Perfect sampling from spatial mixing

Received: 6 October 2020 | Revised: 26 September 2021 | Accepted: 17 December 2021

DOI: 10.1002/rsa.21079

RESEARCH ARTICLE

WILEY

Perfect sampling from spatial mixing

Weiming Feng[1] | **Heng Guo**[1] | **Yitong Yin**[2]

[1]School of Informatics, University of Edinburgh, Informatics Forum, Edinburgh, UK
[2]State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing, Jiangsu Province, China

Correspondence
Weiming Feng, School of Informatics, University of Edinburgh, Informatics Forum, Edinburgh EH8 9AB, UK.
Email: wfeng@ed.ac.uk

Funding information
European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme, Grant/Award Number: 947778

Abstract

We introduce a new perfect sampling technique that can be applied to general Gibbs distributions and runs in linear time if the correlation decays faster than the neighborhood growth. In particular, in graphs with subexponential neighborhood growth like $\mathbb{Z}^d$, our algorithm achieves linear running time as long as Gibbs sampling is rapidly mixing. As concrete applications, we obtain the currently best perfect samplers for colorings and for monomer-dimer models in such graphs.

KEYWORDS

Gibbs distribution, perfect sampling, spatial mixing

1 | INTRODUCTION

Spin systems model nearest neighbor interactions of complex systems. These models originated from statistical physics, and have found a wide range of applications in probability theory, machine learning, and theoretical computer science, often under different names such as *Markov random fields* or *Boltzmann machines*. Given an underlying graph $G = (V, E)$, a *configuration* σ is an assignment from vertices to a finite set of spins, usually denoted by $[q]$. The *weight* of a configuration is specified by the q-dimensional vector b_v assigned to each vertex $v \in V$ and the q-by-q symmetric interaction matrix A_e assigned to each edge $e \in E$, namely,

$$w(\sigma) = \prod_{v \in V} b_v(\sigma_v) \prod_{e=\{u,v\} \in E} A_e(\sigma_u, \sigma_v). \tag{1}$$

The equilibrium state of the system is described by the *Gibbs distribution* μ, where the probability of a configuration is proportional to its weight.

A central algorithmic problem related to spin systems is to sample from the Gibbs distribution. A canonical Markov chain for sampling approximately from the Gibbs distribution is the *Gibbs sampler* (a.k.a. *heat bath* or *Glauber dynamics*). The efficiency of the sampler is determined by its *mixing time*, namely how long it takes to converge to the desired distribution. One (conjectured) general criterion for rapid mixing of such Markov chains is the *spatial mixing* property [52], which roughly states that the correlation among variables decays rapidly in the system as their distances increase. It is widely believed that spatial mixing (in some form) implies the rapid mixing of the Gibbs sampler. However, rigorous implications have only been established for special classes of graphs or systems, such as for lattice graphs [13, 38], for colorings on neighborhood amenable graphs [21], for ferromagnetic Ising models [41], and very recently, for anti-ferromagnetic two-spin systems [3, 7–9].

One main drawback for Gibbs samplers or Markov chains, in general, is that one needs to know the mixing time in advance to implement the algorithms with provably small errors. The mixing time is usually hard to analyze and very pessimistic. The so-called *perfect samplers* are thus more desirable, which run in a Las Vegas fashion and return exact samples upon halting. There have been a number of techniques available to design perfect samplers, such as Coupling From The Past (CFTP) [42], including the monotone and anti-monotone CFTP [28, 42] and the bounding chains [5, 31], Randomness Recycler (RR) [18], and Partial Rejection Sampling (PRS) [16, 27]. Nevertheless, none of these techniques address general spin systems or relate to the important spatial mixing properties of the system.

In this paper, we introduce a new technique to perfectly sample from Gibbs distributions of spin systems. The correctness of our algorithm relies on only the conditional independence property of Gibbs distributions. Moreover, the expected running time is linear in the size of the system, when the correlation decays more rapidly than the growth of the neighborhood.

Theorem 1.1 (informal)**.** *For any spin system with bounded maximum degree,*[1] *if strong spatial mixing holds with a rate faster than the neighborhood growth of the underlying graph, then there exists a perfect sampler with running time $O(n)$ in expectation, where n represents the number of vertices of the graph.*

More details and undefined terms are explained in Section 2. Formal statements of our results are given in Theorem 2.4 for spin systems on subexponential neighborhood growth graphs, and in Theorem 2.6 for spin systems on general graphs. Applications on list colorings and on monomer-dimer models are given in Theorems 2.9 and 2.10.

Lattice graphs, such as $\mathbb{Z}^d$, are of special interests in statistical physics and combinatorics. These graphs have subexponential neighborhood growth, which implies that temporal mixing is equivalent to spatial mixing on them [13]. Therefore our sampler runs in linear time as long as the standard Glauber dynamics has $O(n \log n)$ mixing time. This is a direct strengthening of aforementioned results [13, 21, 38] from approximate to perfect sampling, with an improved running time.

Corollary 1.2 (informal)**.** *For spin systems on graphs with subexponential neighborhood growth, if the Gibbs sampler has $O(n \log n)$ mixing time, where n is the number of vertices, or the system shows strong spatial mixing, then there exists a perfect sampler with running time $O(n)$ in expectation.*

[1] We remark that our algorithm remains in polynomial-time (but not in linear time) for graphs with unbounded degrees, as long as the degree does not grow too quickly. This requirement on the degree comes from the cost of updating a block of certain radius, similar to the cost of block dynamics, and the exact upper bound needed varies from problem to problem. See Theorem 2.6 and the discussion thereafter. For the most part, we state our results for bounded degree cases to keep the statements clean.

It is worth noting that many traditional perfect sampling algorithms, especially those rely on CFTP [5, 31, 42], suffer from "non-interruptibility". That is, early termination of the algorithm induces a bias on the sample. In contrast, our algorithm is interruptible in the following sense: conditioned on its termination at any particular step, the algorithm guarantees to return a perfect sample. Therefore, had the algorithm been running for too long, one can simply stop it and restart. One can also run many independent copies in parallel and output the earliest returned sample without biasing the sample.

In addition, our algorithm can be used to solve the recently introduced dynamic sampling problem [15, 16], where the Gibbs distribution itself changes dynamically and the algorithm needs to efficiently maintain a sample from the current Gibbs distribution. The detail of this part is given in Section 7. Our perfect sampler also generalizes straightforwardly to Gibbs distributions with multi-body interactions (namely spin systems on hypergraphs / constraint satisfaction problems), and similar efficiency can be achieved when some appropriate variant of spatial mixing holds.

1.1 | Algorithm overview

We give an overview of our main new idea. We briefly review the *Gibbs sampler*, which is a Markov chain on the state space $[q]^V$. At each step, a vertex $u \in V$ is picked uniformly at random and the current configuration $\boldsymbol{X} \in [q]^V$ is updated by the simple rule:

- the spin X_u is redrawn according to the marginal distribution $\mu_u^{X_{\Gamma(u)}}$;

where $\Gamma(u) \triangleq \{v \in V | \{u, v\} \in E\}$ denotes the neighborhood of u, and $\mu_u^{X_{\Gamma(u)}}$ denotes the marginal distribution induced at vertex u by the Gibbs distribution μ, conditioned on the current spins of the neighborhood $\Gamma(u)$. It is a basic fact that this chain converges to the Gibbs distribution μ.

Our perfect sampler makes use of the same update rule. To expose our main idea, we first consider the single-site version of our perfect sampler, which works for systems with soft constraints (where all A_e and b_v are positive). The sampling algorithm is quite simple, described in Algorithm 1.

Algorithm 1: Perfect Gibbs sampler (single-site version)

```
Start from an arbitrary initial configuration X ∈ [q]^V and R ← V;
while R ≠ ∅ do
    pick a u ∈ R uniformly at random;
    let μ_min be the minimum value of μ_u^σ(X_u) over all σ ∈ [q]^Γ(u) that σ_{R∩Γ(u)} = X_{R∩Γ(u)};
    with probability μ_min / μ_u^{X_Γ(u)}(X_u) do        ▷ Bayes filter
        update X by redrawing X_u ~ μ_u^{X_Γ(u)};       ▷ Gibbs sampler update
        R ← R \ {u};
    else
        R ← R ∪ Γ(u);
return X;
```

The algorithm starts from an arbitrary initial configuration $\boldsymbol{X} \in [q]^V$, and gradually "repairs" $\boldsymbol{X}$ to a perfect sample drawn from the Gibbs distribution μ. We maintain a set $\mathcal{R} \subseteq V$ of vertices that are currently "incorrect," initially set as $\mathcal{R} = V$. At each step, a random vertex u is picked from $\mathcal{R}$, and we try to remove u from $\mathcal{R}$ while maintaining the following invariant, denoted by the *conditional Gibbs*

property:

$$\begin{aligned} &X_S \text{ always follows the law } \mu_S^{X_{\mathcal{R}}}, \\ &\text{which is the marginal distribution induced by } \mu \text{ on } S \triangleq V \setminus \mathcal{R} \text{ conditioned on } X_{\mathcal{R}}. \end{aligned} \tag{$\star$}$$

This property ensures that the configuration on S follows the desired distribution conditioned on the configuration on $\mathcal{R}$. In particular, when $\mathcal{R} = \emptyset$, $S = V$, $\mu_S^{X_{\mathcal{R}}} = \mu$, and the sample $\boldsymbol{X}$ follows precisely the distribution μ. This is the goal of our algorithm: reduce $\mathcal{R}$ to the empty set.

To do this, we try to eliminate a vertex u from $\mathcal{R}$. Since $(\boldsymbol{X}, \mathcal{R})$ satisfies the invariant ($\star$), it holds that $X_S \sim \mu_S^{X_{\mathcal{R}}}$, where $S = V \setminus \mathcal{R}$. To satisfy the invariant ($\star$), we update the configuration $\boldsymbol{X}$ as follows:

- use a filter $\mathcal{F}$ to change the distribution of X_S from $\mu_S^{X_{\mathcal{R}}}$ to $\mu_S^{X_{\mathcal{R}\setminus\{u\}}}$;
- resample X_u from distribution $\mu_u^{X_{\mathcal{R}\setminus\{u\}} \wedge X_S} = \mu_u^{X_{V\setminus\{u\}}} = \mu_u^{X_{\Gamma(u)}}$, where last equation holds due to the *conditional independence property* (formally, Property 4.3) of the Gibbs distribution.

After the two steps above, it is straightforward to verify the new pair $(\boldsymbol{X}, \mathcal{R} \setminus \{u\})$ satisfies the invariant ($\star$). We call $\mathcal{F}$ the *Bayes filter*, which is determined by a biased coin depending on only part of $\boldsymbol{X}$. For any $X_S \in [q]^S$, the success probability should be proportional to the following quantity:

$$\begin{aligned} \Pr[\ \mathcal{F} \text{ succeeds }] &\propto \frac{\mu_S^{X_{\mathcal{R}\setminus\{u\}}}(X_S)}{\mu_S^{X_{\mathcal{R}}}(X_S)} \overset{(\star)}{=} \frac{\mu_u^{X_{\mathcal{R}\setminus\{u\}}}(X_u)}{\mu_u^{X_{\Gamma(u)}}(X_u)} \\ \left(\text{as } \mu_u^{X_{\mathcal{R}\setminus\{u\}}}(X_u) \text{ is independent from } X_S\right) &\propto \frac{1}{\mu_u^{X_{\Gamma(u)}}(X_u)}. \end{aligned} \tag{2}$$

In Algorithm 1 we set $\Pr[\ \mathcal{F} \text{ succeeds}] = \mu_{\min} / \mu_u^{X_{\Gamma(u)}}(X_u)$ to satisfy the condition in (2). Here we choose $\mu_{\min}$ in the numerator so that the probability is bounded by 1. The key step in the above formula is the equation ($\star$), which holds because

$$\mu_S^{X_{\mathcal{R}}}(X_S) = \mu_S^{X_{\mathcal{R}\setminus\{u\}} \wedge X_u}(X_S) \overset{(\diamond)}{=} \frac{\mu_u^{X_{\mathcal{R}\setminus\{u\}} \wedge X_S}(X_u) \cdot \mu_S^{X_{\mathcal{R}\setminus\{u\}}}(X_S)}{\mu_u^{X_{\mathcal{R}\setminus\{u\}}}(X_u)},$$

together with the fact that $\mu_u^{X_{\mathcal{R}\setminus\{u\}} \wedge X_S}(X_u) = \mu_u^{X_{V\setminus\{u\}}}(X_u) = \mu_u^{X_{\Gamma(u)}}(X_u)$. The equation ($\diamond$) is obtained by applying Bayes' theorem on the distribution $\mu^{X_{\mathcal{R}\setminus\{u\}}}$. This explains the name "Bayes filter".

If the filter succeeds, the distribution of X_S follows the law $\mu_S^{X_{\mathcal{R}\setminus\{u\}}}$, and then we continue as above. Meanwhile, since the filter only reveals the neighborhood spins $X_{\Gamma(u)}$, upon the failure of $\mathcal{F}$, the invariant ($\star$) remains to hold as long as the revealed sites $\Gamma(u)$ are added into $\mathcal{R}$ and $\boldsymbol{X}$ is unchanged.

This sampler is valid for general Gibbs distributions, since the only property we require is conditional independence. However for efficiency purposes, our general algorithm, Algorithm 2, uses *block* updates. A suitable block radius (chosen according to spatial mixing) is the key to efficiency in our analysis. Moreover, to make sure that marginal distributions are well-defined, we restrict our attention to *permissive* systems, which contain all soft constraint systems as well as all hard constraint systems of interest. The details are given in Sections 2 and 3.

The algorithm is efficient as long as the size of $\mathcal{R}$ shrinks in expectation in every step. For the more general Algorithm 2, this holds true when the correlation decays faster than the neighborhood growth. The details are in Section 5.

1.2 | Related work

The conditional Gibbs property has been used implicitly or explicitly in previous works such as PRS [22, 24–27], dynamic sampling [16], and RR [18]. Furthermore, the invariance of the conditional Gibbs property ensures that the sampling algorithm is correct even when the input spin system is dynamically changing over time [16].

Before our work, all previous resampling algorithms [16, 18, 22, 24–27] fall into the paradigm of rejection sampling: a new sample is generated, usually from modifying the old sample, and (part of) the new sample is rejected independently with some probability determined by the new sample. In our algorithm, the filtration is executed *before* the generation of the new sample, with a bias independent of the new sample.

Spatial mixing properties were known to imply rapid mixing of the Gibbs sampler for some particular systems. See for example [41]. For anti-ferromagnetic two-spin systems, weak spatial mixing corresponds to the optimal threshold for efficient samplability [3, 8, 14, 19, 36, 44, 46, 52]. It remains an interesting open problem whether spatial mixing implies the existence of efficient samplers in general, and whether these samplers can be perfect.

2 | OUR RESULTS

2.1 | Model and definitions

Let $G = (V, E)$ be an undirected graph, and $[q] = \{1, 2, \dots, q\}$ a finite domain of $q \geq 2$ spins. An instance of q-state spin system is specified by a tuple $\mathcal{I} = (G, [q], \boldsymbol{b}, \boldsymbol{A})$, where $\boldsymbol{b} = (b_v)_{v\in V}$ assigns every vertex $v \in V$ a vector $b_v \in \mathbb{R}^q_{\geq 0}$ and $\boldsymbol{A} = (A_e)_{e\in E}$ assigns every edge $e \in E$ a symmetric matrix $A_e \in \mathbb{R}^{q\times q}_{\geq 0}$. The *Gibbs distribution* μ over $[q]^V$ is defined as

$$\forall \sigma \in [q]^V : \quad \mu(\sigma) \triangleq \frac{w(\sigma)}{Z} = \frac{1}{Z}\prod_{v\in V} b_v(\sigma_v) \prod_{e=\{u,v\}\in E} A_e(\sigma_u, \sigma_v), \tag{3}$$

where $w(\sigma)$ is the *weight* defined in (1) and $Z \triangleq \sum_{\sigma\in[q]^V} w(\sigma)$ is the *partition function*.

We restrict our attention to the so-called permissive spin systems, where the marginal distributions are always well-defined. Let $\mathcal{I} = (G, [q], \boldsymbol{b}, \boldsymbol{A})$ be an instance of spin system. A configuration on V is called *feasible* if its weight is positive, and a partial configuration is *feasible* if it can be extended to a feasible configuration. For any (possibly empty) subset $\Lambda \subseteq V$ and any (not necessarily feasible) partial configuration $\sigma \in [q]^\Lambda$, we use $w^\sigma(\tau)$ to denote the weight of $\tau \in [q]^{V\setminus\Lambda}$ conditional on σ:

$$w^\sigma(\tau) = \prod_{v\in V\setminus\Lambda} b_v(\tau_v) \prod_{\substack{e=\{u,v\}\in E\\ u,v\in V\setminus\Lambda}} A_e(\tau_u, \tau_v) \prod_{\substack{e=\{u,v\}\in E\\ u\in\Lambda, v\in V\setminus\Lambda}} A_e(\sigma_u, \tau_v). \tag{4}$$

Define the partition function Z^σ conditional on σ as $Z^\sigma \triangleq \sum_{\tau\in[q]^{V\setminus\Lambda}} w^\sigma(\tau)$.

Definition 2.1 (permissive)**.** A spin system $\mathcal{I} = (G, [q], \boldsymbol{b}, \boldsymbol{A})$, where $G = (V, E)$, is called *permissive* if $Z^\sigma > 0$ for any partial configuration $\sigma \in [q]^\Lambda$ specified on any subset $\Lambda \subseteq V$.

Permissive systems are very common, including, for examples, uniform proper q-coloring when $q \geq \Delta + 1$, where Δ is the maximum degree, and spin systems with soft constraints, for example, the

Ising model, or with a "permissive" state that is compatible with all other states, for example, the hardcore model.

For permissive systems, a feasible configuration is always easy to construct by greedy algorithm. More importantly, with permissiveness, marginal probabilities are always well defined, which is crucial for Gibbs sampler and spatial mixing property.

Formally, we use μ^σ to denote the conditional distribution over $[q]^{V\setminus\Lambda}$ given $\sigma \in [q]^\Lambda$, that is,

$$\forall \tau \in [q]^{V\setminus\Lambda}, \quad \mu^\sigma(\tau) \triangleq \frac{w^\sigma(\tau)}{Z^\sigma}. \tag{5}$$

And for any $v \in V \setminus \Lambda$, we use μ_v^σ to denote the marginal distribution at v projected from μ^σ.

For any $u, v \in V$, we use $\mathrm{dist}_G(u, v)$ to denote the shortest-path distance between v and u in G.

Definition 2.2 (strong spatial mixing [51, 52])**.** Let $\delta : \mathbb{N} \to \mathbb{R}^+$. A class $\mathfrak{F}$ of permissive spin systems is said to exhibit strong spatial mixing with rate $\delta(\cdot)$ if for every instance $\mathcal{I} = (G, [q], \boldsymbol{b}, \boldsymbol{A}) \in \mathfrak{F}$, where $G = (V, E)$, for every $v \in V$, $\Lambda \subseteq V$, and any two partial configurations $\sigma, \tau \in [q]^\Lambda$,

$$d_{\mathrm{TV}}(\mu_v^\sigma, \mu_v^\tau) \le \delta(\ell), \tag{6}$$

where $\ell = \min\{\mathrm{dist}_G(v, u) | u \in \Lambda,\ \sigma_u \neq \tau_u\}$, and $d_{\mathrm{TV}}(\mu_v^\sigma, \mu_v^\tau) \triangleq \frac{1}{2}\sum_{a\in[q]} |\mu_v^\sigma(a) - \mu_v^\tau(a)|$ denotes the total variation distance between μ_v^σ and μ_v^τ. In particular, $\mathfrak{F}$ exhibits strong spatial mixing with exponential decay if (6) is satisfied for $\delta(\ell) = \alpha \exp(-\beta\ell)$ for some constants $\alpha, \beta > 0$.

Our first result holds for spin systems on graphs with bounded neighborhood growth.

Definition 2.3 (subexponential neighborhood growth)**.** A class $\mathfrak{G}$ of graphs is said to have subexponential neighborhood growth if there is a function $s : \mathbb{N} \to \mathbb{N}$ such that $s(\ell) = \exp(o(\ell))$ and for every graph $G = (V, E) \in \mathfrak{G}$,

$$\forall v \in V, \forall \ell \ge 0, \quad |B_\ell(v)| \le s(\ell),$$

where $B_\ell(v) \triangleq \{u \in V | \mathrm{dist}_G(v, u) \le \ell\}$ denotes the ball of radius ℓ centered at v in G.

Here $s(\ell) = \exp(o(\ell))$ is a subexponential function, for example, $s(\ell) = \mathrm{poly}(\ell)$ or $s(\ell) = \exp(\ell^{1-\epsilon})$. Note that graphs with subexponential neighborhood growth necessarily have bounded maximum degree because we can set $\ell = 1$ and get $s(1) = O(1)$.

2.2 | Main results

Our first result shows that for spin systems on graphs with subexponential neighborhood growth, strong spatial mixing implies the existence of linear-time perfect sampler.

Theorem 2.4 (main theorem: bounded-growth graphs)**.** *Let $q > 1$ be a finite integer and $\mathfrak{F}$ a class of permissive q-state spin systems on graphs with subexponential neighborhood growth. If $\mathfrak{F}$ exhibits strong spatial mixing with exponential decay, then there exists an algorithm such that given any instance $\mathcal{I} = (G, [q], \boldsymbol{b}, \boldsymbol{A}) \in \mathfrak{F}$, the algorithm outputs a perfect sample from μ within $O(n)$ time in expectation, where n is the number of vertices in G.*

The factor in $O(\cdot)$ is $q^{O(\Delta^{\ell_0})}$, which is the cost for an update in (ℓ_0-block) Gibbs sampler, where $\ell_0 = O(1)$ is the radius at which the decay catches up with neighborhood growth, as formally given in (33).

It is already known that for spin systems on subexponential neighborhood growth graphs, the strong spatial mixing with exponential decay condition implies $O(n \log n)$ mixing time for block Gibbs sampler [13], which generates approximate samples. We give a perfect sampler with $O(n)$ expected running time under the same condition. The notion of subexponential neighborhood growth is related to, but should be distinguished from neighborhood-amenability (see e.g. [21]), which says that in an infinite graph, for any constant real $c > 0$, there is an ℓ such that $\frac{|S_{\ell+1}(v)|}{|B_\ell(v)|} \leq c$ holds everywhere.

Our main result on general graphs assumes the following strong spatial mixing condition.

Condition 2.5. Let $\mathcal{I} = (G, [q], \boldsymbol{b}, \boldsymbol{A})$ be a permissive spin system where $G = (V, E)$. There is an integer $\ell = \ell(q) \geq 2$ such that the following holds: for every $v \in V$, $\Lambda \subseteq V$, for any two partial configurations $\sigma, \tau \in [q]^\Lambda$ satisfying $\min\{\mathrm{dist}_G(v, u) | u \in \Lambda,\ \sigma(u) \neq \tau(u)\} = \ell$,

$$d_{\mathrm{TV}}(\mu_v^\sigma, \mu_v^\tau) \leq \frac{\gamma}{5|S_\ell(v)|}, \tag{7}$$

where $S_\ell(v)$ denotes the sphere of radius ℓ centered at v in G, and

$$\gamma = \gamma(v, \Lambda) \triangleq \min\left\{\mu_v^\rho(a) | \rho \in [q]^\Lambda, a \in [q] \text{ that } \mu_v^\rho(a) > 0\right\}, \tag{8}$$

denotes the lower bound of positive marginal probabilities at v.

The above condition basically says that the spin systems exhibit strong spatial mixing with a decay rate faster than that of neighborhood growth, given that the marginal probabilities are appropriately lower bounded (which holds with $\gamma = \Omega(1)$ when entries of $\boldsymbol{A}$ and $\boldsymbol{b}$ are of finite precision and the maximum degree Δ is finitely bounded). Our result on general graphs is stated as the following theorem.

Theorem 2.6 (main theorem: general graphs)**.** *Let $\mathfrak{F}$ be a class of permissive spin systems satisfying Condition 2.5. There exists an algorithm which given any instance $\mathcal{I} = (G, [q], \boldsymbol{b}, \boldsymbol{A}) \in \mathfrak{F}$, outputs a perfect sample from μ within $n \cdot q^{O(\Delta^\ell)}$ time in expectation, where n is the number of vertices in G, Δ is the maximum degree of G, and $\ell = \ell(q)$ is determined by Condition 2.5.*

The $q^{O(\Delta^\ell)}$ factor in the time cost is contributed by the block Gibbs sampler update on ℓ-radius blocks. This extra cost could remain polynomial if $q = \omega(1)$, but the upper bound on q for that will vary from problem to problem. See for example, the discussion after Theorem 2.9.

We remark that Theorem 2.6 is not a generalization of Theorem 2.4, since we need the marginal lower bound γ in (7). In fact, The conditions of both Theorems 2.4 and 2.6 are special cases of the more technical Condition 5.1, stated in Section 5. Condition 5.1 is a multiplicative version of strong spatial mixing (SSM). In Theorem 2.6, the marginal lower bound is used to derive that from the standard total variation distance SSM. For graphs with subexponential neighborhood growth in Theorem 2.4, the two versions are equivalent, as shown in Proposition 6.1.

2.3 | Applications on specific systems

Our results can be applied on various spin systems. We consider four important examples:

- **Uniform list coloring:** A list coloring instance is specified by $\mathcal{I} = (G, [q], \mathcal{L})$, where $\mathcal{L} \triangleq \{L_v \subseteq [q] | v \in V\}$ assigns each vertex $v \in V$ a list of colors $L_v \subseteq [q]$. A proper list coloring of instance

$\mathcal{I}$ is a $\sigma \in [q]^V$ that $\sigma_v \in L_v$ for all $v \in V$ and $\sigma_u \neq \sigma_v$ for all $\{u, v\} \in E$. Let μ denote the uniform distribution over all proper list colorings of $\mathcal{I}$.

- **The monomer-dimer model:** The monomer-dimer model defines a distribution over matchings in graphs. An instance is specified by $\mathcal{I} = (G, \lambda)$, where $G = (V, E)$ is a graph and $\lambda > 0$. Each matching $M \subseteq E$ in G is assigned a weight $w(M) = \lambda^{|M|}$. Let μ be the distribution over all matchings in G such that $\mu(M) \propto w(M)$.
- **The hardcore model:** The hardcore model defines a distribution over independent sets in graphs. An instance is specified by $\mathcal{I} = (G, \lambda)$, where $G = (V, E)$ is a graph and $\lambda > 0$. Each independent set $S \subseteq V$ in G is assigned a weight $w(S) = \lambda^{|S|}$. Let μ be the distribution over all independent sets in G such that $\mu(S) \propto w(S)$.
- **The Ising model:** An Ising model instance is specified by $\mathcal{I} = (G, \beta, \lambda)$, where $G = (V, E)$ is a graph, $\beta > 0$ is the edge activity, and $\lambda > 0$ is the external field. Each configuration $\sigma \in \{-1, +1\}^V$ is assigned a weight $w(\sigma) = \beta^{|m(\sigma)|}\lambda^{|n_+(\sigma)|}$, where $m(\sigma) = |\{\{u, v\} \in E | \sigma_u = \sigma_v\}|$ denotes the number of monochromatic edges and $n_+(\sigma)$ denotes the number vertices that take the value $+1$. The Gibbs distribution μ is defined by $\mu(\sigma) \propto w(\sigma)$. The Ising model is ferromagnetic if $\beta > 1$ and is anti-ferromagnetic if $\beta < 1$.
- **The anti-ferromagnetic two-spin system:** An anti-ferromagnetic two-spin system is specified by $\mathcal{I} = (G, \lambda, \beta, \gamma)$, where $G = (V, E)$ is a graph, $0 \leq \beta \leq \gamma$, $\beta\gamma < 1$, $\gamma > 0$ and $\lambda > 0$. For any configuration $\sigma \in \{-1, +1\}^V$, its weight is defined by $w(\sigma) = \lambda^{n_+(\sigma)}\beta^{m_+(\sigma)}\gamma^{m_-(\sigma)}$, where $n_+(\sigma) = |\{v \in V | \sigma_v = +1\}|$, $m_+(\sigma) = |\{\{u, v\} \in E | \sigma_u = \sigma_v = +1\}|$ and $m_-(\sigma) = |\{\{u, v\} \in E | \sigma_u = \sigma_v = -1\}|$. The Gibbs distribution μ is defined by $\mu(\sigma) \propto w(\sigma)$.

We use $\deg_G(v)$ to denote the degree of v in G and $\Delta \triangleq \max_{v \in V} \deg_G(v)$ the maximum degree.

First, for the list coloring problem, we define the following two conditions for instance $\mathcal{I} = (G, [q], \mathcal{L})$. Let $\deg_G(v)$ denote the degree of $v \in V$ in graph G, and $\Delta \triangleq \max_{v \in V} \deg_G(v)$ the maximum degree.

Condition 2.7. For every $v \in V$, $|L(v)| \geq \alpha \deg_G(v) + \beta$, where either one of the followings holds:

- $\alpha = 2$ and $\beta = 0$;
- G is triangle-free, $\alpha > \alpha^*$ where $\alpha^* = 1.763\cdots$ is the positive root of $x^x = \mathrm{e}$, and $\beta \geq \frac{\sqrt{2}}{\sqrt{2}-1}$ satisfies $(1 - 1/\beta)\alpha \mathrm{e}^{\frac{1}{\alpha}(1-1/\beta)} > 1$.

Condition 2.8. For every $v \in V$, $|L(v)| \geq \Delta^2 - \Delta + 2$.

Our perfect sampler for list coloring runs in linear time in either above condition.

Theorem 2.9. *Let $\mathfrak{L}$ be a class of list coloring instances with at most q colors for a constant $q > 0$. If either of the two followings holds for all instances $\mathcal{I} = (G, [q], \mathcal{L}) \in \mathfrak{L}$:*

- *Condition 2.7 and G has sub-exponential neighborhood growth; or*
- *Condition 2.8,*

then there exists a perfect sampler for μ that runs in expected $O(n)$ time, where n is the number of vertices.

The constant factor in $O(\cdot)$, can be determined in the same way as in Theorem 2.4 in the case of Condition 2.7, but is much higher in the case of Condition 2.8 due to the arbitrary neighborhood growth.

Nevertheless, even in this costly case, $\ell = O(q^2 \log q)$ and the overall overhead can by upper bounded by a rough estimate $\exp(\exp(\text{poly}(q)))$. In fact, our proof works for q up to $O\left(\left(\frac{\log\log n}{\log\log\log n}\right)^{1/2}\right)$ so that the run time remains a polynoimal in n.

Sampling proper q-colorings (where the lists are identical for all vertices) has been extensively studied, especially using Markov chains. Approximate sampling received considerable attention [11, 12, 29, 33, 40]. The current best result [6, 50] is the $O(n \log n)$ mixing time of the flip chain if $q \geq \left(\frac{11}{6} - \epsilon_0\right)\Delta$ for a small constant $\epsilon_0 \approx 10^{-5}$. For perfect sampling q-colorings, Huber introduced a bounding chain [31] based on CFTP [53], which terminates within $O(n \log n)$ steps in expectation if $q \geq \Delta^2 + 2\Delta$. In a recent breakthrough, Bhandari and Chakraborty [5] introduced a novel bounding chain that has expected running time $O\left(n \log^2 n\right)$ in a substantially broader regime $q > 3\Delta$. Very recently, the regime for bounding chain was further improved to $q \geq \left(\frac{8}{3} + o(1)\right)\Delta$ [32]. Another way to obtain perfect samplers is to use standard reductions between counting and sampling [35]. Using this technique, any FPTAS for the number of colorings can be turned into a polynomial-time perfect sampler. (It is important that the approximate counting algorithm is deterministic, or at least with errors that can be detected.) Currently, the FPTAS with the best regime for general graphs is due to Liu et al. [37], which requires $q \geq 2\Delta$, and it runs in time $n^{\text{EXP}(\Delta)}$.

Comparing to the results above, our algorithm draws perfect samples and achieves the $O(n)$ expected running time. In case of sub-exponential growth graphs such as $\mathbb{Z}^d$, it improves the result of [5] by requiring only $q > \alpha\Delta + O(1)$, where $\alpha > \alpha^* = 1.763\cdots$.

Next we consider the monomer-dimer model. It was proved in [4, 47] that instances $\mathcal{I} = (G, \lambda)$ on graphs G with maximum degree Δ exhibits strong spatial mixing with rate $(1 - \Omega(1/\sqrt{1 + \lambda\Delta}))^{-\ell}$. Applying our algorithm yields the following perfect sampling result.

Theorem 2.10. *Let $\mathfrak{M}$ be a class of monomer-dimer instances $\mathcal{I} = (G, \lambda)$ on graphs G with sub-exponential neighborhood growth and $\lambda = O(1)$. There exists an algorithm which given any instance $\mathcal{I} = (G, \lambda) \in \mathfrak{M}$, outputs a perfect sample from μ within expected $O(n)$ time.*

Note that, if $\lambda = \omega(1)$, then the rate of correlation decay is not fast enough for our need.

Previously, Markov chains were the most successful techniques for sampling weighted matchings. The Jerrum-Sinclair chain [34] on a monomer-dimer model $\mathcal{I} = (G, \lambda)$, generates approximate samples from μ within $\tilde{O}(n^2 m)$ steps, where $m = |E|$. This chain also mixes in $O\left(n \log^2 n\right)$ time for finite subgraphs of the 2D lattice $\mathbb{Z}^2$ [49].

It is difficult to convert the Jerrum-Sinclair chain to perfect samplers. Before our work, the only perfect sampler for the monomer-dimer model we are aware of is the one obtained via standard reductions from sampling to counting [35] (similar to the case of colorings), together with deterministic approximate counting algorithms [4]. This is a perfect sampler with running time $n^{\text{Poly}(\Delta,\lambda)}$.

Our algorithm is the first linear-time perfect sampler for the monomer-dimer model on graphs with sub-exponential neighborhood growth, such as finite subgraphs of lattices $\mathbb{Z}^d$ for any constant d and constant weight λ.

The hardcore model is another widely studied spin system in the sampling context. Due to the strong spatial mixing result established by Weitz [51], we have the following perfect sampling algorithm.

Theorem 2.11. *Let $\mathfrak{M}$ be a class of hardcore instances $\mathcal{I} = (G, \lambda)$ on graphs G with sub-exponential neighborhood growth and constant λ such that $\lambda < \lambda_c(\Delta) = \frac{(\Delta-1)^{(\Delta-1)}}{(\Delta-2)^{\Delta}}$, where Δ is the maximum degree of G. There exists an algorithm which given any instance $\mathcal{I} = (G, \lambda) \in \mathfrak{M}$, outputs a perfect sample from μ within expected $O(n)$ time.*

The condition $\lambda < \lambda_c(\Delta) = \frac{(\Delta-1)^{(\Delta-1)}}{(\Delta-2)^{\Delta}}$ is known as the uniqueness condition for hardcore model. In general graphs, if $\lambda > \lambda_c(\Delta)$, the sampling problem is intractable [45]; if $\lambda < \lambda_c(\Delta)$, the sampling problem can be solved in polynomial time [52]. Specifically, the Glauber dynamics is rapid mixing up to the uniqueness threshold in general graphs [3, 7–9].

For perfect sampling, the PRS works for general graphs but it requires a strong condition $\lambda \leq \frac{1}{\sqrt{2}\Delta-1}$ [16, 27]. The bounding chain also requires $\lambda \leq \frac{2}{\Delta-2}$ [30]. Our algorithm requires only the uniqueness condition when restricted to sub-exponential neighborhood growth graphs.

The last model is the Ising model. Due to the strong spatial mixing result in [36, 41, 43], we have the following perfect sampling algorithm.

Theorem 2.12. *Let $\mathfrak{M}$ be a class of Ising instances $\mathcal{I} = (G, \beta, \lambda)$ on graphs G with sub-exponential neighborhood growth and constant β such that $\frac{\Delta-2}{\Delta} < \beta < \frac{\Delta}{\Delta-2}$, where Δ is the maximum degree of G. There exists an algorithm which given any instance $\mathcal{I} = (G, \beta, \lambda) \in \mathfrak{M}$, outputs a perfect sample from μ within expected $O(n)$ time.*

Note that Theorem 2.12 holds for arbitrary external fields $\lambda_v > 0$ (which may vary for different vertices $v \in V$). The condition $\frac{\Delta-2}{\Delta} < \beta < \frac{\Delta}{\Delta-2}$ corresponds to the uniqueness condition for arbitrary external fields $\lambda_v > 0$, under which the Glauber dynamics is rapid mixing in general graphs [7–9, 41]. For uniform external field $\lambda > 0$ and the anti-ferromggnetic Ising model ($\beta < 1$), the same bound in Theorem 2.12 holds for a refined regime for the uniqueness as formulated in [44].

For perfecting sampling, in monotone systems, the CFTP technique is efficient as long as the corresponding Markov chain is rapid mixing [42]. The Ising model is monotone if it is ferromagnetic ($\beta \geq 1$). Thus, rapid mixing of the Glauber dynamics implies efficient perfect sampling when $1 \leq \beta < \frac{\Delta}{\Delta-2}$. Alternatively, via CFTP of the random cluster dynamics [23], efficient perfect sampling exists for all $\beta \geq 1$ and $\lambda = 1$. However, the expected running time of CFTP above is $\tilde{O}(n^2)$ for Glauber dynamics or $\tilde{O}(n^{10})$ for random cluster dynamics. PRS runs in expected linear time for both ferromagnetic and anti-ferromagnetic Ising models, which requires $\lambda = 1$ and $1 - \frac{1}{2.22\Delta+1} \leq \beta \leq 1 + \frac{1}{2.22\Delta}$ [16]. Our algorithm achieves the same linear expected run time for the whole uniqueness regime in sub-exponential neighborhood growth graphs.

In fact, our result holds more generally for all anti-ferromagnetic two-spin systems in the uniqueness regime as formulated in [36]. Let $\mathcal{I} = (G, \lambda, \beta, \gamma)$ be an anti-ferromagnetic two-spin system. For any integer $d \geq 1$, let x_d be the unique positive fixed point of the function $f_d(x) = \lambda\left(\frac{\beta x+1}{x+\gamma}\right)^d$. Let Δ denote the maximum degree of G. We say $\mathcal{I}$ is up-to-Δ unique if for all $1 \leq d < \Delta$, $|f_d'(x_d)| < 1$.

Theorem 2.13. *Let $\mathfrak{M}$ be a class of anti-ferromagnetic two-spin system instances $\mathcal{I} = (G, \lambda, \beta, \gamma)$ on graphs G with sub-exponential neighborhood growth and constants λ, β, γ such that $\mathcal{I}$ is up-to-Δ unique, where Δ is the maximum degree of G. There exists an algorithm which given any instance $\mathcal{I} = (G, \lambda, \beta, \gamma) \in \mathfrak{M}$, outputs a perfect sample from μ within expected $O(n)$ time.*

3 | THE ALGORITHM

We now describe our general perfect Gibbs sampler. It generalizes the single-site version (Algorithm 1) by allowing block updates. This generalization allows us to bypass some pathological situations, and to greatly improve the efficiency of the algorithm. The pseudocode is given in Algorithm 2.

Algorithm 2: Perfect Gibbs sampler (general version)

Parameter: an integer $\ell \geq 0$;

1 Start from an arbitrary feasible configuration $\boldsymbol{X} \in [q]^V$, i.e.$\sim w(\boldsymbol{X})>0$;
2 $\mathcal{R} \leftarrow V$;
3 **while** $\mathcal{R} \neq \emptyset$ **do**
4 pick a $u \in \mathcal{R}$ uniformly at random and let $B \leftarrow (B_\ell(u) \setminus \mathcal{R}) \cup \{u\}$;
5 let $\mu_{\min}$ be the minimum value of $\mu_u^\sigma(X_u)$ over all $\sigma \in [q]^{\partial B}$ that $\sigma_{\mathcal{R}\cap\partial B} = X_{\mathcal{R}\cap\partial B}$;
6 **with probability** $\frac{\mu_{\min}}{\mu_u^{X_{\partial B}}(X_u)}$ **do** ▷ `Bayes filter`
7 update $\boldsymbol{X}$ by redrawing $X_B \sim \mu_B^{X_{\partial B}}$; ▷ `block Gibbs sampler update`
8 $\mathcal{R} \leftarrow \mathcal{R} \setminus \{u\}$;
9 **else**
10 $\mathcal{R} \leftarrow \mathcal{R} \cup \partial B$;
11 **return** $\boldsymbol{X}$;

Let $\mathcal{I} = (G, [q], \boldsymbol{b}, \boldsymbol{A})$ be a permissive spin system instance and $G = (V, E)$. For any $u \in V$ and integer $\ell \geq 0$, recall that $B_\ell(u) \triangleq \{v \in V | \mathrm{dist}_G(u, v) \leq \ell\}$ denotes the ℓ-ball centered at u in G. And for any $B \subseteq V$, we use $\partial B \triangleq \{v \in V \setminus B | \exists w \in B, \{v, w\} \in E\}$ to denote the vertex boundary of B in G.

The algorithm is parameterized by an integer $\ell \geq 1$, which is set in Section 5.

The initial $\boldsymbol{X} \in [q]^V$ is an arbitrary feasible configuration, which is easy to construct by greedy algorithm since $\mathcal{I}$ is permissive (Definition 2.1). After each iteration of the **while** loop, either $\boldsymbol{X}$ is unchanged or X_B is redrawn from $\mu_B^{X_{\partial B}}$, which is the marginal distribution of μ, conditioned on the current $\boldsymbol{X}_B$. Thus, we have the following observation.

Observation 3.1. *In Algorithm 2, the configuration $\boldsymbol{X}$ is always feasible, i.e. $w(\boldsymbol{X}) > 0$.*

The observation implies that $\mu_u^{X_{\partial B}}(X_u) > 0$ all along. The Bayes filter in Line 5 is always well-defined. If the filter succeeds, X_B is resampled according to the correct marginal distribution $\mu_B^{X_{\partial B}}$ and u is removed from $\mathcal{R}$ (that is, u has been successfully "fixed"); otherwise, $\boldsymbol{X}$ is unchanged and $\mathcal{R}$ is enlarged by ∂B (because variables in ∂B are revealed and no longer random).

The key to the correctness of Algorithm 2 is the conditional Gibbs property ($\star$): the law of $\boldsymbol{X}$ over $\overline{\mathcal{R}} \triangleq V \setminus \mathcal{R}$ is always the conditional distribution $\mu^{X_{\mathcal{R}}}$. By similar argument as in Section 1.1, just redrawing X_B from $\mu_B^{X_{\partial B}}$ will introduce a bias $\propto \mu_u^{X_{\partial B}}(X_u)$ to the sample $\boldsymbol{X}$, relative to its target distribution $\mu^{X_{\mathcal{R}\setminus\{u\}}}$. In the algorithm, we use the Bayes filter that succeeds with probability $\propto 1/\mu_u^{X_{\partial B}}(X_u)$ to cancel this bias, with the risk of enlarging $\mathcal{R}$ by ∂B upon failure. Balancing the success probability and the size of ∂B is the key to getting an efficient algorithm, and this depends on choosing an appropriate ℓ according to the spatial mixing rate.

The efficiency of the algorithm, on the other hand, depends on the success rate for the filter at Line 6: if its success probability p is always close enough to 1, so that $(1-p)|\partial B| < p$, then set $\mathcal{R}$ shrinks in expectation in the worst case, and the algorithm converges geometrically. Meanwhile, the success probability $p = \mu_{\min}/\mu_u^{X_{\partial B}}(X_u)$ is always close enough to 1 when the values of the marginal probability $\mu_u^\sigma(X_u)$ vary little for all $\sigma \in [q]^{\partial B}$ that $\sigma_{\mathcal{R}\cap\partial B} = X_{\mathcal{R}\cap\partial B}$, which is a decay of correlation property.

The correctness and efficiency of the algorithm are analyzed respectively in next two sections.

4 | CORRECTNESS OF THE PERFECT SAMPLING

In this section, we prove the correctness of Algorithm 2, stated by the following theorem.

Theorem 4.1 (correctness theorem)**.** *Given any permissive spin system $\mathcal{I} = (G, [q], \boldsymbol{b}, \boldsymbol{A})$, Algorithm 2 with any parameter $\ell \geq 1$ terminates with probability 1, and outputs $\boldsymbol{X}$ that follows the law of its Gibbs distribution μ.*

The theorem is implied by two key properties of the Gibbs distribution μ.

4.1 | Key properties of Gibbs distributions

Note that the $\mu_{\min}$ in Algorithm 2 is determined by the set $\mathcal{R}$, the vertex $u \in \mathcal{R}$, and the partial feasible configuration $X_{\mathcal{R}}$. Formally, fixing the parameter $\ell \geq 0$ in Algorithm 2,

$$\mu_{\min}(\mathcal{R}, u, X_{\mathcal{R}}) \triangleq \min\left\{\mu_u^{\sigma}(X_u) | \sigma \in [q]^{\partial B} \text{ s.t. } \sigma_{\mathcal{R}\cap\partial B} = X_{\mathcal{R}\cap\partial B}, \text{ where } B \triangleq (B_\ell(u) \setminus \mathcal{R}) \cup \{u\}\right\}.$$

Property 4.2 (positive lower bound of $\mu_{\min}$)**.** *The lower bound γ of $\mu_{\min}$ is positive:*

$$\gamma \triangleq \min\left\{\mu_{\min}(\mathcal{R}, u, X_{\mathcal{R}}) | \mathcal{R} \subseteq V, u \in \mathcal{R}, X_{\mathcal{R}} \in [q]^{\mathcal{R}} \text{ s.t. } X_{\mathcal{R}} \text{ is feasible}\right\} > 0. \tag{9}$$

To state the next property, we introduce some notations: For any $\Lambda \subseteq V, \sigma \in [q]^{\Lambda}$ and $S \subseteq V \setminus \Lambda$, we use $\mu_S^{\sigma}(\cdot)$ to denote the marginal distribution on S projected from μ^{σ}. For any disjoint sets $\Lambda, \Lambda' \subseteq V$, $\sigma \in [q]^{\Lambda}$ and $\sigma' \in [q]^{\Lambda'}$, we use $\sigma \uplus \sigma'$ to denote the configuration on $\Lambda \uplus \Lambda'$ that is consistent with σ on Λ and consistent with σ' on Λ'.

Property 4.3 (conditional independence)**.** *Suppose $A, B, C \subset V$ are three disjoint non-empty subsets such that the removal of C disconnects A and B in G. For any $\sigma_A \in [q]^A, \sigma_B \in [q]^B$ and $\sigma_C \in [q]^C$,*

$$\mu_B^{\sigma_A \uplus \sigma_C}(\sigma_B) = \mu_B^{\sigma_C}(\sigma_B).$$

Theorem 4.1 is proved for general distribution μ over $[q]^V$ relying only on these two properties. Thus, Algorithm 2 is correct for general permissive Gibbs distributions.

In particular, we verify that all permissive spin systems satisfy these two properties. First, the conditional independence (Property 4.3) holds generally for Gibbs distributions [39]. Next, for the positive lower bound of $\mu_{\min}$ (Property 4.2): for spin systems with soft constraints, clearly Property 4.2 holds for all $\ell \geq 0$; and for general permissive spin systems $\mathcal{I}$, we need to verify that Property 4.2 holds if $\ell \geq 1$. Fix a tuple $(\mathcal{R}, u, X_{\mathcal{R}})$ in (9). The following fact follows from the definition of set B.

Fact 4.4. $\partial B \subseteq S_{\ell+1}(u) \cup \mathcal{R}$, where $B = (B_\ell(u) \setminus \mathcal{R}) \cup \{u\}$.

The fact implies $\partial B \setminus \mathcal{R} \subseteq S_{\ell+1}(u)$. Since $\ell \geq 1$, u is not adjacent to any vertex in $\partial B \setminus \mathcal{R}$. Since $\mathcal{I}$ is permissive and $X_{\mathcal{R}}$ is feasible, $\mu_u^{\sigma}(X_u) > 0$ for all $\sigma \in [q]^{\partial B}$ such that $\sigma_{\mathcal{R}\cap\partial B} = X_{\mathcal{R}\cap\partial B}$. This implies $\mu_{\min}(\mathcal{R}, u, X_{\mathcal{R}})$ is positive, thus the Property 4.2 holds.

We then prove Theorem 4.1 assuming only Property 4.2 and Property 4.3. More specifically, termination of the algorithm is guaranteed by Property 4.2, and correctness of the output upon termination is guaranteed by Property 4.3.

4.2 | Termination of the algorithm

Denote by T the number of iterations of the **while** loop in Algorithm 2. To prove that the algorithm terminates with probability 1, we show that T is stochastically dominated by a geometric distribution. We use $\mathcal{F}$ to denote the Bayes filter in Algorithm 2. Then,

$$\Pr[\mathcal{F} \text{ succeeds}] = \frac{\mu_{\min}(\mathcal{R}, u, X_{\mathcal{R}})}{\mu_u^{X_{\partial B}}(X_u)} \geq \mu_{\min}(\mathcal{R}, u, X_{\mathcal{R}}).$$

If $\mathcal{F}$ succeeds for $n = |V|$ consecutive iterations of the **while** loop, then the set $\mathcal{R}$ must become empty and the algorithm terminates. By Property 4.2, we have

$$\forall k \geq 0, \quad \Pr[T \geq kn] \leq (1 - \gamma^n)^k. \tag{10}$$

This implies T is stochastically dominated by a geometric distribution. Each iteration of the **while** loop terminates within finite number of steps. Thus, the algorithm terminates with probability 1.

4.3 | Correctness upon termination

We show that upon termination, the output follows the correct distribution. Let $(\boldsymbol{X}, \mathcal{R}) \in [q]^V \times 2^V$ be the random pair maintained by the algorithm. The following condition is the "loop invariant" of the random pair $(\boldsymbol{X}, \mathcal{R})$.

Condition 4.5 (conditional Gibbs property)**.** For any $R \subseteq V$ and $\sigma \in [q]^R$, conditioned on $\mathcal{R} = R$ and $X_R = \sigma$, the random configuration $X_{V \setminus R}$ follows the law μ^σ.

Condition 4.5 is satisfied initially by the initial pair $(\boldsymbol{X}, \mathcal{R}) = (\boldsymbol{X}, V)$. Furthermore, consider the **while** loop that transforms

$$(\boldsymbol{X}, \mathcal{R}) \to (\boldsymbol{X}', \mathcal{R}').$$

Then next lemma shows that Condition 4.5 holds inductively assuming Property 4.3.

Lemma 4.6. *Suppose that $(\boldsymbol{X}, \mathcal{R}) \in [q]^V \times 2^V$ is a random pair such that $\boldsymbol{X}$ is feasible and the pair $(\boldsymbol{X}, \mathcal{R})$ satisfies Condition 4.5. Then, the random pair $(\boldsymbol{X}', \mathcal{R}')$ satisfies Condition 4.5.*

Our algorithm can be viewed as a Markov chain $(\boldsymbol{X}_t, \mathcal{R}_t)_{t \geq 0}$. By Observation 3.1, the random configuration $\boldsymbol{X}_t \in [q]^V$ maintained by algorithm is always feasible. Initially, $\mathcal{R}_0 = V$. Thus, Lemma 4.6 guarantees that for any $t \geq 0$, the $(\boldsymbol{X}_t, \mathcal{R}_t)$ maintained by Algorithm 2 satisfies Condition 4.5. In particular, for any $t \geq 0$, conditional on $\mathcal{R}_t = \emptyset$, the output $\boldsymbol{X}_t$ follows the correct distribution μ. This proves Theorem 4.1.

Proof of Lemma 4.6. It is sufficient to prove that for any $R \subseteq V$, any feasible partial configuration $\rho \in [q]^R$ and any vertex $u \in R$, conditioned on $\mathcal{R} = R$, $X_R = \rho$, and the vertex picked in Line 4 being u, the new random pair $(\boldsymbol{X}', \mathcal{R}')$ after one iteration of the **while** loop satisfies Condition 4.5.

Fix $R \subseteq V$ and a feasible partial configuration $\rho \in [q]^R$. Let $\mathbf{u} \in R$ denote the uniform random vertex picked in Line 4. Fix a vertex $u \in R$. Let $\mathcal{E}$ denote the event

$$\mathcal{E} : \quad X_R = \rho \wedge \mathcal{R} = R \wedge \mathbf{u} = u.$$

Since $(\boldsymbol{X}, \mathcal{R})$ satisfies Condition 4.5 and given the set R, $\mathbf{u}$ is independent from $\boldsymbol{X}$, we have

$$\forall \tau \in [q]^{V\setminus R}: \quad \Pr\left[X_{V\setminus R} = \tau | \mathcal{E}\right] = \mu^{\rho}(\tau). \tag{11}$$

Recall that $\mathcal{F}$ is the Bayes filter. Depending on whether $\mathcal{F}$ succeeds or not, we have two cases.

The easier case is when $\mathcal{F}$ fails. Recall that the set B is fixed by R and u. In this case, $\mathcal{R}' = R \cup \partial B$ and $\boldsymbol{X}' = \boldsymbol{X}$. Conditioned on $\mathcal{E}$, we know that $X_u = \rho_u$ and $X_{\partial B \cap R} = \rho_{\partial B \cap R}$, the filter $\mathcal{F}$ depends only on the partial configuration $X_{\partial B \setminus R}$. For any configuration $\sigma \in [q]^{\partial B \setminus R}$, conditioned on $\mathcal{E}$ and $X_{\partial B \setminus R} = \sigma$, the failure of $\mathcal{F}$ is independent from $X_{V\setminus(R\cup\partial B)} = X_{V\setminus\mathcal{R}'}$. Thus, by (11), conditioned on $\mathcal{E}$, $X_{\partial B\setminus R} = \sigma$ and the failure of $\mathcal{F}$, we have that $X'_{V\setminus\mathcal{R}'} = X_{V\setminus\mathcal{R}'} \sim \mu^{\rho\uplus\sigma}$, i.e. $(\boldsymbol{X}', \mathcal{R}')$ satisfies Condition 4.5.

Now we analyze the main case that $\mathcal{F}$ succeeds. If this case does occur, we must have

$$\mu_{\min}(R, u, \rho) \triangleq \min\{\mu_u^{\sigma}(\rho_u) | \sigma \in [q]^{\partial B} \text{ s.t. } \sigma_{R\cap\partial B} = \rho_{R\cap\partial B}\} > 0. \tag{12}$$

Define $R_u \triangleq R \setminus \{u\}$. The fact that $\mathcal{F}$ succeeds means $\mathcal{R}' = R \setminus \{u\} = R_u$ and $X'_{R_u} = X_{R_u} = \rho_{R_u}$. Hence, we only need to show that

$$\forall \tau \in [q]^{V\setminus R_u}: \quad \Pr\left[X'_{V\setminus R_u} = \tau | \mathcal{E} \wedge \mathcal{F} \text{ succeeds}\right] = \mu^{\rho(R_u)}(\tau). \tag{13}$$

Recall that $B = (B_\ell(u) \setminus R) \cup \{u\} = B_\ell(u) \setminus R_u$. We define the following set:

$$H \triangleq V \setminus \{B_\ell(u) \cup R\} = V \setminus \{R_u \cup B\}.$$

Notice that the the whole set V is partitioned into three disjoint sets $H \uplus B \uplus R_u$. Namely, B is the set whose configuration is resampled, and $H \uplus R_u$ is the set whose configuration is untouched, i.e. $X'_{H\cup R_u} = X_{H\cup R_u}$. Note that $B \uplus H \uplus R_u = V$. By the chain rule:

$$\begin{aligned}\Pr\left[X'_{V\setminus R_u} = \tau \wedge \mathcal{F} \text{ succeeds} | \mathcal{E}\right] &= \Pr\left[X'_H = \tau_H \wedge X'_B = \tau_B \wedge \mathcal{F} \text{ succeeds} | \mathcal{E}\right] \\ &= \Pr[X'_H = \tau_H | \mathcal{E}] \cdot \Pr[\mathcal{F} \text{ succeeds} | \mathcal{E} \wedge X'_H = \tau_H] \cdot \Pr[X'_B = \tau_B | \mathcal{E} \wedge X'_H = \tau_H \wedge \mathcal{F} \text{ succeeds}].\end{aligned} \tag{14}$$

As $X'_H = X_H$, (11) implies that

$$\Pr[X'_H = \tau_H | \mathcal{E}] = \mu_H^{\rho}(\tau_H).$$

By Line 7 of Algorithm 2, X'_B is redrawn from the distribution $\mu_B^{X_{\partial B}}(\cdot)$. By conditional independence property (Property 4.3), we have $\mu_B^{X_{\partial B}}(\cdot) = \mu^{X_{V\setminus B}}(\cdot)$. Note that $V \setminus B = R_u \uplus H$. Conditioned on $\mathcal{E} \wedge X'_H = \tau_H$, $X_{R_u} = \rho_{R_u}$ and $X_H = X'_H = \tau_H$, thus $\mu_B^{X_{\partial B}}(\cdot) = \mu^{\rho(R_u)\uplus\tau(H)}(\cdot)$. Hence,

$$(14) = \mu_H^{\rho}(\tau_H) \cdot \mu^{\rho(R_u)\uplus\tau(H)}(\tau_B) \cdot \Pr[\mathcal{F} \text{ succeeds} | \mathcal{E} \wedge X'_H = \tau_H]. \tag{15}$$

To finish the proof, we need to calculate $\Pr[\mathcal{F} \text{ succeeds} | \mathcal{E} \wedge X'_H = \tau_H]$. This is done by the following claim, whose proof is deferred to the end of the section. Recall that ρ is a configuration on R.

Claim 4.7. Assume (12). It holds that

$$\mu_H^{\rho}(\tau_H) > 0 \iff \mu_H^{\rho(R_u)}(\tau_H) > 0. \tag{16}$$

Furthermore, for τ_H such that $\Pr[X'_H = \tau_H|\mathcal{E}] = \mu_H^{\rho}(\tau_H) > 0$,

$$\Pr[\mathcal{F} \text{ succeeds}|\mathcal{E} \wedge X'_H = \tau_H] = C \cdot \frac{\mu_H^{\rho(R_u)}(\tau_H)}{\mu_H^{\rho}(\tau_H)}, \tag{17}$$

where $0 < C = C(R, u, \rho) \leq 1$ is a constant depending only on R, u, ρ but not on τ.

Combining (15) and Claim 4.7, we have

$$\forall \tau \in [q]^{V \setminus R_u}, \quad \Pr\left[X'_{V \setminus R_u} = \tau \wedge \mathcal{F} \text{ succeeds}|\mathcal{E}\right] = C \cdot \mu^{\rho(R_u)}(\tau). \tag{18}$$

This equation can be verified in two cases:

- If $\mu_H^{\rho}(\tau_H) = 0$, then by (16), $\mu_H^{\rho(R_u)}(\tau_H) = 0$, thus LHS = RHS = 0.
- If $\mu_H^{\rho}(\tau_H) > 0$, by (15) and (17), we have LHS $= C \cdot \mu_H^{\rho(R_u)}(\tau_H) \cdot \mu^{\rho(R_u) \uplus \tau(H)}(\tau_B) =$ RHS, where the last equation holds because $\tau \in [q]^{V \setminus R_u}$ and $V \setminus R_u = H \uplus B$.

Thus, the probability that $\mathcal{F}$ succeeds is

$$\Pr[\mathcal{F} \text{ succeeds}|\mathcal{E}] = \sum_{\sigma \in [q]^{V \setminus R_u}} \Pr\left[X'_{V \setminus R_u} = \sigma \wedge \mathcal{F} \text{ succeeds}|\mathcal{E}\right] = \sum_{\sigma \in [q]^{V \setminus R_u}} C \cdot \mu^{\rho(R_u)}(\sigma) = C, \tag{19}$$

where the last equation holds because $\sum_{\sigma \in [q]^{V \setminus R_u}} \mu^{\rho(R_u)}(\sigma) = 1$ and $C = C(R, u, \rho) > 0$ is a constant depending only on R, u, ρ. Thus, for any $\tau \in [q]^{V \setminus R_u}$, combining (18) and (19), we have

$$\Pr\left[X'_{V \setminus R_u} = \tau|\mathcal{F} \text{ succeeds} \wedge \mathcal{E}\right] = \frac{\Pr\left[X'_{V \setminus R_u} = \tau \wedge \mathcal{F} \text{ succeeds}|\mathcal{E}\right]}{\Pr[\mathcal{F} \text{ succeeds}|\mathcal{E}]} = \frac{C \cdot \mu^{\rho(R_u)}(\tau)}{C} = \mu^{\rho(R_u)}(\tau),$$

where the last equation holds due to $C = C(R, u, \rho) > 0$. This proves (13). ∎

Proof of Claim 4.7. We first introduce the following definitions. Recall that $R_u \uplus B \uplus H = V$. We further partition ∂B into two disjoint sets $\partial B \cap R_u$ and $\partial B \setminus R_u$. Define

$$\begin{aligned} S &\triangleq \partial B \setminus R_u = \partial B \cap H, \\ \Psi &\triangleq \partial B \cap R_u = \partial B \cap R. \end{aligned} \tag{20}$$

We now prove (16). Since $\rho = \rho(R_u) \uplus \rho(u)$, by the Bayes law, we have the following relation between $\mu_H^{\rho(R_u)}(\tau_H)$ and $\mu_H^{\rho}(\tau_H)$:

$$\mu_H^{\rho}(\tau_H) = \mu_H^{\rho(R_u) \uplus \rho(u)}(\tau_H) = \frac{\mu_u^{\rho(R_u) \uplus \tau(H)}(\rho_u)}{\mu_u^{\rho(R_u)}(\rho_u)} \cdot \mu_H^{\rho(R_u)}(\tau_H). \tag{21}$$

Note that $\rho \in [q]^R$ is a feasible configuration, thus $\mu_u^{\rho(R_u)}(\rho_u) > 0$ and the above ratio is well-defined. Note that $u \in B$ and $R_u \uplus B \uplus H = V$. The set ∂B separates u from $(R_u \uplus H) \setminus \partial B$. Note that $\partial B = S \uplus \Psi$, where S and Ψ is defined in (20). By the conditional independence property (Property 4.3), we have

$$\mu_u^{\rho(R_u) \uplus \tau(H)}(\rho_u) = \mu_u^{\rho(\Psi) \uplus \tau(S)}(\rho_u) \geq \mu_{\min}(R, u, \rho) > 0, \tag{22}$$

where the first inequality is because $\mu_{\min}(R,u,\rho)$ in (12) can be rewritten as $\min_{\eta\in[q]^S}\mu_u^{\rho(\Psi)\uplus\eta}(\rho_u)$, and the second inequality is because $\mu_{\min}(R,u,\rho)>0$ due to the lower bound in (12).

Next, we prove (17). Suppose $\mu_H^\rho(\tau_H)>0$. Combining (21) and (22), it remains to prove that

$$\Pr[\mathcal{F}\ \text{succeeds}|\mathcal{E}\wedge X'_H=\tau_H]=C\cdot\frac{\mu_H^{\rho(R_u)}(\tau_H)}{\mu_H^{\rho}(\tau_H)}\overset{(\star)}{=}C\cdot\frac{\mu_u^{\rho(R_u)}(\rho_u)}{\mu_u^{\rho(\Psi)\uplus\tau(S)}(\rho_u)}. \tag{23}$$

To verify the equality $(\star)$, we write $\mu_H^\rho(\tau_H)=\mu_H^{\rho_u\uplus\rho(R_u)}(\tau_H)=\Pr_{X\sim\mu^{\rho(R_u)}}[X_H=\tau_H|X_u=\rho_u]$. By Bayes's theorem, we have the following identity

$$\begin{aligned}\mu_H^\rho(\tau_H)&=\Pr_{X\sim\mu^{\rho(R_u)}}[X_H=\tau_H|X_u=\rho_u]=\frac{\Pr_{X\sim\mu^{\rho(R_u)}}[X_u=\rho_u|X_H=\tau_H]\Pr_{X\sim\mu^{\rho(R_u)}}[X_H=\tau_H]}{\Pr_{X\sim\mu^{\rho(R_u)}}[X_u=\rho_u]}\\&=\frac{\mu_u^{\rho(R_u)\uplus\tau_H}(\rho_u)\cdot\mu_H^{\rho(R_u)}(\tau_H)}{\mu_u^{\rho(R_u)}(\rho_u)}\overset{(*)}{=}\frac{\mu_u^{\rho(\Psi)\uplus\tau(S)}(\rho_u)\cdot\mu_H^{\rho(R_u)}(\tau_H)}{\mu_u^{\rho(R_u)}(\rho_u)},\end{aligned}$$

where the equality $(*)$ holds due to the conditional independence property (since $V=H\uplus B\uplus R_u$).

Conditional on $\mathcal{E}$, we have $X_\Psi=\rho_\Psi$ and $X_u=\rho_u$. Recall that $X'_H=X_H$, $S\subseteq H$ and $S\uplus\Psi=\partial B$. Conditional on $X'_H=\tau_H$, it holds that $X_S=\tau_S$. By the definition of the filter $\mathcal{F}$ in Line 5 of Algorithm 2, we have that

$$\Pr\left[\mathcal{F}\ \text{succeeds}|\mathcal{E}\wedge X'_H=\tau_H\right]=\frac{\mu_{\min}(\mathcal{R},u,X_{\mathcal{R}})}{\mu_u^{X_{\partial B}}(X_u)}=\frac{\mu_{\min}(R,u,\rho)}{\mu_u^{\rho(\Psi)\uplus\tau(S)}(\rho_u)}. \tag{24}$$

Combining (24) and (23), we can set $C=C(R,u,\rho)$ in (23) as

$$C=C(R,u,\rho)\triangleq\frac{\mu_{\min}(R,u,\rho)}{\mu_u^{\rho(R_u)}(\rho_u)}=\frac{1}{\mu_u^{\rho(R_u)}(\rho_u)}\cdot\min_{\eta\in[q]^S}\mu_u^{\rho(\Psi)\uplus\eta}(\rho_u)\le 1. \tag{25}$$

Note that $\mu_u^{\rho(R_u)}(\rho_u)>0$ because ρ is feasible, and $\mu_{\min}(R,u,\rho)>0$ due to the lower bound in (12). This implies $C(R,u,\rho)>0$. Remark the sets S and Ψ are determined by R and u. This implies that the $C(R,u,\rho)$ defined as above depends only on R,u,ρ. This proves (17). ∎

5 | EFFICIENCY UNDER STRONG SPATIAL MIXING

In this section, we prove the following result.

Condition 5.1. Let $\mathcal{I}=(G,[q],\boldsymbol{b},\boldsymbol{A})$ be a permissive spin system where $G=(V,E)$. There is an integer $\ell=\ell(q)\ge 2$ such that the following holds: for every $v\in V$, $\Lambda\subseteq V$, and any two partial configurations $\sigma,\tau\in[q]^\Lambda$ satisfying $\min\{\mathrm{dist}_G(v,u)|u\in\Lambda,\ \sigma_u\neq\tau_u\}=\ell$,

$$\forall a\in[q]:\quad\left|\frac{\mu_v^\sigma(a)}{\mu_v^\tau(a)}-1\right|\le\frac{1}{5\,|S_\ell(v)|}\quad(\text{with the convention }\ 0/0=1), \tag{26}$$

where $S_\ell(v)\triangleq\{u\in V|\mathrm{dist}_G(v,u)=\ell\}$ denotes the sphere of radius ℓ centered at v in G.

Theorem 5.2. *Let $\mathfrak{F}$ be a class of permissive spin systems satisfying Condition 5.1. Given any instance $\mathcal{I} = (G, [q], \boldsymbol{b}, \boldsymbol{A}) \in \mathfrak{F}$, the Algorithm 2 with parameter $\ell = \ell^* - 1$ outputs a perfect sample from μ within $O\left(n \cdot q^{2\Delta^{\ell^*}}\right)$ time in expectation, where n is the number of vertices in G, Δ is the maximum degree of G, $\ell^* = \ell^*(q) \geq 2$ is determined by Condition 5.1, and $O(\cdot)$ hides only absolute constants.*

The correctness part of Theorem 5.2 follows from Theorem 4.1, we focus on the efficiency part. The proof sketch is that if $\mathfrak{F}$ satisfies Condition 5.1 with parameter ℓ^*, we set the parameter ℓ in Algorithm 2 as $\ell = \ell^* - 1$. We prove that after each iteration of the **while** loop, the size of $\mathcal{R}$ decreases by at least $\frac{1}{5}$ in expectation. Note that the initial $\mathcal{R} = V$, thus the initial size of $\mathcal{R}$ is n. By the optional stopping theorem, the number of iterations of the **while** loop is $O(n)$ in expectation. One can verify that the time complexity of the **while** loop is $O\big(q^{2\Delta^{\ell+1}}\big) = O\big(q^{2\Delta^{\ell^*}}\big)$. This proves the running time result.

To analyze the expected size of $\mathcal{R}$ after each iteration of the **while** loop, we analyze the Bayes filter $\mathcal{F}$ in Line 5. The probability that $\mathcal{F}$ fails is $1 - \mu_{\min}/\mu_u^{X_{\partial B}}(X_u)$. Suppose $\sigma \in [q]^{\partial B}$ achieves $\mu_{\min} = \mu_u^{\sigma}(X_u)$. By Fact 4.4, we can verify that σ and $X_{\partial B}$ can be differ only at $\partial B \setminus \mathcal{R} \subseteq S_{\ell+1}(u) = S_{\ell^*}(u)$. By Condition 5.1, we know that $\Pr[\mathcal{F} \text{ fails}] \leq \frac{1}{5|S_{\ell^*}(u)|}$. In addition, we have

- if $\mathcal{F}$ succeeds, the size of $\mathcal{R}$ decreases by 1;
- if $\mathcal{F}$ fails, the size of $\mathcal{R}$ increases by $|\partial B \setminus \mathcal{R}|$, it easy to verify $\partial B \setminus \mathcal{R} \subseteq S_{\ell^*}(u)$ by Fact 4.4, thus, the size of $\mathcal{R}$ increases by at most $|S_{\ell^*}(u)|$.

Thus, after each iteration of the **while** loop, the size of $\mathcal{R}$ decreases by at least $\frac{1}{5}$ in expectation.
In the formal proof, we actually prove a stronger result. We first introduce the following condition.

Condition 5.3. Let $\mathcal{I} = (G, [q], \boldsymbol{b}, \boldsymbol{A})$ be a permissive spin system where $G = (V, E)$. There is an integer $\ell = \ell(q) \geq 2$ such that the following holds: for every $v \in V$, any two disjoint sets $A, B \subseteq V$ with $\mathrm{dist}_G(v, B) = \min\{\mathrm{dist}_G(v, u) | u \in B\} = \ell$, and any partial configuration $\sigma \in [q]^A$,

$$\forall a \in [q], \quad 1 - \frac{\min\limits_{\tau \in [q]^B} \mu_v^{\sigma \uplus \tau}(a)}{\mu_v^{\sigma}(a)} \leq \frac{1}{5\,|S_\ell(v)|} \quad \text{(with the convention } 0/0 = 1\text{)}, \tag{27}$$

where $S_\ell(v) \triangleq \{u \in V | \mathrm{dist}_G(v, u) = \ell\}$ denotes the sphere of radius ℓ centered at v in G.

It is straightforward to verify that Condition 5.1 implies Condition 5.3. In the rest of this section, we prove that the efficiency result in Theorem 5.2 holds under Condition 5.3.

Let $\mathcal{I} = (G, [q], \boldsymbol{b}, \boldsymbol{A}) \in \mathfrak{F}$ be the input instance satisfying Condition 5.3 with some $\ell^* \geq 2$. Set the parameter ℓ in Algorithm 2 as $\ell = \ell^* - 1$. Denote by T the number of iterations of the **while** loop in Algorithm 2. To prove the efficiency result in Theorem 5.2, we bound the maximum running time of the **while** loop and the expectation of T in the following two lemmas.

Lemma 5.4. *The running time of each **while** loop is at most $O(q^{2\Delta^{\ell+1}}) = O(q^{2\Delta^{\ell^*}})$.*

Lemma 5.5. $\mathbb{E}[T] \leq 5n$.

Since the input instance $\mathcal{I}$ is permissive (Definition 2.1) , the initial feasible configuration can be constructed by a simple greedy algorithm. The running time of the first two lines in

Algorithm 2 is $O(n\Delta)$. Combining this with Lemma 5.4 and Lemma 5.5 proves the efficiency result in Theorem 5.2.

Proof of Lemma 5.4. We first show that $\mu_u^{X_{\partial B}}(X_u)$ can be computed in time $O(q^{\Delta^{\ell+1}})$, where $O(\cdot)$ hides an absolute constant. Let $\tilde{G} = G[B \cup \partial B]$ and $\tilde{\mathcal{I}}$ be the instance restricted to $\tilde{G}$. Let $\tilde{\mu}$ denote the Gibbs distribution defined by $\tilde{\mathcal{I}}$. By the conditional independence property (Property 4.3), it is straightforward to verify

$$\mu_u^{X_{\partial B}}(X_u) = \tilde{\mu}_u^{X_{\partial B}}(X_u) \tag{28}$$

since ∂B separates B from $V \setminus (B \cup \partial B)$ and $u \in B$. Since $|B| \leq |B_\ell(u)| \leq \frac{\Delta^{\ell+1}-1}{\Delta-1} \leq \Delta^{\ell+1}$, it takes at most $O(q^{\Delta^{\ell+1}})$ to enumerate all possibilities and compute $\mu_u^{X_{\partial B}}(X_u)$ using (28). By Fact 4.4, $\partial B \subseteq S_{\ell+1}(u) \cup R$. Since $|\partial B \setminus R| \leq |S_{\ell+1}(u)| \leq \Delta^{\ell+1}$, we can enumerate all $[q]^{\partial B \setminus R}$ to compute $\mu_{\min}$ in time $O(q^{2\Delta^{\ell+1}})$. The total running time for the first three lines of the **while** loop is at most $O(q^{2\Delta^{\ell+1}})$.

Another non-trivial computation is to sample $X(B)$ from $\mu^{X_{\partial B}}$. Similar to (28), conditional independence implies that this can be done by straightforward enumeration in time $O(q^{\Delta^{\ell+1}})$. The total running time of the **while** loop is thus $O(q^{2\Delta^{\ell+1}}) = O(q^{2\Delta^{\ell^*}})$. ∎

Proof of Lemma 5.5. Define a sequence of random pairs $(\boldsymbol{X}_0, \mathcal{R}_0), (\boldsymbol{X}_1, \mathcal{R}_1), \ldots, (\boldsymbol{X}_T, \mathcal{R}_T)$, where each $(\boldsymbol{X}_t, \mathcal{R}_t) \in [q]^V \times 2^V$. The initial $(\boldsymbol{X}_0, \mathcal{R}_0)$ is constructed by the first two lines of Algorithm 2. In t-th **while** loop, Algorithm 2 updates $(\boldsymbol{X}_{t-1}, \mathcal{R}_{t-1})$ to $(\boldsymbol{X}_t, \mathcal{R}_t)$. For any $t \geq 0$, we use a random variable $Y_t \triangleq |\mathcal{R}_t|$ to denote the size of $\mathcal{R}_t$. The stopping time T is the smallest integer such that $Y_t = 0$.

Define the *execution log* of Algorithm 2 up to time t as

$$\mathcal{L}_t \triangleq (X_0(\mathcal{R}_0), \mathcal{R}_0), (X_1(\mathcal{R}_1), \mathcal{R}_1), \ldots, (X_t(\mathcal{R}_t), \mathcal{R}_t).$$

Here $X_i(\mathcal{R}_i)$ is the restriction of X_i on $\mathcal{R}_i$. Note that as $R_0 = V$, $(X_0(\mathcal{R}_0), \mathcal{R}_0)$ is the same as $(X_0, \mathcal{R}_0)$. The algorithm terminates at time T if and only if $\mathcal{R}_T = \emptyset$. In the t-th iteration of the **while** loop, we use $\mathcal{F}_t$ to denote the Bayes filter and $\mathbf{u}_t$ to denote the random vertex picked in Line 4. We have the following claim.

Claim 5.6. Given any execution log $\mathcal{L}_{t-1}$ created by Algorithm 2 such that $\mathcal{R}_{t-1} \neq \emptyset$, for any $u \in \mathcal{R}_{t-1}$,

$$\Pr[\mathcal{F}_t \text{ succeeds} | \mathcal{L}_{t-1}, \mathbf{u}_t = u] \geq \begin{cases} 1 & \text{if } |S_t| = \emptyset; \\ 1 - \frac{2}{5|S_t|} & \text{if } |S_t| \neq \emptyset, \end{cases}$$

where $S_t = \partial B_t \setminus \mathcal{R}_{t-1}$ and $B_t = (B_\ell(u) \setminus \mathcal{R}_{t-1}) \cup \{u\}$ is the set B in the t-th iteration the **while** loop.

Note that given $\mathcal{L}_{t-1}$, the vertex $\mathbf{u}_t \in \mathcal{R}_{t-1}$ is sampled uniformly and independently. For any $1 \leq t \leq T$ and any execution log $\mathcal{L}_{t-1}$ created by Algorithm 2, if $S_t = \emptyset$, by Claim 5.6, we have

$$\mathbb{E}\left[Y_t | \mathcal{L}_{t-1}, S_t = \emptyset\right] = Y_{t-1} - 1;$$

Suppose $S_t \neq \emptyset$. If $\mathcal{F}_t$ fails, then $\mathcal{R}_t = \mathcal{R}_{t-1} \cup \partial B_t = \mathcal{R}_{t-1} \cup (\partial B_t \setminus \mathcal{R}_{t-1})$. In other words, $|S_t|$ new vertices will be added into $\mathcal{R}_{t-1}$ if $\mathcal{F}_t$ fails. We have

$$\begin{aligned}\mathbb{E}\left[Y_t|\mathcal{L}_{t-1}, S_t \neq \emptyset\right] &\leq Y_{t-1} - \Pr[\mathcal{F}_t \text{ succeeds}|\mathcal{L}_{t-1}, S_t \neq \emptyset] + \Pr[\mathcal{F}_t \text{ fails}|\mathcal{L}_{t-1}, S_t \neq \emptyset] \cdot |S_t| \\ &\leq Y_{t-1} + \frac{2}{5|S_t|} - \frac{3}{5} \qquad \text{(by Claim 5.6)} \\ &\leq Y_{t-1} - \frac{1}{5}.\end{aligned}$$

Combining two cases together implies

$$\mathbb{E}[Y_t|\mathcal{L}_{t-1}] = \mathbb{E}[Y_t|(X_0(\mathcal{R}_0), \mathcal{R}_0), (X_1(\mathcal{R}_1), \mathcal{R}_1), \ldots, (X_{t-1}(\mathcal{R}_{t-1}), \mathcal{R}_{t-1})] \leq Y_{t-1} - \frac{1}{5}.$$

We now define a sequence $Y'_0, Y'_1, \ldots, Y'_T$ where each $Y'_t = Y_t + \frac{t}{5}$. Thus, $Y'_0, Y'_1, \ldots, Y'_T$ is a super-martingale with respect to $(X_0(\mathcal{R}_0), \mathcal{R}_0), (X_1(\mathcal{R}_1), \mathcal{R}_1), \ldots, (X_T(\mathcal{R}_T), \mathcal{R}_T)$ and T is a stopping time. Note that each $|Y'_t - Y'_{t-1}| \leq n+1$ is bounded and $\mathbb{E}[T]$ is finite due to (10). Due to the optional stopping theorem [10, chapter 5], we have $\mathbb{E}\left[Y'_T\right] \leq \mathbb{E}\left[Y'_0\right] = \mathbb{E}[Y_0]$. Hence

$$\mathbb{E}[T] \leq 5\mathbb{E}[Y_0] = 5n,$$

where the last eqaution is because $\mathbb{E}[Y_0] = \mathbb{E}[|\mathcal{R}_0|] = n$. ∎

Proof of Claim 5.6. Suppose $S_t = \partial B_t \setminus \mathcal{R}_{t-1} = \emptyset$. This implies $\partial B_t \subseteq \mathcal{R}_{t-1}$. By the definition of $\mu_{\min}$, we have $\mu_u^{X_{t-1}(\partial B_t)}(X_{t-1}(u)) = \mu_{\min}$ and $\Pr[\mathcal{F}_t \text{ succeeds}] = 1$. In the following proof, we assume $S_t \neq \emptyset$.

We need the following property to prove the claim. Fix an execution log L_t up to time $t \geq 0$:

$$L_t = (\rho_0, R_0), (\rho_1, R_1), \ldots, (\rho_t, R_t),$$

where each $R_i \subseteq V$, each $\rho_i \in [q]^{R_i}$. Assume $R_t \neq \emptyset$ and L_t is a feasible execution log, i.e. $\Pr[\mathcal{L}_t = L_t] > 0$. We claim that given the log L_t, the random $\boldsymbol{X}_t \in [q]^V$ satisfies $X_t(R_t) = \rho_t$ and

$$\forall \tau \in [q]^{V \setminus R_t}, \quad \Pr[X_t(V \setminus R_t) = \tau | \mathcal{L}_t = L_t] = \mu^{\rho_t}(\tau). \tag{29}$$

equation (29) is proved by the induction on t. If $t = 0$, we have $R_0 = V$, equation (29) holds trivially. Assume equation (29) holds for all $t < k$. Fix any feasible execution log $L_k = (\rho_0, R_0), (\rho_1, R_1), \ldots, (\rho_k, R_k)$ such that $R_k \neq \emptyset$. Since L_k is feasible, we have $R_{k-1} \neq \emptyset$. Consider the k-th iteration of the **while** loop. The k-th iteration exists because $R_{k-1} \neq \emptyset$. By induction hypothesis, conditioning on the execution log $\mathcal{L}_{k-1} = (\rho_0, R_0), (\rho_1, R_1), \ldots, (\rho_{k-1}, R_{k-1})$, the random pair $(\boldsymbol{X}_{k-1}, \mathcal{R}_{k-1})$ satisfies the Condition 4.5 and $\boldsymbol{X}_{k-1}$ is a feasible configuration (since L_k is a feasible execution log, thus ρ_{k-1} is feasible). By Lemma 4.6, conditioning on the execution log $\mathcal{L}_{k-1} = (\rho_0, R_0), (\rho_1, R_1), \ldots, (\rho_{k-1}, R_{k-1})$, the random pair $(\boldsymbol{X}_k, \mathcal{R}_k)$ satisfies the Condition 4.5. By Condition 4.5, assuming the further condition $\mathcal{R}_k = R_k$ and $X_k(R_k) = \rho_k$, equation (29) holds for $t = k$. This proves equation (29).

Consider a feasible execution log up to time $t - 1 \geq 0$:

$$L_{t-1} = (\rho_0, R_0), (\rho_1, R_1), \ldots, (\rho_{t-1}, R_{t-1})$$

satisfying $R_{t-1} \neq \emptyset$, where each $R_i \subseteq V$ and each $\rho_i \in [q]^{R_i}$. Given the execution log $\mathcal{L}_{t-1} = L_{t-1}$, we fix a vertex $u \in R_{t-1}$ and assume $\mathbf{u}_t = u$. We analyze the t-th iteration of the **while** loop. To simplify the notation, we drop the index and denote

$$\boldsymbol{X} = \boldsymbol{X}_{t-1}, \quad R = R_{t-1}, \quad \rho = \rho_{t-1}, \quad B = B_t = (B_\ell(u) \setminus R) \cup \{u\}, \quad S = S_t = \partial B \setminus R.$$

Note that the vertex $\mathbf{u}_t$ is sampled from R uniformly and independently. By (29), given $\mathcal{L}_{t-1} = L_{t-1}$ and $\mathbf{u}_t = u$, it holds that $X(R) = \rho$ and $(\boldsymbol{X}, R)$ satisfies Condition 4.5. By Property 4.2, we know $\mu_{\min}(R, u, \rho) > 0$, thus the lower bound in (12) holds. According to the proof of Lemma 4.6, combining (19) and (25), we have

$$\Pr[\mathcal{F}_t \text{ succeeds} | \mathcal{L}_{t-1} = L_{t-1}, \mathbf{u}_t = u] = \frac{1}{\mu_u^{\rho(R_u)}(\rho_u)} \cdot \min_{\eta \in [q]^S} \mu_u^{\rho(\Psi) \uplus \eta}(\rho_u),$$

where $R_u = R \setminus \{u\}$, $S = \partial B \setminus R$ and $\Psi = \partial B \cap R$. Note that $\partial B = S \uplus \Psi$ and $u \in B$, the set ∂B separates u from $V \setminus (B \cup \partial B)$. Since $\Psi \subseteq R_u$ and two sets R_u and B are disjoint, by the conditional independence property (Property 4.3), we have $\mu_u^{\rho(\Psi) \uplus \eta}(\rho_u) = \mu_u^{\rho(R_u) \uplus \eta}(\rho_u)$. This implies

$$\Pr[\mathcal{F}_t \text{ succeeds} | \mathcal{L}_{t-1} = L_{t-1}, \mathbf{u}_t = u] = \frac{1}{\mu_u^{\rho(R_u)}(\rho_u)} \cdot \min_{\eta \in [q]^S} \mu_u^{\rho(R_u) \uplus \eta}(\rho_u).$$

By Fact 4.4, we have $S = \partial B \setminus R \subseteq S_{\ell+1}(u)$. We take $A = R_u, B = S, v = u, \sigma = \rho(R_u)$ and $a = \rho_u$ in Condition 5.3, since $\mathrm{dist}_G(u, S) = \ell + 1 = \ell^*$ and $|S| \leq |S_{\ell^*}(u)|$, this proves that

$$\Pr[\mathcal{F}_t \text{ succeeds} | \mathcal{L}_{t-1} = L_{t-1}, \mathbf{u}_t = u] \geq 1 - \frac{1}{5|S_{\ell^*}(u)|} \geq 1 - \frac{1}{5|S|} \geq 1 - \frac{2}{5|S|}.$$

■

Remark 5.7. Suppose the input instance from the class $\mathfrak{F}$ satisfies Condition 5.1 with some $\ell^* \geq 2$ and take $\ell = \ell^* - 1$ in Algorithm 2. We could tweak Algorithm 2 to reduce its running time to $O\left(n \cdot q^{\Delta^{\ell^*}}\right)$. Let $\Psi \triangleq \partial B \cap \mathcal{R}$ and $S \triangleq \partial B \setminus \mathcal{R}$. Note that $S \subseteq S_{\ell^*}(u)$ by Fact 4.4. The idea is that instead of calculating $\mu_{\min}$, we may simply compute $\mu_u^{X(\Psi) \uplus \sigma}(X_u)$ where $\sigma = \mathbf{1} \in [q]^S$ is a one-vector, then use (26) to get a lower bound μ_{low} of $\mu_{\min}$ as

$$\mu_{\text{low}} \triangleq \left(1 - \frac{1}{5\,|S_{\ell^*}(u)|}\right) \mu_u^{X(\Psi) \uplus \sigma}(X_u).$$

By Condition 5.1, $\mu_{\text{low}} \leq \mu_u^{X(\partial B)}(X_u)$. Then we use μ_{low} instead of $\mu_{\min}$ in the definition of $\mathcal{F}$. It is straightforward to check that Algorithm 2 is still correct with this tweak, and for each iteration of the **while** loop, given any $\boldsymbol{X}, \mathcal{R}$, it holds that

$$\begin{aligned}
\Pr[\mathcal{F} \text{ succeeds} | \boldsymbol{X}, \mathcal{R}] &= \frac{\mu_{\text{low}}}{\mu_u^{X(\partial B)}(X_u)} = \left(1 - \frac{1}{5|S_{\ell^*}(v)|}\right) \frac{\mu_u^{X(\Psi) \uplus \sigma}(X_u)}{\mu_u^{X(\partial B)}(X_u)} \\
&= \left(1 - \frac{1}{5|S_{\ell^*}(v)|}\right) \frac{\mu_u^{X(\Psi) \uplus \sigma}(X_u)}{\mu_u^{X(\Psi) \uplus X(S)}(X_u)} \\
&\geq \left(1 - \frac{1}{5|S_{\ell^*}(v)|}\right)^2 \qquad (\text{by } S \subseteq S_{\ell^*}(u) \text{ and Condition 5.1}) \\
&\geq 1 - \frac{2}{5|S|}.
\end{aligned}$$

This proves Claim 5.6. Besides, we do not need to enumerate all configurations in $[q]^S$ to compute $\mu_{\min}$, the expected running time of Algorithm 2 can be reduced to $O\left(n \cdot q^{\Delta^{\ell^*}}\right)$.

6 | ANALYSIS OF STRONG SPATIAL MIXING

In this section, we use Theorem 5.2 to prove other results mentioned in Section 2. We analyze the strong spatial mixing properties for the classes of spin systems mentioned in Section 2, so that we can use Theorem 5.2 to prove the existences of perfect samplers.

6.1 | Spin systems on subexponential neighborhood growth graphs

In this section, we prove Theorem 2.4 using Theorem 5.2. We need the following proposition, which explains the relation between the multiplicative form of decay in (26) and the additive form of decay in (6). Similar results appeared in [1, 2, 17, 48].

Proposition 6.1. *Let $\delta : \mathbb{N} \to \mathbb{N}$ be a nonincreasing function. Let $\mathfrak{F}$ be a class of permissive spin system instances exhibiting strong spatial mixing with decay rate δ. For every instance $\mathcal{I} = (G, [q], \boldsymbol{b}, \boldsymbol{A}) \in \mathfrak{F}$, where $G = (V, E)$, for every $v \in V$, $\Lambda \subseteq V$, and any two partial configurations $\sigma, \tau \in [q]^\Lambda$ with $\ell \geq 2$,*

$$\forall a \in [q], \quad \min\left(\left|\frac{\mu_v^\sigma(a)}{\mu_v^\tau(a)} - 1\right|, 1\right) \leq 10q \cdot |S_{\lfloor \ell/2 \rfloor}(v)| \cdot \delta(\lfloor \ell/2 \rfloor) \quad \text{(with the convention } 0/0 = 1\text{)}, \tag{30}$$

where $\ell \triangleq \min\{\mathrm{dist}_G(v, u) | u \in \Lambda,\ \sigma(u) \neq \tau(u)\} \geq 2$ and recall that $S_{\lfloor \ell/2 \rfloor}(v) = \{u \in V | \mathrm{dist}_G(v, u) = \lfloor \ell/2 \rfloor\}$ denotes the sphere of radius $\lfloor \ell/2 \rfloor$ centered at v in G.

The choice of the parameter $\lfloor \ell/2 \rfloor$ in above proposition is not essential. In the proof, we fix the configuration on the sphere of radius $\lfloor \ell/2 \rfloor$ centered at v, then utilize the conditional independence property to prove the proposition. One can modify our proof by replacing $\lfloor \ell/2 \rfloor$ with $\lfloor \theta \ell \rfloor$ for any constant $\theta \in (0, 1)$, which gives a similar result characterized by parameter θ. We first use Proposition 6.1 to prove Theorem 2.4, and then show Proposition 6.1.

Proof of Theorem 2.4. Let q be a finite integer. Suppose $\mathfrak{F}$ is a class of q-state spin systems that is defined on a class of graphs that have subexponential neighborhood growth in Definition 2.3 with function $s : \mathbb{N} \to \mathbb{N}$. Suppose $\mathfrak{F}$ exhibits strong spatial mixing with exponential decay with some constants $\alpha > 0, \beta > 0$. Let δ be the function $\delta(x) = \alpha \exp(-\beta x)$.

Fix a instance $\mathcal{I} = (G, [q], \boldsymbol{b}, \boldsymbol{A}) \in \mathfrak{F}$, where $G = (V, E)$. By Proposition 6.1, we have for any $\Lambda \subseteq V$, any $v \in V$, and any two partial configurations $\sigma, \tau \in [q]^\Lambda$ satisfying $\ell \triangleq \min\{\mathrm{dist}_G(v, u) | u \in \Lambda,\ \sigma(u) \neq \tau(u)\} \geq 2$,

$$\begin{aligned} \forall a \in [q], \quad \min\left(\left|\frac{\mu_v^\sigma(a)}{\mu_v^\tau(a)} - 1\right|, 1\right) &\leq 10q \cdot |S_{\lfloor \ell/2 \rfloor}(v)| \cdot \delta(\lfloor \ell/2 \rfloor) \\ \text{(by } |S_r(v)| \leq |B_r(v)| \leq s(r)) &\leq 10q \cdot s(\lfloor \ell/2 \rfloor) \cdot \delta(\lfloor \ell/2 \rfloor). \end{aligned} \tag{31}$$

Recall that $s(\cdot)$ is the subexponential function in Definition 2.3. Note that $\delta(\ell) = \alpha \exp(-\beta\ell)$. We take $\ell_0 = \ell_0(q, \alpha, \beta, s)$ sufficiently large such that $\ell_0 \geq 2$ and

$$10\alpha q \cdot s(\lfloor \ell_0/2 \rfloor) \exp(-\beta \lfloor \ell_0/2 \rfloor) \leq \frac{1}{5s(\ell_0)} \leq \frac{1}{5\left|B_{\ell_0}(v)\right|} \leq \frac{1}{5\left|S_{\ell_0}(v)\right|}. \tag{32}$$

Note that (32) is equivalent to

$$\delta(\lfloor \ell_0/2 \rfloor) = \alpha \exp(-\beta \lfloor \ell_0/2 \rfloor) \leq \frac{1}{50q \cdot s(\lfloor \ell_0/2 \rfloor) \cdot s(\ell_0)}. \tag{33}$$

A sufficiently large $\ell_0 = \ell_0(q, \alpha, \beta, s)$ to satisfy the above must exist because $s(r) = \exp(o(r))$. Combining (31) and (32) implies that $\mathcal{I}$ satisfies Condition 5.1 with $\ell_0 \geq 2$. By Theorem 5.2, if the parameter ℓ in Algorithm 2 is set so that $\ell = \ell_0 - 1$, given $\mathcal{I}$, the expected running time of Algorithm 2 is $n \cdot q^{O(\Delta^{\ell_0})}$. Since $\ell_0 = O(1)$, $q = O(1)$ and $\Delta \leq s(1) = O(1)$, the expected running time of Algorithm 2 is $O(n)$. ∎

We now prove Proposition 6.1. Similar results are proved in [1, 2, 17, 48].

Proof of Proposition 6.1. Fix a instance $\mathcal{I} = (G, [q], \boldsymbol{b}, \boldsymbol{A}) \in \mathfrak{F}$, where $G = (V, E)$. Fix two partial configurations $\sigma, \tau \in [q]^{\Lambda}$ with $\ell \triangleq \min\{\mathrm{dist}_G(v, u) | u \in \Lambda,\ \sigma(u) \neq \tau(u)\}$. We use $D \triangleq \{v \in \Lambda | \sigma(v) \neq \tau(v)\}$ to denote the set at which σ and τ disagree. Fix a spin $a \in [q]$. Since $\mathcal{I}$ is permissive (Definition 2.1), we have

$$\begin{aligned}
\mu_v^{\sigma}(a) = 0 \quad &\Longleftrightarrow \quad b_v(a) \prod_{u \in \Gamma_G(v) \cap \Lambda} A_{uv}(a, \sigma(v)) = 0, \\
\mu_v^{\tau}(a) = 0 \quad &\Longleftrightarrow \quad b_v(a) \prod_{u \in \Gamma_G(v) \cap \Lambda} A_{uv}(a, \tau(v)) = 0,
\end{aligned}$$

where $\Gamma_G(v)$ is the neighborhood of v in G. Since $\ell \geq 2$, there is no edge between v and D, we have $\Gamma_G(v) \cap D = \emptyset$. This implies $\mu_v^{\sigma}(a) = 0$ if and only if $\mu_v^{\tau}(a) = 0$. If $\mu_v^{\sigma}(a) = \mu_v^{\tau}(a) = 0$, the proposition holds trivially. We assume

$$\mu_v^{\sigma}(a) > 0 \wedge \mu_v^{\tau}(a) > 0. \tag{34}$$

Define the set of vertices $H \triangleq S_{\lfloor \ell/2 \rfloor}(v) \setminus \Lambda$, where $S_{\lfloor \ell/2 \rfloor}(v) = \{u \in V | \mathrm{dist}(u, v) = \lfloor \ell/2 \rfloor\}$ is the sphere of radius $\lfloor \ell/2 \rfloor$ centered at v in graph G. By the definitions, we have $H \cap D = \emptyset$. If $H = \emptyset$, then $S_{\lfloor \ell/2 \rfloor}(v) \subseteq \Lambda$. This implies that all vertices in $S_{\lfloor \ell/2 \rfloor}(v)$ are fixed to the same values in σ and τ. Note that $S_{\lfloor \ell/2 \rfloor}(v)$ separates v from D in G. By the conditional independence property, $\mu_v^{\sigma}(a) = \mu_v^{\tau}(a)$, the proposition holds. In the rest of the proof, we assume $H \neq \emptyset$.

For any $\rho \in [q]^H$ satisfying $\mu_H^{\sigma \uplus \{v \leftarrow a\}}(\rho) > 0$ and $\mu_H^{\tau \uplus \{v \leftarrow a\}}(\rho) > 0$ we have

$$\mu_v^{\sigma}(a) = \frac{\mu_H^{\sigma}(\rho) \cdot \mu_v^{\sigma \uplus \rho}(a)}{\mu_H^{\sigma \uplus \{v \leftarrow a\}}(\rho)}, \quad \mu_v^{\tau}(a) = \frac{\mu_H^{\tau}(\rho) \cdot \mu_v^{\tau \uplus \rho}(a)}{\mu_H^{\tau \uplus \{v \leftarrow a\}}(\rho)}.$$

The first equation holds since $\mu_H^{\sigma}(\rho) \cdot \mu_v^{\sigma \uplus \rho}(a) = \mu_v^{\sigma}(a) \cdot \mu_H^{\sigma \uplus \{v \leftarrow a\}}(\rho)$ and $\mu_H^{\sigma \uplus \{v \leftarrow a\}}(\rho) > 0$; the second equation holds similarly. Note that $\mu_v^{\sigma}(a) > 0$ and $\mu_v^{\tau}(a) > 0$. We have

$$\frac{\mu_v^{\sigma}(a)}{\mu_v^{\tau}(a)} = \left(\frac{\mu_v^{\sigma \uplus \rho}(a)}{\mu_v^{\tau \uplus \rho}(a)}\right) \left(\frac{\mu_H^{\sigma}(\rho)}{\mu_H^{\tau}(\rho)}\right) \left(\frac{\mu_H^{\tau \uplus \{v \leftarrow a\}}(\rho)}{\mu_H^{\sigma \uplus \{v \leftarrow a\}}(\rho)}\right).$$

Note that $(\Lambda \setminus D) \cup H$ separates v from D in graph G, and the two configurations $\sigma \uplus \rho$ and $\tau \uplus \rho$ disagree only at D. By the conditional independence property, we have $\mu_v^{\sigma \uplus \rho}(a) = \mu_v^{\tau \uplus \rho}(a)$. Hence, we have

$$\frac{\mu_v^{\sigma}(a)}{\mu_v^{\tau}(a)} = \left(\frac{\mu_H^{\sigma}(\rho)}{\mu_H^{\tau}(\rho)}\right)\left(\frac{\mu_H^{\tau \uplus \{v \leftarrow a\}}(\rho)}{\mu_H^{\sigma \uplus \{v \leftarrow a\}}(\rho)}\right). \tag{35}$$

Note that (35) holds for any $\rho \in [q]^H$ satisfying $\mu_H^{\sigma \uplus \{v \leftarrow a\}}(\rho) > 0$ and $\mu_H^{\tau \uplus \{v \leftarrow a\}}(\rho) > 0$. Our goal is to choose a suitable ρ and bound the RHS. Let

$$\epsilon \triangleq \delta(\lfloor \ell/2 \rfloor).$$

Without loss of generality, we assume

$$10q \cdot |S_{\lfloor \ell/2 \rfloor}(v)| \cdot \delta(\lfloor \ell/2 \rfloor) = 10q\epsilon \cdot |S_{\lfloor \ell/2 \rfloor}(v)| < 1. \tag{36}$$

If (36) does not hold, then the inequality (30) holds trivially. We have the following claim.

Claim 6.2. Assume (36). There exists a configuration $\rho \in [q]^H$ satisfying $\mu_H^{\sigma \uplus \{v \leftarrow a\}}(\rho) > 0$ and $\mu_H^{\tau \uplus \{v \leftarrow a\}}(\rho) > 0$ such that

$$\left(1 - \frac{2q\epsilon}{1+q\epsilon}\right)^{2m} \le \left(\frac{\mu_H^{\sigma}(\rho)}{\mu_H^{\tau}(\rho)}\right)\left(\frac{\mu_H^{\tau \uplus \{v \leftarrow a\}}(\rho)}{\mu_H^{\sigma \uplus \{v \leftarrow a\}}(\rho)}\right) \le \left(1 + \frac{2q\epsilon}{1-q\epsilon}\right)^{2m},$$

where $m \triangleq |S_{\lfloor \ell/2 \rfloor}(v)|$ and $\epsilon \triangleq \delta(\lfloor \ell/2 \rfloor)$.

The inequality (36) implies that

$$q\epsilon m \le \frac{1}{10}. \tag{37}$$

Combining Claim 6.2 with the above, we have

$$\begin{aligned}\left(\frac{\mu_H^{\sigma}(\rho)}{\mu_H^{\tau}(\rho)}\right)\left(\frac{\mu_H^{\tau \uplus \{v \leftarrow a\}}(\rho)}{\mu_H^{\sigma \uplus \{v \leftarrow a\}}(\rho)}\right) &\le \exp\left(\frac{4q\epsilon m}{1-q\epsilon}\right) \\ &\le \exp(5q\epsilon m) && \text{(by(37))} \\ &\le 1 + 10q\epsilon m. && \text{(by(37))}\end{aligned}$$

Similarly, we have

$$\begin{aligned}\left(\frac{\mu_H^{\sigma}(\rho)}{\mu_H^{\tau}(\rho)}\right)\left(\frac{\mu_H^{\tau \uplus \{v \leftarrow a\}}(\rho)}{\mu_H^{\sigma \uplus \{v \leftarrow a\}}(\rho)}\right) &\ge (1 - 2q\epsilon)^{2m} \\ &\ge \exp(-8q\epsilon m) && \text{(by(37))} \\ &\ge 1 - 10q\epsilon m.\end{aligned}$$

Recall $m \triangleq |S_{\lfloor \ell/2 \rfloor}(v)|$ and $\epsilon \triangleq \delta(\lfloor \ell/2 \rfloor)$. This proves the proposition. ∎

Proof of Claim 6.2. Recall $H \triangleq S_{\lfloor \ell/2 \rfloor}(v) \setminus \Lambda$. Suppose $|H| = h \geq 1$. Let $H = \{v_1, v_2, \ldots, v_h\}$. Define a sequence of subsets $H_0, H_1, \ldots, H_h$ as $H_i \triangleq \{v_j | 1 \leq j \leq i\}$. Note that $H_0 = \emptyset$ and $H_h = H$. We now construct the configuration $\rho \in [q]^H$ by the following h steps.

- initially, $\rho = \emptyset$ is an empty configuration;
- in ith step, given $\rho \in [q]^{H_{i-1}}$, choose $c_i \in [q]$ that maximizes $\mu_{v_i}^{\sigma \uplus \rho \uplus \{v \leftarrow a\}}(c_i)$ (break tie arbitrarily), extend ρ further at position v_i and set $\rho(v_i) = c_i$, thus $\rho \in [q]^{H_i}$ after the ith step.

By the construction, we have

$$\forall 1 \leq i \leq h, \quad \mu_{v_i}^{\sigma \uplus \rho(H_{i-1}) \uplus \{v \leftarrow a\}}(\rho(v_i)) \geq \frac{1}{q} > 0.$$

We have $\mathrm{dist}_G(H, D) \geq \ell - \lfloor \ell/2 \rfloor \geq \lfloor \ell/2 \rfloor$, where D is the set at which σ and τ disagree, and $\mathrm{dist}_G(H, D) \triangleq \min\{\mathrm{dist}_G(u_1, u_2) | u_1 \in H \wedge u_2 \in D\}$. Recall $\epsilon \triangleq \delta(\lfloor \ell/2 \rfloor)$ and δ is a nonincreasing function. By the strong spatial mixing property in Definition 2.2, we have

$$\forall 1 \leq i \leq h, \quad \mu_{v_i}^{\tau \uplus \rho(H_{i-1}) \uplus \{v \leftarrow a\}}(\rho(v_i)) \geq \frac{1}{q} - \epsilon > 0,$$

where $\frac{1}{q} - \epsilon > 0$ holds due to (36). By the chain rule, we have $\mu_H^{\sigma \uplus \{v \leftarrow a\}}(\rho) > 0$ and $\mu_H^{\tau \uplus \{v \leftarrow a\}}(\rho) > 0$.

We now prove that ρ satisfies the inequalities in Claim 6.2. For any $1 \leq i \leq h$, define

$$p_i \triangleq \mu_{v_i}^{\sigma \uplus \rho(H_{i-1}) \uplus \{v \leftarrow a\}}(\rho(v_i)). \tag{38}$$

Recall that $\mathrm{dist}_G(H, v) \geq \lfloor \ell/2 \rfloor$ and $\mathrm{dist}_G(H, D) \geq \lfloor \ell/2 \rfloor$. Recall $\epsilon \triangleq \delta(\lfloor \ell/2 \rfloor)$ and δ is a nonincreasing function. By the strong spatial mixing property in Definition 2.2, we have for any $c \in [q]$ and any $1 \leq i \leq h$,

$$0 < p_i - \epsilon \leq \mu_{v_i}^{\sigma \uplus \rho(H_{i-1}) \uplus \{v \leftarrow c\}}(\rho(v_i)) \leq p_i + \epsilon, \tag{39}$$

$$0 < p_i - \epsilon \leq \mu_{v_i}^{\tau \uplus \rho(H_{i-1}) \uplus \{v \leftarrow c\}}(\rho(v_i)) \leq p_i + \epsilon. \tag{40}$$

Note that $p_i - \epsilon \geq \frac{1}{q} - \epsilon > 0$ due to (36). Combining (39), (40) and the chain rule implies

$$\forall c, c' \in [q], \quad \prod_{i=1}^{h} \left(\frac{p_i - \epsilon}{p_i + \epsilon} \right) \leq \frac{\mu_H^{\tau \uplus \{v \leftarrow c\}}(\rho)}{\mu_H^{\sigma \uplus \{v \leftarrow c'\}}(\rho)} \leq \prod_{i=1}^{h} \left(\frac{p_i + \epsilon}{p_i - \epsilon} \right).$$

Note that $p_i \geq \frac{1}{q}$ for all $1 \leq i \leq h$ due the construction of ρ, and $q\epsilon < 1$ due to (36). We have

$$\forall c, c' \in [q], \quad \left(1 - \frac{2q\epsilon}{1 + q\epsilon} \right)^h \leq \frac{\mu_H^{\tau \uplus \{v \leftarrow c\}}(\rho)}{\mu_H^{\sigma \uplus \{v \leftarrow c'\}}(\rho)} \leq \left(1 + \frac{2q\epsilon}{1 - q\epsilon} \right)^h, \tag{41}$$

and

$$\forall c, c' \in [q], \quad \left(1 - \frac{2q\epsilon}{1 + q\epsilon} \right)^h \leq \frac{\mu_H^{\sigma \uplus \{v \leftarrow c\}}(\rho)}{\mu_H^{\tau \uplus \{v \leftarrow c'\}}(\rho)} \leq \left(1 + \frac{2q\epsilon}{1 - q\epsilon} \right)^h. \tag{42}$$

Note that

$$\mu_H^\sigma(\rho) = \sum_{c\in[q]} \mu_v^\sigma(c)\mu_H^{\sigma\uplus\{v\leftarrow c\}}(\rho), \quad \mu_H^\tau(\rho) = \sum_{c\in[q]} \mu_v^\tau(c)\mu_H^{\tau\uplus\{v\leftarrow c\}}(\rho).$$

$\mu_H^\sigma(\rho)$ is a convex combination of $\mu_H^{\sigma\uplus\{v\leftarrow c\}}(\rho)$, and $\mu_H^\tau(\rho)$ is a convex combination of $\mu_H^{\tau\uplus\{v\leftarrow c\}}(\rho)$. By (41) and (42), it holds that

$$\left(1-\frac{2q\epsilon}{1+q\epsilon}\right)^{2h} \le \left(\frac{\mu_H^\sigma(\rho)}{\mu_H^\tau(\rho)}\right)\left(\frac{\mu_H^{\tau\uplus\{v\leftarrow a\}}(\rho)}{\mu_H^{\sigma\uplus\{v\leftarrow a\}}(\rho)}\right) \le \left(1+\frac{2q\epsilon}{1-q\epsilon}\right)^{2h}.$$

Note that $h = |H|$ and $H \subseteq S_{\lfloor \ell/2\rfloor}(v)$, then $m = |S_{\lfloor \ell/2\rfloor}(v)| \ge h$. This proves the claim. ∎

6.2 | Spin systems on general graphs

In this section, we prove Theorem 2.6 by showing that Condition 2.5 implies Condition 5.1.

Proof of Theorem 2.6. Fix a instance $\mathcal{I} = (G, [q], \boldsymbol{A}, \boldsymbol{b}) \in \mathfrak{F}$ satisfying Condition 2.5 with parameter $\ell = \ell(q) \ge 2$. Fix subset $\Lambda \subseteq V$ and vertex $v \in V \setminus \Lambda$. For any two partial configurations $\sigma, \tau \in [q]^\Lambda$ satisfying $\min\{\mathrm{dist}_G(v,u) | u \in \Lambda,\ \sigma(u) \ne \tau(u)\} = \ell \ge 2$, we claim

$$\forall a \in [q], \quad \mu_v^\sigma(a) = 0 \iff \mu_v^\tau(a) = 0. \tag{43}$$

Let $D \triangleq \{u \in \Lambda | \sigma(u) \ne \tau(u)\}$, $H \triangleq \Lambda \setminus D$ and $\rho \triangleq \sigma_H = \tau_H$. Since $\ell \ge 2$, $\Gamma_G(v) \cap \Lambda = \Gamma_G(v) \cap H$, where $\Gamma_G(v)$ is the neighborhood of v in G. Since $\mathcal{I}$ is a permissive spin system (Definition 2.1), $\mu_v^\sigma(a) = 0$ if and only if $b_v(a)\prod_{u\in\Gamma_G(v)\cap H} A_{uv}(a,\rho_u) = 0$; similarly, $\mu_v^\tau(a) = 0$ if and only if $b_v(a)\prod_{u\in\Gamma_G(v)\cap H} A_{uv}(a,\rho_u) = 0$. This proves (43).

If $\mu_v^\sigma(a) = \mu_v^\tau(a) = 0$, then (26) holds trivially. Otherwise, by Condition 2.5, $\mu_v^\sigma(a) \ge \gamma$ and $\mu_v^\tau(a) \ge \gamma$, where $\gamma = \gamma(\Lambda, v) > 0$ is positive and depends only on Λ and v. By (8) and (7), we have

$$\left|\frac{\mu_v^\sigma(a)}{\mu_v^\tau(a)} - 1\right| \le \frac{\gamma + d_{\mathrm{TV}}(\mu_v^\sigma, \mu_v^\tau)}{\gamma} - 1 \le \frac{1}{5|S_\ell(v)|}.$$

This implies that any instance $\mathcal{I} = (G, [q], \boldsymbol{A}, \boldsymbol{b}) \in \mathfrak{F}$ satisfies Condition 5.1 with parameter $\ell = \ell(q) \ge 2$. Theorem 2.6 is a corollary of Theorem 5.2. ∎

6.3 | Uniform list coloring

We now prove Theorem 2.9. Let $\mathfrak{L}$ be a class of list coloring instances with at most q colors for a finite $q > 0$. Let $\alpha^* \approx 1.763\ \ldots$ be the positive root of the equation $x^x = \mathrm{e}$. Suppose there exist $\alpha > \alpha^*$ and $\beta \ge \frac{\sqrt{2}}{\sqrt{2}-1}$ satisfying $(1-1/\beta)\alpha \mathrm{e}^{\frac{1}{\alpha}(1-1/\beta)} > 1$ such that for all $\mathcal{I} = (G = (V,E), [q], \mathcal{L}) \in \mathfrak{L}$, the graph G is triangle-free and

$$\forall v \in V, \quad |L(v)| \ge \alpha \deg_G(v) + \beta.$$

Gamarnik, Katz, and Misra [20] proved that $\mathfrak{L}$ exhibits the strong spatial mixing with exponential decay. If $\mathfrak{L}$ is defined on subexponential neighborhood growth graphs, then by Theorem 2.4, the linear time perfect sampler exists for every instance in $\mathfrak{L}$.

There are two remaining cases in Theorem 2.9. We now assume that the class $\mathfrak{L}$ of list coloring instances satisfies one of the following two conditions.

(I) there is an $s : \mathbb{N} \to \mathbb{N}$ with $s(\ell) = \exp(o(\ell))$ such that for any $\mathcal{I} = (G = (V, E), [q], \mathcal{L}) \in \mathfrak{L}$,

$$\forall v \in V, \ell \geq 0, \quad |S_\ell(v)| \leq s(\ell),$$
$$\forall v \in V, \quad |L(v)| \geq 2 \deg_G(v);$$

(II) for any $\mathcal{I} = (G = (V, E), [q], \mathcal{L}) \in \mathfrak{L}$,

$$\forall v \in V, \quad |L(v)| \geq \Delta^2 - \Delta + 2.$$

Lemma 6.3. *Let $\mathfrak{L}$ be a class of list coloring instances with at most q colors for a finite $q > 0$. Suppose $\mathfrak{L}$ satisfies (I) or (II). There exist finite $A > 0$ and $\theta > 0$ such that for every $\mathcal{I} = (G, [q], \mathcal{L}) \in \mathfrak{L}$, where $G = (V, E)$, for any $v \in V$, any $\Lambda \subseteq V$, and any $\sigma, \tau \in [q]^\Lambda$ with $\ell \triangleq \min\{\mathrm{dist}_G(v, u) | u \in \Lambda,\ \sigma(u) \neq \tau(u)\} = \Omega(q \log q)$, it holds that*

$$\forall a \in [q] : \quad \left| \frac{\mu_v^\sigma(a)}{\mu_v^\tau(a)} - 1 \right| \leq \frac{A\mathrm{e}^{-\theta\ell}}{|S_\ell(v)|} \quad \text{(with the convention } 0/0 = 1),$$

where $A = A(q, s) > 0$ and $\theta = \frac{1}{2q} > 0$ if $\mathfrak{L}$ satisfies (I) with the function $s : \mathbb{N} \to \mathbb{N}$; or $A = \mathrm{poly}(q)$ and $\theta = \frac{1}{2q^2} > 0$ if $\mathfrak{L}$ satisfies (II).

Theorem 5.2 together with Lemma 6.3 proves the remaining two cases in Theorem 2.9. We take a sufficiently large ℓ^* such that $\ell^* = \Omega(q \log q)$ and $A\mathrm{e}^{-\theta\ell^*} \leq \frac{1}{5}$. By Lemma 6.3, instances of $\mathfrak{L}$ satisfy Condition 5.1 with this $\ell^* \geq 2$. Thus the perfect sampler exists due to Theorem 5.2. Note that ℓ^* depends only on q and the function s, and for any instance $\mathcal{I} \in \mathfrak{L}$, the maximum degree $\Delta \leq q$. Thus, the expected running time of our algorithm is $n \cdot q^{O(q^{\ell^*})} = O(n)$. Furthermore, if $\mathfrak{L}$ satisfies (II), then $\ell^* = \Theta(q^2 \log q)$, thus the expected running time is $n \cdot \exp(\exp(\mathrm{poly}(q)))$.

6.3.1 | The multiplicative SSM of list coloring (proof of Lemma 6.3)

In [20, theorem 3], Gamarnik, Katz, and Misra established the best-known strong spatial mixing result for list colorings in bounded degree graphs. This is almost what we need, except that we want to control the decay rate under conditions (I) and (II). Going through the proof of [20, theorem 3] and keeping track of the decay rate, we have the proposition below. The similar analysis technique are also used in [37].

Proposition 6.4 ([20])**.** *Let $\mathcal{I} = (G, [q], \mathcal{L})$ be a list coloring instance, where $G = (V, E)$. Assume that $\mathcal{I}$ satisfies $|L(v)| \geq \deg_G(v) + 1$ for all $v \in V$. Suppose*

$$\max_{u \in V} \frac{\deg_G(u) - 1}{|L(u)| - \deg_G(u)} \leq \chi < 1.$$

Then for any $\Lambda \subseteq V$, any vertex $v \in V \setminus \Lambda$, and any two partial colorings $\sigma, \tau \in [q]^\Lambda$ satisfying $\ell \triangleq \min\{\mathrm{dist}_G(v, u) | u \in \Lambda,\ \sigma(u) \neq \tau(u)\} = \Omega\left(\frac{\log q}{\log(1/\chi)}\right)$, it holds that

$$\forall a \in [q] : \quad \left| \frac{\mu_v^\sigma(a)}{\mu_v^\tau(a)} - 1 \right| \leq B\chi^\ell, \quad \text{(with convention } 0/0 = 1),$$

where $B = \mathrm{poly}(q/\chi)$ depends only on q and χ.

Proof of Lemma 6.3. Fix a instance $\mathcal{I} = (G, [q], \mathcal{L}) \in \mathfrak{L}$, where $G = (V, E)$. Suppose $\mathfrak{L}$ satisfies Condition in (I). We have

$$\max_{u\in V} \frac{\deg_G(u) - 1}{|L(u)| - \deg_G(u)} \le \max_{u\in V} \frac{\deg_G(u) - 1}{\deg_G(u)} = \frac{\Delta - 1}{\Delta} \le \frac{q-1}{q}.$$

The χ and B in Proposition 6.4 can be set as $\chi = \frac{q-1}{q}$ and $B = \mathrm{poly}(q/\chi) \le B_{\max} = \mathrm{poly}(q)$. Then for any subset $\Lambda \subseteq V$, any vertex $v \in V \setminus \Lambda$, any two colorings $\sigma, \tau \in [q]^\Lambda$ that disagree on $D \subseteq \Lambda$ satisfying $\ell \triangleq \min\{\mathrm{dist}_G(u, v) | u \in D\} = \Omega\Big(\frac{\log q}{\log 1/\chi}\Big) = \Omega(q \log q)$, it holds that

$$\forall a \in [q] : \quad \left|\frac{\mu_v^\sigma(a)}{\mu_v^\tau(a)} - 1\right| \le B_{\max}\chi^\ell \le B_{\max} \cdot \frac{|S_\ell(v)|}{|S_\ell(v)|} \cdot \chi^\ell.$$

Since G has subexponential growth, we have that $|S_\ell(v)| \le s(\ell) = \exp(o(\ell))$. Thus,

$$\forall a \in [q] : \quad \left|\frac{\mu_v^\sigma(a)}{\mu_v^\tau(a)} - 1\right| \le B_{\max} \cdot \frac{s(\ell)}{|S_\ell(v)|} \cdot \left(\frac{q-1}{q}\right)^\ell \le \frac{s(\ell)B_{\max}}{|S_\ell(v)|} \cdot \mathrm{e}^{-\ell/q} \le \frac{A\mathrm{e}^{-\theta\ell}}{|S_\ell(v)|},$$

for some $A = A(q, s) > 0$ and $\theta = \frac{1}{2q} > 0$.

Suppose $\mathfrak{L}$ satisfies (II). Recall that Δ is the maximum degree of graph G. we have

$$\max_{u\in V} \frac{\deg_G(u) - 1}{|L(u)| - \deg_G(u)} \le \frac{\Delta - 1}{(\Delta - 1)^2 + 1}.$$

The χ and B in Proposition 6.4 can be set as $\chi = \frac{\Delta-1}{(\Delta-1)^2+1}$ and $B = \mathrm{poly}(q/\chi)$. Thus $1/\chi \le \Delta^2 \le q^2$. We have $B = \mathrm{poly}(q/\chi) \le B_{\max} = \mathrm{poly}(q)$. For any subset $\Lambda \subseteq V$, any vertex $v \in V \setminus \Lambda$, any two colorings $\sigma, \tau \in [q]^\Lambda$ that disagree on $D \subseteq \Lambda$ satisfying $\ell \triangleq \min\{\mathrm{dist}_G(u, v) | u \in D\} = \Omega(\frac{\log q}{\log 1/\chi}) = \Omega(\log q)$, it holds that

$$\forall a \in [q] : \quad \left|\frac{\mu_v^\sigma(a)}{\mu_v^\tau(a)} - 1\right| \le B_{\max}\chi^\ell \le B_{\max} \cdot \frac{\Delta(\Delta-1)^{\ell-1}}{|S_\ell(v)|} \cdot \chi^\ell,$$

where the last inequality due to $|S_\ell(v)| \le \Delta(\Delta - 1)^{\ell-1}$. Since $\chi = \frac{\Delta-1}{(\Delta-1)^2+1}$, we have

$$\begin{aligned}
\forall a \in [q] : \quad \left|\frac{\mu_v^\sigma(a)}{\mu_v^\tau(a)} - 1\right| &\le \frac{B_{\max}\Delta}{\Delta - 1} \cdot \frac{1}{|S_\ell(v)|} \cdot \left(\frac{(\Delta-1)^2}{(\Delta-1)^2+1}\right)^\ell \\
(\text{by } \Delta \le q) &\le \frac{2B_{\max}}{|S_\ell(v)|} \cdot \left(\frac{(q-1)^2}{(q-1)^2+1}\right)^\ell \\
&\le \frac{2B_{\max}}{|S_\ell(v)|} \cdot \mathrm{e}^{-\frac{\ell}{2q^2}} = \frac{A\mathrm{e}^{-\theta\ell}}{|S_\ell(v)|},
\end{aligned}$$

where $A = 2B_{\max} = \mathrm{poly}(q)$ and $\theta = \frac{1}{2q^2} > 0$. ∎

6.4 | The monomer-dimer model

We now prove Theorem 2.10. We first present the monomer-dimer model instance as a spin system instance, then we use Theorem 2.4 to prove Theorem 2.10.

Given a graph $G = (V, E)$, we use $G^* = (V^*, E^*) = \mathrm{Lin}(G)$ to denote the *line graph* of G. Each vertex $v_e \in V^*$ in line graph G^* represents an edge $e \in E$ in the original graph G, and two vertices $v_e, v_{e'}$ in G^* are adjacent if and only if e and e' share a vertex in G. We call $S \subseteq V^*$ an *independent set* in G^* if no two vertices in S are adjacent in G^*. It is easy to verify that there is a one-to-one correspondence between the matchings in G and the independent sets in G^*.

Given a monomer-dimer model instance $\mathcal{I} = (G, \lambda)$, we define a *hardcore model* instance $\mathcal{I}^* = (G^*, \lambda)$ in the line graph $G^* = \mathrm{Lin}(G)$. Each independent set S in G^* is assigned a weight $w_{\mathcal{I}^*}(S) = \lambda^{|S|}$. Let $\mu_{\mathcal{I}^*}$ be a distribution over all independent sets in G^* such that $\mu_{\mathcal{I}^*}(S) \propto w_{\mathcal{I}^*}(S)$. Hence, $\mathcal{I}^*$ is a spin system instance and $\mathcal{I}^*$ is permissive. Besides, if we can sample independent sets from $\mu_{\mathcal{I}^*}$, then we can sample matchings from $\mu = \mu_{\mathcal{I}}$.

Suppose the class of monomer-dimer model instances $\mathfrak{M}$ satisfies the condition in Theorem 2.10. Then, there exist a constant C and a function $s : \mathbb{N} \to \mathbb{N}$ with $s(\ell) = \exp(o(\ell))$ such that for all $\mathcal{I} = (G, \lambda) \in \mathfrak{M}$, $\lambda \le C = O(1)$, $|S_\ell(v)| \le s(\ell) = \exp(o(\ell))$ for all $v \in V$ and $\ell \ge 0$, and $\Delta_G \le s(1) = O(1)$. Thus, $\mathfrak{M}$ exhibits strong spatial mixing with exponential decay with constants $\alpha = \alpha(C, s) > 0$ and $\beta = \beta(C, s) > 0$ [4, 47]. Observe that if $e_1, e_2, \ldots, e_\ell$ is a path of edges in G, then $v_{e_1}, v_{e_2}, \ldots, v_{e_\ell}$ is a path of vertices in G^*, and vice versa. Hence, the following results hold for the class of hardcore instances $\mathfrak{H} = \{\mathcal{I}^* = (G^*, \lambda) | \mathcal{I} \in \mathfrak{M}\}$.

- The class of hardcore instances $\mathfrak{H}$ exhibits strong spatial mixing with exponential decay with constants $\alpha' = \alpha'(C, s) > 0$ and $\beta = \beta(C, s) > 0$.
- for any instance $(G^*, \lambda) \in \mathfrak{H}$, the graph G^* has subexponential growth. Suppose $G^* = (V^*, E^*)$ is the line graph of $G = (V, E)$. For all $e = \{u, v\} \in E$, $\ell \ge 1$, it holds that $|S^*_\ell(v_e)| \le \Delta_G(|S_{\ell-1}(u)| + |S_{\ell-1}(v)|) \le 2s(1)s(\ell - 1) = \exp(o(\ell))$, where $v_e \in V^*$ represents the edge e, $S^*_\ell(v_e)$ is the sphere of radius ℓ centered at v_e in G^*, $S_{\ell-1}(v)$ is the sphere of radius $\ell - 1$ centered at v in G.

Note that the number of vertices in G^* is at most $n\Delta_G = O(n)$, where n is number of vertices in G. Theorem 2.10 is a corollary of Theorem 2.4.

6.5 | The hardcore model, Ising model, and anti-ferromagnetic two-spin system

Theorem 2.11, Theorem 2.12, and Theorem 2.13 follow immediately from Theorem 2.4 and strong spatial mixing results in [36, 41, 43, 52].

7 | DYNAMIC SAMPLING

In this section, we use our algorithm to solve the dynamic sampling problem [15, 16]. In this problem, the Gibbs distribution itself changes dynamically and the algorithm needs to maintain a random sample efficiently with respect to the current Gibbs distribution.

We first define the update for the spin system instance. Let $\mathcal{I} = (G, [q], \boldsymbol{b}, \boldsymbol{A})$ be a spin system instance, where $G = (V, E)$. We use $\mu_{\mathcal{I}}$ to denote the Gibbs distribution defined by $\mathcal{I}$.

- updates for vertices: modifying the vector b_v of vertex $v \in V$;
- updates for edges: modifying the matrix A_e of edge $e \in E$; or adding new edge $e \notin E$.

We use (D_V, D_E, C) to denote the update for instance $\mathcal{I}$, where $D_V \subseteq V, D_E \subseteq \{\{u,v\} | u,v \in V, u \neq v\}$, and $\mathcal{C} = (b_v)_{v\in D_V} \cup (A_e)_{e\in D_E}$. For each $v \in D_V$, we modify its vector to $b_v \in \mathcal{C}$, and for each $e \in D_E$, we either add the new edge e with matrix $A_e \in \mathcal{C}$ (if $e \notin E$), or modify its matrix to $A_e \in \mathcal{C}$ (if $e \in E$).

Definition 7.1 (dynamic sampling problem)**.** Given a spin system instance $\mathcal{I} = (G, [q], \boldsymbol{b}, \boldsymbol{A})$ where $G = (V, E)$, a random sample $\boldsymbol{X} \in [q]^V$ such that $\boldsymbol{X} \sim \mu_{\mathcal{I}}$, and an update $(D_V, D_E, \mathcal{C})$ that modifies the instance $\mathcal{I}$ to an updated instance $\mathcal{I}' = (G', [q], \boldsymbol{b}', \boldsymbol{A}')$ where $G' = (V, E')$, the algorithm updates $\boldsymbol{X}$ to a new sample $\boldsymbol{X}' \in [q]^V$ such that $\boldsymbol{X}' \sim \mu_{\mathcal{I}'}$.

Theorem 7.2. *Let $\mathfrak{F}$ be a class of permissive spin systems satisfying Condition 5.1. There exists an algorithm such that if the updated instance $\mathcal{I}' = (G', [q], \boldsymbol{b}', \boldsymbol{A}') \in \mathfrak{F}$, then the algorithm solves the dynamic sampling problem within $\Delta(|D_V| + |D_E|)q^{O(\Delta^\ell)}$ time in expectation, where Δ is the maximum degree of G' and $\ell = \ell(q) \geq 2$ is determined by Condition 5.1.*

Suppose $q, \Delta, \ell = O(1)$. By Theorem 7.2, the running time of our algorithm is linear in the size of the update. Hence, the efficient dynamic sampling algorithm exists if strong spatial mixing holds with a rate faster than the neighborhood growth. The relation between the spatial mixing property and the static sampling is well studied, we extend such relation further to the dynamic setting.

The dynamic sampling algorithm is given in Algorithms 3 and 4.

In Algorithm 3, the set $\mathcal{D} \subseteq V$ contains all the vertices incident to the update. Note that the input $\boldsymbol{X}$ can be an infeasible configuration for $\mathcal{I}'$, that is, $w_{\mathcal{I}'}(\boldsymbol{X}) = 0$, because the configuration $X_{\mathcal{D}}$ may

Algorithm 3: Dynamic perfect Gibbs sampler

Input: a spin system instance $\mathcal{I} = (G = (V, E), [q], \boldsymbol{b}, \boldsymbol{A})$, a random sample $\boldsymbol{X} \sim \mu_{\mathcal{I}}$, and an update $(D_V, D_E, \mathcal{C})$ that modifies $\mathcal{I}$ to $\mathcal{I}' = (G' = (V, E'), [q], \boldsymbol{b}', \boldsymbol{A}')$.

$\mathcal{D} \leftarrow D_V \cup \left(\bigcup_{e\in D_E} e\right)$ and $\partial\mathcal{D} \leftarrow \{v \in V \setminus \mathcal{D} \mid \exists u \in \mathcal{D}$ s.t. $\{u, v\} \in E'\}$;

based on $X_{\partial\mathcal{D}}$, modify the partial configuration $X_{\mathcal{D}}$ so that $w_{\mathcal{I}'}(\boldsymbol{X})>0$;

$\mathcal{R} \leftarrow \mathcal{D} \cup \partial\mathcal{D}$;

while $\mathcal{R} \neq \emptyset$ **do**

  $(\boldsymbol{X}, \mathcal{R}) \leftarrow \mathsf{Fix}(\mathcal{I}', \boldsymbol{X}, \mathcal{R})$;

return $\boldsymbol{X}$;

Algorithm 4: Subroutine $\mathsf{Fix}(\mathcal{I}, \boldsymbol{X}, \mathcal{R})$

Input: a spin system instance $\mathcal{I} = (G = (V, E), [q], \boldsymbol{b}, \boldsymbol{A})$, a configuration $\boldsymbol{X} \in [q]^V$, a nonempty subset $\mathcal{R} \subseteq V$, and an integer parameter $\ell \geq 0$;

pick a $u \in \mathcal{R}$ uniformly at random and let $B \leftarrow (B_\ell(u) \setminus \mathcal{R}) \cup \{u\}$;

let $\mu_{\min}$ be the minimum value of $\mu^{\sigma}_{u,\mathcal{I}}(X_u)$ over all $\sigma \in [q]^{\partial B}$ that $\sigma_{\mathcal{R}\cap\partial B} = X_{\mathcal{R}\cap\partial B}$;

with probability $\frac{\mu_{\min}}{\mu^{X_{\partial B}}_{u,\mathcal{I}}(X_u)}$ **do**

  update $\boldsymbol{X}$ by redrawing $X_B \sim \mu^{X_{\partial B}}_{B,\mathcal{I}}$;

  $\mathcal{R} \leftarrow \mathcal{R} \setminus \{u\}$;

else

  $\mathcal{R} \leftarrow \mathcal{R} \cup \partial B$;

return $(\boldsymbol{X}, \mathcal{R})$

violate the new constraints in $\mathcal{C}$. Hence, in line 2, we modify the configuration $X_{\mathcal{D}}$ so that $w_{\mathcal{I}'}(\boldsymbol{X}) > 0$. Given the $X_{\partial\mathcal{D}}$, this step can be achieved by a simple greedy algorithm since $\mathcal{I}'$ is permissive. Then, we construct the initial $\mathcal{R}$ as $\mathcal{D} \cup \partial\mathcal{D}$. In line 5, we call the subroutine Fix on the updated instance $\mathcal{I}'$.

Note that $\mathcal{R} = \mathcal{D} \cup \partial\mathcal{D}$ and $\partial\mathcal{D}$ separates $\mathcal{D}$ from $V \setminus \mathcal{R}$ in both G and G'. In line 2, we only modify the partial configuration $X_{\mathcal{D}}$. Such modification only reviews the information of $\boldsymbol{X}$ in $\mathcal{D} \cup \partial\mathcal{D}$. Thus, after the modification, the $X_{V\setminus\mathcal{R}}$ follows the distribution $\mu_{\mathcal{I}}^{X_{\mathcal{R}}} = \mu_{V\setminus\mathcal{R},\mathcal{I}}^{X_{\partial\mathcal{D}}}$ due to the conditional independence property. Since two instances $\mathcal{I}$ and $\mathcal{I}'$ differ only at the subset $\mathcal{D}$, due to the conditional independence property, two distributions $\mu_{\mathcal{I}}^{X_{\mathcal{R}}}$ and $\mu_{\mathcal{I}'}^{X_{\mathcal{R}}}$ are identical. Thus, the initial $\boldsymbol{X}, \mathcal{R}$ satisfies Condition 4.5 with respect to $\mathcal{I}'$, and $\boldsymbol{X}$ is a feasible configuration for $\mathcal{I}'$. In each iteration of the **while** loop, we call the subroutine Fix on $\mathcal{I}'$. By the identical proof in Section 4, the output $\boldsymbol{X} \sim \mu_{\mathcal{I}'}$.

Let Δ denote the maximum degree of graph G'. Note that $|\mathcal{D}| = O(|D_V| + |D_E|)$. The time complexity of the first three lines of Algorithm 3 is $O(\Delta|\mathcal{D}|)$. Note that the size of the initial $\mathcal{R}$ is $O(\Delta|\mathcal{D}|)$. The efficiency result in Theorem 7.2 follows from the identical proof in Section 5.

8 | CONCLUSION AND OPEN PROBLEMS

The connection between efficient algorithms and spatial mixing (decay of correlation) has been a long-lasting theme in the study of sampling and approximate counting algorithms. In this work, we introduce a new approach for perfect sampling, that relates efficient perfect sampling to strong spatial mixing, for Gibbs distributions on graphs with subexponential neighborhood growth.

Our perfect sampling approach is generic. It is based on the classic Gibbs sampler, while previous perfect sampling techniques were designed for specific systems or subclasses of systems. It is surprising to us that the Gibbs sampler, studied for decades as the go-to algorithm for approximate sampling, can be turned into a perfect sampler by simply adding a filter that accesses only local information.

One key insight in designing our algorithm is to preserve the so called conditional Gibbs property ($\star$), a strong invariant property that implies interruptible perfect sampling as well as dynamic perfect sampling. An important open problem is to establish the same implication from spatial mixing to efficient perfect sampling on general graphs. This is interesting even for special systems, for example, the hardcore or monomer-dimer models.

We want to point out that our current algorithm preserves the conditional Gibbs invariant ($\star$) in a quite pessimistic way: the invariant holds for any conditioning of the random configuration of the "incorrect" variables. We see that even such a straightforward implementation of the invariant is sufficient to give efficient perfect sampling under spatial mixing on graphs of bounded neighborhood growth. It is promising to have cleverer algorithms that exploit the true power of the conditional Gibbs property by implementing this invariant in a more global fashion, on average over randomness.

In general, do efficient perfect sampling algorithms exist whenever efficient approximate samplers exist? Our results provide evidence towards a positive answer, especially for spin systems on graphs with bounded neighborhood growth. However, in general the gap between approximate and perfect sampling persists (e.g. for sampling proper graph colorings [5]). Designing efficient perfect sampling algorithms matching their approximate counterparts remains an interesting research direction.

ACKNOWLEDGMENTS

We thank Jingcheng Liu for helpful comments. Heng Guo and Yitong Yin want to thank the hospitality of the Simons Institute for the Theory of Computing at UC Berkeley, where part of the work was done.

This research has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 947778).

REFERENCES

1. K. S. Alexander, *On weak mixing in lattice models*, Probab. Theory Related Fields **110** (1998), no. 4, 441–471.
2. K. S. Alexander, *Mixing properties and exponential decay for lattice systems in finite volumes*, Ann. Probab. **32** (2004), no. 1A, 441–487.
3. N. Anari, K. Liu, and S. O. Gharan, *Spectral independence in high-dimensional expanders and applications to the hardcore model*, Proc. 2020 IEEE 61st Annu. Sympos. Found. Comput. Sci., IEEE, Virtual Conference, 2020, pp. 1319–1330.
4. M. Bayati, D. Gamarnik, D. Katz, C. Nair, and P. Tetali, "*Simple deterministic approximation algorithms for counting matchings*," *STOC '07*, D. S. Johnson and U. Feige (eds.), ACM, New York, NY, 2007, pp. 122–127.
5. S. Bhandari and S. Chakraborty, *Improved bounds for perfect sampling of k-colorings in graphs*, Proc. 52nd Annu. ACM SIGACT Sympos. Theory Comput., ACM, New York, NY, 2020, pp. 631–642.
6. S. Chen, M. Delcourt, A. Moitra, G. Perarnau, and L. Postle, *Improved bounds for randomly sampling colorings via linear programming*, Proc. 30th Annu. ACM-SIAM Sympos. Discrete Algorithms. SIAM, Philadelphia, PA, 2019, pp. 2216–2234.
7. X. Chen, W. Feng, Y. Yin, and X. Zhang, Rapid mixing of Glauber dynamics via spectral independence for all degrees, 2021, *arXiv preprint arXiv:2105.15005*.
8. Z. Chen, K. Liu, and E. Vigoda. *Rapid mixing of Glauber dynamics up to uniqueness via contraction*, Proc. 2020 IEEE 61st Annu. Sympos. Found. Comput. Sci., Virtual Conference, 2020, 1307–1318.
9. Z. Chen, K. Liu, and E. Vigoda, *Optimal mixing of Glauber dynamics: Entropy factorization via high-dimensional expansion*, Proc. 53rd Annu. ACM SIGACT Sympos. Theory Comput., ACM, New York, NY, 2020, pp. 1537–1550.
10. R. Durrett, *Probability: Theory and examples*, 5th ed., Cambridge University Press, Cambridge, UK, 2019.
11. M. E. Dyer and A. M. Frieze, *Randomly coloring graphs with lower bounds on girth and maximum degree*, Random Struct. Algorithms **23** (2003), no. 2, 167–179.
12. M. E. Dyer, A. M. Frieze, T. P. Hayes, and E. Vigoda, *Randomly coloring constant degree graphs*, Random Struct. Algorithms **43** (2013), no. 2, 181–200.
13. M. E. Dyer, A. Sinclair, E. Vigoda, and D. Weitz, *Mixing in time and space for lattice spin systems: A combinatorial view*, Random Struct. Algorithms **24** (2004), no. 4, 461–479.
14. C. Efthymiou, T. P. Hayes, D. Štefankovic, E. Vigoda, and Y. Yin, "*Convergence of MCMC and loopy BP in the tree uniqueness region for the hard-core model*," *FOCS '16*, I. Dinur (ed.), IEEE, New Brunswick, NJ, 2016, pp. 704–713.
15. W. Feng, K. He, X. Sun, and Y. Yin, Dynamic MCMC sampling, 2019, arXiv preprint arXiv:1904.11807.
16. W. Feng, N. K. Vishnoi, and Y. Yin, "*Dynamic sampling from graphical models*," *STOC '19*, M. Charikar and E. Cohen (eds.), ACM, New York, NY, 2019, pp. 1070–1081.
17. W. Feng and Y. Yin, *On local distributed sampling and counting*, Proc. 2018 ACM Sympos. Principles Distrib. Comput., ACM, New York, NY, 2018, pp. 189–198.
18. J. A. Fill and M. Huber, The randomness recycler: A new technique for perfect sampling, Proc. 41st Annu. Sympos. Found. Comput. Sci., IEEE, Redondo Beach, CA, 2000, pp. 503-511.
19. A. Galanis, D. Štefankovič, and E. Vigoda, *Inapproximability of the partition function for the antiferromagnetic Ising and hard-core models*, Combin. Probab. Comput. **25** (2016), no. 4, 500–559.
20. D. Gamarnik, D. Katz, and S. Misra, *Strong spatial mixing of list coloring of graphs*, Random Struct. Algorithms **46** (2015), no. 4, 599–613.
21. L. A. Goldberg, R. A. Martin, and M. Paterson, *Strong spatial mixing with fewer colors for lattice graphs*, SIAM J. Comput. **35** (2005), no. 2, 486–517.
22. H. Guo and K. He, *Tight bounds for popping algorithms*, Random Struct. Algorithms **57** (2020), no. 2, 371–392.
23. H. Guo and M. Jerrum, *Random cluster dynamics for the Ising model is rapidly mixing*, Ann. Appl. Probab. **28** (2018), no. 2, 1292–1313.
24. H. Guo and M. Jerrum, *A polynomial-time approximation algorithm for all-terminal network reliability*, SIAM J. Comput. **48** (2019), no. 3, 964–978.
25. H. Guo and M. Jerrum, *Approximately counting bases of bicircular matroids*, Combin. Probab. Comput. **30** (2021), no. 1, 124–135.
26. H. Guo and M. Jerrum, *Perfect simulation of the hard disks model by partial rejection sampling*, Ann. Inst. Henri Poincaré Comb. Phys. Interact. **8** (2021), no. 2, 159–177.
27. H. Guo, M. Jerrum, and J. Liu, *Uniform sampling through the Lovász local lemma*, J. ACM **66** (2019), no. 3, 18:1–18:31.
28. O. Häggström and K. Nelander, *Exact sampling from anti-monotone systems*, Stat. Neerl **52** (1998), no. 3, 360–380.

29. T. P. Hayes and E. Vigoda, A nonMarkovian coupling for randomly sampling colorings, Proc. 44th Annu. IEEE Sympos. Found. Comput. Sci., IEEE, Cambridge, MA, 2003, pp. 618-627.
30. M. Huber, A faster method for sampling independent sets, Proc. 11th Annu. ACM-SIAM Sympos. Discrete Algorithms, ACM/SIAM, San Francisco, CA, 2000, pp. 625-626.
31. M. Huber, *Perfect sampling using bounding chains*, Ann. Appl. Probab. **14** (2004), no. 2, 734–753.
32. V. Jain, A. Sah, and M. Sawhney, "*Perfectly sampling* $k \geq (8/3 + o(1))\Delta$*-colorings in graphs*," *STOC '21*, S. Khuller and V. V. Williams (eds.), ACM, New York, NY, 2021, pp. 1589–1600.
33. M. Jerrum, *A very simple algorithm for estimating the number of k-colorings of a low-degree graph*, Random Struct. Algorithms **7** (1995), no. 2, 157–165.
34. M. Jerrum and A. Sinclair, *Approximating the permanent*, SIAM J. Comput. **18** (1989), no. 6, 1149–1178.
35. M. Jerrum, L. G. Valiant, and V. V. Vazirani, *Random generation of combinatorial structures from a uniform distribution*, Theoret. Comput. Sci. **43** (1986), 169–188.
36. L. Liang, P. Lu, and Y. Yin, "*Correlation decay up to uniqueness in spin systems*," *SODA '13*, S. Khanna (ed.), SIAM, New Orleans, LA, 2013, pp. 67–84.
37. J. Liu, A. Sinclair, and P. Srivastava, "*A deterministic algorithm for counting colorings with* 2Δ *colors*," *FOCS '19*, D. Zuckerman (ed.), IEEE Computer Society, Baltimore, MD, 2019, pp. 1380–1404.
38. F. Martinelli, "*Lectures on Glauber dynamics for discrete spin models*," *Lectures on probability theory and statistics (Saint-Flour, 1997)*, Lecture Notes in Mathematics, Vol **1717**, Springer, Berlin, 1999, pp. 93–191.
39. M. Mezard and A. Montanari, *Information, physics, and computation*, Oxford University Press, Oxford, UK, 2009.
40. M. Molloy, *The Glauber dynamics on colorings of a graph with high girth and maximum degree*, SIAM J. Comput. **33** (2004), no. 3, 721–737.
41. E. Mossel and A. Sly, *Exact thresholds for Ising-Gibbs samplers on general graphs*, Ann. Probab. **41** (2013), no. 1, 294–328.
42. J. G. Propp and D. B. Wilson, *Exact sampling with coupled Markov chains and applications to statistical mechanics*, Random Struct. Algorithms **9** (1996), no. 1-2, 223–252.
43. S. Shao and Y. Sun, "*Contraction: A unified perspective of correlation decay and zero-freeness of 2-spin systems*," *ICALP '20*, *LIPIcs*, Vol **168**, A. Czumaj, A. Dawar, and E. Merelli (eds.), Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2020, pp. 96:1–96:15.
44. A. Sinclair, P. Srivastava, and M. Thurley, *Approximation algorithms for two-state anti-ferromagnetic spin systems on bounded degree graphs*, J. Stat. Phys. **155** (2014), no. 4, 666–686.
45. A. Sly, Computational transition at the uniqueness threshold. Proc. 51st Annu. IEEE Sympos. Found. Comput. Sci., IEEE, Las Vegas, NV, 2010, pp. 287-296.
46. A. Sly and N. Sun, *Counting in two-spin models on d-regular graphs*, Ann. Probab. **42** (2014), no. 6, 2383–2416.
47. R. Song, Y. Yin, and J. Zhao, *Counting hypergraph matchings up to uniqueness threshold*, Inform. Comput. **266** (2019), 75–96.
48. Y. Spinka, *Finitary codings for spatial mixing Markov random fields*, Ann. Probab. **48** (2020), no. 3, 1557–1591.
49. J. van den Berg and R. Brouwer, *Random sampling for the monomer-dimer model on a lattice*, J. Math. Phys. **41** (2000), no. 3, 1585–1597.
50. E. Vigoda, *Improved bounds for sampling colorings*, J. Math. Phys. **41** (2000), no. 3, 1555–1569.
51. D. Weitz, *Mixing in time and space for discrete spin systems,* Ph.D. thesis, University of California, Berkeley, ProQuest LLC, Ann Arbor, MI, 2004.
52. D. Weitz, "*Counting independent sets up to the tree threshold*," *STOC '06*, ACM, Seattle, WA, 2006, pp. 140–149.
53. D. B. Wilson, "*Generating random spanning trees more quickly than the cover time*," *STOC '96*, G. L. Miller (ed.), ACM, New York, NY, 1996, pp. 296–303.

How to cite this article: W. Feng, H. Guo, and Y. Yin, *Perfect sampling from spatial mixing*, Random Struct. Algorithms. (2022), 1–32. https://doi.org/10.1002/rsa.21079