The Annals of Applied Probability
2014, Vol. 24, No. 5, 1835–1849
DOI: 10.1214/13-AAP964

MIXING TIME OF THE CARD-CYCLIC-TO-RANDOM SHUFFLE

BY BEN MORRIS[1], WEIYANG NING AND YUVAL PERES

University of California, Davis, University of Washington and Microsoft Research

The card-cyclic-to-random shuffle on n cards is defined as follows: at time t remove the card with label t mod n and randomly reinsert it back into the deck. Pinsky [Probabilistic and combinatorial aspects of the card-cyclic-to-random shuffle (2011). Unpublished manuscript] introduced this shuffle and asked how many steps are needed to mix the deck. He showed n steps do not suffice. Here we show that the mixing time is on the order of $\Theta(n \log n)$.

1. Introduction. In many Markov chains, such as Glauber Dynamics for the Ising model, the state space is a set of configurations, and at each step a location is chosen and updated. An important general question about such chains is what happens when we move from the world of *random updates*, where at each step a location is chosen at random and updated, to *systematic scan*, when the updates are done in a more deterministic fashion; see, for example, [2]. On the one hand, systematic scan is "less random," so one might expect that the mixing time is larger. On the other hand, systematic scan can update n sites on n steps, whereas with random updates $n \log n$ steps are required by the coupon collector problem, so one might expect systematic scan to have a smaller mixing time.

This question has been investigated in the context of the random transpositions shuffle. In this shuffle, at each step a pair of cards is chosen uniformly at random and interchanged. In a classical result of Diaconis and Shahshahani [4], the mixing time of the random transposition shuffle is shown to be asymptotically $\frac{1}{2} n \log n$. Mironov [7], Mossel, Peres and Sinclair [8], and Saloff-Coste and Zuniga [10] analyzed the cyclic-to-random shuffle, which is a systematic scan version of the random transposition shuffle: at step t the card in position t mod n is interchanged with a randomly chosen card. They found that the mixing time for this chain is still on the order of $n \log n$.

In the present paper, we study a systematic scan version of the random-to-random insertion shuffle. In the random-to-random insertion shuffle, at each step a card is chosen uniformly at random and then inserted in a uniform random position. It was shown in [3, 12] and [11] that the mixing time of this shuffle is on the order of $n \log n$. Pinsky [9] introduced the following model, called the card-cyclic-to-random shuffle: at time t remove the card with the label t mod n and insert it in

Received August 2012; revised August 2013.

[1]Supported by NSF Grants DMS-10-07739 and CNS-1228828.

MSC2010 subject classifications. 60J10.

Key words and phrases. Markov chain, mixing time.

a uniform random position. (It is a unique feature of this shuffle that both the labels of the cards and their positions play a role.) It is not obvious that the mixing time is greater than n: after n steps the location of each card has been randomized, so one might expect the whole deck to be close to uniform at time n. However, Pinsky showed that the mixing time is indeed greater than n, since the total variation distance to stationarity at this time converges to 1 as n goes to infinity.

We show that in fact the mixing time is on the order of $n \log n$. To prove the lower bound we introduce the concept of a barrier between two parts of the deck that moves along with the cards as the shuffling is performed. Then we show that the trajectory of this barrier can be well-approximated by a deterministic function f satisfying

$$f(x) = \int_{x-1}^{x} f(s)\, ds \tag{1}$$

and we relate the mixing rate of the chain to the rate at which f converges to a constant. To prove the upper bound, we use the path coupling method of Bubley and Dyer [1].

REMARK. One can study another systematic scan version of the random-to-random shuffle where one cycles through the cards by position rather than label. Consider the shuffle where at time t, the card in position $t \bmod n$ is removed and inserted in a uniform random position. Call this the position-cyclic-to-random insertion shuffle. For this shuffle the coupon collector problem implies a lower bound for the mixing time of order $n \log n$ by the following argument (Ross Pinsky, personal communication): note that the time-reversal of this chain is the shuffle which at time t picks a uniform card and inserts it to location $t \bmod n$. Thus if we start with the cards in increasing order, then the cards that have never been chosen for re-insertion by time t form an increasing subsequence of the permutation at time t. Since the longest increasing subsequence of a uniform permutation is $\mathrm{O}(\sqrt{n})$ with high probability, the mixing time must be at least of order $n \log n$, since this is the number of steps required to ensure that the number of "uncollected coupons" in the coupon collector problem is $\mathrm{O}(\sqrt{n})$. A matching upper-bound of $\mathrm{O}(n \log n)$ follows from the work of Saloff-Coste and Zuniga; see [10], Theorem 4.8.

2. Statement of main results.

Let X_t be a Markov chain on a finite state space V that converges to the uniform distribution. For probability measures μ and ν on V, define the *total variation distance* $\|\mu - \nu\| = \frac{1}{2}\sum_{x \in V} |\mu(x) - \nu(x)|$, and define the ε-mixing time

$$t_{\mathrm{mix}}(\varepsilon) = \min\{t : \|\mathbb{P}_x(X_t = \cdot) - \mathcal{U}\| \le \varepsilon, \text{ for all } x \in V\},$$

where $\mathcal{U}$ denotes the uniform distribution on V.

Recall that in the card-cyclic-to-random shuffle, at time t we remove the card with label $t \bmod n$ and then reinsert it into a uniform random location.

Define a round to be n consecutive such shuffles. Note that the Markov chain that performs a round of the card-cyclic-to-random shuffle at each step is time-homogeneous with a doubly-stochastic transition matrix, irreducible and aperiodic, and hence converges to the uniform stationary distribution. It follows that the card-cyclic-to-random shuffle converges to uniform as well. Our main results show that the mixing time is on the order of $\log n$ rounds.

THEOREM 1. *There exists c_0 such that for any $c < c_0$ and $0 < \varepsilon < 1$, when n is sufficiently large, we have*

$$t_{\mathrm{mix}}(\varepsilon) \geq cn \log n.$$

Here $c_0 = \frac{1}{2+2a}$ where a is the smallest positive solution of equations $b = e^a \sin b$ and $a = e^a \cos b - 1$. Numerically $c_0 = 0.161875162\ldots$.

THEOREM 2. *For any $\varepsilon > 0$ and $n \geq 4$, we have*

$$t_{\mathrm{mix}}(\varepsilon) \leq C(n \log n - 2n \log \varepsilon),$$

where

$$C = \frac{1}{\log 2 - \log(e-1)} = 6.58664655\ldots.$$

REMARK. Theorems 1 and 2 together establish that the card-cyclic-to-random shuffle has a pre-cutoff in total variation distance. It is an interesting open problem to determine if cutoff occurs in this shuffle. For reference on cutoff phenomenon, see [6], Chapter 18.

Theorems 1 and 2 will be proved in Sections 3 and 4, respectively.

3. Lower bound.

3.1. *The barrier.* The key idea for the lower bound is to imagine a barrier between two parts of the deck, that moves along with the cards as the shuffling is performed. If a card is inserted into the gap that the barrier occupies, we use the convention that the card is inserted on the same side of the barrier as it was in the previous step. We illustrate this with the following example. Suppose there is a deck of 8 cards with a barrier between cards 3 and 5. In the next step, card 7 is inserted between cards 3 and 5.

$$\begin{array}{ccccccccc} 2 & 1 & 3 & | & 5 & 4 & 6 & 8 & 7 \\ 2 & 1 & 3 & | & 7 & 5 & 4 & 6 & 8 \end{array}$$

Let $\{\sigma_t\}_{t=0}^{\infty}$ be a card-cyclic-to-random shuffle. We think of $\sigma_t(i)$ as the position of card i at time t, where the positions range from 1 at the left to n at the right.

Define the position of the barrier as the position of the card immediately to its left, and throughout the present chapter, let B_t be the position of the barrier at time t. Use the convention that $B_t = 0$ if at time t the barrier is to the left of all cards. We will call the pair process (σ_t, B_t) the *auxiliary process*.

Note that at any time $t > n$, every card has been reinserted exactly once in the previous n steps. Furthermore, if a card is reinserted to the left of the barrier, then it stays there until it is reinserted again. Hence

$$B_t = \sum_{i=1}^{n} \mathbf{1} \text{ (the card moved at time } t-i \text{ is inserted to the left of barrier).} \tag{2}$$

Since the conditional probability that the card at time t is inserted to the left of the barrier, given B_t, is $\frac{1}{n} B_t$, taking expectations in (2) gives

$$\mathbb{E}(B_t) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}(B_{t-i}). \tag{3}$$

Define $g(t) = \mathbb{E}(\frac{1}{n} B_t)$. Then g satisfies the following *moving average* condition:

$$g(t) = \frac{1}{n} \sum_{i=1}^{n} g(t-i) \tag{4}$$

for $t > n$. We shall approximate $g(t)$ by $f(t/n)$, where $f : \mathbf{R} \to [0, 1]$ is a continuous function satisfying (1). Our first lemma gives an example of such a function, which is graphed in Figure 1.

LEMMA 3. *There exist $a > 0$ and $b > 2\pi$ such that $f(x) = \frac{1}{2} + \frac{1}{2} e^{-ax} \sin(bx)$ satisfies*

$$f'(x) = f(x) - f(x-1). \tag{5}$$

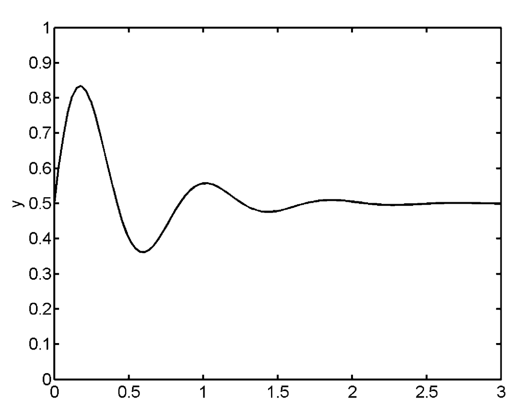

FIG. 1. *Graph of $f(s)$.*

Moreover,

$$f(x) = \int_{x-1}^{x} f(s)\,ds \tag{6}$$

for all x.

PROOF. Since properties (5) and (6) are preserved under shifting and scaling, it is enough to show that they apply to $h(x) = e^{-ax}\sin(bx)$, for suitable a and b.

First, we show that for suitable choice of a and b we have $h'(x) = h(x) - h(x-1)$. By the product rule,

$$h'(x) = -ae^{-ax}\sin(bx) + be^{-ax}\cos(bx) \tag{7}$$

and a calculation shows that

$$h(x) - h(x-1) = \bigl(1 - e^a\cos b\bigr)e^{-ax}\sin(bx) + \bigl(e^a\sin b\bigr)e^{-ax}\cos(bx). \tag{8}$$

The quantities (7) and (8) are equal if $b = e^a\sin b$ and $-a = 1 - e^a\cos b$. Solving for a in the first equation gives

$$a = \log\frac{b}{\sin b}$$

and substituting this into the second one gives

$$\log\frac{\sin b}{b} = 1 - \frac{b\cos b}{\sin b}.$$

By the intermediate value theorem, this equation has a solution with b in the interval $[2\pi + \frac{\pi}{4}, 2\pi + \frac{\pi}{2}]$, since when $b = 2\pi + \frac{\pi}{4}$ the right-hand side is smaller than the left-hand side, but when $b = 2\pi + \frac{\pi}{2}$ the right-hand side is larger. Furthermore, since $\sin b < b$ when $b > 0$, we have $a = \log\frac{b}{\sin b} > 0$. (Numerical approximation gives the solution as $b = 7.4615\ldots$ and $a = 2.0888\ldots$.)

Next we claim that since $h'(x) = h(x) - h(x-1)$, we must have $h(x) = \int_{x-1}^{x} h(s)\,ds$. To see this, define $\hat{q}(x) = \int_{x-1}^{x} h(s)\,ds$ and note that $\hat{q}'(x) = h'(x)$. This implies that $h(x) - \hat{q}(x) = C$ for a constant C. But since $a > 0$, we have $h(x) \to 0$ as $x \to \infty$. Consequently $\hat{q}(x) \to 0$ as $x \to \infty$ by the definition of $\hat{q}$, and so $C = 0$. $\square$

Recall that $g(t) = \mathbb{E}(\frac{1}{n}B_t)$, where B_t is the position of the barrier at time t. A key part of our proof will be to show that g closely follows the continuous function f of Lemma 3. However, in order for this to be the case we must start with a permutation chosen from a certain probability distribution. It is most convenient to describe this starting permutation as being generated in the first n time steps, which we call the *startup round*. In the startup round, we begin with only a barrier. At time t, for $1 \le t \le n$, we put card t to the left of the barrier with probability $f(\frac{t}{n})$. The location among the already existing cards in the left-hand (right-hand) side of the barrier is arbitrary. We must modify the definition of g to handle the startup round. Define $g : \{1, 2, \dots\} \to \mathbf{R}$ by

$$g(t) = \begin{cases} f\left(\dfrac{t}{n}\right), & \text{if } 1 \le t \le n; \\ \mathbb{E}\left(\dfrac{1}{n} B_t\right), & \text{otherwise.} \end{cases}$$

Thus g satisfies the moving average condition, and because of the insertion probabilities used in the startup round, g matches f for the first n steps [i.e., $g(\cdot) = f(\frac{\cdot}{n})$ on $\{1, \dots, n\}$]. As we show below, this is enough to ensure that g is well-approximated by f for a number of rounds on the order of $\log n$.

LEMMA 4. *There exists a constant $C > 0$ such that*

$$\left| g(t) - f\left(\frac{t}{n}\right) \right| \le \frac{C}{2n} e^{2(t+1)/n}$$

for all $t > 0$.

PROOF. First, note that if $t > n$, then

$$\begin{aligned} g(t+1) - g(t) &= \frac{1}{n} \sum_{i=1}^{n} g(t+1-i) - \frac{1}{n} \sum_{i=1}^{n} g(t-i) \\ &= \frac{1}{n} \bigl(g(t) - g(t-n) \bigr). \end{aligned}$$

Rearranging terms gives

$$g(t+1) = \left(1 + \frac{1}{n}\right) g(t) - \frac{1}{n} g(t-n). \tag{9}$$

Recall that $f(x) = \frac{1}{2} + \frac{1}{2} e^{-ax} \sin(bx)$ and $a > 0$. Some calculus shows that the second derivative of f is uniformly bounded on $[0, \infty)$. Hence

$$\begin{aligned} f\left(\frac{t+1}{n}\right) - f\left(\frac{t}{n}\right) &= \frac{1}{n} f'\left(\frac{t}{n}\right) + \mathrm{O}\left(\frac{1}{n^2}\right) \\ &= \frac{1}{n}\left(f\left(\frac{t}{n}\right) - f\left(\frac{t}{n} - 1\right)\right) + \mathrm{O}\left(\frac{1}{n^2}\right), \end{aligned}$$

where the first line follows from Taylor's theorem and the second line follows from Lemma 3. Rearranging terms gives

$$f\left(\frac{t+1}{n}\right) = \left(1 + \frac{1}{n}\right) f\left(\frac{t}{n}\right) - \frac{1}{n} f\left(\frac{t-n}{n}\right) + \mathrm{O}\left(\frac{1}{n^2}\right). \tag{10}$$

Combining (9) and (10) and using the triangle inequality gives

$$\left|g(t+1) - f\left(\frac{t+1}{n}\right)\right| \le \left(1+\frac{1}{n}\right)\left|g(t) - f\left(\frac{t}{n}\right)\right| + \frac{1}{n}\left|g(t-n) - f\left(\frac{t-n}{n}\right)\right| + \frac{C}{n^2} \tag{11}$$

for a universal constant C. We claim that for all t we have

$$\left|g(t) - f\left(\frac{t}{n}\right)\right| \le \frac{C}{n^2}\sum_{i=0}^{t}\left(1+\frac{2}{n}\right)^i. \tag{12}$$

We prove this by induction. For the base case, note that $g(t) = f(\frac{t}{n})$ for $t = 1, \dots, n$. Now if we suppose that (12) holds for $1, \dots, t$, then the two absolute values on the right-hand side of (11) can be bounded by $\frac{C}{n^2}\sum_{i=0}^{t}(1+\frac{2}{n})^i$. Hence

$$\begin{aligned}\left|g(t+1) - f\left(\frac{t+1}{n}\right)\right| &\le \left(1+\frac{2}{n}\right)\left[\frac{C}{n^2}\sum_{i=0}^{t}\left(1+\frac{2}{n}\right)^i\right] + \frac{C}{n^2}\\ &= \frac{C}{n^2}\sum_{i=0}^{t+1}\left(1+\frac{2}{n}\right)^i,\end{aligned}$$

which verifies (12) for $t+1$. To finish the proof of the lemma, note that

$$\frac{C}{n^2}\sum_{i=0}^{t}\left(1+\frac{2}{n}\right)^i = \frac{C}{n^2}\frac{(1+(2/n))^{t+1}-1}{2/n} \le \frac{C}{2n}e^{2(t+1)/n}.$$

□

3.2. *Deviation estimates*. In the previous subsection we proved that the expected barrier location is well-approximated by a continuous function. In the present subsection we show that the barrier stays reasonably close to its expectation with high probability when the number of rounds is on the order of $\log n$.

Define a *configuration* as a pair (σ, b), where σ is a permutation and b is a barrier location. (Thus the state space of the auxiliary process is the set of all configurations.) We define the *insertion distance* between two configurations as the minimum number of cards we would need to remove and re-insert to get from one configuration to the other. For example, the insertion distance between the two configurations below is 2. (Move cards 4 and 7.)

$$\begin{array}{cccccccccc} 2 & 1 & 4 & 3 & \mid & 5 & 6 & 8 & 7 \\ 2 & 1 & 3 & 7 & \mid & 5 & 4 & 6 & 8 \end{array}$$

LEMMA 5. *Let (σ_t^1, B_t^1) and (σ_t^2, B_t^2) be auxiliary processes, and define $\hat{\sigma}_t^i = (\sigma_t^i, B_t^i)$ for $i = 1, 2$. Let d be the insertion distance between $\hat{\sigma}_0^1$ and $\hat{\sigma}_0^2$. Then*

$$|\mathbb{E}B_t^1 - \mathbb{E}B_t^2| \le d\left(1+\frac{1}{n}\right)^t.$$

PROOF. There is a natural coupling of σ_t^1 and σ_t^2 that we call *label coupling*. In label coupling, at time t we choose a label X uniformly at random. If $X = t \bmod n$, then we move card $t \bmod n$ to the leftmost position in both processes. Otherwise, we insert card $t \bmod n$ to the right of the card with label X in both processes.

Suppose that $A = \{a_1, \dots, a_d\}$ is a minimal set of cards that can be moved to get from $\hat{\sigma}_0^1$ to $\hat{\sigma}_0^2$. Note that under the label coupling, only in the case when we move a card not in A can the insertion distance be increased. In such moves, if the card is put to the right of a card in A, the insertion distance increases by 1, and otherwise it stays the same. Thus the expected insertion distance after one step is at most

$$(d+1)\frac{d}{n} + d\frac{n-d}{n} = d\left(1 + \frac{1}{n}\right).$$

Iterating this argument shows that the expected insertion distance after t steps is at most $d(1+\frac{1}{n})^t$. The lemma follows from this, since the barrier can move by at most one position with each re-insertion. $\square$

We are now ready to state the main lemma of this subsection.

LEMMA 6. *Let (σ_t, B_t) be an auxiliary process. Fix $c > 0$ and suppose T satisfies $n < T \le cn\log n$. Then for any $x > 0$, we have*

$$\mathbb{P}\left(\left|\frac{1}{n}B_T - g(T)\right| \ge x\right) \le 2\exp(-x^2 n^{1-2c}).$$

PROOF. Fix T with $n < T \le cn\log n$. Since $g(T) = \frac{1}{n}\mathbb{E}(B_T)$, it is enough to show that for any $x > 0$ we have

$$\mathbb{P}(|B_T - \mathbb{E}(B_T)| \ge x) \le 2\exp(-x^2 n^{-(1+2c)}).$$

Let $\mathcal{F}_t$ be the sigma-field generated by the process up to time t, and consider the Doob martingale

$$M_t := \mathbb{E}(B_T|\mathcal{F}_t).$$

Applying Lemma 5 to the case of two configurations that differ by one insertion gives

$$|M_t - M_{t-1}| \le \left(1 + \frac{1}{n}\right)^{T-t}$$

for t with $1 \le t \le T$. Thus the Azuma–Hoeffding bound gives

$$\begin{aligned}\mathbb{P}(|B_T - \mathbb{E}(B_T)| \ge x) &= \mathbb{P}(|M_T - \mathbb{E}(M_T)| \ge x)\\ &\le 2\exp\left(\frac{-x^2}{2\sum_{t=1}^T b_t^2}\right),\end{aligned} \tag{13}$$

where $b_t = (1+\frac{1}{n})^{T-t}$. Let $r = (1+\frac{1}{n})^2$. The sum in (13) can be written as

$$\sum_{i=0}^{T-1} r^i = \frac{r^T - 1}{r-1} \le \frac{n}{2} r^T, \tag{14}$$

since $r - 1 = \frac{2}{n} + \frac{1}{n^2} \ge \frac{2}{n}$. Since $T < cn \log n$, quantity (14) is at most

$$\frac{n}{2}\left(1+\frac{1}{n}\right)^{2cn\log n} \le \frac{1}{2} n^{1+2c}.$$

Substituting this into (13) yields the lemma. □

3.3. *Proof of the lower bound*. Recall that $f(s) = \frac{1}{2} + \frac{1}{2} e^{-as} \sin(bs)$, for $a = 2.0888\ldots$ and $b = 7.4615\ldots$. The rough idea for the lower bound is as follows. Note that if c is sufficiently small and $s < c\log n$, then the fluctuation of $f(s)$ between s and $s+1$ is of higher order than $n^{-1/2}$. Thus in the corresponding round of the card-cyclic-to-random shuffle, there will be an interval of cards where the probability of inserting to the left of the barrier is detectably high. Before we give the proof, we recall Hoeffding's bounds in [5].

THEOREM 7. *Let $X_1, \ldots, X_k$ be samples from a population of* 0*s and* 1*s, and let $p = \mathbb{E}(X_1)$ be the proportion of* 1*s in the population. Then for $\alpha > 0$,*

$$\mathbb{P}\left(\sum_{i=1}^{k} X_i - kp \ge \alpha\right) \le e^{-2\alpha^2/k}. \tag{15}$$

Bound (15) *applies whether the sampling is done with or without replacement.*

PROOF OF THEOREM 1. Let $c > 0$ be small enough so that

$$c < \frac{1}{2+2a}. \tag{16}$$

Fix T with $n < T < cn\log n$ and let $x = T/n$. Suppose that $\sin(bx) \le 0$. The case $\sin(bx) > 0$ is similar. Since $b > 2\pi$, there exist x_1, x_2 with $x - 1 < x_1 < x_2 < x$, such that

$$bx_1 = 2\pi k + \pi/4 \quad \text{and} \quad bx_2 = 2\pi k + 3\pi/4$$

for an integer k. Note that for $s \in [x_1, x_2]$ we have

$$f(s) \ge \tfrac{1}{2} + \beta e^{-as} \ge \tfrac{1}{2} + \beta n^{-ac}, \tag{17}$$

where $\beta = \frac{1}{2}\sin(\pi/4)$. The second inequality holds because $x \le c\log n$.

Let A be the event that $\left|\frac{1}{n} B_t - f(t/n)\right| \le \frac{\beta}{4} n^{-ac}$ for all t with $T - n < t \le T$. Note that since $T < cn\log n$, substituting T into the upper bound of Lemma 4 implies that if $t \le T$, then $|g(t) - f(t/n)| < Bn^{2c-1}$, for a constant $B > 0$. Since $2c - 1 < -ac$ by (16), for sufficiently large n we have

$$Bn^{2c-1} < \frac{\beta}{8} n^{-ac}$$

and hence $|g(t) - f(t/n)| < \frac{\beta}{8} n^{-ac}$ for $t \le T$. Hence

$$\begin{aligned}
\mathbb{P}(A^c) &\le \mathbb{P}\biggl(\biggl|\frac{1}{n} B_t - g(t)\biggr| > \frac{\beta}{8} n^{-ac}, \text{ for some } t \text{ with } T - n < t \le T\biggr) \\
&\le 2n \exp\biggl(-\biggl[\frac{\beta}{8} n^{-ac}\biggr]^2 n^{1-2c}\biggr) \qquad (18)\\
&= 2n \exp\biggl(-\frac{\beta^2}{64} n^{1-2c(a+1)}\biggr),
\end{aligned}$$

where the second inequality follows from Lemma 6 and a union bound. Since $1 - 2c(a + 1) > 0$ by (16), quantity (18), and hence $\mathbb{P}(A^c)$, converges to 0 as $n \to \infty$.

Let $I = \{t \bmod n : nx_1 < t < nx_2\}$ and $m = |I|$. Since $x_2 - x_1 = \pi/2b$, there is a constant $\lambda > 0$ such that $m \ge \lambda n$ for sufficiently large n. Let S be the set of cards in I (i.e., cards whose label is in I) placed in one of the leftmost $\lceil (\frac{1}{2} + \frac{3\beta}{4} n^{-ac}) n \rceil$ positions between times nx_1 and nx_2. Note that the distribution of $|S|$ stochastically dominates the Binomial$(m, \frac{1}{2} + \frac{3\beta}{4} n^{-ac})$ distribution. Thus Hoeffding's bounds give

$$\begin{aligned}
\mathbb{P}\biggl(|S| < \frac{m}{2} + \frac{\beta}{2} mn^{-ac}\biggr) &\le \exp\biggl(\frac{-2((\beta/4)mn^{-ac})^2}{m}\biggr) \qquad (19)\\
&\le \exp\biggl(-\frac{\beta^2}{8} \lambda n^{1-2ac}\biggr),
\end{aligned}$$

where the second line follows from the fact that $m \ge \lambda n$. Since $1 - 2ac > 0$ by (16), the quantity (19) converges to 0 as $n \to \infty$.

By (17), on the event A the position of the barrier is greater that $(\frac{1}{2} + \frac{3\beta}{4} n^{-ac}) n$ between times nx_1 and nx_2. Furthermore, since $f(\frac{T}{n}) \le \frac{1}{2}$, on the event A we have $B_T \le (\frac{1}{2} + \frac{\beta}{4} n^{-ac}) n$. Combining these two facts shows that every card in S has position at most $(\frac{1}{2} + \frac{\beta}{4} n^{-ac}) n$ on the event A. (Here we are using the fact that every card inserted to the left of the barrier between times nx_1 and nx_2 must also be to the left of the barrier at time T.)

Now let Y be the number of cards in I having position at most $(\frac{1}{2} + \frac{\beta}{4} n^{-ac}) n$ at time T. Then

$$\mathbb{P}\biggl(Y \le \frac{m}{2} + \frac{\beta}{2} mn^{-ac}\biggr) \le \mathbb{P}\biggl(|S| \le \frac{m}{2} + \frac{\beta}{2} mn^{-ac}\biggr) + \mathbb{P}(A^c), \qquad (20)$$

which converges to 0 as $n \to \infty$ by (18) and (19).

To complete the proof, let Y_u be the number of cards in I whose position is at most $\frac{n}{2} + \frac{\beta}{4}n^{1-ac}$ in a uniform random permutation.

Hoeffding's bounds imply that

$$\begin{aligned} \mathbb{P}\left(Y_u > \frac{m}{2} + \frac{\beta}{2}mn^{-ac}\right) &\le \exp\left(\frac{-2((\beta/4)mn^{-ac})^2}{m}\right) \\ &\le \exp\left(-\frac{\beta^2}{8}\lambda n^{1-2ac}\right) \end{aligned} \tag{21}$$

for sufficiently large n. Since $1-2ac > 0$, quantity (21) converges to 0 as $n \to \infty$. Combining this with (20), we conclude that $t_{\mathrm{mix}}(\varepsilon) \ge cn\log n$ for large enough n. □

4. Upper bound. We use the path coupling technique introduced by Bubley and Dyer [1]. Let S_n be the permutation group and $G = (S_n, E)$, where an edge exists between two permutations if and only if they differ by an adjacent transposition. The path metric on G is defined by

$$\rho(x, y) = \min\{\text{length of } \eta : \eta \text{ is a path from } x \text{ to } y\}.$$

Define

$$\mathrm{diam}(G) = \sup_{x,y} \rho(x, y).$$

The following theorem is from [1]. See also [6], Chapter 14.

THEOREM 8. *Suppose that there exists $\alpha > 0$ such that for each edge $\{x, y\}$ in G there exists a coupling (X_1, Y_1) of the distributions $\mathbb{P}(x, \cdot)$ and $\mathbb{P}(y, \cdot)$ such that*

$$\mathbb{E}_{x,y}\rho(X_1, Y_1) \le \rho(x, y)e^{-\alpha}.$$

Then

$$t_{\mathrm{mix}}(\varepsilon) \le \frac{-\log\varepsilon + \log(\mathrm{diam}(G))}{\alpha}.$$

For a permutation x, define σ_t^x to be the card-cyclic-to-random shuffle starting at x. Our mixing time upper bound follows from the following lemma.

LEMMA 9. *If permutations x and y differ by an adjacent transposition and $n \ge 4$, there is a coupling of σ_n^x and σ_n^y such that*

$$\mathbb{E}\rho(\sigma_n^x, \sigma_n^y) \le e^{-\alpha},$$

where $\alpha = 2(\log 2 - \log(e-1))$.

PROOF. There is another natural coupling of two card-cyclic-to-random processes besides label coupling; we call this second coupling *position coupling*. In position coupling, the card is inserted into the same locations in both processes. Now assume that for some $i < j$, the permutation x can be obtained from y by transposing the cards with label i and j, as shown below. In the diagram, the kth X in the top row represents the same card as the kth X in the bottom row.

$$\begin{array}{lcccccccc} x: & X & X & X & i & j & X & X & X \\ y: & X & X & X & j & i & X & X & X \end{array}$$

The coupling strategy is divided into 3 stages, corresponding to t in $\{1, \dots, i-1\}$, $\{i, \dots, j-1\}$ and $\{j, \dots, n\}$, respectively.

Stage 1. Moving cards $1, \dots, i-1$. In this stage use position coupling. As is shown by Diagram 1 below, at the end of this stage we still have two permutations that differ only by a transposition of i and j. However, there may have been some cards inserted between cards i and j; we represent these cards with a's.

$$\begin{array}{lcccccccc} \sigma^x_{i-1}: & X & X & i & a & a & j & X & X \\ \sigma^y_{i-1}: & X & X & j & a & a & i & X & X \end{array}$$

Diagram 1

Stage 2. Moving cards $i, \dots, j-1$. In this stage we use label coupling. At the end of this stage, some cards might have been inserted into the group of a's. We denote such cards with α''s. In addition, some cards might have been inserted between card j and the first X to the right of the card j. We represent them with b's. Diagram 2 shows a typical pair of permutations after stage 2.

$$\begin{array}{lcccccccccccc} \sigma^x_{j-1}: & X & X & a & \alpha' & a & \alpha' & j & b & b & X & X \\ \sigma^y_{j-1}: & X & X & j & b & b & a & \alpha' & a & \alpha' & X & X \end{array}$$

Diagram 2

Stage 3. Moving cards $j, \dots, n$. Here we use label coupling again. Cards inserted into the group of a's and α''s are represented with a_{**}'s, and cards inserted into the group of b's are represented with β's. See Diagram 3 below. Notice that the a's, α''s and a_{**}'s maintain the same relative order in σ^x_n and as in σ^n, and similarly for the b's and β''s.

$$\begin{array}{lccccccccccccc} \sigma^x_n: & X & X & a & \alpha' & a_{**} & a & a_{**} & \alpha' & b & \beta' & b & X & X \\ \sigma^y_n: & X & X & b & \beta' & b & a & \alpha' & a_{**} & a & a_{**} & \alpha' & X & X \end{array}$$

Diagram 3

For $t \le n$, let A_t be the number of a's, α''s and a_{**}'s, and let B_t be the number of b's and β''s, after card t has been moved. Note that

$$\rho(\sigma^x_n, \sigma^y_n) \le A_n B_n.$$

Thus we are left to estimate $\mathbb{E}(A_n B_n)$.

Initially we have $A_0 = B_0 = 0$. Recall that in the first stage we use position coupling. For $t \le i-1$ we have $B_t = 0$, and A_t satisfies

$$\mathbb{P}(A_{t+1} = A_t | A_t) = \frac{n - A_t - 1}{n}$$

and

$$\mathbb{P}(A_{t+1} = A_t + 1 | A_t) = \frac{A_t + 1}{n}.$$

This implies

$$\mathbb{E}(A_{t+1} + 1 | A_t) = (A_t + 1)\left(1 + \frac{1}{n}\right). \tag{22}$$

Hence

$$\mathbb{E}A_{i-1} = \left(1 + \frac{1}{n}\right)^{i-1} - 1. \tag{23}$$

Recall that we use label coupling in the second stage. For $i \le t \le j-1$, we have the following transition rule:

$$\mathbb{P}(A_{t+1} = A_t, B_{t+1} = B_t | A_t, B_t) = \frac{n - A_t - B_t - 1}{n}$$

and

$$\mathbb{P}(A_{t+1} = A_t + 1, B_{t+1} = B_t | A_t, B_t) = \frac{A_t}{n}$$

and

$$\mathbb{P}(A_{t+1} = A_t, B_{t+1} = B_t + 1 | A_t, B_t) = \frac{B_t + 1}{n}.$$

This implies

$$\mathbb{E}\big(A_{t+1}(B_{t+1} + 1) | A_t, B_t\big) = A_t(B_t + 1)\left(1 + \frac{2}{n}\right).$$

Recall that $B_t = 0$ for all $t \le i-1$. Thus we have

$$\mathbb{E}A_{j-1}(B_{j-1} + 1) = \mathbb{E}A_{i-1}\left(1 + \frac{2}{n}\right)^{j-i}. \tag{24}$$

Note that for t with $i \le t < j$ we have

$$\mathbb{E}(A_{t+1} | A_t) = A_t\left(1 + \frac{1}{n}\right). \tag{25}$$

Thus $\mathbb{E}A_{j-1} = \mathbb{E}A_{i-1}(1 + \frac{1}{n})^{j-i}$. Combining this with (24) and (23) gives

$$\mathbb{E}A_{j-1}B_{j-1} = \left(\left(1 + \frac{1}{n}\right)^{i-1} - 1\right)\left(\left(1 + \frac{2}{n}\right)^{j-i} - \left(1 + \frac{1}{n}\right)^{j-i}\right). \tag{26}$$

For $j \le t \le n$ we have the following transition probabilities:

$$\mathbb{P}(A_{t+1} = A_t, B_{t+1} = B_t | A_t, B_t) = \frac{n - A_t - B_t}{n};$$

$$\mathbb{P}(A_{t+1} = A_t + 1, B_{t+1} = B_t | A_t, B_t) = \frac{A_t}{n};$$

$$\mathbb{P}(A_{t+1} = A_t, B_{t+1} = B_t + 1 | A_t, B_t) = \frac{B_t}{n}.$$

This implies

$$\mathbb{E}(A_{t+1} B_{t+1} | A_t, B_t) = A_t B_t \left(1 + \frac{2}{n}\right).$$

Using (26), we obtain

$$\mathbb{E} A_n B_n = \left(\left(1 + \frac{1}{n}\right)^{i-1} - 1\right)\left[\left(1 + \frac{2}{n}\right)^{j-i} - \left(1 + \frac{1}{n}\right)^{j-i}\right]\left(1 + \frac{2}{n}\right)^{n-j+1}.$$

Since $1 + \frac{2}{n} \le (1 + \frac{1}{n})^2$, the expression in square brackets is at most $(1 + \frac{1}{n})^{j-i}((1 + \frac{1}{n})^{i-j} - 1)$. Thus if we define β and γ so that $i = \beta n$ and $j = \gamma n$, calculation yields that

$$\mathbb{E} A_n B_n \le (e^{\beta} - 1) e^{\gamma - \beta} (e^{\gamma - \beta} - 1) e^{2(1-\gamma)},$$

if $0 \le \beta \le \log 2$, and

$$\mathbb{E} A_n B_n \le (e^{\beta} - 1) e^{\gamma - \beta} (e^{\gamma - \beta} - 1) e^{2(1-\gamma)} \left(1 + \frac{2}{n}\right),$$

if $\log 2 < \beta \le 1$. The former expression is maximized, for γ and β with $0 \le \beta \le \gamma \le 1$, by $(\frac{e-1}{2})^2$. The maximum occurs when $\gamma = 1$ and $\beta = \log \frac{2e}{e+1}$. Notice that $\log \frac{2e}{e+1} < \log 2$. Therefore, if $\alpha = 2(\log 2 - \log(e-1))$, then

$$\mathbb{E}(A_n B_n) \le e^{-\alpha}$$

for all $0 \le \beta \le \gamma \le 1$ and $n \ge 4$, which completes the proof. $\square$

PROOF OF THEOREM 2. We apply Theorem 8 to a round of the card-cyclic-to-random shuffle. Since the diameter of S_n with respect to adjacent transpositions is $\frac{n(n-1)}{2} < n^2$, substituting the α of Lemma 9 into Theorem 8 gives

$$t_{\text{mix}}(\varepsilon) \le \frac{1}{\log 2 - \log(e-1)} \left(\log n - \frac{1}{2} \log \varepsilon\right). \qquad \square$$

Acknowledgements. We thank Tonći Antunović, John Jiang, Ross Pinsky, Richard Pymar, Miklos Racz, Perla Sousi and Leonard Wong for many helpful comments.

REFERENCES

[1] BUBLEY, R. and DYER, M. (1997). Path coupling: A technique for proving rapid mixing in Markov Chains. In *Proceedings of the* 38*th Annual Symposium on Foundation of Computer Science* 223–231. IEEE Computer Society, Washington, DC.

[2] DIACONIS, P. and RAM, A. (2000). Analysis of systematic scan Metropolis algorithms using Iwahori–Hecke algebra techniques. *Michigan Math. J.* **48** 157–190. MR1786485

[3] DIACONIS, P. and SALOFF-COSTE, L. (1993). Comparison techniques for random walk on finite groups. *Ann. Probab.* **21** 2131–2156. MR1245303

[4] DIACONIS, P. and SHAHSHAHANI, M. (1981). Generating a random permutation with random transpositions. *Z. Wahrsch. Verw. Gebiete* **57** 159–179. MR0626813

[5] HOEFFDING, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–30. MR0144363

[6] LEVIN, D. A., PERES, Y. and WILMER, E. L. (2009). *Markov Chains and Mixing Times*. Amer. Math. Soc., Providence, RI. MR2466937

[7] MIRONOV, I. (2002). (Not so) random shuffles of RC4. In *Advances in Cryptology—CRYPTO* 2002. 304–319. Springer, Berlin. MR2054828

[8] MOSSEL, E., PERES, Y. and SINCLAIR, A. (2004). Shuffling by semi-random transpositions. In *Proceedings of the* 45*th Annual IEEE Symposium on Foundations of Computer Science* (*FOCS'*04) *October* 17–19, 2004, *Rome*, *Italy* 572–581. IEEE Computer Society, Washington, DC.

[9] PINSKY, R. (2013). Probabilistic and combinatorial aspects of the card-cyclic to insertion random shuffle. *Random Structures Algorithms*. DOI:10.1002/rsa.20505.

[10] SALOFF-COSTE, L. and ZÚÑIGA, J. (2007). Convergence of some time inhomogeneous Markov chains via spectral techniques. *Stochastic Process. Appl.* **117** 961–979. MR2340874

[11] SUBAG, E. (2013). A lower bound for the mixing time of the random-to-random insertions shuffle. *Electron. J. Probab.* **18** no. 20, 20. MR3035748

[12] UYEMURA-REYES, J. (2002). Random walk, semi-direct products, and card shuffling. Ph.D. thesis, Stanford Univ.

B. MORRIS
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF CALIFORNIA, DAVIS
ONE SHIELDS AVE
DAVIS, CALIFORNIA 95616
USA
E-MAIL: morris@math.ucdavis.edu

W. NING
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF WASHINGTON
SEATTLE, WASHINGTON 98195-4350
USA

Y. PERES
MICROSOFT RESEARCH
ONE MICROSOFT WAY
RREDMOND, WASHINGTON 98052
USA
E-MAIL: peres@microsoft.com