# Efficient Parallel Ising Samplers via Localization Schemes

**Xiaoyu Chen**
State Key Laboratory for Novel Software Technology, New Cornerstone Science Laboratory, Nanjing University, China

**Hongyang Liu**
State Key Laboratory for Novel Software Technology, New Cornerstone Science Laboratory, Nanjing University, China

**Yitong Yin**
State Key Laboratory for Novel Software Technology, New Cornerstone Science Laboratory, Nanjing University, China

**Xinyuan Zhang**
State Key Laboratory for Novel Software Technology, New Cornerstone Science Laboratory, Nanjing University, China

## Abstract

We introduce efficient parallel algorithms for sampling from the Gibbs distribution and estimating the partition function of Ising models. These algorithms achieve parallel efficiency, with polylogarithmic depth and polynomial total work, and are applicable to Ising models in the following regimes: (1) Ferromagnetic Ising models with external fields; (2) Ising models with interaction matrix $J$ of operator norm $\|J\|_2 < 1$.

Our parallel Gibbs sampling approaches are based on localization schemes, which have proven highly effective in establishing rapid mixing of Gibbs sampling. In this work, we employ two such localization schemes to obtain efficient parallel Ising samplers: the *field dynamics* induced by *negative-field localization*, and *restricted Gaussian dynamics* induced by *stochastic localization*. This shows that localization schemes are powerful tools, not only for achieving rapid mixing but also for the efficient parallelization of Gibbs sampling.

**2012 ACM Subject Classification** Theory of computation → Random walks and Markov chains; Theory of computation → Parallel algorithms

**Keywords and phrases** Localization scheme, parallel sampling, Ising model

**Digital Object Identifier** 10.4230/LIPIcs.APPROX/RANDOM.2025.46

**Category** RANDOM

**Related Version** *Previous Version*: https://arxiv.org/abs/2505.05185

**Funding** This work has been partially supported by the New Cornerstone Science Foundation.
*Xinyuan Zhang*: NSFC Young Student Basic Research Program (PhD candidate, No. 623B2051).

## 1 Introduction

The Ising model, introduced by Lenz and [26] in statistical physics to study the criticality of ferromagnetism, has since found numerous applications and has been extensively studied across various fields, including probability theory, statistical learning, and computer science.

Let $G = (V, E)$ be a connected undirected graph with $n = |V|$ vertices and $m = |E|$ edges. Let $\boldsymbol{\beta} \in (0, +\infty)^E$ represent the edge activities, and $\boldsymbol{\lambda} \in [0, 1]^V$ represent the external fields. The *Gibbs distribution* $\mu^{\text{Ising}}_{\boldsymbol{\beta},\boldsymbol{\lambda}}$ of the Ising model on the graph $G$ with parameters $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}$ is supported on $2^V$ and is given by:


$$\forall S \subseteq V, \quad \mu^{\text{Ising}}_{\boldsymbol{\beta},\boldsymbol{\lambda}}(S) := \frac{1}{Z^{\text{Ising}}_{\boldsymbol{\beta},\boldsymbol{\lambda}}} \prod_{e \in m(S)} \beta_e \prod_{v \in S} \lambda_v,$$

where $m(S) := \{e = (u, v) \in E \mid u, v \in S \text{ or } u, v \notin S\}$ is the set of "monochromatic" edges, i.e., edges where both endpoints are either in $S$ or both outside of $S$. The normalizing factor, called the *partition function*, is given by: $Z^{\text{Ising}}_{\boldsymbol{\beta},\boldsymbol{\lambda}} := \sum_{S \subseteq E} \prod_{e \in m(S)} \beta_e \prod_{v \in S} \lambda_v$.

The Ising model is called *ferromagnetic* if $\boldsymbol{\beta} \in (1, +\infty)^E$, meaning that pairwise interactions favor monochromatic edges, and the model is called *anti-ferromagnetic* if $\boldsymbol{\beta} \in (0, 1)^E$.

A central problem in the study of Ising models is sampling from the Gibbs distribution. This task is essential not only for estimating the partition function but also for various other inference and learning problems associated with the Ising model.

## 1.1 Parallel sampler for ferromagnetic Ising model

In a seminal work by [29], it was shown that for ferromagnetic Ising models, approximately sampling from the Gibbs distribution and approximately computing the partition function are both tractable in polynomial time. This result, along with celebrated breakthroughs in polynomial-time approximations of permanents [28, 30] and volumes of convex bodies [15], remarkably showcased the power of random sampling in approximately solving #P-hard inference problems on polynomial-time Turing machines.

Motivated by emerging applications in large-scale data and large models, parallel Gibbs sampling has recently drawn considerable attention [37, 3, 5, 1, 36, 4, 22, 20, 23, 21]. Despite these advances, the tractability of the ferromagnetic Ising model through parallel algorithms remains unresolved, in contrast to the successes of polynomial-time sequential algorithms.

In the theory of computing, the parallel complexity of the ferromagnetic Ising model is of foundational significance. In a seminal work, [39] asked whether an RNC algorithm (with polylogarithmic depth on polynomial processors) exists for sampling bipartite perfect matchings, which would imply an RNC approximation algorithm for the permanent of Boolean matrices. To this day, this remains a major open problem. [41] conjectured that no RNC algorithm exists for this task, making it a rare example of a polynomial-time tractable problem suspected to be intrinsically sequential, yet not known to be P-complete. The challenge of parallelizing the ferromagnetic Ising sampler is closely tied to the problem of sampling matchings, as both problems were resolved sequentially using the same canonical path argument [28, 29].

In this work, we present an efficient parallel sampler in the RNC class for the general ferromagnetic Ising model with nonzero external fields. To the best of our knowledge, this is the first RNC sampler for Ising model beyond the critical phase-transition threshold.

▶ **Theorem 1** (ferromagnetic Ising sampler). *Let $\delta \in (0, 1)$ be a constant. There is a parallel algorithm that, given $\epsilon \in (0, 1)$ and an Ising model on a graph $G = (V, E)$ with parameters $\boldsymbol{\beta} \in [1+\delta, +\infty)^E$ and $\boldsymbol{\lambda} \in [0, 1-\delta]^V$, outputs a sample from the Gibbs distribution $\mu^{-Ising}_{\boldsymbol{\beta},\boldsymbol{\lambda}}$ within total variation distance $\epsilon$ in $(\epsilon^{-\frac{1}{\log n}} \log n)^{O_\delta(1)}$ parallel time using $O_\delta(m^2 \log \frac{n}{\epsilon})$ processors.*

One can view this result as a parallel counterpart to [29]. Through a standard reduction via non-adaptive annealing, this RNC Ising sampler can be turned into an RNC algorithm for approximating the Ising partition function $Z^{\text{Ising}}_{\boldsymbol{\beta},\boldsymbol{\lambda}}$. Specifically, applying the work-efficient parallel annealing algorithm recently developed in [38], the parallel Ising sampler stated

in Theorem 1 can be transformed to a randomized parallel algorithm which returns an $(1 \pm \epsilon)$-approximation of the Ising partition function in $(\epsilon^{-\frac{1}{\log n}} \cdot \log n)^{O_\delta(1)}$ parallel time on $\tilde{O}_\delta(m^2 n/\epsilon^2)$ machines.

The parallel Ising sampler in Theorem 1 is implied by a parallel sampler for the *random cluster model* (see Section 2.2 for definition). The correspondence between the Ising model and the random cluster model is due to the well-known *Edwards-Sokal coupling* [16] (see Lemma 6).

▶ **Theorem 2** (random cluster sampler). *Let $\delta \in (0,1)$ be a constant. There is a parallel algorithm that, given $\epsilon \in (0,1)$ and a random cluster model on a graph $G = (V, E)$ with parameters $\boldsymbol{p} \in [\delta, 1)^E$ and $\boldsymbol{\lambda} \in [0, 1-\delta]^V$, outputs a sample from the random cluster distribution $\mu^{\mathrm{RC}}_{E,\boldsymbol{p},\boldsymbol{\lambda}}$ within total variation distance $\epsilon$ in $(\epsilon^{-\frac{1}{\log n}} \cdot \log n)^{O_\delta(1)}$ parallel time on $O_\delta(m^2 \log \frac{n}{\epsilon})$ processors.*

Theorem 1 follows from Theorem 2 via the Edwards-Sokal coupling, as formally described in Section 3. Our main technical contribution in this part is a parallel sampler for the random cluster model, as stated in Theorem 2, whose efficiency is proved in Section 4.

## 1.2 Parallel sampler for Ising models with contracting interaction matrix

An Ising model can be formulated through the interaction matrix. Let $V$ be a set of $n$ vertices. The *Gibbs distribution* $\mu^{\mathrm{Ising}}_{J,h}$ of the Ising model, supported on $\{\pm 1\}^V$, is defined by the *interaction matrix* $J \in \mathbb{R}^{V \times V}$, which is a symmetric positive semi-definite matrix, and the *external fields* $h \in \mathbb{R}^V$:

$$\forall \sigma \in \{\pm 1\}^V, \quad \mu^{\mathrm{Ising}}_{J,h} \propto \exp\left(\frac{1}{2}\sigma^\top J \sigma + h^\intercal \sigma\right). \tag{1}$$

Recent studies have established rapid mixing for Ising models in terms of their interaction norm [18, 6], showing that Glauber dynamics mixes rapidly when $\|J\|_2 < 1$, whereas for $\|J\|_2 > 1$ sampling becomes intractable [34, 24, 31]. For parallel sampling, [36] proposed an RNC algorithm under a stricter condition that requires $J$ to have a constant $\ell_\infty$-norm. However, under the broader condition $\|J\|_2 < 1$, the best known parallel algorithms [32, 37] achieve only a depth of $\widetilde{O}(\sqrt{n})$, highlighting a significant gap in achieving optimal parallelism.

In this work, we present an RNC parallel sampler for the Ising model under the condition $\|J\|_2 < 1$. To the best of our knowledge, this is the first RNC Ising sampler that matches this critical threshold for rapid mixing in terms of the interaction norm.

▶ **Theorem 3** (Ising sampler). *Let $\eta \in (0,1)$ be a constant. There is a parallel algorithm that, given $\epsilon \in (0,1)$ and an Ising model with an interaction matrix $J \in \mathbb{R}^{V \times V}$ satisfying*

$$0 \prec \frac{\eta}{2} I \preceq J \preceq (1 - \frac{\eta}{2}) I$$

*and arbitrary external fields $h \in \mathbb{R}^V$, outputs a sample from the Gibbs distribution $\mu^{\mathrm{Ising}}_{J,h}$ within total variation distance $\epsilon$ in $O_\eta(\log^4(\frac{n}{\epsilon}))$ parallel time using $\widetilde{O}_\eta(n^3/\epsilon^2)$ processors.*

▶ Remark 4. Note that for the Ising model with interaction matrix $J$, for every constant $C$, taking $J' = J + CI$ results in the same Gibbs distribution. Thus, in the regime where $\|J\|_2 \le 1 - \eta$, for which Glauber dynamics is known to mix rapidly [18, 6], one can add a diagonal matrix to the interaction matrix $J$ in order to satisfy the condition of Theorem 3.

The RNC sampler stated in Theorem 3 is presented in Section 5.

## 1.3 Technique overview

Classic local Markov chains, such as Glauber dynamics, are inherently sequential, updating the spin of a single site at each step based on its neighbors. Recently, [36, 37] proposed a generic parallelization framework using *correlated sampling* (or *universal coupling*), enabling polylogarithmic-depth simulation of Glauber dynamics under a relaxed Dobrushin condition. When applied to the Ising model, this yields an RNC sampler in the uniqueness regime $\boldsymbol{\beta} \in \left(\frac{\Delta-2}{\Delta}, \frac{\Delta}{\Delta-2}\right)^E$, where $\Delta$ is the maximum degree. However, these conditions do not hold for the Ising models considered in this work.

Instead of directly parallelizing a local Markov chain, our approach focuses on simulating "global" Markov chains that update $O(n)$ spins in each step, satisfying the following:

- The chain mixes in polylogarithmic steps;
- Each step of the chain has an efficient parallel implementation in RNC.

These together ensure an RNC sampler with polylog depth and polynomial total work.

For the ferromagnetic Ising model, this global chain corresponds to the *field dynamics* introduced by [10], while for the Ising model with a contracting interaction matrix, the global chain is the *restricted Gaussian dynamics* (also known as the *proximal sampler*) introduced by [33]. Both of these global Markov chains can be viewed as arising from different localization schemes for the Ising Gibbs sampling.

Proposed by [13], the localization schemes provide an abstract framework that generalizes high-dimensional expander walks and encompasses a wide range of stochastic processes. This framework can be interpreted in terms of noising and denoising processes, as discussed in [17, 9]. Let $X \sim \mu$ be drawn from the target distribution $\mu$. A noisy channel $D$, which is a Markov chain, is applied to $X$ to obtain a "noised" sample $Y \sim D(X, \cdot)$. The joint distribution of $(X, Y)$ is denoted as $\pi$. The denoising process $U$ represents the time reversal of $D$, where it draws $Z$ from the posterior distribution $U(Y, \cdot) := \pi(\cdot \mid Y)$. The Markov chains associated with the localization scheme update the current state $X$ to a new state $Z$ according to the rules:

- Add noise to $X$ through the noisy channel: sample $Y \sim D(X, \cdot)$;
- Denoise $Y$ via the posterior distribution: sample $Z \sim U(Y, \cdot)$.

In particular, when the noisy channel $D$ corresponds to the continuous-time down walk channel, which drops each element in a set $X$ with a fixed probability, the localization process is called the *negative-field localization*, and the above associated global Markov chain is the *field dynamics*. Alternatively, when $D$ corresponds to a Gaussian noise channel, which adds Gaussian noise to the sample $X$, the localization process is called *stochastic localization*, and the resulting global Markov chain becomes the *restricted Gaussian dynamics*.

The localization scheme has previously proven remarkably effective in establishing rapid mixing of Glauber dynamics up to criticality [10, 11, 7, 13, 12, 8, 9]. Specifically, the localization process can effectively "tilt" the model parameters towards sub-critical directions. This allows the mixing properties established in sub-critical regimes to be conserved up to criticality, provided the associated global chains mix rapidly.

In the current work, we explore another perspective on localization schemes: leveraging the associated global Markov chains to achieve parallel efficiency in sampling. This task is highly non-trivial, as these chains were originally designed as analytical tools for studying mixing times, and their efficient parallel simulation poses significant challenges.

For the ferromagnetic Ising model with external fields, we leverage the field dynamics of the random-cluster model to design a parallel sampler. The field dynamics is the global Markov chain induced by negative-field localization. Each update consists of two steps:

1. A noising step $Y \sim D(X, \cdot)$ which drops each element in $X$ with a fixed probability (as in Line 4 of Algorithm 1), and is straightforward to implement in parallel.
2. A denoising step $Z \sim U(Y, \cdot)$, which requires sampling from a tilted posterior distribution $U(Y, \cdot)$ and is non-trivial. We simulate the denoising step via Glauber dynamics on $U(Y, \cdot)$, which corresponds to a sub-critical (low-temperature) random-cluster model, where rapid mixing is easier to establish.

A similar idea was employed in [12] to develop a near-linear time sequential sampler. Here, we further parallelize this Glauber dynamics on $U(Y, \cdot)$ using the parallel simulation algorithm introduced in [37], which yields Algorithm 2. A key challenge, however, is that despite being tilted to the low-temperature regime, the Dobrushin condition required in [37] to ensure efficient parallel simulation does not hold. To overcome this, we establish a new "*coupling with the stationary*" criterion, which significantly relaxes the Dobrushin condition and guarantees the efficiency of parallel simulation of Glauber dynamics. Altogether, this yields parallel random-cluster and ferromagnetic-Ising samplers with polylogarithmic depth and polynomial total work.

For the Ising model with a contracting interaction matrix $J$, where $\|J\|_2 < 1$, we leverage the restricted Gaussian dynamics (also known as the proximal sampler) to design a parallel sampler. This dynamics defines a global Markov chain, which is the associated chain of the stochastic localization for the Ising model. Each update consists of two steps:

1. A noising step $Y \sim D(X, \cdot)$, which introduces Gaussian noise to the current sample.
2. A denoising step $Z \sim U(Y, \cdot)$, which resamples from a posterior distribution $U(Y, \cdot)$.

With appropriately chosen parameters, $U(Y, \cdot)$ becomes a product distribution, allowing an efficient parallel implementation of the denoising process. The non-trivial step is the noising step $Y \sim D(X, \cdot)$, which samples from a high-dimensional Gaussian distribution with mean $X$ and covariance $J^{-1}$ (as in Line 3 of Algorithm 3). A key observation is that the Gaussian distribution is log-concave, so the noising step can be efficiently approximated using recently developed parallelizations of Langevin Monte Carlo [2].

In a related prior work [5], an efficient parallel algorithm was developed for transport-stable distributions by implementing stochastic localization in discrete steps. However, their approach is fundamentally different from ours. Their method leverages the log-concavity of the tilted (or "noised") distribution $D(X, \cdot)$, which enables efficient parallel sampling. In contrast, our approach relies on the rapid mixing of the proximal sampler, which also admits an efficient parallel implementation.

## 2 Preliminaries

### 2.1 Notation

Given a finite nonempty ground set $E$, we will use boldface letters, such as $\boldsymbol{p}$, to denote a vector in $\mathbb{R}^E$. Given a vector $\boldsymbol{a} \in \mathbb{R}^E$ and function $f : \mathbb{R} \to \mathbb{R}$, we write $\boldsymbol{b} = f(\boldsymbol{a})$ for the vector $\boldsymbol{b} \in \mathbb{R}^E$ satisfying that $b_e = f(a_e)$ for all $e \in E$.

Throughout the paper, we use log to denote the natural logarithm.

### 2.2 Random cluster model

Let $G = (V, E)$ be a connected, undirected graph. The random cluster model on $G = (V, E)$ with edge probabilities $\boldsymbol{p} \in [0, 1]^E$ and vertex weights $\boldsymbol{\lambda} \in [0, 1]^V$ is defined as follows. The random cluster distribution $\mu^{\mathrm{RC}}_{E,\boldsymbol{p},\boldsymbol{\lambda}}$ is supported on the power set $2^E$, where each $S \subseteq E$ is assigned a weight:

$$w^{\mathrm{RC}}_{E,\boldsymbol{p},\boldsymbol{\lambda}}(S) := \prod_{e\in S} p_e \prod_{e\in E\setminus S} (1-p_e) \prod_{C\in\kappa(V,S)} \left(1+\prod_{j\in C}\lambda_j\right), \tag{2}$$

where $\kappa(V,S)$ is the set of connected components in the graph $(V,S)$. The distribution $\mu^{\mathrm{RC}}_{E,\boldsymbol{p},\boldsymbol{\lambda}}$ is then given by:

$$\forall S\subseteq E,\quad \mu^{\mathrm{RC}}_{E,\boldsymbol{p},\boldsymbol{\lambda}}(S) := \frac{w^{\mathrm{RC}}_{E,\boldsymbol{p},\boldsymbol{\lambda}}(S)}{Z^{\mathrm{RC}}_{E,\boldsymbol{p},\boldsymbol{\lambda}}},$$

where the normalizing factor $Z^{\mathrm{RC}}_{E,\boldsymbol{p},\boldsymbol{\lambda}} := \sum_{S\subseteq E} w^{\mathrm{RC}}_{E,\boldsymbol{p},\boldsymbol{\lambda}}(S)$ is the partition function.

## 2.3 Markov chains and mixing time

Let $\Omega$ be a finite state space, and let $(X_t)_{t\geq 0}$ be a Markov chain over $\Omega$ with transition matrix $P$. We use $P$ to represent the Markov chain for convenience. It is well known that a finite Markov chain $P$ converges to a unique stationary distribution $\mu$ if $P$ is irreducible and aperiodic (see [35] for the definitions of these concepts). The *mixing time* of a Markov chain $P$ is defined by

$$T_{\mathrm{mix}}(\epsilon) := \max_{x_0\in\Omega} \min\{t\in\mathbb{N} \mid d_{\mathrm{TV}}(P^t(x_0,\cdot),\mu) < \epsilon\},$$

where $d_{\mathrm{TV}}(\boldsymbol{p},\boldsymbol{q}) := \frac{1}{2}\|\boldsymbol{p}-\boldsymbol{q}\|_1$ is the total variation distance.

### 2.3.1 Glauber dynamics

The Glauber dynamics is a canonical Markov chain for sampling from joint distributions. Let $\mu$ be a distribution over $2^E$ on a finite ground set $E$. For any $X\subseteq E$ and $u\in E$, let $\mu^X_u$ be the marginal probability of $u$ given $X$, formally defined as:

$$\mu^X_u := \frac{\mu(X\cup\{u\})}{\mu(X\cup\{u\})+\mu(X\setminus\{u\})}.$$

The Glauber dynamics $P^{\mathrm{GD}}$ updates a configuration $X\subseteq E$ according to the following rule:

- Pick an element $u\in E$ uniformly at random;
- With probability $\mu^X_u$, replace $X$ with $X\cup\{u\}$; otherwise, replace $X$ with $X\setminus\{u\}$.

It is well known that $\mu$ is the unique stationary distribution of this chain [35].

### 2.3.2 Field dynamics

Field dynamics is a novel Markov chain introduced in [10]. Let $\mu$ be a distribution over $2^E$ on a finite ground set $E$. The field dynamics $P^{\mathrm{FD}}_\theta$ with parameter $\theta\in(0,1)$ updates a configuration $X\subseteq E$ according to the following rule:

- Add each element $u\in E$ into a set $S$ independently with probability $\theta$;
- Replace $X$ with a random $Y\subseteq E$ that follows the law of $\theta^{-1}*\mu$ conditioned on $Y\subseteq X\cup S$, where $\theta^{-1}*\mu$ denotes the distribution supported on $2^E$ and defined as:

$$\forall T\subseteq E,\quad (\theta^{-1}*\mu)(T)\propto\theta^{-|T|}\mu(T).$$

The field dynamics can be thought of as an adaptive block dynamics, where a block of sites is chosen for updating, adapted to the current configuration. It was shown in [10] that $\mu$ is the unique stationary distribution of the field dynamics.

For the random cluster model on graph $G = (V, E)$ with parameters $\boldsymbol{p}$ and $\boldsymbol{\lambda}$, the second step of the field dynamics corresponds to:

- Replace $X$ with a random subgraph $Y$ distributed as $\mu^{\mathrm{RC}}_{S\cup X,\boldsymbol{p}^\star,\lambda}$, where $\mathbf{p}^\star = \frac{\boldsymbol{p}}{\boldsymbol{p}+\theta(1-\boldsymbol{p})}$.

## 2.4 Model of computation

We assume the concurrent-read exclusive-write (CRCW) parallel random access machine (PRAM) [27] as our model of computation, where concurrent writes to the same memory location are allowed, and an arbitrary value written concurrently is stored. The computational complexity is measured by the number of rounds (parallel time steps) and the number of processors (or machines) used in the worst case.

We use NC to refer to both the class of algorithms that run in polylogarithmic time using a polynomial number of processors and the class of problems solvable by such algorithms. RNC denotes the randomized counterpart of NC.

# 3 Parallel random-cluster and Ising samplers via field dynamics

We propose a parallel algorithm for sampling from the random-cluster model, based on the simulation of field dynamics. Consider a random-cluster model defined on a graph $G = (V, E)$ with edge probabilities $\boldsymbol{p} \in (0, 1]^E$ and vertex weights $\boldsymbol{\lambda} \in [0, 1)^V$.

The main algorithm is described in Algorithm 1.

**Algorithm 1** Random cluster field dynamics.

**Input** : Graph $G(V, E)$, $\mathbf{p} \in (0, 1)^E, \boldsymbol{\lambda} \in [0, 1)^V$, and error bound $\epsilon \in (0, 1)$.
**Parameter** : $\theta \in (0, 1)$, $N_0 > 0$ and integer $T^{\mathrm{FD}} \geq 1$.
**Output** : A random configuration $X \subseteq E$ satisfying $d_{\mathrm{TV}}\left(X, \mu^{\mathrm{RC}}_{E,\mathbf{p},\boldsymbol{\lambda}}\right) \leq \epsilon$.

initialize $X \leftarrow E$;
**if** $|V| \leq N_0$ **then return** $X \sim \mu^{RC}_{E,\boldsymbol{p},\boldsymbol{\lambda}}$ by brute force ;
**forall** $i = 1, 2, \ldots, T^{\mathrm{FD}}$ **do**
  construct $S \subseteq E$ by including each $e \in E$ independently with probability $\theta$;
  $X \leftarrow$ parallel-RC $\left((V, S \cup X), \mathbf{p}^\star, \boldsymbol{\lambda}, \left(2T^{\mathrm{FD}}\right)^{-1} \epsilon\right)$, where $\boldsymbol{p}^\star = \frac{\boldsymbol{p}}{\boldsymbol{p}+\theta(1-\boldsymbol{p})}$;
**end**
**return** $X$;

Algorithm 1 simulates the *field dynamics* (defined in Section 2.3.2), the Markov chain induced by the *negative-field localization process* for the random cluster model $\mu^{\mathrm{RC}}_{E,\mathbf{p},\boldsymbol{\lambda}}$. At each transition step, the algorithm samples from $\mu^{\mathrm{RC}}_{S\cup X,\mathbf{p}^\star,\lambda}$, where $\mathbf{p}^\star = \frac{\boldsymbol{p}}{\boldsymbol{p}+\theta(1-\boldsymbol{p})}$. This step is implemented using the parallel-RC subroutine, described in Algorithm 2.

The subroutine parallel-RC $(G, \mathbf{p}, \boldsymbol{\lambda}, \epsilon)$, as presented in Algorithm 2, is a parallel algorithm designed for sampling from the random cluster distribution $\mu^{\mathrm{RC}}_{E,\mathbf{p},\boldsymbol{\lambda}}$. It simulates the Glauber dynamics for $\mu = \mu^{\mathrm{RC}}_{E,\mathbf{p},\boldsymbol{\lambda}}$ through a parallel simulation approach introduced in [36], using *correlated sampling*, also known as *universal coupling*. Specifically, Algorithm 2 employs inverse transform sampling as the universal coupler. However, prior analyses in [36, 37] rely on a variant of the Dobrushin condition, which does not hold in our context, and thus overcoming this barrier require novel techniques.

**Table 1** Parameters assumed by Algorithm 1.

| parameter | value |
|---|---|
| $\theta$ | $\mathrm{e}^{-100}\exp\left(\frac{10\log(\epsilon/2)}{\log n}\right)\exp\left(-140(1-\lambda_{\max})^{-2}\right)\cdot p_{\min}/\log n$ |
| $N_0$ | $\max\left\{\exp\left(60\ (1-\lambda_{\max})^{-2}\right), \frac{3}{p_{\min}}, \sqrt{\frac{\log(2/\epsilon^2)}{\log n}}\right\}$ |
| $T^{\mathrm{FD}}$ | $\left\lceil\left(\frac{\mathrm{e}}{\theta}\right)^{5(1-\lambda_{\max})^{-2}}\left(2\log n+\log\log\frac{2}{p_{\min}}+\log\frac{2}{\epsilon^2}\right)\right\rceil$ |
| $T^{\mathrm{GD}}$ | $\left\lceil 2m(\log m+\log(8T^{\mathrm{FD}}/\epsilon))\right\rceil$ |
| $T^{\mathrm{PA}}$ | $\left\lceil 3\log\left(\frac{8T^{\mathrm{GD}}T^{\mathrm{FD}}}{\epsilon}\right)\right\rceil$ |

Let $n=|V|$ and $m=|E|$. The parameters $\theta$, $N_0$, $T^{\mathrm{FD}}$, $T^{\mathrm{GD}}$, and $T^{\mathrm{PA}}$ assumed by the algorithms are specified in Table 1. We will show that the algorithm is both efficient and correct under these parameters.

**Algorithm 2** parallel-RC $(G,\mathbf{p},\boldsymbol{\lambda},\epsilon)$.

**Input** : Graph $G(V,E)$, $\mathbf{p}\in(0,1)^E$, $\boldsymbol{\lambda}\in[0,1)^V$, and error bound $\epsilon\in(0,1)$.
**Parameter** : Integers $T^{\mathrm{GD}},T^{\mathrm{PA}}\geq 1$.
**Output** : A random configuration $X\subseteq E$ satisfying $d_{\mathrm{TV}}\left(X,\mu^{\mathrm{RC}}_{E,\mathbf{p},\boldsymbol{\lambda}}\right)\leq\epsilon$.

1 generate $X'\subseteq E$ by including each $e\in E$ independently with probability $\mu^E_e$;
2 **forall** $i=1,2,...,T^{\mathrm{GD}}$ **in parallel do**
3  generate $e_i\in E$ and $\mathcal{R}_i\in[0,1]$ uniformly at random;
4  $Y_i^0\leftarrow I[e_i\in X']$;
5 **end**
6 **forall** $i=1,2,...,T^{\mathrm{GD}}$ *and* $e\in E$ **in parallel do**
7  calculate $\mathrm{pred}_i(e)\leftarrow\max\{j\geq 1\mid j<i\wedge e_j=e\}\cup\{0\}$;
8 **end**
9 **for** $t=1,2,...,T^{\mathrm{PA}}$ **do**
10  **forall** $i=1,2,...,T^{\mathrm{GD}}$ **in parallel do**
11  construct $\sigma=\sigma_i^t\subseteq E$ as:
 $\sigma_i^t\leftarrow\{e\in E\mid(\mathrm{pred}_i(e)=0\wedge e\in X')\vee(\mathrm{pred}_i(e)\geq 1\wedge Y^{t-1}_{\mathrm{pred}_i(e)}=1)\}$;
12  calculate marginal probability $\mu^{\sigma}_{e_i}\leftarrow\frac{\mu(\sigma\cup\{e_i\})}{\mu(\sigma\cup\{e_i\})+\mu(\sigma\setminus\{e_i\})}$;
13  **if** $\mathcal{R}_i<\mu^{\sigma}_{e_i}$ **then** $Y_i^t\leftarrow 1$ **else** $Y_i^t\leftarrow 0$;
14  **end**
15 **end**
16 $X\leftarrow$
 $\left\{e\in E\mid(j_e=0\wedge e\in X')\vee(j_e\geq 1\wedge Y_{j_e}^{T^{\mathrm{PA}}}=1), j_e\leftarrow\max\{j\geq 1\mid e_j=e\}\cup\{0\}\right\}$;
17 **return** $X$;

Specifically, we show that if the random cluster model is in the low-temperature regime (i.e., corresponding to an Ising model with large $\beta>1$), then it holds simultaneously:

1. The Glauber dynamics gets sufficiently close to $\mu^{\mathrm{RC}}_{E,\mathbf{p},\boldsymbol{\lambda}}$ within $T^{\mathrm{GD}}$ steps.
2. The parallel iterative updates (the **for** loop in Line 9 of Algorithm 2) stabilize in $T^{\mathrm{PA}}$ rounds, faithfully simulating the $T^{\mathrm{GD}}$-step Glauber dynamics with high probability.

Therefore, Algorithm 2 is an efficient parallel sampler in the low-temperature regime.

As a result of the localization scheme, the field dynamics in Algorithm 1 "tilts" the current instance $\mu^{\mathrm{RC}}_{E,\mathbf{p},\boldsymbol{\lambda}}$ to a new random cluster distribution $\mu^{\mathrm{RC}}_{S\cup X,\mathbf{p}^\star,\boldsymbol{\lambda}}$ at low temperature in each round. Within this regime, the Glauber dynamics is rapidly mixing and can be faithfully simulated by Algorithm 2 in parallel. The field dynamics mixes in polylogarithmic rounds according to [12].

Combining everything together, we obtain the following theorem for parallel sampler.

▶ **Theorem 5.** *Let $\delta \in (0,1)$ be a constant, and let $\boldsymbol{p} \in [\delta,1)^E$ and $\boldsymbol{\lambda} \in [0,1-\delta]^V$. Then:*
1. *Algorithm 1 halts in $(\epsilon^{-\frac{1}{\log n}} \cdot \log n)^{O_\delta(1)}$ parallel time using $O_\delta(m^2 \log \frac{n}{\epsilon})$ processors.*
2. *Algorithm 1 outputs a random $X \subseteq E$ with $d_{\mathrm{TV}}\left(X, \mu^{\mathrm{RC}}_{E,\boldsymbol{p},\boldsymbol{\lambda}}\right) \le \epsilon$.*

Our main theorem for the random cluster sampler (Theorem 2) follows directly from Theorem 5. Our main theorem for the Ising sampler (Theorem 1) follows from Theorem 5 as well, through the Edwards-Sokal coupling [16], which connects the distributions of the Ising and random cluster models. For our purposes, we use a variant of the ES coupling from [19].

▶ **Lemma 6** ([19, Proposition 2.3])**.** *For the Ising model on a graph $G=(V,E)$ with parameters $\boldsymbol{\beta} \in (1,+\infty)$ and $\boldsymbol{\lambda} \in (0,1)$, the following process generate a sample $X \sim \mu^{\mathrm{Ising}}_{\boldsymbol{\beta},\boldsymbol{\lambda}}$:*
1. *Sample $S \subseteq E$ according to the distribution $\mu^{\mathrm{RC}}_{E,\boldsymbol{p},\boldsymbol{\lambda}}$, where $\boldsymbol{p} = 1-\boldsymbol{\beta}^{-1}$;*
2. *For each connected component $C$ in the graph $(V,S)$, add $C$ to $X$ with probability $\frac{\prod_{u\in C}\lambda_u}{1+\prod_{u\in C}\lambda_u}$.*

Furthermore, the ES coupling can be efficiently implemented in parallel.

▶ **Lemma 7.** *Item 2 of Lemma 6 can be computed in $O(\log n)$ parallel time using $O(m)$ processors.*

**Proof.** Use the connected component algorithm in [40], to compute all components of the graph $G'=(V,S)$ in $O(\log n)$ parallel time using $O(n+m)$ processors. For each connected component $C$, the product $\prod_{u\in C}\lambda_u$ can be computed in $O(\log |C|)$ parallel time using $O(|C|)$ processors by divide-and-conquer. The overall complexity is $O(\log n)$ parallel time using $O(m)$ processors. ◀

**Proof of Theorem 1.** By Lemma 6, a sample from $\mu^{\mathrm{Ising}}_{\boldsymbol{\beta},\boldsymbol{\lambda}}$ can be generated as follows:
1. Generate a random subset $S \subseteq E$ using Algorithm 1, where $\boldsymbol{p} = 1-\boldsymbol{\beta}^{-1}$;
2. Apply Item 2 of Lemma 6 to construct $X$.

By Theorem 5, we have $d_{\mathrm{TV}}\left(S, \mu^{\mathrm{RC}}_{E,\boldsymbol{p},\boldsymbol{\lambda}}\right) \le \epsilon$, which, combined with Lemma 6, ensures that $d_{\mathrm{TV}}\left(X, \mu^{\mathrm{Ising}}_{\boldsymbol{\beta},\boldsymbol{\lambda}}\right) \le \epsilon$. The efficiency follows from Theorem 5 and Lemma 7. ◀

It remains to prove Theorem 5. The key steps of the proof are outlined in Section 4, while the full proof is deferred to Section C.

## 4 A coupling-with-stationary criterion for parallel simulation

In this section, we outline the key ideas underlying the proof of Theorem 5.

Consider the random cluster model on $G=(V,E)$ with parameters $\boldsymbol{p} \in (0,1]^E$ and $\boldsymbol{\lambda} \in [0,1)^V$. The distribution $\mu = \mu^{\mathrm{RC}}_{E,\boldsymbol{p},\boldsymbol{\lambda}}$ is defined over subgraphs $S \subseteq E$ and is proportional to the weight function $w = w^{\mathrm{RC}}_{E,\boldsymbol{p},\boldsymbol{\lambda}}$, given by:

$$\mu(S) \propto w(S) := \prod_{e\in S} p_e \prod_{f\in E\setminus S}(1-p_e) \prod_{C\in\kappa(V,S)}\left(1+\prod_{j\in C}\lambda_j\right),$$

where $\kappa(V,S)$ denotes the set of all connected components in the graph $(V,S)$.

Consider the Glauber dynamics $(X_t)_{t\geq 0}$ for sampling from $\mu$, starting from an initial state $X_0 \subseteq E$. At each step, the process updates $X_{i-1}$ to $X_i$ as follows: Select an edge $e_i \in E$ and a real number $\mathcal{R}_i \in [0,1]$ uniformly at random, and update $X_i$ according to:

$$X_i = \begin{cases} X_{i-1} \cup \{e_i\} & \text{if } \mathcal{R}_i < \mu_{e_i}^{X_{i-1}} \\ X_{i-1} \setminus \{e_i\} & \text{otherwise} \end{cases}, \tag{3}$$

where $\mu_{e_i}^{X_{i-1}}$ represents the marginal probability of $e_i$ given $X_{i-1}$, formally defined as

$$\mu_e^X := \frac{\mu(X \cup \{e\})}{\mu(X \cup \{e\}) + \mu(X \setminus \{e\})} = \frac{w(X \cup \{e\})}{w(X \cup \{e\}) + w(X \setminus \{e\})}.$$

It is straightforward to verify that $(X_t)_{t\geq 0}$ is the Glauber dynamics defined in Section 2.3.1.

We will show that Algorithm 2 faithfully simulates the Glauber dynamics, provided that a "coupling with stationarity" condition holds, as formally stated below.

▶ **Condition 1.** Let $\eta \in (0,1)$ be a parameter. There is a $\mathcal{G} \subseteq 2^E$ satisfying the following conditions:

1. For any $X, Y \in \mathcal{G}$, the total variation in marginal probabilities across edges satisfies

$$\sum_{e\in E} \left|\mu_e^X - \mu_e^Y\right| \leq \frac{|X \bigoplus Y|}{2}.$$

2. For any $S : E \to 2^E$, consider the random configuration $\sigma \subseteq E$ generated by including each edge $e \in E$ independently with probability $\mu_e^{S(e)}$. Then, we have

$$\mathbf{Pr}\left[\sigma \notin \mathcal{G}\right] \leq \eta.$$

Condition 1 describes a "coupling with stationary" style criterion, akin to the one introduced in [25]. This criterion significantly relaxes the Dobrushin condition used in [36, 37] for ensuring the efficiency of parallel simulation, such as Algorithm 2. Rather than requiring distance decay between all pairs of configurations, it only enforces distance decay in expectation between good configurations $X, Y \in \mathcal{G}$ within the one-step optimal coupling of Glauber dynamics, while ensuring that these good configurations appear frequently throughout the dynamics.

We now state the main theorem of this section.

▶ **Theorem 8.** *If Condition 1 holds with parameter $\eta$, then for any $T^{\mathrm{GD}}, T^{\mathrm{PA}} \geq 1$, the output $X^{\boldsymbol{ALG}}$ of Algorithm 2 and the Glauber dynamics $(X_t)_{0\leq t\leq T^{\mathrm{GD}}}$ defined in (3), with initial state $X_0 \subseteq E$ generated as in Line 1 of Algorithm 2, satisfy the following bound on the total variation distance:*

$$d_{\mathrm{TV}}\left(X^{\boldsymbol{ALG}}, X_{T^{\mathrm{GD}}}\right) \leq T^{\mathrm{GD}} \cdot \left(2^{-(T^{\mathrm{PA}}-1)} + 4\eta\right).$$

▶ Remark 9 (generality of the criterion). Note that Algorithm 2, Condition 1, and Theorem 8 are stated abstractly for Glauber dynamics applied to a general joint distribution $\mu$ over variables with a Boolean domain. Therefore, these results are applicable to any such distribution.

The following theorem is an application of Theorem 8 to low-temperature random cluster models, establishing that Algorithm 2 is both accurate and efficient in this regime.

▶ **Theorem 10.** *Let $G = (V, E)$, $\boldsymbol{p}, \boldsymbol{\lambda}$, and $\epsilon$ be the input to Algorithm 2. Assume that $n = |V| \geq 3$, $\epsilon \leq \frac{1}{2}$, and the following condition holds:*

$$(1 - p_{\min}) \log n \leq \min \left\{ \mathrm{e}^{-40} \exp\left( -\frac{5 \log(1/\epsilon)}{\log n} \right), \frac{1 - \lambda_{\max}}{27} \right\}.$$

*Then, for any $T^{\mathrm{GD}} \geq \lceil 2m(\log m + \log(4/\epsilon)) \rceil$ and $T^{\mathrm{PA}} \geq \lceil 3 \log\left(4T^{\mathrm{GD}}/\epsilon\right) \rceil$, the output $X^{\boldsymbol{ALG}}$ of Algorithm 2 satisfies*

$$d_{\mathrm{TV}}\left(X^{\boldsymbol{ALG}}, \mu\right) \leq \epsilon. \tag{4}$$

Theorem 10 is a key step in establishing the accuracy and efficiency of Algorithm 1. Under the parameterization given in Table 1, each instance of the random cluster model passed to Algorithm 2 within Algorithm 1 falls within the low-temperature regime required by Theorem 10. The theorem holds because these low-temperature random cluster models satisfy Condition 1.

The proofs of Theorem 8 and Theorem 10 are deferred to Section A and Section B, respectively. Finally, the proof of Theorem 5 is completed in Section C.

## 5 Parallel Ising sampler via restricted Gaussian dynamics

We introduce a parallel sampler for the Ising model with contracting interaction norm. Our algorithm approximately implements the following *restricted Gaussian dynamics* (also known as the *proximal sampler*, described in Algorithm 3), the Markov chain induced by the *stochastic localization process* for Ising Gibbs sampling.

**Algorithm 3** Restricted Gaussian Dynamics.

**Input** : Interaction matrix $J \in \mathbb{R}^{V \times V}$, external fields $h \in \mathbb{R}^V$, and $\epsilon \in (0, 1)$.
**Output**: A random configuration $X \subseteq E$ satisfying $d_{\mathrm{TV}}\left(X, \mu_{J,h}^{Ising}\right) \leq \epsilon$.

1 initialize $x_0 \in \{\pm 1\}^V$ arbitrarily;
2 **forall** $i = 1, 2, \ldots, T^{\mathrm{RGD}}$ **do**
3  draw $y_i \sim \mathcal{N}(x_{i-1}, J^{-1})$;
4  draw $x_i = x \in \{\pm 1\}^V$ with prob. $\propto \mu_{J,h}^{Ising}(x) \cdot \exp\left(-\frac{1}{2}(y_i - x)^\top J (y_i - x)\right)$;
5 **end**
6 $X \leftarrow x_{T^{\mathrm{RGD}}}$;
7 **return** $X$;

The efficient implementation of Algorithm 3 follows from the following observations:

1. Suppose $0 \prec \frac{\eta}{2} I \preceq J \preceq (1 - \frac{\eta}{2}) I$. Given an error bound $0 < \epsilon_0 < 1$, the Gaussian sampling step (Line 3) of Algorithm 3 can be approximated within total variation error $\epsilon_0$. Specifically, there exists an algorithm that, given $x_{i-1}$ and $J^{-1}$, produces samples from a distribution $\pi$ satisfying $d_{\mathrm{TV}}(\pi, \mathcal{N}(x_{i-1}, J^{-1})) \leq \epsilon_0$. This approximation can be achieved in $O_\eta(\log^3(n/\epsilon_0^2))$ parallel time using $\widetilde{O}(n^3/\epsilon_0^2)$ processors.
2. The denoising step (Line 4) of Algorithm 3 involves sampling from a product distribution whose marginals can be computed in $O(\log n)$ parallel time. This step can be implemented faithfully with no error using $O(\log n)$ parallel time.
3. Assuming perfect simulations of Line 3 and Line 4 in Algorithm 3, the total variation distance between the output $X$ and the target distribution $\mu_{J,h}^{\mathrm{Ising}}$ does not exceed $\epsilon/2$ after $T^{\mathrm{RGD}} = O_\eta(\log(n/\epsilon))$ iterations of the outer loop in Algorithm 3.

The detailed analysis of Item 1, Item 2, and Item 3 is provided in Section D. Assuming these properties hold, we now proceed with the proof of Theorem 3.

**Proof of Theorem 3.** Let $X$ be the distribution of output generated by Algorithm 3. Under the assumption that the entire algorithm can be perfectly simulated, Item 3 ensures that the total variation distance between $X$ and the target distribution $\mu_{J,h}^{\text{Ising}}$ is at most $\epsilon/2$. Now, we implement Line 3 of Algorithm 3 using an approximate oracle provided by Item 1, setting the parameter $\epsilon_0 = \epsilon/2T^{\text{RGD}}$. Let $Y$ denote the output produced by our implementation of Algorithm 3. By a simple coupling argument, we obtain $d_{\text{TV}}(X,Y) \leq T^{\text{RGD}} \cdot \epsilon_0 = \epsilon/2$. Applying the triangle inequality, we have $d_{\text{TV}}(Y, \mu_{J,h}^{\text{Ising}}) \leq d_{\text{TV}}(X, \mu_{J,h}^{\text{Ising}}) + d_{\text{TV}}(X,Y) \leq \epsilon$. Thus, our implementation produces an approximate sample within $O_\eta(\log^4(n/\epsilon))$ parallel time using $\widetilde{O}_\eta(n^3/\epsilon^2)$ processors, as guaranteed by Item 1, Item 2, and Item 3. ◀

## References


1 Nima Anari, Callum Burgess, Kevin Tian, and Thuy-Duong Vuong. Quadratic speedups in parallel sampling from determinantal distributions. In *Proceedings of the 35th ACM Symposium on Parallelism in Algorithms and Architectures (SPAA 2023)*, pages 367–377. ACM, 2023. doi:10.1145/3558481.3591104.

2 Nima Anari, Sinho Chewi, and Thuy-Duong Vuong. Fast parallel sampling under isoperimetry. In *The Thirty Seventh Annual Conference on Learning Theory (ICML 2024)*, volume 247 of *Proceedings of Machine Learning Research*, pages 161–185. PMLR, 2024. URL: https://proceedings.mlr.press/v247/anari24a.html.

3 Nima Anari, Ruiquan Gao, and Aviad Rubinstein. Parallel sampling via counting. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing (STOC 2024)*, pages 537–548. ACM, 2024. doi:10.1145/3618260.3649744.

4 Nima Anari, Nathan Hu, Amin Saberi, and Aaron Schild. Sampling arborescences in parallel. In *12th Innovations in Theoretical Computer Science Conference (ITCS 2021)*, volume 185 of *LIPIcs*, pages 83:1–83:18. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2021. doi:10.4230/LIPIcs.ITCS.2021.83.

5 Nima Anari, Yizhi Huang, Tianyu Liu, Thuy-Duong Vuong, Brian Xu, and Katherine Yu. Parallel discrete sampling via continuous walks. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing (STOC 2023)*, pages 103–116. ACM, 2023. doi:10.1145/3564246.3585207.

6 Nima Anari, Vishesh Jain, Frederic Koehler, Huy Tuan Pham, and Thuy-Duong Vuong. Entropic independence I: Modified log-Sobolev inequalities for fractionally log-concave distributions and high-temperature Ising models. *arXiv preprint*, 2021. arXiv:2106.04105.

7 Nima Anari, Vishesh Jain, Frederic Koehler, Huy Tuan Pham, and Thuy-Duong Vuong. Entropic independence: optimal mixing of down-up random walks. In *Proceedings of the 54th Annual ACM Symposium on Theory of Computing (STOC 2022)*, pages 1418–1430. ACM, 2022. doi:10.1145/3519935.3520048.

8 Nima Anari, Frederic Koehler, and Thuy-Duong Vuong. Trickle-down in localization schemes and applications. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing (STOC 2024)*, pages 1094–1105. ACM, 2024. doi:10.1145/3618260.3649622.

9 Xiaoyu Chen, Zongchen Chen, Yitong Yin, and Xinyuan Zhang. Rapid mixing at the uniqueness threshold. In *Proceedings of the 57th Annual ACM Symposium on Theory of Computing (STOC 2025)*, pages 879–890. ACM, 2025. doi:10.1145/3717823.3718260.

10 Xiaoyu Chen, Weiming Feng, Yitong Yin, and Xinyuan Zhang. Rapid mixing of Glauber dynamics via spectral independence for all degrees. In *62nd IEEE Annual Symposium on Foundations of Computer Science (FOCS 2021)*, pages 137–148. IEEE, 2021. doi:10.1109/FOCS52979.2021.00022.

11 Xiaoyu Chen, Weiming Feng, Yitong Yin, and Xinyuan Zhang. Optimal mixing for two-state anti-ferromagnetic spin systems. In *63rd IEEE Annual Symposium on Foundations of Computer Science (FOCS 2022)*, pages 588–599. IEEE, 2022. doi:10.1109/FOCS54457.2022.00062.

12 Xiaoyu Chen and Xinyuan Zhang. A near-linear time sampler for the Ising model with external field. In *Proceedings of the 2023 ACM-SIAM Symposium on Discrete Algorithms (SODA 2023)*, pages 4478–4503. SIAM, 2023. doi:10.1137/1.9781611977554.CH170.

13 Yuansi Chen and Ronen Eldan. Localization schemes: A framework for proving mixing bounds for markov chains (extended abstract). In *63rd IEEE Annual Symposium on Foundations of Computer Science (FOCS 2022)*, pages 110–122. IEEE, 2022. doi:10.1109/FOCS54457.2022.00018.

14 Richard Cole. Parallel merge sort. *SIAM J. Comput.*, 17(4):770–785, 1988. doi:10.1137/0217049.

15 Martin Dyer, Alan Frieze, and Ravi Kannan. A random polynomial-time algorithm for approximating the volume of convex bodies. *Journal of the ACM (JACM)*, 38(1):1–17, 1991. doi:10.1145/102782.102783.

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

17 Ahmed El Alaoui and Andrea Montanari. An information-theoretic view of stochastic localization. *IEEE Trans. Inform. Theory*, 68(11):7423–7426, 2022. doi:10.1109/TIT.2022.3180298.

18 Ronen Eldan, Frederic Koehler, and Ofer Zeitouni. A spectral condition for spectral gap: fast mixing in high-temperature Ising models. *Probability Theory and Related Fields*, pages 1–17, 2021. doi:10.1007/s00440-021-01085-x.

19 Weiming Feng, Heng Guo, and Jiaheng Wang. Swendsen-wang dynamics for the ferromagnetic ising model with external fields. *Inf. Comput.*, 294:105066, 2023. doi:10.1016/J.IC.2023.105066.

20 Weiming Feng, Thomas P. Hayes, and Yitong Yin. Distributed Metropolis sampler with optimal parallelism. In *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms (SODA 2021)*, pages 2121–2140. SIAM, 2021. doi:10.1137/1.9781611976465.127.

21 Weiming Feng, Yuxin Sun, and Yitong Yin. What can be sampled locally? *Distributed Comput.*, 33(3-4):227–253, 2020. doi:10.1007/S00446-018-0332-8.

22 Weiming Feng, Nisheeth K. Vishnoi, and Yitong Yin. Dynamic sampling from graphical models. *SIAM J. Comput.*, 50(2):350–381, 2021. doi:10.1137/20M1315099.

23 Manuela Fischer and Mohsen Ghaffari. A simple parallel and distributed sampling technique: Local glauber dynamics. In *32nd International Symposium on Distributed Computing (DC 2018)*, volume 121 of *LIPIcs*, pages 26:1–26:11. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2018. doi:10.4230/LIPIcs.DISC.2018.26.

24 Andreas Galanis, Alkis Kalavasis, and Anthimos Vardis Kandiros. On sampling from Ising models with spectral constraints. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2024)*, volume 317 of *LIPIcs*, pages 70:1–70:14. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2024. doi:10.4230/LIPIcs.APPROX/RANDOM.2024.70.

25 Thomas P. Hayes and Eric Vigoda. Coupling with the stationary distribution and improved sampling for colorings and independent sets. *Ann. Appl. Probab.*, 16(3):1297–1318, 2006. doi:10.1214/105051606000000330.

26 Ernst Ising. *Beitrag zur theorie des ferro-und paramagnetismus*. PhD thesis, Grefe & Tiedemann Hamburg, 1924.

27 Joseph JéJé. An introduction to parallel algorithms. *Reading, MA: Addison-Wesley*, 10:133889, 1992.

28 Mark Jerrum and Alistair Sinclair. Approximating the permanent. *SIAM J. Comput.*, 18(6):1149–1178, 1989. doi:10.1137/0218077.

29 Mark Jerrum and Alistair Sinclair. Polynomial-time approximation algorithms for the Ising model. *SIAM J. Comput.*, 22(5):1087–1116, 1993. doi:10.1137/0222066.

30 Mark Jerrum, Alistair Sinclair, and Eric Vigoda. A polynomial-time approximation algorithm for the permanent of a matrix with nonnegative entries. *Journal of the ACM (JACM)*, 51(4):671–697, 2004. doi:10.1145/1008731.1008738.

31 Dmitriy Kunisky. Optimality of Glauber dynamics for general-purpose Ising model sampling and free energy approximation. In *Proceedings of the 2024 ACM-SIAM Symposium on Discrete Algorithms (SODA 2024)*, pages 5013–5028. SIAM, 2024. doi:10.1137/1.9781611977912.180.

32 Holden Lee. Parallelising glauber dynamics. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2024)*, volume 317 of *LIPIcs*, pages 49:1–49:24. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2024. doi:10.4230/LIPIcs.APPROX/RANDOM.2024.49.

33 Yin Tat Lee, Ruoqi Shen, and Kevin Tian. Structured logconcave sampling with a restricted Gaussian oracle. In *The Annual Conference on Learning Theory (COLT 2021)*, pages 2993–3050, 2021. URL: http://proceedings.mlr.press/v134/lee21a.html.

34 David A. Levin, Malwina J. Luczak, and Yuval Peres. Glauber dynamics for the mean-field Ising model: cut-off, critical power law, and metastability. *Probab. Theory Related Fields*, 146(1-2):223–265, 2010. doi:10.1007/s00440-008-0189-z.

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

36 Hongyang Liu and Yitong Yin. Simple parallel algorithms for single-site dynamics. In *STOC*, pages 1431–1444. ACM, 2022. doi:10.1145/3519935.3519999.

37 Hongyang Liu and Yitong Yin. Parallelize single-site dynamics up to dobrushin criterion. *J. ACM*, 72(1), January 2025. doi:10.1145/3708558.

38 Hongyang Liu, Yitong Yin, and Yiyao Zhang. Work-efficient parallel counting via sampling. *arXiv preprint arXiv:2408.09719*, 2024. doi:10.48550/arXiv.2408.09719.

39 Ketan Mulmuley, Umesh V Vazirani, and Vijay V Vazirani. Matching is as easy as matrix inversion. In *STOC*, pages 345–354. ACM, 1987. doi:10.1145/28395.383347.

40 Yossi Shiloach and Uzi Vishkin. An $O(\log n)$ parallel connectivity algorithm. *J. Algorithms*, 3(1):57–67, 1982. doi:10.1016/0196-6774(82)90008-6.

41 Shang-Hua Teng. Independent sets versus perfect matchings. *Theoret. Comput. Sci.*, 145(1-2):381–390, 1995. doi:10.1016/0304-3975(94)00289-U.


## A Parallel simulation of Glauber dynamics (Proof of Theorem 8)

We assume that Condition 1 holds for a given parameter $\eta \in [0,1]$.

Let $Y^\ell = (Y_i^\ell)_{1\le i\le T^{\mathrm{GD}}}$, where $Y_i^\ell$ is constructed in Line 9 of Algorithm 2, which indicates the result of the $i$-th update in the $\ell$-th iteration. Note that $Y^\ell$ for $1 \le \ell \le T^{\mathrm{PA}}$ are determined by the random choices $X'$ and $(e_i, \mathcal{R}_i)_{i=1}^{T^{\mathrm{GD}}}$ generated in Line 1 and Line 3 of Algorithm 2, respectively. The next lemma shows that $Y^\ell$ stabilizes fast with high probability.

▶ **Lemma 11.** *Given parameters $T^{\mathrm{PA}}, T^{\mathrm{GD}} \ge 1$, for any $1 \le \ell \le T^{\mathrm{PA}}$, it holds that*

$$\mathbf{Pr}\left[Y^\ell \neq Y^{\ell-1}\right] \le \left(2^{-(\ell-1)} + 4\eta\right) \cdot T^{\mathrm{GD}}. \tag{5}$$

**Proof.** For $1 \le i \le T^{\mathrm{GD}}$, define

$$\mathrm{Pred}_i := \{\mathrm{pred}_i(e) : e \in E\} \setminus \{0\}.$$

Intuitively, $\mathrm{Pred}_i$ is the set of updates that have impact on the result of the $i$-th update.

We claim that for $\ell \ge 2$ and $1 \le i \le T^{\mathrm{GD}}$,

$$\mathbf{Pr}\left[Y_i^\ell \neq Y_i^{\ell-1}\right] \le \left(\frac{1}{2m} \sum_{j \in \mathrm{Pred}_i} \mathbf{Pr}\left[Y_j^{\ell-1} \neq Y_j^{\ell-2}\right]\right) + 2\eta. \tag{6}$$

Assuming (6), the following can be proved by an induction on $1 \le \ell \le T^{\mathrm{PA}}$:

$$\mathbf{Pr}\left[Y_i^{\ell} \neq Y_i^{\ell-1}\right] \le 2^{-(\ell-1)} + 4\eta. \tag{7}$$

For the induction basis $\ell = 1$, (7) holds trivially since $2^{-(\ell-1)} + 4\eta > 1$. We now assume (7) holds for $\ell - 1 \ge 1$. By (6) and the induction hypothesis,

$$\begin{aligned}\mathbf{Pr}\left[Y_i^{\ell} \neq Y_i^{\ell-1}\right] &\le \frac{|\mathrm{Pred}_i|}{2m}\left(2^{-(\ell-2)} + 4\eta\right) + 2\eta \\ \text{(Since } |\mathrm{Pred}_i| \le m\text{)} \quad &\le 2^{-(\ell-1)} + 4\eta.\end{aligned}$$

Lemma 11 follows from (7) by taking the union bound over all $1 \le i \le T^{\mathrm{GD}}$.

It remains to prove the claim (6), which completes the proof of Lemma 11.

Let $\mathfrak{R}_{<i}$ denote the random choices used by the first $i-1$ updates, formally

$$\mathfrak{R}_{<i} := (e_j, \mathcal{R}_j)_{1 \le j < i}.$$

Recall the rules for $Y_i$ being updated in Line 13. The probability of $Y_i^{\ell} \neq Y_i^{\ell-1}$ conditioning on $\mathfrak{R}_{<i}$ and $e_i$ is given by

$$\mathbf{Pr}\left[Y_i^{\ell} \neq Y_i^{\ell-1} \mid \mathfrak{R}_{<i}, e_i\right] = \left|\mu_{e_i}^{\sigma_i^{\ell}} - \mu_{e_i}^{\sigma_i^{\ell-1}}\right|, \tag{8}$$

where $\mu_e^S = \frac{\mu(S\cup\{e\})}{\mu(S\cup\{e\})+\mu(S\setminus\{e\})}$ is the marginal probability.

Let $\mathcal{G} \subseteq 2^E$ be the good event in Condition 1. When $\sigma_i^{\ell}, \sigma_i^{\ell-1} \in \mathcal{G}$, by Condition 1,

$$\mathbf{Pr}\left[Y_i^{\ell} \neq Y_i^{\ell-1} \mid \mathfrak{R}_{<i}\right] = \frac{1}{m}\sum_{e\in E}\left|\mu_e^{\sigma_i^{\ell}} - \mu_e^{\sigma_i^{\ell-1}}\right| \le \frac{|\sigma_i^{\ell} \bigoplus \sigma_i^{\ell-1}|}{2m}. \tag{9}$$

Recall the construction of $\sigma_i^{\ell}$ in Line 11. Configurations $\sigma_i^{\ell}$ and $\sigma_i^{\ell-1}$ differ at site $e \in E$ if and only if $\mathrm{pred}_i(e) > 1$ and $Y^{\ell-1}_{\mathrm{pred}_i(e)} \neq Y^{\ell-2}_{\mathrm{pred}_i(e)}$. Note that for different $e_1, e_2 \in E$ with $\mathrm{pred}_i(e_1), \mathrm{pred}_i(e_2) > 0$, it holds that $\mathrm{pred}_i(e_1) \neq \mathrm{pred}_i(e_2)$. Therefore, $|\sigma_i^{\ell} \bigoplus \sigma_i^{\ell-1}|$ can be bounded by

$$\left|\sigma_i^{\ell} \bigoplus \sigma_i^{\ell-1}\right| \le \sum_{j\in\mathrm{Pred}_i} I\left[Y_j^{\ell-1} \neq Y_j^{\ell-2}\right]. \tag{10}$$

Combining (9), (10), and by the law of total probability, we have

$$\mathbf{Pr}\left[Y_i^{\ell} \neq Y_i^{\ell-1}\right] \le \frac{1}{2m}\sum_{j\in\mathrm{Pred}_i}\mathbf{Pr}\left[Y_j^{\ell-1} \neq Y_j^{\ell-2}\right] + \mathbf{Pr}\left[(\sigma_i^{\ell} \notin \mathcal{G}) \vee (\sigma_i^{\ell-1} \notin \mathcal{G})\right]. \tag{11}$$

Now, fix any $1 \le t \le T^{\mathrm{PA}}$, $1 \le i \le T^{\mathrm{GD}}$. By the definition of $\sigma_i^t$ and the updating rule of $Y_i^t$,

$$\forall e \in E, \quad I[e \in \sigma_i^t] = \begin{cases} Y^{t-1}_{\mathrm{pred}_i(e)}, & t > 1 \wedge \mathrm{pred}_i(e) \ge 1, \\ I[e \in X'], & \text{otherwise.}\end{cases}$$

It is straightforward to verify that, in Algorithm 2, $Y^{t-1}_{\mathrm{pred}_i(e)}$ and $I[e \in X']$ are independently generated according to $\mu_e^{\sigma^{t-1}_{\mathrm{pred}_i(e)}}$ and $\mu_e^E$, respectively. We can define the function $S : E \to 2^E$ as

$$\forall e \in E, \quad S(e) := \begin{cases} \sigma^{t-1}_{\mathrm{pred}_i(e)}, & t > 1 \wedge \mathrm{pred}_i(e) \ge 1, \\ E, & \text{otherwise.}\end{cases}$$

Note that $\sigma_i^t$ is generated by the same rule as Item 2 of Condition 1 using this $S$. This, according to Item 2 of Condition 1, implies that $\mathbf{Pr}\left[\sigma_i^t \notin \mathcal{G}\right] \leq \eta$. Note that $i, t$ are fixed arbitrarily. By a union bound, we have

$$\mathbf{Pr}\left[(\sigma_i^\ell \notin \mathcal{G}) \vee (\sigma_i^{\ell-1} \notin \mathcal{G})\right] \leq \mathbf{Pr}\left[\sigma_i^\ell \notin \mathcal{G}\right] + \mathbf{Pr}\left[\sigma_i^{\ell-1} \notin \mathcal{G}\right] \leq 2\eta. \tag{12}$$

Therefore, (6) holds. This completes the proof of Lemma 11. ◀

Theorem 8 is a corollary of Lemma 11.

**Proof of Theorem 8.** We couple $X^{\mathbf{ALG}}$ and $X_{T^{\mathrm{GD}}}$ by using the same random choices $X'$ and $(e_i, \mathcal{R}_i)_{i=1}^{T^{\mathrm{GD}}}$ in the two processes. Let $(Z_i)_{i=0}^{T^{\mathrm{GD}}}$ be the sequence of configurations defined by

$$Z_i := \left\{e \in E \mid (j_i(e) = 0 \wedge e \in X') \vee (j_i(e) \geq 1 \wedge Y_{j_i(e)}^{T^{\mathrm{PA}}} = 1)\right\},$$

where $j_i(e) = \max\{j \geq 1 \mid j \leq i \wedge e_j = e\} \cup \{0\}$. Recall the rules for $Y_i$ being updated in Line 13, and the construction of $\sigma_i^\ell$ in Line 11. It holds that for all $1 \leq i \leq T^{\mathrm{GD}}$,

$$Z_i = \begin{cases} Z_{i-1} \cup \{e_i\} & \mathcal{R}_i < \frac{\mu(\sigma_i^{T^{\mathrm{PA}}} \cup \{e_i\})}{\mu(\sigma_i^{T^{\mathrm{PA}}} \cup \{e_i\}) + \mu(\sigma_i^{T^{\mathrm{PA}}} \setminus \{e_i\})}, \\ Z_{i-1} \setminus \{e_i\} & \textit{otherwise}. \end{cases} \tag{13}$$

Now assume that $Y^{T^{\mathrm{PA}}} = Y^{T^{\mathrm{PA}}-1}$. For all $1 \leq i \leq T^{\mathrm{GD}}$, $Z_{i-1}$ satisfies

$$\begin{aligned} Z_{i-1} &= \left\{e \in E \mid (j_{i-1}(e) = 0 \wedge e \in X') \vee (j_{i-1}(e) \geq 1 \wedge Y_{j_{i-1}(e)}^{T^{\mathrm{PA}}} = 1)\right\} \\ &= \left\{e \in E \mid (\mathrm{pred}_i(e) = 0 \wedge e \in X') \vee (\mathrm{pred}_i(e) \geq 1 \wedge Y_{\mathrm{pred}_i(e)}^{T^{\mathrm{PA}}} = 1)\right\} \\ &= \left\{e \in E \mid (\mathrm{pred}_i(e) = 0 \wedge e \in X') \vee (\mathrm{pred}_i(e) \geq 1 \wedge Y_{\mathrm{pred}_i(e)}^{T^{\mathrm{PA}}-1} = 1)\right\} = \sigma_i^{T^{\mathrm{PA}}}. \end{aligned}$$

Therefore, the $\sigma_i^{T^{\mathrm{PA}}}$ in (13) can be replaced with $Z_i$, and hence the sequence $(Z_i)_{i=0}^{T^{\mathrm{GD}}}$ is identical to $(X_i)_{i=0}^{T^{\mathrm{GD}}}$ because the two processes have the same transition rules (13) and (3) and initial state $Z_0 = X_0$. Hence,

$$\mathbf{Pr}\left[X^{\mathbf{ALG}} \neq X_{T^{\mathrm{GD}}}\right] \leq \mathbf{Pr}\left[Y^{T^{\mathrm{PA}}} \neq Y^{T^{\mathrm{PA}}-1}\right] \leq \left(2^{-(T^{\mathrm{PA}}-1)} + 4\eta\right) \cdot T^{\mathrm{GD}},$$

where the last inequality follows from Lemma 11. This concludes the proof. ◀

## B Low-temperature random clusters (Proof of Theorem 10)

Recall that $G = (V, E)$, $\boldsymbol{p}, \boldsymbol{\lambda}$, and $\epsilon$ are the input of Algorithm 2, where it holds that $n = |V| \geq 3$, $\epsilon \leq 1/2$, and

$$(1 - p_{\min}) \log n \leq \min\left\{\mathrm{e}^{-40} \exp\left(-\frac{5\log(1/\epsilon)}{\log n}\right), \frac{1 - \lambda_{\max}}{27}\right\}.$$

Furthermore, $T^{\mathrm{GD}} \geq \lceil 2m(\log m + \log(4/\epsilon))\rceil$ and $T^{\mathrm{PA}} \geq \lceil 3\log\left(4T^{\mathrm{GD}}/\epsilon\right)\rceil$ are parameters used in Algorithm 2. These assumptions will be used throughout.

First, under the above assumption, it is known that the Glauber dynamics for the distribution $\mu = \mu_{E,\boldsymbol{p},\boldsymbol{\lambda}}^{RC}$ of the random cluster model is rapidly mixing.

▶ **Lemma 12** ([12, Theorem 5.1]). *Let $(X_i)_{i=0}^{T^{\mathrm{GD}}}$ be the Glauber dynamics on the distribution $\mu$ with the initial state $X_0$ generated as in Line 1 of Algorithm 2, then it holds that*

$$d_{\mathrm{TV}}\left(X_{T^{\mathrm{GD}}}, \mu\right) \leq \epsilon/2.$$

▶ Remark 13. In [12], the initial state of the Glauber dynamics was set to $E$ instead of the random $X_0$. Though, the proof of [12, Theorem 5.1] holds so long as the initial state follows the law of a product distribution whose marginal probabilities are bounded from below by $1 - 3(1 - p_{\min}) \log n$. By the argument in [12, Lemma 5.5], this is satisfied by the random initial state $X_0$ in Lemma 12.

We are now ready to prove Theorem 10.

**Proof of Theorem 10.** We first prove Theorem 10 by assuming Condition 1 with parameter $\eta = \frac{\epsilon}{16T^{\mathrm{GD}}}$. By Theorem 8 and the assumption of parameters,

$$d_{\mathrm{TV}}\left(X^{\mathbf{ALG}}, X_{T^{\mathrm{GD}}}\right) \leq 2T^{\mathrm{GD}} \cdot 2^{-3\log(4T^{\mathrm{GD}}/\epsilon)} + 4T^{\mathrm{GD}} \cdot \frac{\epsilon}{16T^{\mathrm{GD}}} \leq \epsilon/2.$$

Therefore, by Lemma 12 and a triangle inequality, it holds that

$$d_{\mathrm{TV}}\left(X^{\mathbf{ALG}}, \mu\right) \leq d_{\mathrm{TV}}\left(X^{\mathbf{ALG}}, X_{T^{\mathrm{GD}}}\right) + d_{\mathrm{TV}}\left(X_{T^{\mathrm{GD}}}, \mu\right) \leq \epsilon.$$

The remaining proof is dedicated to verifying Condition 1 with parameter $\eta = \frac{\epsilon}{16T^{\mathrm{GD}}}$.

### Verifying Item 1 of Condition 1

let $\mathcal{C}$ be the family of vertex sets with large expansion:

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

where $E(S, V \setminus S)$ is the set of edges between $S$ and $V \setminus S$. We define the good event

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

We remark that this is the same good event used in [12, Section 5]. In has been established in [12, Eq. (14)] that

$$\sum_{e \in E \setminus \{f\}} \left|\mu_e^X - \mu_e^Y\right| = 0.$$

Note that we also have $\left|\mu_f^X - \mu_f^Y\right| = 0$ by definition. Therefore,

$$\sum_{e \in E} \left|\mu_e^X - \mu_e^Y\right| \leq \frac{1}{2}. \tag{14}$$

For $X, Y \in \mathcal{G}$ with $|X \oplus Y| > 1$, let $e_1, e_2, \cdots, e_i$ be edges in $X \setminus Y$ and let $f_1, f_2, \cdots, f_j$ be edges in $Y \setminus X$. Construct following configuration path $P := (P_0, P_1, \cdots, P_{i+j})$ from $X$ to $Y$:

- let $P_0 = X$;
- for $1 \leq t \leq j$, let $P_t = P_{t-1} \cup \{f_t\}$;
- for $j < t$, let $P_t = P_{t-1} \setminus \{e_{t-j}\}$.

Note that $|P_{t-1} \oplus P_t| = 1$ for all $1 \leq t \leq i + j$, and $P_t \in \mathcal{G}$ for all $0 \leq t \leq i + j$. Hence,

$$\sum_{e \in E} \left|\mu_e^X - \mu_e^Y\right| \leq \sum_{e \in E} \sum_{1 \leq t \leq i+j} \left|\mu_e^{P_{t-1}} - \mu_e^{P_t}\right| = \sum_{1 \leq t \leq i+j} \sum_{e \in E} \left|\mu_e^{P_{t-1}} - \mu_e^{P_t}\right| \leq \frac{|X \oplus Y|}{2},$$

where the last inequality follows from (14) and the fact that $i + j = |X \oplus Y|$.

### Verifying Item 2 of Condition 1

Fix $S : E \to 2^E$, the random configuration $\sigma$ is generated by including each $e \in E$ independently with probability $\mu_e^{S(e)}$.

▶ **Lemma 14** ([12, Lemma 5.6]). *For any $T \subseteq E$ and $e \in E$, it holds that*

$$\mu_e^T \geq 1 - 3(1 - p_{\min}) \log n.$$

For simplicity, let $K = (1 - p_{\min}) \log n$, and Lemma 14 actually implies a marginal lower bound for $\sigma$, that is for each $e \in E$, $\mathbf{Pr}[e \in \sigma] \geq 1 - 3K$.

Since each edge $e$ is added to $\sigma$ independently, by the definition of $\mathcal{G}$,

$$\mathbf{Pr}[\sigma \notin \mathcal{G}] \leq \sum_{S \in \mathcal{C}} (3K)^{|E(S, V \setminus S)|} \leq \sum_{S \in \mathcal{C}} (3K)^{|S| \log n} \leq \sum_{j=1}^{+\infty} n^j n^{j \log(3K)} \leq n^{\log(27K)}.$$

Recall that we choose

$$T^{\mathrm{GD}} = \lceil 2m(\log m + \log(2/\epsilon)) \rceil \leq 4n^2 \log\left(\frac{4n^2}{\epsilon}\right) \leq \frac{16n^4}{\epsilon} \leq \frac{n^{10}}{\epsilon},$$

where in the last inequality, we use the assumption that $n \geq 3$. By assumption, it holds that

$$\begin{aligned} n^{\log(27K)} &= \exp\left(\log n \cdot \log(27K)\right) \\ &\leq \exp\left(\log n \cdot \log\left(\mathrm{e}^{-40} \exp\left(-\frac{3\log(1/\epsilon)}{\log n}\right)\right)\right) \\ &= \exp\left(\log n \cdot \left(-40 - \frac{3\log(1/\epsilon)}{\log n}\right)\right) \\ &= \exp\left(-40 \log n - 3\log(1/\epsilon)\right) = \frac{\epsilon^3}{n^{40}} \leq \frac{\epsilon}{16 T^{\mathrm{GD}}}, \end{aligned}$$

where in the last inequality, we use $n \geq 3$. This finishes the verification. ◀

## C Analysis of field dynamics (Proof of Theorem 5)

We now prove Theorem 5. The Item 1 (efficiency) of Theorem 5 will be proved in Section C.1, and Item 2 (accuracy) of Theorem 5 will be proved in Section C.2.

### C.1 Efficiency

We first bound the efficiency of Algorithm 2, assuming the setting of parameters in Table 1.

▶ **Proposition 15.** *Algorithm 2 teriminates in $O(T^{\mathrm{PA}} \cdot \log m)$ parallel time steps on $O(m \cdot T^{\mathrm{GD}})$ machines.*

**Proof.** The parallel complexity of Algorithm 2 is dominated by Line 7 and Line 12.

Line 7 requires to compute $\mathrm{pred}_i(e) = \max\{j \geq 1 \mid j < i \wedge e_j = e\} \cup \{0\}$ for each $1 \leq i \leq T^{\mathrm{GD}}$ and $e \in E$ in parallel. For each edge $e \in E$, the update list $\mathcal{L}_e$ is defined as:

$$\mathcal{L}_e := \{j \geq 1 \mid e_j = e\} \cup \{0\} \tag{15}$$

To implement Line 7, for each $e \in E$, we first store a sorted update list $\mathcal{L}_e$ using parallel merge sort [14], which costs $O(\log T^{\mathrm{GD}})$ parallel time on $O(T^{\mathrm{GD}})$ machines. Then each predecessor $\mathrm{pred}_i(e)$ can be computed in $O(\log T^{\mathrm{GD}})$ parallel time on $O(1)$ machines by a binary search on sorted $\mathcal{L}_e$. The whole process costs $O(\log T^{\mathrm{GD}})$ parallel time on $O(m \cdot T^{\mathrm{GD}})$ machines in total.

In Line 12, the marginal distribution $\mu^{\sigma}_{e_i} = \frac{\mu(\sigma \cup \{e_i\})}{\mu(\sigma \cup \{e_i\}) + \mu(\sigma \setminus \{e_i\})} = \frac{w(\sigma \cup \{e_i\})}{w(\sigma \cup \{e_i\}) + w(\sigma \setminus \{e_i\})}$ can be computed by calculating the weights of $\sigma \cup \{e_i\}$ and $\sigma \setminus \{e_i\}$ (defined in (2)), whereas each weight can be computed by a connected component algorithm such as [40], which costs $O(\log m)$ parallel time on $O(n+m)$ machines. Thus, each iteration of Line 9 costs $O(\log m)$ parallel time on $O(m \cdot T^{\mathrm{GD}})$ machines, and these $O(m \cdot T^{\mathrm{GD}})$ machines can be re-used in the next iteration. In Line 1, the marginal distribution $\mu^{E}_{e}$ can be computed in the same method, which costs $O(\log m)$ parallel time on $O(n+m)$ machines for each edge $e \in E$.

For other costs, the preprocessing part (from Line 0 to Line 5) of Algorithm 2 can be computed in $O(1)$ parallel time on $O(T^{\mathrm{GD}})$ machines; and in Line 16, the result $X$ of Algorithm 2 can be computed in $O(1)$ parallel time on $O(m)$ machines, since $j_e$ is maximum of the sorted list $\mathcal{L}_e$.

Overall, Algorithm 2 runs in $O(T^{\mathrm{PA}} \cdot \log m)$ parallel time on $O(m \cdot T^{\mathrm{GD}})$ machines. ◀

**Proof of Item 1 of Theorem 5.** When $|V| \le N_0$, Algorithm 1 generates $X$ by brute force, which costs $2^{O(N_0^2)} = \exp\left(O_\delta\left(\frac{\log(1/\epsilon)}{\log n}\right)\right)$ in total work. When $|V| > N_0$, Algorithm 1 terminates within $T^{\mathrm{FD}}$ iterations, and in each iteration, it generates a random $S$ using $O(1)$ parallel time on $O(m)$ machines and calls Algorithm 2. By Proposition 15, we have that Algorithm 1 terminates within $O(T^{\mathrm{PA}} \cdot \log m) \cdot T^{\mathrm{FD}} = \left(\exp\left(\frac{\log(1/\epsilon)}{\log n}\right) \cdot \log n\right)^{O_\delta(1)}$ parallel time steps on $O(m \cdot T^{\mathrm{GD}}) = O_\delta(m^2 \log(n/\epsilon))$ machines. ◀

## C.2 Accuracy of sampling

We now prove Item 2 of Theorem 5, the accuracy of sampling of Algorithm 1, still assuming the setting of parameters in Table 1. A key step has already been provided in Theorem 10. To complete the proof, we are going to establish the followings.

▶ **Lemma 16.** *If for every $U \subseteq E$, the output $Y$ of* parallel-RC $\left((V,U), \boldsymbol{p}^\star, \boldsymbol{\lambda}, (2T^{\mathrm{FD}})^{-1}\epsilon\right)$ *always satisfies $d_{\mathrm{TV}}\left(Y, \mu^{\mathrm{RC}}_{U,\boldsymbol{p}^\star,\boldsymbol{\lambda}}\right) \le (2T^{\mathrm{FD}})^{-1}\epsilon$, then the output $X$ of Algorithm 1 satisfies $d_{\mathrm{TV}}\left(X, \mu^{\mathrm{RC}}_{E,\boldsymbol{p},\boldsymbol{\lambda}}\right) \le \epsilon$.*

▶ **Lemma 17.** *Assuming $n = |V| \ge N_0$, it holds that for every $U \subseteq E$, the output $Y$ of the subroutine* parallel-RC $\left((V,U), \boldsymbol{p}^\star, \boldsymbol{\lambda}, (2T^{\mathrm{FD}})^{-1}\epsilon\right)$ *satisfies*

$$d_{\mathrm{TV}}\left(Y, \mu^{\mathrm{RC}}_{U,\boldsymbol{p}^\star,\boldsymbol{\lambda}}\right) \le \frac{\epsilon}{2T^{\mathrm{FD}}}.$$

Item 2 of Theorem 5 follows directly by combining Lemma 16 and Lemma 17.

Lemma 16 follows from the rapid mixing of the field dynamics.

▶ **Lemma 18** ([12, Lemma 3.5])**.** *The mixing time of the field dynamics satisfies*

$$\forall \epsilon \in (0,1), \quad T_{\mathrm{mix}}(\epsilon) \le \left(\frac{\mathrm{e}}{\theta}\right)^{5(1-\lambda_{\max})^{-2}} \left(2\log n + \log\log\frac{2}{p_{\min}} + \log\frac{1}{2\epsilon^2}\right).$$

**Proof of Lemma 16.** Without loss of generality, we assume that $|V| > N_0$. For $0 \le t \le T^{\mathrm{FD}}$, let $X_t$ be the configuration $X$ after the $t$-th iteration of Algorithm 1, and let $Y_t$ be the configuration $Y$ after the $t$-th iteration of the field dynamics $P_\theta^{\mathrm{FD}}$ starting from the same initial state $X_0 = Y_0 = E$.

Let $\mu = \mu^{\mathrm{RC}}_{E,\boldsymbol{p},\boldsymbol{\lambda}}$ for short. By triangle inequality and the coupling lemma, it holds that

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

for any coupling $(X_t, Y_t)$ of the two processes $X_t$ and $Y_t$. And it is obvious that if $X_{T^{\mathrm{FD}}} \neq Y_{T^{\mathrm{FD}}}$, then there must exist $1 \le i \le T^{\mathrm{FD}}$ such that $X_i \neq Y_i$ but $X_j = Y_j$ for all $0 \le j < i$. Hence, it holds that

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

For any $1 \le i \le T^{\mathrm{FD}}$, conditioned on $X_{i-1} = Y_{i-1}$, we construct a coupling of $X_i$ and $Y_i$:

1. generate a random $S \subseteq E$ by adding each $e \in E$ independently with probability $\theta$;
2. generate $(X_i, Y_i)$ according to the optimal coupling of $\widehat{\mu}^{\mathrm{RC}}_{U,\boldsymbol{p}^\star,\boldsymbol{\lambda}}$ and $\mu^{\mathrm{RC}}_{U,\boldsymbol{p}^\star,\boldsymbol{\lambda}}$, where $U = S \cup X_{i-1}$ and $\widehat{\mu}^{\mathrm{RC}}_{U,\boldsymbol{p}^\star,\boldsymbol{\lambda}}$ is the distribution of the output of $\mathsf{parallel\text{-}RC}\left((V,U), \boldsymbol{p}^\star, \boldsymbol{\lambda}, (2T^{\mathrm{FD}})^{-1}\epsilon\right)$.

By the coupling lemma and the assumption of Lemma 16, it holds that

$$\mathbf{Pr}\left[X_i \neq Y_i \mid X_{i-1} = Y_{i-1}\right] \le \max_{U \subseteq E} d_{\mathrm{TV}}\left(\widehat{\mu}^{\mathrm{RC}}_{U,\boldsymbol{p}^\star,\boldsymbol{\lambda}}, \mu^{\mathrm{RC}}_{U,\boldsymbol{p}^\star,\boldsymbol{\lambda}}\right) \le \frac{\epsilon}{2T^{\mathrm{FD}}}.$$

By Lemma 18 and our choice of $T^{\mathrm{FD}}$, we have

$$d_{\mathrm{TV}}\left(X_{T^{\mathrm{FD}}}, \mu\right) \le T^{\mathrm{FD}} \frac{\epsilon}{2T^{\mathrm{FD}}} + \frac{\epsilon}{2} = \epsilon. \qquad \blacktriangleleft$$

Now, it only remains to prove Lemma 17.

**Proof of Lemma 17.** It is sufficient to verify the condition of Theorem 10 with error bound $\frac{\epsilon}{2T^{\mathrm{FD}}}$ on the new instance $\mu^{\mathrm{RC}}_{U,\boldsymbol{p}^\star,\boldsymbol{\lambda}}$:

$$(1 - p^\star_{\min}) \log n \le \min\left\{ \mathrm{e}^{-40} \exp\left(-\frac{5\log(2T^{\mathrm{FD}}/\epsilon)}{\log n}\right), \frac{1-\lambda_{\max}}{27} \right\}. \tag{16}$$

By our assumption $n \ge N_0 \ge \max\left\{4, \frac{3}{p_{\min}}, \sqrt{\frac{\log(2/\epsilon^2)}{\log n}}\right\}$, it holds that

$$\log T^{\mathrm{FD}} \le 6(1-\lambda_{\max})^{-2} \log\left(1/K\right) + 14(1-\lambda_{\max})^{-2} \log n, \tag{17}$$

where for convenience, we denote

$$K := \theta \cdot \frac{\log n}{p_{\min}} = \mathrm{e}^{-100} \exp\left(\frac{10\log(\epsilon/2)}{\log n}\right) \exp\left(-140(1-\lambda_{\max})^{-2}\right).$$

The first term in the right hand side of (16) can be bounded by

$$\begin{aligned}&\mathrm{e}^{-40} \exp\left(-\frac{5\log(2T^{\mathrm{FD}}/\epsilon)}{\log n}\right) = \mathrm{e}^{-40} \exp\left(-\frac{5\log T^{\mathrm{FD}}}{\log n} + \frac{5\log(\epsilon/2)}{\log n}\right) \\ (\text{by (17)}) \quad &\ge \mathrm{e}^{-50} \exp\left(\frac{5\log(\epsilon/2)}{\log n}\right) \exp\left(-70(1-\lambda_{\max})^{-2} + \frac{30(1-\lambda_{\max})^{-2}\log K}{\log n}\right) \\ &\ge \mathrm{e}^{-50} \exp\left(\frac{5\log(\epsilon/2)}{\log n}\right) \exp\left(-70(1-\lambda_{\max})^{-2}\right) K^{1/2},\end{aligned}$$

where the last inequality holds since $n \ge N_0 \ge \exp\left(60(1-\lambda_{\max})^{-2}\right)$.

Note that the $\boldsymbol{p}^\star = \frac{\boldsymbol{p}}{\boldsymbol{p}+\theta(1-\boldsymbol{p})}$ constructed in Algorithm 1 satisfies

$$p^\star_{\min} = \min_{e\in E} p^\star_e = \frac{p_{\min}}{p_{\min}+\theta(1-p_{\min})} \geq \frac{\log n}{\log n + K} \geq 1 - \frac{K}{\log n}.$$

Therefore, we have the following two bounds on $(1-p^\star_{\min})\log n$:

$$\begin{aligned}(1-p^\star_{\min})\log n \leq K &\leq \mathrm{e}^{-50}\exp\left(\frac{5\log(\epsilon/2)}{\log n}\right)\exp\left(-70(1-\lambda_{\max})^{-2}\right)K^{1/2}\\ &\leq \mathrm{e}^{-40}\exp\left(-\frac{5\log(2T^{\mathrm{FD}}/\epsilon)}{\log n}\right), \qquad (18)\end{aligned}$$

$$(1-p^\star_{\min})\log n \leq K \leq \exp\left(-140(1-\lambda_{\max})^{-2}\right) \leq \frac{(1-\lambda_{\max})^2}{140} \leq \frac{1-\lambda_{\max}}{27}. \qquad (19)$$

Combining (18) and (19), we conclude the proof of Lemma 17. ◀

## D Analysis of restricted Gaussian dynamics

In this section, we verify Item 1, Item 2, and Item 3 stated in Section 5, completing the overall proof.

Item 1 requires an efficient method for approximately sampling in parallel from the Gaussian distribution $\mathcal{N}(x_{i-1}, J^{-1})$. The Langevin process is a well-established technique for Gaussian sampling, and recent work has introduced a parallelization for this approach.

▶ **Theorem 19** ([2, Corollary 14]). *Let $\alpha, \beta > 0$ be constants. Define $\kappa := \beta/\alpha$. Let $\pi = \exp(-V)$ be a distribution over $\mathbb{R}^d$ such that $V : \mathbb{R}^d \to \mathbb{R}$ satisfies*

$$0 \prec \alpha I \preceq \nabla^2 V(x) \preceq \beta I$$

*for all $x \in \mathbb{R}^d$. Given the minimizer $x^\star$ of $V$ and an error bound $\epsilon_0$, there exists a parallel sampling algorithm $\mathcal{A}$ that generates $Y$ such that $d_{\mathrm{TV}}(Y, \pi) \leq \epsilon_0$ by using $O(\kappa \log \kappa \log^2(d/\epsilon_0^2))$ parallel rounds and at most $7\max\left\{\frac{\kappa d}{\epsilon_0^2}, \kappa^2\right\}$-gradient evaluations per round.*

Simulating Line 3 requires sampling from $\mathcal{N}(0, J^{-1})$. By Theorem 19, we define the potential function $V(x) = \frac{1}{2}x^\top Jx$. Under the constraint $0 \prec \frac{\eta}{2}I \preceq J \preceq (1-\frac{\eta}{2})I$, we set $\kappa = 2/\eta$ and $d = n$. Evaluating the gradient $\nabla V(x)$ takes $O(\log n)$ parallel rounds on $O(n^2)$ machines. Thus, the total complexity of the parallelized Langevin algorithm is $O(\kappa \log \kappa \log^2(d/\epsilon_0^2))) \cdot O(\log n) = O_\eta(\log^3(n/\epsilon))$ and each parallel round requires $7\max\left\{\frac{\kappa d}{\epsilon_0^2}, \kappa^2\right\} \cdot O(n^2) = \widetilde{O}_\eta(n^3/\epsilon_0^2)$ machines. Thus, Item 1 is verified.

Verifying Item 2 follows directly by expanding the distribution formula in Line 4. Since the interaction matrix $J$ is symmetric, we have

$$\begin{aligned}\mu^{\mathrm{Ising}}_{J,h}(x)\cdot\exp\left(-\frac{1}{2}(y_i-x)^\top J(y_i-x)\right) &= \exp\left(\frac{1}{2}x^\top Jx + h^\intercal\sigma - \frac{1}{2}(y_i-x)^\top J(y_i-x)\right)\\ &= \exp\left(\langle y_i^\top J + h^\top, x\rangle - \frac{1}{2}y_i^\top Jy_i\right)\end{aligned}$$

It follows that each dimension in the above distribution is independent. The cost of computing $y_i^\top J + h^\top$ is $O(\log n)$ parallel rounds on $O(n^2)$ machines.

Verifying Item 3 follows from the rapid mixing of the restricted Gaussian dynamics, which has been established recently.

▶ **Theorem 20** ([13, 9]). *Let $\delta \in (0,1)$ be a constant, and let $P$ be the transition matrix of the restricted Gaussian dynamics. If $\|J\|_2 \le 1-\delta$, then the restricted Gaussian dynamics exhibits entropy decay with rate $\delta$, i.e., for all $f : \Omega(\mu) \to \mathbb{R}_{\ge 0}$,*

$$\mathrm{Ent}_\mu [Pf] \le (1-\delta)\mathrm{Ent}_\mu [f] .$$

*Consequently, by Pinsker's inequality, we obtain*

$$d_{\mathrm{TV}}(X,\mu) \le \sqrt{d_{\mathrm{KL}}(X,\mu)/2} \le \sqrt{\exp(-\delta T^{\mathrm{RGD}}) \log(1/\mu(x_0))},$$

*where $X$ is the output of Algorithm 3 and $x_0$ is the initial state.*

Since $0 \prec \frac{\eta}{2} I \preceq J \preceq (1-\frac{\eta}{2})I$, it follows that for all $x, y \in \{\pm 1\}^n$,

$$\frac{\mu(x)}{\mu(y)} \le \exp\left(\frac{2}{\eta}\right).$$

Thus, for any $x_0 \in \Omega(\mu)$, we have $1/\mu(x_0) \le 2^n \cdot \exp(2/\eta)$. Applying Theorem 20, we ensure $d_{\mathrm{TV}}(X,\mu) \le \epsilon/2$ by choosing

$$T^{\mathrm{RGD}} = \frac{2}{\eta}\left(\log\frac{4}{\epsilon^2} + \log\left(n\log 2 + \frac{2}{\eta}\right)\right) = O_\eta(\log(n/\epsilon)) .$$

Thus, Item 3 is verified.