac{\sum_{0 \leqslant i \leqslant k-2 : d_i \geqslant 2C_2C_3} d_i}{2C_2}\right) \cdot O_{\beta,\gamma,\lambda}(1) \cdot (\log n)^3 \\ \leqslant & \exp\left(-\frac{C_5 \log\log n}{2C_2}\right) \cdot O_{\beta,\gamma,\lambda}(1) \cdot (\log n)^3 < \frac{1}{(\log n)^{1.5}}.\end{aligned}$$

where the last inequality holds because $\exp(-\frac{C_5 \log\log n}{2C_2}) \leqslant \exp(-5 \log\log n)$.

We have shown that $\sum_{w \in L_k(v)} K_{v,k}^{w} < \frac{1}{(\log n)^{1.5}}$ for all cases. Using Lemma 24, the sum of the influence without potential function is at most $O\left(\frac{1}{(\log n)^{1.5}}\right) < \frac{1}{\log n}$.

8.2. **Find the worst pinning.** We now give the proof of Lemma 42. First, we show the following property of the pinning σ^* constructed in Definition 41. Recall that $L_{<k}(v) = \cup_{j<k} L_j(v)$ and $L_{\geqslant k}(v) = \cup_{j \geqslant k} L_j(v)$.

Lemma 49. *Let $\sigma : \Lambda \to \{0, \infty\}$, where $\sigma \in \Omega_\Lambda$[3]. Let $w \in \Lambda$ and $c \in \{0, \infty\}$. Let $k \geqslant 1$ be an integer. Define a pinning $\tau : \Lambda \to \{0, \infty\}$ such that*

$$\forall u \in \Lambda \setminus \{w\}, \quad \tau(u) = \begin{cases} \sigma^*(u) & \text{if } u \in \Lambda \cap L_{<k}(v), \\ \sigma(u) & \text{if } u \in \Lambda \cap L_{\geqslant k}(v). \end{cases}$$

[3] By definition, Ω_Λ contains all pinnings σ such that σ fixes the value of each $w \in \Lambda$ to either 0 or 1, which is equivalent to fixing the ratio at each w to either ∞ or 0.

For any non-leaf vertex u,

$$R_u^{\sigma\wedge w\leftarrow c} \leqslant R_u^{\tau\wedge w\leftarrow c},$$

where $\sigma\wedge w\leftarrow c$ *is the pinning obtained from* σ *by overwriting the value of* w *to* c.

Remark. By Definition 20, the ratio $R_u^{\sigma\wedge w\leftarrow c}$ is computed via tree recursion given the initial value $\sigma\wedge w\leftarrow c$ at leaves Λ. Note that $R_u^{\sigma\wedge w\leftarrow c} = R_u^{\bar{\sigma}}$, where $\bar{\sigma}$ is the pinning obtained from $\sigma\wedge w\leftarrow c$ by removing the pinning outside the subtree of u. This is because the value computed at u is independent of the pinning outside the subtree of u.

Proof of Lemma 49. We prove it by induction on u from bottom to top. For the base case, all children of u are leaf vertices. If $u \in L_{\geqslant k-1}(v)$, then for σ and τ, the pinning on the subtree of u is the same, and hence $R_u^{\sigma\wedge w\leftarrow c} = R_u^{\tau\wedge w\leftarrow c}$. Suppose $u \in L_{<k-1}(v)$. If $|N_\Lambda^T(u)| \leqslant D_2/3$, then for all children $x\in\Lambda$ of u, we have $\sigma(x)\leqslant\tau(x)=\sigma^*(x)=\infty$, and w has the same value in the two pinnings (if w is a child of u). Since the tree recursion is monotone, we have $R_u^{\sigma\wedge w\leftarrow c} \leqslant R_u^{\tau\wedge w\leftarrow c}$. If $|N_\Lambda^T(u)| > D_2/3$, let $w_1, w_2, \dots, w_d \in \Lambda$ be the children of u in the SAW tree T, where $d = |N_\Lambda^T(u)|$. Let $\gamma_i = \gamma_{u,w_i}$ and $\beta_i = \beta_{u,w_i}$. Suppose all w_i are sorted in decreasing order of $\beta_i\gamma_i$ (breaking ties arbitrarily). Let $w'_1, w'_2, \dots, w'_{d'} \notin \Lambda$ be the other children of u in the SAW tree T. Let $\gamma'_i = \gamma_{u,w'_i}$ and $\beta'_i = \beta_{u,w'_i}$. Note that $w'_1,\dots,w'_{d'}$ must be unpinned leaves. Using the tree recursion, we have

$$R_u^{\sigma\wedge w\leftarrow c} = \lambda_u \prod_{1\leqslant i\leqslant d:\sigma(w_i)=0}\frac{1}{\gamma_i}\prod_{1\leqslant i\leqslant d:\sigma(w_i)=\infty}\beta_i\prod_{1\leqslant j\leqslant d'}\frac{\beta'_j\lambda_{w'_j}+1}{\lambda_{w'_j}+\gamma'_j}\cdot W$$

$$= \lambda_u\left(\prod_{1\leqslant i\leqslant d}\frac{1}{\gamma_i}\right)\prod_{1\leqslant i\leqslant d:\sigma(w_i)=\infty}\beta_i\gamma_i\prod_{1\leqslant j\leqslant d'}\frac{\beta'_j\lambda_{w'_j}+1}{\lambda_{w'_j}+\gamma'_j}\cdot W, \tag{46}$$

where $W=1$ if w is not a child of u and $W=\frac{\beta_{u,w}c+1}{\gamma_{u,w}+c}$ if w is a child of u. Similarly,

$$R_u^{\tau\wedge w\leftarrow c} = \lambda_u\left(\prod_{1\leqslant i\leqslant d}\frac{1}{\gamma_i}\right)\prod_{1\leqslant i\leqslant d:\sigma^*(w_i)=\infty}\beta_i\gamma_i\prod_{1\leqslant j\leqslant d'}\frac{\beta'_j\lambda_{w'_j}+1}{\lambda_{w'_j}+\gamma'_j}\cdot W. \tag{47}$$

Let $N_\Lambda^T(u)$ be the set of children of u that are in Λ. Note that $N_\Lambda^T(u)$ must contain $w_1,\dots,w_d$ and may contain w if w is a child of u. By the definition of Ω_Λ, at least $\lfloor|N_\Lambda^T(u)|/(\log n)\rfloor+1$ children in $N_\Lambda^T(u)$ have $\sigma(w_i)=0$ (that is, the value of w_i is pinned to 1). Hence, at most $|N_\Lambda^T(u)|-\lfloor|N_\Lambda^T(u)|/(\log n)\rfloor-1$ children in $N_\Lambda^T(u)$ have $\sigma(w_i)=\infty$. Let us consider two cases.

- Case I: w is not a child of u. Note that $\beta_i\gamma_i\geqslant 1$ for all i. By definition, σ^* picks exactly $|N_\Lambda^T(u)|-\lfloor|N_\Lambda^T(u)|/(\log n)\rfloor$ children w_i with the largest $\beta_i\gamma_i$ and sets $\sigma^*(w_i)=\infty$. By (46) and (47), $R_u^{\sigma\wedge w\leftarrow c}\leqslant R_u^{\tau\wedge w\leftarrow c}$.
- Case II: w is a child of u. Note that W is the same factor in both $R_u^{\sigma\wedge w\leftarrow c}$ and $R_u^{\tau\wedge w\leftarrow c}$ by (46) and (47). In $R_u^{\sigma\wedge w\leftarrow c}$, at most $|N_\Lambda^T(u)|-\lfloor|N_\Lambda^T(u)|/(\log n)\rfloor-1$ children among $\{w_1,w_2,\dots,w_d\}\setminus\{w\}$ contribute a factor $\beta_i\gamma_i$ because the pinning on w has been overwritten. In $R_u^{\tau\wedge w\leftarrow c}$, we may set $\sigma^*(w)=\infty$, but we set $\sigma^*(w')=\infty$ for at least $|N_\Lambda^T(u)|-\lfloor|N_\Lambda^T(u)|/(\log n)\rfloor$ children $w'\in\{w_1,w_2,\dots,w_d\}$. At least $|N_\Lambda^T(u)|-\lfloor|N_\Lambda^T(u)|/(\log n)\rfloor-1$ children among $\{w_1,w_2,\dots,w_d\}\setminus\{w\}$ satisfy $\sigma^*(w_i)=\infty$. These children contribute the $|N_\Lambda^T(u)|-\lfloor|N_\Lambda^T(u)|/(\log n)\rfloor-1$ largest factors $\beta_i\gamma_i$ among all children in $\{w_1,w_2,\dots,w_d\}\setminus\{w\}$. Hence, $R_u^{\sigma\wedge w\leftarrow c}\leqslant R_u^{\tau\wedge w\leftarrow c}$.

For a general non-leaf vertex u, where u may have non-leaf children w' and children w_i in the set Λ, the induction hypothesis gives $R_{w'}^{\sigma\wedge w\leftarrow c}\leqslant R_{w'}^{\tau\wedge w\leftarrow c}$ for every non-leaf child w'. For all children $w_i\in\Lambda$ of u, we can use the same analysis as in the base case. Since the recursion is monotone, it follows that $R_u^{\sigma\wedge w\leftarrow c}\leqslant R_u^{\tau\wedge w\leftarrow c}$. □

We next prove the following technical lemma.

Lemma 50. *Let* $\beta \leqslant 1 < \gamma$, $\beta\gamma > 1$ *and* $\lambda < \lambda_0(\beta,\gamma) := \sqrt{\gamma/\beta}$. *Let* $\lambda \geqslant x > y > 0$ *and* $\lambda \geqslant x' > y' > 0$, *satisfying* $x \geqslant x'$, $y \geqslant y'$, *and* $x/y \geqslant x'/y'$. *Then*

$$\frac{\beta x + 1}{x+\gamma} \cdot \frac{y+\gamma}{\beta y + 1} \geqslant \frac{\beta x' + 1}{x'+\gamma} \cdot \frac{y'+\gamma}{\beta y' + 1}. \tag{48}$$

Proof. Subtracting 1 from both the left and right sides of (48), we only need to show that

$$\frac{(\beta\gamma-1)(x-y)}{(x+\gamma)(\beta y+1)} \geqslant \frac{(\beta\gamma-1)(x'-y')}{(x'+\gamma)(\beta y'+1)}. \tag{49}$$

It is easy to see that the right-hand side is monotone decreasing in y'. We only need to consider the case $y' = x'y/x$. In this case, we can set $1 \geqslant c = x'/x = y'/y$. Then (49) is equivalent to

$$\frac{1}{(x+\gamma)(\beta y+1)} \geqslant \frac{c}{(cx+\gamma)(\beta c y+1)},$$

which, in turn, is equivalent to $(1-c)(\gamma - c\beta xy) \geqslant 0$. The last inequality holds because $\gamma - c\beta xy \geqslant \gamma - \beta\lambda^2 > 0$. □

Now, we are ready to prove Lemma 42.

Proof of Lemma 42. We first consider the following definition of the pinning σ^w on $\Lambda \setminus \{w\}$:

$$\forall u \in \Lambda \setminus \{w\}, \quad \sigma^w(u) = \begin{cases} \sigma^*(u) & \text{if } u \in \Lambda \cap L_{<k}(v), \\ \rho^w(u) & \text{if } u \in \Lambda \cap (L_{\geqslant k}(v) \setminus \{w\}). \end{cases} \tag{50}$$

We first show that (41) holds for this pinning σ^w. Note that the pinning σ^w in the lemma is a pinning on the subset $(L_k(v) \setminus \{w\}) \cup (\Lambda \cap L_{<k}(v))$. After proving (41), we explain how to modify σ^w so that it satisfies the condition in the lemma.

Let the path from w to v in the SAW tree T be $w = u_0, u_1, \dots, u_{k-1}, u_k = v$. By the monotonicity of the recursion function, for all $1 \leqslant j \leqslant k$, we have

$$x_j := R_{u_j}^{\sigma^w \wedge w \leftarrow \infty} > y_j := R_{u_j}^{\sigma^w \wedge w \leftarrow 0}, \quad x_j' := R_{u_j}^{\rho^w \wedge w \leftarrow \infty} > y_j' := R_{u_j}^{\rho^w \wedge w \leftarrow 0}. \tag{51}$$

By definition, $x_j = R_{u_j}^{\sigma_j^w \wedge w \leftarrow \infty}$, where σ_j^w is the pinning σ^w projected on vertices in $\cup_{\ell \geqslant k-j+1} L_\ell(v)$. This is because when computing the tree recursion for u_j, we only need to use all pinnings at the subtree rooted at u_j. Note that the vertex u_j is in $L_{k-j}(v)$. Hence, the value of x_j depends only on σ_j^w. Similar results apply to y_j, x_j', y_j'. By applying Lemma 49 to $u_1, \cdots, u_k$, we have

$$\forall 1 \leqslant j \leqslant k, \quad x_j \geqslant x_j' \text{ and } y_j \geqslant y_j'.$$

We claim that

$$\forall 1 \leqslant j \leqslant k, \quad \frac{x_j}{y_j} \geqslant \frac{x_j'}{y_j'}. \tag{52}$$

We prove inequality (52) by induction on j. For $j = 1$, note that x_1, y_1 depend only on σ^w projected on vertices in $L_{\geqslant k}(v)$ (denoted by σ_1^w), and y_1, y_1' depend only on ρ^w projected on vertices in $L_{\geqslant k}(v)$ (denoted by ρ_1^w). By (50), $\sigma_1^w = \rho_1^w$. Hence, $x_1 = x_1'$ and $y_1 = y_1'$, so the claim holds. Now fix $1 < j \leqslant k$ and assume the claim holds for $j-1$. Note that x_j, y_j, x_j', y_j' can all be computed by tree recursion. Let β_j and γ_j be the parameters on the edge $\{u_j, u_{j-1}\}$. By comparing the tree recursion for x_j and y_j, we have

$$\frac{x_j}{y_j} = \frac{\beta_j x_{j-1} + 1}{x_{j-1} + \gamma_j} \cdot \frac{y_{j-1} + \gamma_j}{\beta_j y_{j-1} + 1}.$$

Similarly, we can write

$$\frac{x'_j}{y'_j} = \frac{\beta_j x'_{j-1} + 1}{x'_{j-1} + \gamma_j} \cdot \frac{y'_{j-1} + \gamma_j}{\beta_j y'_{j-1} + 1}.$$

By the definition of the recursion function, all $x_{j-1}, y_{j-1}, x'_{j-1}, y'_{j-1} \leqslant \lambda$. Note $x_{j-1} > y_{j-1}$, $x'_{j-1} > y'_{j-1}$, $x_{j-1} \geqslant x'_{j-1}$, and $y_{j-1} \geqslant y'_{j-1}$. By induction hypothesis, $\frac{x_{j-1}}{y_{j-1}} \geqslant \frac{x'_{j-1}}{y'_{j-1}}$. Using Lemma 50,

$$\frac{x_j}{y_j} = \frac{\beta_j x_{j-1} + 1}{x_{j-1} + \gamma_j} \cdot \frac{y_{j-1} + \gamma_j}{\beta_j y_{j-1} + 1} \geqslant \frac{\beta_j x'_{j-1} + 1}{x'_{j-1} + \gamma_j} \cdot \frac{y'_{j-1} + \gamma_j}{\beta_j y'_{j-1} + 1} = \frac{x'_j}{y'_j}.$$

Finally, we have $\frac{R_v^{\sigma^w \wedge w \leftarrow \infty}}{R_v^{\sigma^w \wedge w \leftarrow 0}} = \frac{x_k}{y_k} \geqslant \frac{x'_k}{y'_k} = \frac{R_v^{\rho^w \wedge w \leftarrow \infty}}{R_v^{\rho^w \wedge w \leftarrow 0}}$. We can compute that

$$|x_k - y_k| = y_k \left| \frac{x_k}{y_k} - 1 \right| \geqslant y'_k \left| \frac{x'_k}{y'_k} - 1 \right| = |x'_k - y'_k|,$$

where the inequality holds because $y_k \geqslant y'_k$, $\frac{x_k}{y_k}, \frac{x'_k}{y'_k} \geqslant 1$, and $\frac{x_k}{y_k} \geqslant \frac{x'_k}{y'_k}$.

To obtain the pinning σ^w in the lemma, we compute the tree recursion from the bottom up to the level k conditional on σ^w, except for vertex w (note that $w \in \Lambda$ is a leaf at level k). After the computation, every vertex $u \in L_k(v) \setminus \{w\}$ gets a ratio. We set this value as the pinning value of $\sigma^w(u)$ and remove all the pinnings below the level k. Therefore, we get a pinning σ^w defined on the subset $(L_k(v) \setminus \{w\}) \cup (\Lambda \cap L_{<k}(v))$. By definition, for all $u \in \Lambda \cap L_{<k}(v)$, we have $\sigma^w(u) = \sigma^*(u)$. For all siblings $u \in \Lambda$ of the vertex w, note that u is in the level k and u must be a leaf node because $u \in \Lambda$. When computing the tree recursion for u, we simply let u take the pinning value $\rho^w(u)$. For all siblings $u \notin \Lambda$ of the vertex w, their values are not fixed by ρ^w,

- if u is a leaf, then the ratio value at u is $\lambda_u < \lambda$ (note that u cannot be a cycle-closing vertex because we have pruned all cycle-closing vertices when constructing the tree T);
- if u is not a leaf, then the ratio value at u is computed by tree recursion. The range of the tree recursion function implies that $\sigma^w(u) \in (0, \lambda)$.

In both cases, we have $\sigma^w(u) \in (0, \lambda)$. This verifies the two properties of σ^w in the lemma. □

9. Proof of main results

In this section we show the main theorems, namely Theorem 3, Theorem 4, and Theorem 5. Note that Theorem 1 is implied by Theorem 3. We first show the slightly easier Theorem 4 in Section 9.1. Then, in Section 9.2, we show Theorem 3 via a similar approach. We conclude by proving Theorem 5 in Section 9.3.

9.1. Mixing of Glauber dynamics when $\lambda < \lambda_0$.

Theorem 4 is proved by applying Theorem 34. Recall that Glauber dynamics is a special case of the heat-bath block dynamics in Theorem 34, where each block is a single vertex. We verify the conditions in Definition 33 and (23) in Theorem 34 for a (β, γ, λ)-ferromagnetic two-spin system on a graph G with $\beta \leqslant 1 < \gamma$, $\beta\gamma > 1$, and $\lambda < \lambda_0 := \sqrt{\gamma/\beta}$. The definition of good boundary conditions is given in Definition 37. We first show the following lemma.

Lemma 51. *For any $v \in V$, any $S_v \ni v$, and any $\sigma, \tau \in \Omega_{\partial S_v}$, where $\Omega_{\partial S_v}$ is defined in Definition 37, there exists a path $\eta_0, \eta_1, \ldots, \eta_t \in \Omega_{\partial S_v}$ such that $\eta_0 = \sigma$, $\eta_t = \tau$, and for any $0 \leqslant i < t$, η_i and η_{i+1} differ at exactly one vertex, where $t = |\{u \in \partial S_v : \sigma(u) \neq \tau(u)\}|$ is the Hamming distance between σ and τ.*

Proof. To move from σ to τ, define the following two sets of vertices:

$$S_1 = \{u \in \partial S_v : \sigma(u) = 0, \tau(u) = 1\},$$
$$S_2 = \{u \in \partial S_v : \sigma(u) = 1, \tau(u) = 0\}.$$

Starting from σ, we first change all $v \in S_1$ from the value 0 to the value 1, and then change all $v \in S_2$ from the value 1 to the value 0. For any η_i in the path, it is straightforward to see that for any $u \in S_v$ with $|N^G_{\partial S_v}(u)| > D_2/3$, it satisfies

$$\begin{aligned}|\{w \in N^G_{\partial S_v}(u) : \eta_i(w) = 1\}| &\geqslant \min\{|\{w \in N^G_{\partial S_v}(u) : \sigma(w) = 1\}|, |\{w \in N^G_{\partial S_v}(u) : \tau(w) = 1\}|\} \\ &\geqslant |N^G_{\partial S_v}(u)|/(\log n) + 2.\end{aligned}$$

Hence, η_i is a good boundary configuration. The length of the path is $|S_1| + |S_2| = t$. $\square$

Lemma 51 proves the first property of Definition 33. The second property of Definition 33 is proved by Lemma 38. We next verify the condition (23) in Theorem 34. Consider the monotone coupling $(X^+_t, X^-_t)_{t \geqslant 0}$ of the Glauber dynamics in Definition 31. We show that there exists

$$T_{\text{burn-in}} = O(n \log n)$$

such that for any $t \geqslant T_{\text{burn-in}}$ and any $v \in V$, it holds that

$$\Pr[X^+_t(\partial S_v) \notin \Omega_{\partial S_v} \vee X^-_t(\partial S_v) \notin \Omega_{\partial S_v}] \leqslant \frac{1}{n^3}. \tag{53}$$

Fix any time $t \geqslant T_{\text{burn-in}}$. If $T_{\text{burn-in}}$ is a sufficiently large multiple of $n \log n$, then with probability at least $1 - \frac{1}{n^{10}}$, each vertex $u \in V$ has been updated at least once during the time interval $[t - T_{\text{burn-in}}, t]$. For each vertex $u \in \partial S_v$, consider the last time in the interval $[t - T_{\text{burn-in}}, t]$ at which u is updated, and denote this time by t_u. For every edge $e \in E$, we have $\beta_e \leqslant 1$ and $\gamma_e \geqslant 1$. Hence, whenever u is updated, the conditional probability that it is set to 1 is at least $\frac{1}{1+\lambda_u} \geqslant \frac{1}{1+\lambda} = \Omega(1)$. Consider a vertex $w \in S_v$ with $d > D_2/3 = (\log n)^3/3$ neighbors in ∂S_v. Since a good boundary configuration requires at least $d/\log n + 2$ neighbors of w in state 1, a Chernoff bound shows that, with probability at least $1 - \frac{1}{n^{10}}$, at least $d/\log n + 2$ neighbors u of w are set to 1 at their respective times t_u. Taking a union bound over the two chains X^+_t and X^-_t, and over all relevant vertices $w \in S_v$, yields (53).

Finally, we claim the local mixing time for censored Glauber dynamics on $\mu^\sigma_{S_v}$ is

$$T_{\text{local}} = n \cdot (\log n)^{C''}, \tag{54}$$

where $C'' = C''(\beta, \gamma, \lambda) > 0$ is a constant depending on β, γ, λ. Assume the above local mixing time bound holds. Let $t^{\text{Glauber}}_{\text{mix}}$ denote the mixing time of Glauber dynamics. By Theorem 34, we have

$$\begin{aligned}t^{\text{Glauber}}_{\text{mix}}\left(\frac{1}{4e}\right) &= O\left(T_{\text{burn-in}} + T_{\text{local}} \cdot \max_{v \in V} \log |R_v| \cdot \log n\right), \quad \text{where } |R_v| = |S_v \cup \partial S_v| \leqslant n \\ &\leqslant n \cdot (\log n)^{C(\beta,\gamma,\lambda)}.\end{aligned}$$

Then, Theorem 4 follows from the standard decay in ϵ for mixing times, namely $t^{\text{Glauber}}_{\text{mix}}(\epsilon) \leqslant t^{\text{Glauber}}_{\text{mix}}(\frac{1}{4e}) \log \frac{1}{\epsilon}$.

We use the following result to show the local mixing bound in (54).

Theorem 52. *Let $\beta, \gamma, \lambda > 0$ be three constants such that $\beta \leqslant 1 < \gamma$, $\beta\gamma > 1$, and $\lambda < \lambda_c := (\gamma/\beta)^{\frac{\sqrt{\beta\gamma}}{\sqrt{\beta\gamma}-1}}$. For any (β, γ, λ)-ferromagnetic two-spin system with vertex set V, the spectral gap of the Glauber dynamics on the Gibbs distribution μ is at least $\frac{1}{|V|^C}$, where $C = C(\beta, \gamma, \lambda) > 0$ is a constant depending on β, γ, λ.*

Remark. The above theorem only requires a weaker condition $\lambda < \lambda_c$. Note that

$$\lambda_c = (\gamma/\beta)^{\frac{\sqrt{\beta\gamma}}{\sqrt{\beta\gamma}-1}} > \sqrt{\gamma/\beta} = \lambda_0.$$

Hence, we can use the above theorem to prove the local mixing bound when $\lambda < \lambda_0$. Theorem 52 can also be viewed as a weaker version of Theorem 5 when $\lambda < \lambda_c$ as it only provides a $\text{poly}(n) \cdot \log \frac{1}{\mu(\sigma)}$ mixing time bound instead of the $n^3 \cdot \text{polylog}(n)$ mixing time bound in Theorem 5.

To prove Theorem 52, we need the following mixing result obtained from the spectral independence.

Proposition 53 ([ALO24]). *Let μ be a distribution over $\{0,1\}^V$. If there exists a constant $\eta > 0$ such that for any pinning $\sigma \in \{0,1\}^\Lambda$, the conditional distribution $\mu^\sigma_{V\setminus\Lambda}$ has η-bounded all-to-one influence, then, the spectral gap of the Glauber dynamics on μ is at least $\frac{1}{n^{O(\eta)}}$.*

Proof of Theorem 52. Using Observation 8, any conditional distribution μ^σ also induces a Gibbs distribution of a (β,γ,λ)-ferromagnetic two-spin system on a subgraph. By Theorem 19, all conditional distributions have C_{inf}-bounded all-to-one influence for some constant $C_{\mathrm{inf}} = C_{\mathrm{inf}}(\beta,\gamma,\lambda) > 0$ depending on β,γ,λ. The theorem then follows from Proposition 53. □

We use Theorem 52 to prove the local mixing bound. Fix any vertex $v \in V$ and any outside configuration $\sigma \in \{0,1\}^{V\setminus S_v}$. The censored Glauber dynamics on $\mu^\sigma_{S_v}$ updates as follows: in each step, it picks a vertex $u \in V$ uniformly at random; if $u \notin S_v$, then the dynamics does nothing; otherwise, it resamples the value at u conditional on the current configuration of the other variables. It is straightforward to see that the censored Glauber dynamics on $\mu^\sigma_{S_v}$ is at most a factor of n slower than the Glauber dynamics on $\pi = \mu^\sigma_{S_v}$, where in each step, the Glauber dynamics picks a vertex $u \in S_v$ uniformly at random and resamples the value. Using Lemma 36 and (33), we know that $|S_v| \leqslant (\log n)^{C'}$, where $C' = C'(\beta,\gamma,\lambda) > 0$ is a constant. We prove the following mixing result. Note that (54) is a simple corollary of this lemma.

Lemma 54. *Let $\pi = \mu^\sigma_{S_v}$. Let P^{Glauber}_π be the Glauber dynamics on π. Starting from an arbitrary configuration in $\{0,1\}^{S_v}$, after running P^{Glauber}_π for $(\log n)^{C''}$ steps, the total variation distance between the resulting distribution and the stationary distribution π is at most $\frac{1}{4e}$, where $C'' = C''(\beta,\gamma,\lambda) > 0$ is a constant.*

By Observation 8, the conditional distribution π is a Gibbs distribution of a (β,γ,λ)-ferromagnetic two-spin system on $G[S_v]$. If we directly apply Theorem 52 and (6), then we need to bound $\log\frac{1}{\pi_{\min}}$, where $\pi_{\min} = \min_{x\in\{0,1\}^{S_v}} \pi(x)$. However, for some edge e and vertex u, the parameters β_e and λ_u can be arbitrarily small and the parameter γ_e can be arbitrarily large. Hence, $\log\frac{1}{\pi_{\min}}$ can be larger than $\mathrm{polylog}(n)$. To resolve this issue, we use Theorem 52 after reaching a warm-start configuration. We give the following general result.

Lemma 55. *Let $\beta \leqslant 1 < \gamma$, $\beta\gamma > 1$ and $\lambda > 0$ be three constants. Let μ be a Gibbs distribution of a (β,γ,λ)-ferromagnetic two-spin system on a graph $G = (V,E)$. Let P^{Glauber}_μ be the Glauber dynamics on μ. Suppose the spectral gap of the Glauber dynamics is at least $0 < g < 1$. Then, the mixing time of the Glauber dynamics on μ satisfies*

$$t^{\mathrm{Glauber}}_{\mathrm{mix}}\left(\frac{1}{4e}\right) \leqslant O_\lambda\left(|V|\log|V| + \frac{|V|^2}{g}\log|V|\right).$$

Assume that Lemma 55 holds. We apply Lemma 55 to the distribution π defined on the subgraph $G[S_v]$. Note that $|S_v| \leqslant (\log n)^{C'}$, where $C' = C'(\beta,\gamma,\lambda) > 0$ is a constant. Using Theorem 52 on the subgraph $G[S_v]$, the spectral gap of the Glauber dynamics on π is at least $\frac{1}{(\log n)^C}$, where $C = C(\beta,\gamma,\lambda)$ is a constant. Hence, the mixing time of the Glauber dynamics on π is at most $(\log n)^{C''}$. This proves Lemma 54. Finally, we prove Lemma 55.

Proof of Lemma 55. Let $N = |V|$. Let $N_0(\lambda)$ be a sufficiently large constant depending only on λ. First we consider the case when $N \leqslant N_0(\lambda) = O_\lambda(1)$. In each update of the Glauber dynamics, we have a chance at least $\frac{1}{1+\lambda}$ to update the value of a vertex to 1. We run the Glauber dynamics for some $O_\lambda(1)$ steps, so that with probability $\Omega_\lambda(1)$, all vertices takes the value 1. Let $T_0 = O_\lambda(1)$ be a sufficiently large constant. With probability at least $1 - \frac{1}{10e}$, we can find a time $t < T_0$ such that all vertices take the value 1. For each edge, $\gamma_e > 1 \geqslant \beta_e$. It holds that $\mu(\mathbf{1}) = \Omega_\lambda(1)$ if $N \leqslant N_0(\lambda)$. Using (6), starting from all-1 configuration, we only need to run Glauber dynamics for $O_\lambda(1/g)$ steps to get a configuration with total variation distance at most $\frac{1}{10e}$ to the stationary distribution μ. A simple coupling argument shows that the total variation distance between the resulting distribution and μ is at most $\frac{1}{4e}$ after $T_0 + O_\lambda(1/g) = O_\lambda(1/g)$ steps.

Now, we assume $N \geqslant N_0(\lambda)$ is large enough. Fix $\tau \in \{0,1\}^V$. We say that a vertex $u \in V$ is *bad* in τ if

$$\lambda_u \leqslant \frac{1}{100N^5} \qquad \text{and} \qquad \tau(u) = 0.$$

For any edge $e = \{u, w\} \in E$, we say that e is *bad* in τ if

$$\gamma_e \geqslant 100N^5 \text{ and } (\tau(u) = 0 \text{ or } \tau(w) = 0),$$

and we say that τ is a *warm-start configuration* if no vertex or edge is bad in τ.

We prove the following two claims.

- Starting from an arbitrary configuration $X_0 \in \{0,1\}^V$, after running P_μ^{Glauber} for $T_0 = O_\lambda(N(\log N)^2)$ steps, with probability at least $1 - \frac{1}{10e}$, the configuration X_{T_0} is a warm-start configuration.
- Starting from any warm-start configuration X_{T_0}, after running the Glauber dynamics for $T_1 = O_\lambda\left(\frac{N^2}{g}\log N\right)$ steps, where g is a lower bound of the spectral gap, the total variation distance between the resulting distribution and μ is at most $\frac{1}{10e}$.

If these two claims hold, we can construct a coupling between the law of $X_{T_0+T_1}$ and the stationary distribution μ such that the coupling fails with probability at most

$$\begin{aligned}&\Pr[X_{T_0} \text{ is not a warm-start configuration}] + \Pr[\text{coupling fails} \mid X_{T_0} \text{ is a warm-start configuration}]\\ &= \frac{1}{10e} + \frac{1}{10e} < \frac{1}{4e},\end{aligned}$$

which finishes the proof.

Now we prove the first claim. Let $M = C_1 N \log N$ and $L = C_0 \log N$, where $C_1 > 0$ is a sufficiently large absolute constant and $C_0 = C_0(\lambda) > 0$ is a sufficiently large constant depending only on λ. Set

$$T_0 = LM = O_\lambda(N(\log N)^2).$$

Partition the time interval $[T_0]$ into L consecutive blocks, each of length M. We list the sequence of updated vertices as

$$v_1, v_2, \ldots, v_{T_0}.$$

An update sequence is *good* if every vertex is updated at least once in every block. By the coupon collector bound and a union bound over all L blocks, the update sequence is good with probability at least $1 - \frac{1}{20e}$.

Fix a good update sequence. We first bound the probability that a vertex is bad in X_{T_0}. Fix any vertex $u \in V$, and let t_u be the last time at which u is updated. We must have $\lambda_u \leqslant \frac{1}{100N^5}$, since otherwise u cannot be bad. Fix all the updates before time t_u. Let $u_1, u_2, \ldots, u_d$ denote all neighbors of u, let β_i, γ_i denote the parameters of the edge $\{u_i, u\}$, and let ρ denote the configuration of the other variables at time t_u. Then

$$\frac{\Pr[u \text{ is updated to } 0]}{\Pr[u \text{ is updated to } 1]} = \lambda_u \prod_{i\in[d]:\rho(u_i)=1} \frac{1}{\gamma_i} \prod_{i\in[d]:\rho(u_i)=0} \beta_i. \tag{55}$$

Since $\frac{1}{\gamma_i} \leqslant 1$ and $\beta_i \leqslant 1$, we have

$$\Pr[X_{t_u}(u) = 0] \leqslant \lambda_u \leqslant \frac{1}{100N^5}.$$

Hence, the probability that u is bad in X_{T_0} is at most $\frac{1}{100N^5}$.

Now fix a bad edge $e = \{u, w\} \in E$ with $\gamma_e \geqslant 100N^5$, as otherwise, e cannot be bad. We call a pair of times (t, t') a *clean pair* for e if $t < t'$, $\{v_t, v_{t'}\} = \{u, w\}$, $v_t \neq v_{t'}$, and for all $t < \ell < t'$ we have $v_\ell \notin \{u, w\}$. Since the update sequence is good, both u and w are updated at least once in every block. Fix any block. Since both vertices appear in the block, there must be a clean pair of times for e. We list all clean pairs in the update sequence: $(t_j, t_j')_{j=1}^K$ with $t_j' < t_{j+1}$, where $K \geqslant L \geqslant C_0 \log N$ since there is at least one clean pair in each block.

Fix all randomness used to update vertices in $V \setminus \{u, w\}$. Let $p_\lambda := \frac{1}{1+\lambda}$. For each clean pair (t_b, t'_b), define the event A_b by

$$A_b = \{X_{t'_b}(u) = X_{t'_b}(w) = 1\}.$$

By (55), at every update of either u or w, the chosen vertex is updated to 1 with probability at least p_λ. Therefore, conditional on all past update on $\{u, w\}$ before time t_b, we have that the probability of the event A_b is at least p_λ^2. By iterating this bound over all clean pairs and choosing C_0 sufficiently large as a function of λ, we obtain

$$\Pr\left[\bigcap_{b=1}^{K} \overline{A_b}\right] \leqslant (1 - p_\lambda^2)^K \leqslant N^{-6}.$$

If any event A_b occurs, then the following event A holds:

- A: there exists the first time $t_e < T_0$ such that $X_{t_e}(u) = X_{t_e}(w) = 1$.

Hence, t_e exists with probability at least $1 - N^{-6}$. Furthermore, the random variable t_e is independent from the updates after t_e. Suppose $t_e = s'$ and let $s > s'$ be the first time after t_e at which the edge u or w is updated to 0. At time s, one of the endpoints, say u, is updated to 0 while the other endpoint is still equal to 1. Hence, by (55),

$$\Pr[X_s(u) = 0 \mid X_{s-1}(w) = 1] \leqslant \frac{\lambda_u}{\gamma_e} \leqslant \frac{\lambda}{100N^5}.$$

A union bound over all times s' for $t_e = s'$ and all times s for $s > s'$ yields

$$\Pr[e \text{ is bad in } X_{T_0} \mid A] \leqslant \frac{\lambda T_0^2}{100N^5}.$$

Therefore, since $T_0 = O_\lambda(N(\log N)^2)$, we have

$$\Pr[e \text{ is bad in } X_{T_0}] \leqslant \Pr[\neg A] + \Pr[e \text{ is bad in } X_{T_0} \mid A] \leqslant N^{-6} + \frac{\lambda T_0^2}{100N^5} \leqslant \frac{1}{100N^{2.5}}.$$

Taking a union bound over all vertices and edges, conditioned on the update sequence fixed above, the probability that X_{T_0} is not a warm-start configuration is at most

$$\frac{N}{100N^5} + \frac{N^2}{100N^{2.5}} < \frac{1}{20e}. \tag{56}$$

Combining this with the probability $\frac{1}{20e}$ that the update sequence is not good proves the first claim.

For the second claim, we show a lower bound on $\mu(\tau)$ for each warm-start configuration τ. For any configuration $\tau' \in \{0,1\}^V$, not necessarily a warm-start configuration, we give a lower bound on the ratio $\frac{\mu(\tau)}{\mu(\tau')}$. We analyze the contribution of every vertex and every edge in G. Formally, the ratio $\frac{\mu(\tau)}{\mu(\tau')}$ can be written as the following ratio of products:

$$\frac{\mu(\tau)}{\mu(\tau')} = \frac{\prod_{u \in V} a_u(\tau(u)) \prod_{e \in E} b_e(\tau(e))}{\prod_{u \in V} a_u(\tau'(u)) \prod_{e \in E} b_e(\tau'(e))},$$

where, for each vertex $u \in V$,

$$a_u(\tau(u)) := \begin{cases} \lambda_u & \text{if } \tau(u) = 0; \\ 1 & \text{if } \tau(u) = 1, \end{cases}$$

and for each edge $e = \{u, w\} \in E$,

$$b_e(\tau(e)) := \begin{cases} \beta_e & \text{if } \tau(u) = \tau(w) = 0; \\ \gamma_e & \text{if } \tau(u) = \tau(w) = 1; \\ 1 & \text{if } \tau(u) \neq \tau(w). \end{cases}$$

We analyse each ratio as follows.

- If $\lambda_u \leqslant \frac{1}{100N^5}$, then $\tau(u) = 1$ because τ is warm-start. Hence, $\frac{f_u(\tau(u))}{f_u(\tau'(u))} \geqslant \min\{1, \lambda^{-1}\}$.
- If $\lambda_u > \frac{1}{100N^5}$, then $\frac{f_u(\tau(u))}{f_u(\tau'(u))} \geqslant \min\{1/(100N^5), \lambda^{-1}\}$.
- If $\gamma_e \geqslant 100N^5$, then $\tau(u) = \tau(w) = 1$ because τ is warm-start. Therefore, $\frac{f_e(\tau(e))}{f_e(\tau'(e))} \geqslant 1$.
- If $\gamma_e < 100N^5$, then $\beta_e > \frac{1}{\gamma_e} > \frac{1}{100N^5}$ because $\beta_e\gamma_e > 1$. Therefore,
$$\frac{f_e(\tau(e))}{f_e(\tau'(e))} \geqslant \frac{\beta_e}{\gamma_e} > \frac{1}{10^4 N^{10}}.$$

The total number of edges in E is at most N^2. Hence, the ratio $\frac{\mu(\tau)}{\mu(\tau')}$ can be bounded as follows:

$$\frac{\mu(\tau)}{\mu(\tau')} \geqslant \left(\min\{1/(100N^5), \lambda^{-1}\}\right)^N \cdot \left(\frac{1}{10^4 N^{10}}\right)^{N^2} \geqslant \exp(-O_\lambda(N^2 \log N)).$$

Since the above lower bound holds for every $\tau' \in \{0,1\}^V$, summing over all 2^N choices of τ' gives

$$\mu(\tau) \geqslant \exp(-O_\lambda(N^2 \log N)) \cdot 2^{-N} = \exp(-O_\lambda(N^2 \log N)). \tag{57}$$

Let

$$T_1 := O\left(\frac{1}{g} \log \frac{1}{(1/10e)^2 \mu(\tau)}\right) = O_\lambda\left(\frac{N^2}{g} \log N\right).$$

The second claim follows from (6) with $\epsilon = \frac{1}{10e}$ for the warm-start configuration τ. □

9.2. **Mixing of alternating-scan sampler.** In this section, we prove the mixing result in Theorem 3. Our proof applies to a general family of ferromagnetic two-spin systems, of which RBMs in Theorem 1 are a special case. Consider a (β,γ,λ)-ferromagnetic two-spin system on a bipartite graph $G = (V_0, V_1, E)$ with $V = V_0 \uplus V_1$, where $\beta \leqslant 1 < \gamma$, $\beta\gamma > 1$, and $\lambda < \lambda_0 := \sqrt{\gamma/\beta}$. We prove a mixing time bound of $(\log n)^{O_{\beta,\gamma,\lambda}(1)} \log \frac{1}{\epsilon}$ for the alternating-scan sampler on the Gibbs distribution μ.

The proof strategy here is the same as that in Section 9.1. The alternating-scan sampler is a special case of the systematic-scan block dynamics in Theorem 34 with two blocks, namely $\mathcal{B} = \{V_0, V_1\}$. The definition of good boundary conditions is given in Definition 37. Lemma 51 proves the first property of Definition 33. The second property of Definition 33 is proved by Lemma 38. For the burn-in estimate in (53), we can simply set $T_{\text{burn-in}} := 2$. In the alternating-scan sampler, after two steps all vertices have been updated exactly once. The bound in (53) follows from the same Chernoff-bound argument used in Section 9.1.

Finally, we claim that the local mixing time for the censored alternating-scan sampler on $\mu^{\sigma}_{S_v}$ is

$$T_{\text{local}} = (\log n)^{C''}, \tag{58}$$

where $C'' = C''(\beta,\gamma,\lambda) > 0$ is a constant depending on β,γ,λ. Let $t^{\text{AS}}_{\text{mix}}$ denote the mixing time of the alternating-scan sampler. Assuming this local mixing bound, Theorem 34 implies

$$\begin{aligned} t^{\text{AS}}_{\text{mix}}\left(\frac{1}{4e}\right) &= O\left(T_{\text{burn-in}} + T_{\text{local}} \cdot \max_{v\in V} \log |R_v| \cdot \log n\right), \quad \text{where } |R_v| = |S_v \cup \partial S_v| \leqslant n \\ &\leqslant (\log n)^{C(\beta,\gamma,\lambda)}. \end{aligned}$$

Theorem 1 then follows from the standard ϵ decay in mixing times $t^{\text{AS}}_{\text{mix}}(\epsilon) \leqslant t^{\text{AS}}_{\text{mix}}\left(\frac{1}{4e}\right) \log \frac{1}{\epsilon}$.

Fix a set S_v with size $|S_v| = N \leqslant (\log n)^{C'}$ and a boundary configuration $\sigma \in \{0,1\}^{\partial S_v}$. By Observation 8, $\mu^{\sigma}_{S_v}$ is a Gibbs distribution of a (β,γ,λ)-ferromagnetic two-spin system on $G[S_v]$. Note that $G[S_v]$ is a bipartite graph. Let $U_0 = V_0 \cap S_v$ and $U_1 = V_1 \cap S_v$. Using Theorem 19 and Observation 8, every conditional distribution induced by $\pi = \mu^{\sigma}_{S_v}$ has C_{inf}-bounded all-to-one influence. By Proposition 53, the spectral gap of the Glauber dynamics on π is at least $N^{-O(C_{\mathrm{inf}})} = \frac{1}{\mathrm{polylog}(n)}$. Then Propositions 6 and 7 imply that, starting from any configuration $\tau \in \{0,1\}^{S_v}$, after running the alternating-scan sampler on π for $2N^{O(C_{\mathrm{inf}})} \log \frac{4e^2}{\epsilon^2 \pi(\tau)}$ steps, the total variation distance between the resulting distribution and the stationary distribution is at most ϵ. We prove the local mixing bound in (58) using a warm-start argument similar to that in Section 9.1. The case $N = O_\lambda(1)$ can be handled by the same argument. For large N, recall the definition of a warm-start configuration from the proof of Lemma 55. Let

$$T_0^{\mathrm{AS}} = O_\lambda(\log N).$$

In every two consecutive steps of the alternating-scan sampler, every vertex in S_v is updated exactly once, and every edge receives a clean ordered pair of endpoint updates. Therefore, the same argument as in the proof of Lemma 55 shows that, starting from any configuration $X_0 \in \{0,1\}^{S_v}$, after running the alternating-scan sampler on π for T_0^{AS} steps, the probability that $X_{T_0^{\mathrm{AS}}}$ is a warm-start configuration is at least $1 - \frac{1}{10e}$.

For any warm-start configuration $\tau \in \{0,1\}^{S_v}$, by (57), we have $\pi(\tau) \geqslant \exp(-\Omega(N^2 \log N)) = \exp(-\mathrm{polylog}(n))$. Starting from any warm-start configuration $X_{T_0^{\mathrm{AS}}} = \tau$, after T_1 additional steps, where

$$T_1 = 2N^{O(C_{\mathrm{inf}})} \log \frac{4e^2}{(1/10e)^2 \pi(\tau)} \leqslant \mathrm{polylog}(n),$$

the resulting distribution is within $\frac{1}{10e}$ in total variation distance from the stationary distribution.

Hence, starting from any configuration $X_0 \in \{0,1\}^{S_v}$, we can couple $X_{T_0^{\mathrm{AS}}+T_1}$ with the stationary distribution π successfully with probability at least $1 - 1/(10e) - 1/(10e) > 1 - 1/(4e)$. By the coupling inequality,

$$D_{\mathrm{TV}}\left(X_{T_0^{\mathrm{AS}}+T_1}, \pi\right) \leqslant \frac{1}{4e}.$$

This proves the local mixing time bound in (58).

9.3. **Mixing of Glauber dynamics when $\lambda < \lambda_c$.** To prove Theorem 5, we use the field dynamics technique introduced in [CFYZ21]. Let μ be a distribution over $\{0,1\}^V$, and let $\boldsymbol{\theta} = (\theta_v)_{v \in V}$ be a vector of real numbers. The tilted distribution $\boldsymbol{\theta} * \mu$ is defined by

$$\forall \sigma \in \{0,1\}^V, \quad (\boldsymbol{\theta} * \mu)(\sigma) \propto \mu(\sigma) \cdot \prod_{v \in V: \sigma_v = 0} \theta_v.$$

In particular, if $\theta_v = \theta$ for all $v \in V$, then we denote $\boldsymbol{\theta} * \mu = \theta * \mu$.

The field dynamics on μ is defined as follows. Let $\theta \in (0,1)$. Starting from an arbitrary configuration $X \in \{0,1\}^V$, in each step, it updates the current configuration X as follows:

- construct a random subset $S \subseteq V$ by selecting each vertex $v \in V$ independently with probability p_v, where $p_v = 1$ if $X(v) = 1$ and $p_v = \theta$ if $X(v) = 0$;
- resample $X(S) \sim (\theta * \mu)_S^{X(V \setminus S)}$, where $(\theta * \mu)_S^{X(V \setminus S)}$ is the marginal distribution on S induced by $(\theta * \mu)$ conditioned on the configuration $X(V \setminus S)$ on the variables outside S.

Compared with the original version of the field dynamics in [CFYZ21], the above definition swaps the roles of 0 and 1. The two versions are essentially equivalent. The spectral gap of the field dynamics can be analyzed using the *complete spectral independence* property. We have the following proposition.

Proposition 56 ([CFYZ21])**.** *Let $\eta > 0$ be a constant. If the distribution μ over $\{0,1\}^V$ satisfies the following condition: for any $\boldsymbol{\phi} \in (0,1]^V$ and any pinning $\sigma \in \{0,1\}^\Lambda$, the conditional distribution $(\boldsymbol{\phi} * \mu)_{V \setminus \Lambda}^{\sigma}$ has η-bounded all-to-one influence, then for any $\theta \in (0,1)$, the spectral gap of the field dynamics on μ with parameter θ is at least $\theta^{O(\eta)}$.*

Let μ be a Gibbs distribution of a (β, γ, λ)-ferromagnetic two-spin system on a graph G, where $\lambda < \lambda_c := (\gamma/\beta)^{\frac{\sqrt{\beta\gamma}}{\sqrt{\beta\gamma}-1}}$. By Definition 2, the tilted distribution $\theta * \mu$ is again a ferromagnetic two-spin system with the same edge parameters and external fields satisfying $\lambda_v \theta_v < \lambda < \lambda_c$. By Observation 8 and Theorem 19, the distribution μ satisfies the condition in Proposition 56 with $\eta = C_{\text{inf}}$. Let $\gamma_{\text{field}}(\mu, \theta)$ denote the spectral gap of the field dynamics on μ with parameter θ. Then

$$\gamma_{\text{field}}(\mu, \theta) \geqslant \theta^{O(C_{\text{inf}})}. \tag{59}$$

To relate the field dynamics to the Glauber dynamics, we need the following definition. Let $\sigma \in \{0, 1\}^\Lambda$ be a configuration, where $\Lambda \subseteq V$ is a subset of vertices. Consider the distribution $(\theta * \mu)^\sigma$, obtained by pinning all variables in Λ according to σ. The Glauber dynamics on $(\theta * \mu)^\sigma$ is defined as follows. Starting from an arbitrary configuration $X \in \{0, 1\}^V$ with $X(\Lambda) = \sigma$, in each step, pick a vertex $v \in V$ uniformly at random. If $v \in \Lambda$, then do nothing; if $v \notin \Lambda$, then resample $X(v) \sim (\theta * \mu)_v^{X(V \setminus \{v\})}$. In particular, we take the parameter θ as follows:

$$\theta = \frac{1}{2\lambda_c} = \Theta_{\beta,\gamma,\lambda}(1). \tag{60}$$

Note that $(\theta * \mu)^\sigma$ coincides with the conditional distribution $(\theta * \mu)^\sigma_{V \setminus \Lambda}$ because all variables in Λ are pinned. Furthermore, $(\theta * \mu)^\sigma_{V \setminus \Lambda}$ is a Gibbs distribution of a ferromagnetic two-spin system on the induced subgraph $G[V \setminus \Lambda]$ with the same edge parameters and with external fields bounded by

$$\lambda_v \theta < \lambda_c \cdot \theta = \frac{1}{2} < 1 < \lambda_0.$$

Using Theorem 4, for any $\Lambda \subseteq V$ and any $\sigma \in \{0, 1\}^\Lambda$, the mixing time of the Glauber dynamics on $(\theta * \mu)^\sigma$ started from an arbitrary configuration is at most

$$\forall \epsilon > 0, \quad t_{\text{mix}}^{\text{Glauber}}((\theta * \mu)^\sigma, \epsilon) = O\left(n (\log n)^C \log \frac{1}{\epsilon}\right),$$

where $C = C(\beta, \gamma, \lambda) > 0$ is a constant depending on β, γ, λ. As a consequence, the spectral gap of the Glauber dynamics on $(\theta * \mu)^\sigma$ is at least $\Omega(n^{-1}(\log n)^{-C})$ (see [LP17, Theorem 12.5]). Define

$$\gamma_{\min}(\theta) := \min\left\{\gamma_{\text{Glauber}}((\theta * \mu)^\sigma) \mid \sigma \in \{0,1\}^\Lambda, \Lambda \subseteq V\right\} = \Omega\left(\frac{1}{n(\log n)^C}\right), \tag{61}$$

where $\gamma_{\text{Glauber}}((\theta * \mu)^\sigma)$ is the spectral gap of the Glauber dynamics on $(\theta * \mu)^\sigma$. Let $\gamma_{\text{field}}(\mu, \theta)$ denote the spectral gap of the field dynamics on μ with parameter θ. The spectral gap of the Glauber dynamics on μ can be lower-bounded by the following proposition.

Proposition 57 ([CFYZ21])**.** $\gamma_{\text{Glauber}}(\mu) \geqslant \gamma_{\text{field}}(\mu, \theta) \cdot \gamma_{\min}(\theta)$.

Combining (59), (60), (61), and Proposition 57, we obtain the following lower bound on the spectral gap of the Glauber dynamics:

$$\gamma_{\text{Glauber}}(\mu) \geqslant \gamma_{\text{field}}(\mu, \theta) \cdot \gamma_{\min}(\theta) = \Omega_{\beta,\gamma,\lambda}\left(\frac{1}{n(\log n)^C}\right). \tag{62}$$

Finally, we bound the mixing time of the Glauber dynamics on μ. Suppose the starting configuration is the all-1 configuration $X_0 = \mathbf{1}$. For any configuration $\tau \in \{0, 1\}^V$, it holds that

$$\frac{\mu(\mathbf{1})}{\mu(\tau)} \geqslant \min\{1, \lambda^{-1}\}^n \geqslant \lambda_c^{-n}.$$

The above inequality holds because $\mathbf{1}$ maximizes the factors contributed by all edges; for each vertex, the factor contributed by $\mathbf{1}$ is 1, whereas the factor contributed by τ is at most $\max\{1, \lambda\}$. Since there are 2^n configurations in total, we have

$$\mu(\mathbf{1}) \geqslant (2\lambda_c)^{-n}. \tag{63}$$

Combining (62) and (6), the mixing time of the Glauber dynamics starting from the all-1 configuration is

$$t_{\text{mix-}\mathbf{1}}^{\text{Glauber}}(\epsilon) = O\left(\frac{1}{\gamma_{\text{Glauber}}(\mu)} \log \frac{1}{\epsilon^2 \mu(\mathbf{1})}\right) = O_{\beta,\gamma,\lambda}\left(n^2 (\log n)^C \log \frac{1}{\epsilon}\right).$$

To bound the mixing time of the Glauber dynamics on μ starting from an arbitrary configuration, combine (62), Lemma 55, and (5) to obtain

$$t_{\text{mix}}^{\text{Glauber}}(\epsilon) = O_{\beta,\gamma,\lambda}\left(n^3 (\log n)^{C+1} \log \frac{1}{\epsilon}\right).$$

Theorem 5 now follows after increasing the constant C in the theorem by 2. The extra $\log n$ factor absorbs the constants hidden in the notation $O_{\beta,\gamma,\lambda}(\cdot)$.

Acknowledgement

We thank Konrad Anand and Graham Freifeld for useful discussions at an early stage of this paper.

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 947778). Weiming Feng acknowledges the support of ECS grant 27202725 from Hong Kong RGC.

References

[AHS85] David H. Ackley, Geoffrey E. Hinton, and Terrence J. Sejnowski. A learning algorithm for Boltzmann machines. *Cogn. Sci.*, 9(1):147–169, 1985. 1

[AJK+22] Nima Anari, Vishesh Jain, Frederic Koehler, Huy Tuan Pham, and Thuy-Duong Vuong. Entropic independence: optimal mixing of down-up random walks. In *STOC*, pages 1418–1430. ACM, 2022. 5

[ALO24] Nima Anari, Kuikui Liu, and Shayan Oveis Gharan. Spectral independence in high-dimensional expanders and applications to the hardcore model. *SIAM J. Comput.*, 53(6):S20–1, 2024. 4, 5, 6, 13, 15, 16, 39

[Bar16] Alexander I. Barvinok. *Combinatorics and Complexity of Partition Functions*, volume 30 of *Algorithms and combinatorics*. Springer, 2016. 4

[BCV20] Antonio Blanca, Zongchen Chen, and Eric Vigoda. Swendsen-Wang dynamics for general graphs in the tree uniqueness region. *Random Struct. Algorithms*, 56(2):373–400, 2020. 22, 49, 50

[CE22] Yuansi Chen and Ronen Eldan. Localization schemes: A framework for proving mixing bounds for markov chains (extended abstract). In *FOCS*, pages 110–122. IEEE, 2022. 5

[CFYZ21] Xiaoyu Chen, Weiming Feng, Yitong Yin, and Xinyuan Zhang. Rapid mixing of Glauber dynamics via spectral independence for all degrees. In *FOCS*, pages 137–148. IEEE, 2021. 4, 5, 43, 44

[CFYZ22] Xiaoyu Chen, Weiming Feng, Yitong Yin, and Xinyuan Zhang. Optimal mixing for two-state anti-ferromagnetic spin systems. In *FOCS*, pages 588–599. IEEE, 2022. 5

[CLV23a] Zongchen Chen, Kuikui Liu, and Eric Vigoda. Optimal mixing of glauber dynamics: Entropy factorization via high-dimensional expansion. *SIAM J. Comput.*, 0(0):STOC21–104–STOC21–153, 2023. 4

[CLV23b] Zongchen Chen, Kuikui Liu, and Eric Vigoda. Rapid mixing of Glauber dynamics up to uniqueness via contraction. *SIAM J. Comput.*, 52(1):196–237, 2023. 4

[CZ23] Xiaoyu Chen and Xinyuan Zhang. A near-linear time sampler for the Ising model with external field. In *SODA*, pages 4478–4503. SIAM, 2023. 3

[DGGJ04] Martin E. Dyer, Leslie Ann Goldberg, Catherine S. Greenhill, and Mark Jerrum. The relative complexity of approximate counting problems. *Algorithmica*, 38(3):471–500, 2004. 4

[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. 3

[FGW23] Weiming Feng, Heng Guo, and Jiaheng Wang. Swendsen-Wang dynamics for the ferromagnetic ising model with external fields. *Inf. Comput.*, 294:105066, 2023. 3

[Fil91] James Allen Fill. Eigenvalue bounds on convergence to stationarity for nonreversible markov chains, with an application to the exclusion process. *The annals of applied probability*, pages 62–87, 1991. 9

[FK72] Cees M. Fortuin and Piet W. Kasteleyn. On the random-cluster model. I. Introduction and relation to other models. *Physica*, 57:536–564, 1972. 3

[FK13] James Allen Fill and Jonas Kahn. Comparison inequalities and fastest-mixing markov chains. *The Annals of Applied Probability*, pages 1778–1816, 2013. 50

[FY26] Weiming Feng and Minji Yang. Rapid mixing of glauber dynamics for monotone systems via entropic independence. In *SODA*, pages 4894–4929. SIAM, 2026. 5

[GJ18] Heng Guo and Mark Jerrum. Random cluster dynamics for the Ising model is rapidly mixing. *Ann. Appl. Probab.*, 28(2):1292–1313, 2018. 3

[GJP03] Leslie Ann Goldberg, Mark Jerrum, and Mike Paterson. The computational complexity of two-state spin systems. *Random Struct. Algorithms*, 23(2):133–154, 2003. 2, 3, 4

[GKZ18] Heng Guo, Kaan Kara, and Ce Zhang. Layerwise systematic scan: Deep boltzmann machines and beyond. In *AISTATS*, volume 84 of *Proceedings of Machine Learning Research*, pages 178–187. PMLR, 2018. 9

[GL18] Heng Guo and Pinyan Lu. Uniqueness, spatial mixing, and approximation for ferromagnetic 2-spin systems. *ACM Trans. Comput. Theory*, 10(4):17:1–17:25, 2018. 2, 4, 5, 7, 11, 12, 15, 47

[GLL20] Heng Guo, Jingcheng Liu, and Pinyan Lu. Zeros of ferromagnetic 2-spin systems. In *SODA*, pages 181–192. SIAM, 2020. 2, 4, 5

[GŠV16] Andreas Galanis, Daniel Štefankovič, and Eric Vigoda. Inapproximability of the partition function for the antiferromagnetic Ising and hard-core models. *Combin. Probab. Comput.*, 25(4):500–559, 2016. 1

[Hin02] Geoffrey E. Hinton. Training products of experts by minimizing contrastive divergence. *Neural Comput.*, 14(8):1771–1800, 2002. 1

[Hin12] Geoffrey E. Hinton. A practical guide to training restricted boltzmann machines. In *Neural Networks: Tricks of the Trade (2nd ed.)*, Lecture Notes in Computer Science, pages 599–619. Springer, 2012. 1, 2

[HOT06] Geoffrey E. Hinton, Simon Osindero, and Yee Whye Teh. A fast learning algorithm for deep belief nets. *Neural Comput.*, 18(7):1527–1554, 2006. 1

[JS93] Mark Jerrum and Alistair Sinclair. Polynomial-time approximation algorithms for the ising model. *SIAM J. Comput.*, 22(5):1087–1116, 1993. 3

[KQWW26] Youngwoo Kwon, Qian Qin, Guanyang Wang, and Yuchen Wei. A phase transition in sampling from restricted Boltzmann machines. *Ann. Appl. Probab.*, 2026+. to appear. 1, 2

[LLZ14] Jingcheng Liu, Pinyan Lu, and Chihao Zhang. The complexity of ferromagnetic two-spin systems with external fields. In *RANDOM*, LIPIcs, pages 843–856. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2014. 2, 3, 4, 5

[LP17] David A. Levin and Yuval Peres. *Markov chains and mixing times*. American Mathematical Society, second edition, 2017. 44, 49

[MH10] Abdel-rahman Mohamed and Geoffrey E. Hinton. Phone recognition using restricted Boltzmann machines. In *ICASSP*, pages 4354–4357. IEEE, 2010. 1

[MS13] Elchanan Mossel and Allan Sly. Exact thresholds for Ising-Gibbs samplers on general graphs. *Ann. Probab.*, 41(1):294–328, 2013. 5, 21

[nob24] Press release for the Nobel prize in physics. `https://www.nobelprize.org/prizes/physics/2024/press-release/`, 2024. Accessed: 2026-03-30. 1

[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. 4

[Sly10] Allan Sly. Computational transition at the uniqueness threshold. In *FOCS*, pages 287–296. IEEE Computer Society, 2010. 1

[SMH07] Ruslan Salakhutdinov, Andriy Mnih, and Geoffrey E. Hinton. Restricted Boltzmann machines for collaborative filtering. In *ICML*, pages 791–798. ACM, 2007. 1

[Smo86] Paul Smolensky. Parallel distributed processing: Explorations in the microstructure of cognition. *Information Processing in Dynamical Systems: Foundations of Harmony Theory*, pages 194–281, 1986. 1

[SS14] Allan Sly and Nike Sun. Counting in two-spin models on d-regular graphs. *Ann. Probab.*, 42(6):2383–2416, 2014. 1

[SS21] Shuai Shao and Yuxin Sun. Contraction: A unified perspective of correlation decay and zero-freeness of 2-spin systems. *J. Stat. Phys.*, 185(2):12, 2021. 5

[Tos16] Christopher Tosh. Mixing rates for the alternating Gibbs sampler over restricted Boltzmann machines and friends. In *ICML*, pages 840–849. JMLR.org, 2016. 1, 2

[Wei06] Dror Weitz. Counting independent sets up to the tree threshold. In *STOC*, pages 140–149. ACM, 2006. 4, 9, 10, 11

Appendix A. One-step decay in general settings

In this section, we prove Lemma 15 by generalising the proof in [GL18]. For any edge $e = (u, u_i)$, define the function $g_{\lambda,e}(x)$ for $x \in (0,\lambda)$ by:

$$g_{\lambda,e}(x) := \frac{(\beta_e\gamma_e - 1)x \log\frac{\lambda}{x}}{(\beta_e x + 1)(x+\gamma_e)\log\frac{x+\gamma_e}{\beta_e x+1}}.$$

We first prove that: for any $\beta_e \leqslant \beta \leqslant 1 < \gamma \leqslant \gamma_e$ and $\beta\gamma \geqslant \beta_e\gamma_e > 1$, there exists a constant $0 < \alpha < 1$ such that $g_{\lambda,e}(x) \leqslant 1 - \alpha$ for all $x \in (0,\lambda)$.

We have following lemmas about the function $g_{\lambda,e}(x)$.

Lemma 58 (Lemma 3.3, [GL18])**.** $g_{\lambda,e}(x) \leqslant g_{\lambda_c,e}(x) \leqslant 1$.

We have $\log\frac{x+\gamma_e}{\beta_e x+1} \geqslant \log\frac{\lambda+\gamma_e}{\beta_e\lambda+1} \geqslant \log\frac{\lambda+\gamma}{\lambda+1}$ for $x \in (0,\lambda)$, where the first inequality holds because $\log\frac{x+\gamma_e}{\beta_e x+1}$ is monotone decreasing in x. We can compute

$$\begin{aligned} g_{\lambda,e}(x) :=& \frac{(\beta_e\gamma_e - 1)x\log\frac{\lambda}{x}}{(\beta_e x+1)(x+\gamma_e)\log\frac{x+\gamma_e}{\beta_e x+1}} \\ \leqslant& \frac{(\beta\gamma-1)x\log\frac{\lambda}{x}}{\log\frac{x+\gamma_e}{\beta_e x+1}} \leqslant \frac{(\beta\gamma-1)x\log\frac{\lambda}{x}}{\log\frac{\lambda+\gamma}{\lambda+1}}, \end{aligned}$$

where the first inequality holds because $\beta_e x + 1 \geqslant 1$, $x + \gamma_e \geqslant 1$ and $0 < \beta_e\gamma_e - 1 \leqslant \beta\gamma - 1$. $(x\log\frac{\lambda}{x})' = \log\frac{\lambda}{x} - 1 \geqslant 0$ for $0 \leqslant x \leqslant \frac{\lambda}{e}$. We also have $x\log\frac{\lambda}{x} \to 0^+$ as $x \to 0^+$. Hence there exists a constant $x_0 = x_0(\lambda,\beta,\gamma) \in (0,\lambda)$ such that for any $x \in (0, x_0]$, we have

$$g_{\lambda,e}(x) \leqslant \frac{(\beta\gamma-1)x\log\frac{\lambda}{x}}{\log\frac{\lambda+\gamma}{\lambda+1}} \leqslant \frac{1}{2}.$$

For $x \in [x_0, \lambda)$, we have

$$g_{\lambda,e}(x) = g_{\lambda_c,e}(x)\cdot\frac{\log\lambda - \log x}{\log\lambda_c - \log x} \leqslant \cdot\frac{\log\lambda - \log x_0}{\log\lambda_c - \log x_0} < 1.$$

Then we can set $\alpha = 1 - \max\{\frac{1}{2}, \frac{\log\lambda - \log x_0}{\log\lambda_c - \log x_0}\}$ so that $g_{\lambda,e}(x) \leqslant 1 - \alpha < 1$ for all $x \in (0,\lambda)$.

Let $t = \frac{(1-\alpha)\gamma}{\beta\gamma-1}\log\frac{\lambda+\gamma}{\beta\lambda+1}$, then for any $x \in (0,\lambda)$, we have

$$t < \frac{(1-\alpha)(\beta_e x+1)(x+\gamma_e)}{\beta_e\gamma_e - 1}\log\frac{x+\gamma_e}{\beta_e x+1}, \tag{64}$$

because $(\beta_e x+1)(x+\gamma_e) > \gamma$, $\beta_e\gamma_e - 1 < \beta\gamma - 1$ and $\frac{\lambda+\gamma}{\beta\lambda+1} \leqslant \frac{\lambda+\gamma_e}{\beta_e\lambda+1} < \frac{x+\gamma_e}{\beta_e x+1}$. Therefore, if $\phi(x) = \frac{1}{t}$, by (64) we have

$$\frac{(\beta_e\gamma_e - 1)}{(\beta_e x+1)(x+\gamma_e)}\frac{1}{\phi(x)} \leqslant (1-\alpha)\log\frac{x_i+\gamma_e}{\beta_e x_i+1}.$$

If $\phi(x) = \frac{1}{x\log\frac{\lambda}{x}}$, because $g_{\lambda,e}(x) \leqslant 1-\alpha$, we also have

$$\frac{(\beta_e\gamma_e - 1)}{(\beta_e x+1)(x+\gamma_e)}\frac{1}{\phi(x)} \leqslant (1-\alpha)\log\frac{x_i+\gamma_e}{\beta_e x_i+1}.$$

Therefore, for any $x \in (0,\lambda)$, we have

$$\frac{(\beta_e\gamma_e - 1)}{(\beta_e x+1)(x+\gamma_e)}\frac{1}{\phi(x)} \leqslant (1-\alpha)\log\frac{x_i+\gamma_e}{\beta_e x_i+1}.$$

Now we can compute that

$$\begin{aligned}C_{\phi,d}(\boldsymbol{x}) =&\phi(F_u(\boldsymbol{x}))\sum_{i=1}^{d}\left|\frac{\partial F_u}{\partial x_i}(\boldsymbol{x})\right|\frac{1}{\phi(x_i)}\\ =&\phi(F_u(\boldsymbol{x}))\sum_{i=1}^{d}F_u(\boldsymbol{x})\frac{(\beta_{e_i}\gamma_{e_i}-1)}{(\beta_{e_i}x_i+1)(x_i+\gamma_{e_i})}\frac{1}{\phi(x_i)}\\ \leqslant&\phi(F_u(\boldsymbol{x}))\sum_{i=1}^{d}F_u(\boldsymbol{x})(1-\alpha)\log\frac{x_i+\gamma_{e_i}}{\beta_{e_i}x_i+1}\\ =&(1-\alpha)\phi(F_u(\boldsymbol{x}))F_u(\boldsymbol{x})\log\frac{\lambda_u}{F_u(\boldsymbol{x})}\\ \leqslant&(1-\alpha)\phi(F_u(\boldsymbol{x}))F_u(\boldsymbol{x})\log\frac{\lambda}{F_u(\boldsymbol{x})}\\ \leqslant&1-\alpha,\end{aligned}$$

where the last inequality holds because $\phi(F_u(\boldsymbol{x})) = \min\left\{\frac{1}{t}, \frac{1}{F_u(\boldsymbol{x})\log\frac{\lambda}{F_u(\boldsymbol{x})}}\right\} \leqslant \frac{1}{F_u(\boldsymbol{x})\log\frac{\lambda}{F_u(\boldsymbol{x})}}$.

Appendix B. Monotone coupling

B.1. **Proof of Proposition 29 and Proposition 30.** We first prove a lemma to show the monotonicity of the conditional marginal distribution induced by a ferromagnetic two-spin system.

Lemma 59. *Let μ be a Gibbs distribution of a ferromagnetic two-spin system on graph $G = (V, E)$. Consider any $\Lambda \subseteq V$ and two partial configurtions $\sigma \preceq \tau \in \{0,1\}^{\Lambda}$, it holds that*

$$\forall v \in V \setminus \Lambda, \quad \mu_v^{\sigma_{V\setminus\Lambda}}(1) \leqslant \mu_v^{\tau_{V\setminus\Lambda}}(1). \tag{65}$$

Proof. We fix a vertex $v \in V \setminus \Lambda$ and prove (65). Consider the SAW tree $T_\sigma = T_{\mathrm{SAW}}(G, v, \sigma)$ and $T_\tau = T_{\mathrm{SAW}}(G, v, \tau)$, which differ only in the pinning of the leaf nodes. For any vertex w in the SAW tree, define $p_w^{T_\sigma}$ and $p_w^{T_\tau}$ as the marginal probabilities of the vertex w in the sub-trees $T_{w,\sigma}$ and $T_{w,\tau}$ rooted at w, respectively. Because $\sigma \preceq \tau$, for any leaf node u with pinning in T_σ, T_τ, the pinning in T_σ is at most[4] the pinning in T_τ. We have $R_u^{T_\sigma} = \frac{p_u^{T_\sigma}(0)}{p_u^{T_\sigma}(1)} \geqslant R_u^{T_\tau} = \frac{p_u^{T_\tau}(0)}{p_u^{T_\tau}(1)}$. For each parameter of the recursion function $F_w(\cdot)$ in (11), where w is any non-leaf node in the SAW tree, the recursion function is monotone increasing with the parameter. We can recursively prove that for any non-leaf node w in the SAW tree, we have $R_w^{T_\sigma} \geqslant R_w^{T_\tau}$, from bottom to top. Using a inductive proof, for the root node v, we can show that $R_v^{T_\sigma} \geqslant R_v^{T_\tau}$, which implies $\mu_v^{\sigma_{V\setminus\Lambda}}(1) \leqslant \mu_v^{\tau_{V\setminus\Lambda}}(1)$. ∎

Now, we first prove Proposition 30 and then Proposition 29.

Let us consider the heat-bath block dynamics at first. Let $b = |\mathcal{B}|$. Assume $V = \{v_1, v_2, \ldots, v_n\}$. We construct the monotone coupling f as follows. For any configuration $\sigma \in \Omega$ and $r = (r_0 r_1 \ldots r_n) \in [0,1]^{n+1}$, we determine the configuration $f(\sigma, r) \in \{0,1\}^V$. There exists $i \in [1, b]$ such that $r_0 \in [(i-1)/b, i/b)$, and we choose the i-th block $B_i = \{v_{i_1}, v_{i_2}, \ldots, v_{i_j}\}$, where $1 \leqslant i_1 < i_2 < \ldots < i_j \leqslant n$. To simplify the notation, let $\rho = f(\sigma, r)$. We set $\rho(V \setminus B_i) = \sigma(V \setminus B_i)$. Let $B_i^k = \{v_{i_k}, v_{i_2}, \ldots, v_{i_j}\}$ for $1 \leqslant k \leqslant j+1$. We need to resample vertices in B_i conditioned on $\sigma(V \setminus B_i)$, and we recursively decide the value of $\rho(v_{i_k})$ in increasing order of k, such that

$$\rho(v_{i_k}) \sim \mu_{v_{i_k}}^{\rho(V\setminus B_i^k)}. \tag{66}$$

[4]Here, we compare the value $\{0,1\}$ of pinning. The value 0 is smaller than the value 1

Assume that we have decided the value of $\rho(V \setminus B_i^k)$. If $r_k \leqslant \mu_{v_{i_k}}^{\rho(V\setminus B_i^k)}(1)$, we set $\rho(v_{i_k}) = 1$. Otherwise, we set $\rho(v_{i_k}) = 0$. It is easy to verify that for any $\sigma \in \Omega$, the distribution of $f(\sigma, \mathbf{r}) = \rho$ is exactly the distribution of one step of the heat-bath block dynamics on μ starting from σ. This proves that f is a valid coupling. We only need to check that $f(\sigma, r) \preceq f(\tau, r)$ with probability 1 if $\sigma \preceq \tau$. To simplify the notation, let $\rho = f(\sigma, r)$ and $\rho' = f(\tau, r)$. We first have $\sigma(V \setminus B_i) = \rho(V \setminus B_i) \preceq \rho'(V \setminus B_i) = \tau(V \setminus B_i)$. Assume that we have $\rho(V \setminus B_i^k) \preceq \rho'(V \setminus B_i^k)$, for some $0 \leqslant k \leqslant j-1$. By Lemma 59, we have $\mu_{v_{i_k}}^{\rho(V\setminus B_i^k)}(1) \leqslant \mu_{v_{i_k}}^{\rho'(V\setminus B_i^k)}(1)$. By our construction, $\rho(v_{i_k}) \leqslant \rho'(v_{i_k})$, so we have $\rho(V \setminus B_i^{k+1}) \preceq \rho'(V \setminus B_i^{k+1})$. By induction, we have $\rho(V \setminus B_i^{j+1}) \preceq \rho'(V \setminus B_i^{j+1})$, which implies $\rho \preceq \rho'$. Hence, f is a monotone coupling of the heat-bath block dynamics on μ. Since our analysis holds for any $B_i \subseteq V$, we can couple two dynamics to pick the same block, then the block dynamics part in Proposition 30 holds.

For systematic scan block dynamics, we can construct the monotone coupling in the same way as above, except that we choose the block B_i according to the systematic scan order instead of the random choice. Using the above argument for each block B_i and doing an induction on all blocks shows that there exists a monotone coupling of the systematic-scan block dynamics on μ. This proves the systematic scan block dynamics part in Proposition 30.

Finally, Proposition 29 is a simple consequence of the above prove. The above proof works for all blocks $B_i \subseteq V$. Given two configurations $\sigma, \tau \in \{0,1\}^\Lambda$, where $\Lambda \subseteq V$ is any subset, if $\sigma \preceq \tau$, we can use the same process (with $B_i = V \setminus \Lambda$) to couple $X \sim \mu_{V\setminus\Lambda}^\sigma$ and $Y \sim \mu_{V\setminus\Lambda}^\tau$ such that $X \preceq Y$ with probability 1. This proves Proposition 29.

B.2. **Proof of Claim 35.** We first list some definitions about the comparison between Markov chains.

Definition 60 (Increasing function)**.** We say a function $f : \{0,1\}^V \to \mathbb{R}$ is increasing if for any $\sigma, \tau \in \{0,1\}^V$ with $\sigma \preceq \tau$, it holds that $f(\sigma) \leqslant f(\tau)$.

Definition 61 (Monotone Markov chain)**.** We say a Markov chain with transition matrix P on $\{0,1\}^V$ is monotone if for any increasing function $f : \{0,1\}^V \to \mathbb{R}$, Pf is also an increasing function.

Definition 62. Let μ be a distribution over $\{0,1\}^V$. For two monotone Markov chains P and Q on $\{0,1\}^V$, we say $P \preceq_{mc} Q$ if for any increasing function $f, g : \{0,1\}^V \to \mathbb{R}_+$, we have

$$\langle Pf, g \rangle_\mu \leqslant \langle Qf, g \rangle_\mu ,$$

where $\langle f_1, f_2 \rangle_\mu := \sum_{x\in\{0,1\}^V} f_1(x) f_2(x) \mu(x)$ for any functions $f_1, f_2 : \{0,1\}^V \to \mathbb{R}$.

Fix a distribution μ over $\{0,1\}^V$. For any block $B \subseteq V$, let P_B be the transition matrix of the block update on B: given any $\sigma \in \{0,1\}^V$, P_B resamples the configuration on B conditional on the current configuration of other variables: $\sigma(B) \sim \mu_B^{\sigma(V\setminus B)}$. Similarly, $P_{B\cap S}$ is the transition matrix of the block update on $B \cap S$. The following monotonicity result is known.

Lemma 63 ([BCV20], Proof of Lemma 15)**.** *For any block* $B \subseteq V$ *and subset* $S \subseteq V$,

$$P_B \preceq_{mc} P_{B\cap S}.$$

Next, recall that $\preceq_D$ is the stochastic dominance relation for two distributions defined in Claim 35.

Proposition 64 ([LP17, Proposition 22.7])**.** *For any Markov chain* P *on* $\{0,1\}^V$*, the following three statements are equivalent:*

- P *is a monotone Markov chain;*
- *For any two configurations* $\sigma, \sigma' \in \{0,1\}^V$ *with* $\sigma \preceq \sigma'$*, we have* $P(\sigma, \cdot) \preceq_D P(\sigma', \cdot)$*;*
- *for any two distributions* $\nu_0 \preceq_D \nu_1$*, we have* $\nu_0 P \preceq_D \nu_1 P$*.*

By the proof of Proposition 29, the second condition in Proposition 64 holds for transition matrix P_B corresponding to any block $B \subseteq V$. Hence, P_B is a monotone Markov chain.

Lemma 65 ([FK13], Prop. 2.3, 2.4)**.** *Let P_i, Q_i are Markov chains that are reversible w.r.t. μ and monotone for $i \in \{1, 2, \dots, \ell\}$, the following statements hold:*

- *If $P_i \preceq_{mc} Q_i$ for each $i \in \{1, 2, \dots, \ell\}$, then $\frac{1}{\ell}\sum_{i=1}^{\ell} P_i \preceq_{mc} \frac{1}{\ell}\sum_{i=1}^{\ell} Q_i$;*
- *If $P_i \preceq_{mc} Q_i$ for each $i \in \{1, 2, \dots, \ell\}$, then $P_1 P_2 \cdots P_\ell \preceq_{mc} Q_1 Q_2 \cdots Q_\ell$.*

The following lemma holds for both the heat-bath and the systematic scan block dynamics.

Lemma 66. *Let P be the transition matrix of a block dynamics on μ with a set of blocks $B = \{B_1, B_2, \dots, B_r\}$. Let P_S^{censored} be the transition matrix of the censored block dynamics on μ w.r.t. $S \subseteq V$, then $P \preceq_{mc} P_S^{\text{censored}}$.*

Proof. By Lemma 63, we have $P_{B_i} \preceq_{mc} P_{B_i \cap S}$ for each $i \in [r]$. We also have P_{B_i} and $P_{B_i \cap S}$ are reversible with stationary distribution μ for each $i \in [r]$. For the heat-bath block dynamics, by the first statement of Lemma 65, we have

$$P = \frac{1}{r}\sum_{i=1}^{r} P_{B_i} \preceq_{mc} \frac{1}{r}\sum_{i=1}^{r} P_{B_i \cap S} = P_S^{\text{censored}}.$$

Hence, the result holds for heat-bath block dynamics. For systematic scan block dynamics. By the second statement of Lemma 65, we have

$$P = P_{B_r} P_{B_{r-1}} \cdots P_{B_1} \preceq_{mc} P_{B_r \cap S} P_{B_{r-1} \cap S} \cdots P_{B_1 \cap S} = P_S^{\text{censored}}.$$

Hence, the result holds for systematic scan block dynamics. □

Let μ be a distribution over $\{0,1\}^V$. Let $\mathcal{A} \subseteq 2^V$ be a collection of censoring subsets. For any block dynamics P on $\{0,1\}^V$ with stationary distribution μ, and $P \preceq_{mc} P_S^{\text{censored}}$ for $S \in \mathcal{A}$. Let $S_1, S_2, \dots$ be a sequence of censoring subsets in $\mathcal{A}$. Let $(X_t)_{t \geqslant 0}$ be the heat-bath or systematic scan block dynamics on μ with transition matrix P and block set $\mathcal{B} = \{B_1, B_2, \dots, B_r\}$. Let $(Y_t)_{t \geqslant 0}$ be the censored block dynamics on μ with transition matrix $P_{S_i}^{\text{censored}}$ in step i. Formally, the transition matrix of $(Y_t)_{t \geqslant 0}$ in i-th step is

$$\begin{cases} P_{S_i}^{\text{censored}} = \frac{1}{r}\sum_{j=1}^{r} P_{B_j \cap S_i} & \text{if } P \text{ is heat-bath block dynamics,} \\ P_{S_i}^{\text{censored}} = P_{B_r \cap S_i} P_{B_{r-1} \cap S_i} \cdots P_{B_1 \cap S_i} & \text{if } P \text{ is systematic scan block dynamics.} \end{cases}$$

We use the following result in our proof.

Lemma 67 ([BCV20], Theorem 7)**.** *Suppose two initial configurations X_0, Y_0 are both sampled from the same distribution ν over $\{0,1\}^V$. The following properties hold:*

- *If ν/μ is increasing, where $\nu/\mu(x) = \frac{\nu(x)}{\mu(x)}$, then for any $t \geqslant 0$, $X_t \preceq_D Y_t$;*
- *If $-\nu/\mu$ is increasing, then for any $t \geqslant 0$, $Y_t \preceq_D X_t$.*

Now we are ready to prove Claim 35. For parameters in Lemma 67, we set $\mathcal{A} = \{V, S_v\}$, $B_i = V$ for $1 \leqslant i \leqslant s$ and $B_i = S_v$ for $i > s$. Applying the result in Lemma 66 we have $P \preceq_{mc} P_V = P$ and $P \preceq_{mc} P_{S_v}$. We first let $\nu = 1^V$ and apply the first statement in Lemma 67 to obtain that for any $j > s$, $X_j^+ \preceq_D Y_j^+$. Then we let $\nu = 0^V$ and apply the second statement in Lemma 67 to obtain that for any $j > s$, $Y_j^- \preceq_D X_j^-$. By inductively applying the third statement in Proposition 64, we have for any $j > s$, $X_j^- \preceq_D X_j^+$. Combining above three relationships, we have

$$\forall j \geqslant 0, \quad Y_j^- \preceq_D X_j^- \preceq_D X_j^+ \preceq_D Y_j^+.$$

Indeed, the above relationships hold for any $j \geqslant 0$.