Markov Chain Algorithms for Planar Lattice Structures*

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 90° in place. Repeat many times. This process is used in practice to generate a random tiling, and is a widely used 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 (e.g., an $n \times n$ chessboard), generate uniformly at random a tiling of S with non-overlapping dominoes, each of which covers two adjacent squares of the lattice. This problem arises in statistical physics, where the tilings correspond to configurations of

*A preliminary version of this paper appeared in *Proceedings of the 36th IEEE Symposium on Foundations of Computer Science*, 1995, pp. 150–159.

†Digital Fountain, Inc., 600 Alabama St., San Francisco, CA 94110. Email: `luby@digitalfountain.com`. Work done while at the International Computer Science Institute. Supported in part by NSF Grant CCR-9304722 and US-Israel Binational Science Foundation Grant No. 92-00226.

‡School of Mathematics and College of Computing, Georgia Institute of Technology, Atlanta GA 30332-0160. Email: `randall@math.gatech.edu`. Supported in part by NSF Career Grant No. CCR-9703206.

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

a *dimer system* on S (see, e.g., [8]). 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. 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 Theorem [10], which began life as a conjecture based on observations of random configurations.

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, also known in the statistical mechanics community as 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.

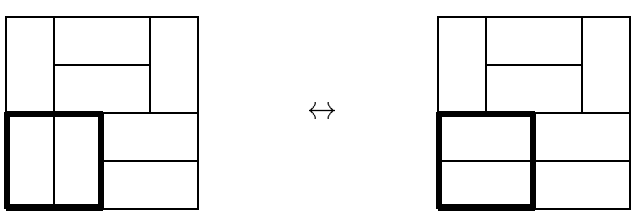


Figure 1: Domino rotations

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) follows from a beautiful result of Thurston [17]. However, no nontrivial 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 ad hoc 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 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.) These correspondences were already known, at least implicitly, but here they play an essential role 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 a coupling argument. 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).

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 [7] 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, 12] (for tilings) and [13] (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, our bounds are in fact quite pessimistic (our main concern is to introduce the methodology rather than to tune the bounds), and can be improved with a more detailed analysis (see the subsequent paper by Wilson [18]). 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, as discussed in [14].

The remainder of the paper is organized as follows. In section 2 we illustrate the correspondence between lattice configurations, routings and height functions

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 sections 4 and 5 we show how to apply the same technology to analogous Markov chains for domino tilings and Eulerian orientations.

2 Lattice routings and height functions

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*. Routings are the key for analyzing the convergence rate, or running time, of our Markov chain algorithms. As we shall see, the routings are closely related to a third representation of the lattice structures known as *height functions*, which arise from the tiling groups of Conway, Lagarias and Thurston [4, 17]. For each of our three examples we briefly outline the bijections between lattice structures, routings and height functions. As we shall see later, this framework will make the analytical tools we develop rather generally applicable.

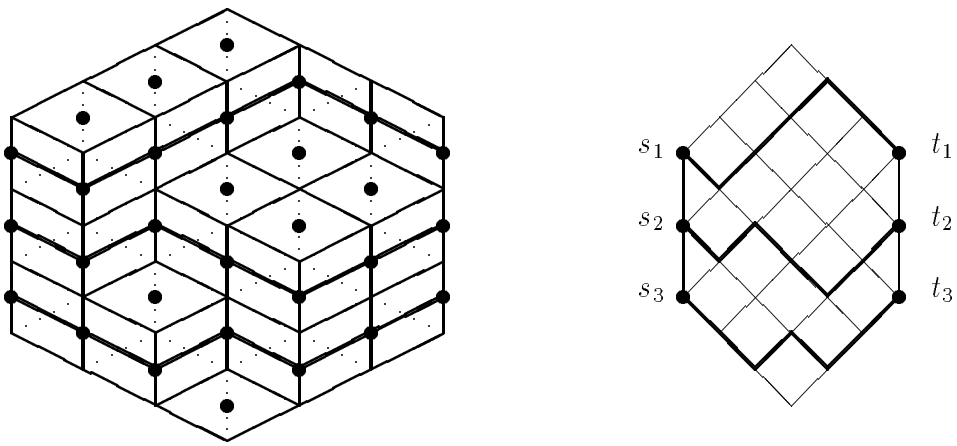
2.1 Lozenge tilings

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. Given a finite, simply connected region S of the triangular lattice, we define an associated region $\hat{S}$ of the Cartesian lattice as follows. The vertices of $\hat{S}$ correspond to the midpoints of the vertical edges in S, and two vertices in $\hat{S}$ are connected if the corresponding points in S lie on adjacent triangles. This mapping is demonstrated in figure 2.

The vertices of $\hat{S}$ that correspond to edges on the boundary of S are called *sources* and *sinks*: a vertex v is a source if the interior of $\hat{S}$ lies to the right of v, and a sink if the interior of $\hat{S}$ lies to its left. It is not hard to check that, if a lozenge tiling of S exists, then the numbers of sources and sinks are necessarily equal. Label the sources $s_1, \ldots, s_k$ and the sinks $t_1, \ldots, t_k$.

A *lozenge routing* of $\hat{S}$ is a set of k non-intersecting (i.e., vertex-disjoint) shortest paths on the Cartesian lattice within $\hat{S}$ from $\{s_1, \ldots, s_k\}$ to $\{t_1, \ldots, t_k\}$.



A lozenge tiling on S The corresponding routing on $\hat{S}$

Figure 2: Lozenge tilings and routings

It is not difficult to see that there is a bijection between lozenge tilings and lozenge routings of corresponding regions. Figure 2 provides a pictorial illustration of this correspondence for a typical region S. An easy way to see this is to "mark" the tiles containing two vertical edges with a stripe connecting the centers of these edges. Now notice that if a tile is placed next to any vertical edge of S, then it must have two vertical edges whose midpoints correspond to adjacent vertices in $\hat{S}$. Furthermore, if the vertical edge of a tile lies in the interior of S, then it must be adjacent to another tile having two vertical edges (so the markings line up). Following such a sequence of tiles, starting from each vertical edge on the boundary of S, defines a set of non-intersecting source-sink paths in $\hat{S}$, i.e., a routing. Conversely, given a routing we may invert the above construction to create a partial tiling using only marked (i.e., non-horizontal) tiles. All vertical edges of these tiles are adjacent to another tile or to the boundary of S. The untiled portion of S therefore consists of regions bounded only by non-vertical edges. It is not hard to see that these can be tiled in only one way, using only horizontal tiles. Hence there is a bijection between the set of tilings and routings of S. Moreover, we can order the sinks so that the path from s_i always ends at t_i, for all i, and the path from s_i to t_i will have the same length in every routing.

This correspondence has been formalized by Sachs *et al*:

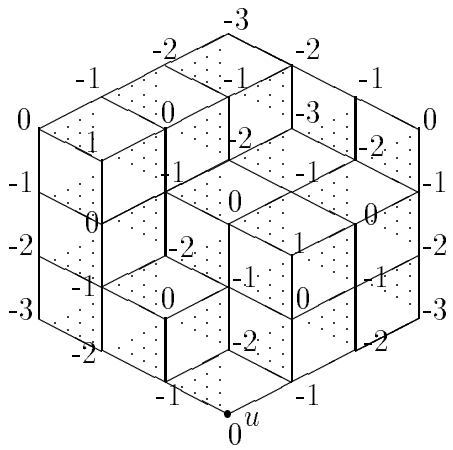
Theorem 1 ([1, 11]) *The set of lozenge tilings of S corresponds bijectively with the set of lozenge routings of $\hat{S}$.* □

The bijection between tilings and routings is closely related to the *height functions* defined by Thurston [17]. Although we don't require it for our analysis, we

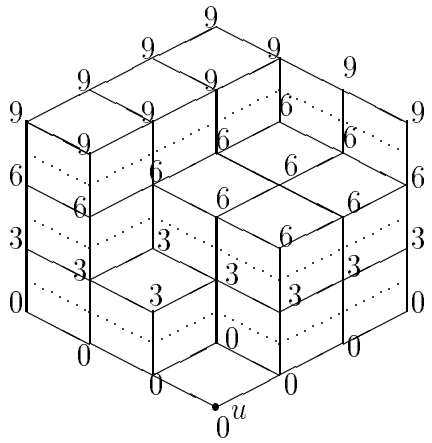
briefly describe this connection here because it sheds further light on the above correspondence. In this representation, the *height* of each vertex in the region S is determined by an underlying algebraic structure known as the tiling group, introduced by Conway and Lagarias [4]. The height function can be extended to all points within the region by a piecewise linear function, thereby defining a three dimensional surface associated with each tiling. It turns out that, by viewing each surface from a certain orientation, the paths of a routing can be interpreted as the level sets of this surface. This is a common feature of all the structures we discuss in this paper; in the case of lozenges the surfaces are immediately apparent and are just the set of three-dimensional boxes that seem to "jump out" of the two-dimensional lozenge tiling, as in the left-hand picture in figure 2.

The heights of the vertices of any tiling can be determined by the following simple rule. First choose a vertex u on the boundary of the region and fix its height $h_{loz}(u) = 0$. To determine the heights of all other vertices we rely on the bipartite structure of the dual lattice, which allows us to color all the triangles pointing to the right white and all those pointing to the left black. Starting at u, walk around the boundaries of tiles until an edge is traversed from a vertex v with known height $h_{loz}(v)$ to a vertex w whose height has not yet been determined. If the triangle to the left of this edge (w.r.t. the direction of traversal) is black, set $h_{loz}(w) = h_{loz}(v) + 1$; if the triangle to the left is white, set $h_{loz}(w) = h_{loz}(v) - 1$. Repeat this process until all of the vertices have been visited. For any tiling of a simply connected region, the heights are always unique and well-defined (up to translation). As an example, figure 3 shows the height function for the tiling of figure 2, with u being the bottom vertex. A simple consequence of the above definition is that the heights along the boundary of any tiling are determined by the region alone and are identical for all tilings of the region.

The connection between height functions and routings in the case of lozenge tilings is quite straightforward. We create an *adjusted* height function $\hat{h}_{loz}$ by letting v_y be the vertical coordinate of vertex v (i.e., the row of the lattice that it lies in) and defining $\hat{h}_{loz}(v) = h_{loz}(v) + v_y$. It is a simple exercise to verify that, for any vertex v, the adjusted height function satisfies $\hat{h}_{loz}(v) = \hat{h}_{loz}(u) + 3k$ where k is the number of paths in the routing which lie between v and u (with k being negative if these paths lie above v and positive if they lie below v); see figure 3. This can be proved using the observation that on every lozenge the heights of the two vertices where the angle is acute must be equal. This implies that horizontal lozenges, where the two vertices with acute angles lie in the same row, will have the same adjusted height at all four corners, while on the other two types of lozenges the lower two vertices will have strictly smaller height than the upper two vertices.



The height function h_{loz}



The adjusted height function $\hat{h}_{loz}$

Figure 3: Lozenge tilings and height functions

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.

Given a finite, simply connected tileable region S in the Cartesian lattice, we define a related region $\hat{S}$ (which lies on a triangular lattice). First color the squares of the Cartesian lattice black and white as on a chessboard. The vertices of the triangular lattice can be defined as the centers of all of the vertical edges of the Cartesian lattice which have a black square to their right, where edges are defined by connecting each vertex (x, y) to $(x+1, y+1)$, $(x+1, y-1)$ and $(x+2, y)$. The region $\hat{S}$ is the part of this triangular lattice which is defined by vertices and edges contained completely within the original region S. Sources and sinks of $\hat{S}$ are defined in similar fashion to the lozenge case: sources are boundary vertices with the interior of $\hat{S}$ 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 $\hat{S}$ is then a collection of non-intersecting shortest paths on the triangular lattice within $\hat{S}$ 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 4: each tiling defines a unique routing using the three permitted paths through the dominoes as shown. This correspondence is formalized in the next theorem.

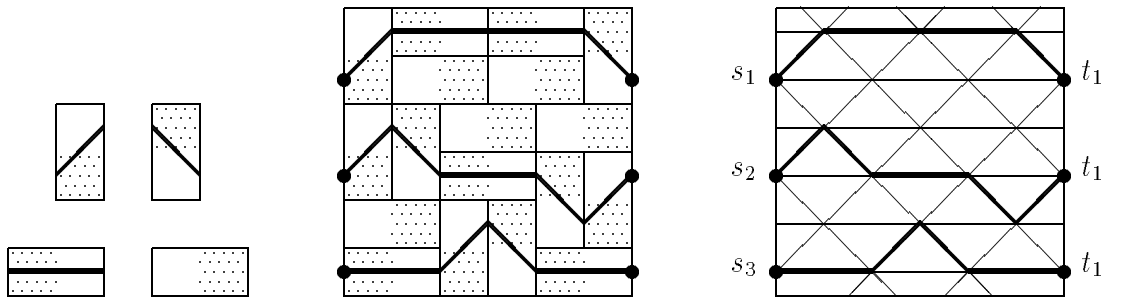
Theorem 2 *There is a bijection between domino tilings of S and domino routings of $\hat{S}$.*

Proof. Figure 4 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 the original Cartesian lattice region 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 $\hat{S}$ in 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 $\hat{S}$ 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 construct the inverse map from routings to tilings as follows. 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 the leftmost square in each row of each untiled sub-region is white, so we cannot complete the tiling if we use any vertical tiles. □

Once again, the routings defined above can be viewed as level sets of a height function. The height function which arises from the tiling group for dominoes can be summarized using a rule based on the bipartition underlying the dual lattice (i.e., the black and white squares of the chessboard). To define the height function h_{dom}, start with some point u on the boundary and assign $h_{dom}(u) = 0$. Now, walking along edges bounding the tiles starting at u, if the square to the left of an edge is black (respectively, white), increase (respectively, decrease) the height by one. An example is illustrated in figure 5, with u being the bottom left vertex.

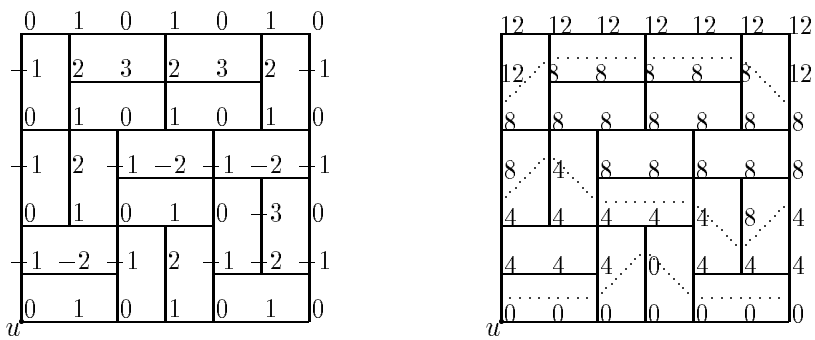
The connection between height functions and routings is analogous to that for lozenge tilings. For any point $v = (v_x, v_y)$, let v_y be the y coordinate of v (i.e., the row of the lattice). Let $\mathrm{par}(v) = 1$ if v is the lower left corner of a white square, and let $\mathrm{par}(v) = 0$ if v is the lower left corner of a black square. Now



Paths through dominoes A domino tiling The corresponding routing

Figure 4: Domino tilings and routings

define the adjusted height function by $\hat{h}_{dom}(v) = -h_{dom}(v) + 2v_y + \mathrm{par}(v)$. It is easy to verify that $\hat{h}_{dom}(v)$ is equal to $\hat{h}_{dom}(u) + 4k$, where k is the number of paths which lie between u and v in the routing (k being negative if these paths lie below u). Figure 5 illustrates this relationship.



The height function h_{dom} The adjusted height function $\hat{h}_{dom}$

Figure 5: Domino tilings and height functions

2.3 Eulerian orientations

A third 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 simply connected 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 6. The sources and sinks in this case are defined by the boundary conditions, as shown. Sources are the vertices on the boundary which are connected to the interior of the region by edges directed towards the interior and which point up or to the right; sinks are the boundary vertices which are connected to the interior by edges directed away from the interior (i.e., towards the boundary) and which also point up or to the right. A necessary condition for the existence of an Eulerian orientation is, of course, that the number of sources and sinks are equal.

An *Eulerian routing* of S is a set of shortest paths in S from sources $\{s_1, \ldots, s_k\}$ to sinks $\{t_1, \ldots, t_k\}$ on the boundary. The paths are permitted to intersect at a vertex but *not* along an edge. As illustrated in figure 6, to get the Eulerian routing corresponding to a given Eulerian orientation, we construct the paths only from edges that are oriented up and those that are oriented to the right. It is straightforward to establish that there is a bijection between the set of Eulerian orientations of a region S and the set of Eulerian routings of S.

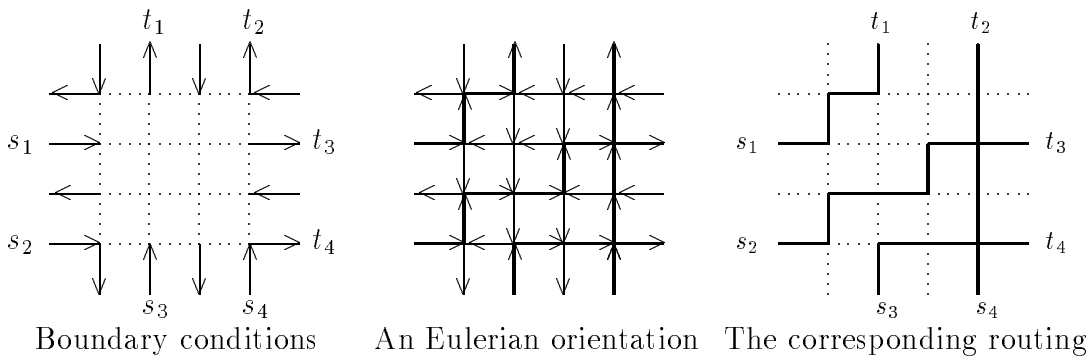
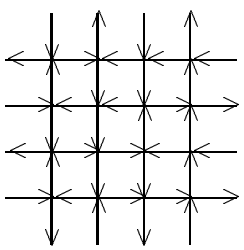


Figure 6: Eulerian orientations and routings

The height function associated with Eulerian orientations is an assignment of integers to the center of each face within a region such that neighboring faces differ in height by one. In the statistical physics community, this is known as a "solid-on-solid surface." To define the heights, start with a face adjacent to the boundary, and assign to its center u the height $h_{eul}(u) = 0$. To define the heights of the other faces, walk along edges in the dual lattice. When traversing a dual edge (v, w), where the height of v is already defined, let $h_{eul}(w) = h_{eul}(v) + 1$ if

the edge of the Eulerian orientation which was crossed points to the right (relative to the path from v to w) and let $h_{eul}(w) = h_{eul}(v) - 1$ if the edge points to the left. Again this height function is unique (up to translation) and well-defined. If we define an adjusted height function $\hat{h}_{eul}(v) = h_{eul}(v) - v_x + v_y$, where v_x and v_y are the Cartesian coordinates of v, then $\hat{h}_{eul}(v) = \hat{h}_{eul}(u) + 2k$ where k is the number of paths of the routing which lie between u and v. See figure 7 for an illustrative example.



Eulerian orientation

$\leftrightarrow$

0	1	0	1	0
1	0	1	2	1
0	1	2	1	0
1	0	1	2	1
0	1	0	1	0

Solid-on-solid surface

$\leftrightarrow$

8	8	6	6	4
8	6	6	6	4
6	6	6	4	2
6	4	4	4	2
4	4	2	2	0

The adjusted height function $\hat{h}_{eul}$

Figure 7: The height function representation of an Eulerian orientation

Remark. The solid-on-solid height function demonstrates the well-known connection between Eulerian orientations and three-colorings of a lattice region, since taking the values of the heights mod 3 always gives a valid three-coloring of the region (see, e.g., [3]). If we fix the colors of all the dual vertices on the boundary of the region (and hence the boundary conditions), then there is a bijection between three-colorings and Eulerian orientations. □

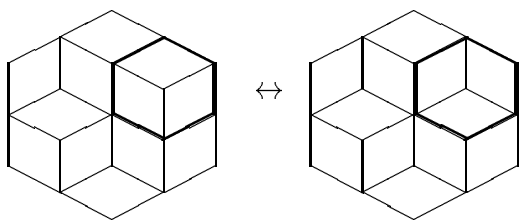
3 Generating lozenge tilings

This section is devoted to an analysis of a 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 discussed in section 2 can be analyzed by more refined applications of the same techniques, as we will demonstrate in section 4.

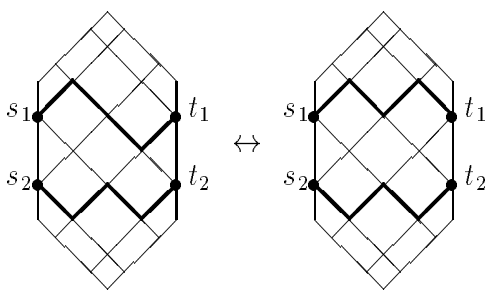
3.1 The Markov chain

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 8(a)). As in the domino case, this chain can also be shown to be connected and to converge to the uniform distribution over tilings (see section 3.2 below).



(a) Tiling rotation



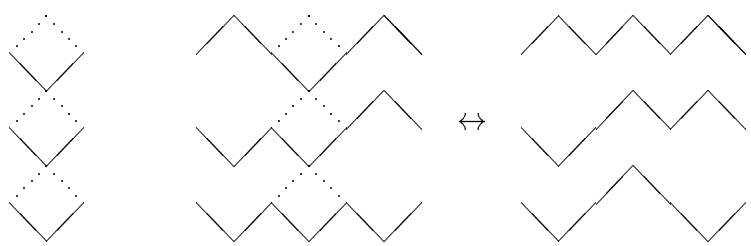
(b) Routing rotation

Figure 8: Lozenge rotations

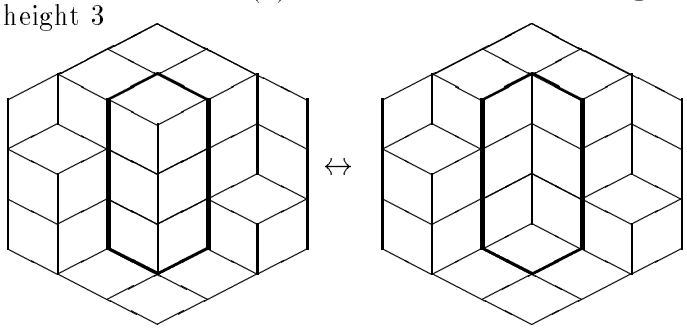
Let us now interpret this Markov chain in the world of routings. It is clear that a lozenge rotation induces a natural local move on the corresponding routing, in which a "peak" or a "valley" is inverted by switching two edges (see figure 8(b)). So we can think of the chain as picking a point uniformly at random on the routing and inverting it if possible.[1] 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 intersect. See, e.g., the second valley on the lower path in the left-hand routing of figure 8(b).) This motivates the introduction of a more general set of moves in which a *tower* is rotated. The original moves will simply correspond to the special case of rotating a tower of height 1.

[1]Strictly speaking, this chain differs slightly from the original one in that it does not attempt to make rotations at points that are not on the current paths, which will always be rejected. This merely reduces the self-loop probabilities of the chain, and hence speeds up the mixing time by a factor that is bounded by the area of the region.

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 $(x, y), (x, y+2), \ldots, (x, y+2h-2)$, where either the points $(x, y), (x, y+2), ..., (x, y+2h-2)$ are all valleys and the point $(x, y+2h)$ does not lie on the routing, or the points $(x, y+2h), (x, y+2h-2), ..., (x, y+2)$ are all peaks and the point (x, y) does not lie on the routing. We call the points (x, y) and $(x, y+2h)$ the *bottom* and *top* of the tower respectively (see figure 9(a)). Provided both the top and bottom of a tower lie in the region, its peaks (respectively, valleys) can be inverted by switching pairs of edges in each of its cells. Such an operation is called a *rotation* of the tower. Figure 9(b) illustrates a rotation of a tower of height 3, and Figure 9(c) shows that tower rotations have a natural counterpart in the original tilings world.



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



(c) Tower rotation in a tiling

Figure 9: A move in the Markov chain $\mathcal{M}_{loz}$ for lozenge routings (and tilings)

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

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

where N is the total number of internal vertices along all the paths in any routing.

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, say); in this case, if $r \leq 1/2h$ then rotate the tower if possible (i.e., if the top of the tower lies in the region). On the other hand, if $r > 1/2$ check whether p is a peak (and hence the top of a unique tower), and if possible rotate this tower if $r \geq 1 - 1/2h$, where h is the height of the tower. In all other cases do nothing. This slightly unusual implementation is a technical device that will prove useful later when we define a coupling for the Markov chain.

As we shall see in section 3.2, this Markov chain is ergodic and 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 3.4 that a small number of steps suffice, or in other words that the Markov chain is "rapidly mixing." In preparation for this we will introduce some general technology in section 3.3.

3.2 Ergodicity of $\mathcal{M}_{loz}$

The fact that $\mathcal{M}_{loz}$ converges to the uniform distribution over tilings follows almost immediately from the fact that the chain is connected, i.e., every state is reachable from every other. It is actually quite straightforward to show that the Markov chain based on *simple* rotations connects the state space of all lozenge tilings. This is sufficient to show the connectedness of the Markov chain based on towers since it includes all the simple moves. Here we sketch a proof whose machinery will be useful to us in other ways.

Lemma 3 *The state space of the Markov chain $\mathcal{M}_{loz}$ is connected.*

Proof. It is conceptually easier to work in the routings world. Note that there is a natural partial order on the set of all lozenge routings of a given region $\hat{S}$, defined as follows. Let R_1, R_2 be two routings, and let P_1, P_2 be a pair of corresponding paths (i.e., having the same source and sink) in R_1, R_2 respectively. We say that $P_1 \succeq P_2$ iff the ith point of P_1 lies on or above the ith point of P_2, for all i. We say that $R_1 \succeq R_2$ iff the relation $P_1 \succeq P_2$ holds for all pairs of corresponding paths P_1, P_2. Since all routings of a region $\hat{S}$ have the same number of paths, this relation is well-defined. If $R_1 \succeq R_2$, we define the distance between them to be the total area enclosed between all pairs of corresponding paths. Now it is not too hard to see that, for any finite, simply connected region $\hat{S}$, there is a unique

minimum routing $R_\perp$, such that $R \succeq R_\perp$ for all routings R of $\hat{S}$.[2] We will show that every routing is connected to $R_\perp$ by a sequence of simple rotations (i.e., of towers of height 1) each of which decreases the distance to $R_\perp$ by one.

Let $R \neq R_\perp$ be an arbitrary routing. Starting with the lowest pair of corresponding paths in $R, R_\perp$ and working upwards, scan left to right along the paths until the first point x at which $R, R_\perp$ deviate from one another. Now continue to follow these two paths (P and $P_\perp$, say) until the first point y where they meet again (this must happen, at the latest, at their common sink). Notice that the union of the segments of P and $P_\perp$ that lie between x and y form a circuit whose interior is entirely contained within the interior of $\hat{S}$, since $\hat{S}$ is simply connected. In addition, since $R_\perp$ is minimal, P must lie strictly above $P_\perp$ along the segment delimited by these two points. Therefore, P must have at least one peak along this segment; let v be the leftmost such peak. It follows that v must be the top of a (rotatable) tower of height 1 in R, since v lies strictly above the corresponding point in $R_\perp$ and R coincides with $R_\perp$ on all lower paths. So if we rotate the tower at v we decrease the distance from $R_\perp$ by one unit. Applying this argument repeatedly, we arrive at $R_\perp$. □

Theorem 4 *The Markov chain $\mathcal{M}_{loz}$ is ergodic and converges to the uniform distribution over lozenge tilings.*

Proof. The Markov chain is clearly aperiodic since it has a holding probability of at least 1/2 in every state. Together with the previous lemma, this implies that the chain is ergodic, i.e., converges to a unique stationary distribution. That this stationary distribution is uniform follows from the fact that the transition probabilities are symmetric: for any pair of adjacent routings R_1, R_2, we have $P(R_1, R_2) = P(R_2, R_1) = 1/2Nh$, where h is the height of the tower by which R_1 and R_2 differ. □

3.3 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*:

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

[2] A careful proof of the existence of a unique highest tiling (in the height function representation) was given by Thurston [17]; this corresponds precisely to the minimum routing defined here.

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 move closer together. By the second condition above, once they have met they must remain together at all future times.

The expected time taken for the processes to meet 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 5 $\tau(\epsilon) \leq T\mathrm{e}\lceil \ln \epsilon^{-1} \rceil$. $\square$

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

Proof. Define the stochastic process $Z(t) = \Phi(t)^2 - 2B\Phi(t) - VT(t)$, where $T(t) = t$ if $\Phi(t) > 0$, and $T(t) = \min\{t' : \Phi(t') = 0\}$ if $\Phi(t) = 0$. Then $Z(t+1) - Z(t) = 0$ whenever $\Phi(t) = 0$, and when $\Phi(t) > 0$ we have

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

Hence $Z(t)$ is a submartingale with respect to the sequence $\{(X_t, Y_t)\}_{t\geq 0}$. 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., [6, 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 $\mathrm{E}T^{x,y}$. □

3.4 $\mathcal{M}_{loz}$ 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 $(i, r) \in \{1, \ldots, N\} \times [0, 1]$ uniformly at random, 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 proceed to bound the expected time it takes the coupled process to cause any two initial routings to agree. To simplify the analysis, we use an observation due to Propp and Wilson [14]: if the state space of a Markov chain is endowed with a partial order with unique maximum and minimum elements, and if the coupling preserves the partial order (in a sense made precise below), then the coupling time is bounded above by the expected coupling time starting from the maximum and minimum states. Recall the partial order defined in the proof of lemma 3: $P_1 \succeq P_2$ iff the ith point of P_1 lies on or above the ith point of P_2 for all i. Recall also that this partial order has a unique minimum (and similarly also a unique maximum) element. To make precise the notion of preserving the partial order, we first extend the above coupling to a random function f on

the entire state space: namely, pick (i, r) as above and, for any path P, let $f(P)$ be the path obtained by moving the ith point of P as specified by r. We say that the coupling is *monotone* with respect to the partial order if $P_1 \succeq P_2$ implies $f(P_1) \succeq f(P_2)$. The next lemma verifies that this condition holds for our coupling.

Lemma 7 *The above coupling is monotone with respect to the partial order* $\succeq$.

Proof. Suppose $P_1 \succeq P_2$, and consider a random move defined by a pair (i, r); let $P_1' = f(P_1)$ and $P_2' = f(P_2)$ be the images of P_1, P_2 respectively. Assume that $r \geq 1/2$ (the case $r < 1/2$ is symmetric). It is straightforward to verify that, if i is a peak in P_1, then either the ith point of P_1 is sufficiently far above the ith point of P_2 to ensure that $P_1' \succeq P_2$ (and therefore $P_1' \succeq P_2'$), or i is also a peak in P_2, in which case the peak is rotated in both paths. In either case we can deduce that $P_1' \succeq P_2'$. $\square$

We now need to bound the time taken for the two extremal paths to meet. To do this, we introduce a distance function as in lemma 6. 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 Cartesian lattice squares) of the region between P_1 and P_2. The crucial observation is that the distance $\Phi(t) = \phi(X_t, Y_t)$ will tend not to increase under our coupling, as the next lemma shows. By lemma 7, we may restrict attention to the case $X_t \succeq Y_t$.

Lemma 8 *Let P_1 and P_2 be any two 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 10, with path P_1 drawn as a solid line and path P_2 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 P_1 is strictly above P_2 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 points on P_1 are strictly above the corresponding points on P_2 (except at the endpoints). On the upper path P_1, label the peaks "good" (G) and the valleys "bad" (B), and vice versa for the lower path P_2. 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 total number of good points is greater than or equal to the total 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. □

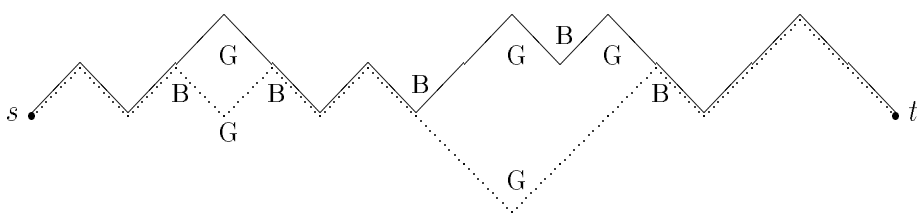


Figure 10: Proof of lemma 8

Theorem 9 *Let S be a region in the triangular lattice such that the corresponding region $\hat{S}$ has one source s and one sink t. Then the mixing time of the Markov chain $\mathcal{M}_{loz}$ on lozenge routings of $\hat{S}$ satisfies $\tau(\epsilon) \leq 2en^3\lceil \ln \epsilon^{-1} \rceil$, where n is the area of $\hat{S}$.*

Proof. By theorem 5, it suffices to show that the coupling time T satisfies $T \leq 2n^3$. To bound T we appeal to lemma 6, restricting attention to extremal initial states following lemma 7. Our distance function ϕ clearly takes integer values in the interval $[0, n]$. Moreover, we have seen in lemma 8 that $\mathrm{E}[\Delta\Phi] \leq 0$. (Note that the monotonicity of the coupling ensures that any pair (P_1, P_2) that is reachable under the coupling satisfies $P_1 \succeq P_2$.) It remains only to bound $\mathrm{E}[(\Delta\Phi)^2]$, assuming that the area between the pair of paths is non-zero. In this case, there must be a segment on which the solid path is strictly above the dotted path (using the terminology in the proof of lemma 8). 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 $\mathrm{E}[(\Delta\Phi)^2] \geq 1/2n$. (Clearly the number of points on a path cannot exceed n.) Plugging all of these quantities into lemma 6 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, say the ith point of the jth path, and the same random bit $r \in [0, 1]$. 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. Recall the partial order defined in the proof of lemma 3, in which routings R_1, R_2 satisfy $R_1 \succeq R_2$ iff the ith path of R_1 lies on or above the ith path of R_2, for all i. As before, there are unique maximum and minimum elements under $\succeq$. And once again it is not hard to check that our coupling is monotone with respect to this partial order:

Lemma 10 *The above coupling for routings is monotone with respect to the partial order* $\succeq$.

Proof. Suppose $R_1 \succeq R_2$, and consider a random move defined by the pair $(p, r) \in [1, \ldots, N] \times \{0, 1\}$. Assume without loss of generality that $r \geq 1/2$. If p is the top of a rotatable tower in R_1, then either R_2 is sufficiently far away from R_1 so that rotating the tower in R_1 does not disturb the order, or p is also the top of a tower in R_2. Because $R_1 \succeq R_2$, it must be the case that the tower defined by p in R_1 has height at least as large as the tower defined by p in R_2. Thus the probability of performing a rotation in R_1 cannot exceed that of performing a rotation in R_2. Therefore, depending on the value of r, we rotate either in both routings, only in R_2, or in neither routing. In each of these cases it is clear that the partial order is preserved. $\square$

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 8, proves that the distance tends not to increase under the coupling. Again, in light of lemma 10 we restrict attention to the case where $R_1 \succeq R_2$.

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

Proof. Consider a pair of routings, the upper one solid and the lower one dotted. Adopting the same terminology as in the proof of lemma 8, 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 11). We can calculate the expected change in area by first considering the expected change if we make a move on the upper (solid) routing only, and then adding to this the expected change in area from a move on the lower (dotted) routing only. Each point labeled G_i has probability $1/2Ni$ of rotating, and such a rotation decreases the area between the

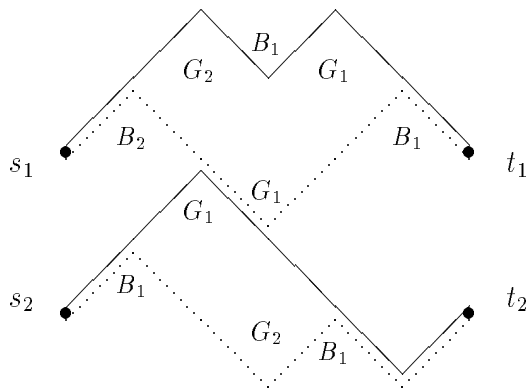


Figure 11: Proof of lemma 11

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 8. □

It is now a short step to the main theorem of this section, which confirms that simulating the Markov chain $\mathcal{M}_{loz}$ for a small number of steps suffices to generate a random lozenge tiling.

Theorem 12 *Let S be a region of the triangular lattice. The mixing time of the Markov chain $\mathcal{M}_{loz}$ on lozenge routings of $\hat{S}$ satisfies $\tau(\epsilon) \le 8en^4\lceil \ln \epsilon^{-1} \rceil$, where n is the area of $\hat{S}$.*

Proof. By theorem 5, it suffices to show that the coupling time satisfies $T \le 8n^4$; and by lemma 10 we need only consider the coupling time for the two extremal initial states. We first bound the distance between these extremal configurations.

Recall two properties of the height function h. First, the heights of all points around the boundary are fixed for all routings of a region. Second, the heights of neighboring vertices differ by at most two. For any internal point v, let k_v be the side length of the largest hexagon centered at v and fully contained within the region. This largest hexagon must share at least one vertex with the boundary of the region, and hence in any tiling the height of v can take on at most $4k_v + 1$

consecutive integer values. Finally, recall that we can walk from the lowest to the highest routing by a sequence of rotations each of which increases the height of one vertex by three (and leaves all other heights unchanged). As we perform this walk, there can be at most $\lceil \frac{4k_v+1}{3} \rceil < 2k_v$ rotations which alter the height of v. If the region being tiled has area n, then $k_v < \sqrt{n}$ for every vertex v. Therefore, summing over all internal vertices, we find that the total distance between the highest and lowest routings is at most $\sum_v 2\sqrt{n} = 2n^{3/2}$.

Lemma 11 confirms that $\mathrm{E}[\Delta\Phi] \leq 0$ for all pairs of states reachable under the coupling. 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 solid and one dotted, and a segment in which the solid path is strictly above the dotted 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 $\mathrm{E}[(\Delta\Phi)^2] \geq h^2/2Nh \geq 1/2n$. Putting all this together, and appealing to lemma 6, we see that the coupling time satisfies $T \leq 8n^4$. The result now follows from theorem 5. □

Remarks. (a) We have made no attempt here to optimize the upper bound on the mixing time in theorem 12; our concern has been to establish that the chain is *rapidly mixing*, in the sense that the mixing grows only *polynomially* with the area n. (Note in contrast that the size of the state space — i.e., the number of lozenge tilings — is in general *exponential* in n.) Using a more detailed analysis of our coupling, Wilson [18] has derived sharper bounds on the mixing time.

(b) The monotonicity property of our coupling allows one to determine bounds on the coupling time experimentally, simply by simulating the coupled process starting at the two extremal states. In practice, this has been found to yield significantly tighter bounds than that of theorem 12. This idea can be further extended using the technique of "coupling from the past" due to Propp and Wilson [14], to obtain a stopping rule for the simulation that eliminates all bias from the samples. Theorem 12 can be viewed as an *a priori* bound on the running time of such experiments.

(c) Randall and Tetali [15] recently demonstrated that the comparison technique of Diaconis and Saloff-Coste [5] can be used to relate the mixing time of $\mathcal{M}_{loz}$ to that of the original Markov chain whose only moves are *simple* rotations. Their results, together with theorem 12, imply that the original Markov chain is rapidly mixing as well. □

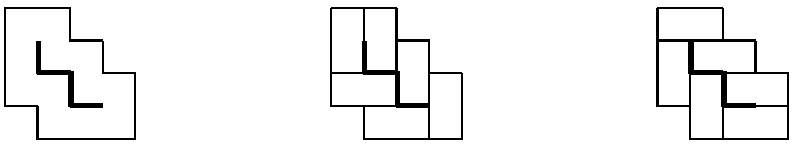
4 Sampling domino tilings

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. 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.

Our task in this section is to construct a random domino tiling of a given region S of the Cartesian lattice. This we achieve using a Markov chain on the space of domino tilings, whose moves correspond to rotations of suitably defined towers. As before, towers are constructed so as to allow, in a single move, a series of domino rotations which enable a rotation at a particular point which would otherwise be disallowed. Consider, for example, a 2×2 block centered at a point $p = (x, y)$ which has one vertical domino to its right, but a horizontal domino covering the square to its lower-left. Then either we can perform a domino rotation at the point $(x - 1, y)$, which would then allow a domino rotation at p, or the square to the upper-left of p is covered by a vertical domino. Continuing in a zigzag fashion to the upper-left, we will eventually find a point $q = (x - k, y - k)$ or $q = (x - k - 1, y - k)$ where we can perform a rotation (assuming that we do not hit the boundary). Rotating dominoes along the zigzag, starting at q, we can eventually perform a rotation at p. See figure 12. As we shall see, allowing these types of tower moves along just one of the two diagonal directions turns out to be sufficient, so we restrict our attention to those running in the NW-SE direction.

To formalize the above, a domino tower is defined by a *spine*, which is a zigzag path (staircase of unit steps along domino boundaries) in the NW-SE direction of a tiling, and the tower includes all the cells that are incident to a vertex on the spine. The endpoints of the spine are the *top* and *bottom* of the tower. Provided that all the vertices along the spine lie in the interior of the region, there are exactly two ways in which the tower can be tiled so that one of the two endpoints of the spine is the center of a 2×2 block tiled with two parallel dominoes. A *rotation* of the tower is an operation that replaces one of these two tilings with the other. There are four distinct types of domino towers, depending on whether the first and last edges of the spine are horizontal or vertical. The *height* of a tower is the number of vertices on the spine.

For the analysis it is more useful to redefine towers in terms of the corre-



A spine defining a tower The two tilings of the tower

Figure 12: Towers moves for domino tilings

sponding routings. Recall that in theorem 2 we showed that generating a random tiling is equivalent to constructing a random domino routing of a region of the triangular lattice with sources $\{s_1, \ldots, s_k\}$ and sinks $\{t_1, \ldots, t_k\}$. The four types of towers can easily be interpreted in terms of routings, as shown in figure 13. A move of the Markov chain consists of identifying a tower and then moving between the solid and dotted local structures shown. In the routings representation, the height is just the number of unit triangles in the tower. The edges marked e in the diagram indicate the bottoms of the corresponding towers in the north-west direction, while the edges marked e' are the tops of towers in the south-east direction. (To relate this definition to tilings, if e is a diagonal edge then it corresponds to a vertical domino such that the center of the left edge is the bottom of the spine defining the tower in the tiling; if e is horizontal then it corresponds to a horizontal tile such that the center of the top edge is the bottom of the spine. For edges e', the center of the left or bottom edge of the corresponding domino is the top of the spine.) We refer to any of the edges e or e' as the "start" of a tower; notice that a tower can be uniquely specified by its start edge and its direction. The type of the tower merely reflects whether the top and bottom edges are horizontal or diagonal. For example, the original Markov chain based on rotating two neighboring domino tiles consists of all Type I and Type IV tower rotations of height 1. Let N be the number of edges in any domino routing of $\hat{S}$. The transition probabilities $P(\cdot, \cdot)$ of $\mathcal{M}_{dom}$ are

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

To implement one step of this Markov chain, starting at a routing R, choose $(e, r) \in \{1, \ldots, N\} \times [0, 1]$ uniformly at random, where e is a random *edge* of R. First suppose $r \leq 1/2$. If the edge e is a diagonal directed up and to the right, check whether 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 if

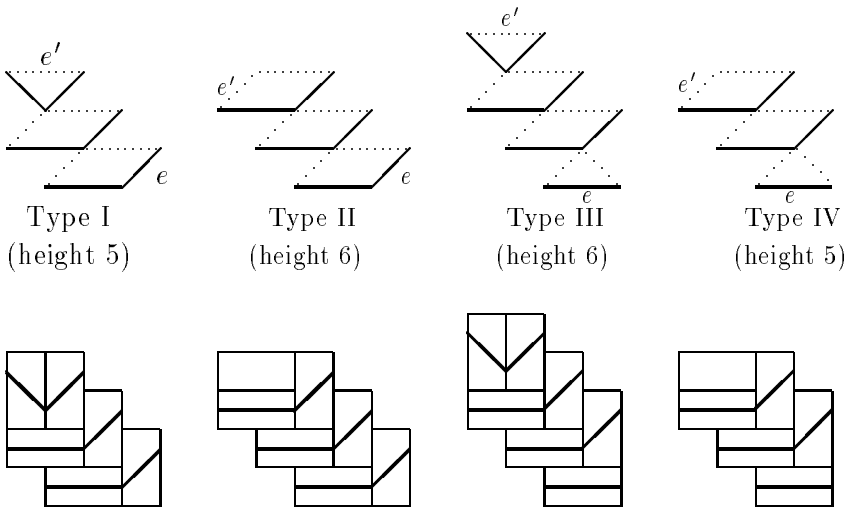


Figure 13: The four types of domino towers

possible (i.e., if the tower lies entirely within the region). The case when $r > 1/2$ is similar: check 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 possible) if $r > 1 - 1/2h$, where h is the height of the tower.

The Markov chain $\mathcal{M}_{dom}$ can be analyzed using arguments similar to those used in the previous section for $\mathcal{M}_{loz}$. As before, we first need to show that the chain is ergodic.

Theorem 13 *The Markov chain $\mathcal{M}_{dom}$ is ergodic and converges to the uniform distribution over domino tilings.*

Proof. It is easy to check that the Markov chain is symmetric and aperiodic. Therefore, it suffices to show that the Markov chain connects the state space of domino tilings of a region. We will show that the state space is connected even if we restrict the set of allowable transitions to simple domino rotations.

We start by defining a partial order on the set of all domino routings of $\hat{S}$, as follows. Let R_1, R_2 be two routings and let P_1, P_2 be a pair of corresponding paths (with the same source and sink). We say that $P_1 \succeq P_2$ iff each vertical line intersecting the paths intersects P_1 at a point at least as high as the intersection with P_2. (Since the paths are piecewise linear, it suffices to check this condition at vertical lines which pass through vertices of the underlying lattice.) As before, we say $R_1 \succeq R_2$ iff the relation $P_1 \succeq P_2$ holds for all pairs of corresponding paths, and define the distance between two routings to be the union of the area between corresponding paths.

Using an argument analogous to that in the proof of lemma 3, it can be verified that there is a unique minimum routing $R_\perp$, and that for any routing

$R \neq R_\perp$ we can find a tower of height 1 whose rotation will reduce the distance between R and $R_\perp$. This moves R closer to $R_\perp$ and inductively guarantees the connectedness, and thus the ergodicity, of $\mathcal{M}_{dom}$. $\square$

It remains for us to show that the Markov chain converges rapidly to stationarity. We can define a coupling in exactly analogous fashion to that for lozenge routings in section 3.4: namely, pick a random pair (e, r) and make the appropriate move in both routings simultaneously. Once again it is not hard to check that this coupling is monotone with respect to the partial order on domino routings defined in the proof of theorem 13, so when analyzing the coupling time we need only consider the two extremal routings. The key fact is the following analog of lemma 11, which says that the area $\Phi(t)$ between routings (defined in the obvious way) is non-increasing in expectation.

Lemma 14 *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. First consider the case when R_1 and R_2 each consist of just a single path. Then, following the ideas from lemmas 8 and 11, we can consider segments $C_1, \ldots, C_\ell$ on which the two paths coincide, and segments $D_1, \ldots, D_m$ 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 decrease by exactly i. Similarly, label an edge B_i if one of the directions could cause the area to increase. When the routings consist of a single path, i is either 1 or 2. Certain edges will receive two labels. See figure 14.

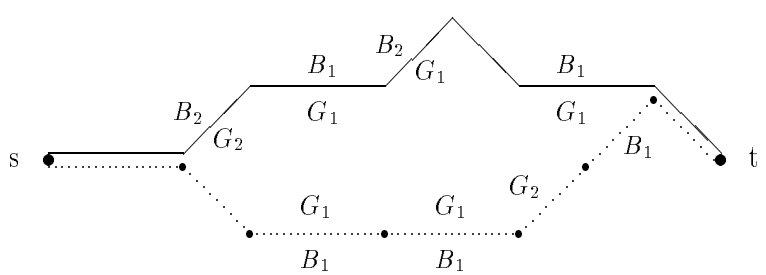


Figure 14: Proof of lemma 14 — single path case

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 less than or equal to the number of good edges. As in lemma 11, 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.

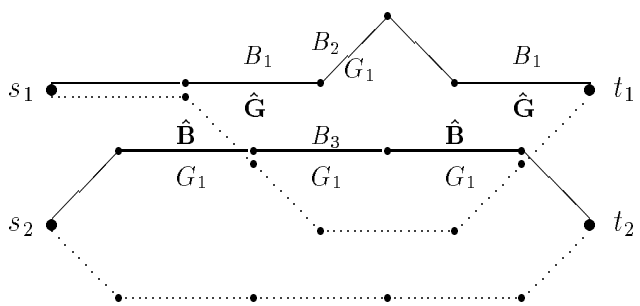


Figure 15: Proof of lemma 14 — multiple path case

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 because of interference with another path. These labels are shown in figure 15 for the solid paths only. The crucial point is that every edge labeled $\hat{G}$ must be paired with a distinct edge labeled $\hat{B}$, and vice-versa; these are unit distance apart in the NE-SW direction. Hence $\#\hat{G} = \#\hat{B}$.

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 as before that the expected change in area is always non-positive. □

As a final ingredient, we need to bound the distance between the two extremal routings. For a fixed region with area n, the distance between the highest and lowest routings is at most $2n^{3/2}$. To see this, recall that the difference between

the heights of neighboring vertices in a domino routing is at most ± 3. If k_v is the side length of the largest square centered at v which lies within the region, then the height of v is an integer value in $[h(u) - 2k_v, h(u) + 2k_v]$, where u is a vertex which lies on both the square and the boundary of the region. Each step in a walk from the lowest to the highest routing increases the height of some vertex by four, so the total number of steps is at most $\sum_v \frac{6k_v+1}{4} \leq \sum_v 2\sqrt{n} = 2n^{3/2}$.

In analogous fashion to the proof of theorem 12, lemma 14, together with lemma 6 and monotonicity of the coupling, implies that the Markov chain $\mathcal{M}_{dom}$ is rapidly mixing:

Theorem 15 *Let S be a region of the Cartesian lattice, and let $\hat{S}$ be the corresponding region containing domino routings. Then the mixing time of the Markov chain $\mathcal{M}_{dom}$ on routings of $\hat{S}$ satisfies $\tau(\epsilon) \leq 8en^4\lceil \ln \epsilon^{-1} \rceil$, where n is the area of S.* $\square$

5 Sampling Eulerian orientations

Let S be a region with specified boundary conditions for Eulerian orientations: recall that these determine the sources and sinks in the routings representation. 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, moves will be defined in terms of "peaks" and "valleys," where in this case a valley is a vertex which is the right endpoint of a horizontal edge and the bottom endpoint of a vertical edge of the routing; a peak is a vertex which is the left endpoint of a horizontal edge and the top endpoint of a vertical edge. Note that a vertex may be both a peak and a valley at the same time. The operation of flipping a peak into a valley (or vice versa) where possible defines a simple Markov chain which we argue in theorem 16 connects the state space.

For our analysis we augment this Markov chain by allowing a move between two routings if they differ by a structure which is either a vertical or horizontal tower, as depicted in figure 16. More precisely, let the *cell at* (x, y) be the four edges of the unit square of the lattice whose lower-right corner is at (x, y). Define a *horizontal tower of height* h to be a set of cells at points $(x, y), (x-1, y), \ldots, (x-h+1, y)$, where either $(x, y), (x-1, y), \ldots, (x-h+1, y)$ are all valleys and $(x, y+1), (x-h, y)$ are not valleys, or $(x-h, y+1), \ldots, (x-1, y+1)$ are all peaks and $(x-h, y), (x, y+1)$ are not peaks. The point (x, y) is the *bottom* of the tower and $(x-h, y+1)$ is the *top* of the tower. Similarly, define a *vertical tower of height* h to be a set of cells at points $(x, y), (x, y+1), \ldots, (x, y+h-1)$, where either $(x, y), \ldots, (x, y+h-1)$ are all valleys and $(x-1, y), (x, y+h)$ are not valleys, or $(x-1, y+h), \ldots, (x-1, y+1)$ are all peaks and $(x, y+h), (x-1, y)$ are not peaks. In this case (x, y) is the bottom of the tower and $(x-1, y+h)$ is

the top (see figure 16). Note that vertical and horizontal towers of height 1 are identical, and that any vertex can be the top (respectively, bottom) of at most one tower. The reader should be able to verify that any tower can be inverted in the obvious way to produce a valid routing provided all of its cells are contained within the region. As usual, we call such an operation a *rotation* of the tower.

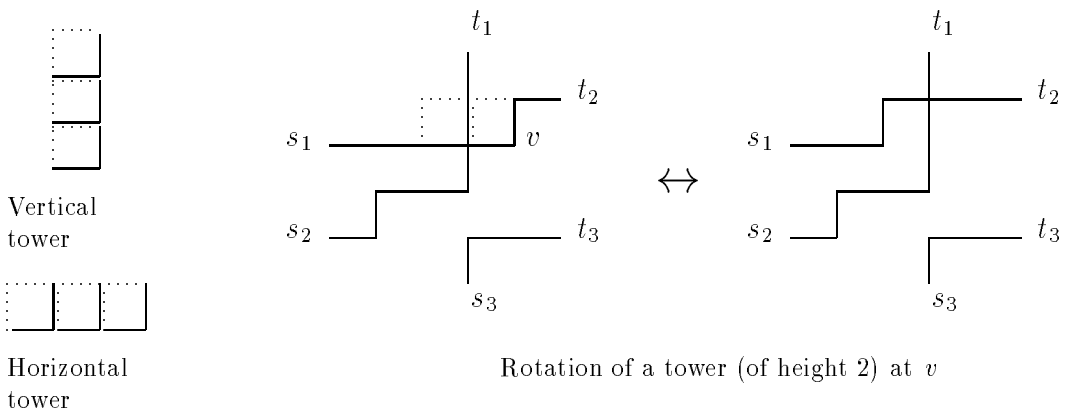


Figure 16: Markov chain for Eulerian orientations

The transition probabilities $P(\cdot\,,\cdot)$ of $\mathcal{M}_{eul}$ are defined as follows:

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

where N is the number of vertices on any routing.

To implement one step of $\mathcal{M}_{eul}$, starting at a routing R, choose $(p, r) \in \{1, \ldots, N\} \times [0, 1]$ uniformly at random, where p denotes a point on the routing. If $r \leq 1/2$, check whether p is the bottom of a tower of height h, and rotate this tower if possible if $r < 1/2h$. Similarly, if $r > 1/2$ check whether p is the top of a tower of height h, and rotate this tower if possible if $r > 1 - 1/2h$.

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 $1 \times h$ or $h \times 1$ 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.

Theorem 16 *The Markov chain $\mathcal{M}_{eul}$ is ergodic and converges to the uniform distribution over Eulerian orientations.*

Proof. Since the Markov chain is symmetric and aperiodic, it is again sufficient to show that it is connected. As in the case of domino and lozenge tilings, the

key to proving this fact is to define a suitable partial order on routings. For corresponding paths P_1, P_2 on two routings R_1, R_2, we say that $P_1 \succeq P_2$ if each diagonal line in the NW-SE direction intersects P_1 at a point at least as far left as P_2. We say $R_1 \succeq R_2$ iff $P_1 \succeq P_2$ holds for each pair of corresponding paths P_1, P_2. Under this partial order there will always be a unique minimum routing. An argument analogous to that in the proof of lemma 3 demonstrates that, for any routing R which is not minimum, we can find a point which is the top of a rotatable tower whose rotation reduces the distance between R and the minimum routing (where again distance is defined as the sum of the area between corresponding paths). This guarantees, inductively, that the state space is connected. $\square$

We now turn to the mixing time of the Markov chain. 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 9 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 construct a coupling that chooses a random point p and a random number r and makes the appropriate move in both routings simultaneously. Again as before, this coupling is easily seen to be monotone with respect to the partial order defined in the proof of theorem 16, so we need only analyze the coupling time starting from the two extremal routings. 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 17 *Let R_1 and R_2 be any two Eulerian routings such that $R_1 \succeq R_2$. Then* $\mathrm{E}[\Delta\Phi | R_1, R_2] \leq 0$.

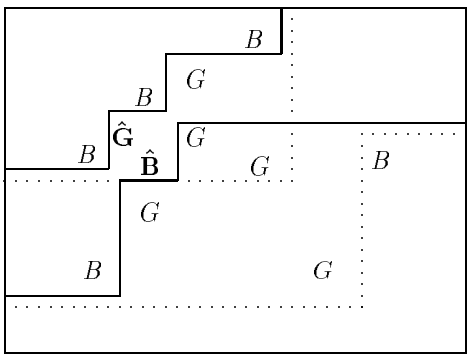


Figure 17: Proof of lemma 17

Proof. We assign labels G and B to peaks and valleys in similar fashion to the lozenge case (see figure 17). 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. Finally we will label all the corners which do not define a tower because of interference with another path. On the lower routing, label $\hat{G}$ any unlabeled corner which goes to the right and then up, if the two edges which complete the unit square are part of another path on the routing. Likewise, label $\hat{B}$ any unlabeled corner which goes up and then to the right, but which is not part of a tower because the two edges completing the square are part of anther path on the routing. On the upper routing, interchange the roles of $\hat{G}$ and $\hat{B}$. Generalizing from the single path case as in lemma 11, we know that $\#B + \#\hat{B} \leq \#G + \#\hat{G}$.

Next observe that each point labeled $\hat{G}$ is paired with a distinct point labeled $\hat{B}$. A point p is labeled $\hat{G}$ if the two edges which complete the unit square defined by the peak or valley at p belong to an adjacent path on the same routing. 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. This pairing implies that $\#\hat{B} = \#\hat{G}$, 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 total number of points on the routing, we have $\mathrm{E}[\Delta\Phi] = (\#B - \#G)/2N \leq 0$. $\square$

A simple calculation based on the height function representation reveals that the distance between extremal Eulerian routings of a region with area n is at most $n^{3/2}$. Combining lemma 17 with lemma 6 and the fact that the coupling is monotone, we can deduce in by now familiar fashion that the Markov chain $\mathcal{M}_{eul}$ is rapidly mixing:

Theorem 18 *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 $\tau(\epsilon) \leq 2en^4\lceil \ln \epsilon^{-1} \rceil$, where n is the area of S.* $\square$

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] Diaconis, P. and Saloff-Coste, L. Comparison theorems for reversible Markov chains. *Ann. Appl. Probability* **3** (1993), pp. 696–730.

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

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

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

[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] Kenyon, C., Randall, D. and Sinclair, A. Approximating the number of dimer coverings of a lattice. *Journal of Statistical Physics* **83** (1996), pp. 637–659.

[13] 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.

[14] Propp, J. and Wilson, D. Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Structures and Algorithms* **9** (1996), pp. 223–252.

[15] Randall, D. and Tetali, P. Analyzing Glauber dynamics by comparisons of Markov chains. *Journal of Mathematical Physics* **41** (2000), pp. 1598–1615.

[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.

[18] Wilson, D.B. Mixing times of lozenge tiling and card shuffling Markov chains. Manuscript, 1997, available at `http://research.microsoft.com/~dbwilson/nlm`.