Rapid mixing of dealer shuffles and clumpy shuffles

Johan Jonasson[*†] and Ben Morris[‡]

March 11, 2014

Abstract

A famous result of Bayer and Diaconis [2] is that the Gilbert-Shannon-Reeds (GSR) model for the riffle shuffle of n cards mixes in $\frac{3}{2}\log_2 n$ steps and that for 52 cards about 7 shuffles suffices to mix the deck. In this paper, we study variants of the GSR shuffle that have been proposed to model more realistically how people actually shuffle a deck of cards. The *clumpy riffle shuffle* and *dealer riffle shuffle* differ from the GSR model in that when a card is dropped from one hand, the conditional probability that the next card is dropped from the same hand is higher/lower than for the GSR model. It is believed that these shuffles mix slightly slower than the GSR shuffle, but still in order $\log n$ steps. However, rigorous results have so far been missing. In this paper we apply the technique of relative entropy and collisions of Morris [5], to show that the clumpy shuffle and the dealer shuffle mix in $\mathrm{O}(\log^4 n)$ steps.

AMS Subject classification : 60J10
Key words and phrases: riffle shuffle, entropy technique, collisions
Short title: Dealer shuffles and clumpy shuffles

1 Introduction

Mixing times for Markov chains is a subject great importance, both from a theoretical point of view and because of its applicability, and has attracted much attention

[*]Chalmers University of Technology and University of Gothenburg
[†]Research supported by the Knut and Alice Wallenberg Foundation
[‡]University of California, Davis. Supported by NSF grants DMS-1007739 and CNS-1228828

over the last decades. A very prominent subclass of mixing time problems is card shuffling, that is, Markov chains on the symmetric group $\mathbb{S}_n$ of permutations of n items that one can think of as the cards of a deck. Perhaps the most famous of card shuffles is the Gilbert-Shannon-Reads (GSR) model for the riffle shuffle for which Bayer and Diaconis [2] proved a remarkably exact result; there is a sharp cutoff at $\frac{3}{2}\log_2 n$ shuffles after which the deck is well mixed and for a standard deck of 52 cards, about 7 shuffles suffices for mixing. Prior to that, Aldous and Diaconis [1] had proved, via a striking strong uniform time argument, that $2\log_2 n$ shuffles is an upper bound on the mixing time.

The riffle shuffle is, together with the inefficient overhand shuffle which mixes in order $n^2 \log n$ steps (see [7] and [4]), the most common way in which people actually shuffle a deck of cards. The model for one step of the GSR shuffle is the following. First the deck is cut into two packets of which one goes into your right hand and the other into your left hand. The number of cards that go into your right (or left if you like) hand is a binomial random variable with parameters n and $1/2$. Then the cards are dropped from the two hands in such a way that whenever there are A cards left in your right hand and B cards left in your left hand, the probability that the next card is dropped from your right hand is $A/(A+B)$.

An equivalent description of the GSR shuffle is as follows. At each step

1. generate a uniform random binary sequence of length n;
2. if the binary sequence has k zeros and $n-k$ ones, cut the deck so that the left pile has k cards and the right pile has $n-k$ cards, and then interleave the two piles by reading the binary sequence from left to right, and dropping from the left pile with each zero and from the right pile with each one.

For example, if $n = 6$ and the binary sequence is 001110, then we first cut the deck into two equal piles, then interleave the piles by dropping the first two cards from the left pile, the next three cards from the right pile, and the last card from the left pile again.

Note that according to the GSR model, when you drop from your right hand, you drop a single card with probability $1/2$, a pair of cards with probability $1/4$, a triple of cards with probability $1/8$, and so on. However, if one analyzes riffle shuffles of a fresh deck of cards in practice, one finds that the shuffles are finer. Cards tend to be dropped in a more alternating fashion, especially with experienced dealers; see Remark (e) and open problem (i) of [1]. Such shuffles are named *dealer riffle shuffles* in [3] and we stick with this term. On the other hand, when the deck has been used for a long time and become sticky, the opposite tends

to occur, namely that cards are dropped in clumps. Hence we call these shuffles *clumpy riffle shuffles*.

A model that includes both the dealer and clumpy shuffles as special cases is the *Markovian model*, which appears in the "open problems" section of [3]. The Markovian model is driven by a two-state Markov chain with transition matrix

$$\begin{bmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{bmatrix}$$

and the transition rule is as follows. At each step

1. run n steps of the two-state Markov chain in stationarity to generate a binary sequence of length n;

2. if the binary sequence has k zeros and $n-k$ ones, cut the deck so that the left pile has k cards and the right pile has $n-k$ cards, and then interleave the two piles by reading the binary sequence from left to right, and dropping from the left pile with each zero and from the right pile with each one.

Note that the Markovian model includes the GSR model as a special case. It is natural to assume a symmetric cut (that is, $p_{01} = p_{10}$, so that the left and right piles have the same expected size) and we shall do this in the present paper. For $p \in (0,1)$ consider the two-state Markov chain with transition matrix

$$K_p := \begin{bmatrix} p & 1-p \\ 1-p & p \end{bmatrix}.$$

We shall call this Markov chain the *two-state chain with parameter* p (or simply the *two-state chain*) and we define the *p-riffle shuffle* as the shuffle driven by this chain. When $p < \frac{1}{2}$ we call the shuffle *dealer* and when $p > \frac{1}{2}$ we call the shuffle *clumpy*. It is widely believed that clumpy riffle shuffles and dealer riffle shuffles may mix more slowly than the GSR shuffle, but still at order $\log n$, but rigorous results have been missing so far. (However for some extremely clumpy shuffles where p is allowed to grow with n a few facts are known; see [9].) In this paper we give the first rigorous results for p-riffle shuffles, showing that $\mathrm{O}(\log^4 n)$ shuffles suffices to mix the deck.

2 The time reversed shuffle and mixing time

Recall that the mixing time of an (aperiodic irreducible) Markov chain is defined in terms of the total variation distance between the distribution at a given time and

the stationary distribution: if X_t is the state of the Markov chain at time t and π is the stationary distribution, then the total variation distance is given by

$$\begin{aligned} \|\mathbb{P}(X_t \in \cdot) - \pi\|_{TV} &:= \max_{A \subseteq S} (\mathbb{P}(X_t \in A) - \pi(A)) \\ &= \frac{1}{2} \sum_{s \in S} |\mathbb{P}(X_t = s) - \pi(s)|, \end{aligned}$$

where S is the state space and $\mathbb{P}$ is the underlying probability measure. The mixing time is then defined by

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

As with the GSR shuffle before it, it turns out that the analysis of the p-riffle shuffle is more conveniently carried out for the time reversed shuffle. Since the GSR shuffle and p-riffle shuffle are random walks on groups (see [8]) each has the same mixing time as its time reversal.

For the GSR shuffle, the time reversal can be described as follows. First give each card an independent 0 or 1 mark, each with probability $1/2$. Then put all cards marked 0 above the cards marked 1, without changing the internal order among cards with the same mark. If we repeat this process and keep track of all the markings that have been given to each card, then after k shuffles each card has an independent iid sequence of $0/1$ marks of length k. A moment's thought reveals that the first time, τ, when all the cards have distinct mark sequences is a strong uniform time, i.e., X_τ is uniformly distributed and independent of τ. Since τ is highly concentrated around $2 \log_n n$, this implies a $\mathrm{O}(\log n)$ upper bound for the mixing time. This argument, which first appeared in [1], relies heavily on the independence between the marks for different cards. The same goes for the more detailed analysis in [2].

For the p-riffle shuffle, the time reversal has the following transition rule. First, generate marks by running n steps of the two state Markov chain in stationarity. That is, the first card is given a mark according to a fair coin flip, and subsequent cards are given the same mark as the previous card with probability p and the opposite mark with probability $1-p$. Then put all cards marked 0 above the cards marked 1, without changing the internal order among cards with the same mark.

Our main result is:

Theorem 2.1 *Fix $p \in (0,1)$. The mixing time τ_{mix} for the p-riffle shuffle satisfies*

$$\tau_{\text{mix}} = O(\log^4 n).$$

Remark. Other models for finer riffle shuffles have been proposed. The most prominent is perhaps the Thorp shuffle, for which the best known upper bound to date is of order $\log^4 n$ and due to Morris [5]. In the special case $n = 2^d$, there is an upper bound of $O(\log^3 n)$, also due to Morris [6]. Both of these papers rely on the same entropy technique from [5] as we do here.

3 The proof

The proof on Theorem 2.1 relies on the entropy technique introduced in [5], so let us first review the parts needed. For two probability measures ν and π on a finite space S, the relative entropy of ν with respect to π is given by

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

Here we will only be concerned with the case when π is uniform. In that case one just speaks of the relative entropy of ν and drops π from the notation, so that

$$\mathrm{ENT}(\nu) = \sum_{s\in S} \nu(s) \log(|S|\nu(s)).$$

For a random variable X, we write $\mathrm{ENT}(X)$ for $\mathrm{ENT}(\mathcal{L}(X))$, where $\mathcal{L}(X)$ is the law of X. The notation $\mathrm{ENT}(X|Y = y)$ then of course stands for the entropy of the conditional law of X given $Y = y$ and $\mathrm{ENT}(X|Y)$ is the random variable that equals $\mathrm{ENT}(X|Y = y)$ when $Y = y$. The following lemma relates relative entropy to total variation. It can be proved by using Schwarz inequality and solving a standard optimization problem.

Lemma 3.1 *Let π be the uniform measure on S. Then*

$$\|\nu - \pi\|_{TV} \le \sqrt{\frac{1}{2}\mathrm{ENT}(\nu)}.$$

Next, recall the chain rule for relative entropies:

$$\mathrm{ENT}(X, Y) = \mathrm{ENT}(X) + \mathbb{E}[\mathrm{ENT}(X, Y|X)],$$

which generalizes to

$$\begin{aligned} \mathrm{ENT}(X_1, X_2, \ldots, X_n) &= \mathbb{E}[\mathrm{ENT}(X_1, \ldots, X_n|X_i, \ldots, X_n)] \\ &+ \sum_{k=i}^{n} \mathbb{E}[\mathrm{ENT}(X_k|X_{k+1}, \ldots, X_n)] \end{aligned}$$

for each $i \in [n]$. Note that the last term in the sum is just $\mathrm{ENT}(X_n)$. We will be concerned with the case when X is a random permutation of n cards. We will write $X(j)$ for the position of card j (i.e. the card that started in position j) after applying X. Consequently $X^{-1}(j)$ is the initial position of the card in position j after applying X. Writing $E_j := \mathbb{E}[\mathrm{ENT}(X^{-1}(j)|\mathcal{F}_{j+1})]$, where $\mathcal{F}_j := \sigma(X^{-1}(j), X^{-1}(j+1), \ldots, X^{-1}(n))$, the chain rule takes on the form

$$\mathrm{ENT}(X) = E[\mathrm{ENT}(X|\mathcal{F}_i)] + \sum_{k=i}^{n} E_k.$$

In particular

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

The key result of [5] states that applying random permutations that involve *collisions* decreases relative entropy by a certain factor. For $a, b \in [n]$, we write $c(a, b)$ for the random permutation that equals id with probability $1/2$ and (a, b) with probability $1/2$ and refer to this random permutation as a collision of positions a and b. For permutations X and Y we write XY for $Y \circ X$. Let Y be a random permutation that can be written as

$$Y = c(a_1, b_1)c(a_2, b_2) \ldots c(a_k, b_k)Z$$

where Z is a random or fixed permutation, the a_i's and b_i's all distinct and the $c(a_i, b_i)$:s mutually independent given Z. (However, the identities of the a_i:s and b_i:s and the number of collisions typically depend on Z.) Let $Y_1, Y_2, \ldots$ be independent copies of Y and write $Y_{(t)} = Y_1 Y_2 \ldots Y_t$, $t = 1, 2, \ldots$. We say that the cards x and y collide at time t if there are two positions i and j, such that $Y_{(t-1)}^{-1}(i) = x$, $Y_{(t-1)}^{-1}(j) = y$ and Y_t contains the collision $c(i, j)$. Fix t and let $T \in [t]$ be a random time independent of the Y_i:s. For a given card x, let $b(x) = y$ if y is the first card that x collides with in $[T, t]$. If also $b(y) = x$, then let $m(x) = y$ (in which case we will also have $m(y) = x$). Otherwise set $m(x) = x$. For the present paper it suffices to note that if x and y collide at time T then $m(x) = y$.

For each x, let

$$A_x = \max\{c : \forall y < x : \mathbb{P}(m(x) = y) \geq c/x\}.$$

Theorem 3.2 *([5]) Let X be a random permutation independent of $Y_1, \ldots, Y_t$. Then*

$$\mathrm{ENT}(X) - \mathrm{ENT}(XY_{(t)}) \geq \frac{C}{\log n} \sum_{k=1}^{n} A_k E_k$$

where C is a universal constant.

We will actually use Theorem 3.2 to analyze the time reversed p-riffle shuffle. Recall that the time reversal has the following transition rule. First, generate marks by running n steps of the two state Markov chain in stationarity. Then put all cards marked 0 above the cards marked 1, without changing the internal order among cards with the same mark.

Fix two cards x and y with $x < y$. Note that if x and y are given the same marks then their distance will typically decrease by a factor of roughly one half after the shuffle. Suppose we associate to each card from x to y a 1 (respectively, 0) if the card is given the same mark as card x, generating a binary sequence of length $y - x + 1$. Call this the *agreement sequence*. If x and y are given the same mark, then we continue and define the agreement sequence for the next step, and so on.

Call a binary sequence *successful* if it ends in a one, and if V is a binary sequence, let $|V|$ denote the Hamming weight of V (that is, the number of ones in V).

The following Markov chain, which we call the *thinning process*, models the process of agreement sequences up until the time when x and y get a different mark. The state space is the set of binary sequences, and the transition rule is as follows. If the current state $V_k = V$, the next state V_{k+1} is defined as follows.

1. if V is not successful, then $V_{k+1} = V$; else
2. let V_{k+1} be the binary sequence of length $|V|$ generated by running the two-state chain starting from a one for $|V| - 1$ steps.

Note that the unsuccessful states are absorbing. For $t \geq 1$ let A_t be the event that V_t is successful.

Lemma 3.3 *Let $V_0, V_1, \ldots$ be the thinning process and for $t \geq 0$ define $L_t = |V_t|$. There exists a universal constant $\gamma > 0$ and positive integers $\widetilde{l}$ and C, which depend only on p, such that if $L_0 = l_0$ and $t = \lfloor \log_2 l_0 - \widetilde{l} \rfloor$ then*

$$\mathbf{P}(A_t \cap \{0 < L_t < C\}) \geq \frac{\gamma}{l_0} \ .$$

Proof. First we give an alternate construction of the thinning process. Note that the trajectory of the two-state chain, starting from a 1, can be generated as follows.

In the dealer case (respectively, clumpy case), start with a sequence of the form $10101\ldots$ (respectively, $111\ldots$) whose length is a geometric random variable of parameter $|1-2p|$. Then in the next step, flip a fair coin to generate the next state and continue with the usual transition rule after that. (Note that from this construction it is clear that for all m the expected number of ones among the first m states is at least $m/2$.) Let T be time when the fair coin is used to generate the next state. We shall call this time the *forget time*.

Suppose that the current state of the thinning process is V, where V is successful and define $L = |V|$. Let $Z_0, Z_1, \ldots$ be the two-state Markov chain constructed using the alternate method described above. We write Z for the sequence $(Z_0, \ldots, Z_{L-1})$. Let $\widetilde{Z}$ be the sequence obtained from Z by reversing every state from time T onward, with $\widetilde{Z} = Z$ if $T > L$. Note that $\widetilde{Z}$ has the same distribution as Z. Define the sequence W by

$$W := \begin{cases} Z & \text{if } Z_L = 1; \\ \widetilde{Z} & \text{otherwise.} \end{cases}$$

Note that $W_L = 1$ unless $T > L$. Let $\widetilde{W}$ be the sequence obtained from W by reversing every state from time T onward,

Now flip a fair coin, which we shall call the *deciding coin.* The next state V' of the thinning process is

$$V' = \begin{cases} W & \text{if the deciding coin lands heads;} \\ \widetilde{W} & \text{if the deciding coin lands tails.} \end{cases}$$

Note that V' is successful whenever the deciding coin lands heads, unless the forget time T is greater than L. (Roughly speaking, the deciding coin "decides" whether the next state will be successful or not.) We call W the *good sequence* in the construction of V' from V.

The main idea of the proof is to use the second moment method to show that, under the assumptions of the Lemma, if we condition on the event that the deciding coin repeatedly lands heads (that is, the good sequence W is chosen repeatedly instead of $\widetilde{W}$) then with probability bounded away from zero we have $0 < L_t < C$.

Fix a state V of the thinning process, let $L = |V|$, and let W be the good sequence in the contruction of the next state V' from V. The key step of the proof is to bound the mean and second moment of $S := |W|$. We claim that

$$\mathbf{E}(S) \geq L/2, \tag{1}$$

and

$$\mathbf{E}(S^2) \leq \frac{L^2}{4} + cL, \tag{2}$$

for a constant c that depends only on p.

First, we verify (1). Since

(1) the sequence $W_0, \ldots, W_{T-1}$ has at least as many ones as zeros;

(2) given $T = k$ where $k \leq L$, the value of $\sum_{i=k}^{L-1} W_i$ has the same distribution as the number of ones in the first $L - k$ states of the two-state chain starting from 1;

equation (1) follows. Next we verify the (2). Note that

$$\mathbf{E}(S^2) \leq \sum_{i=1}^{L} \mathbf{E}(W_i) + 2 \sum_{0 \leq i < j \leq L} \Big(\mathbf{E}(W_i W_j; T \leq i) + \mathbf{P}(T > i) \Big). \tag{3}$$

The first sum can be trivially bounded above by L. For the second sum, note that if $T \leq i$ then $W_i = W_j = 1$ only if $Z_i = Z_j = Z_L$, which occurs with probability

$$\Big[\frac{1}{2} + \frac{1}{2}(p-q)^{j-i}\Big]\Big[\frac{1}{2} + \frac{1}{2}(p-q)^{L-j}\Big],$$

where $q = 1 - p$. (Recall that the probability that a coin of bias q has an even number of heads after m flips is $\frac{1}{2} + \frac{1}{2}(p-q)^m$.) Combining this with the fact that $\mathbf{P}(T > i) = |1 - 2p|^{i-1}$ shows that the terms of the second sum in (3) are at most

$$\frac{1}{4}\Big(1 + (p-q)^{j-i} + (p-q)^{L-j} + (p-q)^{L-i} + (p-q)^{i}\Big) + |1-2p|^{i-1}. \tag{4}$$

Summing this over i and j with $0 \leq i < j \leq L$ gives at most $\frac{L^2}{4} + c'L$, for a constant c' that depends only on p. This verifies (2).

Now let $V_0, V_1, \ldots$ be a thinning process constructed using deciding coins and let E be the event that the deciding coin lands heads for each step up to time t. We write $\widehat{\mathbf{P}}$ and $\widehat{\mathbf{E}}$ for the conditional probability and expectation, respectively, given E. For $k \geq 0$ define $Y_k = L_k^2$. If we define $f : [0, \infty) \mapsto \mathbf{R}$ by $f(x) = \frac{x}{4} + c\sqrt{x}$, then (2) implies that

$$\widehat{\mathbf{E}}(Y_{k+1} \,|\, Y_k = y) \leq f(y). \tag{5}$$

Hence, induction and the fact that f is concave imply that

$$\widehat{\mathbf{E}}(Y_k) \leq f^k(l_0^2), \tag{6}$$

where f^k is the kth iterate of f. Another straightforward calculation and induction imply that

$$f^k(x) \leq h\Big(\frac{x}{4^k}\Big), \tag{7}$$

where $h(x) = x + B\sqrt{x}$ for a sufficiently large constant $B \geq c$ (e.g. $B = 3c^2$ suffices), provided that $x/4^k \geq 1$. It follows that

$$\begin{aligned} \widehat{\mathbf{E}}(L_k^2) &\leq f^k(l_0^2) &(8)\\ &\leq h\Big(\frac{l_0^2}{4^k}\Big), &(9)\end{aligned}$$

since $l_0/2^k \geq 1$ as $k \leq t < \log_2 l_0$. Finally, note that combining (1) with induction gives

$$\widehat{\mathbf{E}}(L_k) \geq \frac{l_0}{2^k}. \tag{10}$$

Combining this with (9) and the definition of h gives

$$\begin{aligned} \widehat{\mathbf{Var}}(L_k) &= \widehat{\mathbf{E}}(L_k^2) - \widehat{\mathbf{E}}(L_k)^2 &(11)\\ &\leq B\widehat{\mathbf{E}}(L_k). &(12)\end{aligned}$$

Let T_k be the forget time in the construction of V_{k+1} from V_k. Recall that on the event E, the step is successful unless $T_k > L_k$. Hence, on E, the step is unsuccessful only if B_k occurs, where

$$B_k = \Big\{T_k > \frac{a_k}{2}\Big\} \cup \Big\{L_k \leq \frac{a_k}{2}\Big\}, \tag{13}$$

where we write a_k for $\frac{l_0}{2^k}$. Combining this with the fact that $L_t = 0$ only if B_t occurs, we get that $\widehat{\mathbf{P}}(A_t^c \cup \{L_t = 0\})$ is at most

$$\sum_{k=0}^{t}\Big[\widehat{\mathbf{P}}(T_k > \frac{a_k}{2}) + \widehat{\mathbf{P}}(L_k \leq \frac{a_k}{2})\Big]. \tag{14}$$

Since T_k is a geometric random variable with parameter $\alpha := |1 - 2p|$, we have

$$\begin{aligned} \widehat{\mathbf{P}}(T_k > \frac{a_k}{2}) &\leq \alpha^{a_k/2} &(15)\\ &\leq D/a_k, &(16)\end{aligned}$$

for a constant D. Furthermore, by (10) and Chebyshev's inequality, we have

$$\begin{aligned} \widehat{\mathbf{P}}(L_k \leq \frac{a_k}{2}) &\leq \frac{4\widehat{\mathbf{V}}\mathbf{ar}(L_k)}{a_k^2} && (17) \\ &\leq 4B/a_k, && (18) \end{aligned}$$

by (12). Thus, the quantity (14) is at most

$$\sum_{k=0}^{t} \frac{D+4B}{a_k}. \tag{19}$$

Recall that $t = \lfloor \log_2 l_0 - \widetilde{l} \rfloor$. Thus if $\widetilde{l}$ is large enough so that

$$2^{\widetilde{l}} > 4(D+4B)$$

then by (10) we have $a_{t-k} > 4(D+4B)2^{-k}$ for all k with $0 \leq k \leq t$, and hence the quantity (19) is at most $\frac{1}{2}$. Hence

$$\widehat{\mathbf{P}}(A_t^c \cup \{L_t = 0\}) \leq \frac{1}{2} \tag{20}$$

Finally, note that (9) implies that $\widehat{\mathbf{E}}(L_t^2) \leq \beta$ for a constant β that depends only on p. Choosing $C > 2\beta^{1/2}$ gives

$$\begin{aligned} \widehat{\mathbf{P}}(L_t \geq C) &= \widehat{\mathbf{P}}(L_t^2 \geq C^2) && (21) \\ &\leq \frac{1}{4} && (22) \end{aligned}$$

by Markov's inequality. Combining this with (20) gives

$$\widehat{\mathbf{P}}\Big(A_t^c \cup \{L_t = 0\} \cup \{L_t \geq C\}\Big) \leq \frac{3}{4}$$

and hence the unconditional probability

$$\begin{aligned} \mathbf{P}(A_t, 0 < L_t < C) &\geq \frac{1}{4}\Big(\frac{1}{2}\Big)^t && (23) \\ &\geq \frac{\gamma}{l_0}, && (24) \end{aligned}$$

for a universal constant γ. □

In order to apply Theorem 3.2 to the reverse p-riffle shuffle we need to generate a step of the shuffle using collisions, and for this we need the following key fact. For binary sequences $M = (M_1, \ldots, M_n)$, let

$$p(M) = \tfrac{1}{2} K_p(M_1, M_2) K_p(M_2, M_3) \cdots K_p(M_{n-1}, M_n)$$

be the probability of generating M as a trajectory of the two-state chain. If we divide M into $\lfloor \frac{M}{4} \rfloor$ blocks of length 4, plus possibly one additional smaller block, then reversing any block of the form $ab(1-b)a$ (e.g., 1011) does not change $p(M)$. Furthermore, the effect of such a change in markings is to interchange the final positions of the middle two cards in the reversed block.

Let M be the random binary sequence generated for a step of the shuffle. We say that positions j and $j+1$ *interact* if

1. j is congruent to 2 modulo 4,
2. $M_j \neq M_{j+1}$,
3. $M_{j-1} = M_{j+2}$.

Let $\mathcal{C} = \{j : j \text{ interacts with } j+1\}$. Note that if Z is the permutation generated from M, then the permutation Y defined by

$$Y := \Big[\prod_{j \in \mathcal{C}} c(j, j+1)\Big] Z$$

has the same distribution as Z, so we can define a step of the shuffle to be the permutation Y.

Now partition the positions in the deck as

$$I_l = \{2^{l-1}, 2^{l-1}+1, \ldots, 2^l - 1\} \cap [n],$$

$l = 1, 2, \ldots, \lceil \log_2(n+1) \rceil$. For each l, let $T = T_l$ be the random time for which $\mathbb{P}(T = 1) = 2^{-l+1}$ and $\mathbb{P}(T = l+1-r) = 2^{-r}$, $r = 1, \ldots, l-1$, so that $l + 1 - T$ is a truncated geometric$(1/2)$ random variable. Now let $t = \lceil \log_2 n \rceil$ and let $Y_1, Y_2, \ldots$ be independent copies of Y. The following lemma ensures that we can apply Theorem 3.2.

Lemma 3.4 *In the above notation, with l fixed and $T = T_l$, there is a constant c independent of l and n such that*

$$\mathbb{P}(m(x) = y) \geq \frac{c}{x}$$

for all $x \in I_l$ and all $y < x$.

Proof. Suppose $x \in I_l$ and $y < x$ and define $d = x - y + 1$. It suffices to find a lower bound for the probability that x and y collide at time T, since this implies that $m(x) = y$. For $k = 0, 1, \ldots$, let S_k be the set of cards in the set consisting of y, x and the cards in between them after k shuffles have been performed. Note that we can couple $\{S_k : k \geq 0\}$ with a thinning process $\{V_k : k \geq 0\}$ in such a way that if V_k is successful then $|S_k| = |V_k|$. It follows that if $\widetilde{l}$ and C are the constants appearing in the statement of Lemma 3.3, then Lemma 3.3 implies that the probability that x and y are within a distance C from each other after $\lfloor \log_2 d \rfloor - \widetilde{l})$ steps is at least $\frac{\gamma}{d}$ for a universal constant γ. Furthermore, if x and y are within distance C of each other, there is probability bounded away from 0 that in the next step all the cards in between them with be removed and that x and y will collide in the step following that. Since

$$\mathbf{P}(T_l - 2 = \lfloor \log_2 d \rfloor - \widetilde{l}) = 2^{\lfloor \log_2 d \rfloor - \widetilde{l} - l + 1},$$

it follows that the probability that x and y collide at time T_l is at least

$$2^{\lfloor \log_2 d \rfloor - \widetilde{l} - l - 3} \Big(\frac{\gamma'}{d} \Big),$$

for a universal constant $\gamma' > 0$. This expression is at least $\frac{c}{x}$ for a constant c that depends only on p. □

Now we are ready to apply Lemma 3.4 to the p-riffle shuffle with $c = \alpha\beta/4$ with α and β as in the proof of the lemma. Let X be a random permutation independent of the Y_i:s. Use the chain rule to write

$$\mathrm{ENT}(X) = \sum_{i=1}^{n} E_i = \sum_{l=1}^{\lceil \log_2(n+1) \rceil} \sum_{i \in I_l} E_i.$$

Since there are at most $\log_2 n + 1 \leq 2 \log_2 n$ of the I_m:s, we must have that

$$\sum_{i \in I_{l^*}} E_i \geq \frac{1}{2 \log_2 n} \mathrm{ENT}(X)$$

where l^* is the l that maximizes the inner sum. Hence by Theorem 3.2 and Lemma 3.4 with $l = l^*$ gives

$$\begin{aligned}
\mathrm{ENT}(XY_{(t)}) &\leq \mathrm{ENT}(X) - \frac{C}{\log n}\sum_{k=1}^{n} A_k E_k \\
&\leq \mathrm{ENT}(X) - \frac{C}{\log n}\sum_{i\in I_{l^*}} A_i E_i \\
&\leq \left(1 - \frac{Cc}{4\log^2 n}\right)\mathrm{ENT}(X).
\end{aligned}$$

Now iterating this for $X = \mathrm{id}$, $X = Y_{(t)}$, $X = Y_{(2t)}$, ... and taking $\gamma = Cc/4$ shows that

$$\begin{aligned}
\mathrm{ENT}(Y_{(Bt\log^3 n)}) &\leq \left(1 - \frac{\gamma}{\log^2 n}\right)^{B\log^3 n}\mathrm{ENT}(\mathrm{id}) \\
&\leq n^{-B\gamma}\log(n!) \leq n^{1-B\gamma}\log n \leq \frac{1}{8}
\end{aligned}$$

as soon as, say, $B\gamma \geq 2$. Then, by Lemma 3.1,

$$\|\mathbb{P}(Y_{(Bt\log^3 n)} \in \cdot) - \pi\|_{TV} \leq \sqrt{\frac{1}{2}\mathrm{ENT}(Y_{(Bt\log^3 n)})} \leq \frac{1}{4}.$$

Since t is order $\log n$ we get

$$\tau_{\mathrm{mix}} = O(\log^4 n).$$

References

[1] Aldous, D. and Diaconis, P. (1986), Shuffling cards and stopping times, *Amer. Math. Monthly* **93** 333-348.

[2] Bayer, D. and Diaconis. P. (1992), Trailing the dovetail shuffle to its lair, *Ann. Appl. Probab.* **2**, 294-313.

[3] Diaconis, P. (2001), Mathematical developments from the analysis of riffle shuffling, *Groups, combinatorics & geometry* (Durham, 2001), 7397, *World Sci. Publ., River Edge, NJ, 2003*.

[4] Jonasson, J. (2006), The overhand shuffle mixes in $\Theta(n^2 \log n)$ steps, *Ann. Appl. Probab.* **16**, 231-243.

[5] Morris, B. (2009), Improved mixing time bounds for the Thorp shuffle and L-reveral chain, *Ann. Probab.* **37**, 453–477.

[6] Morris, B. (2013), Improved mixing time bounds for the Thorp shuffle, *Combin. Probab. Comput.* **22**, 118-132.

[7] Pemantle, R. (1989), Randomization time for the overhand shuffle, *J. Theoret. Probab.* **2**, 37-49.

[8] Saloff-Coste, L. Random walks on finite groups. In *Probability on Discrete Structures*, *Encyclopedia of Mathematical Sciences*, vol. 110, H. Kesten, editor, Springer, pp. 263–346, 2004.

[9] Wager, S. (2011), Clumpy riffle shuffles, PhD Thesis, Stanford.