

Mixing time bounds for overlapping cycles shuffles

Johan Jonasson^{*†‡}

October 26, 2009

Abstract

Consider a deck of n cards. Let p_1, p_2, \dots, p_n be a probability vector and consider the mixing time of the card shuffle which at each step of time picks a position according to the p_i 's and move the card in that position to the top. This setup was introduced in [5], where a few special cases were studied. In particular the case $p_{n-k} = p_n = 1/2$, $k = \Theta(n)$, turned out to be challenging and only a few lower bounds were produced. These were improved in [1] where it was shown that the relaxation time for the motion of a single card is $\Theta(n^2)$ when k/n approaches a rational number.

In this paper we give the first upper bounds. We focus on the case $m = \lfloor n/2 \rfloor$. It is shown that for the additive symmetrization as well as the lazy version of the shuffle, the mixing time is $O(n^3 \log n)$. We then consider two other modifications of the shuffle. The first one is the case $p_{n-k} = p_{n-k+1} = 1/4$ and $p_n = 1/2$. Using the entropy technique developed by Morris [7], we show that mixing time is $O(n^{5/2} \log^3 n)$ and that for the symmetrized shuffle, $O(n^2 \log^3 n)$ steps suffice. The second modification is a variant of the first, where the moves are made in pairs so that if the first move involves position n , then the second move must be taken from positions m or $m + 1$ and vice versa. Interestingly, this shuffle is much slower; the mixing time is at least of order $n^3 \log n$ and at most of order $n^3 \log^3 n$.

It is also observed that results of [1] can be modified to improve lower bounds for some $k = o(n)$.

*Chalmers University of Technology and Göteborg University

†<http://www.math.chalmers.se/~jonasson>

‡Research supported by the Swedish Research Council

AMS Subject classification : 60J10

Key words and phrases: comparison technique, Wilson's technique, relative entropy

Short title: Overlapping cycles shuffles

1 Introduction

How many times does one need to shuffle a deck of n cards to properly randomize it? This intuitively attracting question has turned out to provide one of the most important playgrounds for the development of the more general field of mixing times for Markov chains. The topic dates back to the early 20th century, and has been very lively for the last thirty years or so.

Recent important developments, relevant here, are Wilson's [9] technique for lower bounds and the entropy technique of Morris [7]. Wilson's technique uses test functions based on combinations of eigenvalues and eigenvectors for the motion of a single card, or some other simple Markov chain embedded in the full state space. Often this makes it possible to add the $\log n$ factor that was missing for previous lower bounds. The technique was developed in various directions in [10], [6] and [5]. Morris' entropy technique on the other hand, can in some situations be used to provide good upper bounds where earlier techniques were completely inadequate. In [7], this technique produced upper bounds of $\log^4 n$ for the Thorp shuffle and $(n \vee \frac{n^3}{L^3}) \log^3 n$ for Durrett's L -reversal shuffle (see [4]). These are the by far best bounds to date (but presumably still off by a few $\log n$ -factors; this seems in general to be the price to pay with this technique).

In [5], we introduced the class of "GR-shuffles" as a generalization of the (inverse) Rudvalis shuffle, for which at each step of time, either the bottom or the second bottom card is moved to the top, each with probability $1/2$. For a GR-shuffle, we have a probability vector p_1, p_2, \dots, p_n , and at each step, we pick position i with probability p_i and move the card at that position to the top. In this general form, the problem turns out to be very difficult to analyze, so the focus in [5] was on two special cases. The first of these was the case $p_{n-k+1} = \dots = p_n = \frac{1}{k}$, for some $k = k(n)$, the "bottom-to-top shuffle". Here it was shown, via coupling on one hand and a variant of Wilson's technique on the other, that the mixing time is $\Theta(\frac{n^3}{k^2} \log n)$. The second special case, the topic of this paper, was $p_{n-k} = p_n = 1/2$, $k = k(n)$ (and k odd to avoid parity problems). An application of Wilson's technique gave a lower bound of order $\frac{n^3}{k^2} \log n$, the same as for the bottom-to-top shuffle. This bound, however, is not tight, at least

not for k of large order. Indeed, it was observed that for $k = n/2$, the mixing time for single card motion is $\Omega(n^2)$. Angel, Peres and Wilson [1] developed this observation via a careful analysis of the spectrum of the single card chain. The authors observed how the eigenvalues form two cycles close to the boundary of the unit disc in the complex plane, This, combined with the structure of the shuffle itself, inspired them to propose the name “overlapping cycles shuffle” for this shuffle. In particular, they found that for any rational $\alpha \in (0, 1)$ and $k = \lfloor \alpha n \rfloor$, the relaxation time (i.e. the inverse spectral gap) is $\Theta(n^2)$. Perhaps surprisingly, they also showed that for a.e. α , the relaxation time is in fact of a different order, namely $\Theta(n^{3/2})$. Neither [5] nor [1] gave any upper bounds on the mixing time.

Here we will make progress of various sorts. First, after the necessary preliminaries in Section 2, it will be observed in Section 3 that the results of Angel, Peres and Wilson generalize to give a relaxation time of $\Theta(nk)$ when $k|n - k$ for a single card. When $k = o(n)$, this can then be used together with Wilson’s technique (which does not seem to add any extra information for $k = \Theta(n)$) to give a lower bound of order $nk \log(n/k)$, an improvement over [5] for $k = \Omega(n^{2/3})$.

From Section 4 and on, we will focus solely on the case $k = \lfloor n/2 \rfloor$ and variants on this shuffle, even though most of the results can be obviously generalized to $k = \lfloor \alpha n \rfloor$, α rational. The variant we shall mainly focus on is the one with $p_m = p_{m+1} = 1/4$ and $p_n = 1/2$, $n = 2m$; let us here call this the “triple shuffle”. A variant of the triple shuffle will also be considered. For this variant the shuffles are made in pairs, and if the first shuffle in the pair takes the card in position n to the top, then the next move must pick a card from positions m or $m + 1$ vice versa. This shuffle will be called the “equalized shuffle”. The equalized shuffle turns out to be slower and more amenable to analysis than the original shuffle, and in Section 4 we show that mixing time is $\Omega(n^3 \log n)$. (Note that if one equalizes the original overlapping cycles shuffle in the same way, the resulting shuffle will not mix due to parity problems.)

From Section 5, we turn to upper bounds. The short fifth section is concerned with upper bounds in L^2 . These are derived via the comparison technique of Diaconis and Saloff-Coste [2], which, combined with a simple path-counting argument, yields upper bounds in L^2 of $O(n^3 \log n)$ for the additive symmetrization and the lazy version of the original overlapping cycles shuffle as well as the triple shuffle, and $O(n^5 \log n)$ for the symmetrized or lazy equalized shuffle.

In Section 6, we turn back to upper bounds in total variation. For the triple shuffle and the equalized shuffle the results of Section 5 are improved. The equalized shuffle is shown to have mixing time $O(n^3 \log^3 n)$ whereas the triple shuffle mixes in time $O(n^{5/2} \log^3 n)$. We believe that the $5/2$ -exponent should be possi-

ble to improve to 2. The discrepancy, however, seems to be caused by inherent properties of the entropy technique, which we have not been able to overcome. On the other hand, if we instead consider the symmetrized version of the triple shuffle, these problems vanish and it is shown that in this case, the mixing time is indeed $O(n^2 \log^3 n)$.

2 Preliminaries

2.1 Basics

Let S be a finite set and let μ be a probability measure on S . For a signed measure ν on S and $p \in [1, \infty)$, the L^p -norm of ν with respect to π is given by

$$\|\nu\|_p = \|\nu\|_{L^p(\pi)} := \left(\sum_{i \in S} \left| \frac{\nu(i)}{\pi(i)} \right|^p \pi(i) \right)^{1/p}.$$

The total variation norm of ν is given by

$$\|\nu\|_{TV} := \frac{1}{2} \sum_{i \in S} |\nu(i)| = \max_{A \subseteq S} \nu(A).$$

Obviously $\|\nu\|_1 = 2\|\nu\|_{TV}$. By Cauchy-Schwarz, $\|\nu\|_p \leq \|\nu\|_q$ whenever $p \leq q$.

Let $\{X_t\}_{t=0}^\infty$ be an aperiodic irreducible Markov chain on S with stationary distribution π . It is common to measure the distance between the distribution of X_t and the stationary distribution by some L^p -norm or, most commonly, the total variation norm of their difference. The *mixing time* of the chain is defined as

$$\tau_{\text{mix}} := \min\{t : \|\mathbb{P}(X_t \in \cdot) - \pi\|_{TV} \leq \frac{1}{4}\}.$$

The *convergence time in L^2* is given by

$$\hat{\tau} := \min\{t : \|\mathbb{P}(X_t \in \cdot) - \pi\|_2 \leq \frac{1}{2}\}.$$

By the above, $\tau_{\text{mix}} \leq \hat{\tau}$. The *relaxation time* is defined as

$$\tau_2 = \max_{\lambda \neq 1} \frac{1}{1 - |\lambda|},$$

where the maximum is taken over eigenvalues of the transition matrix. In this presentation, the state space will always be the symmetric group on n cards: $S = S_n$, and the Markov chains will be random walks on this group, so π will be uniform. Hence

$$\|\mathbb{P}(X_t \in \cdot) - \pi\|_2^2 = n! \sum_{i \in S_n} \left(\mathbb{P}(X_t = i) - \frac{1}{n!} \right)^2$$

and

$$\|\mathbb{P}(X_t \in \cdot) - \pi\|_{TV} = \frac{1}{2} \sum_{i \in S_n} \left| \mathbb{P}(X_t = i) - \frac{1}{n!} \right|.$$

2.2 Wilson's technique

Let P be the transition matrix of $\{X_t\}$ and let $((1-\gamma)e^{i\theta}, \phi)$ be an eigenvalue/eigenvector pair for P and suppose that $\gamma \leq 1/2$. Let

$$R := \max_{i \in S} \mathbb{E} \left[|e^{-i\theta} \phi(X_{t+1}) - \phi(X_t)|^2 \mid X_t = i \right]$$

and let

$$T := \frac{1}{-\log(1-\gamma)} \left(\log |\phi(X_0)| - \frac{1}{2} \log \frac{8R}{\gamma} \right).$$

Then (one variant of) Wilson's technique states that:

Lemma 2.1 *With the above setup, $\tau_{\text{mix}} \geq T$.*

The technique was introduced in [9]. A proof for this particular version is found in [5]. The idea of the proof is to use the eigenvalue property $\mathbb{E}[\Phi(X_{t+1}) \mid X_t] = \lambda \phi(X_t)$ and R to bound the variance of $\phi(X_T)$. Then Chebyshev's inequality is used to see that X_T with high probability has a value far from what it should have, had it been stationary.

Note that when $\gamma = o(1)$, then $-\log(1-\gamma) = (1+o(1))\gamma^{-1}$ and Wilson's technique thus states that

$$\tau_{\text{mix}} \geq \gamma^{-1} \left(\log |\phi(X_0)| - \frac{1}{2} \log(8R\gamma^{-1}) \right).$$

3 Relaxation times and lower bounds for $k = o(n)$

Let P_1 be the transition matrix for the motion of a single card. The eigenvalue/eigenvector equation $\lambda\xi = P_1\xi$, setting $\xi(1) = 1$ leads to

$$\xi(j) = \lambda^{j-1}, \quad j = 1, \dots, k,$$

$$\xi(k+j) = (2\lambda - 1)^{j-1}(2\lambda^k - 1), \quad j = 1, \dots, n - k$$

and the characteristic equation

$$f(\lambda) := (2\lambda - 1)^k(2\lambda^{n-k} - 1) - 1 = 0.$$

Assume that $k|n - k$ and let

$$\lambda_0 := \left(1 - \frac{2\pi^2}{nk}\right)e^{2\pi i/k}.$$

The some algebraic manipulation shows that $f(\lambda_0) = O(k^{-2})$. Since, as is readily seen, $f'(\lambda_0) = \Theta(n)$, there is a zero of f with in distance $O(n^{-1}k^{-2})$ of λ_0 . This follows e.g. from Theorem 4.2 of [5]. Hence there is an eigenvalue of the form

$$\lambda = (1 - \gamma)e^{i\theta},$$

where $\gamma = (1 + o(1))2\pi^2/(nk)$ and $\theta = (1 + o(1))2\pi/k$.

If k is also such that $2k|n - k$, then this can be strengthened a bit further, since one can then take $\gamma = (1 + o(1))\pi^2/(2nk)$ with $\theta = (1 + o(1))\pi/k$. In summary:

Theorem 3.1 *Let τ_2^1 be the relaxation time for single card movement. Then, if $k|n - k$,*

$$\tau_2^1 \geq \frac{nk}{2\pi^2}.$$

Moreover, if $2k|n - k$, then

$$\tau_2^1 \geq \frac{2nk}{\pi^2}.$$

Minor adjustments of the above proof also lead to

Theorem 3.2 *For the triple shuffle, i.e. $p_{n-k} = p_{n-k+1} = 1/4$ and $p_n = 1/2$, with $k|n - k$ we have*

$$\tau_2^1 = \Omega((nk)^{-1}).$$

Next we plug this into Wilson's technique under the assumption $k = o(n)$. This will lead to the following result.

Theorem 3.3 *Consider the overlapping cycles shuffle with $k = o(n)$ and $k|n-k$. For this shuffle,*

$$\tau_{\text{mix}} \geq \frac{1}{4\pi^2} nk \log \frac{n}{k}.$$

Moreover, if $2k|n-k$, then

$$\tau_{\text{mix}} \geq \frac{1}{\pi^2} nk \log \frac{n}{k}.$$

Proof. Let Y_t^i be the position of card i at time t and

$$\phi(X_t) = \sum_i \xi(Y_t^i),$$

where the sum is over those i for which $\Re \xi(i) \geq 0$. Then $|\phi(X_0)| = \Theta(n)$. Since cards in positions $1, \dots, n-k-1$ all move deterministically ahead one step for each shuffle, we get $R = O(1)$. Plugging this into Lemma 2.1 finishes the proof. \square

We note that Theorems 3.1 and 3.3 improve over [5] when k is of larger order than $n^{2/3}$.

4 Lower bound for the equalized shuffle

Recall that for the equalized shuffle, the moves are made in pairs, and that the second move in each pair takes the card in position n to the top if and only if the first move does not. To be more precise, counting each pair of moves in this sense as one step, the equalized shuffle is the random walk on S_n generated by the step distribution that gives probability $1/4$ to each of the permutations

$$\begin{aligned} & (1 \ 2 \ \dots \ m)(1 \ 2 \ \dots \ n), \\ & (1 \ 2 \ \dots \ m+1)(1 \ 2 \ \dots \ n), \\ & (1 \ 2 \ \dots \ n)(1 \ 2 \ \dots \ m), \\ & (1 \ 2 \ \dots \ n)(1 \ 2 \ \dots \ m+1). \end{aligned}$$

Recall also that $n = 2m$.

Unlike the unequalized case, the cards on the lower half of deck also move in a mainly deterministic manner. We shall see that this makes the single card chain considerably slower and Wilson's technique to work smoothly.

Let, as in the previous section, P_1 be the transition matrix for a single card. Then spelling out $P_1\xi = \lambda\xi$ gives

$$\begin{aligned}
\lambda\xi(1) &= \xi(3) \\
\lambda\xi(2) &= \xi(4) \\
&\vdots \\
\lambda\xi(m-2) &= \xi(m) \\
\lambda\xi(m-1) &= \frac{1}{4}\xi(1) + \frac{3}{4}\xi(m+1) \\
\lambda\xi(m) &= \frac{1}{4}\xi(1) + \frac{1}{4}\xi(2) + \frac{1}{4}\xi(m+1) + \frac{1}{4}\xi(m+2) \\
\lambda\xi(m+1) &= \frac{1}{4}\xi(2) + \frac{3}{4}\xi(m+2) \\
\lambda\xi(m+2) &= \xi(m+3) \\
\lambda\xi(m+3) &= \xi(m+4) \\
&\vdots \\
\lambda\xi(n-1) &= \xi(n) \\
\lambda\xi(n) &= \frac{1}{2}\xi(1) + \frac{1}{2}\xi(2).
\end{aligned}$$

Let us assume that m is even. (The analysis will take on a slightly different form when m is odd. We leave this easy modification to the reader.) Let $r := (m-2)/2$ and set $\xi(n) := 1$. Then solving backwards equations $n-1, \dots, m+2$ gives $\xi(n-j) = \lambda^{-j}$, $j = 1, 2, \dots, m-1$, in particular

$$\xi(m+2) = \lambda^{-2r}.$$

Solving forward the first $m-2$ equations gives $\xi(2j+1) = \lambda^j\xi(1)$ and $\xi(2j) = \lambda^{j-1}\xi(2)$, $j = 1, 2, \dots, r$, in particular

$$\xi(m-1) = \lambda^r\xi(1)$$

and

$$\xi(m) = \lambda^r\xi(2).$$

Equation n gives $\xi(2) = 2\lambda - \xi(1)$ and equation $m - 1$ gives $x_{m+1} = (4\lambda^{r+1} - 1)\xi(1)/3$. Now equations m and $m + 1$ give, after some algebra,

$$\xi(1) = \frac{24\lambda^{3r+2} - 6\lambda^{2r+1} - 3}{\lambda^{2r}(16\lambda^{r+1} - 1)} = \frac{6\lambda^{2r+1} + 9}{\lambda^{2r}(16\lambda^{r+1} - 4\lambda + 3)}.$$

After some cleaning up, we have the characteristic equation

$$g(\lambda) := 32\lambda^{4r+2} - 8\lambda^{3r+2} - 40\lambda^{3r+1} + 2\lambda^{2r+1} - \lambda^{2r} - 16\lambda^r + 1 = 0.$$

Rewrite this, letting $s := r - 1/3$, to get

$$g(\lambda) := 32\lambda^{4s+10/3} - 8\lambda^{3s+3} - 40\lambda^{3s+2} + 2\lambda^{2s+5/3} - \lambda^{2s+2/3} - 16\lambda^{s+1/3} + 1 = 0.$$

Let $w := 2\pi/s$ and $\lambda_0 := (1 - cw^2/s)e^{iw}$. Then some algebra using Taylor's formula quickly reveals that $\Im g(\lambda_0) = O(n^{-3})$ and

$$\begin{aligned} \Re g(\lambda_0) = & O(n^{-3}) + 32(1 - 4cw^2 - \frac{50}{3}w^2) - 8(1 - 3cw^2 - \frac{9}{2}w^2) - 40(1 - 3cw^2 - 2w^2) \\ & + 2(1 - 2cw^2 - \frac{25}{18}w^2) - (1 - 2cw^2 - \frac{2}{9}w^2) - 16(1 - cw^2 - \frac{1}{18}w^2) + 1 \end{aligned}$$

which also becomes $O(n^{-3})$ on taking $c = 571/270$. Since $g'(\lambda_0) = \Theta(n)$, this means, by [5] Theorem 4.2, that there is an eigenvalue within distance $O(n^{-4})$ of λ_0 .

Summing up, we have shown that there is an eigenvalue of the form $\lambda = (1 - \gamma)e^{i\theta}$, where $\theta = (1 + o(1))8\pi/n$ and $\gamma = (1 + o(1))\frac{18272\pi^2}{135n^3}$.

Next let

$$\phi(X_t) := \sum_i \xi(Y_t^i)$$

where, as in the previous section, Y_t^i is the position of card i at time t and the sum is over i for which $\Re \xi(i) > 0$. We want to apply Wilson's technique, so let us estimate R . Cards in positions $1, \dots, m - 2$ and $m + 1, \dots, n - 1$ move deterministically to a position whose contribution to ϕ differs by a factor λ from the previous contribution. Hence each such card contributes to a change in $|e^{-\theta}\phi(X_{t+1}) - \phi(X_t)|$ by $\gamma = O(n^{-3})$ and hence together at most $O(n^{-2})$. From the above relations between the $\xi(j)$'s it is easily seen that the remaining cards contribute to a change in $|e^{-\theta}\phi(X_{t+1}) - \phi(X_t)|$ which is limited to $O(n^{-1})$. Hence $R = O(n^{-2})$. Lemma 2.1 now gives a lower bound of $\frac{135}{36544\pi^2}n^3 \log n$.

We summarize the results of the present section in the following theorem.

Theorem 4.1 *For the equalized shuffle we have*

$$\tau_2^1 \geq \frac{135}{18272\pi^2} n^3$$

and

$$\tau_{\text{mix}} \geq \frac{135}{36544\pi^2} n^3 \log n.$$

5 Upper bounds in L^2

In this section we will utilize the comparison technique of Diaconis and Saloff-Coste [2]. Let $\{X_t\}$ and $\{Y_t\}$ be two random walks on S_n generated by the symmetric probability measures μ and ν respectively. Write E and F for the supports of μ and ν respectively. For each element $y \in F$, choose a representation $y = x_1 x_2 \dots x_k$ where $x_j \in E$ and k is odd. Write $|y| = k$. Let

$$A_* := \max_{x \in E} \frac{1}{\mu(x)} \sum_{y \in F} |y|^2 \nu(y).$$

Then it is shown in [2] that A_* is an upper bound for the ratio of the Dirichlet forms of ν and μ . For our purposes it suffices to know of the following consequence.

Lemma 5.1 $\|\mathbb{P}(X_t \in \cdot) - \pi\|_2^2 \leq n!e^{-\lfloor t/A_* \rfloor} + \|\mathbb{P}(Y_{\lfloor t/(2A_*)} \in \cdot) - \pi\|_2^2$.

Lemma 5.1 is a special case of Lemma 5 of [2].

Since the comparison technique is restricted to symmetric random walks, we also need a result from Saloff-Coste [8] (Theorem 10.2), which states that if $H := \mu(id)$ is significant, then a random walk generated by μ cannot be much slower than its symmetrized version. (Recall that the additive symmetrization of the walk generated by μ is defined as the walk generated by $(\mu + \mu^*)/2$ where $\mu^*(i) := \mu(i^{-1})$, $i \in S_n$.)

Lemma 5.2 *Let $\{X_t^s\}$ be the additive symmetrization of $\{X_t\}$. Then*

$$\|\mathbb{P}(X_t \in \cdot) - \pi\|_2^2 \leq n!e^{-Ht/2} + \|\mathbb{P}(X_{\lfloor Ht/4 \rfloor}^s \in \cdot) - \pi\|_2^2.$$

The most common benchmark walk, $\{Y_t\}$, to use for comparison is the random transpositions shuffle, i.e. the random walk generated by $\nu(id) = 1/n$ and $\nu(ij) = 2/n^2$, $1 \leq i < j \leq n$. The random transpositions shuffle is very well understood. In particular the next result, due to Diaconis and Shashahani [3], will be of use here.

Lemma 5.3 *Let $\{Y_t\}_{t=0}^\infty$ be the random transpositions shuffle. There exists a constant C such that for $t = \lfloor (1/2)n(\log n + c) \rfloor$,*

$$\|\mathbb{P}(Y_t \in \cdot) - \pi\|_2^2 \leq Ce^{-2c}.$$

We are now ready to prove the main result of this section.

Theorem 5.1 *For the original overlapping cycles shuffle with $p_m = p_n = 1/2$, m odd and $n = 2m$ and the triple shuffle with $p_m = p_{m+1} = 1/4$, $p_n = 1/2$, the lazy versions as well as the additive symmetrizations satisfy*

$$\hat{\tau} = O(n^3 \log n).$$

For the equalized shuffle, the lazy version as well as the additive symmetrization of the lazy version, satisfy

$$\hat{\tau} = O(n^5 \log n).$$

Proof. Let $\{X_t\}$ be the overlapping cycles shuffle and $\{X_t^s\}$ its additive symmetrization. Let $\{L_t\}$ be the lazy version of $\{X_t\}$, i.e. the shuffle which at each step with probability $1/2$ makes a move according to the overlapping cycles shuffle and with probability $1/2$ stays put. Let $\{L_t^s\}$ be the additive symmetrization of $\{L_t\}$ (or, equivalently, the lazy version of $\{X_t^s\}$). Let $\{Y_t\}$ be the random transpositions shuffle.

In order to use the comparison technique, we want to bound $|y|$ for any transposition $y = (i j)$, $i < j$ or $y = id$. For $y = id$, we can take the odd-length representation $y = u_m^m$. Assume now that $y = (i j)$. Write $u_n := (n \ 1 \ 2 \ \dots \ n-1)$, $u_m := (m \ 1 \ 2 \ \dots \ m-1)$, $d_n := u_n^{-1}$ and $d_m := u_m^{-1}$ denote the generators of $\{X_t^s\}$. Then the permutation taking card $n-1$ to position 1 can be written as $x := d_n^{m-1} u_m u_n^{m-1} u_m$. Note that $|x| = n$ and that $(1 \ n) = d_n x$. Hence, if $j \leq m$, we can write

$$y = v d_n x v^{-1}$$

where $v := u_m^{m-j+1} d_n u_m j - i + 1$. Note that this representation of y has odd length. Since $|v| = m + 2$, we have $|y| = 2n + 5$. If $j > m$, we add a prefix to v making the necessary moves to take both i and j to the upper half of the deck. This takes at most a prefix of length $m + 1$. Hence in general $|v| \leq 2m + 3$ and so $|y|$ is still odd and $|y| \leq 3m + 7 \leq 4n$.

Now apply Lemma 5.1. For $\{X_t^s\}$ we get $A_* = 64n^2$, so Lemma 5.1 and Lemma 5.3 give

$$\begin{aligned} \|\mathbb{P}(X_{128n^3 \log n}^s \in \cdot) - \pi\|_2^2 &\leq n! e^{-2n \log n} + \|\mathbb{P}(Y_{n \log n} \in \cdot) - \pi\|_2^2 \\ &= o(1). \end{aligned}$$

For $\{L_t^s\}$, $A_* = 128n^2$ and we get analogously

$$\|\mathbb{P}(L_{256n^3 \log n}^s \in \cdot) - \pi\|_2^2 = o(1).$$

Finally Lemma 5.2 entails

$$\|\mathbb{P}(L_{1024n^3 \log n} \in \cdot) - \pi\|_2^2 = o(1).$$

Since the same bounds on $|y|$ hold for the triple shuffle, the exact same argument goes through with only an adjustment of time by a factor 2.

Next we turn to the equalized shuffles. This is very similar to the above so we will be a bit sketchy. Consider the lazy additive symmetrization of the equalized shuffle. It is an easy exercise to show that a transposition of positions $m - 1$ and m can be made in two moves. A transposition of positions m and $m + 1$ can be brought about in four moves. Now fix two positions $i < j$. Unless $j = i + 1$, a round of n moves can always be made to bring the two cards one step closer together. From this it is easily seen that n^2 moves suffice to bring cards i and j to positions $m - 1$ and m or m and $m + 1$, where they can be transposed whereupon the moves bringing them together can be reversed. To make the number of moves odd, add an extra lazy move. Hence $|(i, j)| = O(n^2)$ and so in Lemma 5.1, $A_* = O(n^4)$. Now an argument analogous to the above gives $\hat{\tau} = O(n^5 \log n)$. Finally an application of Lemma 5.2 takes care of the non-symmetrized case. \square

6 Upper bounds in total variation

We will use the main theorem of Morris [7]. First we extract what we need from [7]. Let μ and ν be two probability measures on S . The *relative entropy* of μ with respect to ν is given by

$$\text{ENT}(\mu\|\nu) := \sum_{s \in S} \mu(s) \log \frac{\mu(s)}{\nu(s)}.$$

An equivalent expression is $\text{ENT}(\mu\|\nu) = \mathbb{E}_\mu[\log(\mu(X)/\nu(X))]$. By Jensen's inequality it follows that $\text{ENT}(\mu\|\nu) \geq 0$ with equality if and only if $\mu = \nu$.

When the measure ν is suppressed from the notation, it will be understood that ν is uniform. In this case we will simply speak of the relative entropy of μ . We note that

$$\text{ENT}(\mu) = \sum_{s \in S} \mu(s) \log(|S|\mu(s)) = \log |S| - H(\mu)$$

where $H(\mu)$ is the usual absolute entropy of μ . Relative entropy relates to total variation norm in the following way.

Lemma 6.1 *Let π be the uniform probability measure on S . Then*

$$\|\mu - \pi\|_{TV} \leq \sqrt{\frac{1}{2}\text{ENT}(\mu)}.$$

Proof. By Cauchy-Schwarz' inequality,

$$\|\mu - \pi\|_{TV} = \sum_s \frac{1}{2} \left| \mu(s) - \frac{1}{|S|} \right| \leq \sqrt{\frac{1}{4}|S| \sum_s \left| \mu(s) - \frac{1}{|S|} \right|^2}.$$

Hence it suffices to show that

$$\frac{1}{2}|S| \sum_s \left| \mu(s) - \frac{1}{|S|} \right|^2 - \sum_s \mu(s) \log(|S|\mu(s)) \leq 0.$$

This is a standard optimization problem over the $\mu(s)$'s. □

When X is a random variable, write for simplicity $\text{ENT}(X)$ for $\text{ENT}(\mathcal{L}(X))$. For two random variables X and Y , $\text{ENT}(X|Y = y)$ and the random variable $\text{ENT}(X|Y)$ then have the obvious interpretations.

Suppose $Z = (Z(1), \dots, Z(n))$ is a random vector. Some algebraic manipulation and induction leads to the well-know chain rule for entropies.

Lemma 6.2 *Let $\mathcal{F}_j = \sigma(Z(j), Z(j+1), \dots, Z(n))$. Then for any $1 \leq i \leq n$,*

$$\text{ENT}(Z) = \mathbb{E}[\text{ENT}(Z|\mathcal{F}_i)] + \sum_{k=i}^n \mathbb{E}[\text{ENT}(Z(k)|\mathcal{F}_{k+1})].$$

In particular, with $i = 1$ and $E_k := \mathbb{E}[\text{ENT}(Z(k)|\mathcal{F}_{k+1})]$, we have

$$\text{ENT}(X) = \sum_{k=1}^n E_k.$$

From now on we take $S = S_n$. For a random permutation $X = (X(1), \dots, X(n))$, the $X(k)$'s will be understood to denote the identity of the card in position k .

Let $c(i, j)$ denote the random permutation that equals *id* with probability $1/2$ and $(i j)$ with probability $1/2$. A permutation of this kind will be called a *collision*

of the positions i and j . Assume that a random permutation Y is expressed on the form

$$Y = Y_0 c(a_1, b_1) c(a_2, b_2) \dots c(a_r, b_r)$$

where the a_j 's and b_j 's are distinct and the collisions are independent given Y_0 , (but where the number of collisions and the identity of a_j and b_j may depend on Y_0). Fix a positive integer t , let Y^1, Y^2, \dots, Y^t be iid copies of Y and let

$$X_s = Y^1 Y^2 \dots Y^s,$$

$s = 1, \dots, t$. Say that the cards x and y collide at time s if there are positions i and j such that $X_s(i) = x$, $X_s(j) = y$ and Y^s contains the collision $c(i, j)$. Let $T \leq t$ be a (possibly) random time, independent of the Y^s 's and let, for each card x , $B(x)$ be the first card in the time interval $[T, t]$ that x collides with (with $B(x) = x$ if no such card exists). Let $M(x) = B(x)$ if $B(B(x)) = x$ and $M(x) = x$ otherwise. The following is the key result of [7].

Lemma 6.3 *For each $k \in [n]$, let $A_k := \min_{1 \leq i < k} \mathbb{P}(M(k) = i)$ and let Z be a random permutation independent of the Y^s 's. Then there exists a constant C independent of n, t, T, X_0 and the Y^s 's such that*

$$\text{ENT}(ZX) \leq \text{ENT}(Z) - \frac{C}{\log n} \sum_{k=1}^n k A_k E_k,$$

where the E_k 's refer to Z .

Next we apply this to the triple shuffle. Write one step, Y , of the triple shuffle on the above form by first letting Y_0 be a step of the original overlapping cycles shuffle $p_m = p_n = 1/2$, and then

$$Y := \begin{cases} Y_0 c(1, m+1), & Y_0 = (1 \ 2 \ \dots \ m) \\ Y_0, & Y_0 = (1 \ 2 \ \dots \ n) \end{cases}$$

Lemma 6.4 *Consider the triple shuffle. Fix $k \in [n]$ and let $l := 2^{\lceil \log_2 k \rceil}$. Let $t = 3n^2$ and $T = 2(l \vee n^{1/2})^2$. Then there exists a constant a independent of n, k and i such that*

$$\mathbb{P}(M(k) = i) \geq \frac{a}{k \vee n^{1/2}}$$

for all $i < k$.

Proof. Note that $k/2 \leq l \leq k$ so in particular k and l are of the same order. Assume first that $k < n^{1/2}$. Fix $i < k$. Up until time $m - k$, cards i and k move deterministically one step down the deck for each step of the shuffle. Then the steps $m - k, m - k + 1$ and $m - i$ may move card k to the bottom half of the deck and card i to the top half, this happens with probability at least $9/16 \cdot 1/4$ (with equality unless $k = i + 1$). Given this, the steps $m - k + 2, \dots, m - i - 1$ and the $m - 1$ steps after step $m - i$, change the position of card k relative to card i by a $\text{binomial}(m + k - i - 2, 1/2)$ random variable and if step $2m - i$ again moves card i to the top (which happens with probability $1/4$ independently of everything else) then this is repeated independently for another $m - 1$ steps unless card k hits position n earlier. However, by the local CLT, there is a conditional probability of order $1/n^{1/2}$ that card k hits position n at exactly the same time as card i hits position m . Given that this happens, then there is a conditional probability at least $1/2 \cdot 1/4 \cdot 1/2 \cdot 1/2$ that the next two moves bring i and k to positions 1 and 2 and that $\{M(k) = i\}$ occurs after another m steps.

The case $k \geq n^{1/2}$ is identical apart from that one has to note that cards i and k will with probability $1/2$ go on different halves of the deck at least half of the time up to time T . Since this event is independent of the relative motion of cards i and k , we may condition on it and use the local CLT to see that $\mathbb{P}(M(k) = i)$ is of order $1/k^2$. \square

Theorem 6.1 *The triple shuffle has a mixing time of $O(n^{5/2} \log^3 n)$.*

Proof. Let $X(t)$ be t steps of the triple shuffle and set $t := 3n^2$. We want to apply Lemma 6.3. By the chain rule we have

$$\text{ENT}(Z) = \sum_{k=1}^n E_k.$$

Partition the indexes k into blocks $I_0, I_1, \dots, I_{\lceil \log_2 n \rceil}$ where $I_j := [n] \cap \{2^j, \dots, 2^{j+1} - 1\}$. Then, since there are no more than $\log_2 n + 2 < 2 \log n$ blocks,

$$\text{ENT}(Z) = \sum_{k=1}^n E_k \leq 2 \log n \sum_{k \in I_{j^*}} E_k$$

where j^* is the index j that maximizes the sum $\sum_{k \in I_j} E_k$. By Lemma 6.4 applied to $k \in I_{j^*}$ and T as in the lemma,

$$\sum_{k \in I_{j^*}} k A_k E_k \geq \frac{a}{n^{1/2}} \sum_{k \in I_{j^*}} E_k.$$

Taken together with Lemma 6.3 with the same T , the last two observations give

$$\text{ENT}(ZX_t) \leq \left(1 - \frac{C}{n^{1/2} \log^2 n}\right) \text{ENT}(Z).$$

Using this inductively yields for $r = 1, 2, \dots$,

$$\text{ENT}(ZX_{rt}) \leq \left(1 - \frac{C}{n^{1/2} \log^2 n}\right)^r \text{ENT}(Z).$$

Taking $Z = id$, noting that $\text{ENT}(id) = \log(n!) < n \log n$, and $r = 2n^{1/2} \log^3 n$ then gives

$$\text{ENT}(ZX_{6n^{5/2} \log^3 n}) \leq \frac{C \log n}{n}.$$

The result now follows from Lemma 6.1. \square

I strongly conjecture that the true mixing time for the triple shuffle does not deviate from n^2 by more than a poly-log n -factor. However the entropy technique can for this shuffle not rule out that the order of the top $\Theta(n^{1/2})$ cards after this many shuffles, is heavily dependent on the other cards. However, if we instead turn to the additive symmetrization of the triple shuffle, then this problem vanishes.

Lemma 6.5 *Fix $k \in [n]$, let $l := 2^{\lceil \log_2 k \rceil}$, $T = 10l^2$ and $t = 100n^2$. Then there is a constant a , independent of n , k and i such that*

$$\mathbb{P}(M(k) = i) \geq \frac{a}{k}$$

for all $i < k$.

Proof. The intuition behind this is simple: with high probability cards k and i will spend a significant proportion of the time up to T on different halves of the deck. During this time they will diffuse with respect to each other a distance which is typically of order k . This will therefore cancel their starting distance with probability of order $1/k$. Doing this properly however, takes some work.

Fix $i < k$. Let $E := \{1, m, m+1, 2m\}$. Let F_s be the difference of the number of bottom-to-top moves and the number of top-to-bottom moves up to time s . Let L_s be the difference of the number of moves taking the bottom card to the top position and the number of moves taking the top card to the bottom position and let $D_s := F_s - L_s$.

Let U_s and V_s be the positions at time s of cards k and i respectively. Regard $\{U_s\}$ and $\{V_s\}$ as random walks on \mathbb{Z}_m , where positions in E are identified with $0 \in \mathbb{Z}_m$ and $j + 1$ and $m + 1 + j$ are identified with $j \in \mathbb{Z}_m$. When card k is in $\{2, \dots, m - 1\}$, U_s behaves like ordinary SRW and when card k is in $\{m + 2, \dots, 2m - 1\}$, it behaves like SRW with holding probability of $1/2$; indeed in the former case U_s moves according to F_s and in the latter it moves along with L_s . When at 0 , i.e. when card k is in E , the random walk behaves differently. Let B be the event that none of the walks spends more than time $C\sqrt{T}$ up to time T at 0 . Then $\mathbb{P}(B)$ is at least, say, $9/10$, once C is properly chosen.

Whenever card k enters E , it will exit E in the upper half of the deck with a probability depending on where it hits E . This probability is readily computed to be $192/253$ from 0 , $200/253$ from m , $12/23$ from $m + 1$ and $96/253$ from $2m$. The particular numbers are not so important, we just note that they are all in the interval $[1/5, 4/5]$. In short, each time U_s hits 0 , it may change state, from going along with F_s to following L_s or vice versa, and a change takes place with probability at least $1/5$. Of course, all this goes for V_t too.

Taken together, U_t and V_t make the exact same moves, except when the cards k and i are on different halves of the deck, in which case the walk corresponding to the card in the lower half, only makes the move with probability $1/2$ and instead holds with probability $1/2$. We also note that if cards i and k are in E at the same time, the conditional probability that i goes to the upper/lower half of the deck given that k goes to the upper/lower half is at least $1/5$.

Now consider for a while conditioning on the set of points in time at which one of the cards k and i hit E and the position in which this happens. For each such time point, s_j , let A_j be the event that the card in question exits on a different half than the other card and that an independent coin flip results in heads. We let this coin be biased in such a way that A_j happens with probability $1/5$. Let $A(p)$ be the event that the proportion of time off 0 between s_1 and some stopping time, that A_j occurred for the latest s_j , is at least p . Then, regardless of the structure of the set of s_j 's, the conditional probability of $A(1/5)$ is at least $1/5$. Let A be $A(1/5)$ intersected with an extra independent coin flip biased in such a way that the conditional probability of A is exactly $1/5$.

The point of the extra coin flips is that A carries no information on the structure of the s_j 's, and since F_s and L_s are independent of which half i and k are on, conditioning on A leaves no information on $\{D_s\}$.

Now $\mathbb{P}(s_1 < T/3) = 1 - o(1)$. On $\{s_1 < T/3\} \cap A \cap B$, which has probability at least $(1 - o(1))1/10$, let X be the sum of the D_s 's for the first $T/15$ time points s for which U_s and V_s are in different states. Note that $s \leq 2T/3 + C\sqrt{T}$ for

all s counted in X . Let Y be the distance, modulo m in the above sense, at time T between cards k and i caused by $k - i$, the rest of the D_s 's and moves when one of the cards is in E . Since $\mathbb{P}(|Y| \leq 10k^2)$ is bounded away from 0 and X is independent of Y , A , B and $\{s_1 > T/3\}$, the local CLT implies that there exists a constant $b > 0$ such that $\mathbb{P}(Y - X = 1) \geq b/k$. Also, given this, there is clearly a probability bounded away from 0 that either i or k hit 0 in the time interval $[2T/3]$ and at the last time before T this happened, i and k both went to the the upper half of the deck. If this happens, then we note that at time T , cards k and i are next to each other on the upper half of the deck with k on top of i .

Finally, given all this, there is a probability at least $1/16$ that $m(k) = i$; this happens e.g. if the two first moves following the first time after T that card i hits position m are favorable. This completes the proof. \square

Theorem 6.2 *The additive symmetrization of the triple shuffle has mixing time $O(n^2 \log^3 n)$.*

Proof. Copying the proof of Theorem 6.1 word for word, but using $T = 100n^2$ and Lemma 6.5 instead of Lemma 6.4 shows that mixing time is bounded by $200n^2 \log^3 n$. \square

Some further small adjustments also lead to the following upper bound for the equalized shuffle, only a factor $\log^2 n$ off from the lower bound in Section 3.

Theorem 6.3 *The mixing time of the equalized shuffle with $n = 2m$ is $O(n^3 \log^3 n)$.*

Proof. An analogous result to Lemma 6.5 goes through, but with $t = 3n^3$ and $T = 2nl^2$. The proof is a copy of the proof of Lemma 6.4 with the difference (and simplification) that cards i and k move non-deterministically with respect to each other only when one of them is at position m , $m + 1$ or n .

Mimicking the proof of Theorem 6.1 once again, now gives a mixing time $O(n^3 \log^3 n)$. \square

References

- [1] Angel O., Peres, Y. and Wilson, D. B.; Card shuffling and diophantine approximation, *Ann. Appl. Probab.* **18** (2008), 1215-1231.

- [2] Diaconis, P. and Saloff-Coste, L.; Comparison techniques for random walk on finite groups, *Ann. Probab.* **21** (1993), 2131-2156.
- [3] Diaconis, P. and Shashahani, M.; Generating a random permutation with random transpositions, *Z. Wahrsch. Verw. Gebiete* **57**, 159-179.
- [4] Durrett, R.; Shuffling chromosomes, *J. Theoret. Probab.* **16** (2003), 725–750
- [5] Jonasson, J.; Biased random-to-top shuffling, *Ann. Appl. Probab.* **16** (2006), 1034-1058.
- [6] Jonasson, J.; The overhand shuffle mixes in $\Theta(n^2 \log n)$ steps, *Ann. Appl. Probab.* **16** (2006), 231-243.
- [7] Morris, B.; Improved mixing time bounds for the Thorp shuffle and L -reveral chain, *Ann. Probab.* **37** (2009), 453–477.
- [8] Saloff-Coste, L.; Random walks on finite groups, *Encyclopaedia Math. Sci.* 110, 263-346, Springer Berlin 2004, Harry Kesten editor.
- [9] Wilson, D. B.; Mixing times of lozenge tiling and card shuffling Markov chains, *Appl. Probab.* **14** (2004), 274-325.
- [10] Wilson, D. B.; Mixing time of the Rudvalis shuffle, *Electron. Commun. Probab.* **8** (2003), 77-85.