Journal of Statistical Mechanics: Theory and Experiment
An IOP and SISSA journal

Rapid mixing for lattice colourings with fewer colours

Dimitris Achlioptas[1], **Mike Molloy**[1,2], **Cristopher Moore**[3,4] **and Frank Van Bussel**[2]

[1] Microsoft Research, USA
[2] Department of Computer Science, University of Toronto, Canada
[3] Computer Science Department, University of New Mexico, NM, USA
[4] Physics and Astronomy Department, University of New Mexico, NM, USA
E-mail: optas@microsoft.com, molloy@cs.toronto.edu, moore@santafe.edu and fvb@cs.toronto.edu

Received 14 December 2004
Accepted 7 October 2005
Published 26 October 2005

Online at stacks.iop.org/JSTAT/2005/P10012
doi:10.1088/1742-5468/2005/10/P10012

Abstract. We provide an optimally mixing Markov chain for 6-colourings of the square lattice on rectangular regions with free, fixed, or toroidal boundary conditions. This implies that the uniform distribution on the set of such colourings has strong spatial mixing, so the six-state Potts antiferromagnet has a finite correlation length and a unique Gibbs measure at zero temperature. Four and five are now the only remaining values of q for which it is not known whether there exists a rapidly mixing Markov chain for q-colourings of the square lattice.

Keywords: analysis of algorithms

J. Stat. Mech. (2005) P10012

 1742-5468/05/P10012+12$30.00

Contents

1. Introduction

Sampling and counting graph colourings is a fundamental problem in computer science and discrete mathematics. It is also of fundamental interest in statistical physics: graph colourings correspond to the zero-temperature case of the *antiferromagnetic Potts model*, a model of magnetism on which physicists have performed extensive numerical experiments (see for instance [11, 18, 17]).

Physicists wish to estimate quantities such as spatial correlations and magnetization, and to do this they sample random states using Markov chains. This is a general technique whereby one starts with an arbitrary state, and then repeatedly modifies it using a random rule. For the zero-temperature Potts model, two standard such rules described below are Glauber dynamics and Kempe chain flips, also known as the Wang–Swendsen–Kotecký algorithm. While these algorithms often appear to work well in practice, we would like to know that their *mixing time*—i.e., the time it takes them to achieve a nearly uniform distribution on the set of states—is polynomially bounded as a function of the size of the lattice. Establishing this rigorously has been a major project in mathematical physics and theoretical computer science; see for example [3, 13, 15, 20, 24].

Moreover, *optimal temporal mixing*, i.e., a mixing time of $O(n \log n)$, is deeply related to the physical properties of the system [10]. In particular, under certain conditions on the Markov chain, it implies *spatial mixing*, i.e., the exponential decay of correlations, and thus the existence of a finite correlation length and the uniqueness of the Gibbs measure. Therefore, optimal mixing of natural Markov chains for q-colourings of the square lattice is considered a major open problem in physics (see e.g. [23]). Physicists have conjectured [11, 23] that the q-state Potts model has spatial mixing for $q \geq 4$, even at zero temperature. As we discuss below, previous results [13, 20, 3, 4] have established this rigorously for $q \geq 7$.

Our main result is that the square lattice has strong spatial mixing for $q = 6$, and in particular that a natural Markov chain has optimal temporal mixing. We consider the so-called *block heat-bath dynamics*, which we call $M(i, j)$. At each step, for fixed integers i and j, we choose an $i \times j$ block S of the lattice G, uniformly at random from among

J. Stat. Mech. (2005) P10012

all such blocks contained in G. Let C be the set of q-colourings of S which are consistent with the colouring of $G \setminus S$. We choose a uniformly random colouring $c \in C$ and recolour S with c. Our main theorem is:

Theorem 1. *$M(2,3)$ on 6-colourings of the square lattice mixes in $\mathrm{O}(n \log n)$ time.*

We prove theorem 1 for finite rectangular regions with free, fixed, or toroidal boundary conditions. Our method is similar to that of [3] in that it consists of a computer-assisted proof of the existence of a path coupling. At the same time, we exploit the specific geometry of the square lattice to consider a greater variety of neighbourhoods. Moreover, the calculations necessary to find a good coupling in our setting are far more complicated than those in [3] and require several new ideas to become computationally tractable.

Using the comparison method of Diaconis and Saloff-Coste [6, 19], theorem 1 implies that the Glauber and Kempe chain Markov chains also mix in polynomial time:

Theorem 2. *The Glauber and Kempe chain Markov chains on 6-colourings of the square lattice mix in $\mathrm{O}(n^2 \log n)$ time.*

Like theorem 1, this result holds on finite rectangular regions with free, fixed, or toroidal boundary conditions.

To discuss spatial mixing, suppose we have a finite region V and two colourings C, C' of its boundary that differ at a single vertex v, and a subregion $U \subseteq V$ such that the distance from v to the nearest point $u \in U$ is ℓ. Let μ and μ' denote the probability distributions on colourings of U, given the uniform distribution on colourings of V conditioned on C and C' respectively. Then we define spatial mixing as follows:

Definition 1 ([10]). *We say that q-colourings have strong spatial mixing if there are constants $\alpha, \beta > 0$ such that the total variation distance between μ and μ' obeys $\|\mu - \mu'\| \leq \beta |U| \exp(-\alpha \ell)$.*

In other words, strong spatial mixing means that conditioning a uniformly random colouring of the lattice on the event that particular colours appear on vertices far away from v has an exponentially small effect on the conditional distribution of the colour of v. Physically, this means that correlations decay exponentially as a function of distance, and that the system has a unique Gibbs measure and no phase transition.

The following recent result of Dyer *et al* [10] (see also the lecture notes by Martinelli [15]) relates optimal temporal mixing with spatial mixing: if the boundary constraints are *permissive*, i.e., if a finite region can always be coloured no matter how we colour its boundary, and if the heat-bath dynamics on some finite block mixes in $\mathrm{O}(n \log n)$ time, then the system has strong spatial mixing. As they point out, q-colourings are permissive for any $q \geq \Delta + 1$. Therefore, theorem 1, which states that $M(2,3)$ has optimal temporal mixing, implies the following result about spatial correlations.

Corollary 1. *The uniform measure on the set of q-colourings of the square lattice, or equivalently the zero-temperature antiferromagnetic q-state Potts model on the square lattice, has strong spatial mixing for $q \geq 6$.*

As mentioned above, physicists conjecture spatial mixing for $q \geq 4$. In the last section we discuss to what extent our techniques might be extended to $q = 4, 5$.

doi:10.1088/1742-5468/2005/10/P10012

1.1. Markov chains, mixing times, and earlier work

Given a Markov chain M, let π be its stationary distribution and P_x^t be the probability distribution after t steps starting with an initial configuration x. Then, for a given $\epsilon > 0$, the ϵ-mixing time of M is

$$\tau_\epsilon = \max_x \min\{t : \|P_x^t - \pi\| < \epsilon\}$$

where $\|P_x^t - \pi\|$ denotes the total variation distance

$$\|P_x^t - \pi\| = \tfrac{1}{2} \sum_y |P_x^t(y) - \pi(y)|.$$

In this paper we will often adopt the common practice of suppressing the dependence on ϵ, which is typically logarithmic, and speak just of the mixing time τ for fixed small ϵ. Thus the mixing time becomes a function of n, the number of vertices, alone. We say that a Markov chain has *rapid mixing* if $\tau = \mathrm{poly}(n)$, and *optimal (temporal) mixing* if $\tau = \mathrm{O}(n \log n)$.

The most common Markov chain for the Potts model is *Glauber dynamics*. There are several variants of this in the literature, but for colourings we fix the following definition, which applies at zero temperature. At each step, choose a random vertex $v \in G$. Let S be the set of colours, and let T be the set of colours taken by vs neighbours. Then choose a colour c uniformly at random from $S \setminus T$, i.e., from among the colours consistent with the colouring of $G - \{v\}$, and recolour v with c.

Independently, Jerrum [13] and Salas and Sokal [20] proved that for q-colourings on a graph of maximum degree Δ the Glauber dynamics (i) is ergodic for $q \geq \Delta + 2$ (this holds for fixed boundary conditions as well) and (ii) has optimal mixing for $q > 2\Delta$. For $q = 2\Delta$, Bubley and Dyer [2] showed that it mixes in $\mathrm{O}(n^3)$ time and Molloy [16] showed that it has optimal mixing. Since the square lattice has $\Delta = 4$, these results imply optimal mixing for $q \geq 8$.

Dyer and Greenhill [8] considered a 'heat-bath' Markov chain which updates both end-points of a random edge simultaneously, and showed that it has optimal mixing for $q \geq 2\Delta$. By widening the updated region to include a vertex and all of its neighbours, Bubley *et al* [3] showed optimal mixing for $q \geq 7$ for triangle-free graphs with maximum degree 4, which includes the square lattice.

Another commonly used Markov chain is the Kempe chain algorithm, known in physics as the zero-temperature case of the *Wang–Swendsen–Kotecký* algorithm [25, 26]. It works as follows. We choose a random vertex v and a colour b which differs from vs current colour a. We construct the largest connected subgraph containing v which is coloured with a and b, and recolour this subgraph by switching a and b. In a major breakthrough, Vigoda [24] showed that a similar Markov chain has optimal mixing for $q > (11/6)\Delta$, and that this implied that the Glauber dynamics and the Kempe chain algorithm both have rapid mixing for $q \geq (11/6)\Delta$. However, for the square lattice this again gives only $q \geq 8$.

For $q = 3$ on the square lattice, Luby *et al* [14] showed that a Markov chain including 'tower moves' has rapid mixing for any finite simply connected region with fixed boundary conditions, and Randall and Tetali [19] showed that this implies rapid mixing for the Glauber dynamics as well. Recently Goldberg *et al* [12] proved rapid mixing for the Glauber dynamics on rectangular regions with free boundary conditions, i.e., with no fixed

J. Stat. Mech. (2005) P10012

colouring of the vertices on their boundary. Unfortunately, the technique of [14, 12] relies on a bijection between 3-colourings and random surfaces through a 'height representation' which does not hold for other values of q.

2. Coupling

We consider two parallel runs of our Markov chain, $M(2,3)$, with initial colourings X_0, Y_0. We will couple the steps of these chains in such a way that (i) each chain runs according to the correct distribution on its choices and (ii) with high probability, $X_t = Y_t$ for some $t = \mathrm{O}(n \log n)$. A now standard fact in this area is that this implies that the chain mixes in time $\mathrm{O}(n \log n)$, i.e. this implies theorem 1; this fact was first proved by Aldous [1] (see also [9]).

Bubley and Dyer [2] introduced the very useful technique of path coupling, via which it suffices to do the following: consider any two colourings X, Y which have a Hamming distance one, i.e., which differ on exactly one vertex, and carry out a single step of the chain on X and on Y, producing two new colourings X', Y'. We will prove that we can couple these two steps such that (i) each step is selected according to the correct distribution, and (ii) the expected Hamming distance between X' and Y' is at most $1 - \epsilon/n$ for some constant $\epsilon > 0$. Thus the expected change in the Hamming distance between the two colourings is negative, $-\epsilon/n$. See, e.g., [9] for the formal (by now standard) details as to why this suffices to prove theorem 1.

We perform the required coupling as follows. Let X and Y be two arbitrary 6-colourings which only disagree at one vertex. We pick a uniformly random 2×3 block S, and let C_X and C_Y denote the set of permissible recolourings of S according to X, Y respectively. For each $c \in C_X$, we define a carefully chosen probability distribution p_c on the colourings of C_Y. We pick a uniformly random colouring $c_1 \in C_X$ and in X we recolour S with c_1 to produce X'. We then pick a random colouring $c_2 \in C_Y$ according to the distribution p_{c_1} and in Y we recolour S with c_2 to produce Y'. Trivially, the marginal distribution $\langle S, c_1 \rangle$ is uniform on S and C_X. In order to ensure that the same is true of $\langle S, c_2 \rangle$, we must have the following property for the set of distributions $\{p_c : c \in C_X\}$:

$$\text{for each } c_2 \in C_Y, \qquad \frac{1}{|C_X|} \sum_{c \in C_X} p_c(c_2) = \frac{1}{|C_Y|}. \tag{1}$$

Suppose that v is the vertex on which X, Y differ. If $v \in S$ then $C_X = C_Y$, so we can simply define $c_2 = c_1$ (i.e., $p_c(c) = 1$ for each c) and this ensures that $X' = Y'$. If S does not contain v or any neighbour of v, then again $C_X = C_Y$ and by defining $c_2 = c_1$ we ensure that X', Y' differ only on v. If v is not in S but is adjacent to a vertex in S, then $C_X \neq C_Y$ and so, depending on our coupling, it is quite possible that $c_2 \neq c_1$ and so X', Y' will differ on one or more vertices of S as well as on v.

For any pair C_X, C_Y, we let $H(C_X, C_Y)$ denote the expected number of vertices in S on which c_1, c_2 differ. For every possible pair C_X, C_Y we obtain a coupling satisfying (1) and

$$H(C_X, C_Y) < 0.52. \tag{2}$$

Thus the expected change in the Hamming distance between the two colourings is -1 if $v \in S$ and less than 0.52 if v is adjacent to S.

J. Stat. Mech. (2005) P10012

Proof of theorem 1. As described above, it suffices to prove that for any choice of X, Y differing only at v, the expected Hamming distance between X' and Y' is less than $1-\epsilon/n$ for some $\epsilon > 0$, or equivalently that the expected change in the Hamming distance is less than $-\epsilon/n$.

Let us consider toroidal boundary conditions first; there are n possible blocks from which the algorithm will choose, one for each vertex. Each vertex is contained in 6 blocks and is adjacent to 10 blocks, so $v \in S$, or v is adjacent to S, with probability $6/n$ or $10/n$ respectively. Thus the expected change in the Hamming distance is less than

$$-1 \times \frac{6}{n} + 0.52 \times \frac{10}{n} = -\frac{0.8}{n}.$$

For rectangular regions with free boundary conditions the situation is a little more complicated. A natural extension of the basic procedure is to chose any block that lies entirely in the region. This, however, means that a region of height h and width w with n vertices will only have $n - 2h - w + 2$ possible 2×3 blocks; furthermore, when the distinguished vertex v is on or close to the boundary, the number of blocks containing v is less than the six that we have when v is snugly in the interior, thereby lowering our chances of decreasing the Hamming distance.

It turns out that if we are dealing with free boundary conditions the natural extension will work anyway, since then the maximum expected change in the Hamming distance when v is on or close to the boundary and *adjacent* to our block is low enough to counteract the reduced chance that our choice of block will contain v. In particular, we obtained these values for the expected change in the Hamming distance when v is adjacent to the block:

v is on boundary, block is not:	< 0.52	*i.e., the interior case*
block is on boundary, v is not:	< 0.503	
both v and block are on boundary:	< 0.390.	

As an example, when v is in the corner of the lattice we have one choice of a block containing v, and two choices of a block adjacent to v; both of the adjacent blocks are also on the boundary too, so the expected change in the Hamming distance is less than

$$-1 \times \frac{1}{n - 2h - w + 2} + 0.39 \times \frac{2}{n - 2h - w + 2} = -\frac{0.22}{n - 2h - w + 2}.$$

It can be easily verified that in the other cases where v is on or close to a free boundary the values given above are also well within what we need for rapid mixing.

If we have to deal with fixed, arbitrary boundary conditions—that is, a fixed proper colouring of the sites just outside the region—applying our Markov process directly will not give us the same result. The problem is that the ratio between the number of blocks containing v and adjacent to v is the same as in the free boundary case, but the expected change in Hamming distance for any block adjacent to v is the same as in the toroidal case, as opposed to the lower values we obtain for free boundary conditions above.

An alternative modification of our basic Markov process that will work is to choose any block that contains *any* vertex in the region, and then (properly) recolour all the vertices in that block which are also in the region. For a lattice of height h and width w with n vertices this gives us $n + 2h + w + 2$ possible 2×3 blocks to chose from; more importantly, for every vertex in the lattice there are six available blocks which contain it.

J. Stat. Mech. (2005) P10012

The expected change in the Hamming distance when the block is on the boundary but fully contained in the lattice is now no different than the expected change for a block on the interior, but we do have to deal with cases where the block may be partially outside the lattice. When this happens we can merely treat the sub-block that we do recolour in exactly the same manner as we would treat a smaller block being recoloured on the interior (that is, we reduce the size of the block to contain only the recolourable part). The calculations for smaller blocks adjacent to v were of course already done in order to confirm previous results or rule out the possibility of rapid mixing for block sizes below 2×3, and the results were:

Sub-block size	Maximum expected change	
1×1	≤ 0.50	
2×1 or 1×2	<0.524	
1×3	<0.514	
2×2	<0.508	
2×3	<0.52	*(i.e. the interior case)*

As an example, when v is in the corner of the lattice we now have six choices of a block containing v, and five choices of a block adjacent to v; two of these choices are full 2×3 blocks, and the remaining choices have sub-block sizes of 2×1, 2×2, and 1×3 respectively. Hence the expected change in the Hamming distance is less than

$$-1 \times \frac{6}{n+2h+w+2} + \frac{0.52 \times 2 + 0.524 + 0.508 + 0.514}{n+2h+w+2} = -\frac{3.414}{n+2h+w+2}.$$

We could work out the cost explicitly for the rest of the cases where the distinguished vertex v is on or near the boundary, but it is easier just to note that in no case can the expected cost of colouring a block adjacent to v be more than 0.524, and there can never be more than 10 adjacent blocks to choose from, so the expected change in the Hamming distance can never be more than

$$-\frac{0.76}{n+2h+w+2}.$$

□

Of course, we still need to prove that the desired couplings exist for each possible C_X, C_Y. These couplings were found with the aid of computer programs. In principle, for any pair C_X, C_Y, searching for the coupling that minimizes $H(C_X, C_Y)$ subject to (1) is simply a matter of solving a linear program and so can be done in polynomial time. However, the number of variables is $|C_X||C_Y|$ which, *a priori*, can be roughly $(5^6)^2$. Furthermore, the number of possible pairs X, Y is roughly 6^{10}, and even after eliminating pairs which are redundant by symmetry, it is enormous. To deal with this combinatorial explosion we designed a fast heuristic which, rather than finding the best coupling for a particular pair, simply finds a very good coupling; i.e., one that satisfies (2). The code used can be found at www.cs.toronto.edu/∼fvb We provide a more detailed description in the next section.

3. The programs used

Method of the computation. Let R denote the *rim* vertices, that is, those vertices which are adjacent to but outside of the block S. We call a colouring of the vertices of R a

rim colouring. For each possible pair of rim colourings X, Y which differ only at a vertex $v \in R$, we need to find a coupling between the extensions C_X and C_Y of X, Y to S, such that the couplings satisfy (1) and (2). These couplings were found with a small suite of programs working in two phases. In the first phase, exhaustive lists of pairs of rim colourings (reduced by equivalence with respect to allowable block colourings) were generated. In the second phase, for each pair X, Y, we generated C_X, C_Y separately; these were then coupled, satisfying (1), in a nearly optimal way to obtain a bound on $H(C_X, C_Y)$ that satisfies (2).

Implementation. All programs take the following parameters: number of colours, block dimensions, and an integer denoting the position of the distinguished vertex v with respect to the block (0 if adjacent to the corner, $+i$ if adjacent to the ith vertex along the top of the block, and $-i$ if adjacent to the ith vertex along the side of the block). We assume that v has colour 0 in X and 1 in Y. Thus, if one specifies a rim colouring X and the position of v, then this determines Y. For each colouring X we determined a good coupling for each non-equivalent position of v.

By default the programs generate rim colourings and couplings on the assumption that the block is not on the boundary of the lattice (i.e., all rim vertices potentially constrain the allowable block colourings). Free boundary cases, however, can easily be simulated by using values in the rim colourings that are outside the range determined by the number-of-colours parameter. These were only checked when the analysis of the non-boundary blocks yielded promising values. As mentioned above, the fixed arbitrary boundary cases required no new calculations, since from the block's 'point of view' there is no difference (in the worst case) between being on an arbitrarily coloured boundary and being in the interior. Thus, we simply reused previously calculated values for smaller blocks on the interior to verify that the modified Markov process would work.

Generating rim colourings. Since the calculations required for phase 2 were much more time-consuming than those for phase 1, the rim colouring generation procedure was designed to minimize the number of colourings output rather than the time used generating them. A rim colouring is represented by a vector of colours used on the rim, starting from the distinguished vertex v and going clockwise around the block. If the set of colours is $\{0, \ldots, 5\}$, we can assume by symmetry that vs colour is 0 in X and 1 in Y. The following reductions were applied to avoid equivalent rim colourings: reduction by colour isomorphism (colours 2 and above), by exchange of colours 0 and 1, by exchange of colours of vertices adjacent to the corners of the block, and by application of flip symmetries where applicable.

Finding a coupling for particular rim colourings. Two programs were used for each rim colouring X, and position i of v. In each, the initial operation is the generation of all compatible lattice colourings; this is done separately for $\mathrm{col}(v) = 0$ and $\mathrm{col}(v) = 1$ (i.e., for C_X and C_Y). The first program creates a set of linear programming constraints that is readable by the program `lp-solve` (by Michel Berkelaar of the Eindhoven University of Technology; it is available with some Linux distributions). As mentioned above, time and space requirements made use of this procedure feasible only for checking individual rim colourings, and even then the block size had to be fairly modest.

The second program calculates an upper bound on the optimal cost using a greedy algorithm to create a candidate coupling. Given sets of colourings C_X and C_Y of size m_X and m_Y respectively, the algorithm starts by assigning 'unused' probabilities of $1/m_X$ and

$1/m_Y$ respectively to the individual colourings. Then, for each distance $d = 0, 1, \ldots, 6$, for each colouring c_1 in C_X it traverses C_Y looking for a colouring c_2 which differs from c_1 on exactly d vertices. When such an c_2 is found it removes the colouring with the lower unused probability p from its list and reduces the unused probability p' of the other to $p' - p$; the distance $d \cdot p$ is added to the total distance so far. The order in which the lists of colourings C_X and C_Y is traversed does affect the solution, so an optional argument is available that allows the user to select one of several alternatives.

This heuristic does not guarantee an optimal solution, and with some blocks and particular rim colourings the coupling it generates is far from the best. However, for the rim colourings we are most interested in (ones where $H(C_X, C_Y)$ is high for all couplings) it seems to consistently give results that are optimal or very close (within 2%). We cannot give a rigorous bound on the running time, but a cursory analysis and empirical evidence suggest that it runs in roughly $\mathcal{O}(m \log m)$ time, where m is the number of compatible block colourings. Because the heuristic is so much faster than the LP solver, our general procedure was as follows. (1) Use the heuristic with the default traversal order to calculate bounds on the expected distance for all the rim colourings generated in phase one. (2) When feasible, use the LP solver on those rim colourings that had the highest value of $H(C_X, C_Y)$, to obtain an exact value for their maximum. (3) For larger blocks or more colours than could be comfortably handled by the LP solver, use all available traversal orders on those rim colourings that had the maximum value of $H(C_X, C_Y)$ to obtain as tight a bound as possible within a feasible time.

Results of the computations. Computations were run on various block sizes and numbers of colours in order to check the correctness of the programs, and also to collect data which could be used to estimate running times and maximum expected distance for larger block dimensions. For 7- and 8-colourings our results corresponded well with previous work on the problem (e.g. [2]).

For 6-colourings, we checked $1 \times k$ blocks for $k \leq 5$, as well as 2×2 and 2×3 blocks. For all but the last of these the maximum expected distance we obtained was too large to give us rapid mixing. The 2×3 subgrid has two non-equivalent positions with respect to the rim: the corner (to which eight rim vertices are adjacent) and the middle of the top or bottom side (to which two rim vertices are adjacent). Denote these positions 1 and 2 respectively. For each X, Y with v in position 1, we obtained a coupling satisfying

$$H(C_X, C_Y) \leq 0.511\,830\,9760.$$

For each X, Y with v in position 2, we obtained a coupling satisfying

$$H(C_X, C_Y) \leq 0.483\,786\,3092.$$

Thus, in each case we satisfy (2) as required.

A slightly stronger output. By examining the problem a bit more closely, we see that condition (2) is sufficient, but not necessary, for our purposes. Let H_i denote the maximum of $H(C_X, C_Y)$ over the couplings found for all pairs X, Y where v is in position i, and let mult_i denote the number of rim vertices adjacent to position i. Then, being more careful about the calculation used in the proof of theorem 1, and extending it to a general $a \times b$ block, we see that the overall expected change in the Hamming distance is at most (in, say, the toroidal case where the number of possible blocks is n)

$$-1 \times \frac{ab}{n} + \sum_i \text{mult}_i \times \frac{H_i}{n},$$

a smaller value than that used in the proof of Theorem 1, where we (implicitly) used $(\max_i H_i) \times \sum_i \mathrm{mult}_i$ rather than $\sum_i \mathrm{mult}_i \times H_i$. Our programs actually compute this smaller value. Even so, we could not obtain suitable couplings for any block size smaller than 2×3.

4. Rapid mixing: Glauber and Kempe chain dynamics

In this section we prove theorem 2, showing rapid mixing for the Glauber and Kempe chain dynamics, by following the techniques and presentation of Randall and Tetali [19].

Suppose P and $\tilde{P}$ are two Markov chains on the same state space with the same stationary distribution π, and that we already have a bound on the mixing time of $\tilde{P}$ while we would like to obtain a bound for that of P. Let $E(P)$ and $E(\tilde{P})$ denote the edges of these Markov chains, i.e., the pairs (x, y) such that the transition probabilities $P(x, y)$ and $\tilde{P}(x, y)$, respectively, are positive. Now, for each edge of $\tilde{P}$, i.e., each $(x, y) \in E(\tilde{P})$, choose a fixed path $\gamma_{x,y}$ using the edges of P: that is, choose a series of states $x = x_0, x_1, x_2, \ldots, x_k = y$ such that $(x_i, x_{i+1}) \in E(P)$ for $0 \le i < k$. Denote the length of such a path as $|\gamma_{x,y}|$. Furthermore, for each $(z, w) \in E(P)$, let $\Gamma(z, w) \subseteq E(\tilde{P})$ denote the set of pairs (x, y) such that $\gamma_{x,y}$ uses the edge (z, w). Finally, let

$$A = \max_{(z,w)\in E(P)} \left[\frac{1}{\pi(z)P(z,w)} \sum_{(x,y)\in\Gamma(z,w)} |\gamma_{x,y}|\pi(x)\tilde{P}(x,y) \right].$$

Note that A depends on our choice of paths.

By combining bounds on the mixing time in terms of the spectral gap [7, 22, 21] with an upper bound on P's spectral gap in terms of $\tilde{P}$'s due to Diaconis and Saloff-Coste [6], we obtain the following upper bound on P's mixing time:

Theorem 3. *Let P and $\tilde{P}$ be reversible Markov chains on q-colourings of a graph of n vertices whose unique stationary distribution is the uniform distribution. Let $\lambda_1(\tilde{P})$ be the largest eigenvalue of $\tilde{P}$'s transition matrix smaller than 1, let τ_ϵ and $\tilde{\tau}_\epsilon$ denote the ϵ-mixing time of P and $\tilde{P}$ respectively, and define A as above. Then for any $\epsilon \le 1/4$,*

$$\tau_\epsilon \le \frac{4 \log q}{\lambda_1(\tilde{P})} An\tilde{\tau}_\epsilon.$$

We omit the proof. The reason for the additional factor of n is the fact that the upper and lower bounds on mixing time in terms of the spectral gap are $\log 1/\pi_*$ apart, where π_* is the minimum of $\pi(x)$ taken over all states x. Since π in this case is the uniform distribution and there are at most q^n colourings, we have $\log 1/\pi(x) \le n \log q$. On the square lattice, it is easy to see that there are an exponentially large number of q-colourings for $q \ge 3$, so removing this factor of n would require a different comparison technique.

Now suppose that $\tilde{P}$ is the block dynamics and P is the Glauber or Kempe chain dynamics. We wish to prove theorem 2 by showing that $\tau_\epsilon = \mathrm{O}(n\tilde{\tau}_\epsilon)$. By adding self-loops with probability greater than $1/2$ to the block dynamics, we can ensure that the eigenvalues of $\tilde{P}$ are positive with only a constant increase in the mixing time. Therefore, it suffices to find a choice of paths for which A is constant. Since, for all three of these Markov chains, each move occurs with probability $\Theta(1/n)$, if $|\gamma_{x,y}|$ and $|\Gamma(z, w)|$ are constant then so is A.

In fact, for $q \geq \Delta + 2$, we can carry out a block move on any finite neighbourhood with Glauber moves. We need to flip each vertex u in the block to its new colour; however, u's flip is blocked by a neighbour v if v's current colour equals u's new colour. Therefore, we first prepare for u's flip by changing v to a colour which differs from u's new colour as well as that of v's Δ neighbours. If the neighbourhood has m vertices, this gives $|\gamma_{x,y}| \leq m(\Delta+1)$, or $|\gamma_{x,y}| \leq 30$ for $M(2,3)$. (With a little work we can reduce this to 13.)

For the Kempe chain dynamics, recall that each move of the chain chooses a vertex v and a colour b other than v's current colour. If b is the colour which the Glauber dynamics would assign to v, then none of v's neighbours are coloured with b, and the Kempe chain move is identical to the Glauber move. Since this happens with probability $1/q$, the above argument applies to Kempe chain moves as well, and again we have $|\gamma_{x,y}| \leq m(\Delta + 1)$. Moreover, we only need to consider moves that use Kempe chains of size 1.

Finally, since each vertex appears in only $m = 6$ blocks, the number of block moves that use a given Glauber move or a given Kempe chain move of size 1 is bounded above by m times the number of pairs of colourings of the block. Thus $|\Gamma(z, w)| \leq m(q^m)^2$, and we are done.

An interesting open question is whether we can prove *optimal* temporal mixing for the Glauber or Kempe chain dynamics. One possibility is to use log-Sobolev inequalities as in [5]. We leave this as a direction for further work.

5. Conclusion: larger blocks and smaller q?

We have run our programs on 2×4 and 3×3 blocks to see whether we could achieve rapid mixing on five colours, but in both cases the largest values of $H(C_X, C_Y)$ were too high. It may be that rapid mixing on five colours is possible by recolouring a 3×4 block, based on the decrease of the ratio of $\max H(C_X, C_Y) \cdot |R|$ to $|S|$ as the dimensions increase; similar reasoning leads us to believe that rapid mixing using a $2 \times k$ block is possible, but we would probably need a 2×10 block or larger to achieve success. Unfortunately, doing the calculations for 3×4 blocks is a daunting proposition. The problem is exponential in two directions at once (number of rim colourings, and number of block colourings for each rim colouring), so this would require a huge increase in the running time.

Acknowledgments

We are grateful to Leslie Goldberg, Dana Randall and Eric Vigoda for helpful discussions, and to an anonymous referee for pointing out a technical error. Recently, L Goldberg, R Martin and M Paterson have reported obtaining the main result of this paper independently.

Some of this work was done while M Molloy was visiting the Fields Institute, Canada, supported by NSERC and a Sloan Research Fellowship. C Moore was supported by NSF grants PHY-0200909, CCR-0220070, and EIA-0218563.

References

[1] Aldous D, *Random walks on finite groups and rapidly mixing Markov chains*, 1986 *Séminaire de Probabilités XVII 1981/82* (*Springer Lecture Notes in Mathematics* vol 986) ed A Dold and B Eckmann (Berlin: Springer) pp 243–97

J. Stat. Mech. (2005) P10012

[2] Bubley R and Dyer M, *Path coupling: a technique for proving rapid mixing in Markov chains*, 1997 *Proc. 28th Ann. Symp. on Foundation of Computer Science* pp 223–31
[3] Bubley R, Dyer M and Greenhill C, *Beating the* 2Δ *bound for approximately counting colourings: a computer-assisted proof of rapid mixing*, 1998 *Proc. 9th Ann. ACM–SIAM Symp. on Discrete Algorithms* pp 355–63
[4] Bubley R, Dyer M, Greenhill C and Jerrum M, *On approximately counting colourings of small degree graphs*, 1999 *SIAM J. Comput.* **29** 387–400
[5] Cesi F, *Quasi-factorization of the entropy and logarithmic Sobolev inequalities for Gibbs random fields*, 2001 *Probab. Theory Rel. Fields* **120** 569–84
[6] Diaconis P and Saloff-Coste L, *Comparison theorems for reversible Markov chains*, 1996 *Ann. Appl. Probab.* **6** 696–730
[7] Diaconis P and Stroock D, *Geometric bounds for eigenvalues of Markov chains*, 1991 *Ann. Appl. Probab.* **1** 36–61
[8] Dyer M and Greenhill C, *A more rapidly mixing Markov chain for graph colorings*, 1998 *Random Struct. Algorithms* **13** 285–317
[9] Dyer M and Greenhill C, *Random walks on combinatorial objects*, 1999 *Surveys in Combinatorics* ed J Lamb and D Preece (Cambridge: Cambridge University Press) pp 101–36
[10] Dyer M, Sinclair A, Vigoda E and Weitz D, *Mixing in time and space for lattice spin systems: a combinatorial view*, 2002 *Proc. RANDOM* pp 149–63
[11] Ferreira S and Sokal A, *Antiferromagnetic Potts models on the square lattice: a high-precision Monte Carlo study*, 1999 *J. Statist. Phys.* **96** 461–530
[12] Goldberg L, Martin R and Paterson M, *Random sampling of 3-colourings in* $\mathbb{Z}^2$, 2004 *Random Struct. Algorithms* **24** 279–302
[13] Jerrum M, *A very simple algorithm for estimating the number of k-colorings of a low-degree graph*, 1995 *Random Struct. Algorithms* **7** 157–65
[14] Luby M, Randall D and Sinclair A, *Markov chain algorithms for planar lattice structures*, 2001 *SIAM J. Comput.* **31** 167–92
[15] Martinelli F, *Lectures on Glauber dynamics for discrete spin models*, 1999 *Lectures on Probability Theory and Statistics (Saint-Flour, 1997)* (*Springer Lecture Notes in Mathematics* vol 1717) (Berlin: Springer) pp 93–191
[16] Molloy M, *Very rapidly mixing Markov Chains for* 2Δ*-colourings and for independent sets in a 4-regular graph*, 2001 *Random Struct. Algorithms* **18** 101–15
[17] Moore C and Newman M, *Height representation, critical exponents, and ergodicity in the four-state triangular Potts antiferromagnet*, 2000 *J. Stat. Phys.* **99** 661–90
[18] Moore C, Nordahl M, Minar N and Shalizi C, *Vortex dynamics and entropic forces in antiferromagnets and antiferromagnetic Potts models*, 1999 *Phys. Rev.* E **60** 5344
[19] Randall D and Tetali P, *Analyzing Glauber dynamics by comparison of Markov chains*, 2000 *J. Math. Phys.* **41** 1598
[20] Salas J and Sokal A, *Absence of phase transition for antiferromagnetic Potts models via the Dobrushin uniqueness theorem*, 1997 *J. Stat. Phys.* **86** 551–79
[21] Sinclair A, 1993 *Algorithms for Random Generation and Counting: a Markov Chain Approach* (Boston, MA: Birkhauser) pp 47–8
[22] Sinclair A and Jerrum M, *Approximate counting, uniform generation, and rapidly mixing Markov chains*, 1989 *Inf. Comput.* **82** 93–133
[23] Sokal A, *A personal list of unsolved problems concerning lattice gases and antiferromagnetic Potts models*, 2001 *Talk presented at the Conf. on Inhomogeneous Random Systems (Université de Cergy-Pontoise, Jan. 2000)*; *Markov Process. Related Fields* **7** 21–38
[24] Vigoda E, *Improved bounds for sampling colorings*, 2000 *J. Math. Phys.* **41** 1555
[25] Wang J, Swendsen R and Kotecký R, 1989 *Phys. Rev. Lett.* **63** 109
[26] Wang J, Swendsen R and Kotecký R, 1990 *Phys. Rev.* B **42** 2465

J. Stat. Mech. (2005) P10012