Sampling Grid Colourings with Fewer Colours

Dimitris Achlioptas[*] Mike Molloy[†] Cristopher Moore[‡] Frank Van Bussel[§]

September 19, 2003

Abstract

We provide an optimally mixing Markov chain for 6-colourings of the square grid. Furthermore, this implies that the uniform distribution on the set of such colourings has strong spatial mixing. 4 and 5 are now the only remaining values of k for which it is not known whether there exists a rapidly mixing Markov chain for k-colourings of the square grid.

Frank Van Bussel is a student.

Contact author: M. Molloy, Dept of Computer Science, University of Toronto, 10 Kings College Rd, Toronto, ON M5S 3G1, Canada. molloy@cs.toronto.edu

[*]Microsoft Research. optas@microsoft.com

[†]Dept of Computer Science, University of Toronto, and Microsoft Research. Some of this work was done while visiting the Fields Institute. Supported by NSERC and a Sloan Research Fellowship. molloy@cs.toronto.edu

[‡]Computer Science Department, University of New Mexico. Supported by NSF grant PHY-0200909. moore@santafe.edu

[§]Dept of Computer Science, University of Toronto. fvb@cs.toronto.edu

1 Introduction

Sampling and counting colourings of a graph is a fundamental problem in computer science and discrete mathematics. Much focus has gone towards attacking this problem using rapidly mixing Markov chains; see for example [3, 7, 8, 10, 11, 15, 16, 17, 19, 20, 24, 28].

These problems are 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 [13, 21, 22, 29, 30]). Physicists wish to estimate physical quantities such as spatial correlations and magnetization, and to do this they attempt to sample random states using simple Markov chains.

However, physicists typically use heuristic methods to guess how many steps of their Markov chain they need to perform in order to be close enough to the uniform distribution to obtain good estimates of these quantities. By proving rapid mixing for these Markov chains, computer scientists can provide a rigorous underpinning for physicists' work, and make it possible for numerical experiments to proceed in a more principled way.

Moreover, *optimal temporal mixing,* i.e. a mixing time of $O(n \log n)$, is deeply related to the physical properties of the system [12]. In particular, 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 grid is considered a major open problem in physics (see e.g. [27]).

1.1 Markov chains, mixing times and statement of results

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 point x. Then, for a given $\epsilon > 0$, the *ϵ-mixing time* of M is

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

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

$$\left\| P_x^t - \pi \right\| = \frac{1}{2} \sum_y \left| P_x^t(y) - \pi(y) \right| \quad .$$

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 $\tau = O(n \log n)$.

The most common Markov chain for this system is *Glauber dynamics.* There are several variants of this in the literature, but for colourings we fix the following definition. 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 v's 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 [16] and Salas and Sokal [24] proved that for q-colourings on a graph of maximum degree Δ the Glauber dynamics has optimal mixing for $q > 2\Delta$, and Bubley and Dyer [2] showed that it mixes in $O(n^3)$ time when $q = 2\Delta$. Molloy [20] showed that in fact it has optimal mixing for $q = 2\Delta$ and then Dyer, Greenhill and Molloy showed that it has optimal mixing for $q \geq (2-\epsilon)\Delta$ where ϵ is a small positive constant. (This latter result came long after Vigoda had proved rapid mixing for $q \geq (11/6)\Delta$ as described below.)

Dyer and Greenhill [10] considered a "heat bath" Markov chain which updates both ends of a random edge simultaneously, and showed that it has optimal mixing for $q \geq 2\Delta$. By widening the neighbourhood to include a site and all of its neighbours, Bubley, Dyer and Greenhill [3] showed optimal mixing for a heat bath algorithm $q \geq 7$ for 4-regular triangle-free graphs, such as the square grid.

Another Markov chain commonly used by physicists is the Kempe chain algorithm, which they call the zero-temperature case of the Wang-Swendsen-Kotecký algorithm [29, 30]. This Markov chain works as follows: we choose a random vertex v and a colour b which differs from v's 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 [28] showed that a similar Markov chain has optimal mixing for $q > (11/6)\Delta$, and 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 grid this gives only $q \geq 8$.

For $q = 3$ on the square grid, Luby, Randall and Sinclair [18] 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 [23] showed that this implies rapid mixing for the Glauber dynamics as well. Recently Goldberg, Martin and Paterson [14] proved rapid mixing for the Glauber dynamics on rectangular regions with free boundary conditions, i.e., with no fixed colouring of the vertices on their boundary. However, the technique of [18, 14] relies on a bijection between 3-colourings and random surfaces through a "height representation" which does not hold for other values of q. Thus, the situation to this point was that we had rapid mixing for $q = 3$ and $q \geq 7$, but not for $q = 4, 5$ or 6.

Here, we narrow this gap by considering the following Markov chain, often called the *block heat-bath dynamics*, which we call $M(i, j)$. At each step, we choose an $i \times j$ subgrid S of G, uniformly at random from amongst all such subgrids (i.e. its upper-left vertex is chosen uniformly from the vertices of 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. Note that $M(1, 1)$ is simply the Glauber dynamics. Our main theorem is:

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

We prove Theorem 1 for a variety of types of boundary conditions: on the torus, on finite rectangular regions with fixed colourings on their boundary, and on finite rectangular regions with free boundary conditions.

While our method is similar to that of [3] in that it consists of a computer-assisted proof of the existence of a path coupling, the specific geometry of the square grid allows us to consider a greater variety of neighbourhoods. Moreover, the calculations are an order of magnitude more complicated, and require several new ideas to carry them out feasibly.

Using the comparison method of Diaconis and Saloff-Coste [5, 23], Theorem 1 easily implies that the Glauber and Kempe chain Markov chains also mix in polynomial time:

Theorem 2 *The Glauber dynamics and the Kempe chain dynamics on 6-colourings of the square grid mix in $O(n^2 \log n)$ time.*

Like Theorem 1, this result holds both on the torus and on finite rectangular regions with fixed or free boundary conditions.

Finally, using recent results of Dyer, Sinclair, Vigoda and Weitz [12] (see also the lecture notes by Martinelli [19]) the fact that $M(2,3)$ has optimal temporal mixing implies a strong result about spatial correlations. Consider a finite region V and two colourings C, C' of its boundary that differ at a single site 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 say that q-colourings have *optimal spatial mixing* if there are constants $\alpha, \beta > 0$ such that $\|\mu - \mu'\| \leq \beta |U| \exp(-\alpha \ell)$. In other words, conditioning on particular colours appearing on vertices far away from v in a uniformly random colouring of the grid has exponentially small effect on the conditional distribution of the colour of v.

Stated in [12] is the following result relating optimal temporal mixing with spatial mixing: in the case where the constraints are *permissive*, i.e. a finite region can always be coloured no matter how we colour its boundary, if heat-bath dynamics on some finite block mixes in $O(n \log n)$ time, then the system has strong spatial mixing. As they point out, q-colourings are permissive for any $q \geq \Delta + 1$, so we have

Corollary 3 *The uniform measure on the set of q-colourings of the square grid, or equivalently the zero-temperature antiferromagnetic q-state Potts model on the square lattice, has strong spatial mixing for $q \geq 6$.*

Physicists have conjectured [13, 27] that the q-state Potts model has spatial mixing for $q \geq 4$. While we have narrowed this gap, we have not closed it, and in the last section we discuss to what extent our techniques might be extended to smaller 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 $t = O(n \log n)$. A now standard fact in this area is that this implies that the chain mixes in time $O(n \log n)$, i.e. this implies Theorem 1; this fact was first proved by Aldous [1] (see also [9, 17]).

Bubley and Dyer [2] introduced the very useful technique of Path Coupling, via which it suffices to do the following: Consider any two *not necessarily proper* colourings X, Y 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 of the two steps is selected according to the correct distribution, and (ii) the expected number of vertices on which X', Y' differ is at most $1 - \epsilon$ for some constant $\epsilon > 0$. See, for example, [9] [17] or [11] for the formal (by now standard) details as to why this suffices to prove Theorem 1.

We perform the required coupling as follows. We pick a uniformly random 2×3 subgrid 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 member $c_1 \in C_X$ and in X we recolour S with c_1 to produce X'. We then pick a random member $c_2 \in C_Y$ according to the distribution p_{c_1} and in Y we recolour S

with c_2 to produce Y'. Trivially, the pair S, c_1 is chosen according to the correct distribution. In order to ensure that the same is true of S, c_2, 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 $C_X \neq C_Y$ so our coupling becomes very complicated and 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:

$$H(C_X, C_Y) < .52. \tag{2}$$

This immediately proves our main theorem:

Proof of Theorem 1 As described above, it suffices to prove that for any choice of X, Y differing only at v, the expected number of vertices on which X', Y' differ is less than $1 - \epsilon/n$ for some $\epsilon > 0$. The probability that S contains v is $|S|/n = 6/n$; for any such choice of S, $X' = Y'$. The probability that S contains a neighbour of v but does not contain v is easily seen to be $10/n$; for any such choice of S, the expected number of vertices on which X', Y' differ is less than 1.52. Therefore, the overall expected number of vertices on which X', Y' differ is less than

$$1 \times (n - 16)/n + 0 \times 6/n + 1.52 \times 10/n = 1 - .8/n.$$

□

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 polytime. However, the number of variables is $|C_X| \times |C_Y|$ which can be as high as 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. Thus, finding these couplings is computationally intensive. To help we designed a fast heuristic which, rather than finding the best coupling for a particular pair, just found a very good coupling; i.e. one that satisfied (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 subgrid S. We call a (not necessarily proper) 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, so that the couplings satisfies (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 grid 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 (close to) 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, grid dimensions, and an integer denoting the position of the distinguished vertex v with respect to the grid (0 if adjacent to the corner, $+i$ if adjacent to the i-th vertex along the top of the grid, and $-i$ if adjacent to the i-th vertex along the side of the grid). Note that if one specifies a rim colouring X and the position of v, then this determines Y (up to equivalence via permutation of colours). 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 subgrid is not on the boundary of the supergrid (i.e. all rim vertices potentially constrain the allowable subgrid colourings). 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. Grids on the boundary were only checked when the analysis of the non-boundary grids yielded promising values; in all cases we found that the maximum cost for boundary subgrids was lower than for the associated non-boundary subgrids.

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 the colours used on the rim, starting from the distinguished vertex v and going clockwise around the subgrid. Since we can assume by symmetry that the colour used for v is 0 in X and 1 in Y, $0|1$ is always the first element (0 is used in the actual output, 1 is understood). 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 subgrid, 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 grid colourings; this is done separately for $\mathrm{col}(v) = 0$ and $\mathrm{col}(v) = 1$ (i.e. for C_X, 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 subgrid 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 $\frac{1}{m_Y}$ and $\frac{1}{m_X}$ to the individual colourings. Then, for each distance $d = 0, 1, ..., n$, 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.

We note that this heuristic does not guarantee an optimal solution, and with some grids 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 something like $\mathcal{O}(m \log m)$ time, where m is the number of compatible grid 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 the maximum. (3) For larger grids / 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 grid sizes and numbers of colours in order to check the correctness of the programs, and as well collect data which could be used to estimate running times and maximum expected distance for larger subgrid dimensions. For 7 and 8 colourings our results corresponded well with previous work on the problem [2, 11].

For 6 colourings, we checked $1 \times k$ subgrids for $k \leq 5$, as well as 2×2 and 2×3 subgrids; 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 2 non-equivalent positions with respect to the rim, the corner (position 0, 8 rim vertices are adjacent to this position) and the side (position 1, 2 rim vertices adjacent). For each X, Y with v in position 0, we obtained a coupling satisfying:

$$H(C_X, C_Y) \leq 0.5118309760.$$

For each X, Y with v in position 1, we obtained a coupling satisfying:

$$H(C_X, C_Y) \leq 0.4837863092.$$

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 $a \times b$ subgrid, we see that the overall expected number of vertices on which X', Y' differ is at most

$$1 \times (n - 2(a+b))/n + 0 \times ab/n + \sum_i mult_i \times H_i/n,$$

a smaller value than that used in the proof of Theorem 1, where we (implicitly) used $(\max_i H_i) \times \sum_i mult_i$ rather than $\sum_i mult_i \times H_i$. Our programs actually compute this smaller value. Even so, we could not obtain suitable couplings for any grid 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 [23].

Suppose Q is a Markov chain whose mixing time we would like to bound, and $\tilde{Q}$ is another Markov chain for which we already have a bound. Let E and $\tilde{E}$ and denote the edges of these Markov chains, i.e. the pairs (x, y) such that the transition probabilities $Q(x, y)$ and $\tilde{Q}(x, y)$, respectively, are positive. Now, for each edge of $\tilde{Q}$, i.e. each $(x, y) \in \tilde{E}$, choose a fixed path $\gamma_{x,y}$ using the edges of Q: that is, choose a series of states $x = x_1, x_2, \ldots, x_k = y$ such that $(x_i, x_{i+1}) \in E$ for $1 \leq i < k$. Denote the length of such a path $|\gamma_{x,y}| = k$. Furthermore, for each $(z, w) \in E$, let $\Gamma(z, w) \subseteq \tilde{E}$ denote the set of pairs (x, y) such that $\gamma_{x,y}$ uses the edge (z, w). Finally, let

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

Note that A depends on our choice of paths.

By combining bounds on the mixing time in terms of the spectral gap [6, 25, 26] with an upper bound on Q's spectral gap in terms of $\tilde{Q}$'s due to Diaconis and Saloff-Coste [5], we obtain the following upper bound on Q's mixing time:

Theorem 4 *Let Q and $\tilde{Q}$ be reversible Markov chains on q-colourings of a graph of n vertices whose unique stationary distribution is the uniform distribution. Let $\tilde{\lambda}_1$ be the largest eigenvalue of $\tilde{Q}$'s transition matrix smaller than 1, let τ_ϵ and $\tilde{\tau}_\epsilon$ denote the ϵ-mixing time of Q and $\tilde{Q}$ respectively, and define A as above. Then for any $\epsilon \leq 1/4$,*

$$\tau_\epsilon \leq \frac{4 \log q}{\tilde{\lambda}_1} A n \tilde{\tau}_\epsilon \ .$$

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 $\log \pi(x)^{-1}$ apart, where π is the uniform distribution. Since there are at most q^n colourings, we have $\log \pi(x)^{-1} \leq n \log q$. On the grid, it is easy to see that there are an exponentially large number of q-colourings for $q \geq 3$, so removing this factor of n would require a different comparison technique.

Now suppose that $\tilde{Q}$ is the block dynamics and Q is the Glauber or Kempe chain dynamics. We wish to prove Theorem 2 by showing that $\tau_\epsilon = O(n\tilde{\tau}_\epsilon)$. By adding self-loops with probability greater than $1/2$ to the block dynamics, we can ensure that the eigenvalues of $\tilde{Q}$ 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 site 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 sites, 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 site 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 [4]. We leave this as a direction for further work.

5 Conclusion: can we reach smaller q?

We have run our programs on 2×4 and 3×3 subgrids to see if we could achieve rapid mixing on 5 colours, but in both cases the largest values of $H(C_X, C_Y)$ were too high. It does seem likely that rapid mixing on 5 colours is possible by recolouring a 3×4 subgrid, based on the decrease of the ratio of $\max E[dist] \cdot |R|$ to $|S|$ as the dimensions increase; similar reasoning leads us to believe that rapid mixing using a $2 \times k$ subgrid is possible, but we would probably need a 2×10 grid or larger to achieve success.

Unfortunately, doing the calculations for 3×4 is a daunting proposition. The problem is exponential in two directions at once (number of rim colourings, and number of grid colourings for each rim colouring), so we get huge increases in running time when we move up a level. For 3×4 subgrids, we know we have approximately 10 to 20 million rim colourings to consider per position, and an estimated average of 275,000 grid colourings for each rim colouring. Timing results for a short prelimary run on 3×4 suggest that full calculation of the greedy bound on $\max E[dist]$ for each rim colouring would require years to complete. However, we note that if there was some way we could remove a majority of the rim colourings from consideration the problem would again become tractable, since we only need to consider those rim colourings which would potentially have maximum expected distance. Currently we are working on a filter that would effectively isolate a small subset of "badly behaved" rim colourings in (hopefully) $\mathcal{O}(|R|)$ time per rim colouring considered.

As for rapid mixing on 4 colours, this does not seem possible with our procedure using any subgrid size. The preliminary data suggests that ratio of $\max E[dist] \cdot |R|$ to $|S|$ converges to a value greater than 1, though admittedly we do not have a great deal of information to extrapolate from.

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 in the Appendix.

Recently, L. Goldberg, R. Martin and M. Paterson have reported obtaining the main result of this paper independently.

References

[1] D. Aldous, *Random walks on finite groups and rapidly mixing Markov chains.* Séminaire de Probabilités XVII 1981/82 (A. Dold and B. Eckmann, eds), Springer Lecture Notes in Mathematics **986** (1986) 243–297.

[2] R. Bubley and M. Dyer. *Path coupling: a technique for proving rapid mixing in Markov chains.* Proc. 28th Ann. Symp. on Foundations of Computer Science (1997) 223–231.

[3] R. Bubley, M. Dyer, and C. Greenhill. *Beating the* 2Δ *bound for approximately counting colourings: A computer-assisted proof of rapid mixing.* Proc. 9th Ann. ACM-SIAM Symposium on Discrete Algorithms (1998) 355–363.

[4] F. Cesi. *Quasi-factorization of the entropy and logarithmic Sobolev inequalities for Gibbs random fields.* Probability Theory and Related Fields **120** (2001) 569–584.

[5] P. Diaconis and L. Saloff-Coste. *Comparison theorems for reversible Markov chains.* Annals of Applied Probability **6** (1996) 696–730.

[6] P. Diaconis and D. Stroock. *Geometric bounds for eigenvalues of Markov chains.* Annals of Applied Probability **1** (1991) 36–61.

[7] M. Dyer and A. Frieze. *Randomly colouring graphs with bounds on girth and maximum degree.* Proceedings of the 32nd Annual Symposium on Foundations of Computer Science (2001).

[8] M. Dyer, L. Goldberg, C. Greenhill, M. Jerrum and M. Mitzenmacher, *An extension of path coupling and its application to the Glauber dynamics for graph colourings,* Proceedings of the 11th Annual ACM-SIAM Symposium on Discrete Algorithms (2000), 355–363.

[9] M. Dyer and C. Greenhill. *Random walks on combinatorial objects.* Surveys in Combinatorics, 1999, eds. J.D. Lamb and D.A. Preece. Cambridge University Press, 1999, 101–136.

[10] M. Dyer and C. Greenhill. *A more rapidly mixing Markov chain for graph colourings.* Random Structures and Algorithms **13** (1998) 285–317.

[11] M. Dyer, C. Greenhill and M. Molloy. *Very rapid mixing of the Glauber dynamics for proper colourings on bounded-degree graphs.* Random Structures and Algorithms (to appear).

[12] M. Dyer, A. Sinclair, E. Vigoda, and D. Weitz. *Mixing in time and space for lattice spin systems: a combinatorial view.* Proc. RANDOM (2002) 149–163.

[13] S. J. Ferreira and A. D. Sokal. *Antiferromagnetic Potts models on the square lattice: a high-precision Monte Carlo study.* J. Stat. Phys. **96** (1999) 461–530.

[14] L.A. Goldberg, R. Martin and M. Paterson. *Random sampling of 3-colourings in* $\mathbb{Z}^2$. Random Structures and Algorithms (to appear).

[15] T. Hayes. *Randomly coloring graphs of girth at least five.* Proc. 35th ACM Symposium on Theory of Computing (2003).

[16] M. Jerrum, *A very simple algorithm for estimating the number of* k*-colourings of a low-degree graph.* Random Structures and Algorithms **7** (1995), 157–165.

[17] M. Jerrum, *Mathematical foundations of the Markov chain Monte Carlo method.* Probabilistic Methods for Algorithmic Discrete Mathematics, eds. M. Habib, C. McDiarmid, J. Ramirez-Alfosin and B. Reed. Springer, 1998, 116–165.

[18] M. Luby, D. Randall and A.J. Sinclair, *Markov chain algorithms for planar lattice structures.* SIAM J. Computing **31** (2001) 167–192.

[19] F. Martinelli. *Lectures on Glauber dynamics for discrete spin models.* Lectures on probability theory and statistics, Saint-Flour 1997, Springer Lecture Notes in Mathematics **1717** (1999) 93-191.

[20] M. Molloy. *Very rapidly mixing Markov Chains for 2Δ-colourings and for independent sets in a 4-regular graph.* Random Structures and Algorithms **18** (2001) 101–115.

[21] C. Moore, M.G. Nordahl, N. Minar, and C. Shalizi, *Vortex dynamics and entropic forces in antiferromagnets and antiferromagnetic Potts models.* Physical Review E **60** (1999) 5344–5351.

[22] C. Moore and M. E. J. Newman, *Height representation, critical exponents, and ergodicity in the four-state triangular Potts antiferromagnet.* Journal of Statistical Physics **99** (2000) 661–690.

[23] D. Randall and P. Tetali. *Analyzing Glauber dynamics by comparison of Markov chains.* J. Mathematical Physics **41** (2000) 1598–1615.

[24] J. Salas and A. D. Sokal. *Absence of phase transition for antiferromagnetic Potts models via the Dobrushin uniqueness theorem.* Journal of Statistical Physics **86** (1997) 551–579.

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

[26] A.J. Sinclair. *Algorithms for random generation and counting: a Markov chain approach.* Birkhauser, Boston, 1993, pp. 47–48.

[27] A.D. Sokal. *A personal list of unsolved problems concerning lattice gases and antiferromagnetic Potts models.* Talk presented at the conference on Inhomogeneous Random Systems, Université de Cergy-Pontoise, January 2000. Markov Processes and Related Fields **7** (2001) 21–38.

[28] E. Vigoda. *Improved bounds for sampling colorings.* J. Mathematical Physics **41** (2000) 1555–1569.

[29] J.-S. Wang, R.H. Swendsen and R. Kotecký. Physical Review Letters **63** (1989), 109–.

[30] J.-S. Wang, R.H. Swendsen and R. Kotecký, Physical Review B **42** (1990), 2465–.

A Proof of Theorem 4

First, we recall the following relationship between the mixing time of a Markov chain and its second eigenvalue. For simplicity we restrict ourselves to the case where the Markov chain converges to the uniform distribution.

Theorem 5 ([6, 25, 26]) *Let Q denote a reversible Markov chain over a set of states G whose unique stationary distribution is the uniform distribution on G, and let $1 = \lambda_0 > \lambda_1 \geq \lambda_2 \geq \cdots$ denote the eigenvalues of its transition matrix. Then its mixing time is bounded above and below by*

$$\frac{\lambda_1}{2(1-\lambda_1)} \log(\epsilon^{-1}/2) \;\leq\; \tau(\epsilon) \;\leq\; \frac{1}{1-\lambda_1} \log \epsilon^{-1} \times \log |G| \;\; .$$

This bounds the mixing time in terms of the spectral gap $1 - \lambda_1$. Using the notation in the text, we have the following bound on Q's spectral gap in terms of $\tilde{Q}$'s:

Theorem 6 ([5, 23]) *Let Q and $\tilde{Q}$ be Markov chains whose unique stationary distribution is the uniform distribution. Let λ_1 and $\tilde{\lambda}_1$ denote their second-largest eigenvalues. Then using the notation above,*

$$1 - \lambda_1 \geq \frac{1}{A}\,(1 - \tilde{\lambda}_1) \ .$$

Theorem 4 follows immediately by combining Theorems 5 and 6, setting $\epsilon \leq 1/4$, and noting that for q-colourings on a graph of n vertices we have $|G| \leq q^n$.