A Bounding Chain for Swendsen-Wang

Mark Huber
Department of Mathematics and Institute of Statistics and Decision Sciences, Duke University, Durham, North Carolina 27713, e-mail: mhuber@math.duke.edu

Received 10 October 2001; revised 13 May 2002; accepted 8 June 2002

DOI 10.1002/rsa.10071

ABSTRACT: The greatst drawback of Monte Carlo Markov chain methods is lack of knowledge of the mixing time of the chain. The use of bounding chains solves this difficulty for some chains by giving theoretical and experimental upper bounds on the mixing time. Moreover, when used with methodologies such as coupling from the past, bounding chains allow the user to take samples drawn exactly from the stationary distribution without knowledge of the mixing time. Here we present a bounding chain for the Swendsen-Wang process. The Swendsen-Wang bounding chain allow us to efficiently obtain exact samples from the ferromagnetic Q-state Potts model for certain classes of graphs. Also, by analyzing this bounding chain, we will show that Swendsen-Wang is rapidly mixing over a slightly larger range of parameters than was known previously. © 2002 Wiley Periodicals, Inc. Random Struct. Alg., 22: 43–59, 2003

1. INTRODUCTION

Many distributions are difficult to sample from directly. One method for approximately sampling from these hard distributions is to run a Markov chain with the desired distribution as its stationary distribution for "a long time." The fundamental problem in Markov chain theory is discovering exactly how long a long time is, known as the mixing time of the chain.

The idea of a *bounding chain* solves this problem for some chains. A bounding chain is itself a Markov chain that in a certain sense contains the original Markov chain. A bounding chain gives a means for determining when *complete coupling* of a Markov chain

has occurred. Complete couplings are stronger than simple couplings, and so, with bounding chains, the user obtains theoretical bounds on the mixing time similar to those obtained using coupling arguments. Moreover, using a bounding chain, simulations can be used to obtain upper bounds on the mixing time that are correct with high probability. Older techniques such as the autocorrelation function can only give evidence that a chain has not mixed, but the bounding chain technique can actually show that a chain has mixed. More importantly, the bounding chain idea can be used together with recent ideas such as coupling from the past [16] to draw samples that come exactly from the stationary distribution of the chain.

Bounding chains were first introduced independently in [9] and [8] for the problem of uniformly generating k-colorings of a graph. Since then, the technique has been applied to a variety of examples [13, 12]. Here we present a bounding chain for the Swendsen-Wang Markov chain. Previously, a bounding chain for Swendsen-Wang was presented in [11, 10]; the one presented here is stronger in the sense that it always finishes at the same time or before the earlier chain. Variants of this work also appeared previously in the author's Ph.D. thesis [11].

The Swendsen-Wang chain is a Markov chain whose stationary distribution is that of the ferromagnetic Potts model [19]. Just running the Swendsen-Wang chain for a fixed length of time will give samples that only approximately come from the stationary distribution, since there is always some bias from the choice of starting state. With the bounding chain we will obtain exact samples; moreover, when a parameter of the chain lies in a specified range, this technique will have a near linear expected running time.

2. SWENDSEN-WANG

The Ising model and its extensions have been extensively used by statistical physicists to model the behavior of magnetic materials and alloys. Simon [17] gives a detailed analysis of the Ising model; we present a summary here. In the Ising model, each node of a graph is assigned one of two values, spin up or spin down, that we record as 0 or 1. Such an assignment for every node is called a configuration. The set of possible configurations is $\{0, 1\}^n$. Let x be a configuration. We say that x has energy $H(x)$, which is equal to the negative of the number of edges in the graph with endpoints labeled the same way. For physical reasons, a probability distribution known as the Gibbs distribution is assigned to the set of states, defined as

$$g(x) = \frac{e^{-JH(x)/(kT)}}{Z_T},$$

where Z_T is the normalizing constant that makes g a probability distribution. Often, Z_T is referred to as the partition function. The parameter T is a measure of the temperature, and k is Boltzmann's constant. When $J = 1$, this measure assigns higher weight to cases where neighboring nodes are colored alike—this situation is known as the ferromagnetic Ising model. When $J = -1$, configurations with large numbers of edges colored differently are more likely. This is the antiferromagnetic Ising model.

While in the ferromagnetic case a polynomial time algorithm exists for drawing approximate samples [14], it is far from linear time, and the practical method for drawing random samples from this model is to run a Markov chain which has the set of configurations as its state space, and the Gibbs distribution as its stationary distribution.

One extension of the Ising model is to the Potts model. Here, each state is assigned a value from $\{1, \ldots, Q\}$. When $Q = 2$, we, of course, just have the standard Ising model. Again $H(x)$ is just the negative of the number of edges with identically labeled endpoints, and the Gibbs distribution is the same as before.

Fortuin and Kastelyn [5] showed that there was another way of looking at the Potts model, known as the random cluster model. Suppose that our graph is $G = (\{1, \ldots, n\}, E)$. Then the random cluster model considers subsets of the edges. Let $B \subseteq E$. The probability assigned to B is just

$$\pi(B) = p^{|B|}(1 - p)^{|E \setminus B|} Q^{C(B)} / Z_{p,Q},$$

where Z is again the normalizing constant that makes this a probability distribution, p is a parameter of the model which varies from 0 to 1, and $C(B)$ is the number of connected components of the subgraph formed by B (note that isolated vertices count as a component in this scheme).

The remarkable fact about the random cluster model is that if B is drawn according to $\pi(B)$, and each resulting connected component is colored uniformly at random from $\{1, \ldots, Q\}$, then the result is a Q-state Potts model configuration with $p = 1 - e^{-1/(kT)}$. Note that the random cluster model defines a distribution even if Q is not integer; hence it is a nontrivial extension of the Potts model.

On the other hand, if we have a sample from the Potts model, we can obtain a sample from the random cluster model as follows. Given the coloring X, consider the set of edges $e = \{v, w\}$ with $X(v) = X(w)$ and retain each of them independently with probability p.

One widely used approach for drawing samples from the random cluster and Potts models is Swendsen-Wang [19]. Let Δ be the maximum degree of the graph. We will show two properties the Swendsen-Wang process in the next section. First, the Swendsen-Wang chain completely couples in polynomial time when p is smaller than $1/[3(\Delta - 1)]$. Under these conditions the chain is rapidly mixing, i.e., the chain need only be run for a polynomial amount of time before the distribution of the particle on the chain is very close to the stationary distribution. (This provides an independent proof of a recent result of Cooper and Frieze [3].) Second, we show that is is possible to determine algorithmically when complete coupling occurs. This allows us to use a procedure known as coupling from the past (described in Section 6) to use Swendsen-Wang to take samples exactly drawn from the stationary distribution. Finally, we also show that Swendsen-Wang couples quickly when p is extremely close to 1.

The Swendsen-Wang approach to sampling from the Potts model works as follows. Given a configuration from the random cluster model it generates a random state from the Potts model. From the Potts model state, Swendsen-Wang then generates a random state from the random cluster model. For a subset of edges A, let $\mathcal{C}_A$ denote the set of connected components of (V, A). Given an ordering of the vertices from 1 through $|V|$, and for each component C, let $\min_v C$ denote the lowest numbered vertex in component C. We say

$a \in_U A$ to mean choose a uniformly at random from the set A. Using this notation, one step of the Swendsen-Wang chain may be written as follows.

Swendsen-Wang Step: (A, p, Q)

1. **For** each edge e of the graph choose $U(e) \in_U [0, 1]$
2. **For** each node v of the graph choose $c(v) \in_U \{1, \ldots, Q\}$
3. **Choose** a uniformly random ordering of the vertices $\{1, \ldots, |V|\}$
4. **For** all $C \in \mathcal{C}_A$ set $X(w) \leftarrow c(\min_v C)$ for all $w \in C$
5. **Set** $A \leftarrow \{e = \{v, w\} : X(v) = X(w)\} \cap \{e : U(e) \leq p\}$

All random choices lie in lines 1, 2, and 3. Line 4 converts the random cluster state to a Potts state and line 5 converts the Potts state back to a random cluster state. Since any state can move to any other state in one step, this chain is ergodic.

The total variation distance is often used to measure how far away two distributions are from one another. Let p and π be two distributions which put probability mass on a finite set. The total variation distance between them is

$$|p - \pi|_{TV} = \frac{1}{2} \sum_x |p(x) - \pi(x)|.$$

If π is the unique stationary distribution of a Markov chain, and $p_{x,t}$ is the distribution of a particle which started in state x and was run forward for t time steps, then $|p_{x,t} - \pi|_{TV}$ goes to 0 as t goes to infinity.

Mixing time is a measure of how fast $p_{x,t}$ converges to π. Let $\tau_{TV}(\epsilon)$ be the smallest t for which $|p_{x,t} - \pi|_{TV} \leq \epsilon$ for all possible starting states x. Then $\tau(\epsilon)$ is the mixing time of the chain. A chain is rapidly mixing if $\tau_{TV}(\epsilon)$ is bounded above by a polynomial in n and $\ln(1/\epsilon)$, where n is a variable that parameterizes the size of the chain.

While not completely understood, a few results are known about the mixing time of the Swendsen-Wang chain. Gore and Jerrum [6] showed that this process is not rapidly mixing in the case of the complete graph. Recently, Cooper and Frieze showed that the chain is not rapidly mixing for a random graph; however, they also gave a positive result that proved the chain to be rapidly mixing for some p. Our result (independently derived) also proves that Swendsen-Wang is rapidly mixing over similar parameters using a different method. In fact, we will show that not only is Swendsen-Wang rapidly mixing, but it is completely coupling over this range of parameters.

3. BOUNDING CHAINS

Bounding chains are a means for detecting complete coupling in a Markov chain. In a complete coupling, a Markov chain step is given as a random function $f : \Omega \to \Omega$. At each time t, we choose function f_t at random from a family of functions and let $X_{t+1} = f_t(X_t)$. The distribution on the functions must satisfy the transition probabilities of the Markov chain, that is, we must have $P(f(X_t) = x) = P(X_{t+1} = x | X_t)$.

Note that the above description of the Swendsen-Wang chain is a complete coupling.

Once the uniform [0, 1] random variables are chosen for each of the edges and the uniform $\{1, \ldots, Q\}$ colors are chosen for each of the nodes, we have defined the function f. Lines 4 and 5 finish the computation of $f(x)$ for the starting coloring x.

Now let F_t be $f_0 \circ f_1 \circ \cdots \circ f_t$. Then $F_t(X_0)$ is equal to X_t. Furthermore, if $F_t(X_0) = F_t(Y_0)$ for some t, then $F_{t'}(X_0) = F_{t'}(Y_0)$ for all $t' \geq t$. A chain with this property is said to be coupling.

We say that a chain has *completely coupled at time t* if F_t is a constant, that is, $F_t(\Omega) = \{x\}$. One way to check if the Swendsen-Wang chain has completely coupled at time t is to do a brute force check that $F_t(x)$ has the same value for every $x \in \Omega$. Since $|\Omega| = 2^{|E|}$, this is not an efficient method. Instead we will utilize the idea of a bounding chain to keep track of how close the chain is to completely coupling.

The benefits of having a means for detecting complete coupling are profound. First, complete coupling is stronger than a simpler coupling. Doeblin [4] and Aldous [1] showed for simple couplings that the total variation distance is bounded above by the probability that a simple coupling has not occurred. Since complete coupling is stronger than simple (pairwise) coupling, this result implies that $|p_t - \pi|_{TV} \leq P(F_t(\Omega) = \{x\})$, giving us a bound on the mixing time of the chain.

Even better, the idea of coupling from the past (CFTP) of Propp and Wilson [16] may be used to generate samples which are exactly drawn from the stationary distribution of the Markov chain. This is true even when the mixing time of the chain is unknown. Although we shall prove that for certain values of the parameter T the Swendsen-Wang chain is rapidly mixing, CFTP gives a way to sample effectively even when the temperature value is outsides these values.

Bounding chains were introduced independently in [9] and [8] as a generalization of antimonotonicity [7], here we give a generalization of the idea. First we set up the type of Markov chains for which bounding chains will be useful. For any set S the state space for chain M must be a subset of the set of c-tuples $(x_1, x_2, \ldots, x_c)$, where each $x_i \subseteq S$. Most sample spaces of interest fall into this category; for instance, in the random cluster model, S is the edges of the graph, $c = 1$, and A is the set x_1. The state space for the random cluster model is exactly the set of subsets of edges.

We could also model the Ising model with $c = 1$, but it is more convenient to use $c = 2$ and set up two subsets one of which corresponds to spin up nodes and one of which corresponds to spin down nodes. That is, we write the state space as partitions of the vertex set of the form $(x_{spin\ up}, x_{spin\ down})$. In fact, any state space which is a coloring of a set S may be represented in this fashion. Given a coloring x, just set $x_i = \{s : x(s) = i\}$.

The advantage of this formulation over simple colorings of S is twofold. First, it easily allows for state spaces where a point can receive more than one color. Second, there is no difficulty dealing with examples where S is countably or uncountably infinite, for instance, point processes in $\mathbf{R}^d$.

In the original chain, each element of the state space is the form $(x_1, \ldots, x_c)$, where $x_i \subseteq S$. To be a bounding chain, the chain M_B must have a state space Ω_{M_B}, that is, a subset of all c-tuples of the form $((y_1, z_1), (y_2, z_2), \ldots, (y_c, z_c))$, where $y_i, z_i \subseteq S$, and y_i and z_i are disjoint. Intuitively, y_i will keep track of points that we know are in the original chain, while z_i will keep track of points that are possibly in the original chain. More precisely,

Definition. *M_B is a bounding chain for M if there exists a coupling between M and M_B so that, for the coupled process $((X_1, \ldots, X_c), ((Y_1, Z_1), \ldots, (Y_c, Z_c))$,*

$$(\forall i)(Y_{i,t} \subseteq X_{i,t} \subseteq Y_{i,t} \cup Z_{i,t}) \Rightarrow (\forall i)(Y_{i,t+1} \subseteq X_{i,t+1} \subseteq Y_{i,t+1} \cup Z_{i,t+1}).$$

In effect, the sets Y_i and $Y_i \cup Z_i$ "sandwich" X_i at each time step. A simple induction tell us that if $Y_{i,0} \subseteq X_{i,0} \subseteq Y_{i,0} \cup Z_{i,0}$ for all i, and when we run the bounding chain according to this coupling, then $Y_{i,t} \subseteq X_{i,t} \subseteq Y_{i,t} \cup Z_{i,t}$ for all i at all times $t \geq 0$. Setting $Y_{i,0} = \varnothing$ and $Z_{i,0} = S$ for all i will meet our starting condition for any $X \in \Omega_M$. In terms of a complete coupling for M, this implies that $Y_t \subseteq F_t(\Omega) \subseteq Y_t \cup Z_t$ for all t. Therefore, if $Z_{i,t} = \varnothing$ for all i, then $F_t(\Omega)$ is a constant, and complete coupling has occurred.

A straightforward way to design a bounding chain M_B given a complete coupling on M is as follows. Suppose M_B at time t is in state $((Y_{1,t}, Z_{1,t}), \ldots, (Y_{c,t}, Z_{c,t}))$. Then choose f_t for M according to the complete coupling. Let $A = \{X_t \in \Omega_M : (\forall i) Y_{i,t} \subseteq X_{i,t} \subseteq Y_{i,t} \cup Z_{i,t}\}$, and $A' = f_t(A)$. Finally, set $Y_{i,t+1} = \cap_{x \in A'} x_i$ and $Z_{i,t+1} = \{\cup_{x \in A'} x_i\} \backslash Y_{i,t+1}$. It is easy to see that this gives a bounding chain. Moreover, a bounding chain constructed in this fashion will have the nice property that if $Z_{i,t} = \varnothing$ for all i at some time t; then $X_t = Y_t$ for all $t' \geq t$.

Utilizing this method of construction, one bounding chain for Swendsen-Wang is given below. To couple the bounding chain with the original Swendsen-Wang chain, we just use the same choices for $U(e)$, $c(v)$, and the random ordering of vertices. We begin with edge sets Y and Z for the bounding chain. Say that $C \sim v$ if component C and node v are connected using edges in $Y \cup Z$. Let $\mathcal{C}_Y$ be the set of connected components using only edges in Y, and C_v be the component in $\mathcal{C}_Y$ which contains v. We can also turn any ordering of vertices into an ordering on $\mathcal{C}_Y$ by saying that $C_{Y_1} < C_{Y_2}$ if and only if $\min_v C_{Y_1} < \min_v C_{Y_2}$. We let $G(v)$ denote the set of possible colors for node v in the intermediate stage of the algorithm where nodes are assigned colors.

Swendsen-Wang Bounding Chain Step: (Y, Z, p, Q)

1. **For** each edge e set $U(e) \in_U [0, 1]$
2. **For** each node v set $c(v) \in_U \{1, \ldots, Q\}$
3. **Choose** a uniformly random ordering of the vertices $\{1, \ldots, |V|\}$
4. **For** all v set $G(v) = \cup_{C \leq C_v, C \sim v} c(\min_v C)$
5. **Let** $Y \leftarrow Y \cup \{\{v, w\} \in E : C_v = C_w\}$
6. **Let** $Y \leftarrow Y \cup \{\{v, w\} \in E : G(v) = G(w), |G(v)| = |G(w)| = 1\}$
7. **Let** $Z \leftarrow \{\{v, w\} \in E : |G(v) \cap G(w)| \geq 1\} \backslash Y$
8. **Let** $Y \leftarrow Y \cap \{e : U(e) \leq p\}$
9. **Let** $Z \leftarrow Z \cap \{e : U(e) \leq p\}$

Lines 1, 2, and 3 make random choices in exactly the same way as the original chain. Suppose that C_1 and C_2 in $\mathcal{C}_Y$ are connected by an edge in Z, and that $\min_v C_1 < \min_v C_2$. Then the nodes in C_1 never receive the color of $\min_v C_2$, but the nodes in C_2 might receive the color of $\min_v C_1$. Line 4 keeps track of all the possible colors that a given node might possibly receive.

Line 5 computes those edges whose endpoints both lie in the same component (with respect to Y) and makes sure that those edges are added to Y. This, of course, does not change the nodes of the connected components of Y.

Now, the set $G(v)$ holds all the colors that a node v could possibly receive. This includes the color assigned component v_C as well as colors assigned components $C < v_C$ that are connected to v by edges in Z. This is accomplished in line 8.

Line 6 adds edges to Y based on the coloring G. Certainly any edge that was in Y before is still in Y, but now we must also include any edge whose endpoints both have the same color. Unlike the original chain, where $X(v)$ was always a single color, here $G(v)$ is a set of colors. The only way that we can be sure that $G(v)$ and $G(w)$ have the same color is if $G(v) = G(w) = \{c\}$ for some color c, which is equivalent to the condition checked for in line 6.

In line 7 we look at new edges that might possibly be in the random cluster set. These are edges whose color sets $G(v)$ and $G(w)$ overlap. Of course, if an edge is in Y it cannot be in Z, so we remove from Z all the edges of Y. Finally lines 8 and 9 implement the edge removal phase. If $U(e) > p$, that edge must be removed whether in set Y or in set Z.

4. BOUNDING CHAIN RUNNING TIME

Our main theorem concerns the running time of this procedure, the time needed for all of the uncertainty to disappear, and Z to reach $\varnothing$.

Theorem 4.1. *Set*

$$\gamma = p\left[\frac{\Delta - 2}{1 - p(\Delta - 1)} + 2\right].$$

If $\gamma < 1$, *the Swendsen-Wang bounding chain will have detected complete coupling by time* $-\log_\gamma 2m$ *with probability at least* 1/2.

Proof. Imagine that in the bounding chain step we insert a line between 8 and 9 that reads: set $Z' \leftarrow Z$. Then Z'_t is also a process that records the subset of edges before the edge retention step. Clearly, if $Z'_t = \varnothing$, then $Z_t = \varnothing$ as well. Similarly, we define Y'_t to be the edges of Y_t remaining before the edge retention step.

We will show that $|Z'_t|$ decreases on average at each step. Consider how an edge $e = \{v, w\}$ in Z'_t leads to an edge in Z'_{t+1}. Edges in Z'_t lead to edges in Z_t which lead to nodes with $|G(i)| > 1$ in the next step which lead to edges in Z'_{t+1}. First, to be in Z_t, the edge must survive the edge retention phase in line 9. Retention occurs with probability p. After this the step terminates, with $Z_t \leftarrow Z$ and $Y_t \leftarrow Y$.

We then begin the next step of the chain with $Y \leftarrow Y_t$ and $Z \leftarrow Z_t$. Let us examine the probability that $|G(i)| > 1$ for a given node i. For each i, there is a collection of components $C_1, \ldots, C_{r(i)}$ in $\mathcal{C}_Y$ connected together by edges in Z, and one of these components contains i. Say, without loss of generality, that $i \in C_1$. Now, for each $a \in \{0, \ldots, r(i) - 1\}$, there is a $1/r(i)$ chance that C_1 is larger than exactly a of these components, with respect to the randomly chosen total ordering.

Note that only components lower than C_1 in the partial ordering can possibly make $|G(i)| > 1$, and only if not all of them choose the same color as C_1. The chance that they all choose the same color for a given value of a is just $(1/Q)^a$. Hence

$$P(|G(i)| > 1 | r(i)) = 1 - \left(\frac{1}{r(i)}\right)\left[\sum_{a=0}^{r(i)-1}\left(\frac{1}{Q}\right)^a\right] = 1 - \frac{1}{r(i)}\left[\frac{1-(1/Q)^{r(i)}}{1-1/Q}\right].$$

Now let e_1 be an edge in Z'_{t+1}. Then e_1 must be adjacent to a node i such that $|G(i)| > 1$. In other words, e_1 is adjacent to a node i that is part of a set of connected components in $\mathcal{C}_Y$, call them $C_1, \ldots, C_{r(i)}$, that are connected via a_{e_1} edges in Z_t. Then, for each such edge e_2 in Z_t that is connected to e_1 by edges in $Y_t \cup Z_t$, we set $g(e_1, e_2) = 1/a_{e_1}$. If e_1 and e_2 are not connected by edges in $Y_t \cup Z_t$, or if $e_1 \in Y'_t$, set $g(e_1, e_2) = 0$.

Using this, note that

$$|Z'_{t+1}| \leq \sum_{e_1 \in E} \sum_{e_2 \in Z_t} g(e_1, e_2)$$

(this would be equality except for the case that edge e_1 connects two components that were previously disconnected in $Y_t \cup Z_t$). All our edge sets are finite, so we can write:

$$E[|Z'_{t+1}| \,|Y'_t, Z'_t] = E\left[\sum_{e_2 \in Z_t} \sum_{e_1 \in E} g(e_1, e_2) | Y'_t, Z'_t\right].$$

We now upper bound $\sum_{e_1} E[\, g(e_1, e_2)|Y'_t, Z'_t]$ by looking at the expected size of the component C connected to edge e_2 using edges in either Y_t or Z_t. Let E_C be the set of edges adjacent to nodes in C that are not in Y'_t; these are the candidates for edges e_1 with $g(e_1, e_2) > 0$. For a node $v \in C$, let $d(v)$ denote the number of edges adjacent to v in E_C. Further, let r_C be the number of components of $\mathcal{C}_Y$ in C. Since these components are connected by edges in Z_t, we know for any relevant e_1 that $a_{e_1} \geq r_C - 1$.

Let Δ be the maximum degree in the graph. There are at most $\Delta|C|$ edges leaving nodes in C. Of these edges, we are counting at least $|C| - 1$ of them twice, and so there are at most $\Delta|C| - (|C| - 1)$ edges touching nodes in C. Recall that if we only include edges in Y_t, C is broken into at most r_c components. Hence $|C| - 1 - r_c$ of these edges must be in Y'_t, and we should not include them in our count of edges in E_C at all. Therefore, we upper bound the number of edges in E_C by $\Delta|C| - (|C| - 1) - (|C| - 1 - r_C) = |C|(\Delta - 2) + r_C + 2$. Let $i \sim e_2$ if node i and edge e_2 are connected via edges in $Z'_t \cup Y'_t$. Then for a given $e_2 \in Z'_t$ that survives the edge retention step,

$$\begin{aligned}
\sum_{e_1 \in E} E[g(e_1, e_2)|Y'_t, Z'_t] &\leq E\left[\sum_{i \in V} d(i)\, \frac{1_{|G(i)|>1} 1_{i \sim e_2}}{r_C - 1} \,|Y'_t, Z'_t\right] \\
&= E\left[E\left[\sum_{i \in V} d(i)\, \frac{1_{|G(i)|>1} 1_{i \sim e_2}}{r_C - 1} \,|Z_t, Y_t\right]|Y'_t, Z'_t\right] \\
&= E\sum_{i \in C} d(i)\, \frac{P(|G(i)| > 1 | Z_t, Y_t)}{r_C - 1} \,|Y'_t, Z'_t\Bigg] \\
&= E\left[\sum_{i \in C} d(i)\, \frac{1}{r_C - 1}\left[1 - \frac{1}{r_C}\left[\frac{1-(1/Q)^{r_C}}{1-1/Q}\right]\right]|Y'_t, Z'_t\right],
\end{aligned}$$

and so our final upper bound for $\sum_{e_1 \in E} E[\, g(e_1, e_2) | Y_t', Z_t']$ is

$$E\left[(|C|(\Delta - 2) + r_C + 2)\left(\frac{1}{r_C - 1}\right)\left[1 - \frac{1}{r_C}\left[\frac{1 - (1/Q)^{r_C}}{1 - 1/Q} \right] \right] \middle| Y_t', Z_t' \right].$$

This last expectation is very difficult to compute directly. Instead, we simply maximize the expression in r_C and then bound $E[|C|]$. Fortunately, the $(r_C + 2)\left(\frac{1}{r_C - 1}\right)[1 - \frac{1}{r_C}\left[\frac{1 - (1/Q)^{r_C}}{1 - 1/Q}\right]]$ term is at most $(r_C + 2)/r_C$ for very large Q, and so is at most 2, given that $r_C \geq 2$. The $\left(\frac{1}{r_C - 1}\right)\left[1 - \frac{1}{r_C}\left[\frac{1 - (1/Q)^{r_C}}{1 - 1/Q}\right]\right]$ term is at most $1 - 1/r_C$, and so at most 1/2. These estimates are worse case bounds for all Q; for a specific value of Q it is possible to do better.

Now we turn to $E[|C|]$. Recall that each edge at the retention phase is kept with probability p. The process where each edge of a graph is kept with probability p is also known as bond percolation. Finding $E[|C|]$ exactly is very difficult for arbitrary general graphs; however, we can get an upper bound by treating it as a branching process. Each node is considered to have from 0 to $\Delta - 1$ "children," where each child is born with probability p. Each child then begins a new branching process. Therefore, the size of a single branching process satisfies the recursion $E[|BP|] \leq 1 + p(\Delta - 1)E[|BP|]$, so $E[|BP|] \leq 1/(1 - p(\Delta - 1))$. We can think of C as a branching process begun from either endpoint of the edge e_2 in Z_t', and so $E[|C|] \leq 2/(1 - p(\Delta - 1))$.

The edge in Z_t' survives the retention phase with probability p, and so

$$E[|Z_{t+1}'| \,|\, |Z_t'|] \leq |Z_t'| p\left[\frac{\Delta - 2}{1 - p(\Delta - 1)} + 2 \right] = \gamma |Z_t'|.$$

An induction gives us that $E[|Z_t'|] \leq Z_0' \gamma^t \leq m\gamma^t$. After $-\log_\gamma 2m$ time, we have that $E[|Z_t'|] < 1/2$, and so by Markov's Inequality, $P[|Z_t'| > 0] < 1/2$, completing the proof. ∎

As noted in the proof, we may do slightly better for specific values of Q. We now compare this with the mixing time result of Cooper and Frieze. Below is a table listing some values for the critical point p where our algorithm is guaranteed to run in polynomial time. Our analysis is slightly stronger, and is best when Q is as small as possible. We give two rows of numbers for our method, one when $Q = 2$, and one which is valid for all values of Q. Naturally for $Q \geq 3$, the actual number will fall somewhere in between these two extremes.

Δ	2	3	4	5	6	7
Cooper/Frieze	0.416	0.209	0.136	0.100	0.079	0.065
$Q = 2$	0.930	0.356	0.231	0.171	0.136	0.113
Any Q	0.500	0.250	0.167	0.125	0.100	0.083

We wish to reiterate, however, that our algorithm gives more than just a mixing time, it gives a complete coupler. This gives a method for estimating the mixing time even when the mixing time cannot be analyzed directly, and through coupling from the past gives a means for obtaining samples drawn exactly from the stationary distribution.

4.1. Low Temperatures

When the temperature is low, the Gibbs distribution for the ferromagnetic Ising model is dominated by states which have most of the nodes colored identically. In contrast, the antiferromagnetic Ising model at low temperatures is dominated by states which represent the maximum size Q-way cut in the graph. In general, finding the max Q-way cut is an NP-complete problem, and so it is unlikely that a method for sampling from the antiferromagnetic Ising model at arbitrarily low temperatures will be found.

Since Swendsen-Wang only samples from the ferromagnetic Ising model, no such restrictions apply, and in fact we may show that it converges when the temperature is extremely low, that is, when the parameter p approaches 1. In the version of Swendsen-Wang above, we labeled the vertices randomly, then colored in order of label. We could have instead picked an arbitrary ordering and colored in that order. We use that idea to prove the following theorem, which applies to all graphs.

Theorem 4.2. *Suppose that* $p \geq 1 - 1/(mQ)$. *Then Swendsen-Wang couples completely with probability at least* 1/2 *in time* $2(mQ)^2$.

The probability of throwing away an edge is very low here, less than $1/(mQ)$. But through coloring, an edge turns to $\{1\}$ when its endpoints are colored the same way. This good event happens with probability $1/Q$. Hence the expected change in the number of edges colored $\{1\}$ (call it $|A_t|$) will be $-|A_t|/(mQ) + 1/Q$, which is positive when $|A_t| < m$. Hence on average more edges become colored until the entire edge set is labeled $\{1\}$, at which point coupling has occurred.

In the case of the complete graph (also known as the Curie-Weiss model), we may make a much stronger statement. Here p does not have to be very large at all before a component which spans the entire graph becomes a real possibility.

Theorem 4.3. *On the complete graph, for* $p > 2\ln(3Qn)Q/n$, *the expected complete coupling time for the Swendsen-Wang chain is at most* $O(Qn)$.

Proof. First note that by the pigeonhole principle the largest component will always have size bounded below by $\lceil n/Q \rceil$. We concentrate on the size of the largest component, and show that with constant probability it grows to encompass the entire graph in polynomial time.

Let A_t denote the set of nodes in this component at time t. The size of A_t can expand and shrink over the course of an iteration of Swendsen-Wang. In the coloring phase, if a node outside of the large component is colored with the same color as the large component, then it becomes part of the large component. Therefore, during this phase, the expected change in the size of A_t is $1/Q$, since that is the probability of a node being added to the large component.

In the pruning phase, removal of edges can shrink the size of A_t. The large component is a complete graph on A_t nodes, where each edge is kept with probability p, so work similar in flavor to results on random graphs (see [2]) can be used. We wish to find the expected size of the large component after pruning.

Pick any node v in A_t. Then $|A_{t+1}|$ is certainly at least as large as the number of nodes in A_t connected to v. Pick any $w \neq v$. The probability that v is not connected to w is just the probability that at least one cut between v and w gets pruned. This is bounded above

by the sum of the probability over all cuts between v and w that the cut gets pruned. Say a cut separating v and w contains $i + 1$ nodes on the side with v and $|A_t| - (i + 1)$ nodes on the w side (there are at most $\binom{|A_t|-2}{i}$ such cuts). Let $D_{v,w}$ denote the event that nodes v and w are disconnected. Then there are $(i + 1)(|A_t| - i - 1)$ edges to prune in order for this cut to be removed, so

$$P(D_{v,w}) \leq \sum_{i=0}^{|A_t|-2} \binom{|A_t| - 2}{i} (1 - p)^{(i+1)(|A_t-2-i+1|)}.$$

Let $m(i) = \min\{i, |A_t| - 2 - i\}$. Then we bound $\binom{|A_t|-2}{i}$ by $(|A_t| - 2)^{m(i)}$ and $(1 - p)$ by e^{-p}. Then

$$\begin{aligned} P(D_{v,w}) &\leq \sum_{i=0}^{|A_t|-2} (|A_t| - 2)^{m(i)} \exp\left\{-\frac{2 \ln(2n)}{n/Q} (m(i) + 1)\left(\frac{|A_t| - 2}{2} + 1\right)\right\} \\ &\leq \sum_{i=0}^{|A_t|-2} \left(\frac{|A_t| - 2}{4nQ}\right)^{m(i)} \frac{1}{(3nQ)^{|A_t|Q/n}} \\ &\leq \frac{1}{(3nQ)^{|A_t|Q/n}}. \end{aligned}$$

There are $|A_t| - 1$ nodes that could be connected to w, so including w we have

$$E[|A_{t+1}||A_t] \geq 1 + (|A_t| - 1)\left(1 - \frac{1}{3Qn}\right) + \frac{1}{3Q}.$$

Hence from Wald's Lemma we have that $P(|A_t| = n) > 1/2$ for $t \geq 3n/Q$. ∎

Therefore, on the complete graph, Swendsen-Wang runs quickly when $p < 1/(3n)$ or $p > 2 \ln(3Qn)Q/n$, but from [6] we know Swendsen-Wang runs slowly when $p = 2\frac{Q-1}{Q-2}\frac{\ln(Q-1)}{n}$.

4.2. Worse Case Running Time

We now compute a worst case bound on the expected running time of this algorithm when the parameter p is neither sufficiently high nor low enough to allow a proof of complete coupling in polynomial time. Our time bound will be exponential, since our only goal here is to show that our algorithm is never worse than the exponential running time of a brute force approach to the problem. Let K_t be the number of components connected by edges in Y_t in the graph at time t. Then with probability $(1/Q)^{K_t}$, each of these components are colored the same way, and the state immediately becomes known. Since $K_t \leq n$, we have that the chain will have coupled with probability at least 1/2 after $Q^n \ln 2$ time steps, regardless of p or Δ. Note that the number of states is Q^n, so Swendsen-Wang is never worse (in order) than the time needed for a direct sample to be taken using the trivial method.

5. SWENDSEN-WANG ON TREES AND CYCLES

Physically, the Ising model does not have a phase transition in the case of a cycle or a tree [17]. This would indicate that algorithms of interest for sampling from the Ising model should run efficiently at all temperatures on these graphs. Cooper and Frieze [3] did in fact show Swendsen-Wang to be rapidly mixing on these graphs. We now modify the bounding chain slightly so that samples can be taken in polynomial time regardless of p, and improve upon the Cooper/Frieze mixing time for cycles.

The bounding chain step again starts with a state Y, Z. We will change the recoloring phase to eliminate as many edges from Z as possible. We do not pick colors directly. Instead we just record for each edge $e = \{v, w\}$ whether or not the component connected to v receives the same color as the component connected to w. With probability $1/Q$, the endpoints of the edge will receive the same color, regardless of whether the edge is in Z or $E\backslash(Y \cup Z)$. This leads to the following bounding chain.

Swendsen-Wang Bounding Chain Step for Trees: (Y, Z, p, Q)

1. **For** each edge e set $U_1(e) \in_U [0, 1]$
2. **For** each edge e set $U_2(e) \in_U [0, 1]$
3. **Let** $Y' \leftarrow \{e : U_2(e) \leq 1/Q\}$, **Let** $Y \leftarrow Y \cup Y'$, **Let** $Z \leftarrow Z\backslash Y'$
4. **Let** $Y \leftarrow Y \cap \{e : U(e) \leq p\}$
5. **Let** $Z \leftarrow Z \cap \{e : U(e) \leq p\}$

This approach does not work with general graphs because of the presence of cycles. For example, supposed that vertices v_1, v_2, and v_3 are all connected by edges in Z, and so form a three cycle. Then, if we say v_1 and v_2 are colored the same way, v_2 and v_3 are colored the same way, and v_2 and v_3 are colored differently, this gives a contradiction. Since there are no cycles in trees, contradictions of this sort cannot occur.

Theorem 5.1. *On trees, let T be the complete coupling time for Swendsen-Wang. Then, for all a,*

$$P\left(T > -(a + \ln n)/\ln\left[\left(1 - \frac{1}{Q}\right)p\right]\right) < e^{-a}.$$

Remark. This implies that the mixing time of the Swendsen-Wang process on trees for fixed Q and p is $O(\ln n)$, an improvement over the $O(n)$ result of Cooper and Frieze [3]. Also note that, for all p, $-1/\ln[(1 - \frac{1}{Q})p] \leq Q$, so this factor is never very large.

Proof. Note that edges in Z are only removed, never created, during the course of a bounding chain run. Edges are removed from Z during the recoloring phase with probability $1/Q$ and during the retention phase with probability $1 - p$. Therefore,

$$E[|Z_{t+1}||Z_t|] = \left(1 - \frac{1}{Q}\right)p|Z_t|,$$

We start with $n - 1$ edges, and so after t steps, $E[|Z_t|] = [(1 - (1/Q))p]^t(n - 1)$. From Markov's Inequality, we know that $P(|Z_t| \geq 1) \leq E[|Z_t|]$, and the theorem follows. ∎

For cycles, we must use several different bounding chains to prove the following result.

Theorem 5.2. *On cycles, the Swendsen-Wang process completely couples in* $O(\ln n)$ *time. That is, letting* τ *be the first time the bounding chain completely couples, for all a there exist constants* C_1 *and* C_2 *such that*

$$P\left(\tau > \frac{\ln n}{\ln(1/p)} + a\right) \leq C_1 e^{-a/C_2}.$$

The chain begins by treating the cycle as though it is a tree; this is phase 1 of the bounding chain. Suppose the edge $\{w, v\}$ is not in Y or Z. Then the graph with $E' = E\backslash\{\{w, v\}\}$ is a tree, and we can run the bounding chain step for trees on it. Effectively, this means that we start at node v and move clockwise around the cycle, at each node rolling to see if it is the same color as the previous node or not. The difference between the cycle and the tree is that when we get back to node w, we do not know whether v and w are colored the same way or differently, so we must add $\{w, v\}$ to Z. We then run the edge retention step as before.

If all the edges of the graph are still in either Y or Z after the edge retention step, then nothing changes between this step and the next. So the worst that can happen is that after the edge retention phase the number of edges in Z increases by 1. Hence

$$E[|Z_{t+1}| \,|\, |Z_t|] \leq p|Z_t| + 1.$$

Since $|Z_0| = m$, after t steps an induction gives us $E[|Z_t|] \leq p^t m + \frac{1}{1-p}$. Let $D_t^1 = \max\{|Z_t| - \frac{1}{1-p}, 0\}$. Then $E[D_t^1] \leq p^t m$ and since $D_t^1 \geq 0$ we may use Markov's Inequality to say that after $(\ln m + a_1)/\ln(1/p)$ steps, the probability that $D_t^1 > 0$ is at most e^{-a_1}. Let τ_1 be the first time that $D_t^1 = 0$, that is, $|Z_t| \leq \frac{1}{1-p}$. The bounding chain then changes approach in phase 2.

At this point we have very few edges in Z_{τ_1}, but they are scattered all over the cycle. We would like them to be together in the cycle so that they form one arc of the cycle (we call a set of edges in a cycle an arc if they form one connected component.) Let e be an edge in Z_{τ_1}. This will be the first edge in $Z_{\tau_1}^{arc}$. Assume without loss of generality that $e = \{1, 2\}$. Then we take a step in the Swendsen-Wang bounding chain as before, but we let our starting edge $\{V, V + 1\}$ be the edge with smallest value of V that is not in $Z_{\tau_1} \cup Y_{\tau_1}$. (It could have been dropped during the edge retention step or just been absent from earlier.) Then we add edges $\{2, 3\}, \ldots, \{V, V + 1\}$ to form $Z_{\tau_1+1}^{arc}$. (Note: we can always add any edge to Z and still have a bounding chain, just usually it isn't a wise move. Here it is.) Because of the edge retention step, the expected value of V is at most $1 + 1/(1 - p)$, so we expect to have added at most $1/(1 - p)$ edges to Z^{arc}. At the next step, we again look for the missing edge to run the bounding chain that increases the size of the arc as little as possible. After taking $t - \tau_1$ steps, the expected number of edges in Z_t^{arc} will be bounded above by $\frac{1}{1-p}(t - \tau_1)$.

While the arc is growing, other edges in Z are being removed through the edge

retention step. Let D_t^3 denote the number of such edges in Z_t that are not in Z_t^{arc}. Ignoring the fact that the Z^{arc} is growing and swallowing edges in D_t^3, we know that the edge retention step is pruning edges, and $E[D_t^3] \leq p^{t-\tau_1}\frac{1}{1-p}$. After $(\ln \frac{1}{1-p} + a_2)/\ln(1/p)$ steps, the probability that $D_t^3 > 0$ is at most e^{-a_2}. Let τ_2 be the first time that $D_t^3 = 0$.

We now begin phase 3, where we try to get the size of Z^{arc} to go to zero. Our main focus in phase 3 is on the number of components of any state bounded by the bounding state. In the Swendsen-Wang process where some edge is not in the state, each edge that is in the state can reduce the number of connected components by 1. Let $D_t^4 = m - |Y_t| - |Z_t|$ and $D_t^5 = m - |Y_t|$. Then, for any bounded state, the number of connected components lies in $[D_t^4, D_t^5]$. Note that the only special case to worry about is when all of the edges are either in $|Y_t|$ or $|Z_t|$. In this case $D_t^4 = 0$, however, the actual minimum number of connected components is 1. Still, D_t^4 does act as a lower bound in this situation, so this is not a problem. Note that $|Y_t| < m$; otherwise $|Z_t| = \varnothing$, so D_t^5 doesn't encounter this difficulty at all.

During phase 3 we abandon our tree bounding chain approach. Instead, at each step we assign to each component from 1 up to D_t^5 a color. Previously, we attached to each edge $e \in E$ a uniform [0, 1] random variable that determined whether or not it survived the edge retention step. Now we only attach uniforms to edges that are in the state. Any bounded state has from $|Y_t|$ to $|Y_t| + |Z_t|$ edges, so generating $|Y_t| + |Z_t|$ uniforms suffices. During the edge coloring step, the first D_t^4 values are the same for all states, and only affects edges in Y_t, so everything is the same up until this point. However, once past the first D_t^4 values, the next colors may collapse components. For instance, suppose $D_t^4 = 2$, $D_t^5 = 5$, $Q = 2$ and the random colors were 1, 2, 1, 1, 2. Then any state having 4 or 5 components would have the third and fourth component colored 1, thereby reducing the number of connected components by 1.

Each of the $D_t^5 - D_t^4$ components has a $1/Q$ chance of being eliminated, so $E[D_{t+1}^5 - D_{t+1}^4 | D_t^5 - D_t^4] \leq (1 - \frac{1}{Q})(D_t^5 - D_t^4)$.

At the same time that the number of components is coming together, the set Z_t is growing. As in phase 2, after each step the first gap (an edge not in Z_t or Y_t) clockwise from the arc Z_t gets added to Z_{t+1}. In fact, we add all the edges between Z_t and this gap to Z_t, removing them from Y_t if necessary. As we noted earlier, this grows Z_t by an expected value of $1/(1 - p)$.

Now for the edge retention step. This works in the opposite direction from the coloring stage, adding new components by removal of edges. Here the number of edges is states bounded by the bounding state go from $m - D_t^5$ up to $m - D_t^4$, so again there is a difference of $D_t^5 - D_t^4$ edges. When an edge is removed with probability $1 - p$ the number of components rises by 1, therefore $E[D_{t+1}^5 - D_{t+1}^4 | D_t^5 - D_t^4] \leq p(1 - \frac{1}{Q})(D_t^5 - D_t^4)$.

At time τ_2, we had $E[D_{\tau_2}^5 - D_{\tau_2}^4 | \tau_2, \tau_1] \leq \frac{1}{1-p}(\tau_2 - \tau_1)$. Therefore, after an induction, we have that $E[D_t^5 - D_t^4 | \tau_2, \tau_1] \leq \frac{1}{1-p}(\tau_2 - \tau_1)[p(1 - \frac{1}{Q})]^{t-\tau_2}$. In particular, taking $t_3 = -[\ln((\tau_2 - \tau_1)/(1 - p))/\ln(p(1 - 1/Q))]$ steps in phase 3 ensures that $D_t^5 \neq D_t^4$ with probability at most e^{-a_3}. Let τ_3 be the first time $D_t^5 = D_t^4$.

To call phase 3 a success, we require that $D_t^5 = D_t^4$, but we also need that $|Z_{\tau_3}|$ is small. Fortunately, $|Z_{\tau_3}|$ is stochastically bounded above by the sum of geometric random variables with mean $1/(1 - p)$, and so standard concentration results (see [15] for details) can be used to show that the probability that $|Z_{\tau_3}| > 2t_3/(1 - p)$ is at most $\exp(-t_3/[4(1 - p)])$.

We are now ready to enter the final phase, phase 4. In this phase, we maintain the

invariant that all bounded states have the same number of components, and work to eliminate Z_t.

Again at each step we choose a color for the set of components, and choose a uniform for each edge in the bounded state in order to determine if that edge is retained. Our situation at this stage is as follows. The edges in Z_t form a contiguous arc of the cycle. Start coloring with the known nodes and edges in Y_t. Consider the first edge $\{v, w\}$ in Z_t. If we roll to color the component connected to v the same color as the next component, then $\{v, w\}$ will always be an edge at the next time step, so it does not appear in Z_{t+1} but does appear in Y_{t+1}.

The edge retention step has a similar story. If the first unknown edge is rolled to remove, then it can be removed from Z_t with probability $1 - p$. Therefore, we expect to remove at least $1/Q$ edges due to coloring and $1 - p$ edges due to edge retention until Z_t is empty. We know that $E[Z_{\tau_3}] \leq (\tau_3 - \tau_1)/(1 - p)$, so we expect to take at most $\frac{\tau_3 - \tau 1}{1-p} \cdot \frac{1}{1-p+1/Q}$ steps before Z_t is empty. Taking $a_4 e \frac{\tau_3 - \tau_2}{1-p} \cdot \frac{1}{1-p+1/Q}$ steps insures that the probability that $Z_t \neq \varnothing$ is at most e^{-a_4} by the end of this stage. Let τ_4 be the first time that $Z_t = \varnothing$.

Putting everything together, we would like to bound the probability that τ_4 is less than the number of steps taken. For $i \in \{1, 2, 3, 4\}$, let E_i denote the event that phase i did not complete successfully. Then we have shown that $P(E_1) \leq e^{-a_1}$, $P(E_2|E_1^C) < e^{-a_2}$, $P(E_3|E_2^C, E_1^C) \leq e^{-a_3}$, and $P(E_4|E_3^C, E_2^C, E_1^C) < e^{-a_4}$, so $P(E_1^C E_2^C E_3^C E_4^C) \leq \sum_i e^{-a_i} \leq e^{-a}$ if we set $a_i = a + \ln 4$.

6. COUPLING FROM THE PAST

We have stated that bounding chains together with coupling from the past (CFTP) may be used to obtain exact samples from the stationary distribution of a Markov chain. CFTP was developed by Propp and Wilson [16], and we now briefly describe the procedure.

First a starting time is chosen, say $t = -1000$. The chain is run from t to 0. If the chain couples completely in that time, we return the value of the chain at time 0 and stop. Otherwise, we run the chain from $2t$ to t, and again see if the chain has coupled completely. If it has, we then run the chain from t to 0 *using the same moves as the first time we ran the chain* from t to 0. This means that at each step in the complete coupling, we must take the same values for our random choices as we did the first time. The state of the chain at time 0 is then our sample. If it has not coupled, then we run from $4t$ to $2t$, and so on. For this algorithm to be correct, it is necessary that the chain couple completely with probability 1 in finite time. We have showed that the Swendsen-Wang chain has this property for all p, and so CFTP may be applied.

This procedure returns a sample that is exactly from the stationary distribution of the Markov chain. Since we have shown that the chain couples in polynomial time with probability 1/2, the total running time of the CFTP procedure is also polynomial.

Utilizing this method has other benefits as well. A metric often used for measuring the closeness of two distributions is the total variation distance. When running the chain a long time, the chain must always be run for the mixing time multiplied by $\ln(1/\epsilon)$, where the distance from the actual distribution to the stationary distribution is bounded above by ϵ. Since CFTP draws exact samples ($\epsilon = 0$), this factor does not enter into the running time.

In addition, when running the chain and just taking the particle, there is no stopping early. Under coupling from the past, we stop whenever we have completely coupled, which may happen sooner than expected (our analysis was a worst-case scenario; the actual complete coupling time may be much lower).

There is one drawback to CFTP, however. The running time of the algorithm is now a random variable, and our polynomial bounds only give an expected running time. We can guarantee that the algorithm stops with probability $1 - \delta$, by adding a factor of $\ln(1/\delta)$ to the running time. However, our requirement for stopping (represented by δ) is often much less stringent than our requirement for accuracy ($\epsilon \ll \delta$), and so this factor is not as important as it might appear at first.

7. CONCLUSIONS

The idea of a bounding chain has proven useful in determining when a Markov chains has completely coupled. In [9] it was shown that the single-site heat bath chain for the Ising model completely couples when $p \leq \frac{2}{\Delta+1}$. This is clearly a much higher value of p than is theoretically known for Swendsen-Wang, but Swendsen-Wang appears to run much faster in practice, indicating that the chain most likely couples more quickly than our analysis indicates. Both Swendsen-Wang and the single-site heat bath chain have a critical value of p at which they do not completely couple quickly. The single-site heat bath does not completely couple quickly for any value of p higher than the critical value. In contrast, we have seen that Swendsen-Wang does completely couple quickly when p is very high.

There are other Markov chains for the Ising and Potts models which do not yet have running time analysis. The method of Sweeny [18] is known to be a complete coupling chain, but no bounds on how quickly it couples are known. The method of Wolff [20] is a variation on Swendsen-Wang, and it is unknown whether it may be run as a complete coupling chain or not. Hopefully bounding chains or similar techniques will be useful in analyzing these chains.

REFERENCES

[1] D. Aldous, Some inequalities for reversible markov chains, J Lond Math Soc 25(2) (1982), 561–576.

[2] B. Bollobás, Random graphs, Academic Press, London, 1985.

[3] C. Cooper and A. Frieze, Mixing properties of the Swendsen-Wang process in classes of graphs, Random Struct Alg 15(3–4) (1999), 242–261.

[4] W. Doeblin, Exposé de la théorie des chains simples constantes de Markov à un nombre fini d'états, Rev Math Union Interbalkanique 2 (1933), 77–105.

[5] C. M. Fortuin and P. W. Kasteleyn, On the random cluster model I: Introduction and relation to other models, Physica 57 (1972), 536–564.

[6] V. Gore and M. Jerrum, The Swendsen-Wang process does not always mix rapidly, J Stat Phys 97(1–2) (1999), 67–86.

[7] O. Häggström and K. Nelander, Exact sampling from antimonotone systems, Statist Neerlandica, 52 (1998), 360–380.

[8] O. Häggström and K. Nelander, On exact simulation from Markov random fields using coupling from the past, Scand J Stat 26(3) (1999), 395–411.

[9] M. L. Huber, Exact sampling and approximate counting techniques, Proc 30th Symp Theory of Computing, Dallas, TX, 1998, pp. 31–40.

[10] M. L. Huber, Exact sampling using Swendsen-Wang, Proc 10th Annu Symp Discrete Algorithms, Baltimore, MD, 1999, pp. 921–922.

[11] M. L. Huber, Perfect sampling with bounding chains, Ph.D. thesis, Cornell University, Ithaca, NY, 1999.

[12] M. L. Huber, A faster method for sampling independent sets, Proc ACM-SIAM Symp Discrete Algorithms, 2000, pp. 625–626.

[13] M. L. Huber and G. Reinert, Exact sampling from the Antivoter model, Proc Workshop Steiner's Method, 1998, to appear.

[14] M. Jerrum and A. Sinclair, Polynomial-time approximation algorithms for the Ising model, SIAM J Comput 22 (1993), 1087–1116.

[15] R. Motwani and P. Raghavan, Randomized algorithms, Cambridge University Press, Cambridge, 1995.

[16] J. G. Propp and D. B. Wilson, Exact sampling with coupled Markov chains and applications to statistical mechanics, Random Struct Alg 9(1–2) (1996), 223–252.

[17] B. Simon, The statistical mechanics of lattice gasses, Volume 1, Princeton University Press, Princeton, NJ, 1993.

[18] M. Sweeny, Monte Carlo study of weighted percolation clusters relevant to the Potts models, Phys Rev B 27(7) (1983), 4445–4455.

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

[20] U. Wolff, Collective Monte Carlo updating for spin systems, Phys Rev Lett 62(4) (1989), 361–364.