Perfect sampling of q-spin systems on $\mathbb{Z}^2$ via weak spatial mixing

Konrad Anand and Mark Jerrum

Abstract. We present a perfect marginal sampler of the unique Gibbs measure of a spin system on $\mathbb{Z}^2$. The algorithm is an adaptation of a previous "lazy depth-first" approach by the authors, but relaxes the requirement of strong spatial mixing to weak. Our result is a step towards methods of efficient sampling using only weak spatial mixing.

The work exploits a classical result in statistical physics relating weak spatial mixing on $\mathbb{Z}^2$ to strong spatial mixing on squares. When the spin system exhibits weak spatial mixing, the runtime of our sampler is linear in the size of sample. Applications of note are the ferromagnetic Potts model at supercritical temperatures and the ferromagnetic Ising model with consistent non-zero external field at any non-zero temperature.

1. Introduction

This paper presents a new version of "lazy depth-first sampling", as presented in [2], which extends the efficacy of the algorithm beyond the domain of the strong spatial mixing.

Originally developed in the study of statistical mechanics, spin systems are now also an active area of research in probability, machine learning, and theoretical computer science. Different communities study the same basic concept using different terminology, an example being Constraint Satisfaction Problems (CSPs), which originated in artificial intelligence. In brief, a spin system is defined by a graph G, a finite (in our case) set of spins, and a local "specification". A configuration of the system is an assignment of spins to the vertices of the graph. The local specification defines a probability distribution on the configurations of the system. Although the specification is local, this Gibbs distribution can exhibit long-range effects. A primary algorithmic goal in this area is to sample configurations from the Gibbs distribution.

The question of sampling from spin systems has produced an array of results recently and over the past decades. Many approaches have been based on Markov

chains, in particular, using Glauber dynamics [25]. Another direction is to compute the sample probabilities, or quantities related to these sample probabilities, and then only as a final step actually produce a sample according to those probabilities. These approaches in some sense reduce the randomness of the algorithm. These results have come from decay of correlations [45], Taylor expansion of the partition function [3, 5, 36], and efficiently computing a Markov chain's history [16].

The strand of research which this paper falls into is perfect sampling. One motivation is obvious: a perfect sampler has no error, but another less obvious motivation is having a definite termination condition, in contrast to Markov chain simulation which has to be carried on for "sufficiently many" steps. This requires an a priori bound on the mixing time of a Markov chain, which must be analytically derived and may be much larger than actually necessary. In contrast, perfect samplers can respond to characteristics of a problem instance and produce results more quickly on favourable instances.

The first demonstration of perfect sampling came with the Coupling From The Past (CFTP) approach of Propp and Wilson [37]. There have been other approaches to the problem of efficient perfect sampling which move away from Markov chains, such as the randomness recycler of Fill and Huber [19], partial rejection sampling of Guo, Jerrum, and Liu [27], and the Bayes filter of Feng, Guo, and Yin [18]. The original lazy depth-first sampling [2] and the adapted version introduced in this paper fall in a similar direction.

Another benefit of lazy depth-first sampling is its applicability to infinite spin systems. Spin systems are defined in a local sense, but under certain conditions, which we assume to hold in this paper, the local definition of the spin system extends to a unique Gibbs measure. Of course, we cannot sample a configuration on an infinite graph, but a lazy depth-first sampler is intrinsically a marginal sampler, so we can sample finite windows into it.

Van den Berg and Steif [44] pioneered perfect sampling from infinite Gibbs measures. They showed that perfect sampling is possible in the case of the ferromagnetic Ising model on $\mathbb{Z}^2$ in the uniqueness regime, i.e., when there is a unique Gibbs measure, using CFTP. Spinka [41], in a wide-ranging investigation, has taken forward this approach. In contrast, lazy depth-first sampling departs from CFTP and takes inspiration from partial rejection sampling and related work.

In many approaches to sampling [12–14, 18, 39–41], the condition of strong spatial mixing is required to ensure efficiency. We relax this condition, in the case of 2-D systems, to weak spatial mixing, which is easier to establish. (Definitions of the mixing conditions will be given in Section 2.) An important question in sampling of spin systems is whether or not weak spatial mixing is enough to allow efficient sampling. In the case of two spins and antiferromagnetic interactions [12, 13], we know that the requirement of strong spatial mixing can be weakened. We take another step in

this direction, presenting an efficient sampling algorithm for a general class of q-spin systems on $\mathbb{Z}^2$ which only requires weak spatial mixing.

1.1. Lazy depth-first sampling

The strengths of lazy depth-first sampling are clear: it is a perfect sampler, it is a marginal sampler which can sample windows into infinite spin systems, and it is efficient (linear time) on any spin system which exhibits strong spatial mixing with the added condition of subexponential growth of the underlying graph G. One can also view it as a sublinear marginal sampler in terms of the size of the whole graph. This is a particularly nice benefit of the algorithm, as most approaches have run-time dependent on the size of the graph, rather than the size of the window.

Lazy depth-first sampling is a simple, recursive algorithm called on a single vertex, $v \in G$. Briefly, we try to guess the spin at a vertex v without looking at spins that live on a boundary around v by computing as much of v's distribution as possible. When we cannot guess, the algorithm recursively determines the spins on a sphere of suitable radius before returning to assign a spin to v. As it runs, the algorithm builds a branching process of recursive calls whose expected size is finite with probability 1 when strong spatial mixing holds.

Relevant conditions for efficiency of lazy depth-first sampling (and many other algorithms) are weak and strong spatial mixing. Informally, weak spatial mixing requires that the marginal distribution of a set of vertices changes little if we change faraway spins and nearby spins are unconstrained. Strong spatial mixing requires that the marginal distribution changes little if we change far away spins even when some local spins are fixed. (We sometimes call the arrangement of fixed spins a "partial configuration".)

In our first presentation of lazy depth-first sampling to spin systems, our sampling procedure did not take into account the precise structure of the underlying graph and relied on strong spatial mixing and subexponential growth to provide a coarse bound to ensure efficiency: a linear time marginal sampler. Our adaptation exploits the 2-D nature of the underlying lattice to carefully choose our sampling patterns and push past strong spatial mixing and require only weak spatial mixing.

The inspiration for the adaption is the classical physics result by Martinelli, Olivieri and Schonnman showing that, for spin systems on $\mathbb{Z}^2$, weak spatial mixing implies strong spatial mixing on squares [33]. In standard strong spatial mixing, the partial configurations are freely arranged, while in strong spatial mixing on squares the partial configurations are restricted to live on a square around the vertices whose change in marginals we are bounding. We use similar arguments to show the efficiency of this adaptation.

This adaptation runs on $\mathbb{Z}^2$ and takes care to live on a mesh (subgraph of $\mathbb{Z}^2$ induced by regularly spaced horizontal and vertical paths) so that all our partial configurations "mostly" live on a square around the vertex we are currently sampling. By doing so, we bypass pathological partial configurations and ensure efficiency of the algorithm with only weak spatial mixing. The algorithm also holds for q-spin systems, rather than only two-spin systems, and it is worth remarking that one can show efficiency of the algorithm on a variety of 2-D lattices—we consider only $\mathbb{Z}^2$ for simplicity.

It is worth remarking that the framework of lazy sampling is not restricted to spin systems. Indeed, two recent papers by He, Wang, and Yin and He, Wu, and Yang have applied these principles to the sampling uniform solutions to constraint satisfaction problems [28, 29]. Their results show the potential in applications of lazy sampling beyond spin systems.

In Section 2, we formally define a spin system in terms of an underlying graph (in this case $\mathbb{Z}^2$), a number of spins q, a "field" b, and an interaction matrix A. A technical condition we require is that the interaction matrix A is "soft". This will be satisfied for all the standard statistical physics models at non-zero temperature, and for some models with hard interactions, such as the hard-core gas model. See Section 2 for all relevant definitions. Of course, we cannot sample a configuration on the whole of $\mathbb{Z}^2$, so our aim is to sample from the marginal distribution on some finite subset $S \subset \mathbb{Z}^2$. Our result, in simplified form, is the following theorem.

Theorem 1.1. *Let $\mathcal{S} = (\mathbb{Z}^2, q, b, A)$ be a q-spin system on $\mathbb{Z}^2$ which exhibits weak spatial mixing, and let $S \subset \mathbb{Z}^2$ be a finite subset of vertices. Then,* LAZY *(Algorithm 1) is a perfect sampler for $\mathcal{S}$, which produces a sample from the marginal distribution on S in expected time linear in $|S|$.*

Theorem 1.1 is split into two parts later on: Theorem 6.1 covers the correctness and Theorem 5.1 covers the efficiency. Please refer to those results for details, such as a precise analysis of the runtime.

1.2. Applications

We focus on two new applications for perfect sampling of spin systems on $\mathbb{Z}^2$: The ferromagnetic Potts model at supercritical temperature and the ferromagnetic Ising model with constant external field at all temperatures. Each of these results leverages previous work on these models.

Sampling from the ferromagnetic Potts model is an active area of research. There are past approximate samplers at subcritical temperatures for general graphs, e.g., Bordewich, Greenhill, and Patel's work on the mixing of Glauber dynamics for maximum degree Δ graphs [9], approximate samplers specifically on $\mathbb{Z}^2$ [8, 22, 23, 43],

and approximate samplers for supercritical temperatures [4, 30, 31]. There have been recent results, including a sampler of the ferromagnetic Potts model at all temperatures for large enough q [10], new results on rapid mixing of the Swendsen–Wang dynamics [42] at low temperatures, relating it to edge-percolation, and the rapid mixing of Glauber dynamics at all temperatures, using careful initialisation [21].

We present the first perfect sampler for the ferromagnetic Potts model on $\mathbb{Z}^2$ at supercritical temperatures that works in constant time per vertex. To guarantee the success and efficiency of the Algorithm 1, we rely on the celebrated work of Beffara and Duminil–Copin [6] in conjunction with the work of Alexander [1].

A major early sampling step for sampling the ferromagnetic Ising model with external field was Mossel and Sly's work on the rapid mixing of the Glauber dynamics provided that the temperature is bounded by a function of the maximum degree of the graph [34]. More recent developments establish rapid mixing of the Swendsen–Wang dynamics on stochastic partitioned graphs [35] as well as a CFTP algorithm [17] and a near linear time field dynamics algorithm [11] for sampling on bounded degree graphs with the external field bounded away from 1.

We present a perfect marginal sampler for the ferromagnetic Ising model on $\mathbb{Z}^2$ with consistent external field that works in constant time per vertex with external field bounded away from 1. While this algorithm does not operate on general bounded degree graphs, it is perfect on an infinite graph and is exactly linear in the size of the window. The efficiency of the algorithm comes from the weak spatial mixing condition established by Schonmann and Schlosman [38].

2. Spin systems

We define spin systems on finite graphs; we will treat their extension to infinite graphs later.

Given a finite graph $G = (V, E)$ and $q \in \mathbb{N}$, we work with the set of configurations $\Omega_V = [q]^V$, where $[q] = \{1, 2, \ldots, q\}$. When there is no room for confusion, we will write Ω instead of Ω_V. Naturally, when $W \subset V$, we write $\Omega_W = [q]^W$ for the partial configurations restricted to W.

For $\sigma \in \Omega$ and $W \subset V$, we denote the restriction of σ to W by σ_W. For simplicity, when $W = \{v\}$, we will write σ_v instead of $\sigma_{\{v\}}$. In this same situation, we denote the spin at v by i and write $\sigma_v = i$ rather than $\sigma_v(v) = i$. For two configurations τ on Λ_1 and σ on Λ_2 which agree on $\Lambda_1 \cap \Lambda_2$, we denote by $\tau \oplus \sigma$ the configuration on $\Lambda_1 \cup \Lambda_2$ which agrees with both τ and σ.

A spin system is defined by a "field" $b : [q] \to \mathbb{R}$ and a symmetric matrix $A : [q] \times [q] \to \mathbb{R}$ of interaction weights. Given a graph G (finite at this stage) the Gibbs

distribution for G gives to each configuration σ the weight

$$\prod_{v \in V} b(\sigma_v) \prod_{(u,v) \in E} A(\sigma_u, \sigma_v).$$

Define the partition function $Z(G)$ to be the sum of all possible weights:

$$Z(G) := \sum_{\sigma \in \Omega} \prod_{v \in V} b(\sigma_v) \prod_{(u,v) \in E} A(\sigma_u, \sigma_v).$$

Then, we define the Gibbs measure to be the measure $\mu : \Omega \to [0, 1]$ where, provided $Z(G) \neq 0$,

$$\mu(\sigma) = \frac{\prod_{v \in V} b(\sigma_v) \prod_{(u,v) \in E} A(\sigma_u, \sigma_v)}{Z(G)}. \tag{2.1}$$

Frequently, in this paper, we will want to look at the marginal distribution induced on a set of vertices by fixing the configuration on the remainder of the vertices. For any W, define the marginal distribution of μ on W by

$$\mu_W(\sigma) = \sum_{\sigma' \in \Omega : \sigma'_W = \sigma} \mu(\sigma')$$

for all $\sigma \in \Omega_W$. Also, for any $\Lambda \subset V$ and $\tau \in \Omega_\Lambda$, define the marginal distribution with boundary condition (Λ, τ) by

$$\mu_W^{(\Lambda, \tau)}(\sigma) := \frac{1}{\mu_\Lambda(\tau)} \sum_{\substack{\sigma' \in \Omega : \\ \sigma'_\Lambda = \tau, \, \sigma'_W = \sigma}} \mu(\sigma'), \tag{2.2}$$

assuming $\mu_\Lambda(\tau) \neq 0$. We abbreviate $\mu_W^{(\Lambda, \tau)}$ to μ_W^τ when the set Λ is clear from the context.

With this definition of the Gibbs measure, we have a fixed probability space determined by a few parameters. We call this a spin system $\mathcal{S} = (G, q, b, A)$. In this paper, we deal exclusively with spin systems in which the underlying graph G is the 2-D grid $\mathbb{Z}^2$, i.e., the graph with vertex and edge sets

$$V_{\mathbb{Z}^2} = \mathbb{Z}^2,$$
$$E_{\mathbb{Z}^2} = \big\{ (u, v) \in V_{\mathbb{Z}^2} : \|u - v\|_2 = 1 \big\},$$

where $\| \cdot \|_2$ is the Euclidean norm. The restriction to the 2-D grid, in particular, is not strictly necessary: with mild modifications to the sampling patterns and analysis the algorithm works on many planar lattices, but for clarity we have restricted our discussion to the most common planar lattice. However, we will see that the restriction to two dimensions is crucial.

Fortunately, when describing and analysing the proposed algorithm, we only encounter finite subgraphs of $\mathbb{Z}^2$. For these, the Gibbs distribution on configurations is defined by (2.1). As usual, the *Gibbs property* of μ is important to us, and is the reason we can concentrate on finite subgraphs. Suppose $W \subset V$ and that the configuration $\tau \in \Omega_{V \setminus W}$ is *feasible*, i.e., $\mu_{V \setminus W}(\tau) > 0$. Define the *boundary* of W by

$$\partial W = \{v \in V \setminus W : \{u, v\} \in E \text{ for some } u \in W\},$$

and let $\tau' \in \Omega_{\partial W}$ be the restriction of τ to ∂W. Then,

$$\mu_W^{(V \setminus W, \tau)} = \mu_W^{(\partial W, \tau')}; \tag{2.3}$$

in other words, conditioning on the complement of W is equivalent to conditioning on its boundary.

As we are claiming that our algorithm samples from the Gibbs measure on an infinite graph, namely, the 2-D grid, we need to say a little about what that means. A full treatment of 'infinite volume' measures is given by Friedli and Velenik [20]: see Section 6. The bare bones are as follows.

If we drop the assumption that V is finite, identity (2.3) still enables us to calculate the marginal distribution $\mu_W^{(V \setminus W, \tau)}$ provided W is finite. The quantities

$$\left\{ \mu_W^{(V \setminus W, \tau)} : W \subset V, W \text{ finite, and } \tau \in \Omega_{V \setminus W} \right\}$$

provide a *specification* of the infinite volume measure [20, Definition 6.11]. (The various marginal distributions have to satisfy a consistency condition, but that is automatic here.) The infinite volume measures we are interested in are ones that are consistent with the specification [20, Definition 6.12]. There may be no, one or many such measures. In our algorithm, we only make use of marginal probabilities on finite subgraphs, which are known by (2.3). Therefore, a proof that our algorithm terminates with probability 1 can be interpreted as a proof that the infinite volume measure is unique, and the output of the algorithm can be interpreted as a construction of (the marginal distribution on a finite subset $U \subset V$ of) the infinite volume measure.

Finally, we say A is a *soft* interaction matrix if there exists $\alpha > 0$ and $i \in [q]$ such that for all $j \in [q]$, $A_{ij} > \alpha$. In this paper, we restrict ourselves to spin systems with soft interaction matrices. It is unnecessarily restrictive but straightforward to handle and enough for our applications.

2.1. Weak spatial mixing

Note that in the following d_G denotes graph distance.

For efficiency of our algorithm, we require weak spatial mixing, which intuitively tells us that, given sets $U \subset \mathbb{Z}^2$ and $W \subset \mathbb{Z}^2$, conditioning on σ_W has little effect

on the distribution of σ_U so long as W is far from U. In order to connect smoothly with previous work, we use essentially the same definition as Martinelli, Olivieri, and Schonmann [33, equation (1.11)]: in our case, we are specialised to nearest-neighbour interactions, so we restrict to $r = 1$.

Definition 2.1. Let S be a spin system and $f : \mathbb{N} \to \mathbb{R}^+$ a function. If for all $W \subset V$ with $V \setminus W$ finite, all $U \subset V \setminus W$, and all feasible $\tau^1, \tau^2 \in \Omega_W$ with $\tau^1 \neq \tau^2$,

$$d_{\text{TV}}\big(\mu_U^{\tau^1}, \mu_U^{\tau^2}\big) \leq \sum_{u \in U, w \in \partial(V \setminus W)} f(d_G(u, w)),$$

then we say that S exhibits *set weak spatial mixing* with rate f.

There are many variants of weak spatial mixing and it is critical, we use a strong enough definition. To obtain $O(n \log n)$ mixing time results for the random cluster model on $\mathbb{Z}^d$, Gheissari and Sinclair require that $U, V \setminus W$ be boxes [24]. In contrast, we need our sets to take any form. In Spinka's perfect sampling paper [41], rather than our sum, the bound is

$$d_{\text{TV}}\big(\mu_U^{\tau^1}, \mu_U^{\tau^2}\big) \leq |U| f(d_G(U, W));$$

however, in our case, this condition is not strong enough to ensure efficiency.

From this point on, weak spatial mixing will be with exponential rate, i.e., $f(x) = Ce^{-\gamma x}$.

3. Martinelli, Olivieri, and Schonmann

From now on, we will drop the subscript G in the graph distance d_G as we are always working on $\mathbb{Z}^2$.

In their 1994 paper [33], obtained a particular result on $\mathbb{Z}^2$ which showed that, for 2-D spin systems, weak spatial mixing implies strong spatial mixing for squares. Specifically, the result is as follows. Let $S \subset \mathbb{Z}^2$ be a square of side length L, that is, for some vertex v_c,

$$S = \{v \in \mathbb{Z}^2 : v = v_c + (a, b), 0 \leq a, b \leq L\}.$$

Theorem 3.1 ([33, Theorem 1.1]). *Let S be a spin system on $\mathbb{Z}^2$ which exhibits weak spatial mixing. Then, there exist constants $C, \gamma > 0$ such that for all $\tau^1, \tau^2 \in \Omega_{\partial S}$ and all $\Delta \subset S$*

$$d_{\text{TV}}\big(\mu_\Delta^{\tau^1}, \mu_\Delta^{\tau^2}\big) \leq Ce^{-\gamma \ell},$$

where $\ell = \min\{d(v, \Delta) : v \in \partial S, \tau^1(v) \neq \tau^2(v)\}$.

Here, $d(v, \Delta) = \min\{d(v, u) : u \in \Delta\}$. Martinelli et al. also note that this result is not specific to $\mathbb{Z}^2$—it can be extended to other regular 2-D graphs—however, there are counterexamples to show that the restriction to 2 dimensions is essential [32].

We will not need the full strength of Theorem 3.1 above for the purposes of our algorithm; we only need to adapt the intermediate Proposition 2.1 from the same paper [33] into our Proposition 3.2. We cannot directly apply their Proposition 2.1, because our algorithm will construct partial configurations which do not restrict themselves to squares around vertices.

Consider a square $S \subset \mathbb{Z}^2$ of side length $2L$ (thus made of 4 squares of side length L). To be precise, for some corner vertex $v_c \in \mathbb{Z}^2$,

$$
\begin{aligned}
S &= \{v \in \mathbb{Z}^2 : v = v_c + (a, b), 0 < a, b < 2L\}, \\
S_0 &= \{v \in \mathbb{Z}^2 : v = v_c + (a, b), 0 < a, b < L\}, \\
S_1 &= \{v \in \mathbb{Z}^2 : v = v_c + (L + a, b), 0 < a, b < L\}, \\
S_2 &= \{v \in \mathbb{Z}^2 : v = v_c + (a, L + b), 0 < a, b < L\}, \\
S_3 &= \{v \in \mathbb{Z}^2 : v = v_c + (L + a, L + b), 0 < a, b < L\}.
\end{aligned}
$$

The notation ∂S will denote the usual boundary of S, and we will also consider the "bisected" boundary of S,

$$
\boxplus S = \partial S_0 \cup \partial S_1 \cup \partial S_2 \cup \partial S_3.
$$

In analysing our algorithm, we will encounter boundary conditions that fix spins on the whole of ∂S and on a subset of $\boxplus S \setminus \partial S$. Such a boundary condition is specified by a partial configuration (T, τ), where

$$
\partial S \subseteq T \subseteq \boxplus S \quad \text{and} \quad \tau \in \Omega_T.
$$

Proposition 3.2. *Let $\mathcal{S}$ be a q-spin system on $\mathbb{Z}^2$ which exhibits weak spatial mixing with constants $C, \gamma > 0$. Then, there exist $L_0 \in \mathbb{N}$ and $\gamma_0 > 0$ such that the following holds: for all squares $S \subset \mathbb{Z}^2$ with side length $2L > 2L_0$, all $y \in \partial S$, all T with $\partial S \subseteq T \subseteq \boxplus S$, and all feasible partial configurations $\tau, \tau' \in \Omega_T$ that agree except at vertex y,*

$$
d_{\mathrm{TV}}(\mu_Q^\tau, \mu_Q^{\tau'}) \le e^{-\gamma_0 L^{1/4}},
$$

where

$$
Q := \{x \in S : \|x - y\|_\infty \ge L^{\frac{1}{2}}\}.
$$

Before proving the result, we run through a simple 1-d analogue of the proof. Recall that for two measures λ and μ on the same state space and random variables $X \sim \lambda$ and $Y \sim \mu$, if we can find a coupling (X, Y) of the random variables such that $\mathbb{P}[X \ne Y] \le \varepsilon$, then

$$
d_{\mathrm{TV}}(\mu, \lambda) \le \varepsilon.
$$

It follows that to bound the total variation distance between two measures, we only need to find a coupling which agrees to an appropriate extent. For the convenience of readers who are familiar with [33], we adopt some of the notation used there. Note that there is no particular significance in the quantities L, $L^{\frac{1}{2}}$, and $L^{\frac{1}{4}}$ beyond the fact that the first grows more quickly than the second, and the second more quickly than the third.

Consider a 2-spin system $\mathcal{S}$ on $S = \{-L, \ldots, L\} \subset \mathbb{Z}$, with $L^{\frac{1}{4}}$ integral, which exhibits weak spatial mixing with constants C, γ. Suppose we want to determine the influence of a fixed spin at 0 on the set

$$Q := \left\{ x \in S : |x| \geq L^{\frac{1}{2}} \right\}.$$

We can do so by building two configurations on S which we attempt to couple at each step; that is, we try to ensure that the two configurations are equal sufficiently far from 0. We denote the two configurations by σ^+ and σ^-, with boundary conditions

$$\sigma_{-L}^+ = \sigma_{-L}^- = +, \quad \sigma_L^+ = \sigma_L^- = +, \quad \sigma_0^+ = +, \quad \sigma_0^- = -.$$

We sample σ^+ and σ^- spin-by-spin, starting at the origin, trying to make them agree, while at the same time ensuring that both σ^+ and σ^- have the correct Gibbs distribution. If we have ever succeeded in coupling the two configurations then the Markov property ensures we can always maintain our coupling. Thus, we only need to analyse the case where we have not already coupled.

We will attempt to couple every $L^{\frac{1}{4}}$ spins. Let $D_i = \{-iL^{\frac{1}{4}}, iL^{\frac{1}{4}}\}$. Suppose that $\sigma_{D_{i-1}}^+ \neq \sigma_{D_{i-1}}^-$, i.e., we have not coupled at the previous step. Since $\mathcal{S}$ exhibits weak spatial mixing, we can couple $\sigma_{D_i}^+$ with $\sigma_{D_i}^-$ so that

$$\mathbb{P}\left[\sigma_{D_i}^+ \neq \sigma_{D_i}^-\right] < Ce^{-\gamma L^{1/4}}.$$

By iterating this procedure, it follows that at step $n = L^{\frac{1}{4}}$, the probability we have not coupled is bounded:

$$\mathbb{P}\left[\sigma_{D_n}^+ \neq \sigma_{D_n}^-\right] < C^n e^{-\gamma n L^{1/4}} = C^{L^{1/4}} e^{-\gamma L^{1/2}}.$$

By choosing L sufficiently large, we can ensure that coupling has taken place with probability at least $\frac{1}{2}$, in which case we can make σ^+ and σ^- agree on Q. The following proof for the 2-D situation is more involved, but is similar.

Proof of Proposition 3.2. Assume again that $L^{\frac{1}{4}}$ is integral. Taking our cue from the warm-up 1-d argument, we will build a sequence of frames $D_i \subset S$ around y, each $L^{\frac{1}{4}}$ further from y than the previous. We will then construct configurations σ, σ', starting with τ and τ', respectively, and with each of these frames we will attempt to
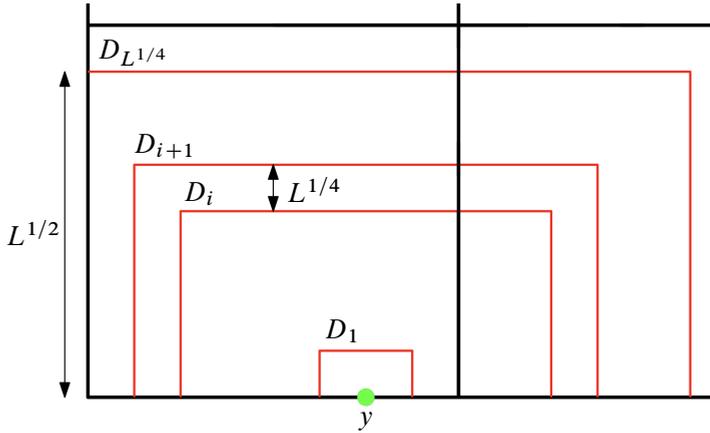
Figure 1. In black $\boxplus S$, the domain of τ and τ'. In green the vertex y on which τ and τ' differ. In red the D_i's, i.e., the $L^{\frac{1}{4}}$ frames on which we will iteratively attempt to couple σ and σ'. Disclaimer: this image is indicative, not to scale.

couple the two configurations. The probability of us coupling will give the bound on $d_{\mathrm{TV}}(\mu_Q^\tau, \mu_Q^{\tau'})$.

Let h be a sufficiently large number, to be chosen later. Define, for $i = 0, \ldots, L^{\frac{1}{4}}$,

$$Q_i = \{x \in S : \|x - y\|_\infty > iL^{\frac{1}{4}}\},$$
$$D_i = \{x \in S : \|x - y\|_\infty = iL^{\frac{1}{4}}\},$$
$$A_i = \{x \in D_i : \min_{z \in \boxplus S} \|x - z\|_\infty \le h\},$$
$$B_i = D_i \setminus A_i.$$

The D_i's will be our frames (see Figure 1). They are split into the A_i's, the pieces close to the boundary, and the B_i's, the pieces far from the boundary. The B_i's are the nice vertices which are far enough from $\boxplus S$ that weak spatial mixing applies. (The parameter h will be chosen so that is the case.) The A_i's are the unpleasant vertices, for which we have no nice trick, but which are few enough in number that we can deal with them quite crudely. We will attempt to couple these two sets separately. Note that, again, once we have matched the configurations on the ith frame, we can always extend the coupling to the whole of Q_i due to the Markov property and the identical boundaries of Q_i on both configurations. Thus, we will only discuss the case where we have failed at every previous step.

Consider the sets A_i and B_i. For a particular h, we may choose L_0 large enough so that $L_0^{\frac{1}{4}} > 2h$. Then, for all but at most one $i = i_0$, the A_i's will be made of (at most) four "feet" of size h which are the only pieces of the D_i's close to $\boxplus S$. We

will skip the attempt at coalescence for i_0 and we will still be able to succeed with probability arbitrarily close to one.

Now, we attempt to couple by first dealing with the B_i's and then the A_i's. For the B_i's, we will need the following technical lemma, which tells us that the influence on B_i of the fixed partial configuration is bounded, independently of h and L, so long as h and L are chosen large enough. This will give us a coupling of

$$\mu_{B_i}^{\tau \oplus \sigma_{D_{i-1}}} \quad \text{and} \quad \mu_{B_i}^{\tau' \oplus \sigma'_{D_{i-1}}},$$

which agrees with probability greater than $\frac{1}{2}$.

Lemma 3.3. *Under the conditions of Proposition* 3.2*, there are* h, L_0 *such that for all* $L > L_0$*, and for all* $i = 1, \ldots, L^{\frac{1}{4}}$ *with* $i \neq i_0$*,*

$$C \sum_{\substack{x \in B_i, \\ w \in (\boxplus S \cup D_{i-1})}} e^{-\gamma d(x,w)} \leq \frac{1}{2}.$$

Proof. We split the sum into two pieces. Recall that $d(\cdot, \cdot)$ is graph distance on $\mathbb{Z}^2$, so our distances are additive. Let

$$P = \left\{ (x, w) \in B_i \times (\boxplus S \cup D_{i-1}) : d(x, w) < L^{\frac{1}{4}} \right\}$$

and

$$P^c = B_i \times (\boxplus S \cup D_{i-1}) \setminus P.$$

Then,

$$\sum_{\substack{x \in B_i, \\ w \in (\boxplus S \cup D_{i-1})}} e^{-\gamma d(x,w)} = \sum_{(x,w) \in P} e^{-\gamma d(x,w)} + \sum_{(x,w) \in P^c} e^{-\gamma d(x,w)}.$$

Note that D_i intersects ∂S twice, and $\boxplus S \setminus \partial S$ at most once. So, P can be partitioned into at most three subsets, each local to some intersection vertex, z. Local to z, we can further partition the pairs (x, w) according to which side of z, x and w lie on. In this way, we partition P into at most eight (broken) L-shaped pieces $\{P_1, \ldots, P_8\}$, where two P_i's are local to each intersection point on ∂S and four are local to the intersection point on $\boxplus S \setminus \partial S$, if it exists. Let P_m be such an L-shaped region local to z. (Refer to Figure 2.) Then, we have

$$\sum_{(x,w) \in P_m} e^{-\gamma d(x,w)} = \sum_{(x,w) \in P_m} e^{-\gamma(d(x,z) + d(w,z))} \leq \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} e^{-\gamma(j+k+h)}$$

$$= \frac{e^{-\gamma h}}{(1 - e^{-\gamma})^2}.$$
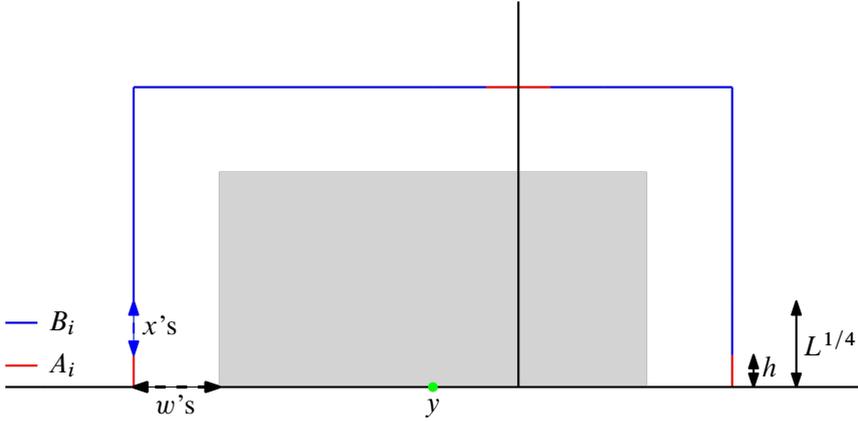
Figure 2. In black, the vertices in $\boxplus S$. In grey, Q^c_{i-1}, i.e., the vertices which have been fixed once we reach the ith frame. In red, A_i, the vertices of D_i which are too close to the boundary to leverage weak spatial mixing. In blue, B_i, the vertices in D_i which are far from $\boxplus S$. Marked by arrows, the $x's$ and w's which form P. The x's are the vertices in B_i that are within $L^{\frac{1}{4}}$ of $\boxplus S$ and the w's are the vertices in $\boxplus S$ which are within $L^{\frac{1}{4}}$ of D_i.

Here, the summations run over $j = d(x, z) - h$ and $k = d(w, z)$. Thus,

$$C \sum_{(x,w) \in P} e^{-\gamma d(x,w)} = C \sum_{m=1}^{8} \sum_{(x,w) \in P_m} e^{-\gamma d(x,w)} \leq \frac{8Ce^{-\gamma h}}{(1 - e^{-\gamma})^2}.$$

We can take h large enough to bound the above by $\frac{1}{4}$.

For P^c we note that every term in the sum has $d(x, w) \geq L^{\frac{1}{4}}$ and the number of terms is bounded by $13L^2$, so

$$C \sum_{(x,w) \in P^c} e^{-\gamma d(x,w)} < 13CL^2 e^{-\gamma L^{1/4}}.$$

Then, taking L_0 large enough, we can bound this term by $\frac{1}{4}$.

Now, with our choice of h and any $L > L_0$, we have that

$$C \sum_{\substack{x \in B_i, \\ w \in (\boxplus S \cup D_{i-1})}} e^{-\gamma d(x,w)} < \frac{1}{4} + \frac{1}{4} = \frac{1}{2}.$$

This completes the proof of Lemma 3.3. ∎

Now, we describe the iterated coupling attempts. Suppose at step i we have thus far failed to couple σ and σ' outside D_{i-1}.

For the B_i's, it is critical that, we have our definition of weak spatial mixing, since it gives us that for every $i = 1, \ldots, L^{\frac{1}{4}}$,

$$d_{\mathrm{TV}}\left(\mu_{B_i}^{\tau \oplus \sigma_{D_{i-1}}}, \mu_{B_i}^{\tau' \oplus \sigma'_{D_{i-1}}}\right) < C \sum_{\substack{x \in B_i, \\ w \in (\partial S \cup B_{i-1})}} e^{-\gamma d(x,w)}.$$

Then, taking h and L_0 from Lemma 3.3, it follows immediately that

$$d_{\mathrm{TV}}\left(\mu_{B_i}^{\tau}, \mu_{B_i}^{\tau'}\right) < \frac{1}{2}.$$

It follows that we can choose a coupling on B_i so that

$$\mathbb{P}\left[\sigma_{B_i} = \sigma'_{B_i}\right] \geq \frac{1}{2}.$$

For the A_i, we are too close to use weak spatial mixing, so we will use a cruder method, sampling σ_{A_i} and σ'_{A_i} independently. Due to our restriction to soft interaction matrices, each time we attempt to couple, we have for some constant $\alpha > 0$,

$$\mathbb{P}\left[\sigma_{A_i} = \sigma'_{A_i}\right] \geq \alpha^{|A_i|} = \alpha^{4h}.$$

Now, combining the two pieces, we find that on every iteration where we have not succeeded, we couple with probability

$$\mathbb{P}\left[\sigma_{D_i} = \sigma'_{D_i}\right] \geq \frac{1}{2}\alpha^{4h}.$$

It follows that

$$d_{\mathrm{TV}}\left(\mu_Q^{\tau}, \mu_Q^{\tau'}\right) \leq \mathbb{P}\left[\sigma_{D_{L^{1/4}}} = \sigma'_{D_{L^{1/4}}}\right] < \left(1 - \frac{1}{2}\alpha^{4h}\right)^{L^{1/4}} < e^{-\frac{1}{2}\alpha^{4h}L^{1/4}},$$

which completes the proof of Proposition 3.2. ∎

We stress again that requiring soft interaction matrices is unnecessarily restrictive. It is simple and satisfactory for the chosen applications, but so long as the spin system has some constant probability of coupling σ_{A_i} and σ'_{A_i}, the argument still holds.

For the purposes of the algorithm, we will not have two configurations that differ on only one vertex, but luckily we can fix this easily. First, define

$$\Delta := \left\{x \in S : \forall y \in \partial S \text{ such that } \tau_y \neq \tau'_y, \|x - y\| \geq L^{\frac{1}{2}}\right\}.$$

Corollary 3.4. *Let S be a q-spin system on the integer lattice $\mathbb{Z}^2$ which exhibits weak spatial mixing. Then, there exists $L_0 \in \mathbb{N}$ and a positive constant γ_0 such that for all squares $S \subset \mathbb{Z}^2$ with side length $2L > 2L_0$, all sets T with $\partial S \subseteq T \subseteq \boxplus S$, and all configurations $\tau, \tau' \in \Omega_T$ which agree on $\boxplus S \setminus \partial S$,*

$$d_{\mathrm{TV}}\left(\mu_\Delta^{\tau}, \mu_\Delta^{\tau'}\right) \leq 8Le^{-\gamma_0 L^{1/4}}.$$

Proof. We simply change τ to τ' by flipping the spin at one boundary vertex at a time. There are $8L$ boundary vertices, so applying the triangle inequality and Proposition 3.2 gives the result. ∎

4. The algorithm

This version of lazy depth-first sampling is very similar to the one we present in [2], but here we take more care in our choices of vertices to sample.

Before we begin the algorithm, we define a (L-)mesh $\Gamma \subset \mathbb{Z}^2$, where each mesh square in Γ has side-length L. To be precise, $\Gamma = (V_\Gamma, E_\Gamma)$ is defined with

$$V_\Gamma = \{v = (i, j) \in \mathbb{Z}^2 : i = 0 \pmod{L}\} \cup \{v = (i, j) \in \mathbb{Z}^2 : j = 0 \pmod{L}\},$$
$$E_\Gamma = \{(u, v) : u, v \in V_\Gamma, d(u, v) = 1\}.$$

To ensure that the bound on total variation distance from Proposition 3.2 applies, LAZY will only sample vertices in Γ. Also, for each $v = (i, j) \in V_\Gamma$, define $\Gamma_v \subset \Gamma$ to be the square of side length $2L$ which contains v and for which $d(\Gamma_v, v) \geq L/2$. Precisely, let $v_c = (k, \ell) \in \Gamma$ be the vertex such that $k, \ell = 0 \pmod{L}$ and

$$k + \frac{L}{2} \leq i < k + \frac{3L}{2},$$
$$\ell + \frac{L}{2} \leq j < \ell + \frac{3L}{2},$$

and define
$$\Gamma_v = \partial\{u \in \mathbb{Z}^2 : u = v_c + (a, b), 0 < a, b < 2L\}.$$

When we recurse in the algorithm called on vertex v, we will always recurse on vertices in Γ_v.

The algorithm takes a q-spin system on $\mathbb{Z}^2$ with a partial configuration σ on some subset of the vertices, $\Lambda \subset \Gamma$. Informally, the partial configuration (Λ, σ) represents the spins that have already been determined. The algorithm is then called on some unassigned vertex $v \in \Gamma$. We compute the following probabilities:

$$p_v^i := \min_{\tau \in \Omega_{\Gamma_v \setminus \Lambda}} \mu_v^{(\Lambda, \sigma) \oplus (\Gamma_v \setminus \Lambda, \tau)}(i) \quad \forall i \in \{1, \dots, q\},$$
$$p_v^0 := 1 - \sum_{i \in [q]} p_v^i.$$

With these values, we partition the unit interval into $q + 1$ subintervals of lengths $(p_v^i : 0 \leq i \leq q)$, and draw a sample $y \sim U[0, 1]$. If y is in a subinterval corresponding to p_v^i, with $i \neq 0$, then we set the spin to i—if not, we recursively call the algorithm

on the vertices in $\Gamma_v \setminus \Lambda$ in turn, each time conditioning on the assignments to the previous vertices in this set. In the case that we need recursion, we use the subroutine BD-CALC to compute the spin at v via this recursion. (See Algorithm 1.)

Note that this differs from the vanilla LAZY in [2] (there referred to as SSMS) by our choice of recursion: here, to take advantage of the total variation bound, we must restrict our samples to a mesh Γ, while on a spin system which exhibits strong spatial mixing we can simply look at the ball of radius L around v instead.

Note also that a spin system is determined by a graph, a number of spins, a vector of weights, and an interaction matrix. Our algorithm is specifically for the case of q spins on the graph $\mathbb{Z}^2$, so we include these in the statement of the algorithm for emphasis.

Algorithm 1 LAZY$(S = (\mathbb{Z}^2, q, b, A), (\Lambda, \sigma), v, \Gamma)$

1: **Input:** The algorithm takes a spin system S, a mesh $\Gamma \subset \mathbb{Z}^2$, a set of known vertices $\Lambda \subset \Gamma$ with a configuration $\sigma \in \Omega_\Lambda$, and a vertex to sample $v \in \Gamma \setminus \Lambda$.
2: **Output:** The algorithm returns the partial configuration passed in with a spin at v as well: $(\Lambda, \sigma) \oplus (v, i)$ for some $i \in \{1, \ldots, q\}$.
3: Define Γ_v as above.
4: **for** $i \in \{1, \ldots, q\}$ **do**
5:     $p_v^i \leftarrow \min_{\tau \in \Omega_{\Gamma_v \setminus \Lambda}} \mu_v^{\sigma \oplus \tau}(i)$
6:     $I_i \leftarrow \left[\sum_{j=1}^{i-1} p_v^j, \sum_{j=1}^{i} p_v^j \right)$
7: **end for**
8: $p_v^0 \leftarrow 1 - \sum_{i \in [q]} p_v^i$
9: $I_0 \leftarrow [1 - p_v^0, 1]$ {Computing the zone of indecision}
10: Sample $y \sim U[0, 1]$, a realisation of a uniform $[0, 1]$ random variable.
11: **for** $i = \{1, \ldots, q\}$ **do**
12:     **if** $y \in I_i$ **then**
13:         **return** $((\Lambda, \sigma) \oplus (v, i))$
14:     **end if**
15: **end for**
16: **return** BD-CALC $\left(S, (\Lambda, \sigma), v, (p_v^1, \ldots, p_v^q), y\right)$

It is worth noting that this algorithm also provides a way to sample vertices outside the mesh Γ. For $v = (i, j) \in \mathbb{Z}^2 \setminus \Gamma$, let $v_c = (k, \ell) \in \Gamma$ be the vertex such that $k, \ell = 0$ (mod L) and

$$k < i < k + L,$$
$$\ell < j < \ell + L,$$

Algorithm 2 BD-CALC $(\mathcal{S}, (\Lambda, \sigma), v, (p_v^1, \ldots, p_v^q), y)$

 1: Give $\Gamma_v \setminus \Lambda$ an ordering $\Gamma_v \setminus \Lambda = \{w_1, \ldots, w_m\}$.
 2: $(\Lambda', \sigma') \leftarrow (\Lambda, \sigma)$
 3: **for** $j \in [1, \ldots, m]$ **do**
 4:    $(\Lambda', \sigma') \leftarrow$ LAZY$(\mathcal{S}, (\Lambda', \sigma'), w_j, \Gamma)$
 5: **end for**
 6: **for** $i \in \{1, \ldots, q\}$ **do**
 7:    $\varrho_v^i \leftarrow \mu_v^{\sigma'}(i) - p_v^i$
 8:    $I_i \leftarrow \left[\sum_{j=1}^q p_v^j + \sum_{k=1}^{i-1} \varrho_v^k, \sum_{j=1}^q p_v^j + \sum_{k=1}^i \varrho_v^k \right)$
 9:    **if** $y \in I_i$ **then**
10:       **return** $((\Lambda, \sigma) \oplus (v, i))$
11:    **end if**
12: **end for**

---

and define
$$B = \partial\{u \in \mathbb{Z}^2 : u = v_c + (a, b), 0 < a, b < L\}.$$

The algorithm allows us to sample the spin of every vertex in B, at which point we can sample the spin at v by using any algorithm (brute force calculation is always an option) to sample spin systems with boundary.

The verification of correctness of LAZY is similar to that presented in [2] as is the analysis of its run-time—for any LAZY-style algorithm, the change in proof amounts to relabelling the recursive vertices. The run-time also follows similar arguments, but our bounds are different and they come from different sources, so we include the updated proof.

5. Run-time

Our run-time analysis is in terms of the number of calls of LAZY. For properly chosen L, and hence Γ, the amount of time each call takes for vertices on Γ will be $O(1)$.

Our viewpoint is that the algorithm behaves similarly to a branching process where each recursive call of the algorithm is a child of the vertex which initiated the call. This gives us an upper bound to the run-time of LAZY on $\mathbb{Z}^2$ in a broad class of spin systems.

Now, note that, although the interior S of the square Γ_v of Γ may not be empty as the algorithm progresses, only the vertical and horizontal bisectors of Γ_v may be occupied. In other words, only the spins of vertices in Γ have been fixed. From our adaptation of Martinelli, Olivieri, and Schonmann's result (Corollary 3.4), we know

that the distribution of the spin at v is only weakly influenced by the boundary condition on Γ_v.

Theorem 5.1. *Let $\mathcal{S} = \left(\mathbb{Z}^2, q, b, A\right)$ be a spin system with a soft interaction matrix A, and which exhibits weak spatial mixing. Take γ_0 and L_0 from Corollary 3.4, and let $L > L_0$ satisfy*

$$64L^2 e^{-\gamma_0 L^{1/4}} < 1.$$

Let Γ be an L-mesh. Then, for all $v \in \Gamma$, $\Lambda \subset \Gamma$, and $\sigma \in \Omega_\Lambda$ such that

$$\mu_\Lambda(\sigma) > 0,$$

the expected number of times LAZY $(\mathcal{S}, (\Lambda, \sigma), v, \Gamma)$ *calls* LAZY *(in total) is* $O(1)$.

Note that we can always choose L large enough to satisfy

$$L > L_0 \quad \text{and} \quad 64L^2 e^{-\gamma_0 L^{1/4}} < 1.$$

Proof. We construct a random tree T_A out of our recursive calls. Our root will be our initial call. Recursively, if at some node w of the tree we fall in the zone of indecision, I_0, we add a child to the node for every recursive call that is initiated by BD-CALC. If we do not fall in the zone of indecision, then w has no children.

Let T_B be the branching process with offspring distribution ξ where

$$\mathbb{P}[\xi = 0] = 1 - 8L e^{-\gamma_0 L^{1/4}}, \quad \mathbb{P}[\xi = 8L] = 8L e^{-\gamma_0 L^{1/4}}.$$

Then, we can couple T_B with T_A so that T_B stochastically dominates T_A. Specifically, T_B's offspring distribution has a higher probability of having children than any of T_A's offspring distributions, as, Corollary 3.4 tells us that at every node $w \in T_A$, the probability of falling in the zone of indecision is bounded by $8L e^{-\gamma_0 L^{1/4}}$. Also, the offspring of T_B always outnumber the offspring of T_A, as when w does have children, the number of children is bounded above by $|\Gamma_v| = 8L$. We can then couple T_B and T_A so that whenever T_A has children, so does T_B. It follows that under this coupling, T_A will always be a subtree of T_B.

Now, we can leverage the theory of branching processes (see [7] for the original statement, or section 5.4 of [26] for a modern discussion) to finish off our bound. We know that

$$\mathbb{E}[\xi] \le 64L^2 e^{-\gamma_0 L^{1/4}} < 1,$$

so by the fundamental theorem of branching processes, it follows that

$$\mathbb{E}[|T_A|] \le \mathbb{E}[|T_B|] \le \frac{1}{1 - 64L^2 e^{-\gamma_0 L^{1/4}}} = O(1). \qquad \blacksquare$$

The qualitative nature of the main result in this section is summarised in the following corollary.

Corollary 5.2. *Let $\mathcal{S} = (\mathbb{Z}^2, q, b, A)$ be a spin system which has a soft interaction matrix A, and which exhibits weak spatial mixing. Let $\Gamma \subset \mathbb{Z}^2$ be an L-mesh with L chosen large enough to satisfy Theorem 5.1 and consider any choices of $v \in V$, $\Lambda \subset \Gamma$, and $\sigma \in \Omega_\Lambda$. Then, the running time R of $\mathrm{LAZY}(\mathcal{S}, (\Lambda, \sigma), v, \Gamma)$ satisfies $\mathbb{E}[R] = O(1)$.*

6. Correctness

The proof of correctness is nearly identical to that in [2], so we include only an intuitive explanation.

Theorem 6.1. *Let $\mathcal{S} = (\mathbb{Z}^2, q, b, A)$ be a spin system which has a soft interaction matrix A and which exhibits weak spatial mixing, and let Γ be the chosen mesh. Suppose for all $v \in \Gamma$, all finite $\Lambda \subset \Gamma$, and all $\sigma \in \Omega_\Lambda$ such that $\mu_\Lambda(\sigma) > 0$, $\mathrm{LAZY}(\mathcal{S}, (\Lambda, \sigma), v, \Gamma)$ terminates with probability 1. Then,*

$$\mathbb{P}\big[\mathrm{LAZY}(\mathcal{S}, (\Lambda, \sigma), v, \Gamma) = i\big] = \mu_v^\sigma(i).$$

We can show that the algorithm is correct in an inductive manner, just one lacking a traditional base case. The issue here is that the tree of recursive calls has paths of unbounded length. In the earlier paper [2], this problem is circumvented by analysing an artificial version of the algorithm in which the recursion depth is restricted to a bound h. Correctness of the actual algorithm follows by taking a limit as $h \to \infty$, assuming termination with probability 1. The key factor in the algorithm's correctness is that each recursive call samples a spin at a vertex in Γ_v from the correct conditional distribution, given the spins already decided. We thus arrive at the correct distribution of spins on the whole of Γ_v by the chain rule for conditional probabilities.

Suppose we know that the recursive calls give us the correct probability, that is, that

$$\mathbb{P}\left[\mathrm{LAZY}\left(\mathcal{S}, (\Lambda, \sigma) \oplus \bigoplus_{k=1}^{j-1}(w_k, \tau_{w_k}), w_j, \Gamma\right) = \tau_{w_j}\right] = \mu_{w_j}^{(\Lambda,\sigma)\oplus\bigoplus_{k=1}^{j-1}(w_k,\tau_{w_k})}(\tau_{w_j}).$$

Then, given the order $\Gamma_v \setminus \Lambda = \{w_1, \ldots, w_m\}$ that the algorithm processes the vertices in, we have

$$\mathbb{P}\big[\mathrm{LAZY}(\mathcal{S}, (\Lambda, \sigma), v, \Gamma) = i\big]$$

$$= p_v^i + \sum_{\tau \in \Omega_{\Gamma_v \setminus \Lambda}} (\mu_v^{\sigma \oplus \tau}(i) - p_v^i) \prod_{j=1}^{m} \mathbb{P}\left[\mathrm{LAZY}\left(\mathcal{S}, (\Lambda, \sigma) \oplus \bigoplus_{k=1}^{j-1}(w_k, \tau_{w_k}), w_j, \Gamma\right) = \tau_{w_j}\right]$$

$$= p_v^i + \sum_{\tau \in \Omega_{\Gamma_v \setminus \Lambda}} (\mu_v^{\sigma \oplus \tau}(i) - p_v^i) \prod_{j=1}^{m} \mu_{w_j}^{(\Lambda,\sigma)\oplus\bigoplus_{k=1}^{j-1}(w_k,\tau_{w_k})}(\tau_{w_j}).$$

Now, we note that

$$p_v^i = \sum_{\tau \in \Omega_{\Gamma_v \setminus \Lambda}} p_v^i \cdot \prod_{j=1}^{m} \mu_{w_j}^{(\Lambda, \sigma) \oplus \bigoplus_{k=1}^{j-1} (w_k, \tau_{w_k})} (\tau_{w_j}),$$

from which it follows that

$$\mathbb{P}\big[\text{LAZY}(\mathcal{S}, (\Lambda, \sigma), v, \Gamma) = i\big]$$

$$= p_v^i + \sum_{\tau \in \Omega_{\Gamma_v \setminus \Lambda}} (\mu_v^{\sigma \oplus \tau}(i) - p_v^i) \prod_{j=1}^{m} \mu_{w_j}^{(\Lambda, \sigma) \oplus \bigoplus_{k=1}^{j-1} (w_k, \tau_{w_k})} (\tau_{w_j})$$

$$= \sum_{\tau \in \Omega_{\Gamma_v \setminus \Lambda}} \mu_v^{\sigma \oplus \tau}(i) \prod_{j=1}^{m} \mu_{w_j}^{(\Lambda, \sigma) \oplus \bigoplus_{k=1}^{j-1} (w_k, \tau_{w_k})} (\tau_{w_j})$$

$$= \mu_v^{\sigma}(i),$$

which gives us correctness.

As promised earlier, Theorem 1.1 follows from Theorems 5.1 and 6.1. In the case that $S \subset \mathbb{Z}^2$ is a single vertex, this is immediate. Otherwise, define

$$\Gamma_S = \bigcup_{v \in S} \Gamma_v,$$

where Γ_v is defined as in Section 4. Then, following the strategy described at the end of that section, we use Algorithm 1 to sample spins for all vertices in Γ_S. Since each vertex $v \in S$ is surrounded by an $L \times L$ square of assigned spins, the distribution of the spin at v can be determined by brute force.

7. Applications

7.1. The ferromagnetic Potts model

An interesting application of this algorithm is that we can sample the ferromagnetic Potts model on $\mathbb{Z}^2$ up to its weak spatial mixing threshold. The Potts model can be characterised by the following interactions:

$$b = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}, \quad A = \begin{pmatrix} e^\beta & 1 & \cdots & 1 \\ 1 & e^\beta & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & e^\beta \end{pmatrix},$$

where the number of spins is $q \geq 2$. To be ferromagnetic, A must be positive definite, i.e., $e^\beta \geq 1$.

In their seminal paper [6], Beffara and Duminil–Copin established for all that the critical inverse temperature of the ferromagnetic Potts model with $q \geq 2$ is

$$\beta_c = \log\left(1 + \sqrt{q}\right),$$

and that correlation decays exponentially for $\beta < \beta_c$.

For our purposes, we require not just correlation decay, but weak spatial mixing (of the set form in particular). Fortunately, Alexander showed that for all $\beta < \beta_c$, weak spatial mixing, as we define it here, occurs in the Potts model for $q \geq 2$ if and only if correlation decays [1, Theorem 3.6]. These two results then ensure that the conditions of Corollary 5.2 hold and that Algorithm 1 has linear run-time.

It is tempting to approach the low temperature regime via the Edwards–Sokal coupling [15] and the dual graph and flows as done by Huijben, Patel, and Regts [31]. Unfortunately, this correspondence breaks down under the behaviour of our algorithm. To determine spins in the dual, we have to pass through the random cluster model and to do so, we must identify connected clusters. If we limit ourselves to boxes of a specific size, what may appear disconnected to us could very well be connected outside our view.

7.2. The ferromagnetic Ising model with external field

A second new application is to the ferromagnetic Ising model with constant external field at any temperature. The Ising model with constant external field h on the positive spins can be characterised by

$$b = \begin{pmatrix} 1 \\ h \end{pmatrix}, \quad A = \begin{pmatrix} e^{\beta} & 1 \\ 1 & e^{\beta} \end{pmatrix}.$$

Schonmann and Schlosman established that at all temperatures, the ferromagnetic Ising model with a constant external field exhibits weak spatial mixing [38, Theorem 2]. It follows that the conditions of Corollary 5.2 hold and that Algorithm 1 has linear run-time on the model.

References

[1] K. S. Alexander, On weak mixing in lattice models. *Probab. Theory Related Fields* **110** (1998), no. 4, 441–471 Zbl 0906.60074 MR 1626951

[2] K. Anand and M. Jerrum, Perfect sampling in infinite spin systems via strong spatial mixing. *SIAM J. Comput.* **51** (2022), no. 4, 1280–1295 Zbl 07571445 MR 4456711

[3] A. Barvinok, Computing the partition function for cliques in a graph. *Theory Comput.* **11** (2015), 339–355 Zbl 1351.05212 MR 3433628

[4] A. Barvinok and G. Regts, Weighted counting of solutions to sparse systems of equations. *Combin. Probab. Comput.* **28** (2019), no. 5, 696–719 Zbl 1433.68166 MR 3991383

[5] A. Barvinok and P. Soberón, Computing the partition function for graph homomorphisms. *Combinatorica* **37** (2017), no. 4, 633–650 Zbl 1399.05209 MR 3694706

[6] V. Beffara and H. Duminil-Copin, The self-dual point of the two-dimensional random-cluster model is critical for $q \geq 1$. *Probab. Theory Related Fields* **153** (2012), no. 3-4, 511–542 Zbl 1257.82014 MR 2948685

[7] M. Bienaymé, *De la loi de multiplication et de la durée des familles probabilités*. Cosson, 1845

[8] A. Blanca and A. Sinclair, Random-cluster dynamics in $\mathbb{Z}^2$. *Probab. Theory Related Fields* **168** (2017), no. 3-4, 821–847 Zbl 1369.60067 MR 3663632

[9] M. Bordewich, C. Greenhill, and V. Patel, Mixing of the Glauber dynamics for the ferromagnetic Potts model. *Random Structures Algorithms* **48** (2016), no. 1, 21–52 Zbl 1332.60132 MR 3432570

[10] C. Borgs, J. Chayes, T. Helmuth, W. Perkins, and P. Tetali, Efficient sampling and counting algorithms for the Potts model on $\mathbb{Z}^d$ at all temperatures. In *STOC'20—Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing*, pp. 738–751, The Association for Computing Machinery, New York, 2020 Zbl 07298284 MR 4141796

[11] X. Chen and X. Zhang, A near-linear time sampler for the Ising model with external field. In *Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pp. 4478–4503, SIAM, Philadelphia, PA, 2023 Zbl 07848130 MR 4538128

[12] Z. Chen, K. Liu, and E. Vigoda, Rapid mixing of Glauber dynamics up to uniqueness via contraction. In *2020 IEEE 61st Annual Symposium on Foundations of Computer Science*, pp. 1307–1318, IEEE Computer Society, Los Alamitos, CA, 2020 Zbl 07672228 MR 4232132

[13] Z. Chen, K. Liu, and E. Vigoda, Optimal mixing of Glauber dynamics: entropy factorization via high-dimensional expansion. In *STOC '21—Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, pp. 1537–1550, The Association for Computing Machinery, New York, 2021 Zbl 07765268 MR 4398939

[14] M. Dyer, A. Sinclair, E. Vigoda, and D. Weitz, Mixing in time and space for lattice spin systems: a combinatorial view. *Random Structures Algorithms* **24** (2004), no. 4, 461–479 Zbl 1126.82021 MR 2060631

[15] R. G. Edwards and A. D. Sokal, Generalization of the Fortuin–Kasteleyn–Swendsen–Wang representation and Monte Carlo algorithm. *Phys. Rev. D (3)* **38** (1988), no. 6, 2009–2012 MR 0965465

[16] W. Feng, H. Guo, C. Wang, J. Wang, and Y. Yin, Towards derandomising Markov chain Monte Carlo. In *2023 IEEE 64th Annual Symposium on Foundations of Computer Science—FOCS 2023*, pp. 1963–1990, IEEE Computer Society, Los Alamitos, CA, 2023 MR 4720362

[17] W. Feng, H. Guo, and J. Wang, Sampling from the ferromagnetic Ising model with external fields. 2022

[18] W. Feng, H. Guo, and Y. Yin, Perfect sampling from spatial mixing. *Random Structures Algorithms* **61** (2022), no. 4, 678–709 Zbl 1534.62010 MR 4504840

[19] J. A. Fill and M. Huber, The randomness recycler: a new technique for perfect sampling. In *41st Annual Symposium on Foundations of Computer Science (Redondo Beach, CA, 2000)*, pp. 503–511, IEEE Computer Society, Los Alamitos, CA, 2000 MR 1931847

[20] S. Friedli and Y. Velenik, *Statistical mechanics of lattice systems. A concrete mathematical introduction*. Cambridge University Press, Cambridge, 2018 Zbl 1407.82001 MR 3752129

[21] A. Galanis, L. A. Goldberg, and P. Smolarova, Sampling from the random cluster model on random regular graphs at all temperatures via Glauber dynamics. In *Approximation, randomization, and combinatorial optimization. Algorithms and techniques*, p. article no. 64, LIPIcs. Leibniz Int. Proc. Inform. 275, Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Wadern, 2023 MR 4640792

[22] R. Gheissari and E. Lubetzky, Mixing times of critical two-dimensional Potts models. *Comm. Pure Appl. Math.* **71** (2018), no. 5, 994–1046 Zbl 1392.82007 MR 3794520

[23] R. Gheissari and E. Lubetzky, Quasi-polynomial mixing of critical two-dimensional random cluster models. *Random Structures Algorithms* **56** (2020), no. 2, 517–556 Zbl 1453.82043 MR 4060355

[24] R. Gheissari and A. Sinclair, Spatial mixing and the random-cluster dynamics on lattices. *Random Structures Algorithms* **64** (2024), no. 2, 490–534 Zbl 07811296 MR 4704276

[25] R. J. Glauber, Time-dependent statistics of the Ising model. *J. Mathematical Phys.* **4** (1963), 294–307 Zbl 0145.24003 MR 0148410

[26] G. R. Grimmett and D. R. Stirzaker, *Probability and random processes*. Oxford University Press, Oxford, 2020 Zbl 0493.60001 MR 4229142

[27] H. Guo, M. Jerrum, and J. Liu, Uniform sampling through the Lovász local lemma. *J. ACM* **66** (2019), no. 3, article no. 18 Zbl 1425.68451 MR 3941342

[28] K. He, C. Wang, and Y. Yin, Sampling Lovász local lemma for general constraint satisfaction solutions in near-linear time. In *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science—FOCS 2022*, pp. 147–158, IEEE Computer Society, Los Alamitos, CA, 2022 MR 4537198

[29] K. He, K. Wu, and K. Yang, Improved bounds for sampling solutions of random CNF formulas. In *Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pp. 3330–3361, SIAM, Philadelphia, PA, 2023 Zbl 07848088 MR 4538081

[30] T. Helmuth, W. Perkins, and G. Regts, Algorithmic Pirogov–Sinai theory. *Probab. Theory Related Fields* **176** (2020), no. 3-4, 851–895 Zbl 1436.82007 MR 4087485

[31] J. Huijben, V. Patel, and G. Regts, Sampling from the low temperature Potts model through a Markov chain on flows. *Random Structures Algorithms* **62** (2023), no. 1, 219–239 Zbl 07751495 MR 4516083

[32] F. Martinelli and E. Olivieri, Approach to equilibrium of Glauber dynamics in the one phase region. I. The attractive case. *Comm. Math. Phys.* **161** (1994), no. 3, 447–486 Zbl 0793.60110 MR 1269387

[33] F. Martinelli, E. Olivieri, and R. H. Schonmann, For 2-D lattice spin systems weak mixing implies strong mixing. *Comm. Math. Phys.* **165** (1994), no. 1, 33–47 Zbl 0811.60097 MR 1298940

[34] E. Mossel and A. Sly, Exact thresholds for Ising–Gibbs samplers on general graphs. *Ann. Probab.* **41** (2013), no. 1, 294–328 Zbl 1270.60113 MR 3059200

[35] S. Park, Y. Jang, A. Galanis, J. Shin, D. Stefankovic, and E. Vigoda, Rapid mixing Swendsen–Wang sampler for stochastic partitioned attractive models. In *International conference on artificial intelligence and statistics*, pp. 440–449, Proceedings of Machine Learning Research, PMLR, 2017

[36] V. Patel and G. Regts, Deterministic polynomial-time approximation algorithms for partition functions and graph polynomials. *SIAM J. Comput.* **46** (2017), no. 6, 1893–1919 Zbl 1383.68099 MR 3738853

[37] J. G. Propp and D. B. Wilson, How to get a perfectly random sample from a generic Markov chain and generate a random spanning tree of a directed graph. *J. Algorithms* **27** (1998), no. 2, 170–217 Zbl 0919.68092 MR 1622393

[38] R. H. Schonmann and S. B. Shlosman, Complete analyticity for 2D Ising completed. *Comm. Math. Phys.* **170** (1995), no. 2, 453–482 Zbl 0821.60097 MR 1334405

[39] A. Sinclair, P. Srivastava, D. Štefankovič, and Y. Yin, Spatial mixing and the connective constant: optimal bounds. *Probab. Theory Related Fields* **168** (2017), no. 1-2, 153–197 Zbl 1373.82026 MR 3651051

[40] A. Sinclair, P. Srivastava, and Y. Yin, Spatial mixing and approximation algorithms for graphs with bounded connective constant (extended abstract). In *2013 IEEE 54th Annual Symposium on Foundations of Computer Science—FOCS 2013*, pp. 300–309, IEEE Computer Society, Los Alamitos, CA, 2013 MR 3246232

[41] Y. Spinka, Finitary codings for spatial mixing Markov random fields. *Ann. Probab.* **48** (2020), no. 3, 1557–1591 Zbl 1457.60134 MR 4112724

[42] R. H. Swendsen and J.-S. Wang, Nonuniversal critical dynamics in Monte Carlo simulations. *Phys. Rev. Lett.* **58** (1987), no. 2, 86–88

[43] M. Ullrich, Comparison of Swendsen–Wang and heat-bath dynamics. *Random Structures Algorithms* **42** (2013), no. 4, 520–535 Zbl 1272.82009 MR 3068036

[44] J. van den Berg and J. E. Steif, On the existence and nonexistence of finitary codings for a class of random fields. *Ann. Probab.* **27** (1999), no. 3, 1501–1522 Zbl 0968.60091 MR 1733157

[45] D. Weitz, Counting independent sets up to the tree threshold. In *STOC'06: Proceedings of the 38th Annual ACM Symposium on Theory of Computing*, pp. 140–149, The Association for Computing Machinery, New York, 2006 Zbl 1301.68276 MR 2277139

Konrad Anand
School of Mathematical Sciences, Queen Mary University of London, Mile End Road, London
E1 4NS, UK; konrad.anand@me.com

Mark Jerrum
School of Mathematical Sciences, Queen Mary University of London, Mile End Road, London
E1 4NS, UK; m.jerrum@qmul.ac.uk