Shuffling by semi-random transpositions

Elchanan Mossel [*]
Statistics
U.C. Berkeley
mossel@stat.berkeley.edu

Yuval Peres [†]
Statistics and Mathematics
U.C. Berkeley
peres@stat.berkeley.edu

Alistair Sinclair [‡]
Computer Science
U.C. Berkeley
sinclair@cs.berkeley.edu

Abstract

In the cyclic-to-random shuffle, we are given n cards arranged in a circle. At step k, we exchange the k'th card along the circle with a uniformly chosen random card. The problem of determining the mixing time of the cyclic-to-random shuffle was raised by Aldous and Diaconis in 1986. Recently, Mironov used this shuffle as a model for the cryptographic system known as RC4, and proved an upper bound of $O(n \log n)$ for the mixing time. We prove a matching lower bound, thus establishing that the mixing time is indeed of order $\Theta(n \log n)$. We also prove an upper bound of $O(n \log n)$ for the mixing time of any "semi-random transposition shuffle", i.e., any shuffle in which a random card is exchanged with another card chosen according to an arbitrary (deterministic or random) rule. To prove our lower bound, we exhibit an explicit complex-valued test function which typically takes very different values for permutations arising from few iterations of the cyclic-to-random-shuffle and for uniform random permutations. Perhaps surprisingly, the proof hinges on the fact that the function $e^z - 1$ has nonzero fixed points in the complex plane. A key insight from our work is the importance of complex analysis tools for uncovering structure in nonreversible Markov chains.

1. Introduction

The *mixing time* of a Markov chain on a finite state space is the number of steps until it is close to its stationary distribution, starting from an arbitrary state. The mixing time is a key parameter in analyzing random sampling algorithms and is of intrinsic interest in probability and statistical physics. For many natural Markov chains, if some of the randomness is removed from the transition rule, resulting in a "more deterministic" process with the same stationary distribution, the chain becomes significantly harder to analyze. Indeed, some of the most challenging problems in the field concern the analysis of such "pseudo-random" variants of well understood chains. Some examples include the riffle shuffle [13, 18] compared to the Thorp shuffle [19], the asymmetric exclusion process [6] compared with its systematic scan version [10], and the comparison between the standard and systematic scan versions of Glauber dynamics for Gaussian fields [14, 4] and for spin systems [12].

Shuffling by random transpositions is one of the simplest random walks on the symmetric group: given n cards in a row, at each step two cards are picked uniformly at random and exchanged. This shuffle was precisely analyzed in 1981 [11]. In the "cyclic-to-random" shuffle (invented by Thorp [20]), at step t a uniformly chosen random card is exchanged with the card at position $t \mod n$. It is easy to see that this semi-random shuffle still converges to the uniform distribution on permutations of n cards. In their landmark 1986 paper on card shuffling [3], Aldous and Diaconis posed as a challenge the analysis of the cyclic-to-random shuffle. More recently, Mironov [16] related this shuffle to the behavior of the cryptographic system RC4 – see Section 4 for a brief discussion of this connection. Mironov showed that a strong uniform time argument due to Broder (as described in [9]) can be adapted to yield an upper bound of $O(n \log n)$ on the mixing time of the cyclic-to-random-shuffle. He posed as an open problem whether this bound is tight, and discussed the relevance of such an analysis of the cyclic-to-random shuffle to potential vulnerabilities in RC4.

In this paper we establish a lower bound of $\Omega(n \log n)$ for the mixing time of the cyclic-to-random shuffle, thus answering the questions posed by Aldous and Diaconis and by Mironov. We also prove a general upper bound of $O(n \log n)$ on the mixing time of *any* semi-random transposition shuffle, i.e., any shuffle in which a random card is exchanged with another card chosen according to an arbitrary (deterministic or random) rule that may vary at each

[*] Supported by a Miller Fellowship in CS and Statistics, U.C. Berkeley.

[†] Research supported in part by NSF Grants DMS-0104073 and DMS-0244479. Part of this work was done while the author was visiting Microsoft Research.

[‡] Supported in part by NSF grant CCR-0121555. Part of this work was done while the author was on sabbatical leave at Microsoft Research and at Ecole Polytechnique, Paris.

step. Previously, the best available upper bound for such a general process was $O(n^2)$, proved by Pak [17].

To prove the lower bound for the cyclic-to-random shuffle $\{\sigma_t\}$, we find an eigenfunction F of the shuffle that mixes slowly. First, we determine the eigenvalues of a nonreversible renewal Markov chain M on the n-cycle which describes the behavior of a single card. The asymptotics for the leading eigenvalues of M depend on the fact that the function $e^z - 1$ has nonzero fixed points in the complex plane. We then pick an eigenfunction f for M and use it to construct a test function F, defined on permutations, which is a weighted sum of f applied to the locations of all cards. To show that the distribution at time t of $F(\sigma_t)$ is far from the distribution of $F(\sigma)$ for a uniform random permutation σ, the key is to estimate the variance. (This approach was used by Wilson [21, 22] to prove $\Omega(n^3 \log n)$ lower bounds for the shuffle generated by transpositions of adjacent cards and several variants.) The variance is a sum of correlations between pairs of cards; to bound these correlations, we couple the shuffle with a system of independent particles evolving according to M. This coupling approach has intuitive appeal, and could potentially be used for other chains on permutations. Alternatively, one could bound the variance of $F(\sigma_t)$ using the martingale decomposition method of Wilson [21, 22].

Our general upper bound for semi-random transpositions is elementary and proved via a strong uniform time argument, extending earlier arguments of Broder and Mironov.

We believe that some of our technical insights may be carried over to other situations where lower bounds for nonreversible or “pseudo-random” Markov chains are sought. These insights include:

- The analysis of a given Markov chain with a transition rule that varies in time can sometimes be reduced to the analysis of an equivalent time-homogeneous chain.
- Coupling arguments, which are often applied to obtain upper bounds for mixing times, can also be used to establish lower bounds.
- When seeking to understand a nonreversible Markov chain, results of classical complex analysis (such as Rouché’s theorem) can be powerful tools. Thus methods from complex analysis should be added to techniques from probability, combinatorics, functional analysis and representation theory in the toolkit of Markov chain analysis.

1.1. Statement of main results

Let $\{L_t\}_{t=1}^{\infty}$ be a sequence of random variables taking values in $[n] = \{0, 1, \ldots, n-1\}$ and let $\{R_t\}_{t=1}^{\infty}$ be a sequence of i.i.d. cards chosen uniformly from $[n]$. The **semi-random transposition shuffle** generated by $\{L_t\}$ is a stochastic process $\{\sigma_t^*\}_{t=0}^{\infty}$ on the symmetric group S_n, defined as follows. Fix the initial permutation σ_0^*. The permutation σ_t^* at time t is obtained from σ_{t-1}^* by transposing the cards at locations L_t and R_t.

The stochastic process $\{\sigma_t^*\}$ is a time-inhomogeneous Markov chain on S_n, and converges to the uniform stationary distribution for any σ_0^* and any choice of $\{L_t\}$. It is a time-homogeneous Markov chain if the L_t are i.i.d. The special case where the L_t are i.i.d. uniform is the random transposition shuffle [2, 3, 11], the random walk on S_n generated by all transpositions; at the other extreme, if all the L_t are identically 0, we get the random walk generated by “star transpositions”, where at each step a randomly chosen card is exchanged with the card at position 0. In the **cyclic-to-random shuffle**, the sequence L_t is given by $L_t = t \mod n$.

Let μ_t^* be the distribution of σ_t^* at time t, and let $\|\mu_t^* - \mathcal{U}\|_{\mathrm{TV}}$ denote the total variation distance between μ_t^* and the uniform distribution $\mathcal{U}$. Define the *mixing time* by

$$\tau(\epsilon) = \max_{\sigma_0} \min\{t : \|\mu_t^* - \mathcal{U}\|_{\mathrm{TV}} \le \epsilon\}.$$

We let $\tau_{\mathrm{mix}} = \tau(\frac{1}{2e})$ and note that $\tau(\epsilon) \le \lceil \log \epsilon^{-1} \rceil \tau_{\mathrm{mix}}$ (see [2]). Therefore, proving an $\Omega(g(n))$ bound on $\tau(\epsilon)$ for fixed $\epsilon > 0$ implies an $\Omega(g(n))$ bound on τ_{mix}. With slight abuse of notation, we say that the mixing time is $\Omega(g(n))$ if there exist constants ϵ, C such that $\tau(\epsilon) \ge C g(n)$ for all sufficiently large n.

Our first result is a lower bound for the mixing time of the cyclic-to-random shuffle, matching (up to a constant factor) the upper bound of Mironov [16]:

Theorem 1.1 *The cyclic-to-random shuffle has mixing time $\Omega(n \log n)$. More precisely,*

$$\tau\left(\frac{\|\chi\|_2^4}{8\|\chi\|_\infty^4} - O\Big(\frac{1}{n}\Big)\right) \ge \frac{n \log n}{2|\Re\zeta + 1|(1 + o(1))}, \qquad (1)$$

where ζ is any nonzero complex root of the equation $\psi(z) = e^z - z - 1 = 0$ and $\chi : [0,1] \to \mathbb{C}$ is defined by

$$\chi(x) := 1 - \frac{\zeta + 1}{\zeta}(e^{\zeta x} - 1). \qquad (2)$$

Remark. Using Mathematica, we find the root $\zeta = 2.088... + 7.461... \times i$ of ψ. This gives $|1 + \zeta| = 8.075..., |\Re\zeta + 1| = 3.088...$ and yield the lower bounds $\tau(0.0095) \ge (.161 + o(1)) n \log n$ and $\tau_{\mathrm{mix}} \ge (0.0345 + o(1)) n \log n$.

We note that a $\Omega(n \log n)$ lower bound for the random transposition shuffle follows easily from a coupon collector argument: up to time $(n \log n)/4$, about $\sqrt{n}$ of the cards have not moved even once. However, such a simple argument cannot work for the cyclic-to-random shuffle as all cards are touched by time n.

Our second result is an upper bound for the mixing time of *any* semi-random transposition shuffle:

Theorem 1.2 *The semi-random transposition shuffle $\{\sigma_t^*\}$ generated by any sequence $\{L_t\}$ has mixing time at most $O(n \log n)$. More precisely, there is a constant C_0 such that, for any $C_1 > C_0$ and any initial configuration σ_0^*, we have*

$$\tau(n^{-\beta}) \leq C_1 n \log n.$$

for some $\beta = \beta(C_1) > 0$.

Remark. The proof shows that we can take $C_0 = 32\theta^{-3} + \theta^{-1}$ where $\theta = e^{-2}(1 - e^{-1})/2$. We do not know the minimal value of C_0; it cannot be strictly less than 1 because of the star transpositions shuffle, whose mixing time is $(1 + o(1))n \log n$ (see [8]).

2. A lower bound for the cyclic-to-random shuffle

2.1. The behavior of a single card via renewals

Fixing a specific card a, it is natural to study the renewal chain on the state space $[n] = \{0, \dots, n-1\}$, where state $i \in [n]$ indicates that the location j of card a satisfies $j + i = t \mod n$. This chain is time-homogeneous. It is obtained from the original chain by rotating all cards to the left after each transposition.

This chain is described by the transition matrix M, where for all $i \in [n]$ we have $M_{0,i} = 1/n$ and $M_{i,1} = 1/n$, while $M_{i,i+1} = 1 - 1/n$, for all $i \geq 1$. (For $i = n-1$, the last equation reads $M_{n-1,0} = 1 - 1/n$.) In other words,

$$M = \begin{pmatrix} \frac{1}{n} & \frac{1}{n} & \frac{1}{n} & \frac{1}{n} & \dots & \frac{1}{n} & \frac{1}{n} \\ 0 & \frac{1}{n} & 1-\frac{1}{n} & 0 & 0 & \dots & 0 \\ 0 & \vdots & 0 & \ddots & 0 & \dots & 0 \\ 0 & \vdots & 0 & 0 & \ddots & \dots & 0 \\ 0 & \vdots & 0 & \dots & 0 & \ddots & 0 \\ 0 & \frac{1}{n} & 0 & \dots & 0 & 0 & 1-\frac{1}{n} \\ 1-\frac{1}{n} & \frac{1}{n} & 0 & \dots & 0 & 0 & 0 \end{pmatrix}.$$

We will now find the eigenfunctions of the chain, that is, the right eigenvectors of the matrix M. Let $f = (f(0), \dots, f(n-1))^T$ be such a (column) eigenvector. Then we obtain the following equations:

$$\frac{1}{n}\sum_{j=0}^{n-1} f(j) = \lambda f(0), \tag{3}$$

and, for $1 \leq i \leq n-1$,

$$\frac{1}{n}f(1) + (1 - \frac{1}{n})f(i+1) = \lambda f(i) \tag{4}$$

(where we set $f(n) = f(0)$). It is easy to check that, up to scaling, $(1, \dots, 1)^T$ is the unique eigenvector corresponding to the eigenvalue $\lambda = 1$, and that $(-1, n-1, -1, \dots, -1)^T$ is the unique eigenvector corresponding to $\lambda = 0$.

We now assume that f is a right eigenvector corresponding to an eigenvalue $\lambda \notin \{0, 1\}$. Since M is doubly stochastic, (3) implies that $\sum_{i=0}^{n-1} f(i) = 0$ and $f(0) = 0$; to verify this, sum (3) and the $n-1$ equations in (4).

Writing $y_i = f(i+1) - f(i)$ for $1 \leq i \leq n-1$ (recall that $f(n) = f(0)$), the equation (4) for $i = 1$ gives

$$y_1 = \frac{n(\lambda - 1)}{n-1} f(1).$$

For $1 \leq i \leq n-2$, subtracting successive equations in (4) yields

$$(1 - \frac{1}{n})y_{i+1} = \lambda y_i.$$

Thus if we set $\gamma = \frac{n\lambda}{n-1}$, then $y_1 = (\gamma - \frac{n}{n-1})f(1)$ and $y_j = \gamma^{j-1} y_1$ for $2 \leq j \leq n-1$. Without loss of generality we may assume that $f(1) = 1$. Therefore,

$$\begin{aligned} f(k) &= 1 + \sum_{j=1}^{k-1} y_j = 1 + y_1 \sum_{j=1}^{k-1} \gamma^{j-1} \\ &= 1 + \left(\gamma - \frac{n}{n-1}\right) \sum_{j=1}^{k-1} \gamma^{j-1} \end{aligned} \tag{5}$$

for $1 \leq k \leq n$. Thus

$$(n-1)(1-\gamma)f(k) = \Big(n - (n-1)\gamma\Big)\gamma^{k-1} - 1$$

for $1 \leq k \leq n$. Since $\sum_{k=0}^{n-1} f(i) = 0$ and $f(n) = f(0)$, we infer that

$$\begin{aligned} 0 &= (n-1)(1-\gamma)^2 \sum_{k=1}^{n} f(k) \\ &= \Big(n - (n-1)\gamma\Big)(1 - \gamma^n) - n(1-\gamma) \\ &= \gamma - n\gamma^n + (n-1)\gamma^{n+1}\,. \end{aligned}$$

Since $\gamma \neq 0$ by assumption, it follows that

$$(n-1)\gamma^n - n\gamma^{n-1} + 1 = 0\,. \tag{6}$$

Note that this equation has a double root at $\gamma = 1$. We therefore conclude that the eigenvalues $\lambda \neq 0, 1$ correspond (via the relation $\gamma = \frac{n\lambda}{n-1}$) to the roots $\gamma \neq 1$ of (6). We investigate these roots next.

2.2. Properties of the roots of equation (6)

Lemma 2.1 *All the roots of equation (6) satisfy $|\gamma| \leq 1$.*

Proof: If $|\gamma| > 1$, then

$$|(n-1)\gamma^{n-1}| > \Big|\sum_{i=0}^{n-2}\gamma^i\Big| = \Big|\frac{\gamma^{n-1}-1}{\gamma-1}\Big|.$$

Multiplying by $|\gamma - 1|$ gives

$$|(n-1)\gamma^n - (n-1)\gamma^{n-1}| > |\gamma^{n-1} - 1|,$$

so γ cannot be a solution to (6). ■

In the other direction, we need to show that (6) has solutions close to 1. We prove:

Lemma 2.2 *There exists a solution of equation (6) which satisfies*

$$1 - \gamma = \frac{\zeta}{n} + O(\frac{1}{n^2}) \qquad (7)$$

$$1 - \lambda = \frac{\zeta + 1}{n} + O(\frac{1}{n^2}), \qquad (8)$$

and

$$1 - |\lambda| = \frac{\Re\zeta + 1}{n} + O(\frac{1}{n^2}), \qquad (9)$$

where ζ is any nonzero root of $e^\zeta - \zeta - 1 = 0$, and $\lambda = (1 - 1/n)\gamma$.

Proof: Defining $\omega = \gamma^{-1}$, we obtain from (6) the equation $\omega^n - n\omega + n - 1 = 0$, or $\omega^n + n(1-\omega) - 1 = 0$. Now write $\omega = 1 + z/n$ to get the asymptotic equation $\psi(z) \equiv e^z - z - 1 = 0$. By Hurwitz's theorem (see [1]), every solution ζ of the equation $\psi(\zeta) = 0$ is a limit of solutions z of the equations $(1 + z/n)^n - z - 1 = 0$. Since $\omega - 1 = z/n$, we obtain

$$\gamma = 1 - z/n + O(\frac{1}{n^2}). \qquad (10)$$

Therefore

$$\lambda = (1 - 1/n)\gamma = 1 - \frac{1+z}{n} + O(\frac{1}{n^2}). \qquad (11)$$

To get more precise estimates, recall that $\psi(z) = e^z - z - 1$ and let $\varphi_n(z) = (1 + z/n)^n - z - 1$. By a Taylor expansion,

$$|n\log(1 + z/n) - z| = \frac{|z|^2}{2n} + O(\frac{1}{n^2}),$$

so in a bounded domain,

$$|\varphi_n(z) - \psi(z)| = |(1 + z/n)^n - e^z| = \frac{|z^2e^z|}{2n} + O(\frac{1}{n^2}).$$

Below we will prove that the equation $e^z - z - 1 = 0$ has nonzero roots. Let ζ be such a root; then ζ is a simple root, since $\psi'(\zeta) = e^\zeta - 1 = \zeta$. Thus for z on the circle $\{|z - \zeta| = b/n\}$, we have

$$|\psi(z)| = |\psi'(\zeta)|\frac{b}{n} + O(\frac{1}{n^2}) = |\zeta|\frac{b}{n} + O(\frac{1}{n^2}).$$

On the other hand, for z on that circle,

$$|\varphi_n(z) - \psi(z)| = \frac{|\zeta^2e^\zeta|}{2n} + O(\frac{1}{n^2}).$$

By Rouché's Theorem (see [1]), it follows that if $b > |\zeta e^\zeta/2|$ and n is large enough, then φ_n has the same number of zeros as ψ in the disk $\{|z - \zeta| < b/n\}$, namely, exactly one zero. We thus obtain (7) from (10). Similarly, (8) follows from (11). To get (9) note that from (8) it follows that $\Im\lambda = O(1/n)$ and $\Re\lambda = \Theta(1)$. Therefore

$$|\lambda| = \sqrt{(\Re\lambda)^2 + (\Im\lambda)^2} = |\Re\lambda| + O(\frac{1}{n^2})$$

which by (8) again implies that

$$1 - |\lambda| = 1 - |\Re\lambda| + O(\frac{1}{n^2}) = \frac{1 + \Re\zeta}{n} + O(\frac{1}{n^2}).$$

It remains to prove that the equation $e^z - z - 1 = 0$ has a nonzero root z. (Plainly $z = 0$ is a root.) To this end, write $z = x + iy$ to get

$$e^x\cos y = 1 + x \quad \text{and} \quad e^x \sin y = y\,.$$

Solve for x to get $x = y\cos y/\sin y - 1$. Inserting this value of x into the second equation we get

$$\frac{y}{\sin y} = \exp\left(\frac{y\cos y}{\sin y} - 1\right). \qquad (12)$$

We will find a solution of the form $y = 2\pi m + a$, where $\pi/4 < a < \pi/2$. Note that if $y = 2\pi m + \pi/4$, then the left hand side of (12) is $\sqrt{2}y$, while the right hand side is $\exp(y - 1)$, which is strictly larger than $\sqrt{2}y$ for all $m \geq 1$. If, on the other hand, $y = 2\pi m + \pi/2$, then the left hand side is y while the right hand side is $\exp(-1)$, which is strictly smaller than y. We conclude that for all integers $m \geq 1$, there exists at least one solution $y = 2\pi m + a$, where $\pi/4 < a < \pi/2$. ■

2.3. The test function

In this subsection we fix an eigenvalue λ of M such that $|\lambda| \geq 1 - O(1/n)$, whose existence is guaranteed by the previous subsection, and let $f : [n] \to \mathbb{C}$ be a corresponding eigenfunction. We will denote the states of the n cards at time t by $\sigma_t(0), \ldots, \sigma_t(n-1)$, and assume that at time 0 we start with the identity permutation, so $\sigma_0(i) = i$ for all i. We emphasize that σ_t is obtained from σ_{t-1} by first transposing the card at state 0 with a uniform random card, and then moving all cards one state up (modulo n). Thus for each i, the sequence $\{\sigma_t(i)\}_{t\geq 0}$ is a Markov chain with transition matrix M. To relate this to the description of the cyclic-to-random shuffle $\{\sigma_t^*\}$ in the introduction, observe that σ_t^* is obtained from σ_t by a rotation through $t \mod n$.

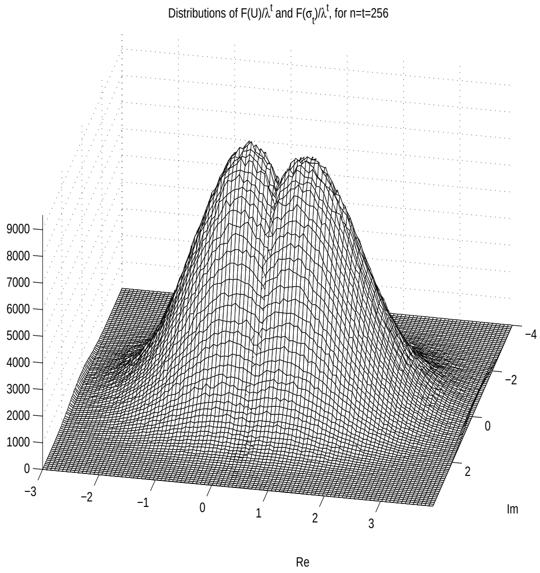

Figure 1. The density functions of the distributions of $F(\mathcal{U})$ and $F(\sigma_t)$ (normalized by λ^t) for $n = t = 256$ drawn in the complex plane

We focus on the following test function $F : S_n \to \mathbb{C}$:

$$F(\sigma) = \frac{1}{n}\sum_{i=0}^{n-1} f(\sigma(i))\overline{f(i)}. \tag{13}$$

We will study the distribution of F under both the uniform distribution and under the distribution of σ_t. See Figure 1 for a numerical sample of F under these two distributions for $t = n = 256$, provided by Ilya Mironov. The fact that these two distributions are significantly different for $t = o(n \log n)$ will yield our lower bound on the mixing time of the cyclic-to-random shuffle.

Since f satisfies $\sum_{i=0}^{n-1} f(i) = 0$, under the uniform distribution $\mathcal{U}$ on S_n we have

$$\mathbf{E}_{\mathcal{U}}[F(\sigma)] = 0. \tag{14}$$

It is also easy to see that F is an eigenfunction of the shuffle, i.e.,

$$\begin{aligned} \mathbf{E}[F(\sigma_{t+1})|\sigma_t] &= \frac{1}{n}\sum_{i=0}^{n-1}\mathbf{E}[f(\sigma_{t+1}(i))|\sigma_t]\overline{f(i)} \\ &= \lambda F(\sigma_t) \end{aligned} \tag{15}$$

and therefore

$$\mathbf{E}[F(\sigma_t)] = \lambda^t F(\sigma_0) = \lambda^t \|f\|_2^2, \tag{16}$$

where $\|\cdot\|_2$ denotes the ℓ_2-norm w.r.t. the uniform distribution on $[n]$, i.e., $\|f\|_2^2 = \frac{1}{n}\sum_{i=0}^{n-1}|f(i)|^2$.

We now calculate the second moment of $F(\sigma)$ under the stationary distribution.

Lemma 2.3

$$\mathbf{E}_{\mathcal{U}}\Big(|F(\sigma)|^2\Big) = \frac{\|f\|_2^4}{n-1}.$$

Proof: Clearly $\mathbf{E}_{\mathcal{U}}\Big(|F(\sigma)|^2\Big)$ equals

$$\begin{aligned} &\frac{1}{n^2}\sum_{i\neq j}\mathbf{E}_{\mathcal{U}}\Big(f(\sigma(i))\overline{f(\sigma(j))}\Big)f(j)\overline{f(i)} \\ +\; &\frac{1}{n^2}\sum_{i}\mathbf{E}_{\mathcal{U}}\Big(|f(\sigma(i))|^2\Big)|f(i)|^2. \end{aligned} \tag{17}$$

The second term in (17) can be evaluated as

$$\begin{aligned} \frac{1}{n^2}\sum_i \mathbf{E}_{\mathcal{U}}\Big(|f(\sigma(i))|^2\Big)|f(i)|^2 &= \frac{\|f\|_2^2}{n^2}\sum_i |f(i)|^2 \\ &= \frac{\|f\|_2^4}{n}. \end{aligned} \tag{18}$$

Now let $i \neq j$ and let η be an independent copy of σ. Then $\mathbf{E}_{\mathcal{U}}[f(\sigma(i))\overline{f(\sigma(j))}]$ is equal to

$$\frac{n}{n-1}\left(\mathbf{E}_{\mathcal{U}}\Big(f(\sigma(i))\overline{f(\eta(j))}\Big)\frac{1}{n}\mathbf{E}_{\mathcal{U}}\Big(|f(\sigma(i))|^2\Big)\right),$$

which in turn is the same as

$$-\frac{\mathbf{E}_{\mathcal{U}}\Big(|f(\sigma(i))|^2\Big)}{n-1} = -\frac{\|f\|_2^2}{n-1}.$$

Similarly,

$$\sum_{i\neq j} f(j)\overline{f(i)} = \sum_i\sum_j f(j)\overline{f(i)} - \sum_i |f(i)|^2 = -n\|f\|_2^2.$$

Therefore, the first term in (17) can be evaluated as

$$-\frac{\|f\|_2^2}{n^2(n-1)}\sum_{i\neq j} f(j)\overline{f(i)} = \frac{\|f\|_2^4}{n(n-1)}. \tag{19}$$

Substituting (18) and (19) into (17) completes the proof of the lemma. ■

For later use, we record here a simple variational bound on f:

Lemma 2.4 *We have*

$$\frac{\|f\|_\infty}{\|f\|_2} = \frac{\|\chi\|_\infty}{\|\chi\|_2} + O(1/n). \tag{20}$$

where χ is defined in (2).

Proof: It follows from (5) that for all $k \neq 0$,

$$\begin{aligned} f(k) &= 1+\frac{n}{n-1}(\lambda-1)\frac{\gamma^{k-1}-1}{\gamma-1} \\ &= 1-\left(\frac{\zeta+1}{n-1}+O(\frac{1}{n^2})\right)\frac{e^{\frac{(k-1)\zeta}{n}}-1+O(\frac{1}{n})}{\frac{\zeta}{n}+O(\frac{1}{n^2})} \\ &= 1-\frac{\zeta+1}{\zeta}(e^{\frac{(k-1)\zeta}{n}}-1)+O(\frac{1}{n}). \end{aligned} \tag{21}$$

Now it is easy to see that $\|f\|_2 = \|\chi\|_2 + O(1/n)$ and that $\|f\|_\infty = \|\chi\|_\infty + O(1/n)$. The proof follows. ■

2.4. The second moment of $F(\sigma_t)$

We begin with an estimate of the contribution to the second moment from a specific pair of cards. Fix two distinct cards, i and j. Denote by $A_i(s) = \{\sigma_s(i) = 0\}$ the event that at step s card i is in state 0 (so it will be transposed with a uniform random card in the next step). Let

$$N_{ij}(t) = \sum_{s=0}^{t-1}(\mathbf{P}[A_i(s)] + \mathbf{P}[A_j(s)])$$

denote the expected number of times $s < t$ where one of cards i, j was at state 0. Since at each step there is exactly one card in state 0, we have $\sum_{i=0}^{n-1}\sum_{s=0}^{t-1}\mathbf{P}[A_i(s)] = t$ and therefore

$$\sum_{i\neq j} N_{ij}(t) \leq 2nt\,. \tag{22}$$

Next, we will couple $\{\sigma_t\}$ with a process $\{(\eta_t, \widetilde{\eta}_t)\}$, where η and $\widetilde{\eta}$ are two *independent* copies of the cyclic-to-random shuffle starting from the identity permutation. We will observe the motions of cards i, j in $\eta, \widetilde{\eta}$ respectively; note that, in contrast to σ_t, these two motions are independent. We use the coupling to bound the dependence between the cards in σ.

Lemma 2.5 *For any two cards $i \neq j$ and all t, the quantity*

$$\left|\mathbf{E}\left[f(\sigma_t(i))\overline{f(\sigma_t(j))}\right] - \mathbf{E}\left[f(\eta_t(i))\overline{f(\widetilde{\eta}_t(j))}\right]\right| \tag{23}$$

is bounded above by

$$\frac{4t+4nN_{ij}(t)}{n^2}\|f\|_\infty^2. \tag{24}$$

Proof: We inductively define a coupling of the process $\{\sigma_t\}$ and the pair process $\{(\eta_t, \widetilde{\eta}_t)\}$. If $(\sigma_s(i), \sigma_s(j)) \neq (\eta_s(i), \widetilde{\eta}_s(j))$ then the updates for the σ and $(\eta, \widetilde{\eta})$ are performed independently. Otherwise, we have

$$(\sigma_s(i), \sigma_s(j)) = (\eta_s(i), \widetilde{\eta}_s(j))\,, \tag{25}$$

and there are three cases to consider in the definition of the coupling at step $s+1$:

Case 1. Card i is in state 0 at time s.

Case 2. Card j is in state 0 at time s.

Case 3. Both cards i, j are not in state 0 at time s.

In Case 1, the transition probabilities to $(\sigma_{s+1}(i), \sigma_{s+1}(j))$ are given by

$$\begin{cases} (\ell, \sigma_s(j)+1) & \text{w.p. } \frac{1}{n} \quad \forall \ell \neq \sigma_s(j)+1, \\ (\sigma_s(j)+1, 1) & \text{w.p. } \frac{1}{n}, \end{cases}$$

and we define $(\eta_{s+1}(i), \widetilde{\eta}_{s+1}(j))$ to be

$$\begin{cases} (\ell, \widetilde{\eta}_s(j)+1) & \text{w.p. } \frac{n-1}{n^2} \quad \forall \ell \neq \widetilde{\eta}_s(j)+1 \\ (\widetilde{\eta}_s(j)+1, 1) & \text{w.p. } \frac{1}{n^2}, \\ (\widetilde{\eta}_s(j)+1, \widetilde{\eta}_s(j)+1) & \text{w.p. } \frac{n-1}{n^2}, \\ (\ell, 1) & \text{w.p. } \frac{1}{n^2} \quad \forall \ell \neq \widetilde{\eta}_s(j)+1 \end{cases}$$

Thus, given that the processes satisfy (25) at time s and that at that time card i is at location 0, we may couple the processes to satisfy (25) at time $s+1$ with conditional probability at least $\frac{(n-1)^2}{n^2} > 1 - \frac{2}{n}$. Similarly, in Case 2, if the coupling satisfies (25) at time s then (25) can be satisfied at time $s+1$ with conditional probability at least $1 - \frac{2}{n}$.

In Case 3, the transition probabilities to $(\sigma_{s+1}(i), \sigma_{s+1}(j))$ are given by

$$\begin{cases} (\sigma_s(i)+1, \sigma_s(j)+1) & \text{w.p. } 1 - \frac{2}{n}, \\ (\sigma_s(i)+1, 1) & \text{w.p. } \frac{1}{n}, \\ (1, \sigma_s(j)+1) & \text{w.p. } \frac{1}{n}. \end{cases}$$

and we define $(\eta_{s+1}(i), \widetilde{\eta}_{s+1}(j))$ to be

$$\begin{cases} (\eta_s(i)+1, \widetilde{\eta}_s(j)+1) & \text{w.p. } 1 - \frac{2}{n} + \frac{1}{n^2}, \\ (\eta_t(i)+1, 1) & \text{w.p. } \frac{1}{n} - \frac{1}{n^2}, \\ (1, \widetilde{\eta}_s(j)+1) & \text{w.p. } \frac{1}{n} - \frac{1}{n^2}, \\ (1, 1) & \text{w.p. } \frac{1}{n^2}. \end{cases}$$

It therefore follows that in Case 3, if the processes satisfy (25) at time s, they may be coupled to satisfy it at time $s+1$ with conditional probability at least $1 - \frac{4}{n^2}$.

It follows that the probability that the processes "unglue" by time t (i.e., (25) fails for some $s \leq t$) is at most

$$\frac{2}{n}N_{ij}(t) + \frac{2t}{n^2}\,. \tag{26}$$

We now estimate the difference of expected values in (23). On the event where the processes satisfy (25) at time t we get a contribution of zero. On the complementary event we get a contribution bounded by $2\|f\|_\infty^2$. We thus obtain the bound (24) from (26). ■

Since the processes η and $\widetilde{\eta}$ defined in the foregoing proof are independent, it follows as in (16) that

$$\begin{aligned} \mathbf{E}\left[f(\eta_t(i))\overline{f(\widetilde{\eta}_t(j))}\right] &= \mathbf{E}[f(\eta_t(i))]\mathbf{E}[\overline{f(\widetilde{\eta}_t(j))}] \\ &= \lambda^t f(i)\overline{\lambda^t f(j)} \\ &= |\lambda|^{2t} f(i)\overline{f(j)}. \end{aligned}$$

Therefore, from Lemma 2.5 we obtain

Corollary 2.6 *For any two cards $i \neq j$ and all t, the quantity*

$$\left|\mathbf{E}\left(f(\sigma_t(i))\overline{f(\sigma_t(j))}\right)\right|$$

is bounded above by

$$\left(|\lambda|^{2t} + \frac{4t + 4nN_{ij}(t)}{n^2}\right)\|f\|_\infty^2.$$

We are now in a position to bound the second moment of F.

Lemma 2.7 $\mathbf{E}\left[|F(\sigma_t)|^2\right]$ *is bounded above by*

$$\left(|\lambda|^{2t} + \frac{12t + n}{n^2}\right)\|f\|_\infty^4.$$

Proof: We have

$$\begin{aligned}\mathbf{E}\left[|F(\sigma_t)|^2\right] &= \frac{1}{n^2}\sum_{i \neq j}\mathbf{E}\left[f(\sigma_t(i))\overline{f(\sigma_t(j))}\right]f(j)\overline{f(i)} \\ &+ \frac{1}{n^2}\sum_{i}\mathbf{E}\left[|f(\sigma_t(i))|^2\right]|f(i)|^2. \qquad (27)\end{aligned}$$

To deal with the second term, note that

$$\frac{1}{n^2}\sum_{i}\mathbf{E}\left[|f(\sigma_t(i))|^2\right]|f(i)|^2 \leq \frac{\|f\|_\infty^4}{n}. \qquad (28)$$

Turning to the first term, by Corollary 2.6, for any $i \neq j$,

$$\left|\mathbf{E}\left[f(\sigma_t(i))\overline{f(\sigma_t(j))}\right]f(j)\overline{f(i)}\right|$$

is bounded above by

$$\left(|\lambda|^{2t} + \frac{4t + 4nN_{ij}(t)}{n^2}\right)\|f\|_\infty^4. \qquad (29)$$

Inserting (28) and (29) into (27) we obtain that $\mathbf{E}\left[|F(\sigma_t)|^2\right]$ is bounded above by

$$\frac{\|f\|_\infty^4}{n^2}\left(n + n^2|\lambda|^{2t} + 4t + \frac{4n}{n^2}\sum_{i \neq j}N_{ij}(t)\right),$$

which is in turn bounded above by

$$\frac{\|f\|_\infty^4}{n^2}\left(n + n^2|\lambda|^{2t} + 12t\right),$$

using (22). This completes the proof. ■

2.5. The mixing time

Given the bound on the second moment of our test function from the previous section, and the bound on the eigenvalue from section 2.2, it is straightforward to derive a lower bound on the mixing time.

Proof of Theorem 1.1 Recall from Lemma 2.2 that the equation $e^z - z - 1 = 0$ has nonzero roots, and let ζ be such a root. By Lemma 2.2 it follows that there exists a solution γ of the equation $(n-1)\gamma^n - n\gamma^{n-1} + 1 = 0$ satisfying (7) and (8). Fix such a γ, let $\lambda = (1 - 1/n)\gamma$ and let f be a corresponding eigenfunction of M.

We use the test function F based on f, as defined in (13). Let μ_t be the distribution of σ_t in the cyclic-to-random shuffle where σ_0 is the identity permutation, and recall that $\mathcal{U}$ denotes the (uniform) stationary distribution on S_n. Let g^2 be the density of μ_t with respect to $\nu = (\mu_t + \mathcal{U})/2$. Let h^2 be the density of $\mathcal{U}$ with respect to ν.

By (14) and (16) we have that

$$\begin{aligned}|\lambda|^t\|f\|_2^2 &= |\mathbf{E}_{\mu_t}[F] - \mathbf{E}_{\mathcal{U}}[F]| \\ &= \left|\int Fg^2\,d\nu - \int Fh^2\,d\nu\right|.\end{aligned}$$

On the other hand,

$$\left|\int Fg^2\,d\nu - \int Fh^2\,d\nu\right|^2 = \left|\int F(g+h)(g-h)\,d\nu\right|^2,$$

which by Cauchy-Schwartz is bounded by

$$\int |F|^2(g+h)^2\,d\nu \times \int (g-h)^2\,d\nu\,.$$

By Lemmas 2.3 and 2.7, $\int |F|^2(g+h)^2\,d\nu$ is bounded above by

$$\begin{aligned}&2\int |F|^2g^2\,d\nu + 2\int |F|^2h^2\,d\nu \\ =\;& 2\mathbf{E}_{\mu_t}\left(|F|^2\right) + 2\mathbf{E}_{\mathcal{U}}\left(|F|^2\right) \\ \leq\;& \frac{2\|f\|_2^4}{n-1} + 2\left(|\lambda|^{2t} + \frac{12t+n}{n^2}\right)\|f\|_\infty^4 \\ \leq\;& 2\left(|\lambda|^{2t}\frac{12t+3n}{n^2}\right)\|f\|_\infty^4.\end{aligned}$$

Moreover,

$$\int (g-h)^2\,d\nu \leq \int |g^2 - h^2|\,d\nu = 2\|\mu_t - \mathcal{U}\|_{\mathrm{TV}}.$$

Thus,

$$\|\mu_t - \mathcal{U}\|_{\mathrm{TV}} \geq \frac{|\lambda|^{2t}\|f\|_2^4}{4\|f\|_\infty^4\left(|\lambda|^{2t} + \frac{12t+3n}{n^2}\right)}.$$

Recalling Lemma 2.4, we conclude that the last expression equals

$$\left(\frac{\|\chi\|_2^4}{4\|\chi\|_\infty^4} + O\left(\frac{1}{n}\right)\right)\left(\frac{|\lambda|^{2t}}{|\lambda|^{2t} + \frac{12t+3n}{n^2}}\right).$$

It follows that (1) holds when $|\lambda|^{2t} \geq \frac{12t+3n}{n^2}$. Note that if $t \geq n$, then $\frac{12t+3n}{n^2} \leq \frac{15t}{n^2}$. Therefore (1) holds if $t \geq n$ and

$$-2t\log|\lambda| + \log t \leq 2\log n - \log(15). \qquad (30)$$

Note that by (9) we have $-\log|\lambda| \le |\Re\zeta + 1|/n + O(1/n^2)$. Therefore, taking

$$t = \frac{1}{2|\Re\zeta + 1|} n \left(\log n - \log\log n - b\right)$$

where b is a large constant we obtain (30) as needed.

3. An upper bound for general semi-random transpositions

In this section we prove Theorem 1.2.

Proof: By the triangle inequality it suffices to prove the theorem assuming that the L_t are deterministic. We thus restrict to that case.

We define a *strong uniform time* for the shuffle, i.e., a stopping time T with the property that, given $T = t$, the random permutation σ_t^* has the uniform distribution over S_n. It is well known (see, e.g., [3]) that, if T is a strong uniform time, then the distribution μ_t^* of σ_t^* satisfies

$$\|\mu_t^* - \mathcal{U}\|_{\mathrm{TV}} \le \mathbf{P}[T > t] \quad \forall t.$$

Following Broder (as described in [9]) and Mironov [16], we define the stopping time in terms of a card marking process as follows. Initially all cards are unmarked. First, the card initially at L_1 is marked. Later, at time t, we mark the card at L_t if it is unmarked and the card at R_t is already marked, and also if $R_t = L_t$ and this location has an unmarked card. Once a card is marked it remains so at all future times. Set T to be the first time t at which all cards are marked. Clearly T is a stopping time. The theorem follows immediately from the following two claims:

Claim 1: T is a strong uniform time.

Claim 2: There exists $C_0 < \infty$ such that for any $C_1 > C_0$ we have

$$\mathbf{P}\Big(T > C_1 n \log n\Big) \le n^{-\beta},$$

for some $\beta = \beta(C_1) > 0$. Specifically, this holds for $C_0 = 32\theta^{-3} + \theta^{-1}$, where $\theta = e^{-2}(1 - e^{-1})/2$.

Proof of Claim 1: By induction, it is easy to check the following. At any time t, given that k cards have been marked, conditional on the set of marked cards and their locations, the mapping between these two sets (assigning to every marked card its location) is uniformly distributed among the $k!$ possibilities. See [16] or [9] for details.

Proof of Claim 2: Divide time into successive *epochs* of length $2n$, starting after the card at L_1 is marked. Denote by u_k the fraction of unmarked cards before epoch k, so $u_1 = 1 - 1/n$. Let $m_k = 1 - u_k$. Let $\mathcal{H}_k$ denote the history of the process prior to epoch k, and note that u_k is a function of $\mathcal{H}_k$.

Claim 3: $\mathbf{E}(u_{k+1}|\mathcal{H}_k) \le u_k[1 - 2\theta m_k]$ for all k, where $\theta = e^{-2}(1 - e^{-1})/2$.

Proof: Consider a card x, unmarked before epoch k. Of the $2n$ prescribed locations $\{L_t\}$ in the epoch, at most n are their last occurrence in the epoch. Thus for $1 \le j \le n$ we can find $t(j) < s(j)$ in the epoch such that $L_{t(j)} = L_{s(j)}$. For each $j \le n$, we have $R_{t(j)} = x$ with probability $1/n$. Therefore, the event A_x that there exists a $j \le n$ satisfying $R_{t(j)} = x$, has probability

$$\mathbf{P}(A_x|\mathcal{H}_k) \ge 1 - (1 - 1/n)^n \ge 1 - e^{-1}.$$

On A_x, we fix j to be minimal such that $R_{t(j)} = x$. Given A_x and $\mathcal{H}_k$, with probability at least $(1 - 1/n)^{2n-2} > e^{-2}$, we have $R_t \ne L_{t(j)}$ for all t such that $t(j) < t < s(j)$. In that case, x is untouched by the random choices between times $t(j)$ and $s(j)$, and then with probability at least m_k the card at $R_{s(j)}$ is one of the nm_k cards marked prior to epoch k. Thus x gets marked with probability at least $2\theta m_k$. The assertion of Claim 3 follows. ■

Proof of Claim 2 continued: Using Claim 3, we first quantify the time to mark at least half the cards (i.e., to achieve $m_k \ge 1/2$), and then the time to mark the remaining cards (i.e., to achieve $u_k < 1/n$). Denote by D_k the number of cards that get marked during epoch k as a result of being transposed with a card that was marked prior to epoch k. Clearly $m_{k+1} \ge m_k + D_k/n$. The proof of Claim 3 implies that

(i) if $m_k < 1/2$, then $\mathbf{E}(D_k|\mathcal{H}_k) \ge \theta n m_k$;

(ii) if $m_k \ge 1/2$, then $\mathbf{E}(u_{k+1}|\mathcal{H}_k) \le (1 - \theta)u_k$.

To bound the number of epochs where $m_k < 1/2$, we need a stochastic lower bound for D_k:

Claim 4: If $m_k < 1/2$, then

$$\mathbf{P}\Big(D_k \ge \frac{\theta n m_k}{2} \,\Big|\, \mathcal{H}_k\Big) \ge \frac{\theta^2}{8}.$$

Proof: Using the notation in the proof of Claim 3, Denote by $\widetilde{D}_k$ the number of $j \le n$ such that $R_{s(j)}$ is one of the nm_k cards marked prior to epoch k. Clearly $D_k \le \widetilde{D}_k$. The distribution of $\widetilde{D}_k$ is Binomial(n, m_k), and this also holds given $\mathcal{H}_k$. Therefore,

$$\mathbf{E}(D_k^2|\mathcal{H}_k) \le \mathbf{E}(\widetilde{D}_k^2|\mathcal{H}_k) \le (nm_k)^2 + nm_k \le 2(nm_k)^2.$$

In conjunction with (i) above, this yields

$$\mathbf{E}(D_k^2|\mathcal{H}_k) \le C_2 \mathbf{E}(D_k|\mathcal{H}_k)^2\,,$$

where $C_2 = 2\theta^{-2}$. A standard second moment bound (see, e.g., [15, p. 8]) now yields Claim 4. ■

Proof of Claim 2 concluded: Call epoch k a "growth epoch" if $m_{k+1} \ge (1 + \theta/2)m_k$. Call epoch k a "good epoch" if it is a growth epoch or it satisfies $m_k \ge 1/2$. Claim 4 implies that the conditional probability that epoch k is a good epoch, given $\mathcal{H}_k$, is at least $\theta^2/8$. Thus the number of good epochs among the first $k_3 = C_3 \log n$ epochs stochastically dominates a Binomial$(k_3, \theta^2/8)$ random variable. Fix $C_3 > 32\theta^{-3}$, and denote by Ω_3 the event that there are at least $(4 \log n)/\theta$ good epochs among the first k_3 epochs. Recall that the probability that a binomial random variable differs from its mean by a constant multiple of the mean decays exponentially in the number of trials $k_3 = C_3 \log n$. We infer that $\mathbf{P}(\Omega_3^c) < n^{-\beta}/2$ for some $\beta > 0$. Moreover, since $(1 + \theta/2)^{4/\theta} > e$ and $m_1 = 1/n$, the number of growth epochs must be smaller than $(4 \log n)/\theta$. Thus on Ω_3 we have $m_{k_3} \geq 1/2$.

Turning now to the second portion, once $m_k \geq 1/2$ we have from (ii) above that $\mathbf{E}(u_{k+1}|u_k) \leq (1 - \theta)u_k$. Therefore, for all $k > 0$ we have

$$\mathbf{E}(u_{k_3+k} \,|\, \Omega_3, u_{k_3}) \leq (1 - \theta)^k u_{k_3} \leq \frac{e^{-\theta k}}{2}\,.$$

Thus if $k = (1 + \beta)\theta^{-1} \log n$, then

$$\begin{aligned} \mathbf{P}(u_{k_3+k} \geq 1/n \,|\, \Omega_3) \quad &\leq \quad \mathbf{E}(n u_{k_3+k} \,|\, \Omega_3) \\ &\leq \quad n \cdot \frac{n^{-1-\beta}}{2} = \frac{n^{-\beta}}{2}\,. \end{aligned}$$

In conjunction with the bound for $\mathbf{P}(\Omega_3^c)$, this implies that $\mathbf{P}(u_{k_3+k} \geq 1/n) \leq n^{-\beta}$ for this value of k. In other words, if $C_1 > C_0 = 32\theta^{-3} + \theta^{-1}$ and $k_1 = C_1 \log n$, then there exists $\beta = \beta(C_1) > 0$ such that $\mathbf{P}(u_{k_1} \geq 1/n) \leq n^{-\beta}$. This completes the proof of Claim 2 and hence of the theorem. ■

4. Concluding remarks and further problems

1. We have shown that the cyclic-to-random shuffle on n cards has mixing time of order $\Theta(n \log n)$. However, the constant in our general upper bound, and that in the specific cyclic-to-random upper bound of Mironov [16], are significantly larger than the constant in our lower bound. We believe that the lower bound is closer to the truth and that this shuffle exhibits the "cutoff phenomenon", i.e., there is a constant C_* such that for $t < (1 - \epsilon)C_* n \log n$ the distribution after t steps, μ_t^*, satisfies $\|\mu_t^* - \mathcal{U}\|_{\mathrm{TV}} = 1 - o(1)$ as $n \to \infty$, while for $t > (1 + \epsilon)C_* n \log n$ we have $\|\mu_t^* - \mathcal{U}\|_{\mathrm{TV}} = o(1)$ as $n \to \infty$. Proving this, and determining C_*, remain a challenge. We note that repeating the proof of Theorem 1.1 one obtains the same lower bound (1) for $\tau(\epsilon)$ for all small ϵ. Is $C_* = (2|\Re\zeta + 1|)^{-1}$?

2. Given the test function F one can define a "distinguisher" between the uniform distribution and the distribution of σ_t by letting $D(\sigma) = 1_{\{\Re(\lambda^{-t}F(\sigma))>0\}}$. Note that $\mathbf{E}_{\mathcal{U}}[\lambda^{-t}F] = 0$ while $\mathbf{E}_{\mu_t^*}[\lambda^{-t}F(\sigma)] = \|f\|_2^2$. Clearly,

   $$\|\mu_t^* - \mathcal{U}\|_{\mathrm{TV}} > Adv_t(D) := \mathbf{E}_{\mu^t}[D] - \mathbf{E}_{\mathcal{U}}[D]$$

   (the quantity $Adv_t(D)$ is often referred to as the *advantage* of the distinguisher D). See Figure 2 for experiments run by Ilya Mironov. These experiments suggest that, for each fixed distinguishing probability for D, the number of steps t should scale as $n \log n$, as expected from the $\Theta(n \log n)$ mixing time of the shuffle.

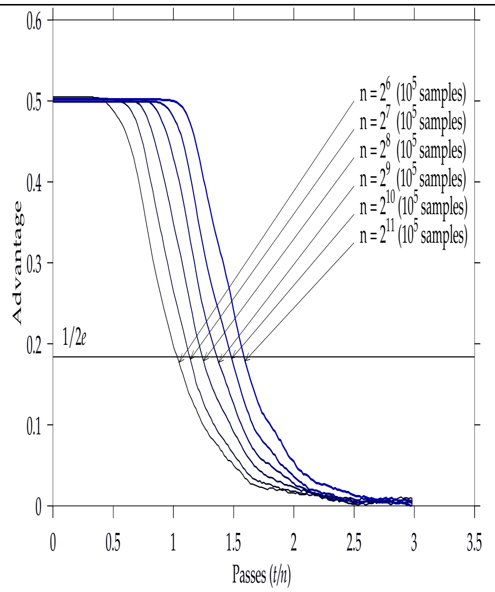

Figure 2. The advantage of the test $D(\sigma) = 1_{\{\Re(\lambda^{-t}F(\sigma))>0\}}$ as function of t/n for various values of n.

3. For which sequence $\{L_t\}$ does the resulting semi-random transposition shuffle on n cards have the largest mixing time?
   We suspect that the slowest shuffle in this class is the "star transpositions" shuffle, for which $L_t = 0$ for all t, and the mixing time is $(1 + o(1))n \log n$ by [8].

4. Is there a universal constant $c > 0$ such that, for any semi-random transposition shuffle on n cards, the mixing time is at least $cn \log n$?
   For this lower bound question there is no obvious reduction to the case where the sequence $\{L_t\}$ is deterministic, so conceivably the question could have different answers for deterministic $\{L_t\}$ and random $\{L_t\}$. Two specific cases of interest are:

   - For each $k \geq 0$, let $\{L_{kn+r}\}_{r=1}^n$ be a uniform random permutation of $\{0, \ldots, n - 1\}$, where these permutations are independent.
   - Let $\{L_t\}$ be a Markov chain with memory 2, where $L_1 = 0, L_2 = 1$ and for each $t \geq 3$ we have $L_{t+1} = 2L_t - L_{t-1} \mod n$ with probability $1 - 1/n$ and $L_{t+1} = L_{t-1}$ with probabil-

    ity $1/n$. This choice of $\{L_t\}$ was suggested to us by Igor Pak (personal communication), motivated by [7].

   Each of these examples has a “quenched” version, where the sequence $\{L_t\}$ is picked in advance and then used as a deterministic sequence, and an “annealed” version, where the $\{L_t\}$ are random variables with the specified distribution.

5. What, if any, are the implications of our results for the analysis of RC4? We first indicate briefly the connection to RC4, as argued by Mironov [16]. RC4 is a stream cipher whose output stream is generated from an internal state consisting of a pseudo-random permutation on $[n] = \{0, 1, \ldots, n-1\}$. (In practice, $n = 256$.) Encryption is performed by XORing the output stream (i.e., elements of $[n]$) with the plaintext. The permutation is initialized by exchanging the number at position t with that at a “pseudo-random” position, for $t = 0, 1, \ldots, n-1$. At each subsequent step, a similar exchange operation is performed. Mironov [16] argues that the essential flavor of RC4 is retained if one replaces the “pseudo-random” position by a truly random one. In this case, the operations performed on the permutation are exactly the cyclic-to-random shuffle on n cards. Before the mixing time of the cyclic-to-random shuffle, “traces” of the initial permutation remain. If one accepts the argument that the shuffle captures the essential features of RC4, then the lower bound of $\Omega(n \log n)$ on the mixing time suggests the presence of a statistical bias in the output stream that persists for a significant number of passes. (A “pass” is a sequence of $n = 256$ outputs.) However, it is not clear how to exploit this bias to produce a computationally efficient test, or “distinguisher”, that could form the basis of an attack.

Acknowledgments: We are grateful to Serban Nacu for his help with some of the complex analysis used in this paper. We heard about the cyclic-to-random mixing problem and the connection to RC4 in a talk by Ilya Mironov at the Bay Area Theory Seminar in Fall, 2002. We are grateful to Ilya Mironov for many helpful comments on an earlier version of the paper, and also for permission to include the results of his computer simulations (Figures 1 and 2).

References

[1] L.V. Ahlfors, *Complex analysis (3rd ed.)*, Mcgraw-Hill, 1979.

[2] D. Aldous, Random walks on finite groups and rapidly mixing Markov chains, *Séminaire de Probabilites XVII*, 1981/82, Springer Lecture Notes in Mathematics **986** (1983), 243–297.

[3] D. Aldous and P. Diaconis, Shuffling cards and stopping times, *American Mathematical Monthly* **93** (1986), 333–348.

[4] Y. Amit, Convergence properties of the Gibbs sampler for perturbations of Gaussians, *Ann. Statist.* **24** (1996), 122–140.

[5] D. Bayer and P. Diaconis, Trailing the dovetail shuffle to its lair, *Annals of Applied Probability* **2** (1992), 294–313.

[6] I. Benjamini, N. Berger, C. Hoffman and E. Mossel, Mixing times of the biased card shuffling and the asymmetric exclusion process, *Trans. Amer. Math. Soc.*, to appear.

[7] F. Chen, L. Lovász and I. Pak, Lifting Markov chains to speed up mixing, *Proceedings of the 31st Annual ACM Symposium on Theory of Computing*, 1999, 275–281.

[8] P. Diaconis, Application of non-commutative Fourier analysis to probability problems. In P. L. Hennequin (ed.), *École d' Été de Probabilités de St. Flour* XV-XVII, 1985–1987, Springer Lecture Notes in Mathematics **1362**, Springer-Verlag, Berlin, 1988, 51–100.

[9] P. Diaconis, *Group representations in probability and statistics*, Institute of Mathematical Statistics Lecture Notes–Monograph Series **11**, Institute of Mathematical Statistics, Hayward CA, 1988.

[10] P. Diaconis and A. Ram, Analysis of systematic scan Metropolis Algorithms Using Iwahori-Hecke Algebra Techniques, *Michigan Math. J.* **48** (2000), 157–190.

[11] P. Diaconis and M. Shahshahani, Generating a random permutation with random transpositions, *Zeitschrift für Wahrscheinlichketistheorie und verwandte Gebiete* **57** (1981), 159–179.

[12] M. Dyer, L.A. Goldberg and M. Jerrum, “Systematic scan for sampling colorings,” Preprint, 2004.

[13] E. Gilbert, Theory of shuffling, Technical Memorandum, Bell Laboratories, 1955.

[14] J. Goodman and A. Sokal, Mutligrid Monte Carlo methods, *Phys. Rev. D.* **40** (1989), 2035–2071.

[15] J.-P. Kahane, *Some random series of functions*, 2nd Ed., Cambridge University Press, 1985.

[16] I. Mironov, (Not So) Random Shuffles of RC4, *Proceedings of CRYPTO 2002*, 304–319.

[17] I. Pak, *Random walks on groups: Strong uniform time approach*, PhD Thesis, Department of Mathematics, Harvard University, May 1997.

[18] J. Reeds, Unpublished manuscript, 1981.

[19] E. Thorp, Nonrandom shuffling with applications to the game of Faro, *Journal of the American Statistical Association* **68** (1973), 842–847.

[20] E. Thorp, Problem E 1763, *Amer. Math. Monthly* **72** (1965), 183.

[21] D. B. Wilson, Mixing times of lozenge tiling and card shuffling Markov chains, *Ann. Applied Probab.* **14** (2004), 274–325.

[22] D. B. Wilson, Mixing Time of the Rudvalis Shuffle, *Electronic Comm. Probab.* **8** (2003), 77–85.