Mixing and Cutoff for the Systematic Scan Dynamics of the Mean-Field Ferromagnetic Potts Model

Antonio Blanca
The Pennsylvania State University, University Park, PA, USA

Md Tahmidur Rafid
The Pennsylvania State University, University Park, PA, USA

Abstract

We study the mixing time of the systematic scan dynamics for the q-state ferromagnetic Potts model on the n-vertex complete graph, known as the mean-field model. This Markov chain updates vertices sequentially according to a fixed predetermined order, in contrast to the Glauber dynamics which updates a uniformly random vertex at each step. Systematic scan dynamics are attractive in practice as they often demonstrate strong empirical performance. However, their theoretical analysis remains far less developed than that of the Glauber dynamics.

We take a step toward addressing this imbalance by showing that for every $q \geq 2$ and $\beta < \beta_s$, where β_s is the metastability threshold associated with the onset of slow mixing for the Glauber dynamics, the systematic scan dynamics for the ferromagnetic mean-field Potts model mixes in $\Theta(\log n)$ scans or, equivalently, in $\Theta(n \log n)$ single site updates. We in fact prove a sharper result; namely, that there exists a constant $c(\beta, q) > 0$ such that the mixing time is $c(\beta, q) \log n + \Theta(1)$, which implies that the Markov chain exhibits the cutoff phenomenon, with the total variation distance to the stationary distribution dropping abruptly from nearly 1 to nearly 0 within a narrow $\Theta(1)$ time window. This result is tight in β as well since the dynamics mixes exponentially slowly for $\beta > \beta_s$. To the best of our knowledge, this is the first general cutoff result for the systematic scan dynamics in the context of spin systems. The result may also be of independent interest in the theory of Markov chains, since the systematic scan dynamics is both global and non-reversible, two settings in which cutoff remains poorly understood.

2012 ACM Subject Classification Theory of computation → Random walks and Markov chains

Keywords and phrases Markov chain, mixing times, systematic scan, Potts model

Digital Object Identifier 10.4230/LIPIcs.APPROX/RANDOM.2026.63

Category RANDOM

Related Version *Full Version*: https://arxiv.org/abs/2607.09841 [5]

Funding Research supported in part by NSF CAREER grant CCF-2143762.

1 Introduction

Sampling high-dimensional probability distributions is an important task in science and engineering. It is, for example, a prevalent problem when running simulations in statistical physics or when solving inference problems in machine learning. As such, there is an array of sophisticated algorithms and heuristics to generate samples from these distributions. One of the most classical and powerful approaches is *Gibbs Sampling*, which iteratively updates one variable at a time conditioned on the values of all other variables.

There are two standard strategies for deciding which variable to update in each step: (i) the Glauber dynamics, where the variable to be updated is chosen uniformly at random, and (ii) systematic scan dynamics, where variables are updated according to a predetermined ordering. In practice, systematic scan dynamics offer several computational advantages over the Glauber dynamics, as they are often amenable to parallelization, can exhibit improved memory locality and thus reduce data-access overhead, and require fewer random bits [29]. Despite these practical advantages, most theoretical analyses focus on the Glauber dynamics, which have historically proven more mathematically tractable, leaving the convergence theory for systematic scan dynamics significantly less developed.

The number of variable updates required to reach stationarity can differ significantly between the Glauber and systematic scan dynamics, sometimes up to polynomial factors in n [21, 27]. For example, even in the trivial case of n independent variables, the systematic scan mixes in exactly n updates, whereas the Glauber dynamics requires $\frac{1}{2}n\log n$ updates. When combined with the implementation benefits noted above, such differences in convergence rates can make the systematic scan dynamics a particularly attractive update strategy.

In this paper, we take a step toward a better theoretical understanding of the systematic scan dynamics by studying them in the context of the *ferromagnetic mean-field Potts model*, a classical spin system model central in statistical physics, theoretical computer science, machine learning, and probability theory. In this model, each variable in $V = \{1, \dots, n\}$ can take a value from $Q = \{1, ..., q\}$. The Gibbs or Boltzmann distribution assigns to each configuration $\sigma \in \Sigma_n = Q^V$ a probability given by:

$$\mu(\sigma) = \frac{1}{Z_{\beta,n}} \exp\left\{ \frac{\beta}{n} \sum_{\{u,v\}\subseteq V} \mathbf{1}\big(\sigma(u) = \sigma(v)\big) \right\},$$

where $\sigma(u)$ is the value (or spin) assigned to vertex u in the configuration σ, and $Z_{\beta,n}$ is the normalization constant or partition function. The parameter β captures the strength of the interaction between the variables and is proportional to the inverse temperature in statistical physics applications. The $q = 2$ case corresponds to the classical Ising model. The Potts model can be defined on any graph by taking the summation in the definition over the edges of the graph instead of over pairs of vertices; the mean-field setting considered here corresponds to the complete graph on n vertices.

For concreteness, we consider the standard variant of the systematic scan dynamics in which site updates are given by the *heat-bath* update, i.e., they are performed according to the conditional distribution at each update site induced by the configuration on the remaining vertices. Formally, if vertex v is to be updated in a configuration σ, then v is assigned a new spin sampled from $\mu(\,\cdot \mid \sigma(V \setminus \{v\}))$. This is a non-reversible ergodic Markov chain with state space Σ_n that converges to μ. We measure rates of convergence to μ in terms of the *mixing time*, which is the most standard notion of speed of convergence for a Markov chain. It is defined as the number of steps until the dynamics is close in total variation distance to μ starting from any initial configuration. Formally, if P denotes the transition matrix of the chain, for $\varepsilon \in (0, 1)$, the ε-mixing time is given by

$$T_{\text{mix}}(\varepsilon) = \min\left\{ t \in \mathbb{Z}_{\geq 0} : \max_{\sigma\in\Sigma_n} \left\| P^t(\sigma, \cdot) - \mu \right\|_{\text{TV}} \leq \varepsilon \right\}.$$

By convention, the mixing time is taken to be $T_{\text{mix}} = T_{\text{mix}}(1/4)$.

A more refined notion of convergence, when present, is the cutoff phenomenon. A Markov chain exhibits *cutoff* when there is a sharp drop in the total variation distance from close to 1 to close to 0 in an interval of time of smaller order than its mixing time; which is called the *cutoff window*. More precisely, we say that there is cutoff if as $n \to \infty$, for any fixed $\varepsilon \in (0, 1)$

$$T_{\text{mix}}(\varepsilon) - T_{\text{mix}}(1 - \varepsilon) = o(T_{\text{mix}}). \tag{1}$$

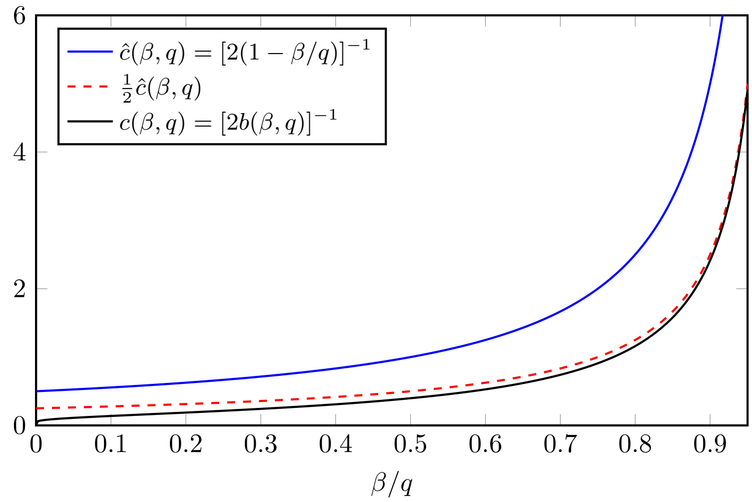

Figure 1 Comparison of mixing times of Glauber (blue curve) and systematic scan dynamics (black curve).

The mixing time and cutoff of the Glauber dynamics for the mean-field ferromagnetic Potts model is by now very well-understood. In the case of the Ising model (the $q = 2$ case) at $\beta < 2$, the Glauber dynamics mixes in $\frac{1}{2(1-\beta/2)} n \log n$ steps with an $O(n)$ cutoff window. The mixing time is $\Theta(n^{3/2})$ at $\beta = 2$ and exponential in n when $\beta > 2$ [13, 12, 23]. For $q \geq 3$, the model exhibits a first-order phase transition and the dynamics undergoes a slowdown at $\beta_s = \beta_s(q)$, which is the spinodal point marking the onset of metastability. Specifically, when $\beta < \beta_s$ the mixing time is $\frac{1}{2(1-\beta/q)} n \log n$ with an $O(n)$ cutoff window. At $\beta = \beta_s$ the dynamics mixes in $\Theta(n^{4/3})$ and no longer exhibits cutoff, and the mixing time is again exponential in n for $\beta > \beta_s$ [10].

In contrast with this fairly complete convergence picture for the mean-field Glauber dynamics, much less is known for systematic scan dynamics, with the best known mixing time bound being $O(n^2 \log n)$ for $\beta < \beta_s$. We provide here the first sharp analysis of the mixing time of the systematic scan dynamics for the mean-field ferromagnetic Potts model, valid for all scan orderings and for all inverse temperatures $\beta < \beta_s$.

▶ **Theorem 1.** *For $q \geq 2$, $\beta < \beta_s$, there exists a constant $c(\beta, q) > 0$ such that the systematic scan dynamics for the mean-field ferromagnetic Potts model exhibits cutoff at mixing time $T_{mix} = c(\beta, q) \log(n) + \Theta(1)$.*

For n large, the leading constant $c(\beta, q)$ corresponds to $\frac{1}{2b(\beta,q)}$ where $b = b(\beta, q)$ is the root of the equation $\beta(1 - e^{-b})/(qb) = e^{-b}$; see Section 4 for additional details. Figure 1 shows a plot of the constant $c(\beta, q)$ as a function of β/q and compares it to the leading constant $\hat{c}(\beta, q) = \frac{1}{2(1-\beta/q)}$ for the mixing time of the Glauber dynamics, revealing at least a twofold speedup. This is consistent with the folklore conjecture that the systematic scan dynamics exhibits such a speedup in several settings.

We note that Theorem 1 is tight in β, since, as mentioned, the Glauber dynamics mixes exponentially slowly for $\beta > \beta_s$ and known mixing time comparison results (see, e.g., [17, Theorem 1.2]) imply that so does the systematic scan. At $\beta = \beta_s$, based on the known behavior of the Glauber dynamics, we conjecture that the systematic scan dynamics mixes $\Theta(n^c)$ for some $c < 1$ without cutoff.

To better contextualize Theorem 1, we briefly comment on prior work on mixing times for the systematic scan dynamics on spin systems. There is a line of work that seeks to compare the convergence rates of Glauber and systematic scan dynamics. This is done with varying degrees of generality in [2, 17, 14, 11, 8, 9, 19]. In parallel, there is work establishing fast mixing for the systematic scan dynamics under model parameter and graph-structural conditions [15, 20, 16], or under monotonicity assumptions [7, 3, 26]. We note that the recent advances in Markov chain mixing via spectral independence have not yet yielded major new results for this dynamics; see [1] for initial progress. This body of work demonstrates sustained interest on the systematic scan dynamics while also underscoring the significant analytical challenges they present, with many questions that are well understood for the Glauber dynamics remaining unresolved in this setting.

Turning to cutoff, results like Theorem 1 are rarer still. To the best of our knowledge, our work provides the first cutoff result for the systematic scan dynamics in the context of spin systems. The cutoff result in [22] for the Ising model considers a Markov chain inspired by systematic scan but that is effectively quite different since in each step, $k = o(n^{1/3})$ vertices are chosen uniformly random and updated according to a random permutation. More generally, the systematic scan dynamics is a global, non-reversible Markov chain, and both features introduce substantial additional analytical challenges. Indeed, the only other cutoff result for a global Markov chain for spin systems we are aware of is the recent work [6] on the Swendsen-Wang dynamics, while the study of the cutoff phenomenon for non-reversible Markov chains is only in its early stages [25].

Proof Sketch

For the rest of the paper, we assume $\beta < \beta_s$, $q \geq 2$ and n to be asymptotically large. We refer to the vertex updates of the dynamics as *site updates*, and n consecutive vertex updates are called a *full-scan*. A step of the systematic scan dynamics Markov chain corresponds to a full-scan starting from the first vertex in the ordering. It will be convenient to index the site updates, and thus we use σ_t for the configuration after the t-th site update and σ_{kn} for the configuration after the k-th step of the chain.

We will show that for any fixed $\varepsilon \in (0, 1)$, there exists a $\kappa(\varepsilon)$ independent of n such that

$$c(\beta, q) \log n - \kappa(\varepsilon) \leq T_{\text{mix}}(\varepsilon) \leq c(\beta, q) \log n + \kappa(\varepsilon). \tag{2}$$

From this, Theorem 1 follows immediately; in particular, this inequality implies (1).

To establish the upper bound on $T_{\text{mix}}(\varepsilon)$ we utilize a multiphase *coupling*, and in that sense, it is reminiscent of the other Markov chain analysis in the mean-field setting [18, 4, 6, 10]. A coupling considers two copies, $\{\sigma_t\}_{t\geq 0}$ and $\{\tilde{\sigma}_t\}_{t\geq 0}$, of the systematic scan dynamics starting from two arbitrary initial configurations σ_0 and $\tilde{\sigma}_0$, respectively. Each copy evolves according to the transition matrix P, but the moves of the two copies can be arbitrarily correlated. The ε-coupling time $T_{\text{coup}}(\varepsilon)$ is the minimum number of full-scans T such that $\Pr[\sigma_{Tn} \neq \tilde{\sigma}_{Tn}] \leq \varepsilon$. A standard result says that $T_{\text{mix}}(\varepsilon) \leq T_{\text{coup}}(\varepsilon)$; for additional details see [24].

To describe the three main phases of our coupling we introduce some useful notation first. For a configuration $\sigma \in \Sigma_n$, let $\alpha(\sigma)$ denote the q-dimensional vector with coordinates

$$\alpha_i(\sigma) = \frac{1}{n} \sum_{j=1}^{n} \mathbf{1}(\sigma(j) = i).$$

We call $\alpha(\sigma)$ the proportions vector of σ. We let $\hat{e} = \big(1/q, \ldots, 1/q\big)$ denote the q-dimensional equiproportions vector that corresponds to the configuration with exactly the same number of vertices assigned each spin (ignoring rounding issues).

The main phases of our coupling are the following:

Phase 1: From arbitrary initial configurations σ_0 and $\tilde{\sigma}_0$, run two copies $\{\sigma_t\}$ and $\{\tilde{\sigma}_t\}$ independently for an initial *burn-in* period of $T = O(1)$ scans. After this initial burn-in, we show that both copies are such that $\|\alpha(\sigma_{Tn}) - \hat{e}\|_\infty \le \rho$ and $\|\alpha(\tilde{\sigma}_{Tn}) - \hat{e}\|_\infty \le \rho$ with probability $1 - o(1)$, for any desired constant $\rho > 0$. In addition, we show that both copies have "good spread" of their spins with respect to the ordering of the systematic scan dynamics.

Phase 2: We can then couple the evolution of the two instances of Markov chain so that after $c(\beta, q)\log n + O(1)$ additional scans: the Hamming distance between the two configurations is $O(\sqrt{n})$ and the ℓ_2 distance between their respective proportions vector and $\hat{e}$ is also $O(1/\sqrt{n})$.

Phase 3: Finally, we couple one final scan in a way that guarantees that the processes coalesce with at least the desired ε probability.

We provide additional details about **Phase 1** next. Let $e_i \in \mathbb{R}^q_{\ge 0}$ be the indicator vector with value 1 in the i-th coordinate and 0 elsewhere. We capture the notion of closeness to $\hat{e}$ with "good spread" by considering for each $\rho > 0$ the set of configurations

$$\Sigma_n^\rho = \left\{\sigma \in \Sigma_n : \left\|\hat{e} - \frac{1}{i}\sum_{j=n-i+1}^{n} e_{\sigma(j)}\right\|_\infty \le \rho \quad \forall\, i \in \mathbb{N} \cap (n^{3/4}, n]\right\}.$$

Observe that $\sigma_t \in \Sigma_n^\rho$ implies not only that $\|\alpha(\sigma_t) - \hat{e}\|_\infty \le \rho$, but also that as we perform site updates to generate σ_{t+n}, the intermediate configurations are also balanced. To see this, note that when updating σ_j for $j \in [t, t+n]$, the fact that $\sigma_t \in \Sigma_n^\rho$ ensures that the partial configuration in the vertices $v_{j+1-t}, \dots, v_n$ is also balanced; in addition, the partial configuration in $v_1, \dots, v_{j-t-1}$ we can guarantee to be balanced since it was just updated. (The $n^{3/4}$ gives a required slack to make such statement possible; note, for example, that a single vertex cannot be balanced.)

This more fine grained notion of closeness to $\hat{e}$ marks a clear distinction between the Glauber and systematic scan analyses. In the former, given a proportions vector α, the dynamics is oblivious to the specific spin assignment to vertices (i.e., from $\alpha(\sigma_{t-1})$ we can generate $\alpha(\sigma_t)$); as such, the speed of convergence of the Markov chain is governed by its ℓ_∞ closeness to $\hat{e}$. However, this is not the case for systematic scan, since this Markov chain updates vertices in a fixed order and thus depends on the specific assignment of spins to vertices along the scan order (not only on the spin counts). We are therefore required to establish this stronger notion of closeness to $\hat{e}$. We show that the following holds after **Phase 1**.

▶ **Lemma 2.** *For any fixed $\rho > 0$, there exist constants $c > 0$ and $k \in \mathbb{Z}_+$ depending only on β, q, and ρ such that*

$$\Pr\left[\sigma_{kn} \notin \Sigma_n^\rho\right] \le 2qke^{-c\sqrt{n}}. \tag{3}$$

Moreover, if $\sigma_0 \in \Sigma_n^\rho$, for any $\hat{k} \in \mathbb{Z}_+$ we have

$$\Pr\left[\max_{1\le i\le n}\left\|\alpha(\sigma_{(\hat{k}-1)n+i}) - \hat{e}\right\|_\infty > q\rho\right] \le 2q\hat{k}e^{-c\sqrt{n}}. \tag{4}$$

We note that (3) establishes that the chain enters the set Σ_n^ρ after a constant number of scans with high probability, whereas (4) shows that once the chain enters Σ_n^ρ, it will remain also with high probability. To prove (3), we show that the proportions vector has drift towards the $\hat{e}$. More precisely, we analyze the evolution of fraction of vertices in a spin class by decomposing into a drift and a fluctuation component. The drift component of each coordinate is then bounded by a one-dimensional recursive function that iterates towards its fixed point $1/q$. For every fixed constant $\rho > 0$ (independent of n), the recursive function enters the interval $[1/q, 1/q + \rho]$ after a constant number of iterations. To analyze the fluctuation component of the process, we show first that it is a supermartingale, and then we use the Doob decomposition to split it into a predictable non-increasing process and a martingale, with the martingale fluctuations controlled via a concentration of its maximum over an interval; more precisely, via a Maximal-Azuma inequality, a tailored version of which we establish here.

The proof of (4) follows from a similar argument. Once the proportions vector is close to $\hat{e}$ and the configuration has a good spread, the drift helps the process to remain close to $\hat{e}$. Then, the corresponding martingale estimate gives the desired high probability statement. The proof of Lemma 2 is fleshed out in Section 2.

In **Phase 2**, we couple the systematic scan dynamics using the optimal single site coupling; i.e., the coupling that maximizes the probability of agreement between the two copies of the chain at the update site conditional on the configuration off the site. We are able to establish that the Hamming distance d_H between the two chains drops from $O(n)$ to $O(\sqrt{n})$ in $O(\log n)$ steps, and that at the same time, the distance of the spin fraction from $\hat{e}$ drops from $O(1)$ to $O(1/\sqrt{n})$.

▶ **Lemma 3.** *Let $\rho > 0$ small enough and $\sigma_0, \tilde{\sigma}_0 \in \Sigma_n^\rho$. Then, for any fixed constants $C, \varepsilon > 0$, there exist an integer $k = k(\beta, q, C, \varepsilon)$ such that with probability at least ε for $T = c(\beta, q)\log(n) + k$ we have $d_H(\sigma_{Tn}, \tilde{\sigma}_{Tn}) \le C\sqrt{n}$.*

To establish this lemma we show that for two copies of the systematic scan Markov chain with starting configurations in Σ_n^ρ, their Hamming distance exhibits an almost uniform geometric decay. To establish this, we show that the distances between the proportions vectors of the copies and $\hat{e}$ contract at exactly same geometric rate. Identifying these matching contraction rates determines the number of steps required for the Hamming distance to reach the desired scale (i.e., $O(\sqrt{n})$) and determines the leading constant $c(\beta, q)$ in the mixing time. The full proof of Lemma 3 is given in Sections 3, 4, and 5.

For **Phase 3** of the coupling we show that the relative entropy between the distribution of σ_{t+n} and $\tilde{\sigma}_{t+n}$ can be bounded in terms of the square of the Hamming distance between σ_t and $\tilde{\sigma}_t$. Then, since we start the phase from two configurations within $O(\sqrt{n})$ Hamming distance, Pinsker's inequality allows to conclude the following.

▶ **Lemma 4.** *There exists a coupling such that for any $\varepsilon > 0$ there is a constant $C = C(\varepsilon, \beta, q)$ such that for $\sigma_0, \tilde{\sigma}_0 \in \Sigma_n$ with $d_H(\sigma_0, \tilde{\sigma}_0) \le C\sqrt{n}$, we have $\Pr\left[\sigma_n \ne \tilde{\sigma}_n\right] \le \varepsilon$.*

The proof of Lemma 4 is deferred to the full version of the paper [5]. We can now provide the proof of the mixing time upper bound in Theorem 1.

Proof of Theorem 1 (upper bound). We show that $T_{\mathrm{mix}}(\varepsilon) \le c(\beta, q)\log n + \kappa(\varepsilon)$ for any $\varepsilon \in (0, 1)$ and a suitable κ independent of n. By Lemma 2, for any $\rho > 0$ there exists $\kappa_1 = \kappa_1(\beta, q, \rho)$ such that $\sigma_{\kappa_1 n}, \tilde{\sigma}_{\kappa_1 n} \in \Sigma_n^\rho$ with probability at least $1 - \varepsilon/3$. Conditioning on $\sigma_{\kappa_1 n}, \tilde{\sigma}_{\kappa_1 n} \in \Sigma_n^\rho$ and choosing ρ small enough, Lemma 3 implies that we can couple the the two copies of the chain so that for any fixed $C > 0$ there exists $\kappa_2 = \kappa_2(\beta, q, C, \varepsilon)$ such that after $T_2 = c(\beta, q)n\log n + \kappa_2 n + \kappa_1 n$ site updates we have that

$$\Pr\left[d_H(\sigma_{T_2}, \tilde{\sigma}_{T_2}) \le C\sqrt{n}\right] \ge 1 - \varepsilon/3.$$

For the last phase, Lemma 4 implies that $\sigma_{T_2+n} = \tilde{\sigma}_{T_2+n}$ with probability at least $1 - \varepsilon/3$ for small enough C. So at $T = T_2 + n$ we obtain $\Pr[\sigma_T \ne \tilde{\sigma}_T] \le \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon$. ◀

To establish the mixing time lower-bound in (2), we use the fact that the distribution μ is well-concentrated around the equiproportion vector $\hat{e}$. Therefore, if the event $\{\|\alpha(\sigma_{kn})-\hat{e}\|_2 > C/\sqrt{n}\}$ has a significant probability for a sufficiently large constant $C>0$, the systematic scan dynamics chain cannot be mixed after k steps. Using our precise understanding of the rate of contraction towards $\hat{e}$ established in the analysis of **Phase 2**, and standard concentration inequalities, we prove that for any $C>0$ and any $\varepsilon \in (0,1)$, there exists κ independent of n such that after $c(\beta,q)\log n-\kappa(\varepsilon,C)$ steps $\Pr[\|\alpha(\sigma_t)-\hat{e}\|_2 > C/\sqrt{n}] > 1-\varepsilon$ where for any fixed ε the function $\kappa(\varepsilon,\cdot)$ is increasing. The lower bound proof is provided in the full version of the paper [5].

2 Independent Burn-in

Our aim in this section is to analyze the **Phase 1** of our coupling and establish Lemma 2. We analyze the evolution of the number of vertices with a given spin (say spin 1 for concreteness) by decomposing this one-dimensional stochastic process into a mostly predictable component determined by a drift function and a fluctuation component that is controlled via martingale concentration bounds.

2.1 Drift analysis

Consider the drift function $G_\beta : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ given by

$$G_\beta(x) = \frac{e^{\beta x}}{e^{\beta x} + (q-1)e^{\beta \frac{1-x}{q-1}}}.$$

This function approximates the probability that a single-site update chooses spin 1 in a configuration that has xn vertices assigned spin 1 and all other $q-1$ spins distributed equally among the remaining $(1-x)n$ vertices. The correspondence is not exact in part because the spin of the selected vertex is taken into consideration; to account for this correction we define

$$G_{\beta,n}(x) = G_\beta(x) + \beta/n.$$

For $q \geq 2$, the spinodal point β_s can be defined in terms of the drift function G_β:

$$\beta_s = \sup\left\{\beta > 0 : G_\beta(x) - x \neq 0 \quad \forall x \in (1/q, 1)\right\};$$

see [10] for additional details.

We use the function G_β to control the evolution of the one-dimensional process corresponding to fraction of vertices assigned spin 1. To formalize this, let J_t be the random variable corresponding to the spin chosen in the t-th site update and let $X_t = \mathbf{1}(J_t = 1)$. We use the convention that for $-n < t \leq 0$, X_t indicates whether the $(t+n)$-th vertex in the initial configuration σ_0 has spin 1. We define $Y_t^{(\ell)}$ as the average of the X_i's for the last ℓ updates:

$$Y_t^{(\ell)} = \frac{1}{\ell}\sum_{i=t-\ell+1}^{t} X_i. \tag{5}$$

Observe that $Y_t^{(n)} = \alpha_1(\sigma_t)$; to simplify the notation we set $Y_t = Y_t^{(n)}$. We decompose the process $Y_t^{(\ell)}$ into the sum of a drift and a fluctuation component as follows. For all $t \geq \ell$ and $1 \leq \ell \leq n$:

$$Y_t^{(\ell)} = \frac{1}{\ell}\sum_{j=t-\ell+1}^{t} G_{\beta,n}(Y_{j-1}) + \frac{1}{\ell}\sum_{j=t-\ell+1}^{t}\left(X_j - G_{\beta,n}(Y_{j-1})\right),$$

and rewrite the fluctuation term as

$$D_t^{(\ell)} = \sum_{j=t-\ell+1}^{t} \big(X_j - G_{\beta,n}(Y_{j-1})\big), \tag{6}$$

so that

$$Y_t^{(\ell)} = \frac{1}{\ell} D_t^{(\ell)} + \frac{1}{\ell} \sum_{j=t-\ell+1}^{t} G_{\beta,n}(Y_{j-1}).$$

Our first lemma shows that under the assumption that the random fluctuations are small in aggregate over all intervals between kn and $(k+1)n$, then $\max_{n^{3/4}<i\leq n} Y_{kn}^{(i)}$ has a contraction given by the function $G_{\beta,n}$ up to a small correction.

▶ **Lemma 5.** *Suppose that for some integer $k \geq 0$, a fixed constant $a \in (1/q, 1]$, and a sufficiently small $\varepsilon > 0$ that satisfies $G_{\beta,n}(a) + 2\varepsilon < a$, we have $\max\limits_{n^{3/4}<i\leq n} Y_{kn}^{(i)} \leq a$, and $\max\limits_{1\leq i\leq j\leq n} D_{kn+j}^{(i)} < \varepsilon n^{3/4}$. Then,*

$$\max_{n^{3/4}<i\leq n} Y_{(k+1)n}^{(i)} \leq 2\varepsilon + G_{\beta,n}(a).$$

Moreover, $\max\limits_{1\leq i\leq n} Y_{kn+i} \leq a + \frac{1}{n^{1/4}}$.

Proof. We start by rewriting $Y_{kn+\ell}$ using (5) and (6) as

$$\begin{aligned} Y_{kn+\ell} &= \frac{1}{n} \sum_{i=(kn+\ell)-n+1}^{kn} X_i + \frac{1}{n} \sum_{i=kn+1}^{kn+\ell} X_i \\ &= \frac{1}{n} \sum_{i=kn-(n-\ell)+1}^{kn} X_i + \frac{1}{n} \sum_{i=kn+1}^{kn+\ell} \Big(X_i - G_{\beta,n}(Y_{i-1})\Big) + \frac{1}{n} \sum_{i=kn+1}^{kn+\ell} G_{\beta,n}(Y_{i-1}) \\ &= \frac{(n-\ell)}{n} Y_{kn}^{(n-\ell)} + \frac{1}{n} D_{kn+\ell}^{(\ell)} + \frac{\ell}{n}\frac{1}{\ell} \sum_{i=kn+1}^{kn+\ell} G_{\beta,n}(Y_{i-1}) \\ &= \frac{1}{n} D_{kn+\ell}^{(\ell)} + \frac{1}{n}\Bigg((n-\ell)Y_{kn}^{(n-\ell)} + \ell\Big(\frac{1}{\ell} \sum_{i=kn+1}^{kn+\ell} G_{\beta,n}(Y_{i-1})\Big)\Bigg). \end{aligned} \tag{7}$$

We show first via induction on ℓ that

$$Y_{kn+\ell} \leq a + \frac{1}{n^{1/4}}. \tag{8}$$

for all $\ell = 1, \dots, n$. This ensures that $Y_{kn+\ell}$ does not deviate much from its initial upper bound of a on the process. Conditioned on this, we can establish the desired contraction $\max_{n^{3/4}<i\leq n} Y_{(k+1)n}^{(i)} \leq 2\varepsilon + G_{\beta,n}(a)$.

For $\ell = 1$, under the assumptions of the lemma, and using the fact that $G_{\beta,n}$ is a monotonically increasing function, we obtain from (7)

$$Y_{kn+1} < \frac{\varepsilon n^{3/4}}{n} + \frac{1}{n}\big((n-1)a + G_{\beta,n}(a)\big) \leq \frac{\varepsilon}{n^{1/4}} + a.$$

For the induction step we consider two cases. First assume that (8) holds for all $\ell \in \{1, \dots, \ell_1 - 1\}$ for some $\ell_1 \in \{1, \dots, n - n^{3/4} - 1\}$. From (7), the monotonicity of $G_{\beta,n}$, the lemma assumptions, and the inductive hypothesis, we get

$$\begin{aligned} Y_{kn+\ell_1} &= \frac{1}{n} D^{(\ell_1)}_{kn+\ell_1} + \frac{1}{n}\Big((n-\ell_1)Y^{(n-\ell_1)}_{kn} + \ell_1\Big(\frac{1}{\ell_1}\sum_{i=kn+1}^{kn+\ell_1} G_{\beta,n}(Y_{i-1})\Big)\Big) \\ &< \frac{\varepsilon}{n^{1/4}} + \frac{1}{n}\Big((n-\ell_1)a + \ell_1 G_{\beta,n}\Big(a + \frac{1}{n^{1/4}}\Big)\Big). \end{aligned}$$

Now, by the mean value theorem, since the derivative of $G_{\beta,n}$ is bounded in the interval $\left[a, a + \frac{1}{n^{1/4}}\right]$, we have $G_{\beta,n}(a + \frac{1}{n^{1/4}}) = G_{\beta,n}(a) + O(\frac{1}{n^{1/4}}) \le a - \varepsilon$ for large enough n. Then,

$$\begin{aligned} Y_{kn+\ell_1} &\le \frac{\varepsilon}{n^{1/4}} + \frac{1}{n}\Big((n-\ell_1)a + \ell_1(a-\varepsilon)\Big) \\ &\le \frac{\varepsilon}{n^{1/4}} + \frac{1}{n}\big(na - \ell_1\varepsilon\big) = a + \frac{\varepsilon}{n^{1/4}} - \frac{\ell_1}{n}\varepsilon \le a + \frac{1}{n^{1/4}}. \end{aligned}$$

Thus, the claim holds for all $1 \le \ell < n - n^{3/4}$. For the remaining range, let us assume that the claim holds for all $\ell \in \{1, \dots, \ell_2 - 1\}$ for some $\ell_2 \in \{n - n^{3/4} - 1, \dots, n\}$. Then, similarly

$$\begin{aligned} Y_{kn+\ell_2} &= \frac{1}{n} D^{(\ell_2)}_{kn+\ell_2} + \frac{1}{n}\Big((n-\ell_2)Y^{(n-\ell_2)}_{kn} + \ell_2\Big(\frac{1}{\ell_2}\sum_{i=kn+1}^{kn+\ell_2} G_{\beta,n}(Y_{i-1})\Big)\Big) \\ &< \frac{\varepsilon}{n^{1/4}} + \frac{1}{n}\Big(n - \ell_2 + \ell_2 G_{\beta,n}\Big(a + \frac{1}{n^{1/4}}\Big)\Big) \\ &\le \frac{\varepsilon}{n^{1/4}} + \frac{1}{n}\Big((n^{3/4}+1) + n(a-\varepsilon)\Big) = a - \varepsilon + \frac{\varepsilon}{n^{1/4}} + \frac{1}{n^{1/4}} + \frac{1}{n} \le a, \end{aligned}$$

and thus (8) follows.

To establish that $\max_{n^{3/4} < i \le n} Y^{(i)}_{(k+1)n} \le 2\varepsilon + G_{\beta,n}(a)$, observe that the lemma assumptions and (8) imply for all $\ell \in \{n^{3/4}+1, \dots, n\}$,

$$\begin{aligned} Y^{(\ell)}_{(k+1)n} &= \frac{1}{\ell} D^{(\ell)}_{(k+1)n} + \frac{1}{\ell}\sum_{j=(k+1)n-\ell+1}^{(k+1)n} G_{\beta,n}(Y_{j-1}) \\ &< \frac{\varepsilon n^{3/4}}{\ell} + \frac{1}{\ell}\sum_{j=(k+1)n-\ell+1}^{(k+1)n} G_{\beta,n}\Big(a + \frac{1}{n^{1/4}}\Big) \le 2\varepsilon + G_{\beta,n}(a), \end{aligned}$$

which proves our claim. ◀

2.2 Fluctuation analysis

Our goal now is to bound the fluctuations $D^{(\ell)}_t$ of the $Y^{(\ell)}_t$ process. For this, consider the process

$$M_t = \sum_{i=1}^{t} (X_i - G_{\beta,n}(Y_{i-1}));$$

observe that $D^{(\ell)}_t = M_t - M_{t-\ell}$. We establish first that the process M_t is a supermartingale.

▶ **Lemma 6.** *$\{M_t\}_{t>0}$ is a supermartingale.*

The proof of this lemma is provided in Section 2.4. We bound the maximum supermartingale difference $M_i - M_j$ over a time interval using a Maximal-Azuma inequality. We use a small variation of a version of this type inequality that appears in [28]. A proof is provided in the full version of the paper [5] for completeness.

▶ **Lemma 7.** *Let $\{Z_t\}_{t\geq 1}$ be a bounded discrete-time supermartingale. Suppose there exists $R \in \mathbb{R}_+$ such that $|Z_{t+1} - Z_t| \leq R$ for all integer $t \geq 1$. Then, for any positive integers t_1, t_2 such that $1 \leq t_1 < t_2$ and $d > 0$ we have*

$$\Pr\left[\max_{t_1 \leq i < j \leq t_2} Z_j - Z_i > d\right] \leq 2e^{-\frac{d^2}{32(t_2-t_1)R^2}}.$$

2.3 Combining drift and fluctuation analysis

The next ingredient of the proof is to argue that repeated applications of Lemma 5 decrease the distance to equiproportionality. For this, we introduce the function $G^{\varepsilon}_{\beta,n} : \mathbb{Z}_+ \to \mathbb{R}_+$ defined recursively by setting $G^{\varepsilon}_{\beta,n}(1) = 1 - \varepsilon$, and for all integer $k \geq 2$:

$$G^{\varepsilon}_{\beta,n}(k) = G_{\beta,n}\left(G^{\varepsilon}_{\beta,n}(k-1) + \varepsilon\right). \tag{9}$$

The function $G^{\varepsilon}_{\beta,n}(k)$ is monotonically decreasing with k and for any $\rho > 0$, it enters the interval $[1/q, 1/q + \rho]$ after constant number of recursive steps. The latter is established in the following lemma which is proved in the full version of the paper [5].

▶ **Lemma 8.** *For any fixed $\rho > 0$, there exist constants $k \in \mathbb{Z}_+$ and $\varepsilon > 0$ such that $G^{\varepsilon}_{\beta,n}(k) + \varepsilon < \frac{1}{q} + \rho$.*

We use the recursive function $G^{\varepsilon}_{\beta,n}$ to probabilistically bound the process $Y^{(\ell)}_{kn}$.

▶ **Lemma 9.** *For any integer $k > 0$ and sufficiently small $\varepsilon > 0$, we have*

$$\Pr\left[\max_{n^{3/4} < i \leq n} Y^{(i)}_{kn} > G^{\varepsilon}_{\beta,n}(k) + \varepsilon\right] \leq 2ke^{-\frac{\varepsilon^2\sqrt{n}}{128}}.$$

Proof. Consider the events

$$A_k = \left\{\max_{n^{3/4} < i \leq n} Y^{(i)}_{kn} \leq G^{\varepsilon}_{\beta,n}(k) + \varepsilon\right\}, \text{ and}$$
$$B_k = \left\{\max_{1 \leq i \leq j \leq n} D^{(i)}_{(k-1)n+j} \leq \frac{\varepsilon n^{3/4}}{2}\right\}.$$

For $k > 1$ we have

$$\Pr[\overline{A_k}] \leq \Pr\left[\overline{A_k} \mid A_{k-1}\right] + \Pr[\overline{A}_{k-1}] \leq \Pr[\overline{A}_1] + \sum_{\ell=2}^{k} \Pr\left[\overline{A}_\ell \mid A_{\ell-1}\right]. \tag{10}$$

Since

$$\Pr[\overline{A}_1] = \Pr\left[\max_{n^{3/4} < i \leq n} Y^{(i)}_{n} > G^{\varepsilon}_{\beta,n}(1) + \varepsilon\right] = \Pr\left[\max_{n^{3/4} < i \leq n} Y^{(i)}_{n} > 1\right] = 0,$$

it suffices to prove that

$$\Pr[\overline{A_\ell} \mid A_{\ell-1}] \leq 2e^{-\frac{\varepsilon^2\sqrt{n}}{128}},$$

for all $\ell \in \{2, \dots, k\}$. For any such ℓ, using the definition of $G^{\varepsilon}_{\beta,n}$ in (9) and taking $a = G^{\varepsilon}_{\beta,n}(\ell - 1) + \varepsilon$, we obtain from Lemma 5 that $A_{\ell-1} \cap B_\ell \subseteq A_\ell$ and thus $\overline{A}_\ell \subseteq \overline{A}_{\ell-1} \cup \overline{B}_\ell$. Then,

$$\begin{aligned}\Pr\left[\overline{A}_\ell \mid A_{\ell-1}\right] &\le \Pr\left[\overline{A}_{\ell-1} \cup \overline{B}_\ell \mid A_{\ell-1}\right] \le \Pr\left[\overline{B}_\ell \mid A_{\ell-1}\right] \\ &= \Pr\left[\max_{1\le i\le j\le n} M_{(\ell-1)n+j} - M_{(\ell-1)n+j-i} > \frac{\varepsilon n^{3/4}}{2} \,\middle|\, A_{\ell-1}\right]. \end{aligned} \tag{11}$$

We consider the following subset of configurations,

$$\Sigma^a_{n,1} = \left\{\sigma \in \Sigma_n : \frac{1}{i}\sum_{j=n-i+1}^{n} \mathbf{1}\big(\sigma(j) = 1\big) \le a \quad \forall i \in \mathbb{N} \cap (n^{3/4}, n]\right\}.$$

It is easy to verify that $A_{\ell-1} = \{\sigma_{(\ell-1)n} \in \Sigma^a_{n,1}\}$. We rewrite (11) as

$$\begin{aligned}&\Pr\left[\overline{A}_\ell \mid A_{\ell-1}\right] \\ &\le \Pr\left[\max_{1\le i\le j\le n} M_{(\ell-1)n+j} - M_{(\ell-1)n+j-i} > \frac{\varepsilon n^{3/4}}{2} \,\middle|\, A_{\ell-1}\right] \\ &\le \max_{\sigma\in\Sigma^a_{n,1}} \Pr\left[\max_{1\le i\le j\le n} M_{(\ell-1)n+j} - M_{(\ell-1)n+j-i} > \frac{\varepsilon n^{3/4}}{2} \,\middle|\, \sigma_{(\ell-1)n} = \sigma\right].\end{aligned}$$

By the Markov property and Lemma 6, $\{M_t\}_{t\ge(\ell-1)n}$ conditioned on $\sigma_{(\ell-1)n}$ is a supermartingale with $|M_{t+1} - M_t| \le 1$. Lemma 7 implies that

$$\Pr\left[\overline{A}_\ell \mid A_{\ell-1}\right] \le 2e^{-\frac{(\varepsilon n^{3/4})^2}{128 n}} = 2e^{-\frac{\varepsilon^2\sqrt{n}}{128}},$$

which yields the result. ◀

We are now ready to prove Lemma 2.

Proof of Lemma 2. By Lemma 8 and 9, for any $\hat{\rho} > 0$, there exist an integer k and $\varepsilon > 0$ such that

$$\Pr\left[\max_{n^{3/4}<i\le n} Y^{(i)}_{kn} > \frac{1}{q} + \hat{\rho}\right] \le 2ke^{-\frac{\varepsilon^2\sqrt{n}}{128}}.$$

Recall that $Y^{(i)}_{kn}$ denotes the fraction of vertices assigned the spin 1, but this bound holds by symmetry for all spins in Q. So, setting

$$\Sigma^{\rho+}_n = \left\{\sigma \in \Sigma_n : \frac{1}{i}\sum_{j=n-i+1}^{n} \mathbf{1}\big(\sigma(j) = \kappa\big) \le \frac{1}{q} + \rho \quad \forall i \in \mathbb{N} \cap (n^{3/4}, n] \text{ and } \forall \kappa \in Q\right\},$$

a union bound over the spins implies

$$\Pr\left[\sigma_{kn} \notin \Sigma^{\hat{\rho}q}_n\right] \le \Pr\left[\sigma_{kn} \notin \Sigma^{\hat{\rho}+}_n\right] \le q \Pr\left[\max_{n^{3/4}<i\le n} Y^{(i)}_{kn} > \frac{1}{q} + \hat{\rho}\right] \le 2qke^{-\frac{\varepsilon^2\sqrt{n}}{128}};$$

taking $\rho = \hat{\rho}q$ establishes (3).

To prove (4), observe that the initial configuration $\sigma_0 \in \Sigma^{\rho}_n \subseteq \Sigma^{\rho+}_n$. Then, as in (10), for any integer $k > 0$

$$\Pr\left[\sigma_{kn} \notin \Sigma^{\rho+}_n\right] \le \sum_{\ell=0}^{k-1} \Pr\left[\sigma_{(\ell+1)n} \notin \Sigma^{\rho+}_n \mid \sigma_{\ell n} \in \Sigma^{\rho+}_n\right].$$

Given that $\sigma_{\ell n} \in \Sigma_n^{\rho+}$, the configuration at the next step can leave the set $\Sigma_n^{\rho+}$ only if the fluctuation is large. In particular, we obtain from Lemma 5 and a union bound over the spins that

$$\Pr\left[\sigma_{kn} \notin \Sigma_n^{\rho+}\right] \leq q \sum_{\ell=0}^{k-1} \Pr\left[\max_{1 \leq j' < j \leq n} M_{\ell n+j} - M_{\ell n+j-j'} > \frac{\varepsilon n^{3/4}}{2} \,\Big|\, \sigma_{\ell n} \in \Sigma_n^{\rho+}\right].$$

Recall that $\{M_t\}_{t \geq \ell n}$ is a supermartingale (see Lemma 6), and thus the Maximal-Azuma inequality in Lemma 7 yields that

$$\Pr\left[\sigma_{kn} \notin \Sigma_n^{\rho+}\right] \leq 2qke^{-\frac{(\varepsilon n^{3/4})^2}{128n}} = 2qke^{-\frac{\varepsilon^2\sqrt{n}}{128}}.$$

Using this bound, we obtain for any integer $\hat{k} > 0$

$$\begin{aligned}
&\Pr\left[\max_{1 \leq i \leq n} \left\|\alpha(\sigma_{(\hat{k}-1)n+i}) - \hat{e}\right\|_\infty > \rho q\right] \\
&\leq \Pr\left[\max_{1 \leq i \leq n} \left\|\alpha(\sigma_{(\hat{k}-1)n+i}) - \hat{e}\right\|_\infty > \rho q \,\Big|\, \sigma_{(\hat{k}-1)n} \in \Sigma_n^{\rho+}\right] + \Pr\left[\sigma_{(\hat{k}-1)n} \notin \Sigma_n^{\rho+}\right] \\
&\leq q \Pr\left[\max_{1 \leq i \leq n} \alpha_1\big(\sigma_{(\hat{k}-1)n+i}\big) - 1/q > \rho + \frac{1}{n^{1/4}} \,\Big|\, \sigma_{(\hat{k}-1)n} \in \Sigma_n^{\rho+}\right] + 2q(\hat{k}-1)e^{-\frac{\varepsilon^2\sqrt{n}}{128}} \\
&\leq q \max_{\sigma \in \Sigma_n^{\rho+}} \Pr\left[\max_{1 \leq i \leq n} \alpha_1\big(\sigma_{(\hat{k}-1)n+i}\big) - 1/q > \rho + \frac{1}{n^{1/4}} \,\Big|\, \sigma_{(\hat{k}-1)n} = \sigma\right] + 2q(\hat{k}-1)e^{-\frac{\varepsilon^2\sqrt{n}}{128}},
\end{aligned} \tag{12}$$

where the second inequality is obtained by taking union bound over each spin. We consider the following events,

$$\begin{aligned}
E_{\hat{k}} &= \left\{\max_{n^{3/4} < i \leq n} Y^{(i)}_{(\hat{k}-1)n} \leq \frac{1}{q} + \rho\right\}, \\
B_{\hat{k}} &= \left\{\max_{1 \leq i \leq j \leq n} D^{(i)}_{(\hat{k}-1)n+j} < \frac{\varepsilon n^{3/4}}{2}\right\}, \text{ and} \\
C_{\hat{k}} &= \left\{\max_{1 \leq i \leq n} Y_{(\hat{k}-1)n+i} \leq \frac{1}{q} + \rho + \frac{1}{n^{1/4}}\right\}.
\end{aligned}$$

Lemma 5 implies that $\overline{C}_{\hat{k}} \subseteq \overline{E}_{\hat{k}} \cup \overline{B}_{\hat{k}}$. We recall that $D_t^\ell = M_t - M_{t-\ell}$ for any positive integer t and integer ℓ such that $1 \leq \ell \leq \min\{t, n\}$. Using the fact $\left\{\sigma_{(\hat{k}-1)n} \in \Sigma_n^{\rho+}\right\} \subseteq \overline{E}_{\hat{k}}$, we obtain for any $\sigma \in \Sigma_n^{\rho+}$

$$\begin{aligned}
\Pr\left[\max_{1 \leq i \leq n} \alpha_1\big(\sigma_{(\hat{k}-1)n+i}\big) - 1/q > \rho + \frac{1}{n^{1/4}} \,\Big|\, \sigma_{(\hat{k}-1)n} = \sigma\right] &= \Pr\left[\overline{C}_{\hat{k}} \,\Big|\, \sigma_{(\hat{k}-1)n} = \sigma\right] \\
\leq \Pr\left[\overline{E}_{\hat{k}} \,\Big|\, \sigma_{(\hat{k}-1)n} = \sigma\right] + \Pr\left[\overline{B}_{\hat{k}} \,\Big|\, \sigma_{(\hat{k}-1)n} = \sigma\right] \\
\leq \Pr\left[\max_{1 \leq i < j \leq n} M_{(\hat{k}-1)n+j} - M_{(\hat{k}-1)n+j-i} > \frac{\varepsilon n^{3/4}}{2} \Big| \sigma_{(\hat{k}-1)n} = \sigma\right] \\
\leq 2e^{-\frac{\varepsilon^2\sqrt{n}}{128}},
\end{aligned}$$

where we apply Lemmas 6 and 7 in the last inequality. Plugging this bound into (12) yields the result. ◀

2.4 Proofs of auxiliary Facts

In this section, we complete the proof of Lemma 2 by providing the proofs of Lemmas 6 and 8. Let us introduce some notation first. Let

$$\mathcal{S}_q = \Big\{x \in \mathbb{R}^q_{\geq 0} : \sum_{i=1}^{q} x_i = 1\Big\},$$

$$\tilde{\mathcal{S}}_q = \Big\{x \in \mathbb{R}^q_{\geq 0} : \sum_{i=1}^{q} x_i = 1 - \frac{1}{n}\Big\}, \text{ and}$$

$$\mathcal{S}'_q = \Big\{x \in \mathbb{R}^q_{+} : \sum_{j=1}^{q} x_j = 1 \text{ and } x_i \geq \frac{1}{e^\beta + (q-1)} \quad \forall i \in Q\Big\}.$$

We define the function $g_\beta : \mathcal{S}_q \cup \tilde{\mathcal{S}}_q \to \mathcal{S}'_q$, as $g_\beta(s) = \big(g^1_\beta(s), \cdots, g^q_\beta(s)\big)$ where

$$g^j_\beta(s) = \frac{e^{\beta s_j}}{\sum_{i=1}^{q} e^{\beta s_i}}.$$

For $i, j \in Q$, let $g^{i\to j}_\beta : \mathcal{S}_q \to \mathcal{S}'_q$ be given by

$$g^{i\to j}_\beta(s) = g^j_\beta\Big(s - \frac{1}{n} e_i\Big),$$

where recall that $e_i \in \{0,1\}^q$ is the indicator vector with 1 in the i-th coordinate and 0 elsewhere. This function corresponds to the probability that a site update flips the vertex from spin i to j from configuration with proportion vector s. We establish the following inequality involving $g^{i\to j}_\beta$ and $G_{\beta,n}$; its proof is given in the full version of the paper [5].

▶ **Lemma 10.** *For $s \in \mathcal{S}_q$ and $i, j \in Q$, we have*

$$g^{i\to j}_\beta(s) \leq g^j_\beta(s) + \frac{\beta}{n} \leq G_{\beta,n}(s_j).$$

We can now provide the proof of Lemma 6.

Proof of Lemma 6. Let $t \geq 1$ be a integer and let j be the spin of the vertex to be update updated at time $t+1$. Then,

$$\begin{aligned}\mathbb{E}[M_{t+1} - M_t | \mathcal{F}_t] &= \mathbb{E}[X_{t+1} - G_{\beta,n}(Y_t) | \mathcal{F}_t] = \mathbb{E}[X_{t+1} | \mathcal{F}_t] - G_{\beta,n}(Y_t) \\ &= g^{j\to 1}_\beta\big(\alpha(\sigma_t)\big) - G_{\beta,n}(Y_t) \leq 0,\end{aligned}$$

where the last inequality follows from Lemma 10. ◀

3 Spin fraction analysis

In this section, we begin our analysis of Phase 2 of the coupling. In particular, we analyze rate of convergence of the systematic scan to a $O(1/\sqrt{n})$ neighborhood of $\hat{e}$ from a configuration $\sigma_0 \in \Sigma^\rho_n$. We start with the following bounds for the function g^ℓ_β which follow from its Taylor expansion; their proofs are provided in the full version of the paper [5].

▶ **Lemma 11.** *For any $\ell \in Q$ and $s, \tilde{s} \in \mathcal{S}_q$,*

(i) $\exists\ R : \mathcal{S}_q \to \mathbb{R}$ *such that* $g^\ell_\beta(s) = \frac{1}{q} + \frac{\beta}{q}\Big(s_\ell - \frac{1}{q}\Big) + R(s)$ *and* $|R(s)| \leq 3q\beta^2 \|s - \hat{e}\|_2^2$ *;*

(ii) $\big\|g_\beta(s) - g_\beta(\tilde{s})\big\|_1 = \frac{\beta}{q}\big\|s - \tilde{s}\big\|_1 + \|s - \tilde{s}\|_1 \Big\{O\big(\|s - \hat{e}\|_1 + \|\tilde{s} - \hat{e}\|_1\big)\Big\}.$

Now, for a vector $X = (X_1, \dots, X_q) \in \mathbb{R}^q$, let

$$\mathrm{Var}(X) = \mathbb{E}\big[\|X - \mathbb{E}X\|_2^2\big] = \sum_{i=1}^{q} \mathbb{E}\big[\big(X_i - \mathbb{E}X_i\big)^2\big] = \sum_{i=1}^{q} \mathrm{Var}(X_i).$$

We prove the following bound on the variance of the q-dimensional process $\alpha(\sigma_t)$, assuming that $\sigma_0 \in \Sigma_n^\rho$; its proof is provided in the full version of the paper [5].

▶ **Lemma 12.** *There exists a constant $\rho = \rho(\beta, q) > 0$ such that for all $0 \le t \le n\sqrt{n}$ and $\sigma_0 \in \Sigma_n^\rho$, we have $\mathrm{Var}\big(\alpha(\sigma_t)\big) = O\big(1/n\big)$.*

We also require a tight control on the following recurrence; its analysis is fleshed out in Section 4.

▶ **Lemma 13.** *Let $\{x_t = (x_t^{(1)}, \dots, x_t^{(q)})\}_{t\ge0}$ be a sequence of vectors in $\mathbb{R}^q_{\ge 0}$ such that for all $j \in Q$ and $t \ge n$, we have the recursion*

$$x_t^{(j)} = \frac{\beta}{nq} \sum_{i=t-n}^{t-1} x_i^{(j)} + \frac{C}{n} \sum_{i=t-n}^{t-1} \|x_i\|_2^2 + \frac{C}{n},$$

where C is a positive constant. Then, if for all $i \in \{0, \dots, n-1\}$ and all $j \in Q$, we have $|x_i^{(j)}| \le \rho$ for a small enough constant $\rho > 0$, then there exists constant $\hat{C} > 0$ such that $\|x_t\|_2^2 \le \hat{C}\rho^2\gamma_n^{2t} + \hat{C}/n$, where γ_n is the unique positive real root of the equation $\sum_{i=1}^{n} \gamma_n^{-i} = nq/\beta$.

We combine the three lemmas above to obtain the correct rate of contraction towards $\hat{e}$ of the proportions vector under the systematic scan dynamics. Let

$$u_t = \begin{cases} e_{J_t} & \text{if } t \ge 1 \\ e_{\sigma_0(n+t)} & \text{if } -n < t \le 0, \end{cases}$$

where recall that J_t denotes the random variable corresponding to the spin chosen in the t-th site update. With this notation, we have

$$\alpha(\sigma_t) = \frac{1}{n} \sum_{i=t-n+1}^{t} u_i. \tag{13}$$

▶ **Lemma 14.** *Let γ_n be the unique positive real root of the equation $\sum_{i=1}^{n} \gamma_n^{-i} = nq/\beta$. Then, for all integer $t \ge 0$ such that $t \le n\sqrt{n}$ there exist constants $\rho = \rho(\beta, q)$ and $C = C(\rho, \beta, q)$ such that for any $\sigma_0 \in \Sigma_n^\rho$*

$$\mathbb{E}\big[\|\alpha(\sigma_t) - \hat{e}\|_2^2\big] \le C\rho^2\gamma_n^{2t} + \frac{C}{n}.$$

Proof. We let $t_m = t_m(t) = \max\{1, t - n + 1\}$. From (13), for all $t \ge 0$ and $\ell \in Q$, we have

$$\alpha_\ell(\sigma_t) = \frac{1}{n} \sum_{i=t-n+1}^{0} \mathbf{1}\big(\sigma_0(n+i) = \ell\big) + \frac{1}{n} \sum_{i=t_m}^{t} \mathbf{1}\big(J_i = \ell\big),$$

and thus

$$\begin{aligned} \mathbb{E}\big[\alpha_\ell(\sigma_t)\big] &= \frac{1}{n} \sum_{i=t-n+1}^{0} \mathbf{1}\big(\sigma_0(n+i) = \ell\big) + \frac{1}{n} \sum_{i=t_m}^{t} \mathbb{E}\big[\mathbf{1}(J_i = \ell)\big] \\ &= \frac{1}{n} \sum_{i=t-n+1}^{0} \mathbf{1}\big(\sigma_0(n+i) = \ell\big) + \frac{1}{n} \sum_{i=t_m}^{t} \mathbb{E}\Big[g_\beta^\ell\Big(\alpha(\sigma_{i-1}) - \frac{1}{n} u_{i-n}\Big)\Big]. \end{aligned} \tag{14}$$

From Lemmas 10 and 11(i), we obtain

$$\begin{aligned}\mathbb{E}\Big[g_\beta^\ell\Big(\alpha(\sigma_{i-1})-\frac{1}{n}u_{i-n}\Big)\Big] \le& \frac{\beta}{n}+\mathbb{E}\Big[g_\beta^\ell\big(\alpha(\sigma_{i-1})\big)\Big]\\ \le& \frac{\beta}{n}+\frac{1}{q}+\frac{\beta}{q}\mathbb{E}\Big[\alpha_\ell(\sigma_{i-1})-\frac{1}{q}\Big]+3q\beta^2\mathbb{E}\big[\|\alpha(\sigma_i)-\hat{e}\|_2^2\big]\\ \le& \frac{\beta}{n}+3q\beta^2\mathrm{Var}\big(\alpha(\sigma_i)\big)+\frac{1}{q}+\frac{\beta}{q}\mathbb{E}\Big[\alpha_\ell(\sigma_{i-1})-\frac{1}{q}\Big]\\ &+3q\beta^2\big\|\mathbb{E}[\alpha(\sigma_i)-\hat{e}]\big\|_2^2.\end{aligned}$$

By Lemma 12, there exists a constant $C_1>0$ such that the following holds for all $t\in\{0,\cdots,n\sqrt{n}\}$

$$\mathbb{E}\Big[g_\beta^\ell\Big(\alpha(\sigma_{i-1})-\frac{1}{n}u_{i-n}\Big)\Big]\le\frac{C_1}{n}+\frac{1}{q}+\frac{\beta}{q}\mathbb{E}\Big[\alpha_\ell(\sigma_{i-1})-\frac{1}{q}\Big]+3q\beta^2\big\|\mathbb{E}[\alpha(\sigma_i)-\hat{e}]\big\|_2^2.$$

Plugging the above inequality into (14)

$$\begin{aligned}\mathbb{E}\big[\alpha_\ell(\sigma_t)\big]\le&\frac{C_1}{n}+\frac{1}{n}\sum_{i=t-n+1}^{0}\Big[\mathbf{1}\big(\sigma_0(n+i)=\ell\big)\Big]\\ &+\frac{t-t_m+1}{nq}+\frac{1}{n}\sum_{i=t_m}^{t}\frac{\beta}{q}\mathbb{E}\Big[\alpha_\ell(\sigma_{i-1})-\frac{1}{q}\Big]+\frac{3q\beta^2}{n}\sum_{i=t_m}^{t}\big\|\mathbb{E}[\alpha(\sigma_{i-1})-\hat{e}]\big\|_2^2,\end{aligned}$$

and subtracting $1/q$ from both sides yields,

$$\begin{aligned}\mathbb{E}\Big[\alpha_\ell(\sigma_t)-\frac{1}{q}\Big]\le&\frac{C_1}{n}+\frac{1}{n}\sum_{i=t-n+1}^{0}\Big[\mathbf{1}\big(\sigma_0(n+i)=\ell\big)-\frac{1}{q}\Big]\\ &+\frac{1}{n}\sum_{i=t_m}^{t}\frac{\beta}{q}\mathbb{E}\Big[\alpha_\ell(\sigma_{i-1})-\frac{1}{q}\Big]+\frac{3q\beta^2}{n}\sum_{i=t_m}^{t}\big\|\mathbb{E}[\alpha(\sigma_{i-1})-\hat{e}]\big\|_2^2.\end{aligned}\qquad(15)$$

Let $\{x_t\}_{t\ge0}$ be a sequence of vectors in $\mathbb{R}^q$ such that for all $j\in Q$ and $t\ge n$, we have the recursion

$$x_t^{(j)}=\frac{\beta}{nq}\sum_{i=t-n}^{t-1}x_i^{(j)}+\frac{C_2}{n}\sum_{i=t-n}^{t-1}\|x_i\|_2^2+\frac{C_2}{n}.$$

where $C_2=\max\{C_1,3q\beta^2\}$. For all $i\in\{0,\dots,n-1\}$ and all $j\in Q$, $x_i^{(j)}=2\rho$. We prove by induction on t that for every $\ell\in Q$ and $t\in\{0,\cdots,n\sqrt{n}\}$, the following holds

$$\Big|\mathbb{E}\Big[\alpha_\ell(\sigma_t)-\frac{1}{q}\Big]\Big|\le x_t^{(\ell)}.\qquad(16)$$

For $t=0$, the claim trivially holds since $\sigma_0\in\Sigma_n^\rho$ and therefore $|\alpha_\ell(\sigma_0)-1/q|\le\rho<2\rho$. We consider three cases for the inductive step. We assume first that the claim holds for all $t\in\{1,\cdots,\kappa-1\}$ for some $\kappa\in\{1,\cdots,n-n^{3/4}-1\}$. For $t=\kappa$, we obtain from (15) and the triangle inequality that

$$\begin{aligned}\Big|\mathbb{E}\Big[\alpha_\ell(\sigma_t)-\frac{1}{q}\Big]\Big|\le&\frac{C_1}{n}+\frac{1}{n}\Big|\sum_{i=t-n+1}^{0}\Big[\mathbf{1}\big(\sigma_0(n+i)=\ell\big)-\frac{1}{q}\Big]\Big|\\ &+\frac{\beta}{nq}\sum_{i=t_m}^{t}\Big|\mathbb{E}\Big[\alpha_\ell(\sigma_{i-1})-\frac{1}{q}\Big]\Big|+\frac{3q\beta^2}{n}\sum_{i=t_m}^{t}\big\|\mathbb{E}[\alpha(\sigma_{i-1})-\hat{e}]\big\|_2^2.\end{aligned}$$

Since $\sigma_0 \in \Sigma_n^\rho$, we observe that $\left|\sum_{i=t-n+1}^{0}\left[\mathbf{1}\left(\sigma_0(n+i)=\ell\right)-\frac{1}{q}\right]\right| \le (n-t)\rho$, using the inductive hypothesis to bound the second and third sums we obtain

$$\begin{aligned}\left|\mathbb{E}\left[\alpha_\ell(\sigma_t)-\frac{1}{q}\right]\right| &\le \frac{C_1}{n}+\frac{(n-t)\rho}{n}+\frac{\beta}{nq}(2t\rho)+\frac{12q\beta^2}{n}(tq\rho^2)\\ &= \frac{C_1}{n}+\rho\Big[1+\frac{t}{n}\Big((2\beta/q)-1+12q^2\beta^2\rho\Big)\Big].\end{aligned}$$

Set $1-\beta/q=\delta$ so that $2\beta/q-1=1-2\delta$ (recall that $\beta/q<1$). Since $12q^2\beta^2$ is a fixed constant, we can choose ρ small enough so that $12q^2\beta^2\rho<\delta$. Plugging these into the above inequality:

$$\left|\mathbb{E}\left[\alpha_\ell(\sigma_t)-\frac{1}{q}\right]\right| \le \frac{C_1}{n}+\rho\Big[1+(1-\delta)\Big]=2\rho+\Big(\frac{C_1}{n}+-\rho\delta\Big)\le 2\rho\le x_t^{(\ell)}$$

provided n is large enough. For the second case, let us assume that the claim holds for all $t\in\{1,\cdots,\kappa-1\}$ for some $\kappa\in\{n-n^{3/4},\cdots,n-1\}$. For $t=\kappa$:

$$\begin{aligned}\left|\mathbb{E}\left[\alpha_\ell(\sigma_t)-\frac{1}{q}\right]\right| &\le \frac{C_1}{n}+\frac{n^{3/4}}{n}+\frac{2\beta t\rho}{nq}+\frac{12q^2\beta^2t\rho^2}{n}\le\frac{C_1}{n}+\frac{1}{n^{1/4}}+\rho\Big(\frac{2\beta}{q}+12q^2\beta^2\rho\Big)\\ &\le 2\rho+\Big(\frac{C_1}{n}+\frac{1}{n^{1/4}}-\rho\delta\Big)\le 2\rho\le x_t^{(\ell)}.\end{aligned}$$

For the third case, We assume that the claim holds for all $t\in\{0,\cdots,\kappa-1\}$ for some $\kappa\in\{n,\cdots,n\sqrt{n}\}$. For $t=\kappa$,

$$\begin{aligned}\left|\mathbb{E}\left[\alpha_\ell(\sigma_t)-\frac{1}{q}\right]\right| &\le \frac{C_1}{n}+\frac{\beta}{qn}\left|\sum_{i=t-n}^{t-1}\frac{\beta}{q}\mathbb{E}\Big[\alpha_\ell(\sigma_i)-\frac{1}{q}\Big]\right|+\frac{3q\beta^2}{n}\sum_{i=t-n}^{t-1}\left\|\mathbb{E}[\alpha(\sigma_i)-\hat{e}]\right\|_2^2\\ &\le \frac{C_2}{n}+\frac{\beta}{qn}\sum_{i=t-n}^{t-1}x_i^{(\ell)}+\frac{C_2}{n}\sum_{i=t-n}^{t-1}\|x_i\|_2^2=x_t^{(\ell)},\end{aligned}$$

which proves the claim (16). Using the Lemma 13, we show that the following holds for some constant $C_3>0$

$$\left\|\mathbb{E}\big[\alpha(\sigma_t)-\hat{e}\big]\right\|_2^2\le\|x_t\|_2^2\le C_3\rho^2\gamma_n^{2t}+\frac{C_3}{n}.$$

Using Lemma 12 and the above equation, we prove this lemma. ◀

4 Recursion

In this section, we prove Lemma 13. For integer $n\ge 1$, we define $\gamma_n\in(0,1)$ as the unique positive root of

$$\frac{\beta}{nq}\sum_{i=1}^{n}\gamma_n^{-i}=1. \tag{17}$$

The left hand side of this equation is continuously decreasing in $\gamma_n\in(0,\infty)$, and it is $\frac{\beta}{q}$ when evaluated at $\gamma_n=1$. For $\gamma_n=\beta/q$, the left hand side is

$$\frac{\beta}{nq}\sum_{i=1}^{n}\Big(\frac{q}{\beta}\Big)^i=\frac{1}{n}\sum_{i=1}^{n}\Big(1+\frac{q}{\beta}+\cdots+\frac{q^{n-1}}{\beta^{n-1}}\Big)>1,$$

where the last inequality follows from $\frac{\beta}{q}<1$. There is indeed a unique positive solution for (17) in $(\beta/q,1)$.

We will show that the unique positive solution of (17) will be asymptotically approximated by $1 - b/n$ where b is the positive real root of the following equation

$$\frac{\beta(1-e^{-b})}{qb} - e^{-b} = 0, \tag{18}$$

while b does not have an elementary closed form, it can be expressed as

$$b = -\frac{\beta}{q} - W_{-1}\left(-\frac{\beta}{q}\exp\left(-\beta/q\right)\right),$$

where W_{-1} is the negative branch of the Lambert W function. Formally, for real numbers x and y, the equation $ye^y = x$ has real roots for y if $-\frac{1}{e} \le x < 0$ and one of its real roots is $W_{-1}(x)$. We establish the following approximation for the root of (17); its proof is provided in the full version [5].

▶ **Lemma 15.** *Let $\hat{\gamma}_n = 1 - b/n$. Then, $|\gamma_n - \hat{\gamma}_n| = O(n^{-2})$ and for every $t \in [0, \dots, n\sqrt{n}]$, we have $e^{-bt/n}/2 \le \gamma_n^t \le 2e^{-bt/n}$.*

A direct consequence of Lemma 15 is that

$$\sum_{i=0}^{\infty} \gamma_n^i = \Theta(n), \quad \text{and} \tag{19}$$

$$\gamma_n^j = \Theta(1) \text{ for all } 0 \le j \le n. \tag{20}$$

We use γ_n to bound the recurrence in the following lemma, which we later use to prove Lemma 13.

▶ **Lemma 16.** *Let $\{b_t\}_{t\ge n}$ be a sequence of nonnegative real numbers and define $\{d_t\}_{t\ge 0}$ as*

$$d_t = \begin{cases} 0 & \text{for } 0 \le t < n, \\ b_t + \frac{\beta}{nq}\sum_{i=t-n}^{t-1} d_i & \text{for } t \ge n. \end{cases}$$

Then, there exists a constant $C \ge 1$ such that for all $t \ge n$, we have $d_t \le b_t + \frac{C\gamma_n^t}{n}\sum_{i=n}^{t} b_i \gamma_n^{-i}$.

The proof of this lemma is deferred to the full version of the paper [5]. We are now ready to provide the proof of Lemma 13.

Proof of Lemma 13. Fix $k \in Q$ and consider the sequence $\{y_t\}_{t\ge 0}$ defined as $y_i = x_i^{(k)}$ for $i \in \{0, \dots, n-1\}$ and $y_t = \frac{1}{n}\frac{\beta}{q}\sum_{i=t-n}^{t-1} y_i$ when $t \ge n$. Let $d_t = x_t^{(k)} - y_t$ for all $t \ge 0$. Observe that $d_t = 0$ for all $0 \le t < n$. For $t \ge n$, we have

$$\begin{aligned} d_t = x_t^{(k)} - y_t &= \frac{\beta}{nq}\sum_{i=t-n}^{t-1}\left(x_i^{(k)} - y_i\right) + \frac{C}{n}\sum_{i=t-n}^{t-1}\|x_i\|_2^2 + \frac{C}{n} \\ &= \frac{\beta}{nq}\sum_{i=t-n}^{t-1} d_i + \frac{C}{n}\sum_{i=t-n}^{t-1}\|x_i\|_2^2 + \frac{C}{n} = b_t + \frac{\beta}{nq}\sum_{i=t-n}^{t-1} d_i, \end{aligned}$$

where we let $b_t = \frac{C}{n}\sum_{i=t-n}^{t-1}\|x_i\|_2^2 + \frac{C}{n}$ for all $t \ge n$. From Lemma 16, there exists $C_1 > 1$ such that the following holds for all $t \ge n$

$$d_t \le b_t + \frac{C_1\gamma_n^t}{n}\sum_{i=n}^{t}\gamma_n^{-i} b_i.$$

For convenience, we choose the constant C_1 large enough such that also $\gamma_n^{-n} \le C_1$ and

$$\sum_{i=0}^{\infty} \gamma_n^i \le C_1 n; \tag{21}$$

see (20) and (19). We expand d_t as follows

$$d_t \le \frac{C}{n} \sum_{i=t-n}^{t-1} \|x_i\|_2^2 + \frac{C}{n} + \frac{CC_1\gamma_n^t}{n^2} \sum_{i=n}^{t} \gamma_n^{-i} \sum_{j=i-n}^{i-1} \|x_j\|_2^2 + \frac{CC_1}{n^2} \sum_{i=n}^{t} \gamma_n^{t-i}.$$

From (21), we see $\sum_{i=n}^{t} \gamma_n^{t-i} \le C_1 n$ and plugging this in the above bound, we obtain

$$\begin{aligned} d_t &\le \frac{C}{n} \sum_{i=t-n}^{t-1} \|x_i\|_2^2 + \frac{C}{n} + \frac{CC_1\gamma_n^t}{n^2} \sum_{i=n}^{t} \gamma_n^{-i} \sum_{j=i-n}^{i-1} \|x_j\|_2^2 + \frac{CC_1^2}{n} \\ &\le \frac{2CC_1^2}{n} + \frac{C}{n} \sum_{i=t-n}^{t-1} \|x_i\|_2^2 + \frac{CC_1\gamma_n^t}{n^2} \sum_{i=n}^{t} \gamma_n^{-i} \sum_{j=i-n}^{i-1} \|x_j\|_2^2 \end{aligned} \tag{22}$$

where we use the fact that $C + CC_1^2 \le 2CC_1^2$. Observe also that

$$\begin{aligned} \sum_{i=n}^{t} \gamma_n^{-i} \sum_{j=i-n}^{i-1} \|x_j\|_2^2 &= \sum_{j=0}^{n-1} \gamma_n^{j-n} \sum_{i=n}^{t} \gamma_n^{-(i+j-n)} \|x_{i+j-n}\|_2^2 \\ &\le \gamma_n^{-n} \sum_{j=0}^{n-1} \sum_{i=n}^{t} \gamma_n^{-(i+j-n)} \|x_{i+j-n}\|_2^2 \le nC_1 \sum_{i=0}^{t-1} \gamma_n^{-i} \|x_i\|_2^2, \end{aligned} \tag{23}$$

where the last inequality follows from $\gamma_n^{-n} \le C_1$. Plugging (23) into (22), we obtain

$$\begin{aligned} d_t &\le \frac{2CC_1^2}{n} + \frac{C}{n} \sum_{i=t-n}^{t-1} \|x_i\|_2^2 + \frac{CC_1^2}{n} \gamma_n^t \sum_{i=0}^{t-1} \gamma_n^{-i} \|x_i\|_2^2 \\ &\le \frac{2CC_1^2}{n} + \frac{C\gamma_n^{-n}}{n} \sum_{i=t-n}^{t-1} \gamma_n^{t-i} \|x_i\|_2^2 + \frac{CC_1^2}{n} \gamma_n^t \sum_{i=0}^{t-1} \gamma_n^{-i} \|x_i\|_2^2 \\ &\le \frac{2CC_1^2}{n} + \frac{C(C_1+1)C_1}{n} \gamma_n^t \sum_{i=0}^{t-1} \gamma_n^{-i} \|x_i\|_2^2 \le \frac{2CC_1^2}{n} + \frac{2CC_1^2}{n} \gamma_n^t \sum_{i=0}^{t-1} \gamma_n^{-i} \|x_i\|_2^2. \end{aligned} \tag{24}$$

We show next by induction on t that the following holds for all $t \ge 0$

$$y_t \le C_1 \rho \gamma_n^t. \tag{25}$$

For all $0 \le t < n$, we have $C_1\rho\gamma_n^t \ge \rho\gamma_n^{-n}\gamma_n^t = \rho\gamma_n^{t-n} \ge \rho \ge y_t$. Therefore, the base case holds. We assume that the claim holds for all $t < \ell$ where $\ell \ge n$. For $t = \ell$,

$$y_t = \frac{\beta}{nq} \sum_{i=t-n}^{t-1} y_i \le \frac{\beta}{nq} \sum_{i=t-n}^{t-1} C_1\rho\gamma_n^i = C_1\rho \frac{\beta}{nq} \sum_{i=t-n}^{t-1} \gamma_n^i = C_1\rho\gamma_n^t \frac{\beta}{nq} \sum_{i=0}^{n-1} \gamma_n^{i-n} = C_1\rho\gamma_n^t,$$

which proves the claim in (25). Now, we show the following using induction

$$\|x_t\|_2^2 < 4C_1^2 q \rho^2 \gamma_n^{2t} + \frac{4C_1^2 q}{n}. \tag{26}$$

We obtain the following for all t such that $0 \le t < n$,

$$4C_1^2 q\rho^2\gamma_n^{2t} + \frac{4C_1^2 q}{n} \ge 4C_1^2 q\rho^2\gamma_n^{2n} \ge 4\gamma_n^{-2n} q\rho^2\gamma_n^{2n} \ge 4q\rho^2 \ge \|x_t\|_2^2$$

Therefore, the base case holds. We assume that the claim holds for all $t < \ell$ where $\ell \ge n$. For $t = \ell$, we obtain the following from (25) and (24),

$$|x_t^{(k)}| \le |y_t| + |d_t| \le C_1\rho\gamma_n^t + \frac{2CC_1^2\gamma_n^t}{n}\sum_{i=0}^{t-1}\gamma_n^{-i}\|x_i\|_2^2 + \frac{2CC_1^2}{n}. \tag{27}$$

The induction hypothesis then implies that

$$\begin{aligned}\gamma_n^t\sum_{i=0}^{t-1}\gamma_n^{-i}\|x_i\|_2^2 &\le \gamma_n^t\sum_{i=0}^{t-1}\gamma_n^{-i}\Big(4C_1^2 q\rho^2\gamma_n^{2i} + \frac{4C_1^2 q}{n}\Big) \le 4C_1^2 q\gamma_n^t\sum_{i=0}^{t-1}\gamma_n^{-i}\big(\rho^2\gamma_n^{2i} + 1/n\big)\\ &\le 4C_1^2 q\Big(\rho^2\gamma_n^t\sum_{i=0}^{t-1}\gamma_n^i + \frac{1}{n}\sum_{i=0}^{t-1}\gamma_n^{t-i}\Big) \le 4C_1^2 q\big(\rho^2\gamma_n^t C_1 n + C_1\big)\end{aligned}$$

We then plug the above bound into (27) to obtain

$$\begin{aligned}|x_t^{(k)}| &\le C_1\rho\gamma_n^t + \frac{2CC_1^2}{n}\Big(4C_1^2 q\big(\rho^2\gamma_n^t C_1 n + C_1\big)\Big) + \frac{2CC_1^2}{n}\\ &\le C_1\rho\gamma_n^t + 8CC_1^5\rho^2 q\gamma_n^t + \frac{8CC_1^5 q + 2CC_1^2}{n}\\ &\le \rho\gamma_n^t C_1\big(1 + 8CC_1^4\rho q\big) + \frac{8CC_1^5 q + 2CC_1^2}{n}.\end{aligned}$$

Let $C_2 = C_1\big(1+8CC_1^4\rho q\big)$ and $C_3 = 8CC_1^5 q + 2CC_1^2$ for ease of notation. Since k is arbitrary, the above inequality holds for all coordinates in Q, hence,

$$\|x_t\|_2^2 \le 2q\rho^2\gamma_n^{2t}C_2^2 + \frac{2qC_3^2}{n^2},$$

where we use the fact that $(w+z)^2 \le 2w^2 + 2z^2$ for $w, z \in \mathbb{R}$. We want to show $2q\rho^2\gamma_n^{2t}C_2^2 + \frac{2qC_3^2}{n^2} \le 4C_1^2 q\rho^2\gamma_n^{2t} + \frac{4C_1^2 q}{n}$ by equating the coefficients of γ_n^{2t} and $1/n$. Comparing the coefficient of γ_n^{2t}, we show $C_2 \le \sqrt{2}C_1$ or equivalently, $1 + 8CC_1^4\rho q \le \sqrt{2}$. We choose $\rho > 0$ small enough to make the above inequality true. Moreover, $2qC_3^2/n \le 4C_1^2 q$ holds trivially. Therefore, we proved our claim (26), hence choosing $\hat{C} = 4C_1^2 q$ proves the lemma. ◀

5 Contraction of Hamming Distance

Our goal in this section is to establish Lemma 3 which ensures that **Phase 2** of our coupling succeeds with the desired ε probability. We use the following lemma to upper bound expectation of Hamming distance; its proof is deferred to the full version of the paper [5].

▶ **Lemma 17.** *Let $\{x_t\}_{t\ge 0}$ be a sequence of positive real numbers, such that $x_t \le n$ for $t < n$, and for $t \ge n$ we have $x_t = \frac{\beta}{nq}\sum_{i=t-n}^{t-1} x_i + C\sum_{i=t-n}^{t-1}\gamma_n^{2i} + C$, where $C > 0$ is a constant. Then, there exists a constant $\hat{C} > 0$ such that for all $t \ge 0$ $x_t \le \hat{C}n\gamma_n^t + \hat{C}$.*

We are now ready to provide the proof of Lemma 3.

Proof of Lemma 3. We couple two instances using the optimal single site coupling for two copies of the dynamics. Formally, we define a distribution $\nu_{\sigma,v}$ on Q such that $\nu_{\sigma,v}(j) = g_\beta^j\big(\alpha(\sigma) - \frac{1}{n}e_{\sigma(v)}\big)$ where $\sigma \in \Sigma_n$, $j \in Q$ and $v \in V$. From a pair of configurations σ_{t-1} and $\tilde{\sigma}_{t-1}$, we couple the t-th site update as follows:

1. Let $v = \big(t \bmod n\big) + 1$ be the vertex to be updated.
2. Draw spins J_t and $\tilde{J}_t$ from an optimal coupling of the distribution $\nu_{\sigma_{t-1},v}$ and $\nu_{\tilde{\sigma}_{t-1},v}$ respectively.
3. Set $\sigma_t\,(v) = J_t$ and $\tilde{\sigma}_t\,(v) = \tilde{J}_t$ leaving all other vertices unchanged.

We show next that under the optimal site coupling, there exists a constant $C_1 > 0$ such that $\forall\ t \geq n$, $\mathbb{E}[d_H(\sigma_t, \tilde{\sigma}_t)]$ is at most

$$C_1 + \frac{\beta}{nq} \sum_{i=t-n}^{t-1} \mathbb{E}[d_H(\sigma_i, \tilde{\sigma}_i)] + C_1 \sum_{i=t-n}^{t-1} \Big(\mathbb{E}\big[\|\alpha(\sigma_i) - \hat{e}\|_2^2\big] + \mathbb{E}\big[\|\alpha(\tilde{\sigma}_i) - \hat{e}\|_2^2\big] \Big). \tag{28}$$

The Hamming distance between two copies after t site updates for $t \geq n$ is $d_H(\sigma_t, \tilde{\sigma}_t) = \sum_{i=t-n+1}^{t} \mathbf{1}\left(J_i \neq \tilde{J}_i\right)$. Taking expectations on both sides we obtain for all $t \geq n$ that

$$\begin{aligned}\mathbb{E}\left[d_H(\sigma_t, \tilde{\sigma}_t)\right] = \sum_{i=t-n+1}^{t} \mathbb{E}\left[\mathbf{1}\left(J_i \neq \tilde{J}_i\right)\right] &= \sum_{i=t-n+1}^{t} \mathbb{E}\left[\mathbb{E}\left[\mathbf{1}\big(J_i \neq \tilde{J}_i\big) \middle| \mathcal{F}_{i-1}\right]\right] \\ &= \sum_{i=t-n+1}^{t} \mathbb{E}\left[\Pr\left[J_i \neq \tilde{J}_i \middle| \mathcal{F}_{i-1}\right]\right]. \end{aligned} \tag{29}$$

Since we are optimally coupling each co-ordinate, the following holds,

$$\Pr\left[J_i \neq \tilde{J}_i \mid \mathcal{F}_{i-1}\right] = \frac{1}{2} \left\| g_\beta\Big(\alpha(\sigma_i) - \frac{1}{n} e_{I_i}\Big) - g_\beta\Big(\alpha(\tilde{\sigma}_i) - \frac{1}{n} e_{\tilde{I}_i}\Big) \right\|_1. \tag{30}$$

For any $I, \tilde{I} \in Q$, we have

$$\begin{aligned}\left\| g_\beta(s - \frac{1}{n} e_I) - g_\beta(\tilde{s} - \frac{1}{n} e_{\tilde{I}}) \right\|_1 &= \|g_\beta(s) - g_\beta(\tilde{s})\|_1 + \frac{2q^2\beta}{n} \\ &\leq \frac{\beta}{q} \|s - \tilde{s}\|_1 + O\big((\|s - \hat{e}\|_1 + \|\tilde{s} - \hat{e}\|_1)^2\big) + \frac{2q^2\beta}{n} \\ &\leq \frac{\beta}{q} \|s - \tilde{s}\|_1 + O\big(\|s - \hat{e}\|_2^2 + \|\tilde{s} - \hat{e}\|_2^2\big) + \frac{2q^2\beta}{n}, \end{aligned} \tag{31}$$

where the first inequality follows from Lemma 11(ii) and the fact that $\|s - \tilde{s}\|_1 \leq \|s - \hat{e}\|_1 + \|\tilde{s} - \hat{e}\|_1$. Since, the total variation distance computes the optimal distance of the spin fractions between two copies, this lower bounds the fraction of the Hamming distance. Therefore, we use the identity $\frac{1}{n} d_H(\sigma_t, \tilde{\sigma}_t) \geq \frac{1}{2} \|\alpha(\sigma_t) - \alpha(\tilde{\sigma}_t)\|_1$, (31), and (30) to deduce that for some constant $C_2 > 2q^2/\beta$

$$\Pr\left[J_i \neq \tilde{J}_i \mid \mathcal{F}_{i-1}\right] \leq \frac{\beta}{nq} d_H(\sigma_i, \tilde{\sigma}_i) + C_2\big(\|\alpha(\sigma_i) - \hat{e}\|_2^2 + \|\alpha(\tilde{\sigma}_i) - \hat{e}\|_2^2\big) + \frac{C_2}{n}.$$

Plugging the above equation in (29), we prove (28). Using Lemma 14 into (28), there exists a constant $C_3 > 0$ such that for any $t \geq n$,

$$\mathbb{E}\big[d_H(\sigma_t, \tilde{\sigma}_t)\big] \leq \frac{\beta}{nq} \sum_{i=t-n}^{t-1} \mathbb{E}[d_H(\sigma_i, \tilde{\sigma}_i)] + 2C_1C_3 \sum_{i=t-n}^{t-1} \gamma_n^{2i} + (2C_1C_3 + C_1).$$

Using the Lemma 17, we get that for some constant $C_4 > 0$, $\mathbb{E}\big[d_H(\sigma_t, \tilde{\sigma}_t)\big] = C_4 n \gamma_n^t + C_4$, and the result follows from Lemma 15 and Markov's inequality. ◀

References

1 Vedat Levi Alev and Ori Parzanchevski. Sequential sweeps and high dimensional expansion. *arXiv preprint*, 2023. doi:10.48550/arXiv.2312.02089.

2 Yali Amit. Convergence properties of the Gibbs sampler for perturbations of Gaussians. *The Annals of Statistics*, 24(1):122–140, 1996.

3 Antonio Blanca, Pietro Caputo, Alistair Sinclair, and Eric Vigoda. Spatial mixing and nonlocal Markov chains. *Random Structures & Algorithms*, 55(3):584–614, 2019. Extended abstract appeared SODA 2018. doi:10.1002/RSA.20844.

4 Antonio Blanca, Reza Gheissari, and Xusheng Zhang. Mean-field Potts and random-cluster dynamics from high-entropy initializations. *The Annals of Applied Probability*, 36(1):393–450, 2026. Extended abstract appeared SODA 2018.

5 Antonio Blanca and Md Tahmidur Rafid. Mixing and cutoff for the systematic scan dynamics of the mean-field ferromagnetic Potts model. *arXiv preprint*, 2026. arXiv:2607.09841.

6 Antonio Blanca and Zhezheng Song. Cutoff for the Swendsen–Wang Dynamics on the Complete Graph. In *45th IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science (FSTTCS 2025)*, pages 17:1–17:21, 2025. doi:10.4230/LIPIcs.FSTTCS.2025.17.

7 Antonio Blanca and Xusheng Zhang. Rapid Mixing of Global Markov Chains via Spectral Independence: The Unbounded Degree Case. *Algorithmica*, 88(1):19, 2026. Extended abstract appeared RANDOM 2023. doi:10.1007/S00453-025-01353-5.

8 J Bricmont. The statistical mechanics of lattice gases: Vol. 1. B. Simon, Princeton University Press, Princeton, New Jersey, 1993, 1996.

9 Iwona Chlebicka, Krzysztof Łatuszyński, and Błażej Miasojedow. Solidarity of Gibbs Samplers: the spectral gap. *The Annals of Applied Probability*, 35(1):142–157, 2025.

10 Paul Cuff, Jian Ding, Oren Louidor, Eyal Lubetzky, Yuval Peres, and Allan Sly. Glauber Dynamics for the Mean-Field Potts Model. *Journal of Statistical Physics*, 149(3):432–477, 2012.

11 Persi Diaconis and Arun Ram. Analysis of systematic scan Metropolis algorithms using Iwahori-Hecke algebra techniques. *Michigan Mathematical Journal*, 48(1):157–190, 2000.

12 Jian Ding, Eyal Lubetzky, and Yuval Peres. Censored Glauber dynamics for the mean field Ising model. *Journal of Statistical Physics*, 137(3):407–458, 2009.

13 Jian Ding, Eyal Lubetzky, and Yuval Peres. The mixing time evolution of Glauber dynamics for the mean-field Ising model. *Communications in Mathematical Physics*, 289(2):725–764, 2009.

14 Martin Dyer, Leslie Ann Goldberg, and Mark Jerrum. Systematic Scan for Sampling Colorings. *The Annals of Applied Probability*, pages 185–230, 2006.

15 Martin Dyer, Leslie Ann Goldberg, and Mark Jerrum. Dobrushin conditions and systematic scan. *Combinatorics, Probability and Computing*, 17(6):761–779, 2008. doi:10.1017/S0963548308009437.

16 Martin Dyer, Leslie Ann Goldberg, and Mark Jerrum. Matrix norms and rapid mixing for spin systems. *The Annals of Applied Probability*, 19(1):71–107, 2009.

17 Jason Gaitonde and Elchanan Mossel. Comparison theorems for the mixing times of systematic and random scan dynamics. In *Proceedings of the 37th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 2575–2587. SIAM, 2026. doi:10.1137/1.9781611978971.93.

18 Andreas Galanis, Daniel Štefankovič, and Eric Vigoda. Swendsen-Wang algorithm on the mean-field Potts model. *Random Structures & Algorithms*, 54(1):82–147, 2019. Extended abstract appeared in RANDOM 2015. doi:10.1002/RSA.20768.

19 Heng Guo, Kaan Kara, and Ce Zhang. Layerwise systematic scan: Deep Boltzmann machines and beyond. In *International Conference on Artificial Intelligence and Statistics*, pages 178–187. PMLR, 2018. URL: http://proceedings.mlr.press/v84/guo18a.html.

20 Thomas P Hayes. A simple condition implying rapid mixing of single-site dynamics on spin systems. In *Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 39–46. IEEE, 2006. doi:10.1109/FOCS.2006.6.

21 Bryan D He, Christopher M De Sa, Ioannis Mitliagkas, and Christopher Ré. Scan order in Gibbs sampling: Models in which it matters and bounds on how much. In *Proceedings of Advances in Neural Information Processing Systems (NeurIPS)*, volume 29, 2016.

22 Sanghak Jeon. Systematic scanning Glauber dynamics for the mean-field Ising model. *arXiv preprint*, 2023. arXiv:2307.10127.

23 David A Levin, Malwina J Luczak, and Yuval Peres. Glauber dynamics for the mean-field Ising model: cut-off, critical power law, and metastability. *Probability Theory and Related Fields*, 146(1):223, 2010.

24 David A. Levin, Yuval Peres, and Elizabeth L. Wilmer. *Markov chains and mixing times*. American Mathematical Society, 2006.

25 Keunwoo Lim. Cutoff Phenomenon for Cyclic Dynamics on Hypercube. *arXiv preprint*, 2024. arXiv:2010.01756.

26 Yuval Peres and Peter Winkler. Can extra updates delay mixing? *Communications in Mathematical Physics*, 323(3):1007–1016, 2013.

27 Gareth O Roberts and Jeffrey S Rosenthal. Surprising Convergence Properties of Some Simple Gibbs Samplers under Various Scans. *International Journal of Statistics and Probability*, 5(4):112–121, 2016.

28 Sebastien Roch. *Modern Discrete Probability: An Essential Toolkit*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2024.

29 Ce Zhang and Christopher Ré. Towards high-throughput Gibbs sampling at scale: A study across storage managers. In *Proceedings of the 2013 ACM SIGMOD international conference on management of data*, pages 397–408, 2013. doi:10.1145/2463676.2463702.