Markov Chain Algorithms for Planar Lattice Structures (Extended Abstract)

Michael Luby* Dana Randall† Alistair Sinclair‡

Abstract

Consider the following Markov chain, whose states are all domino tilings of a $2n \times 2n$ chessboard: starting from some arbitrary tiling, pick a 2×2 window uniformly at random. If the four squares appearing in this window are covered by two parallel dominoes, rotate the dominoes in place. Repeat many times. This process is used in practice to generate a random tiling, and is a key tool in the study of the combinatorics of tilings and the behavior of dimer systems in statistical physics. Analogous Markov chains are used to randomly generate other structures on various two-dimensional lattices. This paper presents techniques which prove for the first time that, in many interesting cases, a small number of random moves suffice to obtain a uniform distribution.

1 Introduction

This paper is concerned with algorithmic problems of the following type: given a simply connected region S of the two-dimensional Cartesian lattice (chessboard), generate uniformly at random a tiling of S with non-overlapping dominoes. (Each domino covers two adjacent squares of the lattice.) This problem arises in statistical physics, where the tilings correspond to configurations of a so-called *dimer system* on S (see, e.g., [7]). Various physical properties of the system are related to the expected value, over the uniform distribution, of some function defined over configurations, such as the number of horizontal dominoes or the correlation between the orientation of dominoes at two given squares.

Evidently, an algorithm for randomly generating configurations allows such expectations to be estimated to any desired accuracy. It also enables one to formulate and test more detailed properties of a "typical" configuration, such as the Arctic Circle Conjecture, which was originally based on observations of random configurations but has recently acquired the status of a theorem [10].

*International Computer Science Institute, 1947 Center Street, Berkeley CA 94704. Email: `luby@icsi.berkeley.edu`. Supported in part by NSF Grant CCR-9304722 and US-Israel Binational Science Foundation Grant No. 92-00226.

†Institute for Advanced Study, Princeton NJ 08540. Email: `randall@cs.princeton.edu`. Supported in part by NSF Grant DMS 9304580 and the Sloan Foundation.

‡Computer Science Division, University of California, Berkeley CA 94720–1776. Email: `sinclair@cs.berkeley.edu`. Supported in part by NSF Grant CCR-9505448 and a UC Berkeley Faculty Research Grant.

A host of other problems of physical and combinatorial interest center around the properties of random structures of various kinds on a two-dimensional lattice. Further examples that we shall consider in this paper are lozenge tilings of a triangular lattice (corresponding to a dimer system with a different underlying geometry), and Eulerian orientations of a Cartesian lattice (the six-point ice model). In all cases, algorithms that randomly generate configurations are the major experimental tool available to researchers interested in the properties of such systems.

Returning to our first example, here is the algorithm that is most widely used in practice to generate a random domino tiling of a region S. Starting from an arbitrary tiling, pick a 2×2 window uniformly at random. If the four squares in this window are covered by two parallel dominoes, rotate the dominoes in place (see figure 1). Repeat this operation a large number of times. The resulting tiling should then be (almost) random.

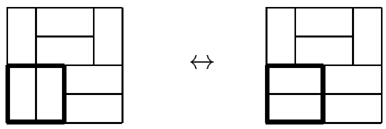


Figure 1: Domino rotations

The fact that this process (a Markov chain on the set of tilings) is connected (i.e., that every tiling is reachable from every other by a sequence of moves of the above kind) is a beautiful result of Thurston [17]. However, no non-trivial upper bound is known on the number of moves needed to achieve a random tiling. In practice, this number is decided by appealing to combinatorial intuition, or experimentally by some form of stopping rule. What is lacking is an analysis of the rate of convergence of the Markov chain to the uniform distribution, which would supply an *a priori* bound on the number of moves. Similar Markov chains, based on analogous local moves, are used to generate other two-dimensional lattice structures in the same way. Like the dominoes chain, they have so far resisted analysis.

In this paper, we develop a unifying combinatorial framework that allows several Markov chains of this kind to be analyzed for the first time. There are two essential ingredients. The first, which we believe to be of independent combinatorial interest, is to establish a 1-1 correspondence between the configurations on a lattice region S and objects which we call *routings* on a related lattice. Informally, a routing is a collection of vertex-disjoint (or edge-disjoint) paths crossing S from left to right. (See Section 2 for precise definitions and examples.) Such a correspondence was already known, at least implicitly, for two of the systems we consider. We extend the correspondence to domino tilings, and identify the common structure of routings as a key ingredient in the analysis of the associated Markov chains.

The second ingredient is to interpret natural Markov chains like the one above on domino tilings in terms of the associated routings. As we shall see, elementary moves on configurations correspond to natural local perturbations of the routings (such as displacing one vertex along a path). The crucial feature of this translation is that, when viewed in the routings world, the rate of convergence of the Markov chain turns out to be amenable to a simple and elegant analysis using coupling arguments. In fact, this analysis leads us to generalize slightly the class of random moves allowed for routings; these in turn map back to natural non-local moves for the configurations themselves. As a result, we obtain new, non-local versions of the Markov chains which are provably "rapidly mixing" (i.e., converge quickly to the uniform distribution), and are also seemingly more efficient than the original versions.

The upshot of all this is low-degree polynomial bounds on the convergence time of these Markov chains for all three of the examples mentioned above. We therefore provide the first rigorous justification for experiments that use short simulations of the chains in order to generate random configurations.

We should mention that these problems can be solved by alternative approaches. A combinatorial trick known as the Gessel-Viennot method [6] allows one to count lattice routings by evaluating a suitable determinant. In conjunction with self-reducibility properties, this allows one to generate configurations uniformly at random (see, e.g., [16]). Other Markov chain algorithms which can be applied to these structures in arbitrary graphs are given in [9, 13] (for tilings) and [14] (for Eulerian orientations). However, in the important special case of planar lattices, the algorithms in this paper have better time bounds than these other methods, and are simpler, more natural and quite widely used in practice. Moreover, simulation studies suggest that our bounds are in fact quite pessimistic, and we are confident that they can be improved with a more careful analysis. We also point out that there is a simple experiment one can perform which provides a reliable estimate of the true convergence rate: this can be used to dramatically reduce the number of simulation steps required in practice.

Finally, we highlight another aspect of our work which we believe to be of wider interest. Up to now, most applications of Markov chains to the design of combinatorial algorithms have utilized indirect geometric techniques based on expansion of the underlying graph to analyze the mixing rate (see, e.g., [12] for a survey). In contrast, our analysis is based on the simpler and more direct probabilistic approach of coupling. Although this technique has long been known in the statistics community, attempts to apply it to the kind of complex Markov chains that arise in Computer Science have generally been unsuccessful (though for another interesting recent exception, see [8]).

The remainder of the paper is organized as follows. In Section 2 we illustrate the correspondence between lattice configurations and routings for each of our examples: lozenge tilings, domino tilings and Eulerian orientations. In Section 3, we show how to analyze the rate of convergence of the natural Markov chain on lozenge tilings by applying a coupling argument to the corresponding chain on routings. In the process, we will enrich the chain with non-local moves. In Section 4 we sketch how to apply similar technology to Markov chains for domino tilings and Eulerian orientations. Owing to space limitations, we concentrate on giving a flavor of the arguments and defer several details to the full paper.

2 Lattice routings

In this section, we consider several important examples of lattice structures and illustrate the correspondence between them and collections of paths which we call *routings*. Lattice routings are a special case of so-called *plane partitions*, which are widely studied in combinatorics. For some of these examples in isolation, the correspondence is already known; however, our aim here is to present the correspondences in a unified setting. As we shall see later, this will make the analytical tools we develop rather generally applicable.

2.1 Lozenge tilings

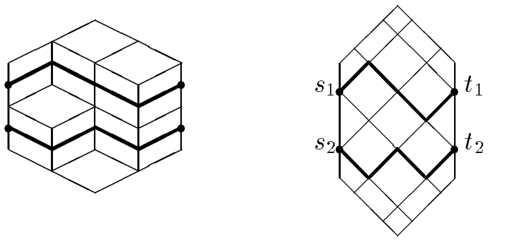
The first structures we consider are lozenge tilings of a finite region of the triangular lattice: we discuss these first because the correspondence with routings is most direct here. A *lozenge* is the analogue of a domino in the Cartesian lattice: each lozenge covers two adjacent triangles in the lattice, and has three possible orientations. Lozenge tilings are configurations of a dimer system on this lattice. As explained in the Introduction, we are interested in the problem of generating a random lozenge tiling of the given region.

The routings corresponding to lozenge tilings are defined on an associated Cartesian lattice. The vertices of this lattice are the midpoints of the vertical edges of the triangular lattice; two vertices are connected by an edge if they lie on adjacent triangles. For a finite, simply-connected region S of the triangular lattice, we designate each vertex of the Cartesian lattice that lies on the boundary of S as either a source or a sink: a vertex v is a *source* if the interior of S is on the right of v, and a *sink* otherwise. If a lozenge tiling of S exists, the numbers of sources and sinks are necessarily equal. Following the boundary in counterclockwise order, starting at a source, we label the sources $\{s_1, \ldots, s_k\}$ and each sink t_i, where s_i was the last unmatched source we labeled.

A *lozenge routing* of S is a set of k non-intersecting shortest paths on the Cartesian lattice within S from s_i to t_i for each i. Note that the length of the path from s_i to t_i is the same in every routing.

Figure 2 provides a pictorial illustration of the correspondence between lozenge tilings and routings for a typical region S. This correspondence has been formalized in general by Sachs *et al*:

Theorem 1 ([1, 11]) *The set of lozenge tilings of S corresponds bijectively with the set of lozenge routings of S.* $\square$



A lozenge tiling The corresponding routing

Figure 2: Lozenge tilings and routings

2.2 Domino tilings

A domino tiling is a covering of a finite region of the Cartesian lattice with dominoes, where each domino covers two adjacent squares of the region. Domino tilings are configurations of dimer systems on this lattice. As in the case of lozenge tilings, there is a family of routings which correspond bijectively to the set of domino tilings. Again, the routings are defined on an associated lattice, which in this case is triangular.

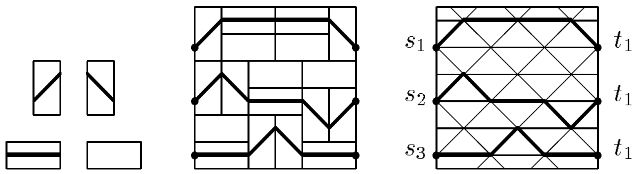
To specify this lattice, first color the squares of the Cartesian lattice black and white as on a chessboard. The vertices of the triangular lattice are the centers of vertical edges that have a black square to their right. Finally, connect a vertex (x, y) to $(x+1, y+1)$, $(x+1, y-1)$ and $(x+2, y)$ to complete the triangular lattice. Sources and sinks are defined in similar fashion to the lozenge case: sources are boundary vertices with the interior of S to their right, and sinks those with the interior to their left. Once again, we pair up sources $\{s_1, \ldots, s_k\}$ and sinks $\{t_1, \ldots, t_k\}$ in the obvious way. A *domino routing* of S is then a collection of non-intersecting shortest paths on the triangular lattice within S from s_i to t_i for each i.

The correspondence between domino tilings and routings is illustrated by means of an example in figure 3: the reader may enjoy convincing himself that a given tiling defines a unique routing using the three permitted paths through the dominoes as shown. This correspondence is formalized in the next theorem.

Theorem 2 *There is a bijection between domino tilings of S and domino routings of S.*

Proof. Figure 3 indicates how to use paths through dominoes to map tilings to routings, as follows. Start at a source vertex s_i. By definition, there must be a black square in the interior of S to the right of s_i. The domino occupying this square determines the first step of our path: we connect s_i to the unique point on the right boundary of the domino which is a vertex of the underlying triangular lattice. We now find ourselves at a new point with a black square to our right, and we can repeat this process. Since we migrate to the right in each step, we eventually hit a point on the right boundary of S which has a black square to its right, and thus must be a sink. The paths must be non-intersecting because the tiles cannot overlap.

To see that the above map is bijective, we will construct the inverse map from routings to tilings. Each path starts at a source s_i (which has a black square to its right) and follows lattice edges to a sink t_i. As we follow the path from left to right, we tile each of the black squares to the right of the lattice points on the path (except the sink t_i). (There are three possible positionings for each tile, corresponding to the three possible types of edge the path can pass through.) Since the paths are non-intersecting, our tiles cannot overlap and we are left with a partial tiling of S. Now there is a unique way to tile the remaining parts of the region, namely using only horizontal tiles (whose left half covers a white square). To see this, consider any untiled black square. The white square to its left must be untiled, for if it were tiled there would be a path exiting its right boundary, and then the black square would be tiled. So every black square can be tiled with the right half of a horizontal domino. This completes the tiling since there must be an equal number of black and white squares in S. The uniqueness comes from the fact that each row of each untiled region starts with a white square, so we cannot complete the tiling if we use any vertical tiles. $\square$



Paths through dominoes A domino tiling The corresponding routing

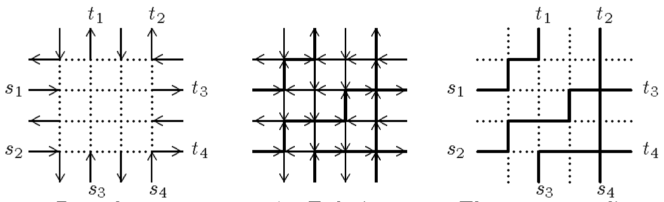
Figure 3: Domino tilings and routings

2.3 Eulerian orientations

Another important set of structures which can be identified with lattice routings are the Eulerian orientations of a region of the Cartesian lattice with specified boundary conditions. An *Eulerian orientation* of an undirected graph is an orientation of its edges so that the in-degree of every vertex is equal to its out-degree. In this problem, the input is a finite region S of the two-dimensional Cartesian lattice, together with a fixed orientation for each of the edges that connects the boundary of S with the interior: these orientations are the *boundary conditions*. Our task is to generate uniformly at random an orientation of the edges in the interior such that all interior vertices have equal in-degree and out-degree. This is the six-point model in statistical mechanics, also known as the "ice model."

The correspondence between Eulerian orientations and a suitable class of routings is well known in the physics community (see, e.g., [3]), and is sketched in figure 4. The sources and sinks in this case are defined by the boundary conditions, as shown.

An *Eulerian routing* of S is a set of shortest paths in S from sources $\{s_1, ..., s_k\}$ to sinks $\{t_1, ..., t_k\}$ on the boundary. The paths are permitted to intersect at a vertex but *not* along an edge. As illustrated in figure 4, to get the Eulerian routing corresponding to a given Eulerian orientation, we construct the paths from edges that are oriented up and to the right.



Boundary conditions — An Eulerian orientation — The corresponding routing

Figure 4: Eulerian orientations and routings

Remark. An alternative unifying framework for these problems is provided by *height functions*, which map two-dimensional structures to three-dimensional surfaces. The correspondence between tilings and height functions is a beautiful result of Thurston, Conway and Lagarias [4, 17] and was previously known for Eulerian orientations (see, e.g., [3]). Lattice routings, which we have derived directly from the underlying combinatorial structures, can also be interpreted as the level sets of appropriate height functions. Although the three-dimensional surfaces supplied the geometric intuition suggesting the Markov chains we study, it is the two-dimensional routings which capture the essential information that admits analysis. □

3 Generating lozenge tilings

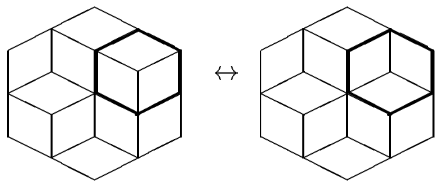
This section is devoted to an analysis of a very natural Markov chain algorithm for generating random lozenge tilings. The analysis will exploit in a crucial way the correspondence with routings established in Section 2.1. We present this example first because it is the most straightforward to deal with. Analogous Markov chains for generating the other structures mentioned in Section 2 can be analyzed by more refined applications of the same techniques (see Section 4).

3.1 The Markov chain

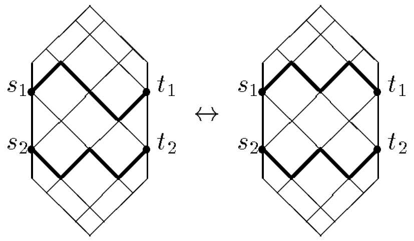
In the Introduction, we discussed a Markov chain on domino tilings based on a local move that rotates a pair of adjacent dominoes. The analogous Markov chain for lozenge tilings has as its local move a rotation of three neighboring lozenges (see figure 5(a)). As in the domino case, this chain can also be shown to be connected, and to converge to the uniform distribution over tilings.

Let us now interpret this Markov chain in the world of routings. It is clear that a lozenge rotation induces a natural local move on the corresponding routing, in which a "peak" or a "valley" is inverted by moving two edges (see figure 5(b)).

It turns out that the Markov chain in the routings world becomes considerably easier to analyze if *every* peak and valley can give rise to a rotation: note that this is not the case for the above chain, since sometimes when we try to invert a point the move will be blocked by the presence of another path. (Recall that the paths in a routing are not allowed to cross.) This motivates the introduction of a more general set of



(a) Tiling rotation

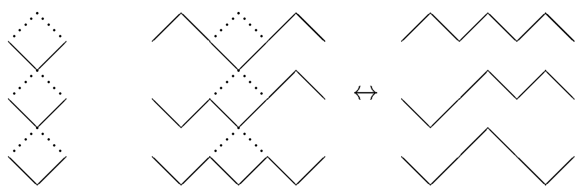


(b) Routing rotation

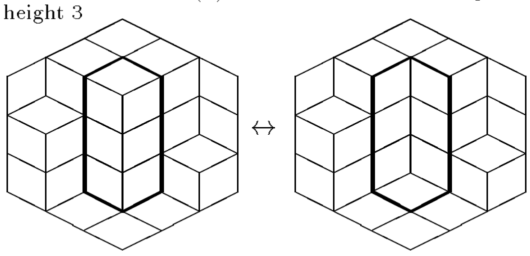
Figure 5: Lozenge rotations

moves in which a *tower* is rotated. The original moves will simply correspond to the special case of rotating a tower of height 1.

In the routings lattice, define the *cell at* (x, y) to be the edges connecting $(x, y), (x+1, y+1), (x, y+2)$ and $(x-1, y+1)$. A *tower of height* h is a connected set of cells at the points $(x, y), (x, y+2), \ldots, (x, y+2h-2)$. We call the points (x, y) and $(x, y+2h)$ the *bottom* and *top* of the tower respectively (see figure 6(a)).



(a) Tower of height 3 — (b) Tower rotation in routing



(c) Tower rotation in tiling

Figure 6: A move in the Markov chain $\mathcal{M}_{loz}$ for lozenge routings, and its counterpart for tilings

Let R_1 and R_2 be lozenge routings of the region S. We define a Markov chain $\mathcal{M}_{loz}$ in which there is a move from R_1 to R_2 if and only if the symmetric difference of the edges of R_1 and R_2 is a tower. The transition probabilities $P(\cdot\,,\cdot)$ of $\mathcal{M}_{loz}$ are defined by

$$P(R_1,R_2)=\begin{cases}1/2Nh, & \text{if } R_1\oplus R_2 \text{ is a tower of height } h;\\ 1-\sum_{R\neq R_1}P(R_1,R), & \text{if } R_2=R_1,\end{cases}$$

where N is the total number of internal vertices along all the paths in any routing. Figure 6(b) illustrates such a move (rotation of a tower of height 3). Figure 6(c) shows that these generalized moves have a natural counterpart in the original tilings world.

Notice that we may implement a move of $\mathcal{M}_{loz}$ as follows. Given a routing R, choose an internal point p on one of the paths in R, and a number $r\in[0,1]$ uniformly at random. Assume first that $r\leq 1/2$. If p is a valley then it is the bottom of a unique tower of height h; in this case, if $r\leq 1/2h$ rotate the tower. On the other hand, if $r>1/2$ check to see if p is a peak (and hence the top of a unique tower of height h), and rotate this tower if $r\geq 1-1/2h$. In all other cases do nothing. This somewhat unusual implementation is a technical device that will prove useful later when we define couplings in Section 3.3.

It is not hard to check that this Markov chain is ergodic and reversible, and that it converges to the uniform distribution over all lozenge routings. Therefore, we can generate a random tiling by simulating $\mathcal{M}_{loz}$ for sufficiently many steps, starting from an arbitrary routing, and outputting the tiling corresponding to the final routing. The efficiency of this algorithm depends on the number of simulation steps necessary to ensure an (almost) uniform distribution, or equivalently on the *rate of convergence* of the Markov chain. We shall see in Section 3.3 that a small number of steps suffice. First, however, we need some technology.

3.2 Coupling and the convergence rate

In this subsection, we establish some general machinery for bounding the rate of convergence of Markov chains, which we shall use repeatedly in the remainder of the paper. Consider an ergodic Markov chain $\mathcal{M}$ with finite state space Ω, transition matrix P and stationary distribution π. Following standard practice, for any given initial state x, we shall measure the deviation of the distribution $P^t(x,\cdot)$ at time t from π by the *variation distance*:

$$\Delta_x(t)=\tfrac{1}{2}\sum\nolimits_{y\in\Omega}|P^t(x,y)-\pi(y)|.$$

The *mixing time* of the Markov chain is defined by the function

$$\tau(\epsilon)=\max_x\min\{t:\Delta_x(t')\leq\epsilon \text{ for all } t'\geq t\}.$$

Our strategy for bounding $\tau(\epsilon)$ is to construct a *coupling* for the Markov chain, i.e., a stochastic process $(X_t,Y_t)_{t=0}^{\infty}$ on $\Omega\times\Omega$ with the properties:

1. Each of the processes X_t and Y_t is a faithful copy of $\mathcal{M}$ (given initial states $X_0=x$ and $Y_0=y$).
2. If $X_t=Y_t$, then $X_{t+1}=Y_{t+1}$.

The idea here is the following. Although each of X_t,Y_t, viewed in isolation, behaves exactly like $\mathcal{M}$, they need not be independent; on the contrary, we will construct a joint distribution for the two processes in such a way that they tend to *couple*, or move closer together. By the second condition above, once they have coupled they must remain coupled at all future times.

The expected time taken for the processes to couple provides a good bound on the mixing time of $\mathcal{M}$. To state this formally, for initial states x,y set

$$T^{x,y}=\min\{t:X_t=Y_t\mid X_0=x,\,Y_0=y\},$$

and define the *coupling time* to be $T=\max_{x,y}\mathrm{E}T^{x,y}$. The following result relating the mixing time to the coupling time is standard (see, e.g., [2]):

Theorem 3 $\tau(\epsilon)\leq 6T(1+\ln\epsilon^{-1})$. □

Next, we introduce some machinery that will help us to bound the coupling time. Suppose we have a distance function ϕ defined on $\Omega\times\Omega$ such that ϕ takes integer values in the range $[0,..,B]$, and $\phi(x,y)=0$ iff $x=y$. In our examples, where the states of the Markov chain are lattice routings, ϕ will be a natural measure of the "area" between a pair of routings. We will measure the distance between a pair of processes (X_t,Y_t) using the stochastic process $\Phi(t)=\phi(X_t,Y_t)$. Our strategy will be to show that, under a suitably defined coupling, the expected change $\Delta\Phi$ in Φ is always non-positive; intuitively, this should enable us to conclude that the coupling time is small. The following lemma makes this intuition precise.

Lemma 4 *With the above notation, suppose the coupling satisfies* $\mathrm{E}\big[\Delta\Phi(t)|X_t,Y_t\big]\leq 0$ *and, whenever* $\Phi(t)>0$, $\mathrm{E}\big[(\Delta\Phi(t))^2|X_t,Y_t\big]\geq V$. *Then the expected coupling time from initial states* x,y *satisfies* $\mathrm{E}T^{x,y}\leq\Phi(0)\big(2B-\Phi(0)\big)/V$.

Proof. Define the stochastic process $Z(t)=\Phi(t)^2-2B\Phi(t)-Vt$. It is easy to check that

$$\begin{aligned}&\mathrm{E}\big[Z(t+1)|Z(t),\ldots,Z(0)\big]-Z(t+1)\;=\\ &2\big(\Phi(t)-B\big)\mathrm{E}\big[\Delta\Phi(t)|X_t,Y_t\big]\\ &\qquad+\big(\mathrm{E}[(\Delta\Phi(t))^2|X_t,Y_t]-V\big)\;\geq\;0,\end{aligned}$$

and hence $Z(t)$ is a submartingale. Moreover, the random time $T^{x,y}=\min\{t:\Phi(t)=0\}$ (where $X_0=x$, $Y_0=y$) is a stopping time for $Z(t)$ with finite expectation, and the differences $|Z(t+1)-Z(t)|$ are bounded. This allows us to apply the Optional Stopping Theorem for submartingales (see, e.g., [5, Chap. 4, Thm. 7.5]), and conclude that $\mathrm{E}Z(T^{x,y})\geq\mathrm{E}Z(0)$. From the definition of $Z(t)$, this implies that

$$-V\,\mathrm{E}T^{x,y}\geq\Phi(0)^2-2B\Phi(0),$$

which gives the desired upper bound on $\mathrm{E}T^{x,y}$. □

3.3 $\mathcal{M}_{loz}$ is rapidly mixing

We first consider the simplified case of routings consisting of a single path P with source s and sink t. Notice that in this case all towers have height 1.

We define a coupling as follows. Consider two copies of the Markov chain, whose states are the paths P_1 and P_2, each containing N internal points. Then in one move we choose $(i, r) \in_u \{1, \ldots, N\} \times [0, 1]$ and simultaneously move the ith points of each of P_1 and P_2 as specified by the random number r. It should be clear that this is a valid coupling. Notice that, because we are using the same value of r for both processes, the two paths will never move in opposite directions. This was the reason we implemented the Markov chain in this fashion in Section 3.1.

We now need to bound the expected time it takes the coupled process to cause any two initial routings to agree. To simplify the problem, we use an observation due to Propp and Wilson [15]: if the state space of a Markov chain forms a distributive lattice, and the coupling preserves the partial order, then the coupling time is bounded above by the expected time for coupling to occur when the initial states are the largest and smallest elements of the lattice. It is easy to see that our set of paths forms a partial order, in which $P_1 \succeq P_2$ iff the ith point of P_1 lies on or above the ith point of P_2 for all i. The following lemma, whose straightforward proof is omitted, establishes that our coupling does indeed preserve this ordering.

Lemma 5 *If, under the above coupling, the pair (P_1, P_2) moves to (P_1', P_2'), and if $P_1 \succeq P_2$, then $P_1' \succeq P_2'$.* □

We now proceed to bound the time taken for the two extremal paths to couple. To do this, we introduce a distance function as in lemma 4. For a pair of paths P_1, P_2, define the distance $\phi(P_1, P_2)$ to be the area (i.e., number of lattice squares) of the region between P_1 and P_2. The crucial observation is that this distance will tend not to increase under our coupling, as the next lemma shows. By lemma 5, we may restrict attention to the case $P_1 \succeq P_2$.

Lemma 6 *Let P_1 and P_2 be any paths such that $P_1 \succeq P_2$. Then $\mathrm{E}[\Delta\Phi | P_1, P_2] \leq 0$.*

Proof. Consider an arbitrary pair of paths $P_1 \succeq P_2$. The typical situation is as depicted in figure 7, with path P_1 (the "red" path) drawn as a solid line and path P_2 (the "blue" path) as a dotted line. We can partition the paths into segments $C_1, \ldots, C_\ell$ on which the two paths coincide, and segments $D_1, \ldots, D_m$ whose endpoints coincide, but for which the red path is strictly above the blue path at all intermediate points.

Consider first a segment C_i. It is clear that, if the point chosen by the coupling lies in this segment, then the point will move with the same probability on both paths, so the paths will still coincide and there will be no change in area.

Now consider a segment D_i, in which the red points are strictly above the blue points (except at the endpoints). On the red (upper) path, label the peaks "good" (G) and the valleys "bad" (B), and vice versa for the blue path. Thus a point is good if a rotation at that point would cause the area between the paths to decrease, and bad if a rotation would case the area to increase. (The boundary of the region might in fact prohibit some of these bad rotations.) It is easy to see that, on each path, the labeled points in the interior of the segment D_i are alternately good and bad, with a net excess of one good point. Moreover, each endpoint of D_i contributes at most one bad point (unless D_i and D_{i+1} meet at a point, in which case there might be two bad points, but we can assign one to each segment). Summing over all segments of the path, we see that the number of good points is greater than or equal to the number of bad points. Since each good or bad point is equally likely to be chosen in the coupling, the expected change in area is at most zero. □

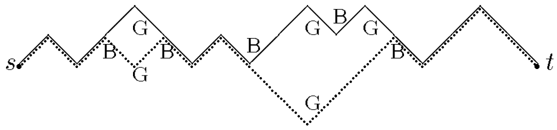


Figure 7: Proof of lemma 6

Theorem 7 *Let S be a region with one source s and one sink t. Then the mixing time of the Markov chain $\mathcal{M}_{loz}$ on lozenge routings of S satisfies $\tau(\epsilon) \leq 12n^3(1 + \ln \epsilon^{-1})$, where n is the area of S.*

Proof. By theorem 3, it suffices to show that the coupling time T satisfies $T \leq 2n^3$. To bound T we appeal to lemma 4, restricting attention to extremal initial states following lemma 5. Our distance function ϕ clearly takes integer values in the interval $[0..n]$. Moreover, we have seen in lemma 6 that $\mathrm{E}[\Delta\Phi] \leq 0$. It remains only to bound $\mathrm{E}[(\Delta\Phi)^2]$, assuming that the area between the pair of paths is nonzero. In this case, there must be a segment on which the red path is strictly above the blue path (using the terminology in the proof of lemma 6). Scanning this segment from left to right, there is a first point where at least one of the paths is good and neither is bad. If this point is chosen at the next time step, then there will be a decrease in area with probability at least $\frac{1}{2}$. Hence Φ decreases strictly with probability at least $1/2n$, so $\mathrm{E}[(\Delta\Phi)^2] \geq 1/2n$. Plugging all of these quantities into lemma 4 yields $T \leq 2n^3$. □

We now extend the above argument to the case of lozenge routings with multiple paths. The coupling we use is the following obvious generalization of our earlier one. Given a pair of routings, choose the same random point p on both routings, say the ith point of the jth path, and the same random bit $r \in_u [0, 1]$. Then update each routing by rotating at point p with the appropriate probability as determined by the random number r.

As in the single-path case, we can argue that it is sufficient to bound the expected coupling time for a pair of extremal routings. It is easy to see that these routings again form a distributive lattice and therefore have a unique highest and lowest element. The partial order is defined in the natural way: we say that a routing $R_1 \succeq R_2$ iff the ith path of R_1 lies on or above the ith path of R_2, for all i. Since all routings of a region S have the same number of paths, this is well-defined. Again, it can be checked that the coupling preserves this ordering:

Lemma 8 *If, under the above coupling, the pair* (R_1, R_2) *moves to* (R_1', R_2')*, and if* $R_1 \succeq R_2$*, then* $R_1' \succeq R_2'$. □

This ensures that we need only consider coupling extremal routings.

For a pair of routings R_1, R_2, we define the distance $\phi(R_1, R_2)$ to be the sum of the areas between corresponding paths in R_1 and R_2. The next lemma, a generalization of lemma 6, proves that the distance tends not to increase under the coupling.

Lemma 9 *Let* R_1 *and* R_2 *be any two routings such that* $R_1 \succeq R_2$*. Then* $\mathrm{E}[\Delta\Phi | R_1, R_2] \le 0$.

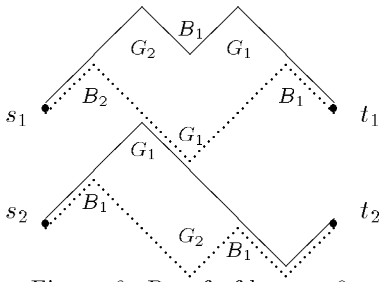


Figure 8: Proof of lemma 9

Proof. Consider a pair of routings, the upper one red and the lower one blue. Adopting the same terminology as in the proof of lemma 6, give a point the label G_i if it defines a tower of height i and is a good point (i.e., choosing it could reduce the area between the two routings). Similarly, label a point B_i if it defines a bad tower of height i (see figure 8). We can calculate the expected change in area by first considering the expected change if we make a move on the red routing only, and then adding to this the expected change in area from a move on the blue routing only. Each point labeled G_i has probability $1/2Ni$ of rotating, and such a rotation decreases the area between the two routings on *each* of the i paths altered by the rotation, a total decrease of i. Summing the expected changes in area over all good and bad points, we have

$$\begin{aligned} \mathrm{E}[\Delta\Phi | R_1, R_2] &= \sum_i \sum_{B_i} \frac{i}{2Ni} - \sum_j \sum_{G_j} \frac{j}{2Nj} \\ &= \frac{1}{2N} \Big(\sum_i \#B_i - \sum_j \#G_j \Big) \le 0, \end{aligned}$$

where $\#B_i$ is the number of points labeled B_i. The final inequality follows from the fact that on each path the number of good points is at least as large as the number of bad points, as argued in lemma 6. □

It is now a short step to the main theorem of this section, which confirms that simulating the Markov chain $\mathcal{M}_{loz}$ provides a reasonable algorithm for generating a random lozenge tiling.

Theorem 10 *Let* S *be a region of the triangular lattice. The mixing time of the Markov chain* $\mathcal{M}_{loz}$ *on lozenge routings of* S *satisfies* $\tau(\epsilon) \le 12n^4(1 + \ln \epsilon^{-1})$*, where* n *is the area of* S.

Proof. The maximum total area between any pair of routings is $n^{1.5}$, so the distance function ϕ takes values in the range $[0, .., n^{1.5}]$. Lemma 9 confirms that $\mathrm{E}[\Delta\Phi] \le 0$. To get a bound on $\mathrm{E}[(\Delta\Phi)^2]$, consider a pair of routings with non-zero area between them. Then there must be a pair of corresponding paths, one red and one blue, and a segment in which the red path is strictly above the blue path. Scanning this segment from left to right, call the first good point we reach on either path p. There is a $1/N$ chance of choosing the point p, and we perform the rotation at p with probability $1/2h$, where h is the height of the tower defined by p. Rotating this tower causes a decrease of h in the total area, one unit for each path included in the tower. Hence we can conclude that $\mathrm{E}[(\Delta\Phi)^2] \ge h^2/2Nh \ge 1/2n$. Putting all this together, and appealing to lemmas 4 and 8 we see that the coupling time satisfies $T \le 2n^4$. The result now follows from theorem 3. □

Remarks. (a) We have made no attempt here to tune the upper bound on the mixing time in theorem 10. This can be improved by more refined arguments, which we defer to the full paper.

(b) In practice, we can run an experiment to bound the coupling time using stopping rules as in [15]. We can run the coupled process starting with the two extremal elements in the set of routings to bound the coupling time T. Experimentally, this has been found to yield considerably tighter bounds; theorem 10 provides, for the first time, an *a priori* bound on the running time of this experiment. □

4 Generating other structures

The machinery presented in the last section provides a general framework which can be applied to the random generation of other lattice structures, including domino tilings and Eulerian orientations (and presumably others). In each case, the development of a provably efficient algorithm follows the same outline: we start with natural local moves connecting the space of configurations which are believed by experimenters to be efficient. We then interpret these moves in terms of the appropriate routings, enrich them with a small set of non-local moves (involving towers), and use a coupling argument to argue that the resulting Markov chain is rapidly mixing: in all these cases, the mixing time is bounded by a low-degree polynomial in the area of the region. The definition of towers is sensitive to the type of routing, and the proofs use slightly more sophisticated arguments. It is interesting that, in each case, as for lozenge tilings, the choice of towers is quite natural in the original setting as well. Owing to limited space, we will only sketch the main ideas here; the details can be found in the full paper.

4.1 Domino tilings

Our task is to construct a random domino tiling of a given region S of the Cartesian lattice. In theorem 2 we showed that this is equivalent to constructing a random domino routing of a region of the triangular lattice with sources $\{s_1, ..., s_k\}$ and sinks $\{t_1, ..., t_k\}$. This we achieve using a Markov chain on the space of routings, whose moves correspond to rotations of suitably defined towers. In this case there are four different types of towers, as indicated in figure 9. The *height* of a domino tower is the number of unit triangles in the tower. Let N be the number of edges in any domino routing of S. The transition probabilities $P(\cdot,\cdot)$ of $\mathcal{M}_{dom}$ are:

$$P(R_1, R_2) = \begin{cases} 1/2Nh, & \text{if } R_1 \oplus R_2 \text{ is a tower of height } h; \\ 1 - \sum\limits_{R \neq R_1} P(R_1, R), & \text{if } R_2 = R_1. \end{cases}$$

To implement one step of this Markov chain, starting at a routing R, choose $(e,r) \in_u R \times [0,1]$. First suppose $r \leq 1/2$. If the edge e is a diagonal directed up and to the right, check to see if there is a tower starting at e and extending in the north-west direction. Notice that this must be a tower of Type I or Type II, and is unique. If on the other hand e is horizontal, check whether there is a tower of Type III or Type IV starting at e and extending north-west (again this is unique). In either case, determine the height h of the tower, and if $r \leq 1/2h$ rotate the tower. The case when $r > 1/2$ is similar: check to see whether there is a tower in the south-east direction starting at e. If e is a diagonal pointing up and to the right this would be a tower of Type II or Type IV; if e is horizontal it would be a tower of Type I or Type III. In either case rotate the tower if $r > 1 - 1/2h$, where h is the height of the tower.

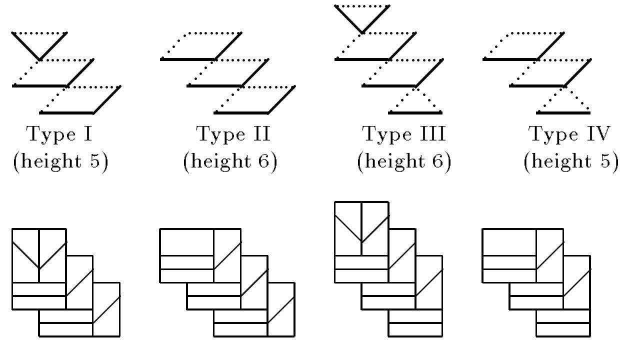


Figure 9: Domino towers

The Markov chain $\mathcal{M}_{dom}$ is ergodic and converges to the uniform distribution over all domino routings of a region S. We can define a coupling in exactly analogous fashion to that for lozenge routings in Section 3.3. Once again it is not hard to check that this coupling preserves the obvious partial order on domino routings, so we may again restrict attention to extremal routings when analyzing the coupling time. The key fact is the following analogue of lemma 9, which says that the area $\Phi(t)$ between routings (defined in the obvious way) is non-increasing in expectation.

Lemma 11 *Let R_1 and R_2 be two domino routings such that $R_1 \succeq R_2$. Then* $\mathrm{E}[\Delta\Phi | R_1, R_2] \leq 0$.

Proof (sketch). First consider the case when R_1 and R_2 each consist of just a single path. Then, following the ideas from lemmas 6 and 9, we can consider segments $C_1, \ldots, C_\ell$ on which the two paths coincide, and segments $D_1, ..., D_m$ on which the two paths coincide at the endpoints, and on which R_1 is strictly above R_2 at all intermediate points. As before, it is clear that choosing an edge on any of the C_j will not change the area between the paths.

Now consider a region D_j. Give an edge the label G_i if, when that edge is chosen by the coupling, the area between the routings could increase by exactly i. Similarly, label an edge B_i if one of the directions could cause the area to decrease. When the routings consist of a single path, i is either 1 or 2. Certain edges will receive two labels. See figure 10.

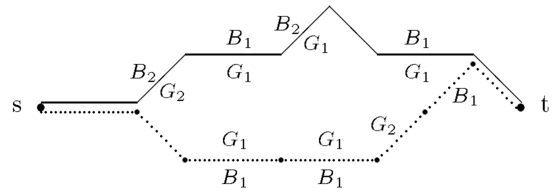


Figure 10: Proof of lemma 11 — single path case

We can match each bad edge to a unique good edge as follows. Each horizontal edge with a bad label also has a good label, so we can pair them off. Any diagonal segment which has a bad label at one end must have a good label at the other end, so we can pair these. (The moves of $\mathcal{M}_{dom}$ are defined so that we can only rotate the routing at the two extremal edges of a diagonal segment.) This tells us that the number of bad edges is no greater than the number of good edges. As in lemma 9, the probability of rotating at a given edge is inversely proportional to the change in area caused by such a rotation. This ensures that the expected change in area is always non-positive.

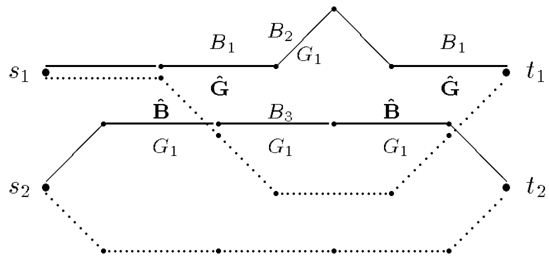


Figure 11: Proof of lemma 11 — multiple path case

To handle multiple paths, we need to modify the above labeling slightly. It is possible that edges which would be labeled good or bad when a path is viewed in isolation might no longer be places where we can perform a rotation. In particular, if horizontal edges from adjacent paths get too close, one will not be able to rotate in the north-west direction, and the other will not be able to rotate in the south-east direction. This will have the net effect of eliminating a good and a bad label. To see this, label an edge $\hat{G}$ if it is a horizontal edge which cannot move in the good direction because of interference with another path, and label it $\hat{B}$ if it cannot move in the bad direction. Any edge which cannot be rotated because the resulting rotation would go outside of the boundary of S is also labeled $\hat{B}$. These labels are shown in figure 11 for the solid paths only. The crucial point is that every edge labeled $\hat{G}$ must be paired with a distinct edge labeled $\hat{B}$. Hence $\#\hat{G} \leq \#\hat{B}$.

Generalizing the argument from the single path case, we have that

$$\#\hat{B} + \sum_i \#B_i \leq \#\hat{G} + \sum_j \#G_j,$$

whence

$$\sum_i \#B_i \leq \sum_j \#G_j.$$

From this we can conclude that the expected change in area is always non-positive. $\square$

Lemma 11, in conjunction with lemma 4, yields:

Theorem 12 *Let S be a region of the Cartesian lattice. Then the mixing time of the Markov chain $\mathcal{M}_{dom}$ on domino routings of S satisfies $\tau(\epsilon) \leq 12n^4(1+\ln\epsilon^{-1})$, where n is the area of S.* $\square$

4.2 Eulerian orientations

Let S be a region with specified boundary conditions for Eulerian orientations: recall that these determine the sources and sinks. To generate a random Eulerian orientation of S, we construct a Markov chain $\mathcal{M}_{eul}$ whose state space is the set of all Eulerian routings on S with these sources and sinks. In similar fashion to the case of lozenge routings, we allow a move between two routings if they differ by a structure which is either a vertical or horizontal tower, as depicted in figure 12. The transition probabilities $P(\cdot,\cdot)$ of $\mathcal{M}_{eul}$ are

$$P(R_1,R_2) = \begin{cases} 1/2Nh, & \text{if } R_1 \text{ and } R_2 \text{ differ by a tower of ht } h; \\ 1 - \sum\limits_{R \neq R_1} P(R_1,R), & \text{if } R_2 = R_1, \end{cases}$$

where N is the number of vertices on any routing.

To implement one step of $\mathcal{M}_{eul}$, starting at a routing R, choose $(p,r) \in_u R \times [0,1]$. If $r \leq 1/2$, check to see whether p is the lower-right point of a vertical or horizontal tower (of height h), and rotate the tower if $r < 1/2h$ and this yields a legal routing. Similarly, if $r > 1/2$, check to see whether p is the upper-leftmost vertex on a vertical or horizontal tower, and rotate this tower if $r > 1 - 1/2h$ and this yields a legal routing. Note that whenever p is part of both a vertical *and* a horizontal tower we will in fact do nothing: rotating either tower would force edges on two paths to coincide, and thus fail to produce a legal routing.

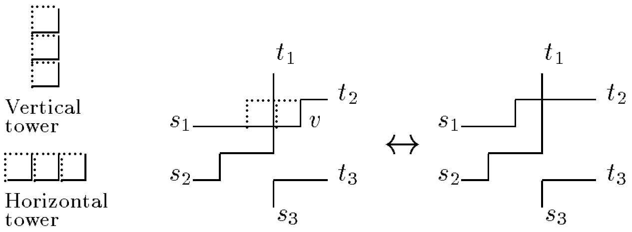


Figure 12: Markov chain for Eulerian orientations

Tower rotations have a simple interpretation in terms of Eulerian orientations. A tower of height h corresponds to a directed cycle in the Eulerian orientation which is a $1 \times h$ rectangle (with its internal edges aligned). The rotation corresponds to reversing all the orientations around this cycle, which of course produces a new Eulerian orientation.

Note that, when the boundary conditions define a set of routings consisting of a single path, the Markov chain is equivalent to the lozenge routing chain $\mathcal{M}_{loz}$ of Section 3. Theorem 7 guarantees that this chain is rapidly mixing. We now extend this result to Eulerian routings that consist of an arbitrary number of paths. As in the previous examples, we will start with the highest and lowest routings and construct a coupling which chooses corresponding points on the two routings and the same random number r. As usual, the following lemma, which says that the area $\Phi(t)$ between routings is non-increasing in expectation, is the key to the proof.

Lemma 13 *Let R_1 and R_2 be any two routings such that $R_1 \succeq R_2$. Then* $\mathrm{E}[\Delta\Phi | R_1, R_2] \leq 0$.

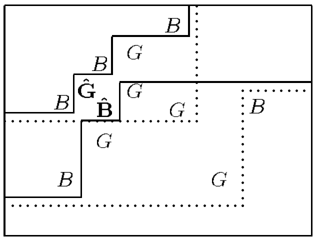


Figure 13: Proof of lemma 13

Proof (sketch). We assign labels G and B to peaks and valleys in similar fashion to the lozenge case (see figure 13). On the lower routing, label G any corner which is the lower-right of a tower, and label B any corner which is the upper-left of a tower; on the upper routing, interchange the roles of G and B. Now we will label all the corners which do not define a tower. On the lower routing, label $\hat{G}$ any unlabeled corner which goes to the right and then up, and label $\hat{B}$ any unlabeled corner which goes up and then to the right; on the upper routing, interchange the roles of $\hat{G}$ and $\hat{B}$. Generalizing from the single path case we know that $\#B + \#\hat{B} \leq \#G + \#\hat{G}$.

Next observe that each point labeled $\hat{G}$ can be paired with a distinct point labeled $\hat{B}$. A point p is labeled $\hat{G}$ for one of two reasons: either the two edges which complete the unit square defined by the peak or valley at p are part of an adjacent path, or p (and its incident edges) lie on the boundary. In the first case, the point defining the opposite corner of the square must be labeled $\hat{B}$, and we can pair this bad point (on the same routing) with p. In the second case, since we have a good point on the boundary, the corresponding point on the other routing must also be on the boundary (and share both incident edges), and therefore must be labeled $\hat{B}$; we pair this bad point with p. This pairing implies that $\#\hat{B} \geq \#\hat{G}$, and hence that $\#B \leq \#G$.

Finally, observe that the weights of the moves are chosen so that each bad and each good point contributes equally to the expected change in area. Hence, if N is the number of points on the routing, we have $\mathrm{E}[\Delta\Phi] = (\#B - \#G)/2N \leq 0$. □

Combining this with lemma 4 we obtain the following:

Theorem 14 *Let S be a region of the Cartesian lattice with specified boundary conditions. Then the mixing time of the Markov chain $\mathcal{M}_{eul}$ on Eulerian routings of S satisfies $\tau(\epsilon) \leq 12n^4(1 + \ln \epsilon^{-1})$, where n is the area of S.* □

5 Concluding remarks

The techniques discussed here also allow us to develop an algorithm for generating a random three-coloring of a rectangular region of the Cartesian lattice, another fundamental problem motivated by statistical mechanics applications. To do this, we identify the set of three-colorings with the union over all possible boundary conditions of the set of Eulerian orientations of the region. We leave this algorithm, which is based on a variant of the Markov chain $\mathcal{M}_{eul}$ of section 4.2, for the full paper.

We conclude by mentioning some directions for further work. Firstly, we suspect that the broad class of plane partition problems, i.e., planar problems which can be expressed in terms of routings, contains several other candidates for similar Markov chains and analysis. Secondly, we note that no comparable machinery exists for lattice structures in three dimensions, despite intense interest. The principle challenge here is to find an analogue of routings that captures naturally occurring three-dimensional structures: this would constitute a major breakthrough. Finally, we should mention the question of whether our arguments could be enhanced to show that the original Markov chains (i.e., our chains with rotations restricted to towers of height 1) are also rapidly mixing. Although we suspect that this would not lead to faster algorithms, it remains interesting for historical reasons.

Acknowledgments

We would like to thank James Akao for useful discussions and for implementing the Eulerian orientations algorithm.

References

[1] Al-Khnaifes, K. and Sachs, H. Graphs, linear equations, determinants, and the number of perfect matchings. In *Contemporary Methods in Graph Theory* (R. Bodendiek ed), Wissenschaftsverlag, Mannheim, 1990, pp. 47-71.

[2] Aldous, D. Random walks on finite groups and rapidly mixing Markov chains. *Séminaire de Probabilités* XVII, 1981/82, Springer Lecture Notes in Mathematics 986, pp. 243–297.

[3] Baxter, R.J. *Exactly solved models in statistical mechanics.* Academic Press, London, 1982.

[4] Conway, J. and Lagarias, J. Tilings with polyominoes and combinatorial group theory. *Journal of Combinatorial Theory A* **53**, 1990, pp. 183-208.

[5] Durrett, R. *Probability: Theory and examples.* Wadsworth & Brooks/Cole, Belmont CA, 1991.

[6] Gessel, I. and Viennot, G. Binomial determinants, paths, and hook length formulae. *Advances in Mathematics* **58**, 1985, pp. 300-321.

[7] Heilmann, O.J. and Lieb, E.H. Theory of monomer-dimer systems. *Comm. in Mathematical Physics* **25**, 1972, pp. 190–232.

[8] Jerrum, M.R. A very simple algorithm for estimating the number of k-colorings of a low-degree graph. *Random Structures and Algorithms,* to appear.

[9] Jerrum, M.R. and Sinclair, A.J. Approximating the permanent. *SIAM Journal on Computing* **18**, 1989, pp. 1149–1178.

[10] Jockusch, W., Propp, J. and Shor, P. *Random domino tilings and the arctic circle theorem.* Preprint, 1995.

[11] John, P. and Sachs, H. Calculating the numbers of perfect matchings and of spanning trees, Pauling's orders, the characteristic polynomial, and the eigenvectors of a benzenoid system. In *Topics in Current Chemistry* **153***: Advances in the Theory of Benzenoid Hydrocarbons* (M.J.S. Dewar *et al*, eds.), Springer Verlag, Berlin, 1990, pp. 145-179.

[12] Kannan, R. Markov chains and polynomial time algorithms. *Proc. 35th IEEE Symposium on Foundations of Computer Science,* 1994, pp. 656–671.

[13] Kenyon, C., Randall, D. and Sinclair, A. Matchings in lattice graphs. *Proc. 25th ACM Symposium on Theoretical Computer Science,* 1993, pp. 738-746.

[14] Mihail, M. and Winkler, P. On the number of Eulerian orientations of a graph. *Proc. 3rd ACM-SIAM Symposium on Discrete Algorithms,* 1992, pp. 138–145.

[15] Propp, J. and Wilson, D. *Exact sampling with coupled Markov chains and applications to statistical mechanics.* Preprint, 1995.

[16] Sinclair, A. *Counting and generating combinatorial structures: a Markov chain approach.* Birkhäuser, Boston, 1993.

[17] Thurston, W. Conway's tiling groups. *American Mathematical Monthly* **97**, 1990, pp. 757-773.