

# RAPID MIXING OF GLAUBER DYNAMICS OF GIBBS ENSEMBLES VIA AGGREGATE PATH COUPLING AND LARGE DEVIATIONS METHODS

YEVGENIY KOVCHEGOV AND PETER T. OTTO

**ABSTRACT.** In this paper, we present a novel extension to the classical path coupling method to statistical mechanical models which we refer to as aggregate path coupling. In conjunction with large deviations estimates, we use this aggregate path coupling method to prove rapid mixing of Glauber dynamics for a large class of statistical mechanical models, including models that exhibit discontinuous phase transitions which have traditionally been more difficult to analyze rigorously. The parameter region for rapid mixing for the generalized Curie-Weiss-Potts model is derived as a new application of the aggregate path coupling method.

## 1. INTRODUCTION

In recent years, mixing times of dynamics of statistical mechanical models have been the focus of much probability research, drawing interest from researchers in mathematics, physics and computer science. The topic is both physically relevant and mathematically rich. But up to now, most of the attention has focused on particular models including rigorous results for several mean-field models. A few examples are (a) the Curie-Weiss (mean-field Ising) model [5, 6, 10, 14], (b) the mean-field Blume-Capel model [9, 13], (c) the Curie-Weiss-Potts (mean-field Potts) model [1, 4, 18]. A good survey of the topic of mixing times of statistical mechanical models can be found in the recent paper by Cuff et. al. [4].

An important question driving the work in the field is the relationship between the mixing times of the dynamics and the equilibrium phase transition structure of the corresponding statistical mechanical models. For example, the Curie-Weiss model, which undergoes a continuous, second-order, phase transition, was one of the first models studied to investigate this relationship and it was found that the mixing times undergo a transition from rapid to slow mixing at precisely the same critical value as the equilibrium phase transition [15]. This property was also shown for the mean-field Blume-Capel model [13] in the parameter regime where the model undergoes a continuous, second-order phase transition.

On the other hand, for models that exhibit a discontinuous, first-order, phase transition, they do not appear to share this same property. This was first verified for the mean-field Blume-Capel model in the discontinuous phase transition parameter regime [13] and recently for the Curie-Weiss-Potts model [4]. For these models, it was shown that the change in mixing times occurs, not at the equilibrium phase transition value, but instead at a smaller parameter value at which metastable states first emerge.

---

*Date:* April 2, 2015.

*2000 Mathematics Subject Classification.* Primary 60J10; Secondary 60K35.

The results for models that exhibit a continuous phase transition were obtained by a direct application of the standard path coupling method that requires contraction of the mean path coupling distance between all neighboring configurations. See [2] and [15]. For models that exhibit a discontinuous phase transition straightforward path coupling methods fail and the results of [13] were obtained by applying a novel extension called *aggregate path coupling* in one dimension and large deviations estimates.

In this paper, we extend the work we did in [13] and provide a single general framework for determining the parameter regime for rapid mixing of the Glauber dynamics for a large class of statistical mechanical models, including all those listed above. The aggregate path coupling method presented here extends the classical path coupling method in two directions. First, we consider macroscopic quantities in higher dimensions and find a monotone contraction path by considering a related variational problem in the continuous space. We also do not require the monotone path to be a nearest-neighbor path. In fact, in most situations we consider, a nearest-neighbor path will not work for proving contraction. Second, the aggregation of the mean path distance along a monotone path is shown to contract for some but not all pairs of configurations. Yet, we use measure concentration and large deviation principle to show that showing contraction for pairs of configurations, where at least one of them is close enough to the equilibrium, is sufficient for establishing rapid mixing.

Our main result is general enough to be applied to statistical mechanical models that undergo both types of phase transitions and to models whose macroscopic quantity are in higher dimensions. Moreover, despite the generality, the application of our results requires straightforward conditions that we illustrate in Section 10. This is a significant simplification for proving rapid mixing for statistical mechanical models, especially those that undergo first-order, discontinuous phase transitions. Lastly, our results also provide a link between measure concentration of the stationary distribution and rapid mixing of the corresponding dynamics for this class of statistical mechanical models. This idea has been previously studied in [17] where the main result showed that rapid mixing implied measure concentration defined in terms of Lipschitz functions. In our work, we prove a type of converse where measure concentration, in terms of a large deviation principle, implies rapid mixing.

The paper is organized as follows. In Section 2 the general construction of the class of the mean-field models considered in this paper is provided. Next, in Section 3 the large deviation principle for the Gibbs measures from [8] that will be used in the main result of this manuscript is reviewed, and the concept of equilibrium macrostates is discussed. In Section 4 and Section 5 the Glauber dynamics is introduced, and its transition probabilities are analyzed. Section 6 provides a greedy coupling construction, standard for the Glauber dynamics of a mean-field statistical mechanical model. Section 7 describes a single time step evolution of the mean coupling distance for two configurations whose spin proportion vectors are  $\varepsilon$ -close. Section 7 is followed by Section 8 which describes a single time step evolution of the mean coupling distance in general by aggregating the mean coupling distances along a monotone path of points connecting two configurations. Also, in Section 8 general conditions for the main result Theorem 9.2 are stated and discussed. The main result is stated and proved in Section 9. The paper concludes with Section 10, where the region of rapid mixing  $\beta < \beta_s(q, r)$  for the generalized Curie-Weiss-Potts model (including the standard Curie-Weiss-Potts model) is proven as an immediate and simple application of the main result of the current paper.

## 2. GIBBS ENSEMBLES

We begin by defining the general class of statistical mechanical spin models for which our results can be applied. As mentioned above, this class includes all of the models listed in the introduction and we illustrate the application of our main result for the particular model: the Curie-Weiss-Potts model, in section 10.

Let  $q$  be a fixed integer and define  $\Lambda = \{e^1, e^2, \dots, e^q\}$ , where  $e^k$  are the  $q$  standard basis vectors of  $\mathbb{R}^q$ . A *configuration* of the model has the form  $\omega = (\omega_1, \omega_2, \dots, \omega_n) \in \Lambda^n$ . We will consider a configuration on a graph with  $n$  vertices and let  $X_i(\omega) = \omega_i$  be the *spin* at vertex  $i$ . The random variables  $X_i$ 's for  $i = 1, 2, \dots, n$  are independent and identically distributed with common distribution  $\rho$ .

In terms of the microscopic quantities, the spins at each vertex, the relevant macroscopic quantity is the *magnetization vector* (aka empirical measure or proportion vector)

$$(1) \quad L_n(\omega) = (L_{n,1}(\omega), L_{n,2}(\omega), \dots, L_{n,q}(\omega)),$$

where the  $k$ th component is defined by

$$L_{n,k}(\omega) = \frac{1}{n} \sum_{i=1}^n \delta(\omega_i, e^k)$$

which yields the proportion of spins in configuration  $\omega$  that take on the value  $e^k$ . The magnetization vector  $L_n$  takes values in the set of probability vectors

$$(2) \quad \mathcal{P}_n = \left\{ \frac{n_k}{n} : \text{each } n_k \in \{0, 1, \dots, n\} \text{ and } \sum_{k=1}^q n_k = n \right\}$$

inside the continuous simplex

$$\mathcal{P} = \left\{ \nu \in \mathbb{R}^q : \nu = (\nu_1, \nu_2, \dots, \nu_q), \text{ each } \nu_k \geq 0, \sum_{k=1}^q \nu_k = 1 \right\}.$$

**Remark 2.1.** For  $q = 2$ , the empirical measure  $L_n$  yields the empirical mean  $S_n(\omega)/n$  where  $S_n(\omega) = \sum_{i=1}^n \omega_i$ . Therefore, the class of models considered in this paper includes those where the relevant macroscopic quantity is the empirical mean, like the Curie-Weiss (mean-field Ising) model.

Statistical mechanical models are defined in terms of the Hamiltonian function, which we denote by  $H_n(\omega)$ . The Hamiltonian function encodes the interactions of the individual spins and the total energy of a configuration. To take advantage of the large deviation bounds stated in the next section, we assume that the Hamiltonian can be expressed in terms of the empirical measures  $L_n$  as stated in the following definition.

**Definition 2.2.** For  $z \in \mathbb{R}^q$ , we define the **interaction representation function**, denoted by  $H(z)$ , to be a differentiable function satisfying

$$H_n(\omega) = nH(L_n(\omega))$$

Throughout the paper we suppose the interaction representation function  $H(z)$  is a finite concave  $\mathcal{C}^3(\mathbb{R}^q)$  function that has the form

$$H(z) = H_1(z_1) + H_2(z_2) + \dots + H_q(z_q)$$

For the Curie-Weiss-Potts (CWP) model discussed in section 10,

$$H(z) = -\frac{1}{2}\langle z, z \rangle = -\frac{1}{2}z_1^2 - \frac{1}{2}z_2^2 - \dots - \frac{1}{2}z_q^2.$$

**Definition 2.3.** *The Gibbs measure or Gibbs ensemble in statistical mechanics is defined as*

$$(3) \quad P_{n,\beta}(B) = \frac{1}{Z_n(\beta)} \int_B \exp\{-\beta H_n(\omega)\} dP_n = \frac{1}{Z_n(\beta)} \int_B \exp\{-\beta n H(L_n(\omega))\} dP_n$$

where  $P_n$  is the product measure with identical marginals  $\rho$  and  $Z_n(\beta) = \int_{\Lambda^n} \exp\{-\beta H_n(\omega)\} dP_n$  is the **partition function**. The positive parameter  $\beta$  represents the inverse temperature of the external heat bath.

**Remark 2.4.** *To simplify the presentation, we take  $\Lambda = \{e^1, e^2, \dots, e^q\}$ , where  $e^k$  are the  $q$  standard basis vectors of  $\mathbb{R}^q$ . But our analysis has a straight-forward generalization to the case where  $\Lambda = \{\theta^1, \theta^2, \dots, \theta^q\}$ , where  $\theta^k$  is any basis of  $\mathbb{R}^q$ . In this case, the product measure  $P_n$  would have identical one-dimensional marginals equal to*

$$\bar{\rho} = \frac{1}{q} \sum_{i=1}^q \delta_{\theta^i}$$

An important tool we use to prove rapid mixing of the Glauber dynamics that converge to the Gibbs ensemble above is the large deviation principle of the empirical measure with respect to the Gibbs ensemble. This measure concentration is precisely what drives the rapid mixing. The large deviations background is presented next.

### 3. LARGE DEVIATIONS

By Sanov's Theorem, the empirical measure  $L_n$  satisfies the large deviation principle (LDP) with respect to the product measure  $P_n$  with identical marginals  $\rho$  and the rate function is given by the **relative entropy**

$$R(\nu|\rho) = \sum_{k=1}^q \nu_k \log \left( \frac{\nu_k}{\rho_k} \right)$$

for  $\nu \in \mathcal{P}$ . Theorem 2.4 of [8] yields the following result for the Gibbs measures (3).

**Theorem 3.1.** *The empirical measure  $L_n$  satisfies the LDP with respect to the Gibbs measure  $P_{n,\beta}$  with rate function*

$$I_\beta(z) = R(z|\rho) + \beta H(z) - \inf_t \{R(t|\rho) + \beta H(t)\}$$

In other words, for any closed subset  $F$ ,

$$(4) \quad \limsup_{n \rightarrow \infty} \frac{1}{n} \log P_{n,\beta}(L_n \in F) \leq - \inf_{z \in F} I_\beta(z)$$

and for any open subset  $G$ ,

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log P_{n,\beta}(L_n \in G) \leq - \inf_{z \in G} I_\beta(z).$$

The LDP upper bound (4) stated in the previous theorem yields the following natural definition of **equilibrium macrostates** of the model.

$$(5) \quad \mathcal{E}_\beta := \{\nu \in \mathcal{P} : \nu \text{ minimizes } R(\nu|\rho) + \beta H(\nu)\}$$

For our main result, we assume that there exists a positive interval  $B$  such that for all  $\beta \in B$ ,  $\mathcal{E}_\beta$  consists of a single state  $z_\beta$ . We refer to this interval  $B$  as the single phase region.

Again from the LDP upper bound, when  $\beta$  lies in the single phase region, we get

$$(6) \quad P_{n,\beta}(L_n \in dx) \implies \delta_{z_\beta} \quad \text{as } n \rightarrow \infty.$$

The above asymptotic behavior will play a key role in obtaining a rapid mixing time rate for the Glauber dynamics corresponding to the Gibbs measures (3).

An important function in our work is the free energy functional defined below. It is defined in terms of the interaction representation function  $H$  and the logarithmic moment generating function of a single spin; specifically, for  $z \in \mathbb{R}^q$  and  $\rho$  equal to the uniform distribution, the **logarithmic moment generating function** of  $X_1$ , the spin at vertex 1, is defined by

$$(7) \quad \Gamma(z) = \log \left( \frac{1}{q} \sum_{k=1}^q \exp\{z_k\} \right).$$

**Definition 3.2.** *The free energy functional for the Gibbs ensemble  $P_{n,\beta}$  is defined as*

$$(8) \quad G_\beta(z) = \beta(-H)^*(-\nabla H(z)) - \Gamma(-\beta \nabla H(z))$$

where for a finite, differentiable, convex function  $F$  on  $\mathbb{R}^q$ ,  $F^*$  denotes its Legendre-Fenchel transform defined by

$$F^*(z) = \sup_{x \in \mathbb{R}^q} \{\langle x, z \rangle - F(x)\}$$

The following lemma yields an alternative formulation of the set of equilibrium macrostates of the Gibbs ensemble in terms of the free energy functional. The proof is a straightforward generalization of Theorem A.1 in [3].

**Lemma 3.3.** *Suppose  $H$  is finite, differentiable, and concave. Then*

$$\inf_{z \in \mathcal{P}} \{R(z|\rho) + \beta H(z)\} = \inf_{z \in \mathbb{R}^q} \{G_\beta(z)\}$$

Moreover,  $z_0 \in \mathcal{P}$  is a minimizer of  $R(z|\rho) + \beta H(z)$  if and only if  $z_0$  is a minimizer of  $G_\beta(z)$ .

Therefore, the set of equilibrium macrostates can be expressed in terms of the free energy functional as

$$(9) \quad \mathcal{E}_\beta = \{z \in \mathcal{P} : z \text{ minimizes } G_\beta(z)\}$$

As mentioned above, we consider only the single phase region of the Gibbs ensemble; i.e. values of  $\beta$  where  $G_\beta(z)$  has a unique global minimum. For example, for the Curie-Weiss-Potts model, the single phase region are values of  $\beta$  such that  $0 < \beta < \beta_c := (2(q-1)/(q-2)) \log(q-1)$ . At this critical value  $\beta_c$ , the model undergoes a first-order, discontinuous phase transition in which the single phase changes to a multiple phase discontinuously. This is discussed in detail in Section 10.

As we will show, the geometry of the free energy functional  $G_\beta$  not only determines the equilibrium behavior of the Gibbs ensembles but it also yields the condition for rapid mixing of the corresponding Glauber dynamics.

#### 4. GLAUBER DYNAMICS AND MIXING TIMES

On the configuration space  $\Lambda^n$ , we define the Glauber dynamics for the class of spin models considered in this paper. These dynamics yield a reversible Markov chain  $X^t$  with stationary distribution being the Gibbs ensemble  $P_{n,\beta}$ .

- (i) Select a vertex  $i$  uniformly,
- (ii) Update the spin at vertex  $i$  according to the distribution  $P_{n,\beta}$ , conditioned on the event that the spins at all vertices not equal to  $i$  remain unchanged.

For a given configuration  $\sigma = (\sigma_1, \sigma_2, \dots, \sigma_n)$ , denote by  $\sigma_{i,e^k}$  the configuration that agrees with  $\sigma$  at all vertices  $j \neq i$  and the spin at the vertex  $i$  is  $e^k$ ; i.e.

$$\sigma_{i,e^k} = (\sigma_1, \sigma_2, \dots, \sigma_{i-1}, e^k, \sigma_{i+1}, \dots, \sigma_n)$$

Then if the current configuration is  $\sigma$  and vertex  $i$  is selected, the probability the spin at  $i$  is updated to  $e^k$ , denoted by  $P(\sigma \rightarrow \sigma_{i,e^k})$ , is equal to

$$(10) \quad P(\sigma \rightarrow \sigma_{i,e^k}) = \frac{\exp\{-\beta n H(L_n(\sigma_{i,e^k}))\}}{\sum_{\ell=1}^q \exp\{-\beta n H(L_n(\sigma_{i,e^\ell}))\}}.$$

The mixing time is a measure of the convergence rate of a Markov chain to its stationary distribution and is defined in terms of the **total variation distance** between two distributions  $\mu$  and  $\nu$  defined by

$$\|\mu - \nu\|_{\text{TV}} = \sup_{A \subset \Omega} |\mu(A) - \nu(A)| = \frac{1}{2} \sum_{x \in \Omega} |\mu(x) - \nu(x)|$$

Given the convergence of the Markov chain, we define the maximal distance to stationary to be

$$d(t) = \max_{x \in \Omega} \|P^t(x, \cdot) - \pi\|_{\text{TV}}$$

where  $P^t(x, \cdot)$  is the transition probability of the Markov chain starting in configuration  $x$  and  $\pi$  is its stationary distribution. Then, given  $\epsilon > 0$ , the **mixing time** of the Markov chain is defined by

$$t_{mix}(\epsilon) = \min\{t : d(t) \leq \epsilon\}$$

See [15] for a detailed survey on the theory of mixing times.

Rates of mixing times are generally categorized into two groups: *rapid mixing* which implies that the mixing time exhibits polynomial growth with respect to the system size  $n$ , and *slow mixing* which implies that the mixing time grows exponentially with the system size. Determining the parameter regime where a model undergoes rapid mixing is of major importance, as it is in this region that the application of the dynamics is physically feasible.

## 5. GLAUBER DYNAMICS TRANSITION PROBABILITIES

In this section, we show that the update probabilities of the Glauber dynamics introduced in the previous section can be expressed in terms of the derivative of the logarithmic moment generating function of the individual spins  $\Gamma$  defined in (7). The partial derivative of  $\Gamma$  in the direction of  $e^\ell$  has the form

$$[\partial_\ell \Gamma](z) = \frac{\exp\{z_\ell\}}{\sum_{k=1}^q \exp\{z_k\}}$$

We introduce the following function that plays the key role in our analysis.

$$(11) \quad g_\ell^{H,\beta}(z) = [\partial_\ell \Gamma](-\beta \nabla H(z)) = \frac{\exp(-\beta [\partial_\ell H](z))}{\sum_{k=1}^q \exp(-\beta [\partial_k H](z))}.$$

Denote

$$(12) \quad g^{H,\beta}(z) := \left( g_1^{H,\beta}(z), \dots, g_q^{H,\beta}(z) \right).$$

Note that  $g^{H,\beta}(z)$  maps the simplex

$$\mathcal{P} = \left\{ \nu \in \mathbb{R}^q : \nu = (\nu_1, \nu_2, \dots, \nu_q), \text{ each } \nu_k \geq 0, \sum_{k=1}^q \nu_k = 1 \right\}$$

into itself and it can be expressed in terms of the free energy functional  $G_\beta$  defined in (8) by

$$\nabla G_\beta(z) = \beta[\nabla(-H)^*(-\nabla H(z)) - g^{H,\beta}(z)]$$

**Lemma 5.1.** *Let  $P(\sigma \rightarrow \sigma_{i,e^k})$  be the Glauber dynamics update probabilities given in (10). Then, for any  $k \in \{1, 2, \dots, q\}$ ,*

$$P(\sigma \rightarrow \sigma_{i,e^k}) = [\partial_k \Gamma] \left( -\beta \nabla H(L_n(\sigma)) - \frac{\beta}{2n} \mathcal{Q}H(L_n(\sigma)) + \frac{\beta}{n} \left\langle \sigma_i, \mathcal{Q}H(L_n(\sigma)) \right\rangle \sigma_i \right) + O\left(\frac{1}{n^2}\right),$$

where  $\mathcal{Q}$  is the following linear operator:

$$\mathcal{Q}F(z) := (\partial_1^2 F(z), \partial_2^2 F(z), \dots, \partial_q^2 F(z)),$$

for any  $F : \mathbb{R}^q \rightarrow \mathbb{R}$  in  $\mathcal{C}^2$ .

*Proof.* Suppose  $\sigma_i = e^m$ . By Taylor's theorem, for any  $k \neq m$ , we have

$$\begin{aligned} H(L_n(\sigma_{i,e^k})) &= H(L_n(\sigma)) + H_m(L_{n,m}(\sigma) - 1/n) - H_m(L_{n,m}(\sigma)) \\ &\quad + H_k(L_{n,k}(\sigma) + 1/n) - H_k(L_{n,k}(\sigma)) \\ &= H(L_n(\sigma)) + \frac{1}{n} [\partial_k H(L_n(\sigma)) - \partial_m H(L_n(\sigma))] \\ &\quad + \frac{1}{2n^2} [\partial_k^2 H(L_n(\sigma)) + \partial_m^2 H(L_n(\sigma))] + O\left(\frac{1}{n^3}\right). \end{aligned}$$

Now, if  $k = m$ ,

$$\begin{aligned} H(L_n(\sigma_{i,e^k})) &= H(L_n(\sigma)) \\ &= H(L_n(\sigma)) + \frac{1}{n} [\partial_k H(L_n(\sigma)) - \partial_m H(L_n(\sigma))] \\ &\quad + \frac{1}{2n^2} [-\partial_k^2 H(L_n(\sigma)) + \partial_m^2 H(L_n(\sigma))]. \end{aligned}$$

This implies that the transition probability (10) has the form

$$P(\sigma \rightarrow \sigma_{i,e^k}) = [\partial_k \Gamma] \left( -\beta \nabla H(L_n(\sigma)) - \frac{\beta}{2n} \mathcal{Q}H(L_n(\sigma)) + \frac{\beta}{n} \partial_m^2 H(L_n(\sigma)) e^m \right) + O\left(\frac{1}{n^2}\right)$$

as  $\exp\left\{O\left(\frac{1}{n^2}\right)\right\} = 1 + O\left(\frac{1}{n^2}\right)$ .  $\square$

The above Lemma 5.1 can be restated as follows using Taylor expansion.

**Corollary 5.2.** *Let  $P(\sigma \rightarrow \sigma_{i,e^k})$  be the Glauber dynamics update probabilities given in (10). Then, for any  $k \in \{1, 2, \dots, q\}$ ,*

$$P(\sigma \rightarrow \sigma_{i,e^k}) = g_k^{H,\beta}(L_n(\sigma)) + \frac{\beta}{n} \varphi_{k,\sigma_i}^{H,\beta}(L_n(\sigma)) + O\left(\frac{1}{n^2}\right),$$

where

$$\varphi_{k,e^r}^{H,\beta}(z) := -\frac{1}{2} \left\langle \mathcal{Q}H(z), [\nabla \partial_\ell \Gamma](-\beta \nabla H(z)) \right\rangle + \left\langle e^r, \mathcal{Q}H(z) \right\rangle \left\langle e^r, [\nabla \partial_\ell \Gamma](-\beta \nabla H(z)) \right\rangle.$$

As indicated in the title of the paper, we employ a coupling method for proving rapid mixing of Glauber dynamics of Gibbs ensembles. In the next section, we define the specific coupling used.

## 6. COUPLING OF GLAUBER DYNAMICS

We begin by defining a metric on the configuration space  $\Lambda^n$ . For two configurations  $\sigma$  and  $\tau$  in  $\Lambda^n$ , define

$$(13) \quad d(\sigma, \tau) = \sum_{j=1}^n 1\{\sigma_j \neq \tau_j\}$$

which yields the number of vertices at which the two configurations differ.



Let  $X^t$  and  $Y^t$  be two copies of the Glauber dynamics. Here, we use the standard greedy coupling of  $X^t$  and  $Y^t$ . At each time step a vertex is selected at random, uniformly from the  $n$  vertices. Suppose  $X^t = \sigma$ ,  $Y^t = \tau$ , and the vertex selected is denoted by  $j$ . Next, we erase the spin at location  $j$  in both processes, and replace it with a new one according to the following update probabilities. For all  $\ell = 1, 2, \dots, q$ , define

$$p_\ell = P(\sigma \rightarrow \sigma_{j,e^\ell}) \quad \text{and} \quad q_\ell = P(\tau \rightarrow \tau_{j,e^\ell})$$

and let

$$P_\ell = \min\{p_\ell, q_\ell\} \quad \text{and} \quad P = \sum_{\ell=1}^q P_\ell.$$

Now, let  $B$  be a Bernoulli random variable with probability of success  $P$ . If  $B = 1$ , we update the two chains equally with the following probabilities

$$P(X_j^{t+1} = e^\ell, Y_j^{t+1} = e^\ell \mid B = 1) = \frac{P_\ell}{P}$$

for  $\ell = 1, 2, \dots, q$ . On the other hand, if  $B = 0$ , we update the chains differently according to the following probabilities

$$P(X_j^{t+1} = e^\ell, Y_j^{t+1} = e^m \mid B = 0) = \frac{p_\ell - P_\ell}{1 - P} \cdot \frac{q_m - P_m}{1 - P}$$

for all pairs  $\ell \neq m$ . Then the total probability that the two chains update the same is equal to  $P$  and the total probability that the chains update differently is equal to  $1 - P$ .

Observe that once  $X^t = Y^t$ , the processes remain matched (coupled) for the rest of the time. In the coupling literature, the time

$$\min\{t \geq 0 : X^t = Y^t\}$$

is referred to as the *coupling time*.

The mean coupling distance  $\mathbb{E}[d(X^t, Y^t)]$  is tied to the total variation distance via the following inequality known as the *coupling inequality*:

$$(14) \quad \|P^t(x, \cdot) - P^t(y, \cdot)\|_{\text{TV}} \leq P(X^t \neq Y^t) \leq \mathbb{E}[d(X^t, Y^t)]$$

The above inequality implies that the order of the mean coupling time is an upper bound on the order of the mixing time. See [16] and [15] for details on coupling and coupling inequalities.

## 7. MEAN COUPLING DISTANCE

Fix  $\varepsilon > 0$ . Consider two configurations  $\sigma$  and  $\tau$  such that

$$d(\sigma, \tau) = d,$$

where  $d(\sigma, \tau) \in \mathbb{N}$  is the metric defined in (13) and  $\varepsilon \leq \|L_n(\sigma) - L_n(\tau)\|_1 < 2\varepsilon$ .

Let  $\mathcal{I} = \{i_1, \dots, i_d\}$  be the set of vertices at which the spin values of the two configurations  $\sigma$  and  $\tau$  disagree. Define  $\kappa(e^\ell)$  to be the probability that the coupled processes update differently when the chosen vertex  $j \notin \mathcal{I}$  has spin  $e^\ell$ . If the chosen vertex  $j$  is such that  $\sigma_j = \tau_j = e^\ell$ , then by Corollary 5.2 of Lemma 5.1,

$$\begin{aligned}
\kappa(e^\ell) &:= \frac{1}{2} \sum_{k=1}^q \left| P(\sigma \rightarrow \sigma_{j,e^k}) - P(\tau \rightarrow \tau_{j,e^k}) \right| \\
&= \frac{1}{2} \sum_{k=1}^q \left| \left( g_k^{H,\beta}(L_n(\sigma)) + \frac{\beta}{n} \varphi_{k,e^\ell}^{H,\beta}(L_n(\sigma)) \right) - \left( g_k^{H,\beta}(L_n(\tau)) + \frac{\beta}{n} \varphi_{k,e^\ell}^{H,\beta}(L_n(\tau)) \right) \right| + O\left(\frac{1}{n^2}\right) \\
(15) \quad &= \frac{1}{2} \sum_{k=1}^q \left| g_k^{H,\beta}(L_n(\sigma)) - g_k^{H,\beta}(L_n(\tau)) \right| + O\left(\frac{\varepsilon}{n} + \frac{1}{n^2}\right).
\end{aligned}$$

Next, we observe that for any  $C^2$  function  $f : \mathcal{P} \rightarrow \mathbb{R}$ , there exists  $C > 0$  such that

$$(16) \quad \left| f(z') - f(z) - \langle z' - z, \nabla f(z) \rangle \right| < C\varepsilon^2$$

for all  $z, z' \in \mathcal{P}$  satisfying  $\varepsilon \leq \|z' - z\|_1 < 2\varepsilon$ .

Therefore for  $n$  large enough, there exists  $C' > 0$  such that

$$(17) \quad \left| \kappa(e^\ell) - \frac{1}{2} \sum_{k=1}^q \left| \langle L_n(\tau) - L_n(\sigma), \nabla g_k^{H,\beta}(L_n(\sigma)) \rangle \right| \right| < C'\varepsilon^2.$$

The above result holds regardless of the value of  $\ell \in \{1, 2, \dots, q\}$ .

Similarly, when the chosen vertex  $j \in \mathcal{I}$ , the probability of not coupling at  $j$  satisfies (17).

We conclude that in terms of  $\kappa_{\sigma,\tau} := \frac{1}{2} \sum_{k=1}^q \left| \langle L_n(\tau) - L_n(\sigma), \nabla g_k^{H,\beta}(L_n(\sigma)) \rangle \right|$ , the mean distance between a coupling of the Glauber dynamics starting in  $\sigma$  and  $\tau$  with  $d(\sigma, \tau) = d$  after one step has the form

$$\begin{aligned}
\mathbb{E}_{\sigma,\tau}[d(X, Y)] &\leq d - \frac{d}{n}(1 - \kappa_{\sigma,\tau}) + \frac{n-d}{n}\kappa_{\sigma,\tau} + c\varepsilon^2 \\
(18) \quad &= d \cdot \left[ 1 - \frac{1}{n} \left( 1 - \frac{\kappa_{\sigma,\tau} + c\varepsilon^2}{d/n} \right) \right]
\end{aligned}$$

for a fixed  $c > 0$  and all  $\varepsilon$  small enough.

## 8. AGGREGATE PATH COUPLING

In the previous section, we derived the form of the mean distance between a coupling of the Glauber dynamics starting in two configurations whose distance is bounded. We next derive the form of the mean coupling distance of a coupling starting in two configurations that are connected by a path of configurations where the distance between successive configurations are bounded.

**Definition 8.1.** Let  $\sigma$  and  $\tau$  be configurations in  $\Lambda^n$ . We say that a path  $\pi$  connecting configurations  $\sigma$  and  $\tau$  denoted by

$$\pi : \sigma = x_0, x_1, \dots, x_r = \tau,$$

is a **monotone path** if

- (i)  $\sum_{i=1}^r d(x_{i-1}, x_i) = d(\sigma, \tau)$
- (ii) for each  $k = 1, 2, \dots, q$ , the  $k$ th coordinate of  $L_n(x_i)$ ,  $L_{n,k}(x_i)$  is monotonic as  $i$  increases from 0 to  $r$ ;

Observe that here the points  $x_i$  on the path are not required to be nearest-neighbors.

A straightforward property of monotone paths is that

$$\sum_{i=1}^r \sum_{k=1}^q L_{n,k}(x_i) - L_{n,k}(x_{i-1}) = L_n(\sigma) - L_n(\tau)$$

Another straightforward observation is that for any given path

$$L_n(\sigma) = z_0, z_1, \dots, z_r = L_n(\tau)$$

in  $\mathcal{P}_n$ , monotone in each coordinate, with  $\|z_i - z_{i-1}\|_1 > 0$  for all  $i \in \{1, 2, \dots, r\}$ , there exists a monotone path

$$\pi : \sigma = x_0, x_1, \dots, x_r = \tau$$

such that  $L_n(x_i) = z_i$  for each  $i$ .

Let  $\pi : \sigma = x_0, x_1, \dots, x_r = \tau$  be a monotone path connecting configurations  $\sigma$  and  $\tau$  such that  $\varepsilon \leq \|L_n(x_i) - L_n(x_{i-1})\|_1 < 2\varepsilon$  for all  $i = 1, \dots, r$ . Equation (18) implies the following bound on the mean distance between a coupling of the Glauber dynamics starting in configurations  $\sigma$  and  $\tau$ :

$$\begin{aligned} & \mathbb{E}_{\sigma, \tau}[d(X, Y)] \\ & \leq \sum_{i=1}^r \mathbb{E}_{x_{i-1}, x_i}[d(X_{i-1}, X_i)] \\ & \leq \sum_{i=1}^r \left\{ d(x_{i-1}, x_i) \cdot \left[ 1 - \frac{1}{n} \left( 1 - \frac{\frac{1}{2} \sum_{k=1}^q \left| \left\langle L_n(x_i) - L_n(x_{i-1}), \nabla g_k^{H, \beta}(L_n(x_{i-1})) \right\rangle \right| + c\varepsilon^2}{d(x_{i-1}, x_i)/n} \right) \right] \right\} \\ & = d(\sigma, \tau) \left[ 1 - \frac{1}{n} \left( 1 - \frac{\sum_{k=1}^q \sum_{i=1}^r \left| \left\langle L_n(x_i) - L_n(x_{i-1}), \nabla g_k^{H, \beta}(L_n(x_{i-1})) \right\rangle \right| + c\varepsilon^2}{2d(\sigma, \tau)/n} \right) \right] \\ (19) \quad & \leq d(\sigma, \tau) \left[ 1 - \frac{1}{n} \left( 1 - \frac{\sum_{k=1}^q \sum_{i=1}^r \left| \left\langle L_n(x_i) - L_n(x_{i-1}), \nabla g_k^{H, \beta}(L_n(x_{i-1})) \right\rangle \right| + c\varepsilon^2}{\|L_n(\sigma) - L_n(\tau)\|_1} \right) \right], \end{aligned}$$

as  $\sum_{i=1}^r d(x_{i-1}, x_i) = d(\sigma, \tau)$ .

From inequality (19), if there exists monotone paths between all pairs of configurations such that there is a uniform bound less than 1 on the ratio

$$\frac{\sum_{k=1}^q \sum_{i=1}^r \left| \left\langle L_n(x_i) - L_n(x_{i-1}), \nabla g_k^{H,\beta}(L_n(x_{i-1})) \right\rangle \right|}{\|L_n(\sigma) - L_n(\tau)\|_1}$$

then the mean coupling distance contracts which yields a bound on the mixing time via coupling inequality (14).

Although the Gibbs measure are distributions of the empirical measure  $L_n$  defined on the discrete space  $\mathcal{P}_n$ , proving contraction of the mean coupling distance is often facilitated by working in the continuous space, namely the simplex  $\mathcal{P}$ . We begin our discussion of aggregate path coupling by defining distances along paths in  $\mathcal{P}$ .

Recall the function  $g^{H,\beta}$  defined in (12) which is dependent on the Hamiltonian of the model through the interaction representation function  $H$  defined in Definition 2.2.

**Definition 8.2.** Define the **aggregate  $g$ -variation** between a pair of points  $x$  and  $z$  in  $\mathcal{P}$  along a continuous monotone (in each coordinate) path  $\rho$  to be

$$D_\rho^g(x, z) := \sum_{k=1}^q \int_\rho \left| \left\langle \nabla g_k^{H,\beta}(y), dy \right\rangle \right|$$

Define the corresponding **pseudo-distance** between a pair of points  $x$  and  $z$  in  $\mathcal{P}$  as

$$d_g(x, z) := \inf_\rho D_\rho^g(x, z),$$

where the infimum is taken over all continuous monotone paths in  $\mathcal{P}$  connecting  $x$  and  $z$ .

Notice if the monotonicity restriction is removed, the above infimum would satisfy the triangle inequality. We will need the following condition.

**Condition 8.3.** Let  $z_\beta$  be the unique equilibrium macrostate. There exists  $\delta \in (0, 1)$  such that

$$\frac{d_g(z, z_\beta)}{\|z - z_\beta\|_1} \leq 1 - \delta$$

for all  $z$  in  $\mathcal{P}$ .

Observe that if it is shown that  $d_g(z, z_\beta) < \|z - z_\beta\|_1$  for all  $z$  in  $\mathcal{P}$ , then by continuity the above condition is equivalent to

$$\limsup_{z \rightarrow z_\beta} \frac{d_g(z, z_\beta)}{\|z - z_\beta\|_1} < 1$$

Suppose Condition 8.3 is satisfied. Then let denote by  $\mathbf{NG}_\delta$  the family of **neo-geodesic** smooth curves, monotone in each coordinate such that for each  $z \neq z_\beta$  in  $\mathcal{P}$ , there is exactly one curve  $\rho = \rho_z$  in the family  $\mathbf{NG}_\delta$  connecting  $z_\beta$  to  $z$ , and

$$\frac{D_{\rho_z}^g(z, z_\beta)}{\|z - z_\beta\|_1} \leq 1 - \delta/2$$

**Condition 8.4.** For  $\varepsilon > 0$  small enough, there exists a neo-geodesic family  $\mathbf{NG}_\delta$  such that for each  $z$  in  $\mathcal{P}$  satisfying  $\|z - z_\beta\|_1 \geq \varepsilon$ , the curve  $\rho = \rho_z$  in the family  $\mathbf{NG}_\delta$  that connects  $z_\beta$  to  $z$  satisfies

$$\frac{\sum_{k=1}^q \sum_{i=1}^r \left| \left\langle z_i - z_{i-1}, \nabla g_k^{H,\beta}(z_{i-1}) \right\rangle \right|}{\|z - z_\beta\|_1} \leq 1 - \delta/3$$

for a sequence of points  $z_0 = z_\beta, z_1, \dots, z_r = z$  interpolating  $\rho$  such that

$$\varepsilon \leq \|z_i - z_{i-1}\|_1 < 2\varepsilon \quad \text{for } i = 1, 2, \dots, r.$$

It is important to observe that Condition 8.3 is often simpler to verify than Condition 8.4. Moreover, under certain simple additional prerequisites, Condition 8.3 implies Condition 8.4. For example, this is achieved if there is a uniform bound on the Cauchy curvature at every point of every curve in  $\mathbf{NG}_\delta$ . So it will be demonstrated on the example of Curie-Weiss-Potts model that the natural way for establishing Condition 8.4 for the model is via first establishing Condition 8.3.

In addition to Condition 8.4 that will be shown to imply contraction when one of the two configurations in the coupled processes is at the equilibrium, i.e.  $L_n(\sigma) = z_\beta$ , we need a condition that will imply contraction between two configurations within a neighborhood of the equilibrium configuration. We state this assumption next.

**Condition 8.5.** Let  $z_\beta$  be the unique equilibrium macrostate. Then,

$$\limsup_{z \rightarrow z_\beta} \frac{\|g^{H,\beta}(z) - g^{H,\beta}(z_\beta)\|_1}{\|z - z_\beta\|_1} < 1.$$

Since  $H(z) \in \mathcal{C}^3$ , the above Condition 8.5 implies that for any  $\varepsilon > 0$  sufficiently small, there exists  $\gamma \in (0, 1)$  such that

$$\frac{\|g^{H,\beta}(z) - g^{H,\beta}(w)\|_1}{\|z - w\|_1} < 1 - \gamma$$

for all  $z$  and  $w$  in  $\mathcal{P}$  satisfying

$$\|z - z_\beta\|_1 < \varepsilon \quad \text{and} \quad \|w - z_\beta\|_1 < \varepsilon.$$

## 9. MAIN RESULT

A sufficient condition for rapid mixing of the Glauber dynamics of Gibbs ensembles is contraction of the mean coupling distance  $\mathbb{E}_{\sigma,\tau}[d(X, Y)]$  between coupled processes starting in all pairs of configurations in  $\Lambda^n$ . The classical path coupling argument [2] is a method of obtaining this contraction by only proving contraction between couplings starting in neighboring configurations. However for some classes of models (e.g. models that undergo a first-order, discontinuous phase transition) there are situations when Glauber dynamics exhibits rapid mixing, but coupled processes do not exhibit contraction between some neighboring configurations. Such models include the mean-field Blume-Capel (in the discontinuous phase transition region) and Curie-Weiss-Potts models. A major strength of the **aggregate path coupling** method introduced in [13] for mean-field Blume-Capel model and further expanded in this study is that, in addition to

its generality, it yields a proof for rapid mixing even in those cases when contraction of the mean distance between couplings starting in all pairs of neighboring configurations does not hold.

The strategy is to take advantage of the large deviations estimates discussed in section 3. Recall from that section that we assume that the set of equilibrium macrostates  $\mathcal{E}_\beta$ , which can be expressed in the form given in (9), consists of a single point  $z_\beta$ . Define an *equilibrium configuration*  $\sigma_\beta$  to be a configuration such that

$$L_n(\sigma_\beta) = z_\beta = ((z_\beta)_1, (z_\beta)_2, \dots, (z_\beta)_q).$$

First we observe that in order to use the coupling inequality (14) we need to show contraction of the mean coupling distance  $\mathbb{E}_{\sigma, \tau}[d(X, Y)]$  between a Markov chain initially distributed according to the stationary probability distribution  $P_{n, \beta}$  and a Markov chain starting at any given configuration. Using large deviations we know that with high probability the former process starts near the equilibrium and stays near the equilibrium for long duration of time.

Our main result Theorem 9.2 states that once we establish contraction of the mean coupling distance between two copies of a Markov chain where one of the coupled dynamics starts near an equilibrium configuration in Lemma 9.1, then this contraction, along with the large deviations estimates of the empirical measure  $L_n$ , yields rapid mixing of the Glauber dynamics converging to the Gibbs measure.

Now, the classical path coupling relies on showing contraction along any monotone path connecting two configurations, in one time step. Here we observe that we only need to show contraction along one monotone path connecting two configurations in order to have the mean coupling distance  $\mathbb{E}_{\sigma, \tau}[d(X, Y)]$  contract in a single time step. However, finding even one monotone path with which we can show contraction in the equation (19) is not easy. The answer to this is in finding a monotone path  $\rho$  in  $\mathcal{P}$  connecting the  $L_n$  values of the two configurations,  $\sigma$  and  $\tau$ , such that

$$\frac{\sum_{k=1}^q \int_{\rho} \left| \left\langle \nabla g_k^{H, \beta}(y), dy \right\rangle \right|}{\|L_n(\sigma) - L_n(\tau)\|_1} < 1$$

Although  $\rho$  is a continuous path in continuous space  $\mathcal{P}$ , it serves as Ariadne's thread for finding a monotone path

$$\pi : \sigma = x_0, x_1, \dots, x_r = \tau$$

such that  $L_n(x_0), L_n(x_1), \dots, L_n(x_r)$  in  $\mathcal{P}_n$  are positioned along  $\rho$ , and

$$\sum_{k=1}^q \sum_{i=1}^r \left| \left\langle L_n(x_i) - L_n(x_{i-1}), \nabla g_k^{H, \beta}(L_n(x_{i-1})) \right\rangle \right|$$

is a Riemann sum approximating  $\sum_{k=1}^q \int_{\rho} \left| \left\langle \nabla g_k^{H, \beta}(y), dy \right\rangle \right|$ . Therefore we obtain

$$\frac{\sum_{k=1}^q \sum_{i=1}^r \left| \left\langle L_n(x_i) - L_n(x_{i-1}), \nabla g_k^{H, \beta}(L_n(x_{i-1})) \right\rangle \right|}{\|L_n(\sigma) - L_n(\tau)\|_1} < 1,$$

that in turn implies contraction in (19) for  $\varepsilon$  small enough and  $n$  large enough. See Figure 1.

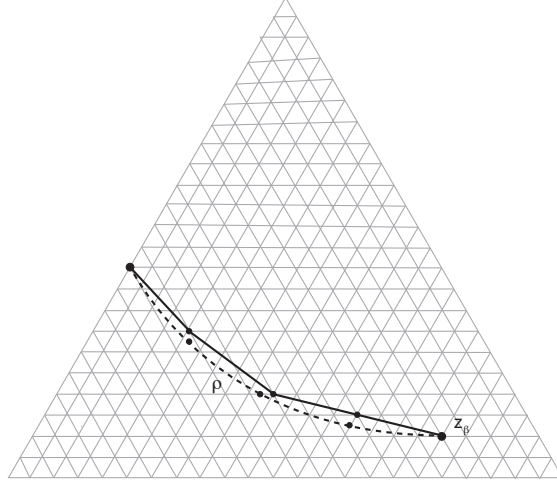


FIGURE 1. Case  $q = 3$ . Dashed curve is the continuous monotone path  $\rho$ . Solid lines represent the path  $L_n(x_0), L_n(x_1), \dots, L_n(x_r)$  in  $\mathcal{P}_n$ .

Observe that in order for the above argument to work, we need to spread points  $L_n(x_i) \in \mathcal{P}_n$  along a continuous path  $\rho$  at intervals of fixed order  $\varepsilon$ . Thus  $\pi$  has to be **not a nearest-neighbor path** in the space of configurations, another significant deviation from the classical path coupling.

**Lemma 9.1.** *Assume Condition 8.4 and Condition 8.5. Let  $(X, Y)$  be a coupling of the Glauber dynamics as defined in Section 6, starting in configurations  $\sigma$  and  $\tau$  and let  $z_\beta$  be the single equilibrium macrostate of the corresponding Gibbs ensemble. Then there exists an  $\alpha > 0$  and an  $\varepsilon' > 0$  small enough such that for  $n$  large enough,*

$$\mathbb{E}_{\sigma, \tau}[d(X, Y)] \leq e^{-\alpha/n} d(\sigma, \tau)$$

whenever  $\|L_n(\sigma) - z_\beta\|_1 < \varepsilon'$ .

*Proof.* Let  $\varepsilon$  and  $\delta$  be as in Condition 8.4, and let  $\varepsilon' = \varepsilon^2 \delta / M$  with a constant  $M \gg 0$ .

**Case I.** Suppose  $L_n(\tau) = z$  and  $L_n(\sigma) = w$ , where  $\|z - z_\beta\|_1 \geq \varepsilon$  and  $\|w - z_\beta\|_1 < \varepsilon'$ .

Then there is an equilibrium configuration  $\sigma_\beta$  with  $L_n(\sigma_\beta) = z_\beta$  such that there is a *monotone path*

$$\pi' : \sigma_\beta = x'_0, x'_1, \dots, x'_r = \tau$$

connecting configurations  $\sigma_\beta$  and  $\tau$  on  $\Lambda^n$  such that  $\varepsilon \leq \|L_n(x'_i) - L_n(x'_{i-1})\|_1 < 2\varepsilon$ , and by Condition 8.4,

$$\frac{\sum_{k=1}^q \sum_{i=1}^r \left| \left\langle L_n(x'_i) - L_n(x'_{i-1}), \nabla g_k^{H, \beta}(L_n(x'_{i-1})) \right\rangle \right|}{\|L_n(\sigma_\beta) - L_n(\tau)\|_1} \leq 1 - \delta/4$$

for  $n$  large enough. Note that the difference between the above inequality and Condition 8.4 is that here we take  $L_n(x'_i) \in \mathcal{P}_n$ .

Now, there exists a monotone path with from  $\sigma$  to  $\tau$

$$\pi : \sigma = x_0, x_1, \dots, x_r = \tau$$

such that

$$\|L_n(x_i) - L_n(x'_i)\|_1 \leq \varepsilon' \quad \text{for all } i = 0, 1, \dots, r.$$

The new monotone path  $\pi$  is constructed from  $\pi'$  by insuring that either

$$0 \leq \langle L_n(x_i) - L_n(x_{i-1}), e^k \rangle \leq \langle L_n(x'_i) - L_n(x'_{i-1}), e^k \rangle$$

or

$$\langle L_n(x'_i) - L_n(x'_{i-1}), e^k \rangle \leq \langle L_n(x_i) - L_n(x_{i-1}), e^k \rangle \leq 0$$

for  $i = 2, \dots, r$  and each coordinate  $k \in \{1, 2, \dots, q\}$ .

Then

$$\left| \frac{\sum_{k=1}^q \sum_{i=1}^r \left| \langle L_n(x_i) - L_n(x_{i-1}), \nabla g_k^{H,\beta}(L_n(x_{i-1})) \rangle \right|}{\|L_n(\sigma) - L_n(\tau)\|_1} - \frac{\sum_{k=1}^q \sum_{i=1}^r \left| \langle L_n(x'_i) - L_n(x'_{i-1}), \nabla g_k^{H,\beta}(L_n(x'_{i-1})) \rangle \right|}{\|L_n(\sigma_\beta) - L_n(\tau)\|_1} \right| \leq C'' r \varepsilon' / \varepsilon$$

for a fixed constant  $C'' > 0$ . Noticing that  $r \varepsilon' / \varepsilon \leq \delta / M$  as  $r \leq 1 / \varepsilon$ , and taking  $M$  large enough, we obtain

$$\frac{\sum_{k=1}^q \sum_{i=1}^r \left| \langle L_n(x_i) - L_n(x_{i-1}), \nabla g_k^{H,\beta}(L_n(x_{i-1})) \rangle \right|}{\|L_n(\sigma) - L_n(\tau)\|_1} \leq 1 - \delta / 4.$$

Thus equation (19) will imply

$$\begin{aligned} \mathbb{E}_{\sigma, \tau}[d(X, Y)] &\leq d(\sigma, \tau) \left[ 1 - \frac{1}{n} \left( 1 - \frac{\sum_{k=1}^q \sum_{i=1}^r \left| \langle L_n(x_i) - L_n(x_{i-1}), \nabla g_k^{H,\beta}(L_n(x_{i-1})) \rangle \right| + c \varepsilon^2}{\|L_n(\sigma) - L_n(\tau)\|_1} \right) \right] \\ &\leq d(\sigma, \tau) \left[ 1 - \frac{1}{n} (1 - (1 - \delta / 4) - \delta / 20) \right] \\ &= d(\sigma, \tau) \left[ 1 - \frac{1}{n} \delta / 5 \right] \end{aligned}$$

as  $\frac{c \varepsilon^2}{\|L_n(\sigma) - L_n(\tau)\|_1} \leq c \varepsilon \leq \delta / 20$  for  $\varepsilon$  small enough.

**Case II.** Suppose  $L_n(\tau) = z$  and  $L_n(\sigma) = w$ , where  $\|z - z_\beta\|_1 < \varepsilon$  and  $\|w - z_\beta\|_1 < \varepsilon'$ .



Similarly to (18), equation (15) implies for  $n$  large enough,

$$\begin{aligned} \mathbb{E}[d(X, Y)] &\leq d(\sigma, \tau) \cdot \left[ 1 - \frac{1}{n} \left( 1 - \frac{\|g^{H, \beta}(L_n(\sigma)) - g^{H, \beta}(L_n(\tau))\|_1}{\|L_n(\sigma) - L_n(\tau)\|_1} \right) \right] + O\left(\frac{1}{n^2}\right) \\ &\leq d(\sigma, \tau) \cdot \left[ 1 - \frac{\gamma}{n} \right] + O\left(\frac{1}{n^2}\right) \\ &\leq d(\sigma, \tau) \cdot \left[ 1 - \frac{\gamma}{2n} \right] \end{aligned}$$

by Condition 8.5 (see also discussion following Condition 8.5).  $\square$

We now state and prove the main theorem of the paper that yields sufficient conditions for rapid mixing of the Glauber dynamics of the class of statistical mechanical models discussed.

**Theorem 9.2.** *Suppose  $H(z)$  and  $\beta > 0$  are such that Condition 8.4 and Condition 8.5 are satisfied. Then the mixing time of the Glauber dynamics satisfies*

$$t_{mix} = O(n \log n)$$

*Proof.* Let  $\varepsilon' > 0$  and  $\alpha > 0$  be as in Lemma 9.1. Let  $(X^t, Y^t)$  be a coupling of the Glauber dynamics such that  $X^0 \stackrel{dist}{=} P_{n, \beta}$ , the stationary distribution. Then, for sufficiently large  $n$ ,

$$\begin{aligned} &\|P^t(Y^0, \cdot) - P_{n, \beta}\|_{\text{TV}} \\ &\leq P\{X^t \neq Y^t\} \\ &= P\{d(X^t, Y^t) \geq 1\} \\ &\leq \mathbb{E}[d(X^t, Y^t)] \\ &= \mathbb{E}[\mathbb{E}[d(X^t, Y^t) \mid X^{t-1}, Y^{t-1}]] \\ &\leq \mathbb{E}[\mathbb{E}[d(X^t, Y^t) \mid X^{t-1}, Y^{t-1}] \mid \|L_n(X^{t-1}) - z_\beta\|_1 < \varepsilon'] \cdot P\{\|L_n(X^{t-1}) - z_\beta\|_1 < \varepsilon'\} \\ &\quad + nP\{\|L_n(X^{t-1}) - z_\beta\|_1 \geq \varepsilon'\}. \end{aligned}$$

By Lemma 9.1, we have

$$(20) \quad \begin{aligned} &\mathbb{E}[\mathbb{E}[d(X^t, Y^t) \mid X^{t-1}, Y^{t-1}] \mid \|L_n(X^{t-1}) - z_\beta\|_1 < \varepsilon'] \\ &\leq e^{-\alpha/n} \mathbb{E}[d(X^{t-1}, Y^{t-1}) \mid \|L_n(X^{t-1}) - z_\beta\|_1 < \varepsilon'] \end{aligned}$$

By iterating (20), it follows that

$$\begin{aligned}
& \|P^t(Y^0, \cdot) - P_{n,\beta,K}\|_{\text{TV}} \\
& \leq e^{-\alpha/n} \mathbb{E}[d(X^{t-1}, Y^{t-1}) \mid \|L_n(X^{t-1}) - z_\beta\|_1 < \varepsilon'] \cdot P\{\|L_n(X_{t-1}) - z_\beta\|_1 < \varepsilon'\} \\
& \quad + nP\{\|L_n(X^{t-1}) - z_\beta\|_1 \geq \varepsilon'\} \\
& \leq e^{-\alpha/n} \mathbb{E}[d(X^{t-1}, Y^{t-1})] + nP\{\|L_n(X^{t-1}) - z_\beta\|_1 \geq \varepsilon'\} \\
& \quad \vdots \\
& \leq e^{-\alpha t/n} \mathbb{E}[d(X^0, Y^0)] + n \sum_{s=0}^{t-1} P\{\|L_n(X^s) - z_\beta\|_1 \geq \varepsilon'\} \\
& = e^{-\alpha t/n} \mathbb{E}[d(X^0, Y^0)] + ntP_{n,\beta}\{\|L_n(X^0) - z_\beta\|_1 \geq \varepsilon'\} \\
& \leq ne^{-\alpha t/n} + ntP_{n,\beta}\{\|L_n(X^0) - z_\beta\|_1 \geq \varepsilon'\}.
\end{aligned}$$

We recall the LDP limit (6) for  $\beta$  in the single phase region  $B$ ,

$$P_{n,\beta}\{L_n(X^0) \in dx\} \implies \delta_{z_\beta} \quad \text{as } n \rightarrow \infty.$$

Moreover, for any  $\gamma' > 1$  and  $n$  sufficiently large, by the LDP upper bound (4), we have

$$\begin{aligned}
\|P^t(Y^0, \cdot) - P_{n,\beta}\|_{\text{TV}} & \leq ne^{-\alpha t/n} + ntP_{n,\beta}\{\|L_n(X^0) - z_\beta\|_1 \geq \varepsilon'\} \\
& < ne^{-\alpha t/n} + tne^{-\frac{n}{\gamma'} I_\beta(\varepsilon')}.
\end{aligned}$$

For  $t = \frac{n}{\alpha}(\log n + \log(2/\varepsilon'))$ , the above right-hand side converges to  $\varepsilon'/2$  as  $n \rightarrow \infty$ .  $\square$

## 10. AGGREGATE PATH COUPLING APPLIED TO THE GENERALIZED POTTS MODEL

In this section, we illustrate the strength of our main result of section 9, Theorem 9.2, by applying it to the generalized Curie-Weiss-Potts model (GCWP), studied recently in [12]. The classical Curie-Weiss-Potts (CWP) model, which is the mean-field version of the well known Potts model of statistical mechanics [19] is a particular case of the GCWP model with  $r = 2$ . While the mixing times for the CWP model has been studied in [4], these are the first results for the mixing times of the GCWP model. Moreover, the application of our methods gives a significantly shorter derivation for the region of rapid mixing than the one used for the CWP model in [4].

Let  $q$  be a fixed integer and define  $\Lambda = \{e^1, e^2, \dots, e^q\}$ , where  $e^k$  are the  $q$  standard basis vectors of  $\mathbb{R}^q$ . A *configuration* of the model has the form  $\omega = (\omega_1, \omega_2, \dots, \omega_n) \in \Lambda^n$ . We will consider a configuration on a graph with  $n$  vertices and let  $X_i(\omega) = \omega_i$  be the *spin* at vertex  $i$ . The random variables  $X_i$ 's for  $i = 1, 2, \dots, n$  are independent and identically distributed with common distribution  $\rho$ .

For the generalized Curie-Weiss-Potts model, for  $r \geq 2$ , the interaction representation function as in Definition 2.2, has the form

$$H^r(z) = -\frac{1}{r} \sum_{j=1}^q z_j^r$$

and the generalized Curie-Weiss-Potts model is defined as the Gibbs measure

$$(21) \quad P_{n,\beta,r}(B) = \frac{1}{Z_n(\beta)} \int_B \exp\{-\beta n H^r(L_n(\omega))\} dP_n$$

where  $L_n(\omega)$  is the empirical measure defined in (1).

In [12], the authors proved that there exists a phase transition critical value  $\beta_c(q, r)$  such that in the parameter regime  $(q, r) \in \{2\} \times [2, 4]$ , the GCWP model undergoes a continuous, second-order, phase transition and for  $(q, r)$  in the complementary regime, the GCWP model undergoes a discontinuous, first-order, phase transition. This is stated in the following theorem.

**Theorem 10.1** (Generalized Ellis-Wang Theorem). *Assume that  $q \geq 2$  and  $r \geq 2$ . Then there exists a critical temperature  $\beta_c(q, r) > 0$  such that in the weak limit*

$$\lim_{n \rightarrow \infty} P_{n,\beta,r}(L_n \in \cdot) = \begin{cases} \delta_{1/q(1,\dots,1)} & \text{if } \beta < \beta_c(q, r) \\ \frac{1}{q} \sum_{i=1}^q \delta_{u(\beta,q,r)e^i + (1-u(\beta,q,r))/q(1,\dots,1)} & \text{if } \beta > \beta_c(q, r) \end{cases}$$

where  $u(\beta, q, r)$  is the largest solution to the so-called mean-field equation

$$u = \frac{1 - \exp(\Delta(u))}{1 + (q-1)\exp(\Delta(u))}$$

with  $\Delta(u) := -\frac{\beta}{q^r-1} [(1+(q-1)u)^{r-1} - (1-u)^{r-1}]$ . Moreover, for  $(q, r) \in \{2\} \times [2, 4]$ , the function  $\beta \mapsto u(\beta, q, r)$  is continuous whereas, in the complementary case, the function is discontinuous at  $\beta_c(q, r)$ .

For the GCWP model, the function  $g_\ell^{H,\beta}(z)$  defined in general in (11) has the form

$$g_k^{H,\beta}(z) = [\partial_k \Gamma](\beta \nabla H(z)) = [\partial_k \Gamma](\beta z) = \frac{e^{\beta z_k^{r-1}}}{e^{\beta z_1^{r-1}} + \dots + e^{\beta z_q^{r-1}}}.$$

For the remainder of this section, we will replace the notation  $H, \beta$  and refer to  $g^{H,\beta}(z) = (g_1^{H,\beta}(z), \dots, g_q^{H,\beta}(z))$  as simply  $g^r(z) = (g_1^r(z), \dots, g_q^r(z))$ . As we will prove next, the rapid mixing region for the GCWP model is defined by the following value.

$$(22) \quad \beta_s(q, r) := \sup \{\beta \geq 0 : g_k^r(z) < z_k \text{ for all } z \in \mathcal{P} \text{ such that } z_k \in (1/q, 1]\}.$$

**Lemma 10.2.** *If  $\beta_c(q, r)$  is the critical value derived in [12] and defined in Theorem 10.1, then*

$$\beta_s(q, r) \leq \beta_c(q, r)$$

*Proof.* We will prove this lemma by contradiction. Suppose  $\beta_c(q, r) < \beta_s(q, r)$ . Then there exists  $\beta$  such that

$$\beta_c(q, r) < \beta < \beta_s(q, r).$$

Then, by Theorem 10.1, since  $\beta_c(q, r) < \beta$ , there exists  $u > 0$  satisfying the following inequality

$$(23) \quad u < \frac{1 - e^{\Delta(u)}}{1 + (q-1)e^{\Delta(u)}},$$

where  $\Delta(u) := -\frac{\beta}{q^{r-1}} [(1 + (q-1)u)^{r-1} - (1-u)^{r-1}]$ . Here, the above inequality (23) rewrites as

$$(24) \quad e^{\Delta(u)} = \exp \left\{ \beta \left[ \left( \frac{1-u}{q} \right)^{r-1} - \left( \frac{1+(q-1)u}{q} \right)^{r-1} \right] \right\} < \frac{1-u}{(q-1)u+1}.$$

Next, we substitute  $\lambda = (1-u)^{\frac{q-1}{q}}$  into the above inequality (24), obtaining

$$(25) \quad \exp \left\{ \beta \left[ \left( \frac{\lambda}{q-1} \right)^{r-1} - (1-\lambda)^{r-1} \right] \right\} < \frac{\lambda}{(1-\lambda)(q-1)}.$$

Now, consider

$$z = \left( 1-\lambda, \frac{\lambda}{q-1}, \dots, \frac{\lambda}{q-1} \right).$$

Observe that  $z_1 = 1-\lambda = 1 - (1-u)^{\frac{q-1}{q}} = \frac{1+u(q-1)}{q} > \frac{1}{q}$  as  $u > 0$ . Here, the inequality (25) can be consequently rewritten in terms of the above selected  $z$  as follows

$$z_1 = 1-\lambda < \frac{e^{\beta(1-\lambda)^{r-1}}}{e^{\beta(1-\lambda)^{r-1}} + (q-1)e^{\beta\left(\frac{\lambda}{q-1}\right)^{r-1}}} = g_1^r(z),$$

thus contradicting  $\beta < \beta_s(q, r)$ . Hence  $\beta_s(q, r) \leq \beta_c(q, r)$ .  $\square$

Combining Theorem 10.1 and Lemma 10.2 yields that for parameter values  $(q, r)$  in the continuous, second-order phase transition region  $\beta_s(q, r) = \beta_c(q, r)$ , whereas in the discontinuous, first-order, phase transition region,  $\beta_s(q, r)$  is strictly less than  $\beta_c(q, r)$ . This relationship between the equilibrium transition critical value and the mixing time transition critical value was also proved for the mean-field Blume-Capel model discussed in [13]. This appears to be a general distinguishing feature between models that exhibit the two distinct types of phase transition. We now prove rapid mixing for the generalized Curie-Weiss-Potts model for  $\beta < \beta_s(q, r)$  using the aggregate path coupling method derived in section 9.

We state the lemmas that we prove below, and the main result for the Glauber dynamics of the generalized Curie-Weiss-Potts model, a Corollary to Theorem 9.2.

**Lemma 10.3.** *Condition 8.3 and Condition 8.4 are satisfied for all  $\beta < \beta_s(q, r)$ .*

**Lemma 10.4.** *Condition 8.5 is satisfied for all  $\beta < \beta_s(q, r)$ .*

**Corollary 10.5.** *If  $\beta < \beta_s(q, r)$ , then*

$$t_{\text{mix}} = O(n \log n).$$

*Proof.* Condition 8.4 and Condition 8.5 required for Theorem 9.2 are satisfied by Lemma 10.3 and Lemma 10.4.  $\square$

*Proof of Lemma 10.4.* Denote  $z' = (z'_1, \dots, z'_q) = z - z_\beta$ . Then by Taylor's Theorem, we have

$$\begin{aligned}
 \limsup_{z \rightarrow z_\beta} \frac{\|g^r(z) - g^r(z_\beta)\|_1}{\|z - z_\beta\|_1} &= \limsup_{z \rightarrow z_\beta} \frac{\sum_{k=1}^q \left| \frac{e^{\beta z_k^{r-1}}}{\sum_{j=1}^q e^{\beta z_j^{r-1}}} - \frac{1}{q} \right|}{\sum_{k=1}^q \left| z_k - \frac{1}{q} \right|} \\
 &= \lim_{z' \rightarrow 0} \frac{\sum_{k=1}^q \left| \frac{\beta(r-1) \left(\frac{1}{q}\right)^{r-2} z'_k + O((z'_1)^2 + \dots + (z'_q)^2)}{q + O((z'_1)^2 + \dots + (z'_q)^2)} \right|}{\sum_{k=1}^q |z'_k|} \\
 (26) \qquad \qquad \qquad &= \frac{\beta(r-1)}{q^{r-1}}.
 \end{aligned}$$

Recall that  $\beta_s(q, r) \leq \beta_c(q, r)$  was shown in Lemma 10.2, and  $\beta_c(q, r) \leq \frac{q^{r-1}}{r-1}$  was shown in the proof of Lemma 5.4 of [12]. Therefore,  $\beta < \frac{q^{r-1}}{r-1}$  and the last expression above is less than 1, and we conclude that

$$\limsup_{z \rightarrow z_\beta} \frac{\|g^r(z) - g^r(z_\beta)\|_1}{\|z - z_\beta\|_1} < 1. \quad \square$$

*Proof of Lemma 10.3.* First, we prove that the family of straight lines connecting to the equilibrium point  $z_\beta = (1/q, \dots, 1/q)$  is a neo-geodesic family as it was defined following Condition 8.3. Specifically, for any  $z = (z_1, z_2, \dots, z_q) \in \mathcal{P}$  define the line path  $\rho$  connecting  $z$  to  $z_\beta$  by

$$(27) \qquad \qquad \qquad z(t) = \frac{1}{q}(1-t) + zt, \quad 0 \leq t \leq 1$$

Then, along this straight-line path  $\rho$ , the aggregate  $g$ -variation has the form

$$D_\rho^g(z, z_\beta) := \sum_{k=1}^q \int_\rho \left| \langle \nabla g_k^r(y), dy \rangle \right| = \sum_{k=1}^q \int_0^1 \left| \frac{d}{dt} [g_k^r(z(t))] \right| dt$$

Next, for all  $k = 1, 2, \dots, q$  and  $t \in [0, 1]$ , denote

$$z(t)_k = \frac{1}{q}(1-t) + z_k t$$

Then

$$(28) \qquad \qquad \qquad g_k^r(z(t)) = \frac{e^{\beta \left(\frac{1}{q}(1-t) + z_k t\right)^{r-1}}}{\sum_{j=1}^q e^{\beta \left(\frac{1}{q}(1-t) + z_j t\right)^{r-1}}}$$

and

$$(29) \qquad \frac{d}{dt} [g_k^r(z(t))] = \beta(r-1) g_k^r(z(t)) \left[ \left( \frac{1}{q}(1-t) + z_k t \right)^{r-2} \left( z_k - \frac{1}{q} \right) - \langle z - z_\beta, g^r(z(t)) \rangle_\rho \right]$$

where  $\langle z - z_\beta, g^r(z(t)) \rangle_\rho$  is the weighted inner product

$$\langle z - z_\beta, g^r(z(t)) \rangle_\rho := \sum_{j=1}^q g_j^r(z(t)) \left( z_k - \frac{1}{q} \right) \left( \frac{1}{q}(1-t) + z_k t \right)^{r-2}$$

Now, observe that for  $z(t)$  as in (27) with  $z \neq z_\beta$ , the inner product  $\langle (z - z_\beta), g^r(z(t)) \rangle_\rho$  is monotonically increasing in  $t$  since

$$\frac{d}{dt} \langle z - z_\beta, g^r(z(t)) \rangle_\rho \geq \beta(r-1) \text{Var}_{g^r} \left( \left( z_k - \frac{1}{q} \right) \left( \frac{1}{q}(1-t) + z_k t \right)^{r-1} \right) > 0$$

where  $\text{Var}_{g^r}(\cdot)$  is the variance with respect to  $g^r$ .

So  $\langle z - z_\beta, g^r(z(t)) \rangle_\rho$  begins at  $\langle z - z_\beta, g^r(z(0)) \rangle_\rho = \langle z - z_\beta, z_\beta \rangle = 0$  and increases for all  $t \in (0, 1)$ .

The above monotonicity yields the following claim about the behavior of  $g_k^r(z(t))$  along the straight-line path  $\rho$ .

- (a) If  $z_k \leq 1/q$ , then  $g_k^r(z(t))$  is monotonically decreasing in  $t$ .
- (b) If  $z_k > 1/q$ , then  $g_k^r(z(t))$  has at most one critical point  $t_k^*$  on  $(0, 1)$ .

The above claim (a) follows immediately from (29) as  $\langle z - z_\beta, g^r(z(t)) \rangle_\rho > 0$  for  $t > 0$ . Claim (b) also follows from (29) as its right-hand side,  $z_k - 1/q > 0$  and  $\langle z - z_\beta, g^r(z(t)) \rangle_\rho$  is increasing. Thus there is at most one point  $t_k^*$  on  $(0, 1)$  such that  $\frac{d}{dt} [g_k^r(z(t))] = 0$ .

Next, define

$$A_z = \{k : z_k > 1/q\}$$

Then the aggregate  $g$ -variation can be split into

$$D_\rho^g(z, z_\beta) = \sum_{k \in A_z} \int_0^1 \left| \frac{d}{dt} [g_k^r(z(t))] \right| dt + \sum_{k \notin A_z} \int_0^1 \left| \frac{d}{dt} [g_k^r(z(t))] \right| dt$$

For  $k \notin A_z$ , claims (a) and (b) imply

$$\int_0^1 \left| \frac{d}{dt} [g_k^r(z(t))] \right| dt = - \int_0^1 \frac{d}{dt} [g_k^r(z(t))] dt = g_k^r(z(0)) - g_k^r(z(1)) = \frac{1}{q} - g_k^r(z)$$

For  $k \in A_z$ , let  $t_k = \max\{t_k^*, 1\}$ , where  $t_k^*$  is defined in (b). Then, we have

$$\int_0^1 \left| \frac{d}{dt} [g_k^r(z(t))] \right| dt = \int_0^{t_k^*} \frac{d}{dt} [g_k^r(z(t))] dt - \int_{t_k^*}^1 \frac{d}{dt} [g_k^r(z(t))] dt = 2g_k^r(z(t_k^*)) - g_k^r(z) - \frac{1}{q}$$

Combining the previous two displays, we get

$$\begin{aligned} D_\rho^g(z, z_\beta) &= \sum_{k \in A} \left( 2g_k^r(z(t_k^*)) - g_k^r(z) - \frac{1}{q} \right) + \sum_{k \notin A} \left( \frac{1}{q} - g_k^r(z) \right) \\ &= 2 \sum_{k \in A} \left( g_k^r(z(t_k^*)) - \frac{1}{q} \right) \end{aligned}$$

Since  $\beta < \beta_s$  and  $k \in A_z$ , we have

$$g_k^r(z(t_k^*)) < z(t_k^*)_k \leq z(1)_k = z_k$$

and we conclude that

$$D_\rho^g(z, z_\beta) < 2 \sum_{k \in A} \left( z_k - \frac{1}{q} \right) = \|z - z_\beta\|_1$$

Thus

$$\frac{d_g(z, z_\beta)}{\|z - z_\beta\|_1} \leq \frac{D_\rho^g(z, z_\beta)}{\|z - z_\beta\|_1} < 1 \quad \text{for all } z \neq z_\beta \text{ in } \mathcal{P}.$$

Next, since we are dealing with the straight line segments  $\rho$ ,

$$\limsup_{z \rightarrow z_\beta} \frac{D_\rho^g(z, z_\beta)}{\|z - z_\beta\|_1} = \limsup_{z \rightarrow z_\beta} \frac{\|g(z) - g(z_\beta)\|_1}{\|z - z_\beta\|_1} < 1$$

by (26), the Mean Value Theorem, and  $H(z) \in \mathcal{C}^3$ . This, in turn, guarantees the continuity required for Condition 8.3:

$$\limsup_{z \rightarrow z_\beta} \frac{d_g(z, z_\beta)}{\|z - z_\beta\|_1} \leq \limsup_{z \rightarrow z_\beta} \frac{D_\rho^g(z, z_\beta)}{\|z - z_\beta\|_1} < 1$$

Thus Condition 8.3 is proved for the CWP model. Moreover this proves that the family of straight line segments  $\rho$  is a neo-geodesic family (see definition following Condition 8.3). Indeed, there is  $\delta \in (0, 1)$  such that

$$\left\{ \rho : z(t) = \frac{1}{q}(1-t) + zt, z \in \mathcal{P} \right\} \quad \text{is a } \mathbf{NG}_\delta \text{ family of smooth curves,}$$

i.e.  $\forall z \neq z_\beta$  in  $\mathcal{P}$ , and corresponding  $\rho : z(t) = \frac{1}{q}(1-t) + zt$ ,

$$\frac{D_\rho^g(z, z_\beta)}{\|z - z_\beta\|_1} \leq 1 - \delta/2$$

Since the family of straight line segments  $\rho$  is a neo-geodesic family  $\mathbf{NG}_\delta$ , the integrals

$$D_\rho^g(x, z) := \sum_{k=1}^q \int_\rho \left| \left\langle \nabla g_k^r(y), dy \right\rangle \right|$$

can be uniformly approximated by the corresponding Riemann sums of small enough step size by the Mean Value Theorem as  $H(z) \in \mathcal{C}^3$  and therefore each  $g_k^r(z) \in \mathcal{C}^2$ . That is, there exists a constant  $C > 0$  that depends on the second partial derivatives of  $g^r(z) = (g_1^r(z), \dots, g_q^r(z))$ , such that for  $\varepsilon > 0$  small enough, the curve  $\rho = \frac{1}{q}(1-t) + zt$  in the family  $\mathbf{NG}_\delta$  that connects  $z_\beta$  to  $z$  satisfies

$$\left| \sum_{k=1}^q \sum_{i=1}^r \left| \left\langle z_i - z_{i-1}, \nabla g_k^r(z_{i-1}) \right\rangle \right| - D_\rho^g(z, z_\beta) \right| < Cr\varepsilon^2 \quad \forall z \in \mathcal{P} \text{ s.t. } \|z - z_\beta\|_1 \geq \varepsilon$$

for a sequence of points  $z_0 = z_\beta, z_1, \dots, z_r = z \in \mathcal{P}$  interpolating  $\rho$  such that

$$\varepsilon \leq \|z_i - z_{i-1}\|_1 < 2\varepsilon \quad \text{for } i = 1, 2, \dots, r.$$

Hence

$$\frac{\sum_{k=1}^q \sum_{i=1}^r \left| \left\langle z_i - z_{i-1}, \nabla g_k^r(z_{i-1}) \right\rangle \right|}{\|z - z_\beta\|_1} \leq 1 - \delta/2 + C\varepsilon \leq 1 - \delta/3$$

for  $\varepsilon \leq \delta/(6C)$ . This concludes the proof of Condition 8.4.  $\square$

Finally, the region of exponentially slow mixing  $\beta > \beta_s(q, r)$  can be shown using the standard approach of bottleneck ratio similar to section 7 in [13].

#### ACKNOWLEDGMENTS

This work was partially supported by a grant from the Simons Foundation (#284262 to Yevgeniy Kovchegov).

#### REFERENCES

1. N. Bhatnagar and D. Randall, Torpid mixing of simulated tempering on the Potts model, *Proceedings of the 15th annual ACM-SIAM symposium on Discrete algorithms*, 478–487 (2004).
2. R. Bubley and M.E. Dyer, Path coupling: a technique for proving rapid mixing in Markov chains, *Proceedings of the 38th IEEE Symposium on Foundations of Computer Science (FOCS)*, 223–231 (1997).
3. M. Costeniuc, R. S. Ellis, and H. Touchette, Complete analysis of phase transitions and ensemble equivalence for the Curie-Weiss-Potts model. *Journal of Mathematical Physics*, Volume 46, 063301 (2005).
4. P. Cuff, J. Ding, O. Louidor, E. Lubetzky, Y. Peres and A. Sly, Glauber dynamics for the mean-field Potts model, *Journal of Statistical Physics*, Volume 149, Number 3, 432–477 (2012).
5. J. Ding, E. Lubetzky, and Y. Peres, The mixing time evolution of Glauber dynamics for the mean-field Ising model, *Comm. Math. Physics*, Volume 289, Number 2, 725–764 (2009).
6. J. Ding, E. Lubetzky, and Y. Peres, Censored Glauber dynamics for the mean-field Ising model, *Journal of Statistical Physics*, Volume 137, Number 1, 161–207 (2009).
7. T. Eisele and R. S. Ellis, Multiple phase transitions in the generalized Curie-Weiss model. *Journal of Statistical Physics*, Volume 52, Number 1/2, 161–207 (1988).
8. R. S. Ellis, K. Haven, and B. Turkington, Large deviation principles and complete equivalence and nonequivalence results for pure and mixed ensembles. *Journal of Statistical Physics*, Volume 101, Numbers 5/6, 999–1064 (2000).
9. M. Ebbers, H. Knöpfel, M. Löwe, and F. Vermet, Mixing times for the swapping algorithm on the Blume-Emery-Griffiths model, *Random Structures and Algorithms*, DOI:10.1002/rsa.20461 (2012).
10. M. Ebbers and M. Löwe, Rapid mixing of the swapping algorithm for the generalized Curie-Weiss model, submitted.
11. R. S. Ellis, P. T. Otto, and H. Touchette, Analysis of phase transitions in the mean-field Blume-Emery-Griffiths model, *Ann. Appl. Prob.* **15** (2005) 2203–2254.
12. B. Jahnke, C. Külske, E. Rudelli, and J. Wegener, Gibbsian and non-Gibbsian properties of the generalized mean-field fuzzy Potts-model, Accepted for publication in *Markov Processes and Related Fields*. Preprint arXiv:1312.5229v1
13. Y. Kovchegov, P. T. Otto, and M. Titus Mixing times for the mean-field Blume-Capel model via aggregate path coupling, *Journal of Statistical Physics*, Volume 144, Numbers 5, 1009–1027 (2011).
14. D. A. Levin, M. Luczak, and Y. Peres Glauber dynamics of the mean-field Ising model: cut-off, critical power law, and metastability, *Probability Theory and Related Fields*, Volume 146, Numbers 1, 223–265 (2010).
15. D. Levin, Y. Peres, E. Wilmer, *Markov Chains and Mixing Times*. American Mathematical Society, USA (2009)
16. T. Lindvall, *Lectures on the Coupling Method*, John Wiley & Sons, New York, 1992. Reprint: Dover paperback edition, 2002.
17. M. J. Luczak, Concentration of measure and mixing times of Markov chains. *Discrete Mathematics and Theor. Comp. Sci.* (Proceedings of the 5th Colloq. Mathem. Comp. Sci. ) 95–120 (2008).



18. A. Galanis, D. Stefankovic, E. Vigoda, Swendsen-Wang Algorithm on the Mean-Field Potts Model, Preprint arXiv:1502.06593v1
19. F. Y. Wu, The Potts model. *Rev. Mod. Phys.* 54:235-268 (1982).

DEPARTMENT OF MATHEMATICS, OREGON STATE UNIVERSITY, CORVALLIS, OR 97331  
*E-mail address:* kovchegy@math.oregonstate.edu

DEPARTMENT OF MATHEMATICS, WILLAMETTE UNIVERSITY, SALEM, OR 97302, PHONE: 1-503-370-6487  
*E-mail address:* potto@willamette.edu