

MIXING TIMES OF LOZENGE TILING AND CARD SHUFFLING MARKOV CHAINS¹

BY DAVID BRUCE WILSON

Microsoft Research

We show how to combine Fourier analysis with coupling arguments to bound the mixing times of a variety of Markov chains. The mixing time is the number of steps a Markov chain takes to approach its equilibrium distribution. One application is to a class of Markov chains introduced by Luby, Randall and Sinclair to generate random tilings of regions by lozenges. For an $\ell \times \ell$ region we bound the mixing time by $O(\ell^4 \log \ell)$, which improves on the previous bound of $O(\ell^7)$, and we show the new bound to be essentially tight. In another application we resolve a few questions raised by Diaconis and Saloff-Coste by lower bounding the mixing time of various card-shuffling Markov chains. Our lower bounds are within a constant factor of their upper bounds. When we use our methods to modify a path-coupling analysis of Bubley and Dyer, we obtain an $O(n^3 \log n)$ upper bound on the mixing time of the Karzanov–Khachiyan Markov chain for linear extensions.

1. Introduction. Using a simple idea, we obtain improved upper and lower bounds on the mixing times of a number of previously studied Markov chains:

- A lozenge is a rhombus with angles of 120° and 60° and sides of unit length. Figure 1 shows a random lozenge tiling of a hexagon. Random lozenge tilings were originally studied in physics as a model for dimer systems, and have more recently served as an exploratory tool for people in combinatorics. Section 5 gives further background. Luby, Randall and Sinclair (1995) proposed a Markov chain to generate random lozenge tilings of regions, proved that it runs in time $O(n^4)$ when there are n lozenges, and in later unpublished work reduced the bound to $O(n^{3.5})$. (They also analyzed domino-tiling and Eulerian-orientation Markov chains.) For a regular hexagon with side length ℓ , for example, their methods give a bound of $O(\ell^7)$. We show here that the correct mixing time of this Markov chain is $\Theta(\ell^4 \log \ell) = \Theta(n^2 \log n)$, by showing the state to be very far from stationarity after $\sim (8/\pi^2)\ell^4 \log \ell$ steps and very close to stationarity after $\sim (48/\pi^2)\ell^4 \log \ell$ steps. The correct constant appears to be $16/\pi^2$. For general regions of size n and width w , our upper bound is $\sim (3/\pi^2)w^2 n \log n$.

Received February 2001; revised September 2002.

¹Supported in part by an NSF postdoctoral fellowship.

AMS 2000 subject classifications. Primary 60J10; secondary 60C05.

Key words and phrases. Random walk, mixing time, card shuffling, lozenge tiling, linear extension, exclusion process, lattice path, cutoff phenomenon.

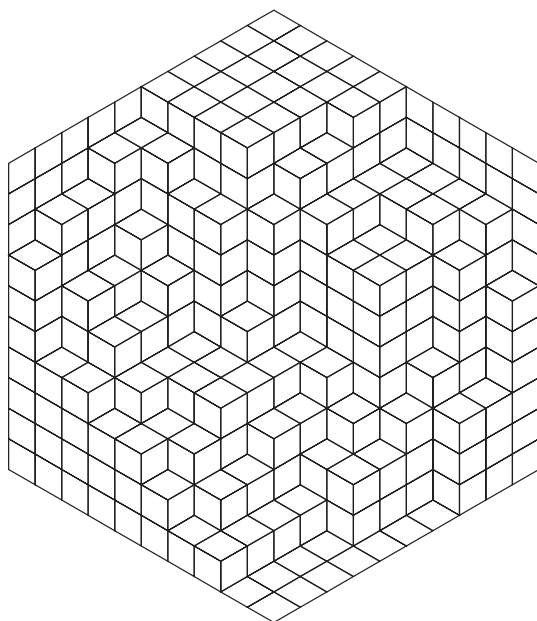


FIG. 1. *Random lozenge tiling of the order 10 hexagon, chosen uniformly at random from the 9,265,037,718,181,937,012,241,727,284,450,000 possible such tilings. Here $\ell = 10$, $w = 20$ and $n = 300$.*

- Consider the following shuffle on a deck of n cards: with probability $1/2$ do nothing; otherwise transpose a random adjacent pair of cards. How many of these operations are needed before the deck becomes well shuffled? In the years since Aldous [(1983), Section 4] showed that $O(n^3 \log n)$ shuffles are enough and that $\Omega(n^3)$ shuffles are necessary to randomize the deck, there have been a couple of heuristic arguments [Aldous (1997) and Diaconis (1997)] for why $\Omega(n^3 \log n)$ shuffles should be necessary, but unruly technical difficulties prevented a rigorous proof from being written down. Using our method these technical difficulties vanish. With little more than algebra and trigonometry, we show that $(1/\pi^2 - o(1))n^3 \log n$ shuffles are necessary to begin to randomize the deck, and that $(2/\pi^2 + o(1))n^3 \log n$ shuffles are enough. The best previous published explicit upper bound was $(4 + o(1))n^3 \log n$ shuffles. The correct constant appears to be $1/\pi^2$.
- We lower bound the mixing time of a few other shuffles analyzed by Diaconis and Saloff-Coste (1993a). They considered shuffles for which the cards appear on the vertices of a graph, and the shuffle picks a random edge and transposes the cards at its endpoints. They obtained upper bounds on the mixing time for shuffles based on the $\sqrt{n} \times \sqrt{n}$ grid and on the $(\log_2 n)$ -dimensional hypercube, but did not have matching lower bounds. We provide lower bounds, showing that their upper bounds are correct to within constant factors.

- Counting the number of linear extensions of a partially ordered set is #P-complete [Brightwell and Winkler (1991)], making it computationally intractable. (One application of counting linear extensions is in a data mining application that infers partial orders [Mannila and Meek (2000)].) One can approximately count linear extensions by randomly generating them, so there have been a number of articles on generating random linear extensions. The latest, by Bublely and Dyer (1998), proposed a Markov chain in which pairs of elements in the linear extension are randomly transposed if doing so respects the partial order. To make the analysis easy, their Markov chain selects a random site with probability proportional to a parabolic curve and then attempts to transpose the elements at this random site. We show here that the uniform distribution on sites works as well, obtaining a constant factor that is only marginally worse.

We give further background on these various Markov chains in later sections where we analyze them. In Section 2 we provide basic definitions, such as what it means formally for the state of a Markov chain to be close to random. We study in Section 3 a Markov chain for generating random lattice paths, since it is simple, yet illustrates the key ideas that we use to analyze the various other Markov chains. We use Fourier analysis to define the state space of the Markov chain a function Φ that has a certain contraction property. With S denoting the current state of the Markov chain and the random variable S' denoting the next state of the chain, we have $\mathbf{E}[\Phi(S')|S] = (1 - \gamma)\Phi(S)$. We derive both upper bounds and lower bounds using this contraction property. After Section 3 the remaining sections may be read in any order. We see in Section 4 how to apply the results about the path Markov chain to the chain for shuffling by random adjacent transpositions. In Section 5 we generalize the upper bound for the path Markov chain to upper bound the mixing time of the lozenge-tiling Markov chain introduced by Luby, Randall and Sinclair (1995). In Section 6 we modify Bublely and Dyer's path-coupling analysis of the Karzanov–Khachiyan Markov chain to obtain the $O(n^3 \log n)$ mixing time bound. When using a local randomizing operation to update a high-dimensional configuration, one typically either updates a random coordinate each step or updates the coordinates in a systematic order. In Section 7 we compare these two methods for the chains that we are studying; our analysis indicates that the second method is better. We take a second look at the lattice path and permutation Markov chains in Section 8, and refine our previous arguments to obtain tighter constants. We consider exclusion and exchange processes in Section 9, where among other things we resolve the aforementioned questions of Diaconis and Saloff-Coste. Many of the mixing time upper and lower bounds we give differ by small constant factors. We give in Section 10 heuristic arguments and present experimental evidence for determining the correct constant factors in the mixing times. We summarize in Table 1 many of these mixing time bounds and their (conjectural) correct values. Section 10 also contains several open problems for further research. We make some concluding remarks in Section 11.

TABLE 1

State space, Markov chain (Summary of mixing time bounds for several classes of Markov chains.)

Parameter	(Conjectural) Correct answer	Rigorous lower bound	Rigorous upper bound
Paths in $n/2 \times n/2$ box, random adjacent transpositions			
Coupling threshold	$2/\pi^2 n^3 \log n$	$2/\pi^2 n^3 \log n$	$2/\pi^2 n^3 \log n$
Variation threshold	$1/\pi^2 n^3 \log n$	$1/\pi^2 n^3 \log n$	$2/\pi^2 n^3 \log n$
Separation threshold	$2/\pi^2 n^3 \log n$	$1/\pi^2 n^3 \log n$	$4/\pi^2 n^3 \log n$
Spectral gap	$\frac{1 - \cos(\pi/n)}{n - 1}$	$\frac{1 - \cos(\pi/n)}{n - 1}$	$\frac{1 - \cos(\pi/n)}{n - 1}$
Permutations S_n , random adjacent transpositions			
Coupling threshold	$4/\pi^2 n^3 \log n$	$2/\pi^2 n^3 \log n$	$4/\pi^2 n^3 \log n$
Variation threshold	$1/\pi^2 n^3 \log n$	$1/\pi^2 n^3 \log n$	$2/\pi^2 n^3 \log n$
Separation threshold	$2/\pi^2 n^3 \log n$	$1/\pi^2 n^3 \log n$	$4/\pi^2 n^3 \log n$
Spectral gap	$\frac{1 - \cos(\pi/n)}{n - 1}$	$\frac{1 - \cos(\pi/n)}{n - 1}$	$\frac{1 - \cos(\pi/n)}{n - 1}$
Lozenge tilings of order ℓ hexagon, Luby–Randall–Sinclair chain			
Coupling threshold	??	$8/\pi^2 \ell^4 \log \ell$	$48/\pi^2 \ell^4 \log \ell$
Variation threshold	$16/\pi^2 \ell^4 \log \ell$	$8/\pi^2 \ell^4 \log \ell$	$48/\pi^2 \ell^4 \log \ell$
Separation threshold	$32/\pi^2 \ell^4 \log \ell$	$8/\pi^2 \ell^4 \log \ell$	$96/\pi^2 \ell^4 \log \ell$
Spectral gap	$\frac{1 - \cos(\pi/(2\ell))}{\ell(2\ell - 1)}$	$\frac{1 - \cos(\pi/(2\ell))}{\ell(2\ell - 1)}$	$\frac{1 - \cos(\pi/(2\ell))}{\ell(2\ell - 1)}$
Linear extensions of partially ordered set, Karzanov–Khachiyan chain			
Variation threshold	Depends on poset	Sometimes tight to within constants	$4/\pi^2 n^3 \log n$
Spectral gap	Depends on poset	$\frac{1 - \cos(\pi/n)}{n - 1}$	Sometimes tight
Lozenge tilings of region of n triangles and width w , Luby–Randall–Sinclair chain			
Variation threshold	Depends on region	Sometimes tight to within constants	$3/\pi^2 w^2 n \log n$
Spectral gap	Depends on region	$\frac{1 - \cos(\pi/w)}{n}$	Sometimes tight to within constants

Notes. The variation and separation thresholds are defined in Section 2. The bounds for lattice paths and permutations are proved in Sections 3 and 8, the bounds for lozenge tilings are proved in Section 5, and the bounds for the Karzanov–Khachiyan chain are proved in Section 6. The coupling times are for the natural monotone grand couplings described in Sections 3–5. The conjectural correct answers are derived in Section 10. The spectral gap for permutations was previously known (Diaconis, unpublished).

2. Preliminaries. Here we review some basic definitions and properties pertaining to mixing times and couplings. For a more complete introduction to these ideas, see Aldous (1983) or Aldous and Fill (2004).

When a Markov chain is started in a state x and run for t steps, we denote the distribution of the state at time t by P_x^t , where P is the state-transition matrix of

the Markov chain. If the Markov chain is connected and aperiodic, then as $t \rightarrow \infty$ the distribution P_x^t converges to a unique stationary distribution which is often denoted by π . Since we use π to denote the ratio of the circumference of a circle to its diameter, here we use μ to denote the stationary distribution. In all our examples, the stationary distribution is the uniform distribution, which we denote with U .

To measure the distance between the distributions P_x^t and μ one usually uses the total variation distance. With \mathcal{X} denoting the state space, the variation distance is defined by

$$\|P_x^t - \mu\|_{TV} = \max_{A \subset \mathcal{X}} |P_x^t(A) - \mu(A)| = \frac{1}{2} \sum_{y \in \mathcal{X}} |P_x^t(y) - \mu(y)| = \frac{1}{2} \|P_x^t - \mu\|_1.$$

The variation distance when the chain is started from the worst start state is denoted by

$$d(t) = \max_x \|P_x^t - \mu\|_{TV},$$

although it is often more convenient to work with

$$\bar{d}(t) = \max_{x,y} \|P_x^t - P_y^t\|_{TV},$$

since $\bar{d}(t)$ is submultiplicative whereas $d(t)$ is not. It is easy to see that $d(t) \leq \bar{d}(t) \leq 2d(t)$.

The (variation) mixing time is the time it takes for $\bar{d}(t)$ to “become small” (say less than $1/e$). It is a surprising fact that for many classes of Markov chains there is a threshold time T such that $\bar{d}((1 - \varepsilon)T) > 1 - \varepsilon$ but $\bar{d}((1 + \varepsilon)T) < \varepsilon$, where ε tends to 0 as the “size” of the Markov chain gets large; see Diaconis (1996) for a survey of this “cutoff phenomenon.”

Most of our mixing time upper bounds are derived via coupling arguments. In a (pairwise) coupling there are two copies X_t and Y_t of the Markov chain that are run in tandem. The X_t ’s by themselves follow the transition rule of the Markov chain, as do the Y_t ’s, but the joint distribution of (X_{t+1}, Y_{t+1}) given (X_t, Y_t) is often contrived to make the two copies of the Markov chain quickly coalesce (become equal). It is a standard fact that

$$\bar{d}(t) \leq \max_{x,y} \Pr[X_t \neq Y_t | X_0 = x, Y_0 = y],$$

so that a coupling which coalesces quickly can give us a good upper bound on the mixing time. Most of the variation threshold upper bounds in Table 1 follow from this relationship and a corresponding coupling time bound, and lower bounds on the coupling time likewise follow from corresponding lower bounds on the variation threshold time. The remaining variation threshold upper bounds and coupling time lower bounds in Table 1 that do not follow from this relationship are derived in Section 8.

Many pairwise couplings can be extended to “grand couplings,” where at each time step there is a random function F_t defined on the whole state space of the Markov chain, and $X_{t+1} = F_t(X_t)$ and $Y_{t+1} = F_t(Y_t)$ for any X_t and Y_t . For example, if the state space is the set of permutations on n cards, then the update rule “pick a random adjacent pair of cards, and flip a coin to decide whether to place them in increasing order or decreasing order” defines a grand coupling; the choice of the adjacent pair and the value of the coin flip define the random function on permutations. In Sections 6 and 8 we also consider pairwise couplings that do not extend to grand couplings.

All of the grand couplings considered in this article are monotone, which is to say that there is a partial order \preceq such that if $x \preceq y$, then also $F_t(x) \preceq F_t(y)$. All of the partial orders considered here have a maximal element, denoted $\hat{1}$, and a minimal element, denoted $\hat{0}$, that is, so that $\hat{0} \preceq x \preceq \hat{1}$ for each x in the state space. Monotone grand couplings are particularly convenient for algorithms [see, e.g., Propp and Wilson (1996) or Fill (1998)].

In Section 10 we consider not only the variation distance, but also the separation distance, which is defined by

$$s(t) = \max_{x,y} \frac{\mu(y) - P_x^t(y)}{\mu(y)}.$$

The function $s(t)$ is also submultiplicative and also often exhibits a sharp threshold. In general, $d(t) \leq s(t)$ and, for reversible Markov chains, $s(2t) \leq 2\bar{d}(t) - \bar{d}(t)^2$ [see Aldous and Fill (2004), Chapter 4, Lemma 7]. The rigorous bounds in Table 1 that pertain to separation distance follow from these relationships and the corresponding bounds for the variation distance.

3. Lattice path Markov chain. Consider an $a \times b$ rectangle composed of 1×1 boxes rotated 45° , so that the sides of length a are oriented north-west/southeast. A lattice path (see Figure 2) is a traversal from the leftmost corner to the rightmost corner of the rectangle, traveling along the borders of the 1×1 boxes, so that each move is either up and right or down and right. Such lattice paths can be encoded as strings of length $a + b$ consisting of a 0’s (down moves) and b 1’s (up moves). There are $n!/(a!b!)$ such lattice paths, where for convenience we let $n = a + b$. These lattice paths correspond to sets of 1×1 boxes which are “stable under gravity,” that is, no box lies above an empty cell.

Consider the following Markov chain for randomly generating a lattice path between the opposite corners of the $a \times b$ rectangle. Given a path, the Markov chain randomly picks one of the $(n - 1)$ internal columns (we assume $n \geq 2$) and then randomly decides whether to try pushing the path up at that point or to try pushing it down. If pushing the path up (or down) would result in an invalid path, the Markov chain simply sits idle during that step. Equivalently, in the binary string representation, the Markov chain picks a random adjacent pair

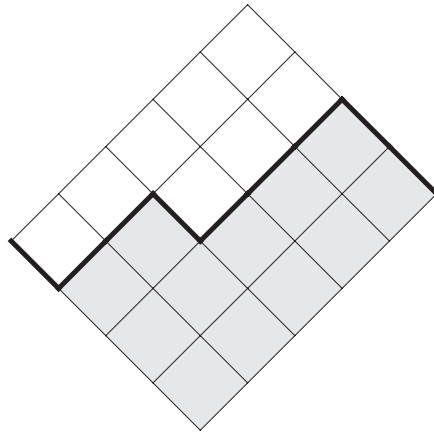


FIG. 2. A lattice path through the 4×5 rectangle, with the 1×1 boxes underneath it shaded. Here $a = 4$, $b = 5$ and $n = 9$, and the lattice path's encoding is 011011100.

of letters in the string and then randomly decides to either sort them or reverse-sort them. Understanding this Markov chain is instrumental to understanding the Markov chains for random lozenge tilings, card shuffling by random adjacent transpositions, and other types of card shuffling.

3.1. *Contraction property.* We analyze this Markov chain by measuring (1) the “displacement” of a given single path on the $a \times b$ square from “equilibrium” and (2) the “gap” between two such paths when one of the paths is entirely above the other. We lower bound the mixing time by computing the displacement of a path when not enough steps have been taken, and showing that it is typically different from the displacement of a random path. We upper bound the mixing time by showing that after enough steps, starting from the top and bottom paths, the expected gap is so small that the paths have almost surely coalesced to the same path.

It will be useful later to use horizontal coordinates that range from $-n/2$ to $n/2$. Let $h(x)$ denote the height of the path at position x relative to the line connecting the opposite corners of the box. That is, $h(x)$ is the number of up moves to the left of position x minus the expected number of such up moves. Thus $h(-n/2) = h(n/2) = 0$ and $h(x) = h(x - 1) + a/n$ if there was an up move between $x - 1$ and x , or else $h(x) = h(x - 1) - b/n$ if there was a down move. For the example in Figure 2, the heights change by $+4/9 = a/(a + b)$ for up moves and $-5/9 = -b/(a + b)$ for down moves, and are as follows:

x	$-9/2$	$-7/2$	$-5/2$	$-3/2$	$-1/2$	$1/2$	$3/2$	$5/2$	$7/2$	$9/2$
$h(x)$	0	$-5/9$	$-1/9$	$3/9$	$-2/9$	$2/9$	$6/9$	$10/9$	$5/9$	0

The displacement function Φ of h that we find useful is

$$(1) \quad \Phi(h) = \sum_{x=-n/2}^{n/2} h(x) \cos \frac{\beta x}{n},$$

where $0 \leq \beta \leq \pi$. This function weighs deviations from expectation more heavily near the middle of the path than near its endpoints. Given two lattice paths with height functions \check{h} and \hat{h} , where $\check{h}(x) \leq \hat{h}(x)$ for all x , define the gap function to be $\hat{h} - \check{h}$ and the gap to be

$$\Phi(\check{h}, \hat{h}) = \Phi(\hat{h} - \check{h}) = \Phi(\hat{h}) - \Phi(\check{h}).$$

Note that since $0 \leq \beta \leq \pi$, the gap is strictly positive when the paths \check{h} and \hat{h} differ, and is 0 otherwise. After the Markov chain has equilibrated, so that each path is equally likely, $\mathbf{E}[h(x)] = 0$, so the expected displacement is $\mathbf{E}[\Phi(h)] = 0$.

LEMMA 1. *Let the displacement function Φ be defined by (1). Suppose h is a height function [so $\Phi(h)$ is the displacement] and $\beta = \pi$ or $h = \hat{h} - \check{h}$ is a gap function [so $\Phi(h)$ is the gap] and $0 \leq \beta \leq \pi$. Let h' be the height or gap function after one step of the Markov chain. Then*

$$\mathbf{E}[\Phi(h') - \Phi(h)|h] \leq \frac{-1 + \cos(\beta/n)}{n - 1} \Phi(h)$$

with equality when $\beta = \pi$. The coefficient on the right-hand side is bounded by

$$-\frac{\beta^2}{2n^2(n - 1)} \leq \frac{-1 + \cos(\beta/n)}{n - 1} \leq -\frac{\beta^2}{2n^3}.$$

PROOF. Suppose we pick a site x , flip a coin, and adjust the height accordingly. Then the expected value of the new height at x is just $[h(x + 1) + h(x - 1)]/2$. Assume that we pick each site (other than $-n/2$ and $n/2$) with probability $1/p$, where $p = n - 1$ is the number of positions that can be picked. Then with primes denoting the updated variables,

$$\mathbf{E}[h'(x)|h] = \frac{p - 1}{p} h(x) + \frac{1}{p} \frac{h(x + 1) + h(x - 1)}{2},$$

when $-n/2 < x < n/2$, so that

$$(2) \quad \mathbf{E}[\Phi(h')|h] = \sum_{x=-n/2+1}^{n/2-1} \mathbf{E}[h'(x)|h] \cos \frac{\beta x}{n},$$

$$(3) \quad \mathbf{E}[\Phi(h')|h] = \frac{p - 1}{p} \Phi(h) + \frac{1/2}{p} \sum_{x=-n/2+1}^{n/2-1} [h(x + 1) + h(x - 1)] \cos \frac{\beta x}{n}$$

and

$$\begin{aligned}
 & \mathbf{E}[\Phi(h') - \Phi(h)|h] \\
 &= \frac{-1}{p} \Phi(h) + \frac{1/2}{p} \sum_{\substack{-n/2+1 \leq x \leq n/2-1 \\ -n/2 \leq y \leq n/2 \\ |x-y|=1}} h(y) \cos \frac{\beta x}{n} \\
 (4) \quad &= \frac{-1}{p} \Phi(h) + \frac{1/2}{p} \sum_{\substack{-n/2+1 \leq x \leq n/2-1 \\ -n/2+1 \leq y \leq n/2-1 \\ |x-y|=1}} h(y) \cos \frac{\beta x}{n} \\
 (5) \quad &\leq \frac{-1}{p} \Phi(h) + \frac{1/2}{p} \sum_{\substack{-n/2 \leq x \leq n/2 \\ -n/2+1 \leq y \leq n/2-1 \\ |x-y|=1}} h(y) \cos \frac{\beta x}{n} \\
 &= \frac{-1}{p} \Phi(h) + \frac{1/2}{p} \sum_{y=-n/2+1}^{n/2-1} h(y) \left[\cos \frac{\beta(y-1)}{n} + \cos \frac{\beta(y+1)}{n} \right] \\
 (6) \quad &= \frac{-1}{p} \Phi(h) + \frac{1/2}{p} \sum_{y=-n/2+1}^{n/2-1} h(y) 2 \cos \frac{\beta y}{n} \cos \frac{\beta}{n} \\
 &= \frac{-1 + \cos(\beta/n)}{p} \Phi(h) = \frac{-1 + \cos(\beta/n)}{n-1} \Phi(h),
 \end{aligned}$$

where we have used $\mathbf{E}[h'(\pm n/2)] = 0$ in (2), $h(\pm n/2) = 0$ in (4), and the trigonometric identity $\cos(\theta - \phi) + \cos(\theta + \phi) = 2 \cos(\theta) \cos(\phi)$ in (6). Inequality (5) is justified if $\beta \leq \pi$, since by assumption $h = \hat{h} - \check{h} \geq 0$ for each x , and $\cos(\beta/2) \geq 0$; when $\beta = \pi$, it becomes an equality.

To upper bound the right-hand side we use $\cos(x) \leq 1 - x^2/2 + x^4/24 = 1 + (x^2/2)(-1 + x^2/12)$,

$$\begin{aligned}
 \frac{-1 + \cos(\beta/n)}{n-1} &\leq \frac{1}{n-1} \frac{(\beta/n)^2}{2} \left(-1 + \frac{(\beta/n)^2}{12} \right) \\
 &\leq -\frac{\beta^2}{2n^3},
 \end{aligned}$$

where to get the last line we used $(1 - c/n^2)/(n - 1) \geq 1/n$ when $n > 1$ and $-c/n \geq -1$. Here $c \leq \pi^2/12$, so these conditions are satisfied whenever $n > 1$. The lower bound is somewhat easier: use the bound $\cos(x) \geq 1 - x^2/2$. \square

3.2. Upper bound.

THEOREM 2. When n is large, after

$$\frac{2 + o(1)}{\pi^2} n^3 \log \frac{ab}{\varepsilon}$$

steps, the variation distance from stationarity is ε and the probability that the two extremal paths have coalesced is $1 - \varepsilon$. [The $o(1)$ term is a function of n alone.]

Felsner and Wernisch (1997) referenced an early version of this article, since they used the upper bound in this theorem to bound the rate of convergence of the Karzanov–Khachiyan Markov chain for generating random linear extensions of a certain class of partially ordered sets. The Karzanov–Khachiyan Markov chain is discussed further in Section 6.

PROOF OF THEOREM 2. To obtain the upper bound, we consider a pair of coupled paths \hat{h}_t and \check{h}_t such that \hat{h}_0 is the topmost path and \check{h}_0 is the bottommost path. The sequences \hat{h}_t and \check{h}_t are generated by the Markov chain using “the same random moves,” so that \hat{h}_{t+1} and \check{h}_{t+1} are obtained from \hat{h}_t and \check{h}_t , respectively, by sorting or unsorting (same random decision made in both cases) at the same random location x . One can check by induction that $\hat{h}_t \geq \check{h}_t$. Let $\Phi_t = \Phi(\hat{h}_t - \check{h}_t)$; $\Phi_t = 0$ if and only if $\hat{h}_t = \check{h}_t$.

We compute $\mathbf{E}[\Phi_t]$; when it is small compared to the minimum possible positive value of Φ_t , it follows that with high probability $\Phi_t = 0$. By choosing β to be slightly smaller than π , we make this minimum positive value not too small and thereby get a somewhat improved upper bound.

From Lemma 1 and induction, we get

$$\mathbf{E}[\Phi_t] \leq \Phi_0 \left[1 - \frac{1 - \cos(\beta/n)}{n - 1} \right]^t \leq \Phi_0 \exp \left[-\frac{\beta^2 t}{2n^3} \right].$$

However, $\mathbf{E}[\Phi_t] \geq \Pr[\Phi_t > 0] \Phi_{\min}$. Thus after $t \geq (2/\beta^2)n^3 \log(\Phi_0/(\Phi_{\min}\varepsilon))$ steps, $\Pr[\Phi_t > 0] \leq \varepsilon$. We have $\Phi_0 \leq ab$ and $\Phi_{\min} = \cos(\beta(n/2 - 1)/n) > \cos(\beta/2) \approx (\pi - \beta)/2$. The optimal choice of β is $\pi - \Theta(1/\log n)$, but all that matters is that $\pi - \beta \rightarrow 0$ as $n \rightarrow \infty$ while $\log(1/(\pi - \beta)) \ll \log(ab)$. Substituting, we find that $t = (2/\pi^2 + \Theta(\log \log n / \log n))n^3 \log(ab/\varepsilon)$ steps are enough to ensure that the probability of coalescence is at least $1 - \varepsilon$. \square

REMARK. We show in Section 8 that when $a = b = n/2$, the coupling time is actually $(2/\pi^2)n^3 \log n$.

REMARK. Proving mixing time upper bounds via a contraction property in the “distance” between configurations is a fairly standard technique. Traditionally the distance has been measured in terms of Hamming distance or other integer-valued distance, which for our applications does not yield the requisite contraction property.

We get the spectral gap entries in Table 1 using similar reasoning [see also, e.g., Chen (1998)]:

PROPOSITION 3. *If a function Φ is strictly monotone increasing in the partial order of a reversible monotone Markov chain with top and bottom state, and whenever $X \preceq Y$, we have $\mathbf{E}[\Phi(Y') - \Phi(X')|X, Y] \leq (1 - \gamma)(\Phi(Y) - \Phi(X))$, then the spectral gap must be at least γ .*

PROOF. Perturb the stationary distribution by an eigenvector associated with the second largest eigenvalue λ and run the Markov chain starting from this distribution. After t steps the variation distance from stationarity is $A\lambda^t$. If we run the Markov chain starting from the top and bottom states, after t steps the states are different with probability at most $(\Phi(\hat{1}) - \Phi(\hat{0}))(1 - \gamma)^t / \min_{X \prec Y} (\Phi(Y) - \Phi(X))$. The coupling time bound on the variation distance gives $\lambda \leq 1 - \gamma$. \square

3.3. Lower bound. We obtain a lower bound on the mixing time when the rectangle is not too narrow:

THEOREM 4. *If $\min(a, b) \gg 1$, then after*

$$\frac{1 - o(1)}{\pi^2} n^3 \log \min(a, b)$$

steps, the variation distance from stationarity is $1 - o(1)$.

We use Lemma 1 with $\beta = \pi$ so that we get an exact expression for $E[\Phi_t]$. Then we bound $\text{Var}[\Phi_t]$ to show that the distribution of Φ_t is sharply concentrated about its expected value. When Φ_t and Φ_∞ are sharply concentrated about values that are far enough apart, the chain is far from equilibrium. [This second-moment approach was also used by Diaconis and Shahshahani (1987) to lower bound the mixing time of random walk on \mathbb{Z}_2^d and by Lee and Yau (1998) to lower bound the mixing time of the exclusion process on a circle.] The following technical lemma formalizes this argument and is used to derive the mixing time lower bounds we give in this article, except the bound proved in Section 9.4, where we need a generalization. (The coupling time lower bound proved in Section 8.2 uses a different approach altogether.)

LEMMA 5. *If a function Φ on the state space of a Markov chain satisfies $\mathbf{E}[\Phi(X_{t+1})|X_t] = (1 - \gamma)\Phi(X_t)$, and $\mathbf{E}[(\Delta\Phi)^2|X_t] \leq R$, where $\Delta\Phi = \Phi(X_{t+1}) - \Phi(X_t)$, then when the number of Markov chain steps t is bounded by*

$$t \leq \frac{\log \Phi_{\max} + \frac{1}{2} \log(\gamma\varepsilon/(4R))}{-\log(1 - \gamma)}$$

and $0 < \gamma \leq 2 - \sqrt{2} \doteq 0.58$ (or else $0 < \gamma \leq 1$ and t is odd), then the variation distance from stationarity is at least $1 - \varepsilon$.

Before proving Lemma 5, we show how to use it to prove Theorem 4.

PROOF OF THEOREM 4. By Lemma 1, our function Φ satisfies the contraction property required by Lemma 5 when $\beta = \pi$, and

$$\gamma = \frac{1 - \cos(\pi/n)}{n - 1} \underset{\sim}{\geq} \frac{\pi^2}{2n^3}.$$

The constraint on γ is satisfied when $n \geq 3$.

To get a bound R , observe that any path h can have at most $2 \min(a, b)$ local extrema, so $\Pr[\Delta\Phi \neq 0] \leq \min(a, b)/(n - 1)$. But, $|\Delta\Phi| \leq 1$, so $\max_h \mathbf{E}[(\Delta\Phi(h))^2] \leq \min(a, b)/(n - 1) \equiv R$.

The maximal path maximizes Φ , giving $\Phi_0 = \Theta(ab)$. Substituting into Lemma 5 and simplifying, $\gamma/R \sim 1/(n^2 \min(a, b))$, so the numerator becomes $\log(ab) - \log n - (1/2) \log \min(a, b) + O(1) = \log \min(a, b) + O(1)$ for bounded values of ε , giving our lower bound of $(1/\pi^2 - o(1))n^3 \log \min(a, b)$. \square

PROOF OF LEMMA 5. Let $\Phi_t = \Phi(X_t)$. By induction,

$$\mathbf{E}[\Phi_t | X_0] = \Phi_0(1 - \gamma)^t.$$

By our assumptions on γ , in equilibrium $\mathbf{E}[\Phi] = 0$.

With $\Delta\Phi$ denoting $\Phi_{t+1} - \Phi_t$, we have

$$\Phi_{t+1}^2 = \Phi_t^2 + 2\Phi_t \Delta\Phi + (\Delta\Phi)^2,$$

$$\mathbf{E}[\Phi_{t+1}^2 | \Phi_t] = (1 - 2\gamma)\Phi_t^2 + \mathbf{E}[(\Delta\Phi)^2 | \Phi_t] \leq (1 - 2\gamma)\Phi_t^2 + R$$

and so, by induction,

$$\mathbf{E}[\Phi_t^2] \leq \Phi_0^2(1 - 2\gamma)^t + \frac{R}{2\gamma},$$

then by subtracting $\mathbf{E}[\Phi_t]^2$,

$$\text{Var}[\Phi_t] \leq \Phi_0^2[(1 - 2\gamma)^t - (1 - \gamma)^{2t}] + \frac{R}{2\gamma} \leq \frac{R}{2\gamma}$$

for each t . To get the last line we used our constraints on γ and t : $(1 - \gamma)^2 = 1 - 2\gamma + \gamma^2 \geq 1 - 2\gamma$, so when t is odd, $(1 - \gamma)^{2t} \geq (1 - 2\gamma)^t$. When t is even, we need $(1 - \gamma)^2 \geq 2\gamma - 1$ as well, which is satisfied when $\gamma \leq 2 - \sqrt{2}$ or $\gamma \geq 2 + \sqrt{2}$.

From Chebychev's inequality,

$$\Pr[|\Phi_t - \mathbf{E}[\Phi_t]| \geq \sqrt{R/(2\gamma\varepsilon)}] \leq \varepsilon.$$

As $\mathbf{E}[\Phi_\infty] = 0$, if $\mathbf{E}[\Phi_t] \geq \sqrt{4R/(\gamma\varepsilon)}$, then the probability that Φ_t deviates below $\sqrt{R/(\gamma\varepsilon)}$ is at most $\varepsilon/2$, and the probability that Φ in stationarity deviates above this threshold is at most $\varepsilon/2$, so the variation distance between the distribution at time t and stationarity must be at least $1 - \varepsilon$. If we take the initial state to be the

one that maximizes Φ_0 , then

$$E[\Phi_t] = \Phi_{\max}(1 - \gamma)^t \geq \sqrt{4R/(\gamma\varepsilon)}$$

when

$$t \leq \frac{\log[\Phi_{\max} \div \sqrt{4R/(\gamma\varepsilon)}]}{-\log(1 - \gamma)}. \quad \square$$

3.4. *Intuition.* Since the expected value of the new height function is a certain local average of the current height function, the evolution of the height function h (or rather its expected value) proceeds approximately according to the rule

$$(7) \quad \frac{\partial h}{\partial t} = \text{const} \times \frac{\partial^2 h}{\partial x^2}.$$

Since the equation is linear in h , it is natural to consider its eigenfunctions, which are just the sinusoidal functions that are zero at the boundaries. We can decompose any given height function into a linear combination of these sinusoidal components and consider the evolution of each component independently. The displacement function Φ (when $\beta = \pi$) is just the coefficient of the principal mode (the sinusoidal function with longest period) when we decompose the height function in this way. The coefficient of the principal mode decays the slowest, making it the most useful for purposes of establishing a lower bound. For purposes of establishing an upper bound, the difference between two height functions also obeys (7). We used the fact that when two paths, one above the other, have the same displacement, then they are the same path. The coefficient of the principal mode is the only one for which we can guarantee this property, since the other eigenfunctions take on both positive and negative values in the interior. Thus the principal mode essentially controls the rate of convergence, making its coefficient a natural displacement function.

4. Card shuffling by adjacent transpositions. Next, we analyze the card shuffling Markov chain that transposes random adjacent pairs of cards. [This Markov chain is a special case of the move-ahead-one update rule that has been studied in self-organizing linear search; Hester and Hirschberg (1985) gave a survey.] We consider this Markov chain to be implemented according to the rule: pick a random adjacent pair $(i, i + 1)$ of cards, flip a coin c , and then sort the items in that adjacent pair if heads; otherwise reverse-sort them. The same random update defined by i and c may be applied to more than one permutation to obtain coupled Markov chains. A permutation on the numbers $1, \dots, n$ has associated with it $n + 1$ threshold functions, where the i th threshold function ($0 \leq i \leq n$) is a string of i 1's and $n - i$ 0's, with the 1's at the locations of the i largest numbers of the permutation. The permutation can be recovered from these threshold functions simply by adding them up (see Figure 3). When a random adjacent pair of numbers

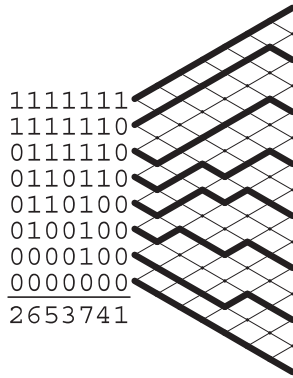


FIG. 3. The permutation 2653741 and its eight threshold functions shown as lattice paths.

within the permutation are transposed, the effect on any given threshold function is to transpose the same adjacent pair of 0's and 1's. The identity permutation $12 \cdots n$ and its reverse $n(n-1) \cdots 1$ give the minimal and maximal paths for any threshold function. So when the coupled Markov chains started at these two permutations coalesce (take on the same value), the grand coupling would take any starting permutation to this same value. We can therefore use our analysis of the Markov chain on lattice paths to analyze the Markov chain on permutations.

THEOREM 6. *After $(1/\pi^2 - o(1))n^3 \log n$ shuffles, the variation distance from stationarity is $1 - o(1)$ and the probability of coalescence is $o(1)$. After $(6/\pi^2 + o(1))n^3 \log n$ shuffles, the variation distance from stationarity is $o(1)$ and the probability of coalescence is $1 - o(1)$.*

We prove better upper bounds in Sections 6 and 8.1; the point here is to give a quick and easy proof.

PROOF OF THEOREM 6. The lower bound comes from considering the $\lfloor n/2 \rfloor$ th threshold function. By Theorem 4, after $[1/\pi^2 - o(1)]n^3 \log n$ steps the variation distance of just this one threshold function from stationarity is $1 - o(1)$. The variation distance from stationarity of the permutation itself is at least as large.

The upper bound follows from Theorem 2 when we take $\varepsilon = \delta/n$. As $ab/(\delta/n) \leq n^3/\delta$, after $[2/\pi^2 + o(1)]n^3 \log(n^3/\delta)$ steps the probability of any one given threshold function differing for the upper and lower permutations is less than or equal to δ/n . The probability that the upper and lower permutations differ is at most the expected number of threshold functions for which they differ, which is at most δ . Taking $\delta \ll 1$ but $\log(1/\delta) \ll \log n$, after $[6/\pi^2 + o(1)]n^3 \log n$ steps coalescence occurs with probability $1 - o(1)$. \square

5. Lozenge tilings.

5.1. *Background.* Random tilings or, equivalently, random perfect matchings, were originally studied in the physics community as a model for dimer systems [see, e.g., Fisher (1961) and Kasteleyn (1963) and references contained therein]. Physicists were interested in properties such as long-range correlations in the dimers. In the case of lozenge tilings, the dimers are the lozenges and the monomers are the two regular triangles contained in a lozenge. In recent years mathematicians have been studying random lozenge tilings (among other types of random tilings) and have proved a number of difficult theorems about their asymptotic properties. For instance, when a very large hexagonal region is randomly tiled by lozenges, with high probability the tiling will exhibit a certain circular shape and the density of each of the three orientations of lozenges, as a function of position, is also known [Cohn, Larsen and Propp (1998)]. [See also Cohn, Kenyon and Propp (2001).] Observations of random lozenge tilings of very large regions played an important role in the history of these theorems, since they indicated what sort of results might be true before they were proved, thereby guiding researchers in their efforts.

Consequently, there have been several articles [Propp and Wilson (1996), Luby, Randall and Sinclair (1995), Wilson (1997a) and Ciucu and Propp (1996)] on techniques to randomly generate lozenge tilings and other types of tilings. The first two of these articles used a Markov chain approach, while the second two used linear algebra. Propp and Wilson (1996) also introduced monotone-CFTP, which lets one efficiently generate random structures (e.g., lozenge tilings) using special Markov chains, without requiring any knowledge about the convergence rate of the Markov chain. It is the article by Luby, Randall and Sinclair (1995) that is most relevant to us here. In it they introduced novel Markov chains for generating lozenge tilings (and two other types of structures). In this case knowledge of the mixing time of these Markov chains does not help with the specific task of random generation, because monotone-CFTP, which determines on its own how long to run a Markov chain, may be applied to each of these Markov chains. Still there are several reasons to determine the mixing time: (1) in the same way that designers of efficient algorithms like to prove that the algorithms actually are efficient, it is desirable to have a proof that the Markov chain is rapidly mixing, (2) there are physical interpretations of the mixing properties of dimer systems [see the discussion of Destainville (2002) and references contained therein] and (3) there has been some speculation [Propp (1995–1997)] that knowledge of the convergence properties of these Markov chains can be converted into knowledge about random tilings of the whole plane (but this remains to be seen). For these reasons, Luby, Randall and Sinclair established polynomial time bounds on the convergence rates of each of their Markov chains. In this section we substantially improve the analysis of the lozenge-tiling Markov chain and in many cases our bounds differ by just a constant factor from the true convergence rate. As discussed

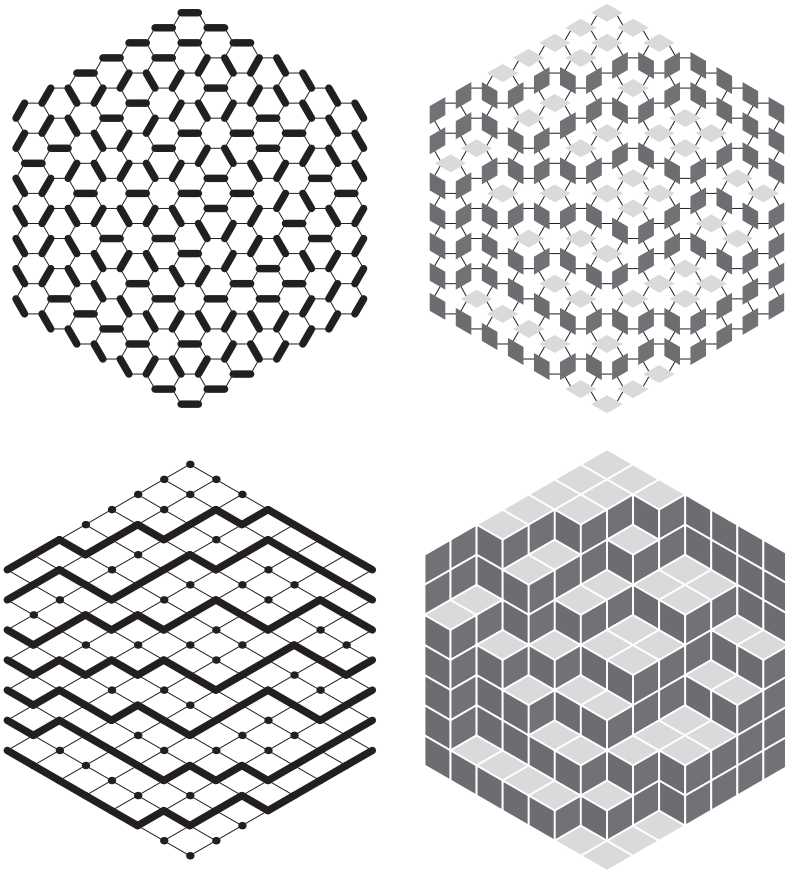


FIG. 4. Shown starting in the upper left and proceeding clockwise are (1) a random perfect matching (every vertex paired with exactly one neighbor) of a portion of the hexagonal lattice, (2) the same perfect matching where edges of the matching are shown as small lozenges, (3) same as 2, but the lozenges are large enough to touch one another, forming a lozenge tiling of a certain region and (4) same as 3, but horizontal lozenges are represented as dots, while the other types are represented as ascending or descending line segments, which form nonintersecting lattice paths. These transformations are bijective, so that any set of nonintersecting lattice paths corresponds to a lozenge tiling, which in turn corresponds to a perfect matching of the original hexagonal lattice graph.

in Section 5.5, Luby, Randall and Sinclair also analyzed other Markov chains for which it is not clear how to apply the methods of this section. Nonetheless, empirical studies suggest that these other Markov chains converge about as quickly as the lozenge-tiling Markov chain.

There is a well-known correspondence between dimers on the hexagonal lattice, lozenge tilings, and nonintersecting lattice paths of the type we considered in Section 3. Figure 4 illustrates this correspondence by showing a random perfect

matching of a region of the hexagonal lattice, an equivalent random tiling of a related region by lozenges, and an equivalent random collection of nonintersecting lattice paths. Following Luby, Randall and Sinclair, we use the lattice path representation of lozenge tilings.

5.2. Displacement function. In this section we apply the same techniques used to upper bound the mixing time of the lattice path Markov chain to upper bound the mixing time of a Markov chain for generating random lozenge tilings. There are several Markov chains for generating random lozenge tilings of regions (each of which possesses the monotonicity property required by monotone-CFTP): the one that we analyze was introduced by Luby, Randall and Sinclair. Two of these Markov chains use the lattice path representation of lozenge tilings and may be viewed as generalizations of the lattice path Markov chain that we studied already.

Consider the Markov chain that picks a random point on some lattice path and then randomly decides whether to try pushing it up or down. The Markov chain is connected because the top path can be pushed to its maximum height, then the next highest path and so on, so that each configuration can reach a unique maximal configuration. (Similarly, there is a unique minimal configuration.) The Markov chain is aperiodic since pushing the same lattice point twice in the same direction results in no change the second time. The Markov chain is symmetric, so its unique stationary distribution is the uniform distribution.

REMARK. Luby, Randall and Sinclair assumed that the region to be tiled is simply connected, since otherwise the lattice paths cannot cross the interior holes in the region, causing the state space to be disconnected. However, if we restrict the state space to configurations with a specified set of lattice paths passing under each interior hole, then this restricted state space is connected by these local moves. Our mixing time upper bounds will apply to each such connected component of the state space if the region to be tiled is not simply connected.

Such is the “local moves” Markov chain for lozenge tilings. Unfortunately, it is difficult to analyze in the same way that we analyzed the path Markov chain because the paths must remain disjoint. We would like to define the displacement Φ of the lozenge tiling to be the sum of the displacements of each of its paths, but computing $\mathbf{E}[\Delta\Phi]$ is difficult. We cannot compute the expected new height of a point on a lattice path simply by looking at the heights of its neighbors, since another lattice path may or may not be nearby and block its movement. Luby, Randall and Sinclair introduced “nonlocal moves” to circumvent this problem and make a Markov chain that is more tractable.

Consider a single lattice path site in isolation. Its height will change only if it is a local extremum and gets pushed up (if it is a minimum) or down (if it is a maximum). The idea behind the nonlocal moves is to preserve the expected change in height even if there are other lattice paths that might block the movement of this

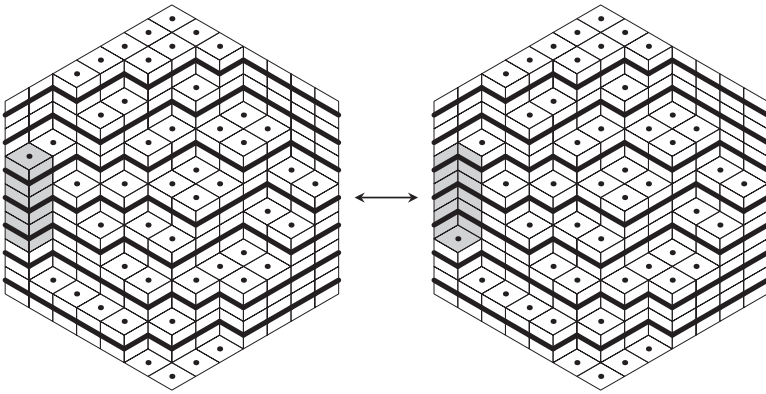


FIG. 5. A tower move for the Luby–Randall–Sinclair Markov chain for lozenge tilings. If the chain attempts to push up the third path from the bottom in the gray area, then it and the two paths blocking it are all three pushed up together with probability $1/3$.

local extremum. If there are k paths blocking the movement of the local extremum, then with probability $1/(k + 1)$ the corresponding points in the $k + 1$ paths are each moved. (Naturally if the border of the region itself prevents the movement of the paths, then the probability remains zero.) See Figure 5. This modified Markov chain is ergodic and symmetric (the probability that these $k + 1$ paths get pushed back to their original positions equals the probability that they were moved in the first place), so the uniform distribution remains the unique stationary distribution. Additionally, if there were no border effects, we would be able to compute $\mathbf{E}[\Delta\Phi]$, since when a site in a lattice path is selected by the Markov chain, the expected change in the total height function is determined by that point and its immediate neighbors on the same path. For hexagonal regions the borders cannot obstruct the movements of the lattice paths, and we are able to get both upper and lower bounds for these chains. For general regions we still obtain a good upper bound bound on the mixing time despite these “border effects.”

The constraints that the borders of the region impose are that certain locations of certain lattice paths have maximal or minimal values. For instance, at the start and end of a lattice path, the maximal and minimal values are identical.

Let w be the width of the lozenge-tiling region in the lattice path representation. That is, w is the distance between the leftmost start of a path and the rightmost end of a path. (For the path Markov chain, w was n .) As in the single path Markov chain, we find it convenient to center the region about the origin, so that the x coordinates of the points on the paths range from $-w/2$ to $w/2$. For a given set of nonintersecting lattice paths (also known as routing), let $h_i(x)$ be the height of the i th path in the routing at the given x coordinate. If the path includes x and $x - 1$, we have $h_i(x) = h_i(x - 1) \pm (1/2)$, so that pushing the path up or down changes its height by 1. For convenience, extend the definition of h_i to locations x where the i th path does not have a point, say by letting $h_i(x)$ take on its maximum possible

value consistent with the constraint $h_i(x) = h_i(x - 1) \pm (1/2)$. The “displacement” function of h that we use is

$$(8) \quad \Phi(h) = \sum_i \sum_{x=-w/2}^{w/2} h_i(x) \cos \frac{\beta x}{w},$$

where $0 \leq \beta \leq \pi$. This function Φ is the natural generalization of our earlier lattice path displacement function, but in the context of lozenge tilings we put “displacement” in quotes because it does not measure displacement from anything in particular.

Suppose that the Markov chain picks a particular location x on the i th path to try randomly pushing up or down. Let $h'_i(x)$ be the updated value of the gap at that location. If the i th path does not have an extremum at x , then $h'_i(x) = \mathbf{E}[h'_i(x)] = h_i(x) = [h_i(x - 1) + h_i(x + 1)]/2$, so

$$(9) \quad \mathbf{E}[\Delta\Phi(h)] = \left[\frac{h_i(x - 1) + h_i(x + 1)}{2} - h_i(x) \right] \cos \frac{\beta x}{w}.$$

Suppose instead that the i th path does have an extremum at x . In the absence of border constraints or interactions with other paths, the extremum is pushed up or down with probability $1/2$, so that we have $\mathbf{E}[h'_i(x)] = [h_i(x - 1) + h_i(x + 1)]/2$. Next we take into account interactions among the paths, but not border effects yet. Suppose the Markov chain attempts to push the given local extremum in the opposite direction that it is pointing, but that there are k paths blocking its movement. Because we are using nonlocal moves, with probability $1/(k + 1)$ each of these $k + 1$ paths is moved at location x , with each affecting $\Phi(h)$ by $\cos(\beta x/w)$. Thus (9) continues to hold true.

The only reason that (9) might fail is if path i has a local maximum at x and pushing it down violates the border constraints [in which case $=$ in (9) becomes $>$] or the path has a local minimum at x and pushing it up violates the border constraints [where $=$ in (9) becomes $<$]. For some regions, such as the hexagon, all the paths start at $-w/2$ and end at $w/2$, and the only border effects are that the endpoints of the paths stay fixed. For such regions (9) always holds for each path and each x such that $-w/2 < x < w/2$, and it holds for $x = \pm w/2$ as well when $\beta = \pi$. Using this equality, we derive a lower bound on the mixing time when the region is a hexagon. For general regions it is difficult to obtain a lower bound on the mixing time, but we still obtain an upper bound.

5.3. *Mixing time upper bound.* To get the upper bound we work with a pair of routings with heights \hat{h}_i and \check{h}_i , such that each path in the first routing lies above the corresponding path in the second routing. How we extended the definition of \hat{h}_i and \check{h}_i to locations x where path i does not exist was a bit arbitrary, but we did it the same way in both routings so that at these locations $\hat{h}_i(x) - \check{h}_i(x) = 0$. Then the gap function between the two routings is $g = \hat{h} - \check{h}$ and the gap is

$\Phi(g) = \Phi(\hat{h}) - \Phi(\check{h})$, which is zero when the routings are the same and positive otherwise.

When we pick the same location x on the same path in both routings and randomly push in the same direction, then from (9) we get

$$\begin{aligned} \mathbf{E}[\Delta\Phi(g)] &= \mathbf{E}[\Delta\Phi(\hat{h}) - \Delta\Phi(\check{h})] \\ (10) \qquad &= \left[\frac{g_i(x-1) + g_i(x+1)}{2} - g_i(x) \right] \cos \frac{\beta x}{w} \end{aligned}$$

unless the borders influence $\mathbf{E}[\Delta\Phi(g)]$. Suppose that the borders influence $\mathbf{E}[\Delta\Phi(\hat{h})]$ to be larger than the value given by (9). Then $\hat{h}_i(x)$ takes on its minimal possible value and is a local maximum, so $\hat{h}_i(x \pm 1)$ also take on their minimum possible values. But, \hat{h}_i dominates \check{h}_i , so $\check{h}_i(x)$ and $\check{h}_i(x \pm 1)$ also assume these same minimum possible values. Thus the right-hand side of (10) is zero. Since $\hat{h}_i(x)$ and $\check{h}_i(x)$ are immobile, the left-hand side is zero as well, so (10) continues to be true. Similarly, if the borders influence $\mathbf{E}[\Delta\Phi(\check{h})]$ to be smaller than the value given by (9), equation (10) continues to hold true. If on the other hand, the borders influence $\mathbf{E}[\Delta\Phi(\hat{h})]$ to be smaller than the value specified by (9) and/or influence $\mathbf{E}[\Delta\Phi(\check{h})]$ to be larger than that specified by (9), then we may replace the second equality in (10) with less than or equal to (\leq) to obtain a true statement.

Let p denote the number of internal points on the paths in the routing, that is, the number of places where the Markov chain might try to push a path up or down. (The p in this section was $n - 1$ in the section on the path Markov chain.) Then when we stop conditioning on a particular site of a certain path getting pushed, we use the same derivation used in the proof of Lemma 1 to conclude that

$$\mathbf{E}[\Delta\Phi] \leq \frac{-1 + \cos(\beta/w)}{p} \Phi$$

with equality when $\beta = \pi$, all paths start at $-w/2$ and end at $w/2$, and the only restrictions on the locations of the paths are that their endpoints are pinned down and they do not intersect. Then we use the same argument used in the proof of Theorem 2 to find that the Markov chain is within ε of uniformity after

$$\frac{2 + o(1)}{\beta^2} pw^2 \log \frac{m}{\Phi_{\min}\varepsilon}$$

steps, where m is the number of (local) moves separating the upper and lower configurations, and $\Phi_{\min} > \cos(\beta/2) \approx (\pi - \beta)/2$. Taking $\beta = \pi - \Theta(1/\log n)$ as before yields the following theorem:

THEOREM 7. *For the Luby–Randall–Sinclair lozenge-tiling Markov chain on a region which has width w , has m (local) moves separating the top and bottom configurations, and contains p places where a lattice path may be moved, after*

$$\frac{2 + o(1)}{\pi^2} pw^2 \log \frac{m}{\varepsilon}$$

steps, coalescence occurs except with probability ε . (For regions which are not simply connected, we mean coalescence within a given connected component of the state space.)

For the bound stated in the Introduction we used $p \leq n$ and $m \leq 2n^{3/2}$ [Luby, Randall and Sinclair (1995)].

For the hexagon of order ℓ , $p = 2\ell(\ell - 1)$, $w = 2\ell$ and $m = \ell^3$, so our mixing time bound is

$$\frac{48 + o(1)}{\pi^2} \ell^4 \log \ell.$$

5.4. Lower bound for the hexagon.

THEOREM 8. *For the regular hexagon with side length ℓ , the mixing time of the Luby–Randall–Sinclair Markov chain is at least $(8/\pi^2 - o(1))\ell^4 \log \ell$.*

PROOF. We apply Lemma 5 here to lower bound the mixing time of the lozenge-tiling Markov chain proposed by Luby, Randall and Sinclair, when the region is a regular hexagon with side lengths ℓ . Our potential function has the required contraction property, with $\gamma \approx \beta^2/(2pw^2) \approx \pi^2/(16\ell^4)$, and $\Phi_{\max} = \ell^3/2$. Since we are using nonlocal moves, $\Delta\Phi$ can be as large as ℓ . Suppose that when the Markov chain picks a site on a path and tries to push it, there are k paths in the way. With probability $1/(k + 1)$, $\Delta\Phi = (k + 1) \cos(\alpha)$; otherwise $\Delta\Phi = 0$. Conditioning on this site being selected, $\mathbf{E}[(\Delta\Phi)^2] \leq k + 1$ and, in general, $\mathbf{E}[(\Delta\Phi)^2] \leq R = \ell$. Applying Lemma 5, we obtain a mixing time lower bound of

$$\frac{(16 + o(1))\ell^4}{\pi^2} \left[3 \log \ell + \frac{1}{2} \log \left(\frac{\varepsilon}{\ell^5} \right) \right] = \frac{8 + o(1)}{\pi^2} \ell^4 \log \ell. \quad \square$$

5.5. *The other Luby–Randall–Sinclair Markov chains.* In addition to the Markov chain for lozenge tilings, Luby, Randall and Sinclair introduced Markov chains for domino tilings and Eulerian orientations. Randall (1998) pointed out that our analysis in Section 5.3 is readily adapted more or less unchanged to their domino-tiling Markov chains. It is much less obvious how to adapt the analysis to the Eulerian-orientation chains. The reason for this difference is that for the Eulerian-orientation Markov chains, the nonlocal “tower moves” overlap one another in a criss-cross fashion, whereas in the lozenge-tiling and domino-tiling chains, the towers are parallel to one another. We effectively gave each local move within a given tower the same weight, and if we do the same for the Eulerian-orientation chain, we get the trivial weighting, which does not have the desired contraction property.

5.6. *The local-moves Markov chains.* For a “normal” $\ell \times \ell$ region one might expect that for typical configurations, the towers in the nonlocal tower moves are fairly short, which suggests that while the Luby–Randall–Sinclair (LRS) Markov chain is much nicer to analyze rigorously, it may not be much faster on these regions than the local-moves Markov chain. (However for certain contrived regions, such as a pencil-shaped region consisting of one long tower, the LRS Markov chain will be much faster than the local-moves Markov chain.) Thus we have a heuristic prediction that the local-moves chain takes $\Theta(\ell^4 \log \ell)$ time to mix in “normal” $\ell \times \ell$ regions. Cohn (1995) tested this prediction by doing coupling time experiments and reported that it was “about right.” Then Henley (1997) did some detailed heuristic calculations and predicted that the relaxation time (reciprocal of the spectral gap) was $\Theta(\ell^4)$ for a variety of models that have associated height functions. In an interesting development, Destainville (2002) experimented with the local-moves chain for rhombus tilings of octagonal regions where there are $6 = \binom{4}{2}$ types of rhombuses and concluded that the $\Theta(\ell^4 \log \ell)$ estimate holds for these tilings as well.

From a rigorous standpoint, Randall and Tetali (2000) established a polynomial time upper bound on the mixing time of the local-moves chain. Their approach was to use the mixing time bound from Theorem 7 on the nonlocal-moves Markov chain to obtain a bound on the chain’s spectral gap, use techniques developed by Diaconis and Saloff-Coste (1993b) to compare the spectral gaps of the local and nonlocal Markov chains, and then derive a mixing time bound for the local-moves chain from its spectral gap. Their local-moves mixing time bound was $O(n^2 w^2 h^2 \log n)$, where n is the area, w is the width and h is the height, or in other words $O(\ell^8 \log \ell)$ for $\ell \times \ell$ regions. If rather than starting from our mixing time bound on the nonlocal Markov chain, one instead starts from our bound on the spectral gap (which was not explicitly given in an earlier version of this article), then the log factors disappear from the mixing time bound of the local-moves Markov chain.

6. The Karzanov–Khachiyan Markov chain. As mentioned in the Introduction, random generation of linear extensions can be used to approximately count the number of linear extensions of a partially ordered set, which is a #P-complete problem [Brightwell and Winkler (1991)]. Dyer, Frieze and Kannan (1991) showed one can generate a (approximately) random linear extension of a partial order in polynomial time using a Markov chain on a certain polytope. Matthews (1991) gave a different geometric Markov chain for random linear extensions that runs in time $O(550n^8 \log^3 n \log 1/\varepsilon)$. Karzanov and Khachiyan (1991) gave a combinatorial Markov chain for linear extensions and showed that it mixes in time $8n^5 \log(|\Omega|/\varepsilon) \leq O(n^6 \log n)$, where $|\Omega|$ is the number of linear extensions. Dyer and Frieze (1991) improved the mixing time bound to $O(n^4 \log(|\Omega|/\varepsilon)) \leq O(n^5 \log n)$. Felsner and Wernisch (1997) showed that

the Karzanov–Khachiyan Markov chain mixes in time $O(n^3 \log n)$ for a certain class of partial orders and that one can obtain an unbiased sample in this time. Bubley and Dyer (1998) showed that a related Markov chain mixes in time $O(n^3 \log n)$ and that the original Karzanov–Khachiyan Markov chain mixes in time $O(n^3 \log n \log(|\Omega|/\varepsilon)) \leq O(n^4 \log^2 n)$. We show here that the Karzanov–Khachiyan Markov chain mixes in time $O(n^3 \log n)$ and we exhibit a partial order for which the Karzanov–Khachiyan Markov chain and Bubley and Dyer’s variation of it both need order $n^3 \log n$ steps before they begin to get close to being random.

(Despite this progress, it still takes $O(n^5 \log^2 n \varepsilon^{-2} \log(n/\varepsilon))$ time to approximately count linear extensions to within a factor $1 + \varepsilon$ [Bubley and Dyer (1998)]. One can count the linear extensions of series-parallel posets much more quickly, so the aforementioned data-mining application [Mannila and Meeck (2000)] restricted its attention to these posets.)

Bubley and Dyer (1998) used their simple yet powerful method of path coupling [Bubley and Dyer (1997)] to bound the mixing time of Markov chains related to the Karzanov–Khachiyan Markov chain for random linear extensions. In their generalization, the items at positions i and $i + 1$ are considered with probability $f(i)$, and the Markov chain transposes these items with probability $1/2$, provided that doing so does not violate the partial order. For the Karzanov–Khachiyan Markov chain, $f(i) = 1/(n - 1)$ for $i = 1, \dots, n - 1$. Bubley and Dyer showed that if $f(i)$ is given by a parabola, $f(i) \propto i(n - i)$, then the Markov chain mixes in time $(1/3 + o(1))n^3 \log n$ and then they argued using eigenvalue comparison techniques that the original Karzanov–Khachiyan chain mixes in time no greater than $O(n^4 \log^2 n)$.

We show here how to generalize Bubley and Dyer’s analysis of these Markov chains and we obtain an upper bound of $(4/\pi^2 + o(1))n^3 \log n$ for the Karzanov–Khachiyan Markov chain. We remark that $4/\pi^2$ is about 22% larger than $1/3$, but selecting a uniformly random location is easier than selecting one according to a parabolic distribution. We will see in Section 7 that by doing updates in “sweeps” rather than at independent uniformly random locations, the required number of transpositions can be cut in half. So this analysis marginally improves but does not significantly impact the time needed to generate random linear extensions. Mainly it serves to illustrate the utility of the technique used throughout this article for analyzing the mixing time of a variety of Markov chains that have been studied before.

What we do is simply add weights to the distance function between linear extensions. If positions i and $j > i$ are transposed, Bubley and Dyer defined the “width” of the transposition to be $j - i$. We define the width to be $w(i, j) = \sum_{i \leq k < j} w(k)$, where the $w(k) \geq 0$ are to be chosen later. Given two linear extensions X and Y of the partial order, a transposition sequence was defined to be a sequence of linear extensions $X = Z_0, Z_1, \dots, Z_r = Y$ such that Z_k and Z_{k+1} differ by a single transposition. The weight of a transposition sequence is the sum

of the widths of the transpositions, and Bubley and Dyer defined the distance $\delta(X, Y)$ between linear extensions X and Y to be the weight of the minimum weight transposition sequence.

Bubley and Dyer showed in an Appendix that [when each $w(i) = 1$] $\delta(X, Y) = \delta_S(X, Y)$, where δ_S is Spearman’s footrule, which is defined by $\delta_S(X, Y) = (1/2) \sum_{i=1}^n |X(i) - Y(i)|$. The proof is also valid in the weighted scenario when the definition of δ_S is generalized to $\delta_S(X, Y) = (1/2) \sum_{i=1}^n w(\min(X(i), Y(i)), \max(X(i), y(i)))$.

Given two permutations A and B that differ by single transposition (i, j) and which are updated to A' and B' using a coupled Markov chain, Bubley and Dyer proved that

$$\begin{aligned} \mathbf{E}[\delta(A', B')] &\leq \delta(A, B) \\ &\quad + \frac{f(i-1)w(i-1) - f(i)w(i) - f(j-1)w(j-1) + f(j)w(j)}{2} \\ &= \delta(A, B)(1 - \gamma_{i,j}), \end{aligned}$$

where

$$\gamma_{i,j} = \frac{1 - f(i-1)w(i-1) + f(i)w(i) + f(j-1)w(j-1) - f(j)w(j)}{w(i) + \dots + w(j-1)}.$$

As always, they showed this assuming constant weights w , but the same proof holds for general positive weights w .

Letting $\gamma = \min_{i,j} \gamma_{i,j}$, their method of path coupling gives an upper bound of $(1/\gamma) \log(D/\varepsilon)$ on the number of steps before the variation distance from stationarity is at most ε , where D is the ratio of the maximum distance to the minimum positive distance. Observe that $\gamma_{i,j} = c$ (resp. $\gamma_{i,j} \geq c$) for each i and j if and only if $\gamma_{i,i+1} = c$ (resp. $\gamma_{i,i+1} \geq c$) for each i .

Given constant weights w , the choice of frequencies f that maximizes γ is given by a parabola. Therefore, Bubley and Dyer chose $f(i) = i(n - i)/K$, where the normalizing constant is $K = (n^3 - n)/6$. It is easily checked that $\gamma_{i,j} = 1/K$ when $j = i + 1$, so this holds for all i and j . With constant weights one can show $D \leq \lfloor n^2/4 \rfloor$, so their bound on the mixing time is $(1/3 + o(1))n^3 \log n$.

Given constant frequencies f , the optimal choice of w is sinusoidal. Let $w(i) = \cos(\beta(i/n - 1/2))$ with $0 \leq \beta \leq \pi$; these weights are positive as required. Since

$$\frac{-\cos(x - \delta) + 2 \cos(x) - \cos(x + \delta)}{2 \cos(x)} = 1 - \cos(\delta),$$

we have

$$\gamma_{i,j} \geq \frac{1 - \cos(\beta/n)}{n - 1} \geq \frac{\beta^2}{2n^3}$$

for $j = i + 1$ [we do not have equality when $i = 1$ or $i = n - 1$ since $f(0) = f(n) = 0$], so this bound holds for all i and j . As in the proof of Theorem 2, we take $\beta = \pi - \Theta(1/\log n)$ so that γ is large, while not making the ratio D of the maximum distance to the minimum positive distance too large. Then the upper bound on the mixing time is

$$\frac{2n^3}{\beta^2} \log \frac{D}{\varepsilon} = \frac{4 + o(1)}{\pi^2} n^3 \log n.$$

Of course when designing a Markov chain, we are free to pick both f and w . Optimizing them together would be an interesting challenge.

7. Sweeps versus independent updates. So far we have focused on updates where a random site is selected and then a local randomizing operation is performed at that site. Often in practice the various sites are updated in systematic “sweeps” rather than at random. For instance, for permutations or linear extensions, rather than randomize a random adjacent pair of items, one may instead randomize the items in positions $(1, 2)$ $(3, 4)$ $(5, 6) \dots$ and then do positions $(2, 3)$ $(4, 5)$ $(6, 7) \dots$. Likewise for lozenge tilings, one may randomize the lattice paths at all places where the x coordinate is even, then afterward randomize at all places where the x coordinate is odd. Call the first set of updates an even sweep and call the second set an odd sweep. In all cases where we have derived upper bounds on the mixing time of a Markov chain (i.e., grand-coupling time for a random path or permutation by random adjacent transpositions, grand-coupling time for Luby, Randall and Sinclair’s chain on lozenge tilings, and the mixing time for the Karzanov–Khachiyan Markov chain for random linear extensions), the same analysis that worked for independent updates at each step also works for sweeps. If one randomly chooses between even sweeps and odd sweeps, then we have the same contraction property, with γ scaled up by a factor of $(n - 1)/2$ (or $p/2$ in the case of lozenge tilings). The mixing time bounds are then roughly the same, though slightly better than the bounds we would get when using the same total number of transpositions (or path pushes) but at independent uniformly random locations. Successive even sweeps are redundant, as are successive odd sweeps, so when we alternate, we perform about half as many moves (in continuous time) to get the same value of the total variation distance or probability of not coupling.

Note that we have not proved that the mixing time is actually twice as fast, merely that our upper bound on it is half as large.

Perhaps more important than this factor of 2 savings is that many fewer random bits are required to do the updates, since the locations of the updates are deterministic. For Markov chains with simple moves such as these, generating pseudorandom bits can take an appreciable fraction of the total running time. Whether for this reason or for simplicity, in practice the algorithms used to generate random tilings have typically used systematic sweeps.

8. Lattice paths and permutations revisited. We have already given a quick-and-easy analysis of the adjacent-transposition Markov chain on lattice paths and on permutations, obtaining upper and lower bounds on the mixing time and coupling time that match to within constants. In this section we give a more refined analysis which improves these constants.

8.1. *Upper bounds.* Consider the Markov chain on permutations that exchanges a random adjacent pair with probability $1/2$ and the pairwise coupling for which the choice of adjacent pair is always the same in the two Markov chains. The decision of whether or not to exchange is also the same in both chains unless an exchange in one chain but not the other would decrease the Hamming distance between the two permutations, in which case the exchange is done in a random one of the two random permutations but not the other. This coupling was also used by Aldous (1983). Let us focus on how a given item moves in the two permutations. The state space is the $n \times n$ grid, which represents the location of the item in the two different permutations. A typical state (x, y) transits to its four neighboring states $(x \pm 1, y)$ and $(x, y \pm 1)$, each with probability $\alpha = 1/(2(n - 1))$, with the following exceptions: (1) if $x = y$, then the transitions are to $(x + 1, y + 1)$ and $(x - 1, y - 1)$, each with probability α , and (2) if the transition would be to a pair outside the $n \times n$ grid, the transition instead self-loops. Observe that the x coordinate is a simple random walk on a chain of length n and similarly for the y coordinate. Furthermore, any state (x, y) with $x \neq y$ is transient, so that eventually $x = y$. This takes about $\Theta(n^2/\alpha)$ steps. We need a good estimate of the probability that coalescence has not occurred after a large multiple of $\Theta(n^2/\alpha)$ steps, so we prove the following lemma:

LEMMA 9. *After T steps $\Pr[x_T \neq y_T] < 10 \exp[-T(1 - \cos(\pi/n))/(n - 1)]$.*

Before proving this lemma, we derive from it our bounds on the coupling times.

THEOREM 10. *For the above pairwise coupling on permutations, the number of steps before the probability of coalescence is at least $1 - \varepsilon$ is at most $(2/\pi^2 + o(1))n^3 \log(10n/\varepsilon)$.*

PROOF. After T steps the probability that the two permutations are still different is at most the expected number of items that are in different positions in the two permutations, which is at most $10n \exp[-T(1 - \cos(\pi/n))/(n - 1)]$. Setting this equal to ε gives the desired bound. \square

For the permutation Markov chain, the $(2/\pi^2 + o(1))n^3 \log n$ upper bound for the variation distance threshold and the $(4/\pi^2 + o(1))n^3 \log n$ upper bound for the separation distance threshold follow immediately. The same bounds for the lattice path Markov chain follow from projecting the permutation to the lattice path.

THEOREM 11. *For the sort/reverse-sort grand coupling on permutations, the number of steps before the probability of coalescence is at least $1 - \varepsilon$ is at most $(4/\pi^2 + o(1))n^3 \log(10n/\varepsilon)$. For lattice paths the sort/reverse-sort grand-coupling time is at most $(2/\pi^2 + o(1))n^3 \log(10n/\varepsilon)$.*

PROOF. What we have analyzed in Theorem 10 is a pairwise coupling of a Markov chain on permutations, that is, an update rule that updates pairs of permutations. This pairwise update rule does not extend to a grand coupling, that is, there is no update rule defined on all permutations such that pairs of permutations evolve according to the above pairwise coupling. However, let us look at the evolution of a threshold function of the two permutations. Normally both permutations are either sorted or reverse-sorted at a location, which corresponds to a push-down or push-up move in the threshold functions. The exceptional case where one permutation is sorted while the other is reverse-sorted occurs when items i and j are in adjacent locations x and $x + 1$ in one permutation, while the other permutation has items j and k at these locations, and either $i < j < k$ or $k < j < i$. Any given threshold function maps j to either 0 or 1. Then one of the two permutations has either an up slope or a down slope at locations x and $x + 1$, and for that permutation an observer is unable to tell whether a sort or reverse-sort operation was performed. Thus from the standpoint of an observer, it appears as if the two threshold functions evolved according to the monotone grand coupling considered earlier, and our bound on the pairwise coupling time for permutations translates to a bound on the grand-coupling time for lattice paths.

Next we can convert the bound on grand-coupling time for lattice paths into an upper bound on the grand-coupling time for permutations for the straightforward sort/reverse-sort coupling. After $(4/\pi^2 + o(1))n^3 \log n$ steps, the pairwise permutation coupling (and hence the lattice path grand coupling) has coalesced except with probability much less than $1/n$, so the permutation grand coupling has coalesced except with probability $\ll 1$. \square

Surprisingly, experiments suggest rather strongly that $4/\pi^2$ is in fact the correct constant, so that essentially nothing was lost in converting the coupling time bounds from permutations to lattice paths and back to permutations.

PROOF OF LEMMA 9. Since we are interested in the probability that the random walk has not hit the diagonal, and the regions below and above the diagonal behave symmetrically, let us consider the state-transition matrix M_n for the random walk above the diagonal ($x < y$), where the random walker reflects off the boundaries of the grid and dies when it hits the diagonal. The matrix M_n resembles a stochastic matrix, except that the row sum is less than 1 for those rows that correspond to states next to the diagonal. For the reader's convenience we proceed to diagonalize the matrix M_n ; other triangular regions with different boundary conditions have been similarly diagonalized, for example, in Kenyon,

Propp and Wilson (2000).

For $0 \leq j < k < n$, define the function

$$f_{j,k}(x, y) = \cos(j\pi x/n) \cos(k\pi y/n) - \cos(j\pi y/n) \cos(k\pi x/n).$$

For convenience let the values of the grid coordinates x and y range from $1/2$ to $n - 1/2$. Since

$$\begin{aligned} & f_{j,k}(x - 1, y) + f_{j,k}(x + 1, y) + f_{j,k}(x, y - 1) + f_{j,k}(x, y + 1) - 4f_{j,k}(x, y) \\ &= [2 \cos(j\pi/n) + 2 \cos(k\pi/n) - 4]f_{j,k}(x, y), \end{aligned}$$

$f_{j,k}$ is an eigenvector of the nearest-neighbor random walk on \mathbb{Z}^2 with transition probabilities α , and its eigenvalue is

$$\lambda_{j,k} = 1 + \alpha[2 \cos(j\pi/n) + 2 \cos(k\pi/n) - 4].$$

Since furthermore $f_{j,k}(x, x) = 0$, $f_{j,k}(x, y) = f_{j,k}(-x, y)$, $f_{j,k}(x, y) = f_{j,k}(x, -y)$, $f_{j,k}(x, y) = f_{j,k}(2n - x, y)$ and $f_{j,k}(x, y) = f_{j,k}(x, 2n - y)$, it follows that $f_{j,k}$ is also an eigenvector of M_n with eigenvalue $\lambda_{j,k}$.

Next we show that any two of these $n(n - 1)/2$ eigenvectors are orthogonal and that each is not identically zero:

$$\begin{aligned} & \sum_{1/2 \leq x < y \leq n-1/2} f_{j_1, k_1}(x, y) f_{j_2, k_2}(x, y) \\ &= \sum_{x < y} \left[\cos\left(\frac{j_1\pi x}{n}\right) \cos\left(\frac{k_1\pi y}{n}\right) \cos\left(\frac{j_2\pi x}{n}\right) \cos\left(\frac{k_2\pi y}{n}\right) \right. \\ & \quad + \cos\left(\frac{j_1\pi y}{n}\right) \cos\left(\frac{k_1\pi x}{n}\right) \cos\left(\frac{j_2\pi y}{n}\right) \cos\left(\frac{k_2\pi x}{n}\right) \\ & \quad - \cos\left(\frac{j_1\pi x}{n}\right) \cos\left(\frac{k_1\pi y}{n}\right) \cos\left(\frac{j_2\pi y}{n}\right) \cos\left(\frac{k_2\pi x}{n}\right) \\ & \quad \left. - \cos\left(\frac{j_1\pi y}{n}\right) \cos\left(\frac{k_1\pi x}{n}\right) \cos\left(\frac{j_2\pi x}{n}\right) \cos\left(\frac{k_2\pi y}{n}\right) \right] \\ &= \sum_{x, y} \left[\cos\left(\frac{j_1\pi x}{n}\right) \cos\left(\frac{k_1\pi y}{n}\right) \cos\left(\frac{j_2\pi x}{n}\right) \cos\left(\frac{k_2\pi y}{n}\right) \right. \\ & \quad \left. - \cos\left(\frac{j_1\pi x}{n}\right) \cos\left(\frac{k_1\pi y}{n}\right) \cos\left(\frac{j_2\pi y}{n}\right) \cos\left(\frac{k_2\pi x}{n}\right) \right] \\ &= \sum_x \cos\left(\frac{j_1\pi x}{n}\right) \cos\left(\frac{j_2\pi x}{n}\right) \sum_y \cos\left(\frac{k_1\pi y}{n}\right) \cos\left(\frac{k_2\pi y}{n}\right) \\ & \quad - \sum_x \cos\left(\frac{j_1\pi x}{n}\right) \cos\left(\frac{k_2\pi x}{n}\right) \sum_y \cos\left(\frac{k_1\pi y}{n}\right) \cos\left(\frac{j_2\pi y}{n}\right) \end{aligned}$$

$$\begin{aligned}
 &= \frac{\mathbb{1}_{j_1=j_2} + \mathbb{1}_{j_1=j_2=0}}{2}n \times \frac{\mathbb{1}_{k_1=k_2} + \mathbb{1}_{k_1=k_2=0}}{2}n \\
 &\quad - \frac{\mathbb{1}_{j_1=k_2} + \mathbb{1}_{j_1=k_2=0}}{2}n \times \frac{\mathbb{1}_{k_1=j_2} + \mathbb{1}_{k_1=j_2=0}}{2}n.
 \end{aligned}$$

Since $j_1 < k_1$ and $j_2 < k_2$, the second term is zero. The first term is also zero unless both $j_1 = j_2$ and $k_1 = k_2$, giving us orthogonality. If $(j_1, k_1) = (j_2, k_2) = (j, k)$, then we find that this inner product is $(1 + \mathbb{1}_{j=0})n^2/4$, so the eigenvectors are nontrivial. Hence, we have an orthogonal eigenbasis of the matrix M_n .

Suppose that the random walker starts at (x_0, y_0) . Let $\delta_{x_0, y_0}(x, y)$ be the function which is 1 at the starting location and 0 elsewhere. Let $J(x, y)$ denote the function which is 1 whenever $x < y$. We have

$$\begin{aligned}
 \Pr[x_T \neq y_T] &= (\delta_{x_0, y_0} M_n^T) \cdot J \\
 &= \left(\sum_{j < k} \frac{\delta_{x_0, y_0} \cdot f_{j, k}}{f_{j, k} \cdot f_{j, k}} f_{j, k} M_n^T \right) \cdot \left(\sum_{j < k} \frac{J \cdot f_{j, k}}{f_{j, k} \cdot f_{j, k}} f_{j, k} \right) \\
 &= \sum_{j < k} \frac{(\delta_{x_0, y_0} \cdot f_{j, k})(J \cdot f_{j, k})}{f_{j, k} \cdot f_{j, k}} \lambda_{j, k}^T \\
 &< \sum_{j < k} \frac{(2)(2\binom{n}{2})}{(1 + \mathbb{1}_{j=0})n^2/4} \lambda_{j, k}^T \\
 &< 8 \sum_{j < k} \lambda_{j, k}^T.
 \end{aligned}$$

We need a bound on $\lambda_{j, k}$ to bound this summation. To this end, consider the line passing through $(0, \cos 0)$ and $(t, \cos t)$. When $t = \pi/2$, the line is at least as high as $\cos s$ when $t \leq s \leq \pi$. If $t < \pi/2$, the line's slope increases toward 0, so it continues to be above $\cos s$ when $\pi/2 \leq s \leq \pi$, and by concavity of $\cos s$ for $0 \leq s \leq \pi/2$, the line is also above $\cos s$ when $t \leq s \leq \pi/2$. Taking $t = \pi/n$ (assume $n \geq 2$; the lemma is trivial if $n = 1$) and $s = j\pi/n$, we have

$$\begin{aligned}
 \cos(j\pi/n) &\leq 1 - (j\pi/n)/(\pi/n)(1 - \cos(\pi/n)), \\
 2 \cos(j\pi/n) - 2 &\leq -2j(1 - \cos(\pi/n)), \\
 2 \cos(j\pi/n) + 2 \cos(k\pi/n) - 4 &\leq -2(j+k)(1 - \cos(\pi/n)), \\
 \lambda_{j, k} &\leq 1 - \alpha 2(j+k)(1 - \cos(\pi/n)).
 \end{aligned}$$

Let $c = 2\alpha(1 - \cos(\pi/n)) \approx \alpha\pi^2/n^2$ so that

$$\lambda_{j, k} \leq 1 - (j+k)c < \exp[-(j+k)c].$$

Since $\alpha = 1/(2(n-1))$, $\lambda_{j, k} \geq 1 - 8/[2(n-1)] \geq 0$ when $n \geq 5$, and $\lambda_{j, k} \geq 0$

for $n = 2, 3, 4$ as well. So we may take the T th power of both sides to get

$$\lambda_{j,k}^T \leq \exp[-(j+k)cT],$$

$$\sum_{k=j+1}^{\infty} \lambda_{j,k}^T \leq \frac{\exp[-(2j+1)cT]}{1 - \exp(-cT)},$$

$$\Pr[x_T \neq y_T] < 8 \sum_{0 \leq j < k} \lambda_{j,k}^T \leq 8 \frac{\exp[-cT]}{[1 - \exp(-2cT)][1 - \exp(-cT)]}.$$

The lemma is trivial unless $\exp[-cT] \leq 1/10$, in which case $8/[1 - \exp(-2cT)]/[1 - \exp(-cT)] \leq 8 \times 100/99 \times 10/9 = 8000/891 < 10$, so that

$$\Pr[x_T \neq y_T] < 10 \exp[-cT]. \quad \square$$

8.2. Lower bounds.

THEOREM 12. *For the Markov chain on lattice paths in an $n/2 \times n/2$ box, the time it takes the top path and the bottom path to coalesce is with high probability at least $(1 - o(1))2/\pi^2 n^3 \log n$.*

PROOF. As an upper lattice path \hat{h} and lower lattice path \check{h} evolve together via the push-down/push-up coupling, let us look at the difference path $h = \hat{h} - \check{h}$. If \hat{h} goes up and \check{h} goes down, which we denote $\overset{U}{D}$, then the difference path h goes up, which we denote with U. If \hat{h} goes down and \check{h} goes up ($\overset{D}{U}$), then h goes down (D). In the remaining two cases ($\overset{U}{U}$ and $\overset{D}{D}$), the difference path remains flat (F). We may view the difference path as a string of U, F, and D particles, and it is easy to check that the evolution of the difference path is a Markov process: If the particles at the updated site are $\overset{UU}{DD}$, then they remain $\overset{UU}{DD} = UU$. If a $\overset{UD}{DU}$ is updated, the result is either $\overset{UD}{UD} = FF$ or $\overset{DU}{DU} = FF$. If a UF is updated, the underlying paths might be $\overset{UU}{DU}$ and then change to $\overset{UU}{UD} = FU$ or $\overset{UU}{DU} = UF$, or the underlying paths might be $\overset{UD}{DD}$ and then change to $\overset{UD}{DD} = UF$ or $\overset{DU}{DD} = FU$. Likewise, if an FF is updated, there are four possibilities for the underlying paths, and in each case the updated configuration is FF. The other cases (DD, DU, DF, FU and FD) are similar and related to the above cases by symmetry. We summarize the update rules for the string of U's, D's and F's as follows: Pick a random adjacent pair and with probability 1/2 exchange them; when a D and U are exchanged past each other, they both turn into F's. If we start with the top path and bottom path, then in the difference path every U is to the left of every D.

We do a number of comparisons between a random permutation σ and the difference lattice path h . For the k th comparison ($0 \leq k \leq n$), look at the locations of cards $1, \dots, k$ and, in particular, their relative order. Let $\tau_U(1)$ denote the first of these cards encountered in a left-to-right scan and, in general, $\tau_U(i)$, $1 \leq i \leq k$, denotes the i th such card encountered. Label the first k U's of the difference path

with the numbers $\tau_U(1), \dots, \tau_U(k)$. Similarly let $\tau_D(i), 1 \leq i \leq k$, denote the i th card from the cards $n + 1 - k, \dots, n$ to be encountered in a right-to-left scan of the permutation and label the i th to last D of the difference path with $\tau_D(i)$. We leave the remaining particles of the difference path unlabeled. When we evolve the difference path via random exchanges, we let labeled particles be exchanged past each other, but a labeled U or D may not be exchanged past an unlabeled U or D. This rule for the labels does not affect the evolution of the unlabeled difference path, but it is important for our understanding of it.

Initially for each $1 \leq i \leq k$, the position of card $\tau_U(i)$ in the difference path is weakly to the left of card $\tau_U(i)$ in the permutation, while the position of card $\tau_D(i)$ in the difference path is weakly to the right of card $\tau_D(i)$ in the permutation. We pick the same random adjacent pair in the permutation as in the labeled difference path and make the same decision as to whether or not to exchange the adjacent items. Consider the first time that the above invariant fails to hold, say that card $\tau_U(i)$ in the labeled difference path moves to the right of card $\tau_U(i)$ in the permutation. On the previous step, card $\tau_U(i)$ was in the same location in the difference path and the permutation. The exchange could not have been to the right of the card $\tau_U(i)$, because exchanges in the permutation always succeed; neither could the exchange be to the left of card $\tau_U(i)$, because any particle to the left of card $\tau_U(i)$ is either an F or a labeled U and such exchanges succeed. Thus the invariant is maintained.

Consider the locations of two cards i and j in a random permutation σ or two labels in the difference path h . Let the weighted gap between them be defined by $\text{wgap}(i, j) = \sum_x \sin(\pi x/n)$, where the sum is taken over positions x between the two cards and is negative if card j occurs before card i . Within a random permutation σ we have

$$\begin{aligned} \mathbf{E}[|\text{wgap}_\sigma(i, j)|] &= \frac{\sum_{x=1}^{n-1} x(n-x) \sin(\pi x/n)}{\binom{n}{2}} \\ &\approx 2n \int_0^1 u(1-u) \sin(\pi u) du = \frac{8}{\pi^3} n. \end{aligned}$$

The area under the difference path is the sum of the locations of the D particles minus the sum of the locations of the U particles. The potential function (the weighted area) is

$$\begin{aligned} \Phi(h) &= \sum_{x=0}^n h(x) \sin \frac{\pi x}{n} \\ &= \sum_{i=1}^{\max_x h(x)} \text{wgap}_h(i\text{th U}, i\text{th D}) \\ &\geq \sum_{i=1}^k \text{wgap}_h(i, n + 1 - i) \end{aligned}$$

if $k \leq \max_x h(x)$. As $\text{wgap}_h(i, n + 1 - i) \geq \text{wgap}_\sigma(i, n + 1 - i)$ and also $\text{wgap}_h(i, n + 1 - i) \geq 0$ whenever $i \leq \max_x h(x)$, we have

$$\Phi(h) \geq \sum_{i=1}^k \max(0, \text{wgap}_\sigma(i, n + 1 - i))$$

for any $k \leq \max_x h(x)$, so

$$\Phi(h) \geq \sum_{i=1}^{n/2} \mathbb{1}_{i \leq \max_x h(x)} \max(0, \text{wgap}_\sigma(i, n + 1 - i)).$$

Since we started with a random permutation and the dynamics are reversible, then even conditional upon all the past moves, the permutation is still uniformly random. In particular, $\text{wgap}_\sigma(i, n + 1 - i)$ is independent of the maximum height $\max_x h(x)$ of the difference path, so that

$$\mathbf{E}[\Phi(h)] \geq \sum_{i=1}^{n/2} \Pr\left[i \leq \max_x h(x)\right] \mathbf{E}[\max(0, \text{wgap}(i, n + 1 - i))],$$

$$\mathbf{E}[\Phi(h)] \geq \mathbf{E}\left[\max_x h(x)\right] \left[\left(\frac{8}{\pi^3} + o(1)\right) \frac{n}{2}\right],$$

$$\mathbf{E}\left[\max_x h(x)\right] \leq (1 + o(1)) \frac{\pi^3}{4n} \Phi(h_0) (1 - \gamma)^t$$

and since $\Phi(h_0) \doteq (2/\pi^2)n^2$,

$$\mathbf{E}\left[\max_x h(x)\right] \leq (1 + o(1)) \frac{\pi}{2} n (1 - \gamma)^t.$$

Note that this gives another proof that coalescence is likely after $t = (2/\pi^2 + o(1))n^3 \log n$ steps.

Notice that the difference path never changes by more than 1 at a time and only if a U or a D particle moves. There are $2 \max_x h(x)$ U and D particles, each particle can move in one of two directions, and a given proposed exchange occurs with probability 1/2. Thus

$$\mathbf{E}[\Delta \Phi^2 | h(\cdot)] \leq 2 \max_x h(x) / (n - 1),$$

$$\mathbf{E}[\Phi^2(h_t) | h_{t-1}] = (1 - 2\gamma)\Phi^2(h_{t-1}) + \mathbf{E}[\Delta \Phi^2 | h_{t-1}(\cdot)],$$

$$\mathbf{E}[\Phi^2(h_t)] \leq (1 - 2\gamma)\mathbf{E}[\Phi^2(h_{t-1})] + \mathbf{E}[\mathbf{E}[\Delta \Phi^2 | h_{t-1}(\cdot)]],$$

$$\mathbf{E}[\Phi^2(h_t)] \leq (1 - 2\gamma)\mathbf{E}[\Phi^2(h_{t-1})] + (\pi + o(1))(1 - \gamma)^{t-1},$$

where here the $o(1)$ term depends only on n . By induction,

$$\mathbf{E}[\Phi^2(h_t)] \leq (1 - 2\gamma)^t \Phi^2(h_0) + (\pi + o(1))(1 - \gamma)^t / \gamma.$$

Subtracting $\mathbf{E}[\Phi(h_t)]^2 = (1 - \gamma)^{2t} \Phi^2(h_0)$,

$$\begin{aligned} \text{Var}[\Phi(h_t)] &\leq (\pi + o(1))(1 - \gamma)^t / \gamma, \\ \text{Var}[\Phi(h_t)] &\leq (2/\pi + o(1))n^3(1 - \gamma)^t, \\ \text{Var}[\Phi(h_t)] &\leq (\pi + o(1))n\Phi(h_t). \end{aligned}$$

Thus if $\Phi_t \gg \pi n$, w.h.p. $\Phi(h_t) > 0$, so that the time until coalescence is w.h.p. at least $(1 - o(1))2/\pi^2 n^3 \log n$. \square

9. Exclusion and interchange processes. In this section we show how to apply Lemma 5 to lower bound the convergence rate of exclusion and interchange processes. Several of these Markov chains were studied by Diaconis and Saloff-Coste (1993a), who derived upper bounds on their mixing times but did not have lower bounds that matched to within constant factors. The lower bounds derived here match (most of) their upper bounds to within constant factors.

The interchange process describes particles moving around on an undirected (but possibly weighted) graph. At each time step, a random edge of the graph is selected and the particles at either endpoint of that edge are exchanged. The particles could be 0's and 1's or they could, for instance, be distinct numbers from 1 up to the number n of vertices.

Lee and Yau (1998) studied the logarithmic Sobolev constants and the L_2 -mixing times of some exclusion and exchange processes. The exclusion process can also be viewed as the infinite-temperature limit of Kawasaki dynamics for the Ising model; see, for example, Cancrini and Martinelli (2000) or Lu and Yau (1993).

Let $Q_{x,y}$ denote the probability that the Markov chain exchanges the particles at locations x and y , and for convenience let $Q_{x,x} = 1 - \sum_{y \neq x} Q_{x,y}$. Then $Q_{x,y}$ is the state-transition matrix for the location of a particular particle. Let v be a right eigenvector of matrix Q (the matrix is symmetric, so the left and right eigenvectors are the same) with eigenvalue $1 - \gamma$. The lemma requires $\gamma > 0$, but, in general, one would expect that eigenvectors with smaller γ 's give better lower bounds.

For convenience let us assume for the moment that all the particles are distinguishable, so that the state of the Markov chain is a permutation σ of size n . Define

$$\Phi(\sigma) = \sum_{i=1}^k v_{\sigma(i)}$$

and $\Phi_{\max} = \max_{\sigma} \Phi(\sigma)$. After one step of the Markov chain we have

$$\begin{aligned} \mathbf{E}[\Phi(\sigma')|\sigma] &= \sum_{i=1}^k \mathbf{E}[v_{\sigma'(i)}|\sigma] \\ &= \sum_{i=1}^k \sum_y Q_{\sigma(i),y} v_y \\ &= \sum_{i=1}^k (1 - \gamma) v_{\sigma(i)} \\ &= (1 - \gamma) \Phi(\sigma) \end{aligned}$$

so that this Φ has the contraction property required by Lemma 5 (assuming $0 < \gamma \leq 2 - \sqrt{2}$). In effect we used an eigenvector of the graph to define an eigenvector of the interchange process. For further information comparing the eigenvalues of the graph with those of the interchange process, see, for example, Handjani and Jungreis (1996).

With $\Delta\Phi = \Phi(\sigma') - \Phi(\sigma)$ denoting the change in Φ that occurs after one step of the Markov chain, let R be an upper bound on the largest value that $\mathbf{E}[(\Delta\Phi)^2]$ can take. In general, we can take

$$R \leq \max_{x,y: Q_{x,y}>0} (v_x - v_y)^2,$$

but in some cases we might find a better bound.

In general it need not be the case that all the particles are distinguishable. If there is a set A of k particles that are distinguishable from the remaining $n - k$ particles, then we can still define

$$\Phi(\text{state}) = \sum_{x: \text{particle of type } A \text{ at location } x} v_x.$$

We arbitrarily label the A particles $1, \dots, k$ and the remaining particles $k + 1, \dots, n$. The evolution of Φ is exactly the same as it was when all the particles were distinguishable.

One may add self-loops to the Markov chain to avoid periodicity problems—say the probability of a nontrivial transition is α . Typical choices for α are $\alpha = 1/2$ or $\alpha \rightarrow 0$ (continuous time).

9.1. *Shuffling cards on a hypercube.* Here we lower bound the mixing time of the Markov chain considered by Diaconis and Saloff-Coste that shuffles cards via random transpositions where each transposition is an edge of the hypercube. The underlying graph from which we need an eigenvector is the hypercube \mathbb{Z}_2^d , and the state space of the Markov chain is $\mathbb{Z}_2^d!$ or $\binom{\mathbb{Z}_2^d}{2^{d-1}}$ depending on whether we want to shuffle distinct particles or 2^{d-1} 0's and 2^{d-1} 1's (the same lower bound applies to

both cases). [Here $V!$ denotes the set of permutations of a set V and $\binom{V}{n}$ denotes the set of subsets of V containing n items.]

We can take our eigenvector to be the function that is 1 on those vertices of the hypercube that have first coordinate 0 and is -1 on the other vertices. When we follow a particular particle, the probability that the particle gets moved across the first coordinate is $\alpha/(d2^{d-1})$, so our function is an eigenvector for which $\gamma = \alpha/(d2^{d-2})$. Here $\Phi_{\max} = 2^{d-1}$ and our bound on R is 4α . Substituting into Lemma 5, we obtain a lower bound of

$$(1 - o(1)) \frac{d2^{d-2}}{\alpha} \left[\log 2^{d-1} + \frac{1}{2} \log \frac{\varepsilon}{16d2^{d-2}} \right] = (1 - o(1)) \frac{\log 2}{8\alpha} d^2 2^d$$

for bounded values of ε . It appears that this bound is correct up to constant factors.

9.2. High-dimensional product graphs. For a fixed connected graph G , consider the nearest-neighbor random walk on G^d . We can imagine that there is a particle in each of d disjoint copies of the graph G , where the particle in the i th copy of G gives the i th coordinate of the walker. At each time step a particle chosen from a random copy of G makes a move. For example, the Ehrenfest urn model from statistical mechanics is essentially a random walk on \mathbb{Z}_2^d , so here G consists of two vertices and an edge. A common choice for the factor α by which to slow down the walk is $\alpha = d/(d + 1)$, in addition to the usual $\alpha = 1/2$ and $\alpha \rightarrow 0$.

The mixing threshold for random walk on G^d has already been determined, so it is an instructive exercise to check that Lemma 5 gives a sharp lower bound in this case. For \mathbb{Z}_2^d , Diaconis and Shahshahani (1987) showed that there is a sharp variation mixing threshold at $(1/4 \pm o(1))d \log d$ steps. Aldous and Fill [(2004), Chapter 7, Section 1.7] stated that the same approach works for G^d for more general graphs G . In continuous time, Diaconis and Saloff-Coste [(1996), Theorem 2.9] proved an upper bound on the variation mixing time for random walk on G^d . The lower bound that we get from Lemma 5 differs from the upper bound by a factor of $1 - o(1)$.

Aldous and Diaconis [(1987), Section 7] determined the mixing time threshold of a related random walk on G^d , where at each time step, the particles in each copy of G get moved. We note that the lower bound from Lemma 5 is tight in this case as well.

To lower bound the mixing time of random walk on G^d , the underlying graph for which we need an eigenvector is $\overbrace{G + \cdots + G}^d$. Suppose that we have an eigenvector v of G with eigenvalue $1 - \gamma_0$, where γ_0 is the spectral gap. We take our eigenvector to be the canonical extension of v , that is, the function that assigns to each vertex x of $G + \cdots + G$ the value of the eigenvector v in the copy of G in which x resides. The probability that a particular particle gets moved at a time step is α/d , so our function is an eigenvector for which $\gamma = \alpha\gamma_0/d$. [For the walks considered in Aldous and Diaconis (1987), we have $\gamma = \alpha\gamma_0$.] Here $\Phi_{\max} = \Theta(d)$

and our bound on R is $\Theta(\alpha)$. [For the walks considered in Aldous and Diaconis (1987), the bound on R is $\Theta(\alpha d)$.] Substituting into Lemma 5, we obtain a lower bound of

$$(11) \quad \frac{\log \Theta(d) + \frac{1}{2} \log(\varepsilon \alpha \gamma_0 / (d \Theta(\alpha)))}{\log 1 / (1 - \alpha \gamma_0 / d)} = (1 - o(1)) \frac{d \log d}{2 \alpha \gamma_0}$$

when ε does not go to zero too quickly.

REMARK. This application of Lemma 5 is the natural generalization of the approach that Diaconis and Shahshahani (1987) used to get their lower bound on the mixing time of \mathbb{Z}_2^d . For the walk on \mathbb{Z}_2^d , the eigenvalues of \mathbb{Z}_2 are 1 and -1 , so $\gamma_0 = 2$, which upon substitution into lower bound (11) gives the familiar $\frac{1}{4} d \log d$. For \mathbb{Z}_2^d the eigenvector used above just counts the difference between the number of 1's and the number of 0's, which is the same test function that Diaconis and Shahshahani (1987) used.

9.3. *Shuffling cards on a grid.* Here the cards (or 0's and 1's) are arranged on an $\ell \times m$ grid. The state space is $([\ell] \times [m])!$ or $\binom{[\ell] \times [m]}{[\ell m / 2]}$, and the graph for which we need an eigenvalue is the $\ell \times m$ grid. (Here $[n]$ denotes $\{1, 2, \dots, n\}$ and we identify the vertices of the grid with $[\ell] \times [m]$.) When $\ell = m$, Diaconis and Saloff-Coste showed that order $(\ell m)^2 \log(\ell m)$ steps suffice for the Markov chain to equilibrate and they conjectured that this is the correct order of magnitude (but see Section 9.4).

Our asymptotic results will be valid as $\max(\ell, m)$ gets large. For convenience suppose that $\ell \geq m$. Suppose the vertices are labeled as (i, j) for $1 \leq i \leq \ell$ and $1 \leq j \leq m$. We can take our eigenvector to be the function that is $\cos(\pi(i - 1/2)/\ell)$ at vertex (i, j) . There are $E = \ell(m - 1) + m(\ell - 1)$ edges of this grid graph. One can verify that this function is an eigenvector with eigenvalue

$$1 - \gamma = 1 - 2\alpha/E + (\alpha/E)2 \cos(\pi/\ell),$$

so $\gamma \sim \alpha\pi^2/(E\ell^2)$. The bound for R is on the order $\alpha(1/\ell)^2$, and Φ_{\max} is on the order ℓm . Applying Lemma 5, we obtain a mixing time lower bound of

$$\begin{aligned} & (1 - o(1)) \frac{E\ell^2}{\alpha\pi^2} \left[\log \Theta(\ell m) + \frac{1}{2} \log \Theta(\varepsilon/E) \right] \\ & = (1 - o(1)) \frac{\ell^2(\ell - 1/2)(m - 1/2)}{\alpha\pi^2} \log(\ell m) \end{aligned}$$

when ε does not get too small too quickly as ℓm gets large.

REMARK. We can recover our mixing time lower bound for shuffling cards by adjacent transpositions by substituting $m = 1$ and $\alpha = 1/2$ into our lower bound for the $\ell \times m$ grid.

9.4. *Diaconis and Saloff-Coste's grid shuffling process.* The actual Markov chain that Diaconis and Saloff-Coste considered had transposition probabilities that were slightly higher along edges that touched the border of the grid. This was because their update rule was to pick a random vertex and exchange the particles along a random edge incident to that vertex. Thus each edge (u, v) is selected with probability proportional to $1/d(u) + 1/d(v)$, where $d(u)$ and $d(v)$ are the degrees of its endpoints. In some sense it is clear that the slight nonuniformity in the probability with which we select edges cannot affect the mixing time too much, but in the Introduction we promised a rigorous lower bound for Diaconis and Saloff-Coste's Markov chain, so we shall supply one. Finding an explicit eigenvector given these boundary conditions seems somewhat painful, as does approximating one with sufficient accuracy, so we take a different approach. We show how to obtain mixing time lower bounds using only an approximate contraction property.

Rather than try to approximate an eigenvector, it is easier to make use of the fact that we have an exact eigenvector Φ to an approximate state-transition matrix, namely the state-transition matrix considered in Section 9.3 and the eigenvector Φ that we used there:

$$\Phi(S) = \sum_{i,j: \text{state } S \text{ has particle at } (i,j)} \cos(\pi(i - 1/2)/\ell).$$

[For convenience we slow down the chain in Section 9.3 by a factor of $\alpha = E/(2\ell m) = 1 - 1/(2\ell) - 1/(2m)$ when doing the comparisons, so that the probability of a transition occurring on a given edge is simply $1/(2\ell m)$.] After one update of the approximate state-transition matrix, we have $\mathbf{E}[\Phi(S')|S] = (1 - \gamma)\Phi(S)$ with γ as given above, but with the actual state-transition matrix, there is an $O((\ell + m)/(\ell m))$ chance that a vertex on the boundary is selected, causing the exchange process to do something different than the approximate exchange process. To bound

$$|\mathbf{E}[\Phi(S')|S] - (1 - \gamma)\Phi(S)|,$$

let us focus on a single particle at a time and then make use of linearity of expectations and the triangle inequality. If the particle is near the boundary, the corresponding difference is at most $O(1/m\ell^3)$; if the particle is not near the boundary, then the corresponding difference is 0. Since there are $O(\ell + m)$ particles near the boundary, with $\delta = O(1/\ell^3 + 1/(m\ell^2))$ we have

$$\mathbf{E}[\Phi(S')|S] = (1 - \gamma)\Phi(S) \pm \delta.$$

By induction,

$$\mathbf{E}[\Phi(S_t)|S_0] = (1 - \gamma)^t \Phi(S_0) \pm \frac{\delta}{\gamma}.$$

Likewise

$$\begin{aligned} \mathbf{E}[\Phi^2(S_{t+1})|S_t] &= \Phi^2(S_t) + 2\Phi(S_t)\mathbf{E}[\Delta\Phi|S_t] + \mathbf{E}[(\Delta\Phi)^2|S_t] \\ &= (1 - 2\gamma)\Phi^2(S_t) \pm 2\Phi(S_t)\delta + \mathbf{E}[(\Delta\Phi)^2|S_t], \\ \mathbf{E}[\Phi^2(S_{t+1})] &= (1 - 2\gamma)\mathbf{E}[\Phi^2(S_t)] \pm 2\Phi(S_0)\delta(1 - \gamma)^t \pm \frac{2\delta^2}{\gamma} + \mathbf{E}[(\Delta\Phi)^2] \end{aligned}$$

and so, by induction,

$$\mathbf{E}[\Phi^2(S_t)] \leq (1 - 2\gamma)^t \Phi(S_0) + \frac{2\Phi(S_0)\delta}{\gamma}(1 - \gamma)^t + \frac{\max \mathbf{E}[(\Delta\Phi)^2]}{2\gamma} + \frac{\delta^2}{\gamma^2}.$$

Subtracting off $\mathbf{E}[\Phi(S_t)]^2$ we get

$$\text{Var}[\Phi(S_t)] \leq \frac{4\Phi(S_0)\delta}{\gamma}(1 - \gamma)^t + \frac{\max \mathbf{E}[(\Delta\Phi)^2]}{2\gamma} + \frac{\delta^2}{\gamma^2},$$

where in the last step we have assumed as in Lemma 5 that $1 - 2\gamma$ is not excessively negative, that is, that $0 < \gamma \leq 2 - \sqrt{2} \doteq 0.58$ (or else $0 < \gamma \leq 1$ and t is odd).

For convenience let us follow Diaconis and Saloff-Coste in assuming $\ell = m$, so that $\gamma = (\pi^2/2 + o(1))/\ell^4$ and $\delta = O(1/\ell^3)$. We also have $\Phi(S_0) = \Theta(\ell^2)$ when all the particles start on one side, and $\max \mathbf{E}[(\Delta\Phi)^2] = \Theta(1/\ell^2)$. Let K be a suitably large parameter to be selected in a moment. Pick $t = \log(\ell/K)/\gamma$, so that $\mathbf{E}[\Phi(S_t)] = \Theta(K\ell)$ and $\text{Var}[\Phi(S_t)] = \Theta(K\ell^2)$. However, in stationarity $\mathbf{E}[\Phi(S)] = O(\ell)$ (indeed it is 0) and $\text{Var}[\Phi(S)] = \Theta(\ell^2)$. Thus for any given ε we can take K large enough so that, when we start the exclusion process with all the particles on one side and run it for $\log(\ell/K(\varepsilon))/\gamma$ steps, with probability $1 - \varepsilon$ we are able to distinguish the configuration from a state drawn from stationarity. This gives us the desired mixing time lower bound of

$$\frac{\log \ell}{\gamma} = \left(\frac{2}{\pi^2} + o(1) \right) \ell^4 \log \ell.$$

10. Heuristic arguments for the true constants. Up until now we have given upper bounds and lower bounds for various mixing times and coupling times, and these bounds have typically differed by small constant factors. In this section we give heuristic arguments and summarize experimental results for determining the true asymptotic constant factors that were given in Table 1. Readers concerned primarily with rigorous arguments will find in this section a few theorems and many open problems.

10.1. *A million shuffles or seven.* It is well known that for any Markov chain, when one considers the distance from stationarity of the distribution at time t , the variation distance decays as $d(t) = (1 + o(1))A_d|\lambda|^t$ and, similarly, the separation distance decays as $s(t) = (1 + o(1))A_s|\lambda|^t$, where λ is the second largest eigenvalue (in absolute value). For the Markov chains considered here, we have rigorous exact values for λ . To paraphrase Diaconis (1996), the goal of finding mixing times is not to determine precisely how far from stationarity a deck of cards is after a million shuffles, but to determine if seven shuffles are enough. For many Markov chains, the variation distance from uniformity stays close to 1 for a time, and then rapidly becomes small and decays exponentially fast [see Diaconis (1988)]. The seven-shuffle question, which is more relevant to practical applications, asks where this cutoff occurs. The million-shuffle question has the virtue of typically being easier to answer and it appears to be relevant to the seven-shuffle question. Diaconis (1996) himself explained that the long-term behavior of the Markov chain can be used as a heuristic for predicting which Markov chains exhibit the “cutoff phenomenon” in the time it takes to randomize. Specifically for reversible Markov chains, he uses the L_2 bound

$$4\|P_x^t - \pi\|^2 \leq \sum_{i=1}^{N-1} v_i(x)^2 \lambda_i^{2t},$$

where the v_i 's are an orthogonal eigenbasis, and uses the lead term for large t ,

$$\sum_{i: |\lambda_i|=\lambda} v_i(x)^2 \lambda^{2t} = [A_{L_2} \lambda^t]^2,$$

to make the prediction: If A_{L_2} is large, then the Markov chain probability exhibits a sharp transition. This extends an earlier heuristic that Aldous and Diaconis [(1987), Section 7] gave for predicting the order of magnitude of the mixing time cutoff for random walk on groups.

In this section we work more directly with the separation and variation distances rather than use the L_2 norm, and we hypothesize that

$$d(t) \approx \min(1, A_d|\lambda|^t) \quad \text{and} \quad s(t) \approx \min(1, A_s|\lambda|^t).$$

A priori it is not clear why the variation distance should be well approximated by $A_d|\lambda|^t$ whenever this approximation is not obviously bad (i.e., when it is not larger than 1). Indeed, there are examples where this type of approximation fails: for walks on random Cayley graphs on \mathbb{Z}_2^d , the variation distance and the L_2 norm bound have essentially the same lead term behavior, but exhibit sharp transitions at different times [Wilson (1997b)]. Nonetheless the above approximation is valid for many Markov chains [see, e.g., Diaconis (1988)] and numerical computations described in Section 10.7 give every indication that this approximation holds for the Markov chains in which we are interested. We therefore take a moment to formalize this observation:

DEFINITION 1. A family of Markov chains exhibits a *clean (variation) cutoff* if for every ε there is a K so that for any Markov chain in the family and for any time t , whenever $|\log(A_d \lambda^t)| > K$,

$$|\log d(t) - \min(0, \log(A_d \lambda^t))| < \varepsilon.$$

A *clean separation cutoff* is defined similarly. If a family of Markov chains exhibits a clean cutoff and $A \rightarrow \infty$, then it necessarily exhibits a sharp cutoff (mixing time threshold) at $\log A / \log(1/|\lambda|)$. Since we already have the second largest eigenvalue for several classes of Markov chains, our goal in this section is to compute A_s and A_d , and report on experiments that suggest rather strongly that the Markov chains we are considering do in fact exhibit clean cutoffs.

10.2. *Preliminaries.* To obtain our conjectured values for the true constant factors in the mixing times of the adjacent transposition Markov chain on permutations and on lattice paths, and the Luby–Randall–Sinclair chain on lozenge tilings of a hexagon, we compute A_s and approximate A_d for these Markov chains. Before working on these specific chains, for the reader’s convenience we start with some basic preliminaries that are common to all these examples.

For the Markov chains considered here, we have a “potential function” Φ [defined in (1), (8) and (15)] such that $\mathbf{E}[\Phi(X_{t+1})|X_t] = \lambda \Phi(X_t)$. If we view Φ as the vector which is $\Phi(s)$ in its s th coordinate (where s is a state of the chain), then Φ is an eigenvector of the state-transition matrix with eigenvalue λ . Since the Markov chains are monotone and Φ is monotone increasing with respect to this partial order, we know that λ is the second largest eigenvalue in absolute value. We do not a priori know the multiplicity of the eigenvalue λ .

Since the Markov chains considered here are reversible, their state-transition matrices are diagonalizable and there is an orthogonal eigenbasis v_i with eigenvalues λ_i . If the Markov chain is started in state s , the distribution at time t is $\sum_i \alpha_i \lambda_i^t v_i$, where $\alpha_i = v_i(s)/(v_i \cdot v_i)$.

To determine the quantity $d(t)$, we need the worst starting state; for $\bar{d}(t)$, we need the worst pair of starting states; and for the separation distance $s(t)$, we need the worst start and destination states. It seems intuitive that for all three measures the worst states must be the top state $\hat{1}$ and bottom state $\hat{0}$, although there are some monotone Markov chains for which this intuition is wrong [Hägström (2001)]. Let us consider first the Markov chain on the symmetric group. Since it is vertex-transitive, all starting states are equivalent, so $\hat{0}$ and $\hat{1}$ are worst starting states. Furthermore, if the chain starts in $\hat{0}$, then the worst destination state is $\hat{1}$. This fact was used in the monotone version of Fill’s algorithm [Fill (1998)]. For the other Markov chains, we can only argue that $\hat{0}$ and $\hat{1}$ are the right states to look at for sufficiently large time t , and then only under the (apparently correct) assumption that the second largest eigenvalue λ has multiplicity 1. While not completely satisfying, this will allow us to compute, for example, A_s under this

one assumption. Assuming that λ has multiplicity 1, from the above formula we see that when t is large enough, the worst starting states are the states that maximize $|\Phi(x)|$ (i.e., states $\hat{0}$ and $\hat{1}$), the worst pair of starting states are those that maximize $|\Phi(x) - \Phi(y)|$ (i.e., $\hat{0}$ and $\hat{1}$), and the worst start and destination states are those that maximize $-\Phi(x)\Phi(y)$ (i.e., $\hat{0}$ and $\hat{1}$).

Let us suppose that the eigenvalue λ has multiplicity 1 (the multiplicity is larger for the permutation chain, but we deal with that later). Recall that P_x^t denotes the distribution at time t when the chain starts in state x and that U denotes the uniform (stationary) distribution. For large times t , $P_{\hat{1}}^t - U \doteq \frac{\Phi(\hat{1})}{\Phi \cdot \Phi} \lambda^t \Phi$. If we view Φ as the random variable obtained by picking a uniformly random state X and returning $\Phi(X)$, then $\Phi \cdot \Phi = \text{Var}[\Phi]N$, where N is the number of states. The vector Φ takes on its most negative value at $\hat{0}$, so it contributes $-\Phi(\hat{0})N$ to the separation distance. Hence for large times t we expect the separation distance to be $-\left(\Phi(\hat{1})\Phi(\hat{0})\right)/\text{Var}[\Phi]\lambda^t$, so that

$$(12) \quad A_s \stackrel{?}{=} -\frac{\Phi(\hat{1})\Phi(\hat{0})}{\text{Var}[\Phi]} = \frac{\Phi(\hat{1})^2}{\text{Var}[\Phi]},$$

where the question mark above the equal sign reminds us that in its derivation we used the assumption that the second largest eigenvalue has multiplicity 1.

The contribution of Φ to the variation distance is $\frac{1}{2}N\mathbf{E}[|\Phi|]$. Typically it is hard to get an analytic expression for $\mathbf{E}[|\Phi|]$, but heuristically it is plausible that the distribution of Φ is Gaussian, so that $\mathbf{E}[|\Phi|] \approx \frac{2}{\sqrt{2\pi}} \int_0^\infty x e^{-x^2/2} dx \sqrt{\text{Var}[\Phi]} = \sqrt{2/\pi} \sqrt{\text{Var}[\Phi]}$. There are interesting Markov chains, such as random transpositions on permutations that were analyzed by Diaconis and Shahshahani (1981), for which the principal eigenvector evaluated at a random state is very far from being approximated by a normal distribution. Thus, in principle, this approximation should be proved for the chains that we are considering, but we simply assert that it is intuitively obvious that this approximate normality holds for these chains. Assuming this approximate normality, we have

$$(13) \quad A_d \stackrel{?}{=} \frac{\Phi(\hat{1})\mathbf{E}[|\Phi|]/2}{\text{Var}[\Phi]} \stackrel{?}{\approx} \frac{\Phi(\hat{1})}{\sqrt{2\pi} \text{Var}[\Phi]} \stackrel{?}{=} \sqrt{\frac{A_s}{2\pi}}$$

(where we used the approximate normality assumption in the $\stackrel{?}{\approx}$ relationship and the multiplicity-1 assumption in the two $\stackrel{?}{=}$ relationships). This relationship between A_d and A_s is consistent with the folkwisdom that (for reversible chains) it usually takes twice as long for the separation distance to become small as it does for variation distance.

REMARK. A notable nonreversible chain where this relationship fails is the riffle-shuffle Markov chain [Bayer and Diaconis (1992)]. What failed is the

relationship $\alpha_i = v_i(s)/(v_i \cdot v_i)$, which assumes reversibility. With the correct α_i 's the above heuristic reasoning gives the right thresholds for the riffle–shuffle chain as well.

10.3. *Lozenge tilings.* For the Luby–Randall–Sinclair Markov chain on lozenge tilings of the order ℓ hexagon, we used $\mathbf{E}[(\Delta\Phi)^2] \leq O(\ell)$ to get a bound on the variance of the height function. While there do exist atypical configurations for which $\mathbf{E}[(\Delta\Phi)^2]$ is this large, it seems that more often $\mathbf{E}[(\Delta\Phi)^2]$ is closer to $O(1)$. If we substitute this bound on $\mathbf{E}[(\Delta\Phi)^2]$ into Lemma 5, we would get that the variance in the potential function is $O(\ell^4)$, with typical deviations from the mean being $O(\ell^2)$. These heuristic bounds are, in fact, correct in stationarity. It is known that the variance in the total height is $\ell^4/4$ and, more generally, $abc(a + b + c)/12$ for the hexagon with side lengths a, b, c, a, b, c [Blum (1996)]. Additionally, as we see from (14) below, in stationarity the variance in Φ is $abc(a + b + c)/(2\pi^2 + o(1))$. Thus for the order l hexagon in a stationarity, the potential function Φ is typically within $O(\ell^2)$ of its expected value 0. If (as seems likely) for each time t , $\Phi(X_t)$ is also typically within $O(\ell^2)$ of its expected value, then we would get a mixing time lower bound of $\log(\ell^3/\ell^2)/\log(1/|\lambda|) = (16/\pi^2 + o(1))\ell^4 \log \ell$. Intuitively the lozenge tiling is random when the distribution of its average height is close to its stationary distribution.

THEOREM 13. *Consider the Luby–Randall–Sinclair lozenge-tiling Markov chain on the hexagon with side lengths a, b, c, a, b, c , with the tower moves parallel to the c sides. Assuming the second largest eigenvalue has multiplicity 1,*

$$A_s = \frac{\Phi(\hat{1})^2}{\text{Var}[\Phi]} = \frac{c[(a + b)^2 - 1]}{ab(a + b + c)} \frac{\sin^2(\pi a/(a + b))}{1 - \cos(\pi/(a + b))}.$$

PROOF. The first relationship is just (12), which is where we use the multiplicity-1 assumption. To compute the variance of Φ in stationarity, we let S_i denote the total height in column i , $-a < i < b$, and write

$$\Phi = \sum_{i=-a}^b (S_i - \mathbf{E}[S_i]) \sin \frac{\pi(i + a)}{a + b},$$

$$\text{Var}[\Phi] = \sum_{-a < i, j < b} \text{Cov}(S_i, S_j) \sin \frac{\pi(i + a)}{a + b} \sin \frac{\pi(j + a)}{a + b}$$

and use a formula [Wilson (2001)] for the covariances of the S_i 's,

$$\text{Cov}(S_i, S_j) = (a + i)(b - j) \frac{abc(a + b + c)}{(a + b)^2((a + b)^2 - 1)}, \quad i \leq j,$$

to find (with the help of Maple) that

$$\begin{aligned}
 \text{Var}[\Phi] &= \frac{1/4}{1 - \cos(\pi/(a+b))} \frac{abc(a+b+c)}{(a+b)^2 - 1} \\
 (14) \qquad &= (1 + o(1)) \frac{abc(a+b+c)}{2\pi^2}.
 \end{aligned}$$

To evaluate $\Phi(\hat{1})$, we write

$$\Phi(\hat{1}) = \sum_{i=-a}^0 \frac{(a+i)bc}{a+b} \sin\left(\frac{\pi(a+i)}{a+b}\right) + \sum_{i=1}^b \frac{(b-i)ac}{a+b} \sin\left(\frac{\pi(a+i)}{a+b}\right),$$

which with the help of Maple simplifies to

$$\Phi(\hat{1}) = \frac{c}{2} \frac{\sin(\pi a/(a+b))}{1 - \cos(\pi/(a+b))}. \qquad \square$$

When $a = b = c = \ell$, Theorem 13 gives $A_s \stackrel{?}{\approx} 32/(3\pi^2)\ell^2$ and $A_d \stackrel{?}{\approx} \sqrt{16/(3\pi^3)}\ell$. Since $\gamma \doteq \pi^2/16/\ell^4$, we estimate the separation threshold to be $(16/\pi^2)\ell^4 \log(32/(3\pi^2)\ell^2) \doteq (32/\pi^2)\ell^4 \log \ell$ and the variation threshold to be $(16/\pi^2)\ell^4 \log \ell$, which matches the intuitive lower bound given above.

As a check of the $\mathbf{E}[|\Phi|] \approx \sqrt{2/\pi} \sqrt{\text{Var}[\Phi]}$ approximation, for the $3 \times 3 \times 3$ cube $\mathbf{E}[|\Phi|] \doteq 2.872$, while $\sqrt{2/\pi} \sqrt{\text{Var}[\Phi]} \doteq 2.892$, an error of about 1%. Even for the $2 \times 2 \times 2$ cube, the error between $\mathbf{E}[|\Phi|] \doteq 1.307$ and $\sqrt{2/\pi} \sqrt{\text{Var}[\Phi]} \doteq 1.319$ is less than 1%.

10.4. *Lattice paths.* We have effectively already computed A_s and approximated A_d for lattice paths in the $a \times b$ box—just set $c = 1$ in the above formulas for the $a \times b \times c$ lozenge-tiling region.

For lattice paths we can give some additional intuition. Consider the lattice path Markov chain on an $n/2 \times n/2$ box. In stationarity, the height fluctuation at a given site near the center of the path is $\Theta(\sqrt{n})$ and the fluctuations in the potential function is $\Theta(n^{3/2})$. Initially the potential function is $\Theta(n^2)$ and at any given time, the fluctuations in the potential function are about $O(n^{3/2})$ about its expected value. We used these facts to obtain the lower bound on the variation mixing time of $\log(n^2/n^{3/2})/\log(1/|\lambda|)$. Intuitively the path is about random when its average height is close to its stationary distribution, which would imply that the above lower bound is tight.

10.5. *Permutations.*

THEOREM 14. *For the random adjacent transposition Markov chain on permutations of order n , assuming the second largest eigenvalue has multiplicity $n - 1$, $A_s = n - 1$.*

REMARK. It is not clear to what extent it is a coincidence that A_s is the multiplicity of λ . This relationship does not hold for the tiling or lattice path Markov chains, but there the state spaces do not have a group structure. For \mathbb{Z}_2^n , $A_s = n$ is the multiplicity of the second largest eigenvalue. However, for the cycle \mathbb{Z}_n , the second largest eigenvalue has multiplicity 2 for $n \geq 3$, while $A_s = 2$ only for even $n > 3$ and $A_s = 2 \cos(\pi/n)$ for odd $n \geq 3$.

PROOF OF THEOREM 14. For random adjacent transpositions on permutations, for any given card i one can define an eigenvector with eigenvalue λ based on the location $\sigma^{-1}(i)$ of that card:

$$f_i(\sigma) = \cos[\pi(\sigma^{-1}(i) - 1/2)/n].$$

There is one linear dependency among these eigenvectors ($\sum_i f_i = 0$) and it appears that there are no other eigenvectors with eigenvalue λ . Note that

$$\begin{aligned} f_i \cdot f_i &= (n - 1)! \sum_{j=1}^n \cos^2\left(\frac{\pi(j - 1/2)}{n}\right) \\ &= (n - 1)! \sum_{j=1}^n \frac{1 + \cos(2\pi(j - 1/2)/n)}{2} = \frac{n!}{2}. \end{aligned}$$

By symmetry considerations, $f_i \cdot f_j = f_i \cdot f_k$ when $k \neq i \neq j$. Since

$$0 = f_i \cdot \left(\sum_j f_j\right) = \frac{n!}{2} + (n - 1)f_1 \cdot f_2,$$

we have that

$$f_i \cdot f_j = -\frac{n!}{2(n - 1)}$$

when $i \neq j$.

As before, to determine A_s we compute the coefficients of the f_i 's in the eigenbasis decomposition of $1_{\hat{1}}$. Since there is a linear relationship among the f_i 's, there will be a one-parameter family of valid sets of coefficients—we just need one such valid set of coefficients. We could be methodical and use the Gram–Schmidt procedure to extract $n - 1$ orthogonal vectors from the f_i 's and then use these to get a valid set of coefficients, but the guess-and-verify method is less messy. Consider the function

$$(15) \quad \Phi = \sum_{i=1}^n f_i(\hat{1}) f_i.$$

We have

$$\begin{aligned} f_i \cdot \Phi &= \frac{n!}{2} \left(f_i(\hat{1}) - \frac{1}{n-1} \sum_{j \neq i} f_j(\hat{1}) \right) \\ &= \frac{n!}{2} \frac{n}{n-1} f_i(\hat{1}). \end{aligned}$$

By comparison

$$f_i \cdot 1_{\hat{1}} = f_i(\hat{1}).$$

Since the dot products with the f_i 's are the same [up to the constant factor $n!n/(2(n-1))$], by linear algebra we conclude that $2(n-1)/(n!n)\Phi$ has the desired coefficients. Next we evaluate this eigenfunction at $\hat{0}$ and multiply by $-N = -n!$ to obtain

$$\begin{aligned} A_s &= -n! \frac{2(n-1)}{n!n} \sum_{i=1}^n f_i(\hat{1}) f_i(\hat{0}) \\ &= -2 \frac{n-1}{n} \sum_{i=1}^n -\cos^2\left(\frac{\pi(i-1/2)}{n}\right) \\ &= n-1. \quad \square \end{aligned}$$

10.6. *Shape of the thresholds.* In Figures 6–8, where we present numerical data for separation and variations distances, we also plot some hypothetical asymptotic curves for the separation and variation distances, in particular,

$$\begin{aligned} s(t) &\doteq 1 - \exp(-A_s \lambda^t), \\ d(t) &\doteq \operatorname{erf}\left(\frac{\sqrt{\pi}}{2} A_d \lambda^t\right), \\ \bar{d}(t) &\doteq \operatorname{erf}(\sqrt{\pi} A_d \lambda^t), \end{aligned}$$

where $\operatorname{erf}(x) = \int_{-x}^x e^{-t^2} dt / \sqrt{\pi}$ is the error function. For random walk on \mathbb{Z}_2^d , the variation distance $d(t)$ was shown to take the above form by Diaconis, Graham and Morrison (1990). The intuition for why it should also hold for the Markov chains that we are interested in is essentially the same as for their proof for \mathbb{Z}_2^d . When X is a random state drawn from the uniform distribution, $\Phi(X)$ is well approximated by a Gaussian. (For \mathbb{Z}_2^d , Φ is the number of 1's.) When t is near the mixing time threshold, $\Phi(X_t)$ should be well approximated by a Gaussian with the same variance as $\Phi(X)$ but with mean $\Phi(X_0)\lambda^t$. The intuition, which was made rigorous for \mathbb{Z}_2^d but appears difficult to prove for the other chains, is that Φ is the

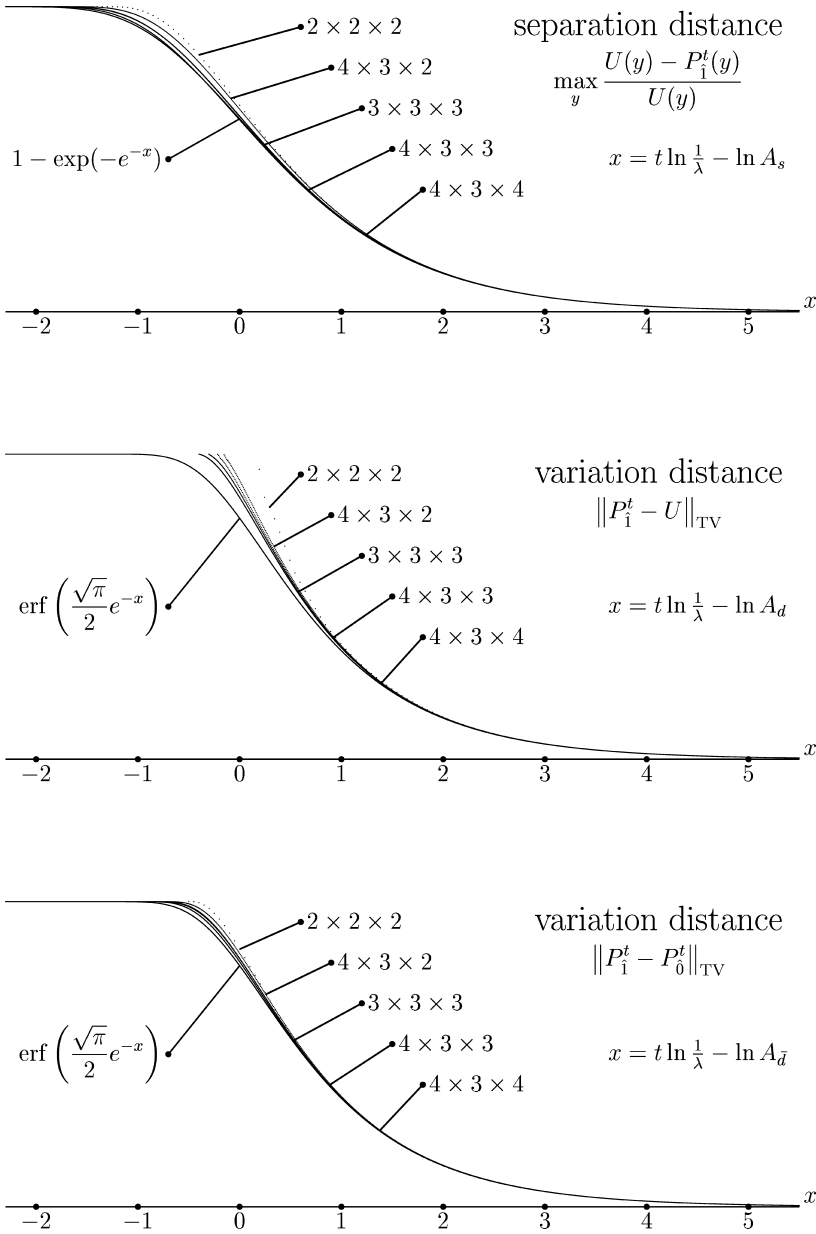


FIG. 6. Data for the Luby–Randall–Sinclair lozenge-tiling chain on the hexagon with side lengths a, b, c, a, b, c .

best test statistic for distinguishing X from X_t . The amount by which these two Gaussians fail to overlap gives the asymptotic curve for $d(t)$. The curve for $\bar{d}(t)$ also follows from this heuristic reasoning.

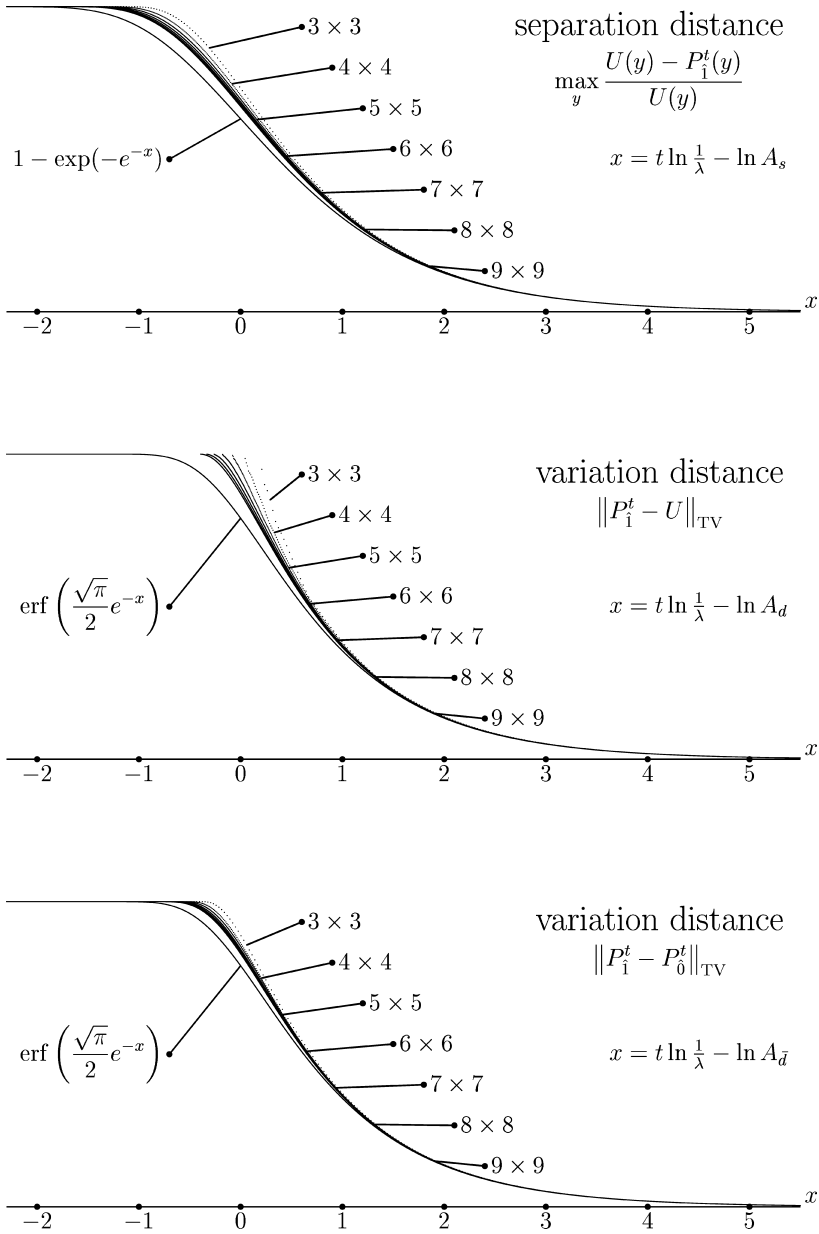


FIG. 7. Data for random adjacent transpositions on $a \times b$ lattice paths.

There does not seem to be any similar intuition for why the asymptotic curve for $s(t)$ should be that which is given above, other than that it holds for \mathbb{Z}_2^d and other high-dimensional product graphs, and it appears to be a good fit, at least for the tiling Markov chains.

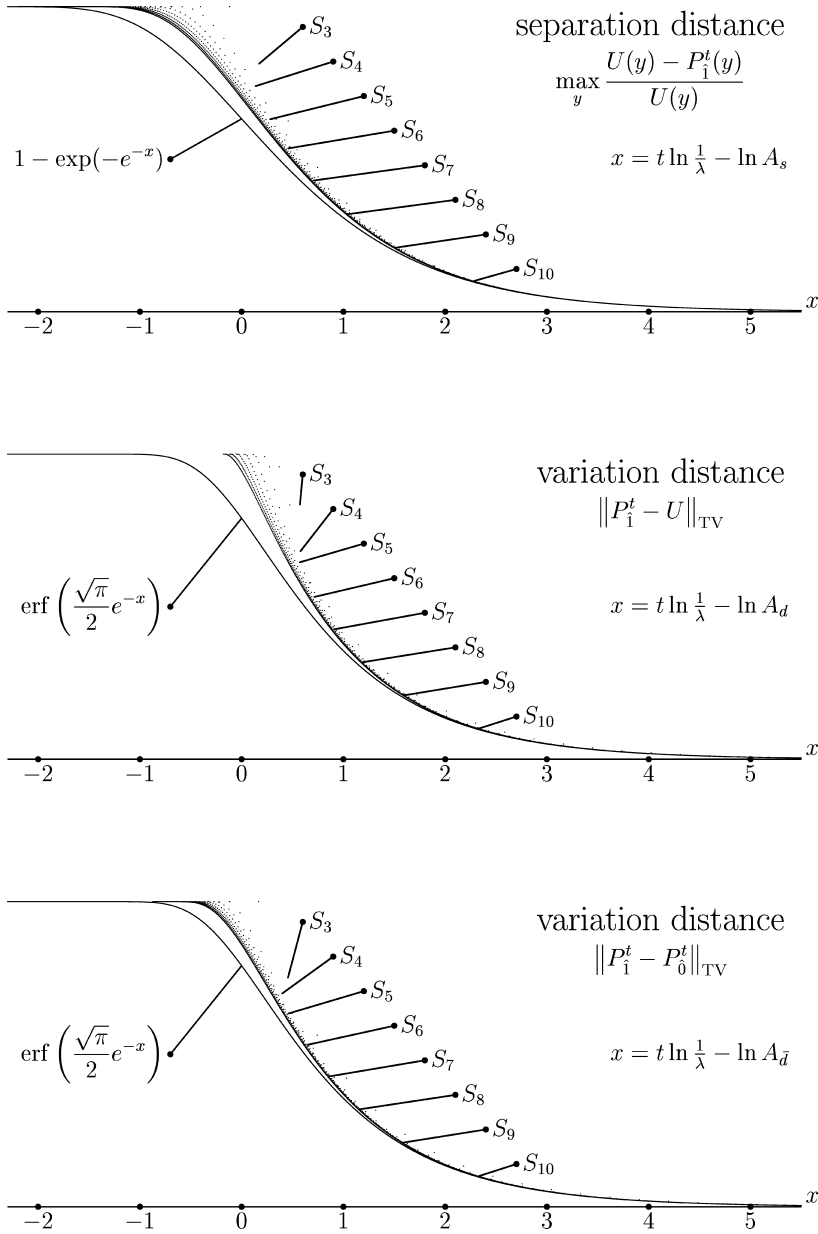


FIG. 8. Data for random adjacent transpositions on the symmetric group S_n .

REMARK. There exist Markov chains for which the asymptotic curves for $d(t)$ are not given by the above formula. For example, Diaconis, Fill and Pitman (1992) analyzed the top-to-random shuffle and determined that the curve for $d(t)$ is given by an explicit piecewise-analytic formula, which in particular is different from the

above formula. Nonetheless, the chains we are interested in seem to have more in common with random walk on \mathbb{Z}_2^d than they do with the top-to-random shuffle, and we are confident that the above formulas for $d(t)$ and $\bar{d}(t)$ are the correct asymptotic shape of the transition.

10.7. *Numerical experiments.* Figures 6–8 show numerical data for the convergence rates of the three classes of Markov chains considered here. The data were obtained by explicit multiplications of the state-transition matrix, and assume that $\hat{0}$ and $\hat{1}$ are the worst starting states for $d(t)$, worst pair of starting states for $\bar{d}(t)$, and worst start and destination states for $s(t)$. For each system size the convergence data were scaled and shifted to make the transitions line up. The amount by which to scale and shift was computed a priori using the preceding formulas for λ , A_s and A_d ; in particular, the values for λ and A_s are exact and the value for A_d was approximated using (13).

The curves line up quite nicely when $d(t)$, $\bar{d}(t)$ or $s(t)$ are small, and they line up progressively better with increasing system size. This fact and additional graphs not shown here indicate that the Markov chains have a clean cutoff and that, hence, the cutoff phenomenon does indeed occur where we expect it. The data look less good when $d(t)$, $\bar{d}(t)$ or $s(t)$ are large; this distortion is due in part to the fact that for finite system sizes there is a small finite x_{\min} , while in the idealized limit $x_{\min} = -\infty$. For $s(t)$, x_{\min} is about twice as large as for $d(t)$, so this distortion is much less pronounced for $s(t)$. For $\bar{d}(t)$, x_{\min} is $\log 2$ more negative than for $d(t)$, which is large enough to make the curves line up noticeably better for $\bar{d}(t)$ than for $d(t)$.

10.8. *Monte Carlo experiments.* To estimate the coupling time for the three classes of Markov chains, one can actually run the Markov chain until the upper and lower configurations coalesce, repeat many times and compare the results for different system sizes. The obvious advantage of Monte Carlo over numerical experiments is that one can do much larger system sizes. However, even with the large system sizes, when comparing different system sizes it is still much better to rescale time by

$$\log \frac{1}{\lambda} = \log \frac{1}{1 - (1 - \cos(\pi/n))/(n - 1)}$$

rather than its asymptotic value $(\pi^2/2)/n^3$.

After rescaling time in this manner it becomes quite clear that the coupling time for $n/2 \times n/2$ lattice paths is about $\log n / \log 1/\lambda \doteq 2/\pi^2 n^3 \log n$, and the coupling time for permutations is $\log n^2 / \log 1/\lambda \doteq 4/\pi^2 n^3 \log n$. Of course this coupling time estimate for lattice paths is actually rigorous thanks to Theorems 11 and 12. Estimating the correct constant for the tiling Markov chain is, however, much more challenging and we have not succeeded in doing this.

Surprisingly one can use Fill's algorithm to do similar Monte Carlo experiments to measure the separation distance [Fill (1998), Section 9]; we did not do this since we already had the numerical data. We are unaware of any similar Monte Carlo method for measuring the variation distance.

11. Concluding remarks.

- Adding weights to distance functions can be useful when proving mixing time upper bounds. While the optimal weighting scheme will be related to an eigenvector of the state-transition matrix, there is no need to diagonalize the matrix, nor is it even necessary to exhibit a single eigenvector to produce an effective weighting scheme that yields good upper bounds. For example, one could use parabolic weights rather than the sinusoidal weights that we did, and still derive mixing time upper bounds that are only a constant factor worse than the ones we derived.
- There are a variety of Markov chains for which a mixing time cutoff phenomenon has been proved. In future work on sharp mixing time thresholds, it would be worthwhile to determine whether or not the Markov chains exhibit a clean cutoff as defined in Section 10.1.

Acknowledgments. I thank Persi Diaconis and David Aldous for useful discussions, and Jim Fill, Chris Henley and a referee for their detailed comments on an earlier draft. The research that led to this article was done while the author was at MIT, DIMACS at Rutgers, Stanford and Microsoft Research.

REFERENCES

- ALDOUS, D. (1983). Random walks on finite groups and rapidly mixing Markov chains. *Séminaire de Probabilités XVII. Lecture Notes in Mathematics* **986** 243–297. Springer, Berlin.
- ALDOUS, D. (1997). Personal communication.
- ALDOUS, D. and DIACONIS, P. (1987). Strong uniform times and finite random walks. *Adv. in Appl. Math.* **8** 69–97.
- ALDOUS, D. J. and FILL, J. A. (2004). Reversible Markov chains and random walks on graphs. Unpublished manuscript. Available at www.stat.berkeley.edu/~aldous/book.html.
- BAYER, D. and DIACONIS, P. (1992). Trailing the dovetail shuffle to its lair. *Ann. Appl. Probab.* **2** 294–313.
- BLUM, M. D. (1996). Unpublished notes.
- BRIGHTWELL, G. and WINKLER, P. (1991). Counting linear extensions. *Order* **8** 225–242.
- BUBLEY, R. and DYER, M. (1997). Path coupling: A technique for proving rapid mixing in Markov chains. In *Proceedings of the 38th Annual Symposium on Foundations of Computer Science* 223–231.
- BUBLEY, R. and DYER, M. (1998). Faster random generation of linear extensions. In *Proceedings of the Ninth Annual ACM–SIAM Symposium on Discrete Algorithms* 350–354.
- CANCRINI, N. and MARTINELLI, F. (2000). On the spectral gap of Kawasaki dynamics under a mixing condition revisited. *J. Math. Phys.* **41** 1391–1423.
- CHEN, M.-F. (1998). Trilogy of couplings and general formulas for lower bound of spectral gap. *Probability Towards 2000. Lecture Notes in Statist.* **128** 123–136. Springer, Berlin.

- CIUCU, M. and PROPP, J. (1996). Unpublished manuscript.
- COHN, H. (1995). Personal communication.
- COHN, H., KENYON, R. and PROPP, J. (2001). A variational principle for domino tilings. *J. Amer. Math. Soc.* **14** 297–346.
- COHN, H., LARSEN, M. and PROPP, J. (1998). The shape of a typical boxed plane partition. *New York J. Math.* **4** 137–165.
- DESTAINVILLE, N. (2002). Flip dynamics in octagonal rhombus tiling sets. *Phys. Rev. Lett.* **88** 30601.
- DIACONIS, P. (1988). *Group Representations in Probability and Statistics*. IMS, Hayward, CA.
- DIACONIS, P. (1996). The cutoff phenomenon in finite Markov chains. *Proc. Nat. Acad. Sci. U.S.A.* **93** 1659–1664.
- DIACONIS, P. (1997). Personal communication.
- DIACONIS, P., FILL, J. A. and PITMAN, J. (1992). Analysis of top to random shuffles. *Combin. Probab. Comput.* **1** 135–155.
- DIACONIS, P., GRAHAM, R. L. and MORRISON, J. A. (1990). Asymptotic analysis of a random walk on a hypercube with many dimensions. *Random Structures Algorithms* **1** 51–72.
- DIACONIS, P. and SALOFF-COSTE, L. (1993a). Comparison techniques for random walk on finite groups. *Ann. Probab.* **21** 2131–2156.
- DIACONIS, P. and SALOFF-COSTE, L. (1993b). Comparison theorems for reversible Markov chains. *Ann. Appl. Probab.* **3** 696–730.
- DIACONIS, P. and SALOFF-COSTE, L. (1996). Logarithmic Sobolev inequalities for finite Markov chains. *Ann. Appl. Probab.* **6** 695–750.
- DIACONIS, P. and SHAHSHAHANI, M. (1981). Generating a random permutation with random transpositions. *Z. Wahrsch. Verw. Gebiete* **57** 159–179.
- DIACONIS, P. and SHAHSHAHANI, M. (1987). Time to reach stationarity in the Bernoulli–Laplace diffusion model. *SIAM J. Math. Anal.* **18** 208–218.
- DYER, M. and FRIEZE, A. (1991). Computing the volume of convex bodies: A case where randomness provably helps. In *Proceedings of the 44th Symposium in Applied Mathematics* 123–169.
- DYER, M., FRIEZE, A. and KANNAN, R. (1991). A random polynomial-time algorithm for approximating the volume of convex bodies. *J. Assoc. Comput. Mach.* **38** 1–17.
- FELSNER, S. and WERNISCH, L. (1997). Markov chains for linear extensions, the two-dimensional case. In *Proceedings of the Eighth Annual ACM–SIAM Symposium on Discrete Algorithms* 239–247.
- FILL, J. A. (1998). An interruptible algorithm for perfect sampling via Markov chains. *Ann. Appl. Probab.* **8** 131–162.
- FISHER, M. E. (1961). Statistical mechanics of dimers on a plane lattice. *Phys. Rev.* **124** 1664–1672.
- HÄGGSTRÖM, O. (2001). Personal communication.
- HANDJANI, S. and JUNGREIS, D. (1996). Rate of convergence for shuffling cards by transpositions. *J. Theoret. Probab.* **9** 983–993.
- HENLEY, C. L. (1997). Relaxation time for a dimer covering with height representation. *J. Statist. Phys.* **89** 483–507.
- HESTER, J. H. and HIRSCHBERG, D. S. (1985). Self-organizing linear search. *Computing Surveys* **17** 295–311. Available at www.acm.org/pubs/contents/journals/surveys/.
- KARZANOV, A. and KHACHIYAN, L. (1991). On the conductance of order Markov chains. *Order* **8** 7–15.
- KASTELEYN, P. W. (1963). Dimer statistics and phase transitions. *J. Math. Phys.* **4** 287–293.
- KENYON, R. W., PROPP, J. G. and WILSON, D. B. (2000). Trees and matchings. *Electron. J. Combin.* **7** R25.

- LEE, T.-Y. and YAU, H.-T. (1998). Logarithmic Sobolev inequality for some models of random walks. *Ann. Probab.* **26** 1855–1873.
- LU, S. and YAU, H.-T. (1993). Spectral gap and logarithmic Sobolev inequality for Kawasaki and Glauber dynamics. *Comm. Math. Phys.* **156** 399–433.
- LUBY, M., RANDALL, D. and SINCLAIR, A. (1995). Markov chain algorithms for planar lattice structures. In *36th Annual Symposium on Foundations of Computer Science* 150–159. [Expanded version *SIAM J. Comput.* (2001) **31**.]
- MANNILA, H. and MEEK, C. (2000). Global partial orders from sequential data. In *Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 161–168.
- MATTHEWS, P. (1991). Generating a random linear extension of a partial order. *Ann. Probab.* **19** 1367–1392.
- NG, L. L. (1996). Heisenberg model, Bethe Ansatz, and random walks. Bachelor's thesis, Harvard Univ.
- PROPP, J. G. (1995–1997). Personal communications.
- PROPP, J. G. and WILSON, D. B. (1996). Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Structures Algorithms* **9** 223–252.
- RANDALL, D. (1998). Personal communication.
- RANDALL, D. and TETALI, P. (2000). Analyzing Glauber dynamics by comparison of Markov chains. *J. Math. Phys.* **41** 1598–1615.
- WILSON, D. B. (1997a). Determinant algorithms for random planar structures. In *Proceedings of the Eighth Annual ACM–SIAM Symposium on Discrete Algorithms* 258–267.
- WILSON, D. B. (1997b). Random random walks on \mathbb{Z}_2^d . *Probab. Theory Related Fields* **108** 441–457.
- WILSON, D. B. (2001). Diagonal sums of boxed plane partitions. *Electron. J. Combin.* **8** N1.

MICROSOFT RESEARCH
ONE MICROSOFT WAY
REDMOND, WASHINGTON 98052
USA
E-MAIL: dbwilson@microsoft.com
URL: <http://dbwilson.com>