Microsoft Technical Report
MSR-TR-99-40

# Torpid Mixing of Some Monte Carlo Markov Chain Algorithms in Statistical Physics

Christian Borgs[1], Jennifer T. Chayes[1,i],
Alan Frieze[2,ii], Jeong Han Kim[1], Prasad Tetali[3,iii],
Eric Vigoda[4,iv] and Van Ha Vu[5]

[1] *Microsoft Research, One Microsoft Way, Redmond, WA 98052*
[2] *Department of Mathematics, Carnegie Mellon, Pittsburgh, PA 15213*
[3] *School of Mathematics, Georgia Tech, Atlanta, GA 30332-0160*
[4] *Department of Computer Science, Berkeley, CA 94720*
[5] *Institute for Advanced Study, Princeton, NJ 08540*

April 23, 1999
Revised: June 3, 1999

**Abstract**

We study two widely used algorithms, Glauber dynamics and the Swendsen-Wang algorithm, on rectangular subsets of the hypercubic lattice $\boldsymbol{Z}^d$. We prove that under certain circumstances, the mixing time in a box of side length $L$ with periodic boundary conditions can be exponential in $L^{d-1}$. In other words, under these circumstances, the mixing in these widely used algorithms is not rapid; instead, it is *torpid*. The models we study are the independent set model and the $q$-state Potts model. For both models, we prove that Glauber dynamics is torpid in the region with phase coexistence. For the Potts model, we prove that Swendsen-Wang is torpid at the phase transition point.

[i] On leave from University of California, Los Angeles
[ii] Supported in part by the NSF Grant No. CCR–9530974, and by a Guggenheim fellowship
[iii] Supported in part by the NSF Grant No. DMS–9800351
[iv] Supported in part by a GAANN graduate fellowship and NSF grant CCR–9505448

# 1 Introduction

Monte Carlo Markov chains (MCMC) are used in computer science to design algorithms for estimating the size of large combinatorially defined structures. In statistical physics, they are used to study the behavior of idealized models of physical systems in equilibrium. In the latter case, the models of interest are usually defined on regular, finite-dimensional structures such as the hypercubic lattice $\boldsymbol{Z}^d$. In both applications, it is necessary to run the chain, $\mathcal{M}$, until it is close enough to its steady state. Thus it is important to design *rapidly mixing* algorithms, i.e. algorithms for which the *mixing time*, $\tau_{\mathcal{M}}$, is small.

In this paper, we study two statistical physics models, the $q$-state Potts model and the independent set problem. We consider these models on the graphs on which they are most often studied in physical applications, namely on subsets of $\boldsymbol{Z}^d$. For the Potts model, we study two types of Monte Carlo Markov chains – Glauber dynamics, and the empirically more rapid Swendsen-Wang dynamics.

Both the Potts model and the independent set model are characterized by non-negative parameters, the former by a so-called inverse temperature $\beta$, and the latter by a so-called fugacity $\lambda$ (see definitions below). Both models are known to undergo phase transitions from a so-called disordered phase with a unique equilibrium state to an ordered phase with multiple equilibrium states. Due to the multiple states, the ordered phase is also known as the region of phase coexistence.

The point of this work is to relate the mixing times of the MCMCs to the phase structures of the underlying equilibrium models. In particular, we show that Glauber dynamics is slow, or *torpid*, for both models in their regions of phase coexistence, while Swendsen-Wang for the Potts model is torpid at the phase transition point. This latter result has apparently come as a surprise to some physicists who use the Swendsen-Wang algorithm to simulate the Potts model, and who have tacitly assumed that it mixes rapidly for all values of the inverse temperature.

In addition to this "physically surprising" result, our work is new in a number of respects. While there has recently been a good deal of work in the theoretical CS community on slowness of Swendsen-Wang dynamics for the Potts model (see citations below), this is the one of the first works to consider the physically relevant case of the hypercubic lattice $\Gamma_d = \boldsymbol{Z}^d$ and finite portions thereof. (In $\Gamma_d$, two lattice points are joined by an edge if they differ by 1 in one coordinate.) From a technical point of view, the hypercubic lattice is much more challenging than the complete graph. However, these technical difficulties give us the opportunity to use some beautiful and non-trivial results. In particular, our work brings to bear, and to some extent extends, very sophisticated statistical physics expansion techniques for the problem of controlling the number of cutsets in graphical expansions of these models. Specifically, we use so-called Pirogov-Sinai theory [22] from the statistical physics literature, in the form adapted to the Potts model by Borgs, Kotecký and Miracle-Sole ([5], [6]). We also use the new and powerful combinatoric estimates of Lebowitz and Mazel [18] for controlling the number of cutsets. Finally, we use the lovely isoperimetric inequalities of Bollobás and Leader [2].

In this introduction, we will first describe our work on MCMC for the Potts model, and then for the independent set problem. In both cases, we will state basic versions of the results, and then discuss generalizations to be derived in a more detailed version of this work [3].

The $q$-state Potts Model (see [27], [28]) on an arbitrary graph $G = (V, E)$, $|V| = n$ is defined as follows: a coloring $\boldsymbol{\sigma}$ is a map from $V \to [q] = \{1, 2, \ldots, q\}$. Let $D(\boldsymbol{\sigma})$ be the set of edges with endpoints of a different color and let $d(\boldsymbol{\sigma}) = |D(\boldsymbol{\sigma})|$. The weight of a coloring $w(\boldsymbol{\sigma})$ is $e^{-\beta d(\boldsymbol{\sigma})}$. We turn this into a probability distribution $\mu$ by normalizing with the *partition function* $Z = \sum_{\boldsymbol{\sigma}} w(\boldsymbol{\sigma})$. To study this model empirically, one needs to be able to generate $\boldsymbol{\sigma}$ with probability (close to)

$$\mu(\boldsymbol{\sigma}) = \frac{w(\boldsymbol{\sigma})}{Z}. \tag{1}$$

The model is said to be *ferromagnetic* if $\beta \geq 0$, otherwise it is *anti-ferromagnetic*. Note that $\beta = -\infty$ corresponds to random *proper* colorings.

The widely used Swendsen-Wang algorithm [25] for the ferromagnetic model uses a Markov chain with state space $[q]^V$ which has steady state $\mu$ – see Section 3. Gore and Jerrum [14] proved that on the complete graph $K_n$ with $q \geq 3$, there is a certain value of $\beta$ (*inverse temperature*) such that the mixing time of the algorithm is exponential in $n$. Jerrum [16] has coined the phrase *torpid* mixing to describe this phenomenon. Cooper and Frieze [8] extended their arguments to show that in the Potts model on the random graph $G_{n,p}$, this phenomenon persists with high probability for $p = \Omega(n^{-1/3})$. Li and Sokal [19] proved a linear (in the number of *sites*) lower bound for finite boxes in $\mathbf{Z}^d$. (For positive results on this algorithm see [8], Cooper, Dyer, Frieze and Rue [7], Huber [15], Martinelli [21].)

Our first result concerns this algorithm and the simpler Glauber dynamics – see Section 3. an Let $T = T_{L,d} = (\boldsymbol{Z}/L\boldsymbol{Z})^d$ be the $d$-dimensional torus of side $L$. We view this as a graph where two points are connected by an edge if they differ by 1 (mod $L$) in one component. It has vertex set $V = V_{L,d}$ and edge set $E = E_{L,d}$. Using the results of Borgs, Kotecký and Miracle-Solé [6], we prove the following negative result:

identified.

**Theorem 1** *For $d \geq 2$ and sufficiently large $q$, there exists $\beta_c = \beta_c(q, d)$ such that:*
**(a)** *The mixing time $\tau_{SW}$ of the Swendsen-Wang algorithm on $T_{L,d}$ at $\beta_c$ satisfies*

$$\tau_{SW} \geq e^{K_1 L/(\log L)^2}$$

*for some absolute constant $K_1 > 0$.*
**(b)** *The mixing time $\tau_{GD}$ of the Glauber dynamics for $\beta \geq \beta_c$ satisfies*

$$\tau_{GD} \geq e^{K_2 L/(\log L)^2}$$

*for some absolute constant $K_2 > 0$.*

It turns out that we can strengthen the bounds of Theorem 1 so that the lower bound is of the form $e^{KL^{d-1}/(\log L)^2}$. In order to do this, we have to refine the geometric analysis of Section 4 to improve the finite-size scaling bounds of [6]. This requires much more involved analysis, which will be carried out in detail in [3].

For an arbitrary graph $G = (V, E)$, an independent set is a set of vertices $I \subset V$ such that no pair of vertices $i, j \in I$ is incident to the same edge $e \in E$. Dyer, Frieze and Jerrum [10] considered the problem of generating a nearly random independent set of a bipartite graph. They prove that Glauber dynamics exhibits torpid mixing on almost all regular graphs of degree 6 or more and that the problem is NP-hard for regular graphs of degree 25 or more. In statistical physics, the independent set problem is called the *hard-core gas* model. In general there is a parameter $\lambda > 0$ called the *fugacity* and one wants to generate independent sets $I$ with probability proportional to $\lambda^{|I|}$ i.e.

$$\mu(I) = \frac{\lambda^{|I|}}{\sum_{J \text{ independent}} \lambda^{|J|}}. \tag{2}$$

Our second result concerns this problem. The Glauber dynamics chain is a simple chain on the independent sets of graph $G$ that selects a random vertex and adds/deletes it to/from the current independent set with some probability dependent on $\lambda$ – see Section 3. Dyer and Greenhill [11], Luby and Vigoda [20] have proved that this chain is rapidly mixing for $\lambda < \frac{2}{\Delta - 2}$, where $\Delta$ denotes the maximum degree of $G$.

We also prove bounds on more general Markov chains. To define this class, let $I, I'$ be independent sets, and let $D(I, I') = |I \setminus I'| + |I' \setminus I|$. For an ergodic Markov chain $\mathcal{M}_L$ on $T_{L,d}$, let $D_{\mathcal{M}_L}$ be the maximum of $D(I, I')$ over all $I$ and $I'$ for which the transition probability is non-zero. We say that $\mathcal{M}_L$ is local if $D_{\mathcal{M}_L}$ is bounded uniformly in $L$, and we say that it is $\rho$-quasi-local if $D_{\mathcal{M}_L} \leq \rho L^d$ for some $\rho < 1$ which is independent of $L$.

**Theorem 2** *For $d \geq 2$ and $\lambda$ sufficiently large, the mixing time $\tau_{GD}$ of the Glauber chain on $T_{L,d}$ satisfies*

$$\tau_{GD} \geq e^{K_3 L^{d-1}/(\log L)^2}$$

*for some constant $K_3 > 0$ depending only on the dimension $d$. More generally, let $\tau_L$ be the mixing time for any ergodic Markov chain on $T_{L,d}$ which is $\rho$-quasi-local for some $\rho < 1$. Then*

$$\tau_L \geq e^{K_4 L^{d-1}/(\log L)^2}$$

*for some constant $K_4$ depending only on $d$ and $\rho$.*

As stated above, Theorem 2 is also not the optimal theorem we can prove. Combining the methods developed in Sections 4 and 5 of this paper with those of [26], we can improve the bound $e^{K_3 L^{d-1}/(\log L)^2}$ on the mixing time of the Glauber dynamics for the independent set problem to a bound of the form $e^{K_3 L^{d-1}}$, see [3] for details. However, at this time, we can strengthen the bound only for Glauber dynamics, not for more general quasi-local Markov chains.

Finally, we want to point out that our techniques for proving slow mixing are quite robust, and can be applied to many models exhibiting the phenomenon of phase coexistence. All that we require is that the equilibrium model have energy barriers between different phases that is high enough to apply the techniques of [4], and that the dynamics is not sufficiently global to permit jumps over these barriers. (An example of a Markov chain which "jumps over" energy barriers is the Swendsen-Wang algorithm at temperatures below the transition temperature.) See [3] for the description of permissible algorithms, and a proof of slow mixing for these more general models.

## 2 Mixing Time

Let $\mathcal{M}$ be an ergodic Markov chain on a finite state space $\Omega$, with transition probabilities $P(x, y),\ x, y \in \Omega$. Let $\pi$ denote the stationary distribution of $\mathcal{M}$.

Let $x \in \Omega$ be an arbitrary fixed state, and denote by $P_{t,x}(\omega)$ the probability that the system is in state $\omega$ at time $t$ given that $x$ is the initial state.

The *variation distance* $\Delta(\pi_1, \pi_2)$ between two distributions $\pi_1, \pi_2$ on $\Omega$ is defined by

$$\Delta(\pi_1, \pi_2) = \max_{S \subseteq \Omega} |\pi_1(S) - \pi_2(S)| = \tfrac{1}{2} \sum_{\omega \in \Omega} |\pi_1(\omega) - \pi_2(\omega)|.$$

The variation distance at time $t$ with respect to the initial state $x$ is then defined as

$$\Delta_x(t) = \Delta(P_{t,x}, \pi).$$

We define the function $d(t) = \max_{x \in \Omega} \Delta_x(t)$ and the *mixing time*

$$\tau = \min\{t : \ 2d(t) \leq e^{-1}\}.$$

A property of $d(t)$ given in [1] is that $d(s + t) \leq 2d(s)d(t)$, implying in particular that $d(t) \leq \exp(-\lfloor t/\tau \rfloor)$. It is therefore both necessary and sufficient that chains be run for some multiple of mixing time in order to get a sample which is close to a sample from the steady state.

For our purposes, the Swendsen-Wang algorithm is *rapidly mixing* if its mixing time $\tau_{SW}$ is bounded by a polynomial in $n$, the number of vertices of $G$. Similarly for the Glauber chain.

Jerrum and Sinclair [24] introduced the notion of conductance to the study of finite *time reversible* Markov chains. A chain is reversible if it satisfies the *detailed balance* equations:

$$\pi(x)P(x, y) = \pi(y)P(y, x), \qquad \text{for all } x, y \in \Omega.$$

Putting $Q(x, y) = \pi(x)P(x, y)$ and $Q(A, B) = \sum_{(x,y) \in A \times B} Q(x, y)$, we define the conductance of a set of states $\emptyset \neq S \subset \Omega$ as

$$\Phi_S = \frac{Q(S, \bar{S})}{\pi(S)\pi(\bar{S})} \qquad \text{where } \bar{S} = \Omega \setminus S. \tag{3}$$

The conductance $\Phi_{\mathcal{M}}$ of the chain itself is simply $\min_{S \neq \emptyset} \Phi_S$. We prove our lower bounds on mixing time by showing that $\Phi_{\mathcal{M}}$ is small and then using the well-known bound [1]

$$e^{-1/\tau_{\mathcal{M}}} \geq 1 - \Phi_{\mathcal{M}}. \tag{4}$$

## 3 MCMC Algorithms

There are several MCMC algorithms that are used to generate a random sample from these distributions corresponding to the hard-core model and ferromagnetic Potts model. The Glauber dynamics is perhaps the simplest such Markov chain. Its transitions are as follows: Choose a vertex at random, and modify the spin of that vertex by choosing from the distribution conditional on the spins of the other vertices remaining the same. We will detail the algorithm for the hard-core model on independent sets.

**Glauber Dynamics:** From an independent set $\sigma$,

**G1** Choose $v$ uniformly at random from $V$.

**G2** Let

$$\sigma' = \begin{cases} \sigma \cup \{v\} & \text{with probability } \lambda/(1+\lambda) \\ \sigma \setminus \{v\} & \text{with probability } 1/(1+\lambda). \end{cases}$$

**G3** If $\sigma'$ is an independent set, then move to $\sigma'$, otherwise stay at the current independent set $\sigma$.

For the ferromagnetic Potts model, an alternative method, the Swendsen-Wang process [25], is often preferred over other dynamics in Markov chain Monte Carlo simulations.
**Swendsen-Wang Algorithm:**

**SW1** Let $B = E \setminus D(\boldsymbol{\sigma})$ be the set of edges joining vertices of the same color. Delete each edge of $B$ independently with probability $1-p$, where $p = 1 - e^{-\beta}$. This gives a subset $A$ of $B$.

**SW2** The graph $(V, A)$ consists of connected components. For each component a colour (spin) is chosen uniformly at random from $[q]$ and all vertices within the component are assigned that colour (spin).

The Swendsen-Wang algorithm was motivated by the equivalence of the ferromagnetic $q$-state Potts model and the *random cluster* model of Fortuin and Kasteleyn [13], which we now describe.

Given a graph $G = (V, E)$, let $G(A) = (V, A)$ denote the subgraph of $G$ induced by the edge set $A \subseteq E$. In the random cluster model, $G(A)$ is the measure given by

$$\mu(G(A)) = \frac{1}{Z} p^{|A|}(1-p)^{|E|-|A|} q^{c(A)}, \tag{5}$$

where $c(A)$ is the number of components of $G(A)$ and $p$ is a probability.

The relationship between the two models is elucidated in a paper by Edwards and Sokal [12]. The Potts and random cluster models are defined on a joint probability space $[q]^n \times 2^E$. The joint probability $\pi(\boldsymbol{\sigma}, A)$ is defined by

$$\pi(\boldsymbol{\sigma}, A) = \frac{1}{Z} \prod_{(i,j)\in E} ((1-p)\delta_{(i,j)\notin A} + p\delta_{(i,j)\in A}\delta_{\sigma_i=\sigma_j})), \tag{6}$$

where $Z$ is a normalizing constant. By summing over $\boldsymbol{\sigma}$ or $A$ we see that the marginal distributions are correct, and (remarkably) the normalising constants in both Potts and Cluster models are the value of $Z$ given in the expression above.

The Swendsen-Wang algorithm can be seen as given $\boldsymbol{\sigma}$, (i) choose a random $A'$ according to $\pi(\boldsymbol{\sigma}, A')$ and then (ii) choose a random $\boldsymbol{\sigma'}$ according to $\pi(\boldsymbol{\sigma'}, A')$.

After Step SW1 we say that we are in the FK representation of the chain.

# 4 Minimal Cutsets

Let $G = (V, E)$ be a connected graph. For $W \subset V$ we define $G_W$ as the graph $(W, E_W)$, where $E_W$ is the set of all edges in $E$ that join two vertices in $W$. We say that $C \subset W$ is a component of $W$ if $C$ is the vertex set of a component of $G_W$. As usual, we define a subset $\gamma \subset E$ to be a cutset if $(V, E \setminus \gamma)$ is disconnected. We define $\gamma$ to be a minimal cutset if all cutsets contained in $\gamma$ are identical to $\gamma$. If $\gamma$ is minimal, $(V, E \setminus \gamma)$ has exactly two components. For $W \subset V$, we let $\overline{W}$ denote the complement of $W$, i.e. $\overline{W} = V \setminus W$. We denote the set of edges between two disjoint sets of vertices $W$ and $W'$ by $(W : W')$. Finally, we use $\mathcal{C}(W)$ to denote the set of components of $W$.

We consider the cutset $\partial W = (W : \overline{W})$ and decompose it as $\partial W = \cup_{C \in \mathcal{C}(W)} \partial C$. We will further decompose $\partial C$ into minimal cutsets, see Lemma 1 below. In order to state the lemma, we introduce the sets

$$\Gamma_C = \{(C : D) | D \in \mathcal{C}(\overline{C})\} = \{(\overline{D} : D) | D \in \mathcal{C}(\overline{C})\}$$

and

$$\Gamma(W) = \bigcup_{C \in \mathcal{C}(W)} \Gamma_C.$$

**Lemma 1** *Consider $W \subset V$.*
**(a)** *Let $C, C'$ be different components of $W$. There exist unique $D \in \mathcal{C}(\overline{C})$ and $D' \in \mathcal{C}(\overline{C'})$ such that $\overline{D} \subseteq D'$ or equivalently $\overline{D'} \subseteq D$.*
**(b)** *For $C \in \mathcal{C}(W)$, $\partial C$ has a unique decomposition into minimal cutsets as $\partial C = \cup_{\gamma \in \Gamma_C} \gamma$.*
**(c)** *If $\gamma$, $\gamma' \in \Gamma(W)$ are distinct then they are disjoint.*
**(d)** *Let $C$ and $C'$ be two (not necessarily distinct) components of $W \subset V$. If $X$ or $\overline{X}$ is a component of $\overline{C}$ and $Y$ or $\overline{Y}$ is a component of $\overline{C'}$ then*

$$X \cap Y = \emptyset, X \cap \overline{Y} = \emptyset, \overline{X} \cap Y = \emptyset, \text{ or } \overline{X} \cap \overline{Y} = \emptyset.$$

**Proof in appendix**

Let $\gamma = (D : \overline{D})$ be a minimal cutset of $G$, in particular $D$ and $\overline{D}$ are connected. We then define $\operatorname{Int}\gamma$ as the smaller (in terms of cardinality) of $D$ and $\overline{D}$. If $D$ and $\overline{D}$ have the same size, we can define $\operatorname{Int}\gamma$ as either $D$ or $\overline{D}$. For definiteness, we define $\operatorname{Int}\gamma$ as the one containing a fixed point $x_o \in V$. For a cutset $\gamma$ we define $\operatorname{Ext}\gamma = V \setminus \operatorname{Int}\gamma$, and for a collection $\Gamma$ of minimal cutsets, we define the *interior of* $\Gamma$ and the *common exterior of* $\Gamma$ as as

$$\operatorname{Int}\Gamma = \bigcup_{\gamma \in \Gamma} \operatorname{Int}\Gamma \qquad \text{and} \qquad \operatorname{Ext}\Gamma = \bigcap_{\gamma \in \Gamma} \operatorname{Ext}\gamma.$$

Note that $\operatorname{Int}\Gamma \cup \operatorname{Ext}\Gamma = V$ for all sets $\Gamma$ of minimal cutsets.

**Lemma 2** *Let $W \subset V$.*
**(a)** *Let $\gamma, \gamma' \in \Gamma(W)$. If $\operatorname{Int}\gamma \cap \operatorname{Int}\gamma' \neq \emptyset$ , then either $\operatorname{Int}\gamma \subset \operatorname{Int}\gamma'$ or $\operatorname{Int}\gamma' \subset \operatorname{Int}\gamma$.*
**(b)** *Either $W$ or $\overline{W}$ is a subset of $\operatorname{Int}\Gamma(W)$.*

**Proof in appendix**

Next we specialize to the torus $T_{L,d} = (V_{L,d}, E_{L,d})$. Consider a set $W \subset V_{L,d}$ and a fixed minimal cutset $\gamma$ corresponding to $W$. For $e \in \gamma$ we define a *dual* $(d-1)$-dimensional cube $e^*$ which is (i) orthogonal to $e$ and (ii) bisects $e$, when $T_{L,d}$ is considered as immersed in the continuum torus $(\boldsymbol{R}/\boldsymbol{Z})^d$. (In dimension $d = 3$, the two-dimensional dual cells are referred to as *plaquettes*). We define a graph $\Gamma^* = (\gamma^*, E^*)$ where $\gamma^* = \{e^* : \ e \in \gamma\}$ and $(e_1^*, e_2^*) \in E^*$ iff $e_1^* \cap e_2^*$ is a cube of dimension $d-2$. The components of $\Gamma^*$ are called the *co-components* of $\gamma$. These co-components are connected hypersurfaces of dual $(d-1)$-dimensional cells.

In the following, we will call cutsets with one co-component topologically trivial, and cutsets with more than one co-component topologically non-trivial. Small components which can be embedded in $Z^d$ give rise to cutsets with only a single co-component, which are therefore topologically trivial. Topologically non-trivial cutsets arise from certain components which are large enough to "feel" the non-trivial topology of the torus. For example, the component $C = \{x \in V_{L,d} \mid 1 \le x_1 \le L/2\}$ gives rise to a cutset whose two co-connected components are two parallel interfaces, each of which has size $L^{d-1}$.

**Lemma 3** **(a)** *Given a fixed edge $e \in E_{L,d}$ there are at most $\nu^k$, $\nu = \min\{3, d^{64/d}\}$, distinct co-components $\gamma$ of size $k$ with $e \in \gamma$.*
**(b)** *If a cutset is non-trivial, each of its co-components contains at least $L^{d-1}$ edges.*

**Proof in appendix**

# 5 Independent Sets

In this section, we give a proof of Theorem 2. We start with some notation. For a bipartite graph $G = (V, E)$ we arbitrarily call the vertices in one partition even, and those in the other partition odd. We write $V_{even}$ for the set of even vertices in $V$, and $V_{odd}$ for the set of odd vertices in $V$. We denote the collection of independent sets of $G$ by $\Omega$. Let $I$ be an independent set in $\Omega$. We then define $W_{odd}(I)$ as the set of vertices in and adjacent to a vertex in the set $I \cap V_{odd}$. Similarly $W_{even}(I)$ is defined for $I \cap V_{even}$. We define the set $\Gamma_{odd}(I)$ as the set of minimal cutsets corresponding to $W_{odd}(I)$, $\Gamma_{odd}(I) = \Gamma(W_{odd}(I))$, and similarly for the set $\Gamma_{even}(I)$. Finally, for a cutset $\gamma$, we define $V(\gamma) = \bigcup_{\{x,y\}\in\gamma}\{x, y\}$.

**Lemma 4** **(a)** *If $\gamma \in \Gamma_{odd}(I)$, then $V(\gamma) \cap I = \emptyset$.*
**(b)** *For $\gamma \in \Gamma_{odd}(I)$, the vertices in the set $V(\gamma) \cap \operatorname{Int}\gamma$ are either all even or all odd.*
**(c)** *For $\gamma \in \Gamma_{odd}(I)$, there exists an independent set $I_\gamma$ such that $\Gamma_{odd}(I_\gamma) = \{\gamma\}$.*
**(d)** *Either $I \cap V_{odd}$ or $I \cap V_{even}$ is a subset of* $\operatorname{Int}\Gamma_{odd}(I)$.

**Proof:**

**(a)** We have to prove that $\{x, y\} \cap I = \emptyset$ whenever $\{x, y\} \in \gamma \subset \partial W_{odd}(I)$. First notice that for an odd vertex $v$, $v \in W_{odd}(I) \Leftrightarrow v \in I$, whereas if $v$ is even then $v \in W_{odd}(I) \Leftrightarrow v$ has a neighbor $w \in I$. Suppose that $x \in I, y \notin I$. If $x$ is odd then $x, y \in W_{odd}(I)$. If $x$ is even, then $x, y \notin W_{odd}(I)$. In either case, we have the contradiction that $\{x, y\} \notin \partial W_{odd}(I)$.

**(b)** If $\gamma \in \Gamma_{odd}(I)$,then $\gamma = (\overline{D} : D) = (C : D)$ for some componenent $C$ of $W_{odd}(I)$ and some component $D$ of $\overline{C}$. As a consequence, either $(V(\gamma) \cap \operatorname{Int}\gamma) \subseteq W_{odd}(I)$, or $(V(\gamma) \cap \operatorname{Int}\gamma) \subseteq \overline{W_{odd}(I)}$. If an odd vertex $v$ is in the set $W_{odd}(I)$ then $v \in I$ and $w \in W_{odd}(I)$ for all neighbors $w$ of $v$. Thus an odd vertex $v \in W_{odd}(I)$ cannot be incident to an edge in $\partial W_{odd}(I)$. As a consequence, the vertices of $V(\gamma) \cap \operatorname{Int}\gamma$ are is even if $(V(\gamma) \cap \operatorname{Int}\gamma) \subseteq W_{odd}(I)$ and odd otherwise.

**(c)** If the vertices of the set $V(\gamma)\cap\operatorname{Int}\gamma$ are even then let $I_\gamma = (V_{odd}\cap\operatorname{Int}\gamma)\cup(V_{even}\cap\overline{\operatorname{Int}\gamma})$. Otherwise, exchange the sets $V_{odd}$ and $V_{even}$ in the definition of $I_\gamma$.

**(d)** Lemma 2 implies that either

$$W_{odd}(I) \subset \operatorname{Int}\Gamma_{odd}(I) \quad \text{or} \quad \overline{W_{odd}(I)} \subset \operatorname{Int}\Gamma_{odd}(I).$$

Since $I \cap V_{odd} \subset W_{odd}(I)$ and $I \cap V_{even} \subset \overline{W_{odd}(I)}$, the result follows. □

From now on, we specialize to the graph $T_{L,d}$. For a vertex $v = (v_1, \dots, v_d) \in V$ and a "direction" $\alpha \in \{\pm 1, \dots, \pm d\}$, we define the *shift* $\sigma_\alpha(v)$ as the vertex with coordinates $v_i$ for $i \neq |\alpha|$ and $v_i + \operatorname{sign}(\alpha) \pmod L$ for $i = |\alpha|$, where $\operatorname{sign}(\alpha) = \alpha/|\alpha|$. For a cutset $\gamma \in \Gamma_{odd}(I)$, we define $\gamma_\alpha = \{(v,w) | (v,w) \in \gamma, v \in \operatorname{Int}\gamma, w = \sigma_\alpha(v)\}$.

**Lemma 5** *For any cutset $\gamma \in \Gamma_{odd}(I)$ and any direction $\alpha$, $|\gamma_\alpha| = |\gamma|/2d$.*

**Proof:** We first prove the lemma for $d = 2$. Let $\gamma^*$ be the set of edges dual to the edges in $\gamma$. The set $\gamma^*$ is a union of cycles, and each edge in the $+1$ or $-1$ direction in any of these loops is followed by an edge in the $+2$ or $-2$ direction by Lemma 4 (b). We therefore have that $|\gamma_i| + |\gamma_{-i}|$ is independent of the direction $i$. Since $\gamma$ is a cutset, $|\gamma_i|$ must be equal to $|\gamma_{-i}|$, which implies the claim. For $d > 2$, we consider the intersection of $\operatorname{Int}\gamma$ with a two-dimensional plane $S(\{k_i\}) = \{x \in T \mid x_i = k_i, i \notin \{1,2\}\}$. Since also the points in $(V(\gamma) \cap \operatorname{Int}\gamma) \cap S(\{k_i\})$ are all even or all odd, the above arguments can be applied to the intersection of $\gamma$ and $S(\{k_i\})$, implying that $|\gamma_1| = |\gamma_{-1}| = |\gamma_2| = |\gamma_{-2}|$ since it is true for the intersection of these sets with any of the hyperplanes $S(\{k_i\})$. Applying this argument for an arbitrary pair of directions, we get the lemma. □

The next lemma is a generalization of a lemma first proved by Dobrushin in [9].

**Lemma 6** *Let $\Gamma$ be a set of minimal cutsets, and let $\Omega_\Gamma = \{I : \Gamma \subset \Gamma_{odd}(I)\}$. Then*

$$\mu(\Omega_\Gamma) \leq \lambda^{-\sum_{\gamma\in\Gamma}(|\gamma|/2d)} \tag{7}$$

**Proof:** We first note that it is enough to prove there exists an injective map $\phi_\Gamma : \Omega_\Gamma \to \Omega$ such that

$$\mu(I) = \lambda^{-\sum |\gamma_i|/2d} \mu(\phi_\Gamma(I)).$$

Indeed, given such a map, we have

$$\mu(\Omega_\Gamma) = \lambda^{-\sum |\gamma_i|/2d} \mu(\phi_\Gamma(\Omega_\Gamma)) \leq \lambda^{-\sum |\gamma_i|/2d}.$$

In order to construct such a map $\phi_\Gamma$, we introduce the partial order $\gamma \le \gamma' \Leftrightarrow \operatorname{Int}\gamma \subset \operatorname{Int}\gamma'$. We then observe that, by induction, it is enough to prove that for any $\Gamma$ and any $\gamma \in \Gamma$ such that $\gamma$ is minimal in $\Gamma$ with respect to the partial order, we have an injective map $\phi_\gamma : \Omega_\Gamma \to \Omega_{\Gamma\setminus\{\gamma\}}$ such that $\mu(I) = \lambda^{-|\gamma|/2d}\mu(\phi_\gamma(I))$.

We will now construct such a map. Consider $I \in \Omega_\Gamma$. Let $\sigma = \sigma_\alpha$. The proof holds for any choice of $\alpha$. Defining

$$\phi_\gamma(I) = (I \cap \overline{\operatorname{Int}\gamma}) \cup \sigma(I \cap \operatorname{Int}\gamma) \cup (\operatorname{Int}\gamma \setminus \sigma(\operatorname{Int}\gamma)),$$

we will have to show that $\phi_\gamma$ is an injection, that $I' = \phi_\gamma(I)$ is an independent setwith $\mu(I') = \mu(I)\lambda^{|\gamma|/2d}$ and that $I' \in \Omega_{\Gamma\setminus\{\gamma\}}$.

The first statement is obvious from the fact that the three sets $I_1 = I \cap \overline{\operatorname{Int}\gamma}$, $I_2 = \sigma(I \cap \operatorname{Int}\gamma)$ and $I_3 = \operatorname{Int}\gamma \setminus \sigma(\operatorname{Int}\gamma)$ are pairwise disjoint (use Lemma 4 (a) to see that $I_1$ and $I_2$ are disjoint).

$I_1, I_2$ are obviously independent and the independence of $I_3$ follows from $I_3 \subseteq V(\gamma)$ and Lemma 4(b). To then prove that $I'$ is an independent set, we use that, again by Lemma 4 (a), the sets $I_1 \cup I_2$ and $I_1 \cup I_3$ are independent sets. It remains to show that $I_2 \cup I_3$ is also an independent set. Consider $v \in \operatorname{Int}\gamma \setminus \sigma(\operatorname{Int}\gamma)$ and $w \in \sigma(I \cap \operatorname{Int}\gamma)$. Then $v \notin \sigma(\operatorname{Int}\gamma)$ and hence $\sigma_{-\alpha}(v) \notin \operatorname{Int}\gamma$. On the other hand, $\sigma_{-\alpha}(w) \in I \cap \operatorname{Int}\gamma$. Therefore, $\sigma_{-\alpha}(v)$ and $\sigma_{-\alpha}(w)$ cannot be adjacent by Lemma 4 (a), which implies that $v$ and $w$ cannot be adjacent.

To prove $\mu(I') = \mu(I)\lambda^{|\gamma|/2d}$, we notice that $|(I \cap \overline{\operatorname{Int}\gamma}) \cup \sigma(I \cap \operatorname{Int}\gamma)| = |I|$. Thus $\phi_\gamma$ has increased the size of the independent set by exactly $|\operatorname{Int}\gamma \setminus \sigma(\operatorname{Int}\gamma)|$ which is $|\gamma_{-\alpha}| = |\gamma|/2d$ by Lemma 5.

To see that $I' \in \Omega_{\Gamma\setminus\{\gamma\}}$ note that $W_{odd}(I') = W_{odd}(I) \setminus \operatorname{Int}\gamma$. There are two possibilities for $\gamma' \in \Gamma \setminus \{\gamma\}$: $\operatorname{Int}\gamma \cap \operatorname{Int}\gamma' = \emptyset$ implying that $\operatorname{dist}(\operatorname{Int}\gamma, \operatorname{Int}\gamma') \ge 2$ and $I \cap \operatorname{Int}\gamma' = I' \cap \operatorname{Int}\gamma'$. Otherwise, $\operatorname{Int}\gamma \subseteq \operatorname{Int}\gamma'$ implying $\operatorname{dist}(\operatorname{Int}\gamma, \operatorname{Ext}\gamma') \ge 2$ and $I \cap \operatorname{Ext}\gamma' = I' \cap \operatorname{Ext}\gamma'$. $\square$

**Lemma 7** *Let $\Omega(k_1, \dots, k_t)$ be the set of independent sets $I \in \Omega$ which contain a set of odd trivial cutsets of sizes $k_1, \dots, k_t$. Then for $a = \min_{1\le i\le t} k_i$, $b = \max_{1\le i\le t} k_i$ and $k = \sum_{i=1}^t k_i$, we have*

$$\mu(\Omega(k_1, \dots, k_t)) \le (e(b-a+1)L^d/t)^t(\nu\lambda^{-1/(2d)})^k.$$

*Let $\Omega_{n.t.}$ be the set of $I \in \Omega$ such that $\Gamma_{odd}(I)$ contains at least one non-trivial cutset. Then*

$$\mu(\Omega_{n.t.}) \le \left(L^d \frac{(\nu\lambda^{-1/2d})^{L^{d-1}}}{1-\nu\lambda^{-1/2d}}\right)^2 \exp\left(L^d \frac{(\nu\lambda^{-1/2d})^{L^{d-1}}}{1-\nu\lambda^{-1/2d}}\right).$$

**Proof:** To generate $\{\gamma_1, ..., \gamma_t\}$ with $|\gamma_i| = k_i$, we first choose edges $e_i$ in a certain fixed direction, e.g. direction 1, and then cutsets $\gamma_i \ni e_i$. (Every cut set contains an edge in direction 1 – see Lemma 5) In this way, each $\{\gamma_1, ..., \gamma_t\}$ is counted $\prod_{j=a}^b t_j!$ times, where $t_j$ is the number of $k_i$ with $k_i = j$. The previous lemma and Lemma 3 (a) yield

$$\mu(\Omega(k_1, \dots, k_t)) \le \frac{L^{dt}}{\prod_{j=a}^b (t_j!)}(\nu\lambda^{-1/(2d)})^k.$$

(Note that it is safe to use the bound from Lemma 3(a) to bound the number of *trivial* cutsets, since for each trivial cutset, the dual is a single co-component.) Since $\sum_{j=a}^{b} t_j = t$,

$$\prod_{j=a}^{b}(t_j!) \geq \prod_{j=a}^{b}\left(\frac{t_j}{e}\right)^{t_j} \geq \left(\frac{t}{e(b-a+1)}\right)^{t}$$

and hence the result follows.

To prove the second statement, we use the previous lemma and the fact that each non-trivial cutset has at least two co-connected components to bound

$$\mu(\Omega_{n.t.}) \leq \sum_{k=2}^{\infty}\sum_{\gamma_{n.t.}^{(k)}} \lambda^{-|\gamma_{n.t.}^{(k)}|/2d}.$$

Here the sum $\sum_{\gamma_{n.t.}^{(k)}}$ goes over minimal cutsets with $k$ co-components. Using Lemma 3, and the fact that there are at most $L^{kd}$ possibilities for the $k$ starting edges for the $k$ co-components of $\gamma_{n.t.}^{(k)}$, we conclude that

$$\mu(\Omega_{n.t.}) \leq \sum_{k=2}^{\infty}\frac{1}{k!}\left(\sum_{\ell=L^{d-1}}^{\infty} L^d(\nu\lambda^{-2d})^{\ell}\right)^{k} \leq \sum_{k=0}^{\infty}\frac{1}{k!}\left(\sum_{\ell=L^{d-1}}^{\infty} L^d(\nu\lambda^{-2d})^{\ell}\right)^{k+2},$$

which concludes the proof of the second statement.

□

**Lemma 8** *Let $0 < \alpha < 1$, and let*

$$\Omega_\alpha = \{I \in \Omega : \Gamma_{odd}(I) \;\; \textit{contains only trivial cutsets, and } \left|\mathrm{Int}\,\Gamma_{odd}(I)\right| \geq \alpha L^d\}.$$

*If $\lambda$ is sufficiently large, say $\lambda^{1/(2d)} \geq 200\nu/\alpha$, then*

$$\mu(\Omega_\alpha) \leq 2^{-c_\alpha L^{d-1}/(\log L)^2}$$

*for some constant $c_\alpha$ depending on $\alpha$ and $d$.*

**Proof:** For $I \in \Omega_\alpha$, the isoperimetric inequality of Bollobás and Leader [2] implies that $|\gamma| \geq |\mathrm{Int}\,\gamma|^{(d-1)/d}$ and hence

$$\sum_{\gamma\in\Gamma_{odd}(I)} |\gamma|^{d/(d-1)} \geq \sum_{\gamma\in\Gamma_{odd}(I)} |\mathrm{Int}\,\gamma| \geq \Big|\bigcup_{\gamma\in\Gamma_{odd}(I)} \mathrm{Int}\,\gamma\Big| \geq \alpha L^d. \tag{8}$$

If there is a cutset in $\Gamma_{odd}(I)$ of size at least $L^{d-1}$, then Lemma 7 directly gives the desired bound. Assume all cutsets are of size at most $L^{d-1}$. Let $\Gamma_i(I) = \{\gamma \in \Gamma_{odd}(I) : 2^{i-1} \leq |\gamma| < 2^i\}$, $i = 1, 2, ..., r = \lceil \log_2 L^{d-1} + 1\rceil$. Then since $\sum_{i=1}^{\infty}\frac{1}{i^2} = \pi^2/6$, there exists $i$ such that

$$\sum_{\gamma\in L_i} |\gamma|^{d/(d-1)} \geq c_\alpha^* L^d/i^2$$

where $c^*_\alpha = 6\alpha/\pi^2$. Thus $I$ is in $\Omega(k_1, ..., k_t)$ for some $t$ and $k_1, ..., k_t$ with $2^{i-1} \le k_j \le 2^i$ and $\sum_{j=1}^t k_j^{d/(d-1)} \ge c^*_\alpha L^d/i^2$. Moreover, $k_j \le 2^i$ implies that $t \ge c^*_\alpha L^d/(i^2 2^{id/(d-1)})$. This together with Lemma 7 gives

$$\begin{aligned}
\mu(\Omega_\alpha) &\le \sum_{i=1}^r \sum_{t \ge c^*_\alpha L^d/(i^2 2^{id/(d-1)})} \sum_{\substack{2^{i-1} \le k_j < 2^i \\ j=1,\dots,t}} \mu(\Omega(k_1, ..., k_t)) \\
&\le \sum_{i=1}^r \sum_{t \ge c^*_\alpha L^d/(i^2 2^{id/(d-1)})} \sum_{\substack{2^{i-1} \le k_j < 2^i \\ j=1,\dots,t}} (e2^i L^d/t)^t (\nu\lambda^{-1/(2d)})^{\sum k_j}.
\end{aligned}$$

Since $\sum k_j \ge 2^{i-1}t$ and there are at most $2^{it}$ choices for $k_1, k_2, \dots, k_t$,

$$\begin{aligned}
\mu(\Omega_\alpha) &\le \sum_{i=1}^r \sum_{t \ge c^*_\alpha L^d/(i^2 2^{id/(d-1)})} (e2^{2i} L^d/t)^t (\nu\lambda^{-1/(2d)})^{2^{i-1}t} \\
&\le \sum_{i=1}^r \sum_{t \ge c^*_\alpha L^d/(i^2 2^{id/(d-1)})} \left(ei^2 2^{2i} 2^{id/(d-1)} (c^*_\alpha)^{-1} (\nu\lambda^{-1/(2d)})^{2^{i-1}}\right)^t \\
&\le \sum_{i=1}^r \sum_{t \ge c^*_\alpha L^d/(i^2 2^{id/(d-1)})} \left(\left[ei^2 2^{2i} 2^{id/(d-1)}\right]^{2^{1-i}} (c^*_\alpha)^{-1} (\nu\lambda^{-1/(2d)})\right)^{2^{i-1}t},
\end{aligned}$$

where we have used the fact that $(c^*_\alpha)^{-1} \ge 1$ in the last step. Bounding $\left[ei^2 2^{2i} 2^{id/(d-1)}\right]^{2^{1-i}} \le \left[ei^2 (16)^i\right]^{2^{1-i}} \le [e(16e)^i]^{2^{1-i}} \le 16e^2$ we see that for $\lambda$ large enough (e.g. for $\lambda^{1/2d} \ge 32e^2\nu/c^*_\alpha$) , one gets

$$\mu(\Omega_\alpha) \le \sum_{i=1}^r \sum_{t \ge c^*_\alpha L^d/(i^2 2^{id/(d-1)})} 2^{-2^{i-1}t} \le r2^{1-c^*_\alpha 2^{r-1} L^d/(r^2 2^{rd/(d-1)})} \le 2^{-c_\alpha L^{d-1}/(\log L)^2}.$$

□

We show next that if $I$ is chosen from the probability distribution (2), then $|I|$ is unlikely to be small.

**Lemma 9** *Let $0 < \delta < 1$. Then*

$$\mu\Big(|I| \le (1-\delta)\frac{L^d}{2}\Big) \le (2\lambda^{-\delta/2})^{L^d}.$$

**Proof:** There are at most $2^{L^d}$ independent subsets in $T_{L,d}$ and so the weight of those of size at most $(1-\delta)L^d/2$ is at most $2^{L^d}\lambda^{(1-\delta)L^d/2}$. On the other hand, the set of all even points has weight $\lambda^{L^d/2}$. The lemma follows immediately. □

**Lemma 10** *For any $\rho < 1$, there is a constant $c^*_\rho$ such that for $\lambda$ sufficiently large,*

$$\mu(||I \cap V_{odd}| - |I \cap V_{even}|| \le \rho L^d/2) \le \exp(-c^*_\rho L^{d-1}/(\log L)^2).$$

**Proof:** Let $\delta = (1-\rho)/2$. Lemma 7 and Lemma 9 imply that $\mu(\Omega_{n.t.})$ and $\mu(|I| \le (1-\delta)L^d/2)$ are small enough. Moreover, Lemma 8 for $\alpha = (1-\rho)/8$ implies that $\mu(\Omega_\alpha)$ is also small enough. If none of the three events whose probabilities we discuss above occurs, then $|I| > (1-\delta)L^d/2$ and $|\mathrm{Int}\,\Gamma_{odd}(I)| < L^d(1-\rho)/8$. The latter and Lemma 4(d) imply that either $|I \cap V_{odd}| < L^d(1-\rho)/8$ or $|I \cap V_{even}| < L^d(1-\rho)/8$. This together with the former yields that either

$$|I \cap V_{odd}| - |I \cap V_{even}| = |I| - 2|I \cap V_{even}| > (1-\delta)L^d/2 - L^d(1-\rho)/4$$

or

$$|I \cap V_{even}| - |I \cap V_{odd}| = |I| - 2|I \cap V_{odd}| > (1-\delta)L^d/2 - L^d(1-\rho)/4.$$

Since $(1-\delta)L^d/2 - L^d(1-\rho)/4 = \rho L^d/2$, this concludes the proof. $\square$

**Proof of Theorem 2:** We now partition $\Omega = \Omega^{(\rho)}_{\mathrm{odd}} \cup \Omega^{(\rho)}_{\mathrm{even}} \cup \Omega^{(\rho)}_{\mathrm{rest}}$ where

$$\begin{aligned}
\Omega^{(\rho)}_{\mathrm{odd}} &= \{I \in \Omega : \ |I \cap V_{odd}| - |I \cap V_{even}| > \rho L^d/2\} \\
\Omega^{(\rho)}_{\mathrm{even}} &= \{I \in \Omega : \ |I \cap V_{even}| - |I \cap V_{odd}| > \rho L^d/2\} \\
\Omega^{(\rho)}_{\mathrm{rest}} &= \Omega \setminus (\Omega^{(\rho)}_{\mathrm{odd}} \cup \Omega^{(\rho)}_{\mathrm{even}})
\end{aligned}$$

By the last lemma $\mu(\Omega^{(\rho)}_{\mathrm{rest}}) \le \exp(-c^*_\rho L^{d-1}/(\log L)^2)$, and by symmetry $\mu(\Omega^{(\rho)}_{\mathrm{odd}}) = \mu(\Omega^{(\rho)}_{\mathrm{even}})$. Now consider Glauber dynamics. Clearly, if $I \in \Omega^{(\rho)}_{\mathrm{odd}}$ and $I'$ is obtained by a single transition then $I' \in \Omega^{(\rho)}_{\mathrm{odd}} \cup \Omega^{(\rho)}_{\mathrm{rest}}$. The same is true if we generalize from Glauber dynamics to an ergodic Markov chain that is $\rho$-quasi-local. (See the paragraph before Theorem 2 for the definition of $\rho$-quasi-local.) To complete our proof by estimating $\Phi_S$ (see (3)) for $S = \Omega_{\mathrm{odd}}$, first notice $\mu(S)\mu(\overline{S}) \ge 1/5$. Furthermore,

$$Q(S,\overline{S}) = \sum_{\substack{I\in\Omega_{\mathrm{odd}}\\ J\in\Omega_{\mathrm{rest}}}} \mu(I)P(I,J) = \sum_{\substack{I\in\Omega_{\mathrm{odd}}\\ J\in\Omega_{\mathrm{rest}}}} \mu(J)P(J,I) \le \mu(\Omega_{\mathrm{rest}}).$$

The theorem now follows. $\square$

# 6 Swendsen-Wang Algorithm on a $d$-dimensional Torus

This section is deferred to the appendix. We combine the methods and results of [6] and [5] with those of the last section to prove Theorem 1.

**Acknowledgements:** The authors wish to thank Marek Biskup, Michael Freedman and Roman Kotecký for numerous helpful discussions. This work began at Microsoft Research, and four of the authors (A.F., P.T., E.V. and V.V.) would like to thank Microsoft Research, especially the theory group, for providing this opportunity.

# References

[1] D. Aldous and J. Fill, Reversible Markov Chains and Random Walks on Graphs, in preparation. Some chapters available at
http://stat-www.berkeley.edu/pub/users/aldous/book.html.

[2] B. Bollobás and I. Leader, *Edge-isoperimetric inequalities in the grid*, Combinatorica 11 (1991) 299-314.

[3] C. Borgs, J. T. Chayes, A. Frieze, J. H. Kim, P. Tetali, E. Vigoda, V. Vu, in preparation.

[4] C. Borgs and J. Imbrie, *A unified approach to phase diagrams in field theory and statistical mechanics*, Commun. Math. Phys. 123 (1989) 305-328.

[5] C. Borgs and R. Kotecký, *A rigorous theory of finite-size scaling at first-order transitions*, Jour. Statist. Phys. 61 (1990) 79–110.

[6] C. Borgs, R. Kotecký and S. Miracle-Solé, *Finite-size scaling for Potts models*, Jour. Statist. Phys. 62 (1991) 529–551.

[7] C. Cooper, M. Dyer, A. M. Frieze and R. Rue, *Mixing properties of the Swendsen-Wang process on classes of graphs II*, in preparation.

[8] C. Cooper and A. M. Frieze, *Mixing Properties of the Swendsen-Wang process on classes of graphs*, to appear in Proceedings of DIMACS Workshop on Statistical Physics Methods in Discrete Probability, Combinatorics and Theoretical Computer Science.

so.

[9] R. L. Dobrushin, *An investigation of Gibbs states for three-dimensional lattice systems* (Russian with English summary), Teor. Verojatnost. i Primenen. 18 (1973) 261–279.

English translation: Theor. Probability Appl. 18 (1974) 253–271.

[10] M. E. Dyer, A. M. Frieze and M. R. Jerrum, *On counting independent sets in sparse graphs*, preprint.

[11] M. E. Dyer and C. Greenhill, *On Markov chains for independent sets*, preprint.

[12] R. G. Edwards and A. D. Sokal *Generalizations of the Fortuin-Kasteleyn-Swendsen-Wang representation and Monte Carlo algorithm.* Phys. Rev. D 38 (1988) 2009-2012.

[13] C. Fortuin and P. Kasteleyn. *On the random cluster model I: Introduction and relation to other models.* Physica 57 (1972) 536-564.

[14] V. Gore and M. Jerrum. *The Swendsen-Wang process does not always mix rapidly.* Proceedings of the 29th Annual ACM Symposium on Theory of Computing, (1997) 674-681.

[15] M. Huber, *Efficient exact sampling from the Ising model using Swendsen-Wang*, Proceedings of the Symposium on Discrete Algorithms, (1999).

[16] M. R. Jerrum, Talk given at Workshop on Randomised Approximation and Stochastic Simulation in Warwick, England, 1998.

[17] L. Laanait, A. Messager, S. Miracle-Sole, J. Ruiz and S. Shlossman, *Interfaces in the Potts models I: Pirogov-Sinai theory of the Fortuin-Kasteleyn representation*, Commun. Math. Phys. 140 (1991) 81.

[18] J. L. Lebowitz and A. E. Mazel, *Improved Peierls argument for higher dimensional Ising models*, Jour. Statist. Phys. 90 (1998) 1051-1059.

[19] X-J. Li and A. D. Sokal, *Rigorous lower bound on the dynamic critical exponents of the Swendsen-Wang algorithm*, Phys. Rev. Lett. 63 (1989) 827-830.

[20] M. Luby and E. Vigoda, *Fast convergence of the Glauber dynamics for sampling independent sets*, to appear in Proceedings of DIMACS Workshop on Statistical Physics Methods in Discrete Probability, Combinatorics and Theoretical Computer Science.

[21] F. Martinelli, *Dynamical Analysis of the low temperature cluster algorithms*, Jour. Statist. Phys., 66 (1992) 1245-1276.

[22] S. A. Pirogov and Ya.G. Sinai, *Phase diagrams of classical lattice systems*, Theor. Math. Phys. 25 (1975) 1185–1192; *Phase diagrams of classical lattice systems. Continuation*, Theor. Math. Phys. 26 (1976) 39–49.

[23] D. Ruelle, *Statistical Mechanics: Rigorous Results*, W. A. Benjamin, 1969.

[24] A. J. Sinclair and M. R. Jerrum, *Approximate counting, uniform generation and rapidly mixing Markov chains*, Information and Computation 82 (1989), 93-133.

[25] R. Swendsen and J-S. Wang, *Non-universal critical dynamics in Monte-Carlo simulation.* Phys. Rev. Lett. 58 (1987) 86-88.

[26] L. Thomas, *Bound on the mass gap for finite volume stochastic Ising models at low temperature*, Commun. Math. Phys. 126 (1989) 1-11.

[27] D. J. A. Welsh, *Complexity: Knots, Colourings and Counting*, London Mathematical Society Lecture Note Series 186, 1993.

[28] F. Y. Wu, *The Potts model*, Rev. Mod. Phys. 54 (1982) 235-268.

# Appenices

## A Proof of Lemma 1

**Proof:**

**(a)** We will first prove uniqueness. Since $C \cap C' = \emptyset$, $C' \subset \overline{C}$ and $C \subset \overline{C'}$. Furthermore, $C$ is connected. Hence, there exists a unique $D' \in \mathcal{C}(\overline{C'})$ such that $C \subset D'$. For all $D \in \mathcal{C}(\overline{C})$, $C \subset \overline{D}$. Therefore, if there exists a $D' \in \mathcal{C}(\overline{C'})$ with $\overline{D} \subset D'$, $D'$ must be the unique component containing $C$. The uniqueness of $D$ is proved similarly. Next, we prove existence. Let $D'$ be as above, so that $\overline{D'} \subset \overline{C}$. Since $C \cup D''$ is connected for all

$D'' \in \mathcal{C}(\overline{C'})$, the set $\overline{D'} = C' \cup \bigcup\{D'' \in \mathcal{C}(\overline{C'}) : D'' \neq D'\}$ is connected. As a consequence, $\overline{D'} \subset \overline{C}$ must lie in one of the components $D$ of $\overline{C}$.

**(b)** Obviously, $\partial C = \cup_{\gamma \in \Gamma_C}$ is a decomposition of $\partial C$ into minimal cutsets of $G$. To prove uniqueness, assume that $\gamma \subset \partial C$ is a minimal cutset of $G$. Then there exists a $D \in \mathcal{C}(\overline{C})$ such that $(D : \overline{D}) \subset \gamma$. Otherwise, $C \cup D$ is connected in $G \setminus \gamma$ for every $D \in \mathcal{C}(\overline{C})$, which would imply that $G \setminus \gamma$ is connected. Since $\gamma$ is minimal, $(D : \overline{D}) \subset \gamma$ implies $(D : \overline{D}) = \gamma$.

**(c)** For cutsets $\gamma$ and $\gamma'$ corresponding to the same component $C$, disjointness follows from the explicit form given in (b). Assume that $\gamma \cap \gamma' \neq \emptyset$ for two different components $C$ and $C'$. This would imply that $\partial C \cap \partial C' \neq \emptyset$, which in turn implies that $C$ and $C'$ are connected in $G$, and hence in $G_W$. But this contradicts the assumption that $C$ and $C'$ are different components of $G_W$.

**(d)** Without loss of generality $X \in \mathcal{C}(\overline{C})$ and $Y \in \mathcal{C}(\overline{C'})$. We consider several cases:

- If $X = Y$ then $X \cap \overline{Y} = \emptyset$.
- If $C = C'$ and $X \neq Y$, then $X$ and $Y$ are different components of $\overline{C}$ which implies that $X \cap Y = \emptyset$.
- If $C \neq C'$ then we use part (a) of this lemma. We condition on whether $X$ and/or $Y$ are the unique $D \in \mathcal{C}(\overline{C})$ and $D' \in \mathcal{C}(\overline{C}')$ such that $\overline{D} \subseteq D'$.
  - $X \neq D, Y \neq D'$ : Since $Y \subset \overline{D'}$ and part (a) implies that $X \subset \overline{D} \subset D'$, so we have that $X \cap Y = \emptyset$.
  - $X \neq D, Y = D'$: We saw in the previous case $X \subset D'$ and thus $X \cap \overline{Y} = \emptyset$. The case when $X = D, Y \neq D'$ is symmetric.
  - $X = D, Y = D'$: Since $\overline{X} \subset Y$ by part (a), $\overline{X} \cap \overline{Y} = \emptyset$.

□

# B Proof of Lemma 2

**Proof: (a)** Let $X = \operatorname{Int}\gamma$ and $Y = \operatorname{Int}\gamma'$, and assume without loss of generality that $X \cap Y \neq \emptyset$. Applying the previous lemma, we have three cases:

(i) $X \cap \overline{Y} = \emptyset$, which is equivalent to $X \subset Y$,

(ii) $\overline{X} \cap Y = \emptyset$, which is equivalent to $Y \subset X$, and

(iii) $\overline{X} \cap \overline{Y} = \emptyset$ which is equivalent to $\overline{X} \subset Y$. Notice that $|\overline{X}| \geq |V|/2$ which implies that $|Y| \geq |V|/2$ and $|\overline{Y}| \leq |V|/2$. This contradicts the fact that $|Y| = |\operatorname{Int}\gamma'| \leq |\overline{Y}|$ unless equality holds, i.e. unless $|Y| = |\overline{Y}| = |X| = |\overline{X}| = |V|/2$. Together with $\overline{X} \subset Y$, this implies $\overline{X} = Y$ in contradiction to our assumption $X \cap Y \neq \emptyset$.

**(b)** We consider two cases. Suppose that for every $C \in \mathcal{C}(W)$ there is a cutset $\gamma \in \Gamma(W)$ with $C \subset \operatorname{Int}\gamma$. Then, clearly

$$W = \bigcup_{C \in \mathcal{C}(W)} C \subset \bigcup_{\gamma \in \Gamma(W)} \operatorname{Int}\gamma.$$

Suppose instead that there is $C \in \mathcal{C}(W)$ such that $C \not\subset \operatorname{Int}\gamma$ for all $\gamma$. Then since $C$ is a subset of $\overline{D}$ for every component $D$ of $\overline{C}$, the interior of the corresponding cutset $\gamma_D = (D : \overline{D})$ must be $D$. Thus $\overline{C} = \cup_{D\in\mathcal{C}(\overline{C})}\operatorname{Int}\gamma_D$. In particular, since $C$ is a component of $W$,

$$\overline{W} \subset \overline{C} \subset \bigcup_{\gamma\in\Gamma(W)} \operatorname{Int}\gamma.$$

□

# C Proof of Lemma 3

**Proof:**
**(a)** This follows from the observation that the proofs in [23] and [18] may be applied without changes to the torus.

**(b)** We need some notation. Consider a set of edges $X$ and its dual $X^*$. Define the boundary $\partial X^*$ of $X^*$ as the set of $(d-2)$-dimensional hypercubes which belong to an odd number of $(d-1)$-dimensional cells in $X^*$. If $\partial X^* = \emptyset$, define the $Z_2$ winding vector of $X^*$ as the vector $\mathbf{N}(X^*) = (N_1, \ldots, N_d)$, where $N_i$ is the number of times $X^*$ intersects an elementary loop in the $i^{\text{th}}$ lattice direction mod 2.

Let $X$ be a cutset, $X = (W : \overline{W})$, where $W \subset V$. Let $\mathcal{W} \subset (\boldsymbol{R}/\boldsymbol{Z})^d$ be the union of all closed unit cubes with center $w \in W$. Then $X^*$ is the boundary of the set $\mathcal{W}$, and hence $\partial X^* = \emptyset$. Obviously, each elementary loop must leave and enter the set $\mathcal{W}$ the same number of times, implying that the winding vector of $X^*$ is 0. On the other hand, it is not difficult to prove that each set of edges $X$ with $\partial X^* = \emptyset$ and $\mathbf{N}(X^*) = 0$ is a cutset for some set of points $W \subset V$, $X = (W : \overline{W})$. Indeed, the assumptions $\partial X^* = \emptyset$ and $\mathbf{N}(X^*) = 0$ imply that every closed loop in $T_{L,d}$ intersects $X^*$ an even number of times. Considering an arbitrary vertex $w_0 \in V$ and the set of all "walks" of the form $(w_0, w_1, \ldots, w_k)$, $\{w_i, w_{i+1}\} \in \mathbf{E}_{L,d}$, we then define $W$ as the set of points which can be reached from $w_0$ by a walk which intersects $X^*$ an odd number of times.

Consider now a non-trivial minimal cutset $\gamma$ and one of its co-components $\tilde{\gamma}$. Since $\gamma$ is a cutset, $\partial\gamma^* = \emptyset$. This property is inherited by all its co-components, implying that $\partial\tilde{\gamma}^* = \emptyset$. Obviously, $\mathbf{N}(\tilde{\gamma}^*)$ is different from zero, since otherwise $\tilde{\gamma}$ would be a cutset itself, in contradiction to the assumption that $\gamma$ is minimal. Let $j$ be a direction for which $N_j(\tilde{\gamma}^*) \neq 0$. Then $\tilde{\gamma}^*$ intersects any fundamental loop in the $j$-direction an odd number of times, giving that $\tilde{\gamma}^*$ contains at least $L^{d-1}$ dual $(d-1)$-dimensional cells.

□

# D Swendsen-Wang Algorithm on a $d$-dimensional Torus

In this section we combine the methods and results of [6] and [5] with those of the last section to prove Theorem 1.

Recall the standard representation of the Potts model in Section 1. On the graph $\mathbf{Z}^d$, $d \geq 2$, this model is known to undergo a phase transition as the inverse temperature, $\beta$, passes through a certain critical temperature $\beta_c = \beta_c(q, d)$. To make this statement precise, we introduce finite-volume distributions with boundary conditions. We consider the graph $G=(V,E)$, where $V=[L]^d$ and $E$ consists of all pairs of vertices in $V$ whose coordinates differ by 1 in one coordinate. We say that a vertex lies in the (inner) boundary of $V$ if one of its coordinates is either 1 or $L$. For a coloring $\boldsymbol{\sigma}$ of $V$, we then introduce the weights $w_{L,k}(\boldsymbol{\sigma})=e^{-\beta d(\boldsymbol{\sigma})+\beta n_k(\boldsymbol{\sigma})}$, where $n_k(\boldsymbol{\sigma})$ is the number of vertices in the boundary of $V$ that have color $k$. With $Z_{L,k}=\sum_{\boldsymbol{\sigma}} w_{L,k}(\boldsymbol{\sigma})$, the finite-volume distributions $\mu_{L,k}$ with boundary condition $k$ are then defined as $\mu_{L,k}(\boldsymbol{\sigma})=w_{L,k}(\boldsymbol{\sigma})/Z_{L,k}$, and the spontaneous magnetization $m^*(\beta)$ is defined as the $L\to\infty$ limit of the finite-volume magnetizations $m_L(\beta)=L^{-d}\sum_{x\in V}(\mu_{L,1}(\sigma_x=1)-1/q)$. The above-mentioned phase transition can then be characterized as a transition between a high-temperature, disordered region $\beta<\beta_c$ where the spontaneous magnetization is zero, and a low-temperature, ordered region $\beta>\beta_c$ where the spontaneous magnetization is positive.

As a first step towards proving Theorem 1, we define the contours corresponding to a configuration $A\in\Omega=2^E$. To this end, we embed the vertex set $V$ of the torus $T=(V,E)$ into the set $\mathbf{V}=(\mathbf{R}/(L\mathbf{Z}))^d$. For a set $X\subset\mathbf{V}$, we define its diameter $\operatorname{diam}(X)=\inf_{y\in\mathbf{V}}\sup_{x\in X}\operatorname{dist}(x,y)$, where $\operatorname{dist}(x,y)$ is the $\ell_\infty$-distance between the two points $x$ and $y$ in the torus $\mathbf{V}$. For an edge $e=\{x,y\}\in E$, let $\mathbf{e}$ be the set of points in $\mathbf{V}$ that lie on the line between $x$ and $y$. Given $A$, we call a closed $k$-dimensional unit hypercube $c\subset\mathbf{V}$ with vertices in $V$ *occupied* if all edges $e$ with $\mathbf{e}\subset c$ are in $A$. We then define the set $\mathbf{V}(A)\subset\mathbf{V}$ as the 1/3-neighbourhood of the union of all occupied $k$-dimensional hypercubes, $k=1,\dots,d$, i.e., $\mathbf{V}(A)=\{x\in\mathbf{V}:\ \exists c \text{ occupied, such that} \operatorname{dist}(x,c)<1/3\}$, and the set $V(A)$ as the intersection of $\mathbf{V}(A)$ with the vertex set $V$ of the discrete torus $T$. Note that $V(A)=\bigcup_{\{x,y\}}\{x,y\}$. The set $\Gamma(A)$ of contours corresponding to a configuration $A\in\Omega$ are then the components of the boundary of $\mathbf{V}(A)$.

Following [6], we decompose the set of configurations $\Omega$ into three sets $\Omega_{\rm ord}$, $\Omega_{\rm dis}$ and $\Omega_{\rm Big}$. To this end, we define a contour $\gamma$ to be small if $\operatorname{diam}(\gamma)\le L/3$. The set $\Omega_{\rm Big}$ is then just the set of configurations $A\in\Omega$ for which $\Gamma(A)$ contains at least one contour that is not small. Next, restricting ourselves to small contours $\gamma$, we define the set $\mathbf{Ext}\,\gamma$ as the larger of the two components of $\mathbf{V}\setminus\gamma$, the set $\operatorname{Ext}\gamma$ as the intersection of $\mathbf{Ext}\,\gamma$ with $V$, and the set $\operatorname{Int}\gamma$ as $V\setminus\operatorname{Ext}\gamma$. For $A\in\Omega\setminus\Omega_{\rm Big}$, let $\operatorname{Int}A=\bigcup_{\gamma\in\Gamma(A)}\operatorname{Int}\gamma$ and $\operatorname{Ext}A=V\setminus\operatorname{Int}A$. The sets $\Omega_{\rm ord}$, $\Omega_{\rm dis}$ and $\Omega_{\rm Big}$ are then defined as

$$\begin{aligned}
\Omega_{\rm Big}&=\{A\subset E:\ \exists\gamma\in\Gamma(A)\text{ such that }\operatorname{diam}(\gamma)>L/3\}\\
\Omega_{\rm ord}&=\{A\subset E:\ \operatorname{diam}(\gamma)\le L/3\ \forall\gamma\in\Gamma(A)\text{ and }V(A)\cap\operatorname{Ext}A\neq\emptyset\}\\
\Omega_{\rm dis}&=\{A\subset E:\ \operatorname{diam}(\gamma)\le L/3\ \forall\gamma\in\Gamma(A)\text{ and }V(A)\cap\operatorname{Ext}A=\emptyset\}.
\end{aligned}$$

**Lemma 11** *Let $A\in\Omega_{\rm ord}$, and let $A_{\operatorname{Ext}A}=\{b\in E:\ b\subset\operatorname{Ext}A\}$. Then*
**(a)** $\operatorname{Ext}A=V(A)\cap\operatorname{Ext}A\neq\emptyset$*, and*
**(b)** $(\operatorname{Ext}A,A_{\operatorname{Ext}A})$ *is connected.*

**Proof:** (a) Proceeding as in the proof of Lemma 2 (b), we obtain that either $V(A)\subset\operatorname{Int}A$ or $\overline{V(A)}\subset\operatorname{Int}A$. Since $A\in\Omega_{\rm ord}$, we conclude that the latter is the case, which is equivalent to the statement that $\operatorname{Ext}A=V(A)\cap\operatorname{Ext}A$.

(b) The proof of this statement, which is implicit in [6], is straightforward but tedious. We leave it to the reader.

$\square$

In the next lemma we summarize some of the results of [6] used in this paper. We need some notation. Let $A \in \Omega \setminus \Omega_{\mathrm{Big}}$, and let $\gamma \in \Gamma(A)$. We say that $\gamma$ is an exterior contour in $\Gamma(A)$ if $\gamma \subset \mathrm{Ext}\,\gamma'$ for all $\gamma' \in \Gamma(A) \setminus \{\gamma\}$, and denote by $\Gamma_{\mathrm{ext}}(A)$ the set of exterior contours in $\Gamma(A)$. Also, we define the size $\|\gamma\|$ of a contour $\gamma$ as the number of times $\gamma$ intersects the set $\bigcup_{e\in E} \mathbf{e}$. In order to motivate this definition, assume for a moment that the definition of the set $\mathbf{V}(A)$ had involved an $\epsilon$-neighborhood, instead of the 1/3-neighborhood used above. With such a definition, the $(d-1)$-dimensional area of a contour $\gamma$ would actually converge to $\|\gamma\|$ as $\epsilon \to 1/2$.

**Lemma 12** *For all $d \geq 2$ there are constants $c > 0$ and $q_0 < \infty$ such that the following statements hold for $q \geq q_0$.*
**(a)** $\beta_c = \log q/d + O(q^{-c})$.
**(b)** *For all $\beta > 0$,*

$$\mu(\Omega_{\mathrm{Big}}) \leq q^{-cL}.$$

**(c)** *If $\beta = \beta_c$, then*

$$\mu(\Omega_{\mathrm{ord}}) = \frac{q}{q+1} + O(q^{-cL}) \qquad \text{and} \qquad \mu(\Omega_{\mathrm{dis}}) = \frac{1}{q+1} + O(q^{-cL}).$$

**(d)** *If $\beta \geq \beta_c$, then*

$$\mu(\Omega_{\mathrm{ord}}) \geq \frac{q}{q+1} + O(q^{-cL}).$$

**(e)** *If $\beta \geq \beta_c$ and $\Gamma$ is a set of contours, then*

$$\mu\Big(A \in \Omega \setminus \Omega_{\mathrm{big}} \text{ and } \Gamma \subset \Gamma_{\mathrm{ext}}(A)\Big) \leq q^{-c\sum_{\gamma\in\Gamma}\|\gamma\|}.$$

Observing that for $A \in \Omega \setminus \Omega_{\mathrm{Big}}$, the set $\mathrm{Ext}\,A$ can be written as $\bigcup_{\gamma\in\Gamma_{\mathrm{ext}}} \mathrm{Ext}\,\gamma$, which in turn implies that $\mathrm{Int}\,A = \bigcap_{\gamma\in\Gamma_{\mathrm{ext}}} \mathrm{Int}\,\gamma$, we can now continue as in Section 5 to prove an analog of Lemma 8. Defining

$$\begin{aligned}
\Omega_{\mathrm{ord}}^{(\alpha)} &= \{A \in \Omega_{\mathrm{ord}} :\ |\{b \in A :\ b \subset \mathrm{Ext}\,A\}| \geq (1-\alpha)dL^d\},\\
\Omega_{\mathrm{dis}}^{(\alpha)} &= \{A \in \Omega_{\mathrm{dis}} :\ |\mathrm{Int}\,A| \leq \alpha L^d\},
\end{aligned}$$

and $\Omega_{\mathrm{Big}}^{(\alpha)} = \Omega \setminus (\Omega_{\mathrm{ord}}^{(\alpha)} \cup \Omega_{\mathrm{dis}}^{(\alpha)})$, we therefore get the following lemma.

**Lemma 13** *Let $d \geq 2$ and $0 < \alpha < 1$. Then there are constants $c > 0$ and $c_\alpha > 0$ such that for $q$ large enough the following statements hold.*
**(a)** *If $\beta \geq \beta_c$, then*

$$\mu(\Omega_{\mathrm{Big}}^{(\alpha)}) = O(q^{-cL}) + O(q^{-c_\alpha L^{d-1}/(\log L)^2})$$

*and*

$$\mu(\Omega_{\mathrm{ord}}^{(\alpha)}) \geq \frac{q}{q+1} + O(q^{-cL}) + O(q^{-c_\alpha L^{d-1}/(\log L)^2}).$$

**(b)** *If $\beta = \beta_c$, then*

$$\mu(\Omega_{\mathrm{ord}}^{(\alpha)}) = \frac{q}{q+1} + O(q^{-cL}) + O(q^{-c_\alpha L^{d-1}/(\log L)^2}).$$

**Proof of Theorem 1(a):** Let $S = \Omega_{\text{ord}}^{(\alpha)}$. The conductance $\Phi_{SW}$ of the Swendsen-Wang chain can then be estimated as follows:

$$\Phi_{SW} \leq \Phi_S = \frac{1}{\mu(\overline{S})} \Pr(A' \notin \Omega_{\text{ord}}^{(\alpha)} \mid A \in \Omega_{\text{ord}}^{(\alpha)}). \tag{9}$$

Here $A$ is chosen according to the measure $\mu$ defined in (5) and $A'$ is constructed from $A$ by one step of the Swendsen-Wang algorithm. We have

$$\Pr(A' \notin \Omega_{\text{ord}}^{(\alpha)} \mid A \in \Omega_{\text{ord}}^{(\alpha)}) = \Pr(A' \in \Omega_{\text{dis}}^{(\alpha)} \mid A \in \Omega_{\text{ord}}^{(\alpha)}) + \Pr(A' \in \Omega_{\text{Big}}^{(\alpha)} \mid A \in \Omega_{\text{ord}}^{(\alpha)}).$$

Observing that $A \in \Omega_{\text{ord}}^{(\alpha)}$ implies $|A| \geq (1-\alpha)dL^d$ while $A' \in \Omega_{\text{dis}}^{(\alpha)}$ implies $|A'| \leq d|V(A')| \leq d|\text{Int}\, A'| \leq d\alpha L^d$, we see that $A'$ can only be in $\Omega_{\text{dis}}^{(\alpha)}$ if at least $(1-2\alpha)dL^d$ edges are deleted in Step (SW1) of Swendsen-Wang. But the number of edges deleted is dominated by the binomial $B(dL^d, 1-p_c)$ and so

$$\begin{aligned}
\Pr(A' \in \Omega_{\text{dis}}^{(\alpha)} \mid A \in \Omega_{\text{ord}}^{(\alpha)}) &\leq \binom{dL^d}{(1-2\alpha)dL^d} (1-p_c)^{(1-2\alpha)dL^d} \\
&\leq \left(\frac{e(1-p_c)}{1-2\alpha}\right)^{(1-2\alpha)dL^d} \\
&= e^{-\Omega((\log q)L^d)},
\end{aligned}$$

where we have used Lemma 12(a) to bound $1 - p_c = e^{-\beta_c} = e^{-\Omega(\log q)}$ . Also

$$\begin{aligned}
\Pr(A' \in \Omega_{\text{Big}}^{(\alpha)} \mid A \in \Omega_{\text{ord}}^{(\alpha)}) &\leq \frac{\Pr(A' \in \Omega_{\text{Big}}^{(\alpha)})}{\Pr(A \in \Omega_{\text{ord}}^{(\alpha)})} \\
&= O(q^{-cL}) + O(q^{-c_\alpha L^{d-1}/(\log L)^2}),
\end{aligned}$$

by Lemma 13(a). Using Lemma 13(b) to bound $\mu(\overline{S}) = 1 - \mu(\Omega_{\text{ord}}^{(\alpha)})$ from below, we obtain that

$$\Phi_{SW} = O(q^{-cL}) + O(q^{-c_\alpha L^{d-1}/(\log L)^2}).$$

$\square$

**Proof of Theorem 1(b):**

Let $\widehat{\Omega} = [q]^V$ be the set of colorings, and let $V_k(\boldsymbol{\sigma}) = \{x \in V : \ \sigma_x = k\}$ be the set of vertices that have color $k$ in the coloring $\boldsymbol{\sigma} \in \widehat{\Omega}$. We then define the sets

$$\widehat{\Omega}_k^{(\alpha)} = \{\boldsymbol{\sigma} \in \widehat{\Omega} : \ |V_k| \geq (1-\alpha)|V|\}, \quad , k \in [q],$$

$$\widehat{\Omega}_{\text{ord}}^{(\alpha)} = \bigcup_{k \in [q]} \widehat{\Omega}_k^{(\alpha)},$$

$$\widehat{\Omega}_{\text{dis}}^{(\alpha)} = \{\boldsymbol{\sigma} \in \widehat{\Omega} : \ |V_k| \geq \frac{(1-\alpha)^2}{q}|V| \text{ for all } k \in [q]\},$$

and

$$\widehat{\Omega}^{(\alpha)}_{\text{Rest}} = \widehat{\Omega} \setminus (\widehat{\Omega}^{(\alpha)}_{\text{ord}} \cup \widehat{\Omega}^{(\alpha)}_{\text{dis}}).$$

To estimate the probability of $\widehat{\Omega}^{(\alpha)}_{\text{ord}}$ in the measure (1), we use the fact that both the measure (1) (denoted $\widehat{\mu}$ in this section) and the measure (5) (denoted $\mu$ in this section) are marginals of the Edwards-Sokal measure (6). Thus

$$\widehat{\mu}(\widehat{\Omega}^{(\alpha)}_{\text{ord}}) = \sum_{A\in\Omega} \pi(\widehat{\Omega}^{(\alpha)}_{\text{ord}} \mid A)\, \mu(A), \tag{10}$$

where $\pi(\widehat{\Omega}^{(\alpha)}_{\text{ord}} \mid A)$ is the conditional measure of $\widehat{\Omega}^{(\alpha)}_{\text{ord}}$, given $A \in \Omega$. Observing that $A \in \Omega^{(\alpha)}_{\text{ord}}$ implies that all vertices in $\operatorname{Ext} A$ have the same color by Lemma 11 and the definition (6) of $\pi$, we have that

$$\pi(\widehat{\Omega}^{(\alpha)}_{\text{ord}} \mid A) = 1 \qquad \text{if} \qquad A \in \Omega^{(\alpha)}_{\text{ord}}. \tag{11}$$

For $A \in \Omega^{(\alpha)}_{\text{dis}}$, on the other hand, all $n_A = |\operatorname{Ext} A| \geq (1-\alpha)|V|$ vertices in $\operatorname{Ext} A$ are colored independently of each other, so that

$$\begin{aligned}
\pi\Big(|V_k(\boldsymbol{\sigma}) \cap V| \leq (1-\alpha)^2 V) \;\Big|\; A\Big) &\leq \pi\Big(|V_k(\boldsymbol{\sigma}) \cap \operatorname{Ext} A| \leq (1-\alpha) n_A) \;\Big|\; A\Big) \\
&= \sum_{k \leq (1-\alpha)|n_A} \binom{n_A}{k} \Big(\frac{1}{q}\Big)^k \Big(1 - \frac{1}{q}\Big)^{n_A - k} \\
&\leq e^{-c^* L^d}
\end{aligned}$$

for some constant $c^*$ depending on $q$ and $\alpha$. As a consequence,

$$\pi(\widehat{\Omega}^{(\alpha)}_{\text{dis}} \mid A) \geq 1 - O(e^{-c^* L^d}) \qquad \text{if} \qquad A \in \Omega^{(\alpha)}_{\text{dis}}. \tag{12}$$

Combining (10) – (12) with Lemma 13 and the fact that $\widehat{\Omega}^{(\alpha)}_{\text{dis}} \cap \widehat{\Omega}^{(\alpha)}_{\text{ord}} = \emptyset$ if $\alpha$ is chosen small enough, we then get

$$\widehat{\mu}(\widehat{\Omega}^{(\alpha)}_k) = \frac{1}{q} \widehat{\mu}(\widehat{\Omega}^{(\alpha)}_{\text{ord}}) = \frac{1}{q} \mu(\Omega^{(\alpha)}_{\text{ord}}) + O(e^{-c^* L^d}) + O(q^{-cL}) + O(q^{-c_\alpha L^{d-1}/(\log L)^2}),$$

$$\widehat{\mu}(\widehat{\Omega}^{(\alpha)}_{\text{dis}}) = \mu(\Omega^{(\alpha)}_{\text{dis}}) + O(e^{-c^* L^d}) + O(q^{-cL}) + O(q^{-c_\alpha L^{d-1}/(\log L)^2})$$

and

$$\widehat{\mu}(\widehat{\Omega}^{(\alpha)}_{\text{Rest}}) = O(e^{-c^* L^d}) + O(q^{-cL}) + O(q^{-c_\alpha L^{d-1}/(\log L)^2}).$$

We complete our proof by estimating $\Phi_S$ (see (3)) for $S = \widehat{\Omega}^{(\alpha)}_k$. First notice $\widehat{\mu}(S)\widehat{\mu}(\overline{S}) \geq (1-1/q)/2q$. Since the heat bath algorithm can only change one vertex at a time, it does not make transitions between the different sets $\widehat{\Omega}^{(\alpha)}_k$, nor does it make transitions between $\widehat{\Omega}^{(\alpha)}_1$ and $\Omega^{(\alpha)}_{\text{dis}}$. Thus

$$Q(S, \overline{S}) = \sum_{I \in \widehat{\Omega}^{(\alpha)}_1, J \in \Omega^{(\alpha)}_{\text{Rest}}} \widehat{\mu}(I) P(I,J) = \sum_{I \in \widehat{\Omega}^{(\alpha)}_1, J \in \Omega^{(\alpha)}_{\text{Rest}}} \widehat{\mu}(J) P(J,I) \leq \widehat{\mu}(\Omega^{(\alpha)}_{\text{Rest}}).$$

The theorem now follows. $\square$