

Efficient Exact Sampling From the Ising Model Using Swendsen-Wang

Mark Huber*

School of Operations Research and Industrial Engineering
Cornell University
Ithaca, NY 14850
mhuber@orie.cornell.edu

October 17, 2000

Abstract

In Monte Carlo Markov Chain theory, a particle on a Markov chain is run for a 'long time' until the particle is nearly in the stationary distribution. Coupling from the past is a methodology which allows samples to be taken exactly from the stationary distribution of the chain, but which requires that we have a completely coupling chain. One chain often used for the ferromagnetic Ising model is the Swendsen-Wang chain. We give the first method for running Swendsen-Wang as a completely coupling chain, thereby giving a procedure for taking exact samples from the ferromagnetic Ising model using Swendsen-Wang. Moreover, this chain completely couples in polynomial time when a parameter of the chain is inversely proportional to the maximum degree of the graph, or when this parameter is close to 1. This has two implications. First, the Swendsen-Wang chain is rapidly mixing over this set of parameter values, a fact recently (and independently) shown by Cooper and Frieze using path coupling. Second, over this set of parameter values the exact sampling procedure will have an expected running time which is polynomial.

*This work was partially supported by an Office of Naval Research Fellowship and NSF grants CCR-9307391, CCR-9700029, and DMS-9505155, and ONR grant N00014-96-1-00500

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”. An example is the Ising model, where we wish to select a state at random according to the Gibbs distribution. Several Markov chains have this distribution as their stationary distribution. Here we show that for one of these chains which is in common use, the Swendsen-Wang chain [13], his technique only requires that we run the chain for a polynomial length of time. Note, however, that this technique only gives approximate samples from the stationary distribution, since there is always some bias from the choice of starting state. We give a new method for running this chain as a *complete coupling* chain, which allows us for the first time to take exact samples from the Gibbs distribution using the Swendsen-Wang chain. In addition, we show that when a parameter of the chain lies in a specified range, this procedure has a polynomial expected running time, which also implies that the chain is rapidly mixing over this set of parameter values.

The Ising model and its extensions have been extensively used by statistical physicists to model the behavior of magnetic materials and alloys. Simon [11] 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, either 0 or 1. Such an assignment for every node is called a configuration. The set of possible configurations is just $\{0, 1\}^n$. Let x be a configuration. Then x has energy $H(x)$, which is equal to the number of edges in the graph with endpoints labeled differently (the value of the cut defined by the configuration). 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 more weight to cases where neighboring nodes are colored alike—this 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 [9], 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 $\{1, \dots, Q\}$. When $Q = 2$ we of course just have the standard Ising model. Again $H(x)$ is just the number of edges in a configuration with differently 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 $G = (\{1, \dots, n\}, E)$. Then the random cluster model considers subsets of the edges. Let $B \subset E$. The probability assigned to B is just

$$q(B) = p^{|B|}(1 - p)^{|E|-|B|}Q^{C(B)}/Z$$

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 amazing thing that Fortuin and Kastelyn were able to show is that if B is drawn according to $q(B)$, and each resulting connected component is colored uniformly at random, 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.

One approach for drawing samples from the random cluster model is known as Swendsen-Wang, and is widely used for drawing samples from the Ising and Potts models. Let Δ be the maximum degree of the graph. We will show two properties of Swendsen-Wang in the next section. First, the Swendsen-Wang chain completely couples in polynomial time when p is smaller than $1/[3(\Delta - 1)]$. This implies that 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 [2].) Secondly, we show that it 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 4) 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 very close to 1.

2 Swendsen-Wang

The Swendsen-Wang approach to sampling from the Potts model rests upon the relationship to the random cluster model [13]. Swendsen-Wang has two steps. In the first step, we have an element of the Potts model, x . Let A be the set of edges whose endpoints are colored the same way in x . Now perform a pass through A removing each edge with probability $p = 1 - e^{-1/(kT)}$. This leaves a new set of edges A' . In the second step, we construct a new coloring from B' by coloring each connected component of B' uniformly at random. We then repeat the process.

For a subset of edges A , let \mathcal{C}_A denote the set of connected components of (V, A) . Number the vertices 1 through $|V|$, and for each component C , let C_v denote the lowest numbered vertex in the component. Using this notation, one step of the Swendsen-Wang chain may be written as follows.

Swendsen-Wang Step
Let $A = \{\{v, w\} \in E : X(v) = X(w)\}$
For each edge e set $U(e) \in_U [0, 1]$
For each node v set $k(v) \in_U \{1, \dots, Q\}$
For each edge $e \in A$
 If $U(e) < 1 - p$, set $A = A \setminus \{e\}$
For all $C \in \mathcal{C}_A$
 Set $X(w) = k(C_v)$ for all $w \in C$

Since any state can move to any other state in one step, this chain is ergodic. Showing that it has the right stationary distribution is not much more difficult (see [4]).

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 $\tau(\epsilon)$ 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 $\tau(\epsilon)$ is bounded above by a polynomial in n and $\ln(1/\epsilon)$, where n is a variable that parameterizes the size of the chain.

Two 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, and gave a positive result that said that the chain is 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.

Complete Coupling In a complete coupling, a Markov chain step is given as a random function $f : \Omega \rightarrow \Omega$. Then at time t , we have $X_{t+1} = f_t(X_t)$. The function f is chosen at random and 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, once the uniform $[0,1]$ random variables are chosen for each of the edges and the uniform $\{1, \dots, Q\}$ colors are chosen for each of the nodes, we have defined $f(x)$ for every Q coloring x .

Now let F_t be $f_0 \circ f_1 \circ \dots \circ f_t$. 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* if F_T is a constant, that is $F_T(\Omega) = \{c\}$. One way to check if the Swendsen-Wang chain has completely coupled at time t is to check that $F_t(x)$ has the same value for every $x \in \Omega$. Since $|\Omega| = 2^{|V|}$, 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, the idea of coupling from the past (CFTP) of Propp and Wilson [10] 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 outside these values.

Bounding Chains Bounding chains were introduced independently in [8] and [7]. The idea is simple. A second *bounding chain* with state space $(2^{\{1, \dots, k\}})^V$ is run alongside the original chain with state space $\{1, \dots, k\}^V$. Let X_0, X_1, \dots be a run of the original chain, and Y_0, Y_1, \dots be a run of the bounding chain. At each time t , we maintain inclusion. That is, $X_t(v) \in Y_t(v) \Rightarrow X_{t+1}(v) \in Y_{t+1}(v)$ for all vertices v .

Note that if $Y_0(v) = \{1, \dots, k\}$ for all v , then we know $F_t(\Omega)(v) \subset Y_t(v)$ for all t . Therefore if $|Y_t(v)| = 1$ for all v , then $F_t(\Omega)$ is a constant and complete coupling has occurred. One such bounding chain for Swendsen-Wang is given below. Say that $w \sim v$ if w and v are connected using edges in B . Let v_C denote the component in \mathcal{C}_A which contains v .

Swendsen-Wang Bounding Chain Step**Let** $A = \{\{v, w\} \in E : Y(v) = Y(w), |Y(v)| = |Y(w)| = 1\}$ **Let** $B = \{\{v, w\} \in E : |Y(v) \cap Y(w)| \geq 1\}$ **For** each edge e set $U(e) \in_U [0, 1]$ **For** each node v set $Q(v) \in_U \{1, \dots, Q\}$ **For** each edge $e \in A$ If $U(e) < 1 - p$ **Set** $A = A \setminus \{e\}$ **Set** $B = B \setminus \{e\}$ **For** all $C \in \mathcal{C}_A$ **Set** $X(w) = k(C_v)$ for all $w \in C$ **Choose** a total order uniformly at random for $C \in \mathcal{C}_A$ **For** all v **Set** $Y(v) = \cup_{v_C < w_C, w \sim v} X(w)$

The intuition behind this chain is the following. The set $Y(v)$ denotes the set of colors which are possible for v . The actual coloring received depends upon X_0 , but we know that $X_t(v) \in Y_t(v)$ for all v and all t .

Then if $|Y(v) \cup Y(w)| \geq 1$ for some edge $\{v, w\}$ in E , then there might be an edge between v and w . The set B is the set of all possible edges. On the other hand, if an edge is in B and $|Y(v)| = |Y(w)| = 1$, then there is only one common color possible for v and w , and we know that this edge will be definitely turned on, no matter what the value of X_0 . These are the edges that are in A .

In other words, the edges in A are those edges that we know for sure are included in the Swendsen-Wang step, and edges in $B \setminus A$ are edges which we are not sure whether they should be included or not.

So now the recoloring phase begins. We start by giving each component connected by edges in A a single color. If $B \setminus A$ was empty, then this step is exactly the step taken in the basic Swendsen-Wang chain. Unfortunately, $B \setminus A$ might not be empty, and edges in this set might need to be included, in which case a component would receive the color of another component connected by edges in B . Therefore the set of possible colors that a component might receive is just the union over the color choices for components connected by edges in B , which is exactly the statement of the last line of the bounding chain step.

Now we that we have our algorithm, we are able to state our main theorem.

Theorem 1 *Set*

$$\beta = 1 - (1 - p)^\Delta + \frac{Q - 1}{2Q} \cdot \frac{p\Delta}{1 - p(\Delta - 1)}$$

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

Proof. Let D denote the set of vertices where $|Y(v)| > 1$ at the beginning of a bounding chain step, and D' the set of vertices where $|Y(v)| > 1$ at the end of the bounding chain step. Clearly, when $|D| = 0$ then $|Y(v)| = 1$ for all v and the chain has completely coupled. We will show that $E[|D'| | D] < |D|$. That is, on average the size of D shrinks at each step, and so eventually with positive probability will be 0.

The key point is that for a vertex to receive more than one color, it must be connected to another vertex using edges in B , and at least one edge in $B \setminus A$. But for an edge to be in $B \setminus A$,

it had to have been adjacent to a node in D . Hence nodes in D' must be connected through edges in B to a node in D . This is necessary but not sufficient; a node v may be connected to a node in D and still end up not being placed in D' depending on the color choices. Three possibilities that would preclude v adding w to D' are: w is already in D' because of another node, the color chosen for v and w is the same, and $w_C > v_C$, so w does not add the color of v to its edge set. While we cannot analyze the first exclusion, it is easy to quantify the last two.

We shall write $w \sim v$ if nodes v and w are connected using edges in B . Every node placed in D' is either in D to start with or connected to some node in D using edges in B . Therefore we have that

$$|D'| \leq \sum_{v \in D} \left[1_{v \in D'} + \sum_{w \sim v} 1_{k(w_C) \neq k(v_C)} 1_{w_C < v_C} \right].$$

The ordering of components is uniform over all possible orderings, and so $P(w_C < v_C) \leq 1/2$. The probability that the components which v and w lie in receiving different colors is $(Q-1)/Q$. By linearity of expectations

$$E[|D'|] \leq \sum_{v \in D} \left[P(v \in D') + \sum_{w \sim v} \frac{Q-1}{2Q} \right].$$

Now, for v to be in D' , at least one edge adjacent to v must have survived the edge removal phase. This occurs with probability $1 - (1-p)^\Delta$. We bound the number of $w \sim v$ by using a branching process argument (see [3]).

After edge removal, the number of nodes adjacent to w is bounded above stochastically by a binomial random variable with parameters p and Δ . Each of these is a separate branching process, with number of children distributed as a binomial random variable with parameters p and $\Delta - 1$. (The possible number of children is $\Delta - 1$ rather than Δ because one edge must be used as the parent). Let h denote the expected size of each of these child processes.

Each one of these children processes has $(\Delta - 1)p$ expected number of children, and so the expected size of each child branching process satisfies the recursion:

$$E[h] \leq 1 + (\Delta - 1)pE[h].$$

Solving, we find that $E[h] \leq 1/[1 - p(\Delta - 1)]$.

The node v has $p\Delta$ expected such branching processes as children, and so altogether the expected number of $w \sim v$ is $p/[1 - p(\Delta - 1)]$. Summing up, we find that

$$E[|D'|] \leq \sum_{v \in D} \beta = |D'| \beta$$

where

$$\beta = \left[1 - (1-p)^\Delta + \frac{Q-1}{2Q} \cdot \frac{p\Delta}{1 - p(\Delta - 1)} \right].$$

What we have shown is that $E[|D'| | D] \leq \beta |D|$. If $|D_t|$ is the size of D at time t , then $E[|D_{t+1}| | D_t] \leq \beta |D_t|$ and a simple conditioning argument gives $E[|D_t| | D_0] \leq \beta^t |D_0|$. D_0 is just the entire set of vertices V , and so $E[|D_t|] \leq \beta^t n$. Hence after $-\log_\beta 2n$ time steps, $E[|D_t|] \leq 1/2$. Since $|D_t|$ is integral, we have that $P(|D_t| = 0) > 1/2$ by Markov's inequality, and we are done. \square

How does this result compare with the 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. For Δ at most three, the Cooper/Frieze result is stronger, but for larger Δ our result gives the

currently fastest mixing time. 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
$Q = 2$	0.410	0.260	0.188	0.148	0.121	0.103
Any Q	0.318	0.202	0.147	0.116	0.095	0.081

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.

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

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

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 [1]) can be used. We wish to find the expected size of the large component after pruning. Consider the probability that this size, call it C , is greater than some i . Pick a node in the original component. Since this node will be in the new component, $P(C \geq 1) = 1$. There are at least $|A_t|/2$ nodes that this first node could connect to, so we have that

$$P(C \geq 2) \geq (1 - (1 - p)^{|A_t|/2})P(C \geq 1).$$

Since we know that $|A_t| \geq n/Q$, and $p > \ln(Qn^2)2Q/n$, we have that $(1 - p)^{|A_t|/2} \leq e^{-\ln(Qn^2)} = 1/(Qn^2)$.

Similarly, $P(C \geq i) \geq (1 - (1 - p)^{|A_t|/2})^{i-1}$ as long as $i \leq |A_t|/2$. Once we know $C \geq |A_t|/2$, consider one of the remaining nodes that we do not yet know is in C . If the node is connected to any of the nodes we know are in the new component, then the node will also be in the new component. Since $C \geq |A_t|/2$, we again see that the probability that this node will be joined to one of the old nodes is at least $1 - (1 - p)^{|A_t|/2}$, and $P(C \geq i) \geq (1 - 1/(Qn^2))^{i-1}$ for all $1 \leq i \leq |A_t|$.

Now

$$\begin{aligned} E[C] &= \sum_{i=1}^{|A_t|} P(C \geq i) \\ &\geq \sum_{i=1}^{|A_t|} (1 - 1/(Qn^2))^{i-1} \\ &\geq |A_t| - \sum_{i=1}^{|A_t|} (i-1)/(Qn^2) \\ &\geq |A_t| - 1/(2Q) \end{aligned}$$

Adding both the pruning and the expansion of the giant component gives

$$E[|A_{t+1}|] \geq |A_t| - 1/(2Q) + 1/Q = |A_t| + 1/(2Q).$$

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

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

Worse Case Running Time We now compute a worst case bound on the 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. Let K be the number of components (connected by known edges) in the graph at any time. Then with probability $1/K^Q$, each of these components are colored the same way, and the state immediately becomes known. Since $K \leq n$, we have that the chain will have coupled with probability at least $1/2$ after $n^Q \ln 2$ time steps, irregardless of p or Δ .

Note that the number of states is n^Q , so Swendsen-Wang is never worse than the time needed for a direct sample to be taken using the trivial method.

3 Trees and Cycles

Physically, the Ising model does not have a phase transition in the case of a cycle or a tree [11]. 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 [2] did in fact show Swendsen-Wang to be rapidly mixing on these graphs. We now show that Swendsen-Wang completely couples on trees.

Theorem 4 *On trees, let T be the complete coupling time for Swendsen-Wang. Then*

$$\frac{\ln(2)}{-p \ln(1 - 1/Q)} \leq T \leq \frac{\ln(2n)}{-p \ln(1 - 1/Q)} < Q \ln(2n)$$

PROOF: Note that unknown edges are only created during the coloring phase of Swendsen-Wang. Since we are in a tree, each edge defines a cut. During the coloring phase, we record for each edge whether the endpoints were colored the same way or differently. In an arbitrary graph, this might lead to a contradiction on a cycle, but since we are working with trees it does not.

Using this method, edges never move from known to unknown, and can move from unknown to known by either being pruned with probability $1 - p$ or colored the same way with probability $1/Q$. We start with $n - 1$ unknown edges. After t steps, the expected number of unknown edges left is

$$E[|D_t|] = \left[\left(1 - \frac{1}{Q}\right) p \right]^t (n - 1).$$

We know $0 \leq |D_t| \leq n$. So as long as this expectation is above $n/2$, we know that complete coupling hasn't occurred with probability at least $1/2$. When the expectation has fallen below $n/2$, however, complete coupling must have occurred with probability $1/2$ by Markov's Inequality. This gives the bounds above. (Note that we get equality for the expectation by considering two states, one with all edges labeled $\{1\}$ at the start, and one with all edges labeled $\{0\}$.) \square

For cycles, note that if $p \geq 1 - 1/(nQ)$, the high temperature results of the previous section apply. If p is low, we can remove an edge to leave a tree, sample, and then reject the sample if the two endpoints are colored differently with probability $1/(1 - p)$. Since $1 - p \geq 1/(nQ)$ the expected number of samples needed to be taken is nQ , and the total running time is $O(nQ^2 \ln(n))$. This is a polynomial time method for exact sampling from the Gibbs distribution on cycles.

4 Coupling From the Past

We have stated that coupling from the past (CFTP) may be used to obtain exact samples from the stationary distribution of a Markov chain. This is a technique developed by Propp and Wilson [10]. We 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. 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.

5 Open Questions

The idea of a bounding chain has proven useful in showing that several Markov chains are in fact complete coupling chains. In [8] 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 [12] is known to be a complete coupling chain, but no bounds on how quickly it couples are known. The method of Wolffe [14] 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] B. Bollobás. *Random Graphs*. Academic Press, London, 1985.
- [2] Colin Cooper and Alan Frieze. Mixing properties of the Swendsen-Wang process in classes of graphs. *Random Structures & Algorithms*, 15(3-4):242-261, 1999.

- [3] Luc Devroye. *Branching Processes and Their Applications in the Analysis of Tree Structures and Tree Algorithms*, pages 249–306. Springer, 1998.
- [4] R.G. Edwards and A.D. Sokal. Fortuin-Kasteleyn-Swendsen-Wang representation and Monte Carlo algorithm. *Physical Review*, 38:2009–2012, 1988.
- [5] C.M. Fortuin and P.W. Kasteleyn. On the random cluster model I: Introduction and relation to other models. *Physica*, 57:536–564, 1972.
- [6] V. Gore and M. Jerrum. The Swensen-Wang process does not always mix rapidly. *Journal of Statistical Physics*, 97(1–2):67–86, 1999.
- [7] Olle Häggström and Karin Nelander. On exact simulation from Markov random fields using coupling from the past. *Scand. J. Statist.*, 26(3):395–411, 1999.
- [8] Mark L. Huber. Exact sampling and approximate counting techniques. In *Proc. of the 30th Symposium on the Theory of Computing*, pages 31–40, 1998.
- [9] M. Jerrum and A. Sinclair. Polynomial-time approximation algorithms for the Ising model. *SIAM Journal on Computing*, 22:1087–1116, 1993.
- [10] James G. Propp and David B. Wilson. Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Structures & Algorithms*, 9(1–2):223–252, 1996.
- [11] Barry Simon. *The Statistical Mechanics of Lattice Gasses*, volume 1. Princeton University Press, 1993.
- [12] Mark Sweeny. Monte Carlo study of weighted percolation clusters relevant to the Potts models. *Physical Review B*, 27(7):4445–4455, 1983.
- [13] R. Swendsen and J-S. Wang. Non-universal critical dynamics in Monte Carlo simulation. *Physical Review Letters*, 58:86–88, 1987.
- [14] Ulli Wolff. Collective Monte Carlo updating for spin systems. *Physical Review Letters*, 62(4):361–364, 1989.