The Annals of Applied Probability
2019, Vol. 29, No. 6, 3861–3892
https://doi.org/10.1214/19-AAP1495
© Institute of Mathematical Statistics, 2019

CUTOFF FOR THE CYCLIC ADJACENT TRANSPOSITION SHUFFLE

BY DANNY NAM[1] AND EVITA NESTORIDI

Princeton University

We study the cyclic adjacent transposition (CAT) shuffle of n cards, which is a systematic scan version of the random adjacent transposition (AT) card shuffle. In this paper, we prove that the CAT shuffle exhibits cutoff at $\frac{n^3}{2\pi^2}\log n$, which concludes that it is twice as fast as the AT shuffle. This is the first verification of cutoff phenomenon for a time-inhomogeneous card shuffle.

1. Introduction. How long does it take to shuffle a deck of cards sufficiently well? Mixing time of card shuffling schemes and Markov chains in general is a widely studied subject in probability. Recently, there has been a lot of interest in understanding the behavior of time-inhomogeneous chains and in sharpening the techniques that have been developed in the time-homogeneous case (see [3, 8, 9, 11–15]). In the present paper, we study the mixing time of the cyclic adjacent transposition shuffle and show that it exhibits cutoff, which is the first demonstration of cutoff phenomenon for a time-inhomogeneous card shuffle.

The cyclic adjacent transposition (CAT) shuffle is a systematic scan version of the adjacent transposition shuffle. In the CAT shuffle, we start with a deck of n cards, that are placed on the vertices on a path of length $(n-1)$. At the beginning of the first step, we flip a fair coin, which determines if we are going to move from left to right or from right to left. If we do the former, then at time $t=1$ with probability $1/2$ we transpose the cards at the ends of the first edge, otherwise we stay fixed. For $t=2,\dots n-1$, with probability $1/2$ we transpose cards at the ends of the tth edge, otherwise we stay fixed, etc. If we move from right to left, at time $t=1,\dots n-1$, with probability $1/2$ we transpose the cards that lie on the ends of the $(n-t)$-th edge, otherwise we do nothing.

In other words, we explore the deck from the first card to the last card with respect to the direction we choose at the beginning, and independently at each step either swap the positions of the neighboring ones or stay fixed according to a fair coin toss. When $t \equiv 1 \bmod (n-1)$, we repeat the first $(n-1)$ steps of the chain independently, that is, pick the orientation (either from 1 to n or from

Received September 2018; revised April 2019.
[1]Supported in part by a Samsung scholarship.
MSC2010 subject classifications. Primary 60J10; secondary 60C05, 60G42.
Key words and phrases. Markov chains, mixing time, adjacent transpositions, time inhomogeneous card shuffles, cutoff phenomenon.

n to 1) uniformly at random, move from the first card to the last one according to the chosen direction and at each step either transpose or do nothing uniformly independently at random.

The configuration space of the CAT shuffle is the symmetric group S_n. Let $x, y \in S_n$ and let $P_x^t(y)$ be the probability of moving from the x to y in t steps. Then the basic limit theorem of Markov chains tells us that P_x^t converges to the uniform measure μ as $t \to \infty$ with respect to the total variation distance

$$d_x(t) := \|P_x^t - \mu\|_{\text{T.V.}} := \frac{1}{2} \sum_{y \in S_n} |P_x^t(y) - \mu(y)|.$$

The mixing time of this Markov chain is defined as

$$t_{\text{mix}}(\varepsilon) = \min\Big\{t \in \mathbb{N} : \max_{x \in S_n}\{d_x(t)\} \le \varepsilon\Big\}.$$

Our main result provides sharp bounds for the mixing time of the CAT shuffle.

THEOREM 1. *For the cyclic adjacent transposition shuffle, we have that for any* $\varepsilon > 0$:

(a) *There is a universal constant* c, *such that* $t_{\text{mix}}(1-\varepsilon) \ge \frac{n^3}{2\pi^2}\log n - \frac{n^3}{2\pi^2}\log(\frac{c\log n}{\varepsilon})$.

(b) $t_{\text{mix}}(\varepsilon) \le (1+o(1))\frac{n^3}{2\pi^2}\log n$.

Theorem 1 says that the cyclic adjacent transposition shuffle exhibits cutoff at $\frac{n^3}{2\pi^2}\log n$, that is, that there is window $w_n = o(n^3 \log n)$ such that

$$\lim_{k\to\infty}\lim_{n\to\infty} d\left(\frac{n^3}{2\pi^2}\log n - kw_n\right) = 1 \quad \text{and}$$

$$\lim_{k\to\infty}\lim_{n\to\infty} d\left(\frac{n^3}{2\pi^2}\log n + kw_n\right) = 0.$$

As mentioned above, the CAT shuffle is a systematic scan version of the adjacent transposition (AT) shuffle. In the AT shuffle, with probability $1/2$ we transpose a random adjacent pair of cards and otherwise do nothing. It is an important card shuffling model mainly because of its connection to the exclusion process. Only recently, Lacoin [6] proved the sharp upper bound for the mixing time of the AT shuffle, which combined with the sharp lower bound of Wilson [17] concluded the proof of cutoff for this model. They also established the same result for the simple exclusion process, verifying the close connections between the two models.

The first time-inhomogeneous card shuffle to be studied is the semirandom transposition card shuffle, which suggests that at time t we transpose the card in position $t \bmod n$ with a uniformly random card. It was introduced by Thorp [16], and

Aldous and Diaconis [1] first raised the question of determining the mixing time of semirandom transpositions. Mironov [7] used this model for a cryptographic system and proved that the mixing time is at most $O(n \log n)$. Mossel, Peres and Sinclair [9] established a matching lower bound of order $\Theta(n \log n)$. This lower bound was obtained using Wilson's method [17], which relies on finding an appropriate eigenfunction.

Another well-studied time-inhomogeneous card shuffle is the card-cyclic-to-random shuffle. In this model, at time t we remove the card with the label $t \bmod n$ and insert it to a uniformly random position of the deck. This model was introduced by Pinsky [11], who showed that n steps are not sufficient to shuffle the deck well enough. Morris, Ning, Peres [8] later proved both a lower and an upper bound of order $n \log n$.

Saloff-Coste and Zuniga [12–15] studied time-inhomogeneous Markov chains via singular value decomposition. In their work, they find better constants for the upper bound for both semirandom transpositions and card-cyclic-to-random shuffles. Their result is based on bounding the singular values of the transition matrix of the time-inhomogeneous chains by the eigenvalues of the time-homogeneous card shuffles. Although very useful in some models, their technique does not work well in our case.

Very recently, Angel and Holroyd [2] asked a different question concerning a similar model; given a sequence of parameters $S = (a_i, b_i, p_i)_{i=1}^{\ell}$, at time $t = 1, \dots, \ell$ with probability p_t they transpose card a_t with the card b_t, otherwise do nothing. They study the question of finding the minimum length ℓ such that the resulting permutation of n cards is uniformly random. They prove that the for the case that $b_i = a_i + 1$, this minimum length is exactly $\binom{n}{2}$.

Another model one can consider is the *single-directional CAT shuffle*, which at time t swaps the cards at positions t, $t + 1 \bmod n - 1$. In other words, it is a variant of the CAT shuffle that explores the deck in a single direction rather than renewing it at every $n - 1$ steps. In this model, we have the same upper bound on the mixing time as part (b) of Theorem 1, and indeed the proof works analogously for this case. However, the techniques used to prove part (a) no longer applies to this model due to lack of symmetry. In the CAT shuffle, setting a random direction of exploration at every $n - 1$ steps provides some amount of symmetry which makes it more convenient to carry out our approach. We conjecture that the single-directional CAT shuffle exhibits cutoff at $\frac{n^3}{2\pi^2} \log n$, the same location as the CAT shuffle.

1.1. *Proof outline*. The main difficulty of studying the CAT shuffle comes from its deterministic selections of update locations. Due to this aspect, it seems impossible to write down the closed formula of the transition using eigenvalues and eigenfunctions, although most of the properties of the AT shuffle can be deduced by this approach [6, 17]. To overcome this difficulty, we rely on the following observations:

(i) We can compute "approximate eigenfunctions," which behave like the actual eigenfunctions but with errors.

(ii) When n is large enough, each card follows a Brownian-type move under an appropriate scaling of n and t.

To prove the lower bound on the mixing time, we derive a generalized version of Wilson's lemma [17] which enables to implement the "approximate eigenfunctions" obtained from observation (i). Using this lemma, we conclude the first part of Theorem 1 by showing that the errors of the approximate eigenfunctions are small enough.

For the upper bound, we rely on the idea of monotone coupling and censoring from Lacoin [6]; by defining the "height" of card decks, we can construct a monotone coupling of the system and take advantage of the censoring inequality following the approach of [6].

In this procedure, a crucial ingredient we need is that the height of the deck decays exponentially in time according to the correct rate. In the AT shuffle [6], this property is derived based on the algebraic relations of the model. Since this approach seems impossible for the CAT shuffle, we take account of (ii) to deduce such condition.

1.2. *Organization of the paper.* In Section 2, we derive a generalized Wilson's lemma that works for approximate eigenfunctions. Then in Section 3.1, we introduce the appropriate approximator to study and show that the error is small enough to deduce the correct lower bound. Based on this result, we conclude the proof of part (a) of Theorem 1 in Section 3.2.

Section 4 is devoted to understanding the movement of a single card. Here, we explain the precise meaning of observation (ii) above and deduce hitting time estimates of a single card. The monotone coupling, the censoring inequality and the exponential decay of the "height" are explained in Section 5.1, and we prove part (b) of Theorem 1 in Section 5.2.

In the final section, Section 6, as an application of our main theorem we study the systematic simple exclusion process which is the particle system version of the CAT shuffle.

2. Generalizing Wilson's lemma. For the lower bound, we will need a generalization of Wilson's lemma [17]. The main difference is that we do not use the precise eigenfunctions of the transition matrix P, but rather functions that behave sufficiently like eigenfunctions.

LEMMA 2. *Let X_t be a Markov chain on a state space Ω_n, with stationary distribution μ. Let $x_0 \in \Omega$. Suppose that there are parameters $\gamma, \delta, R > 0$ and a function $\Phi : \Omega_n \to \mathbb{R}$ such that $\Phi(x_0) > 0$, satisfying the following*:

(a) *The mean of Φ under stationarity is zero, that is $\mu(\Phi) = 0$.*

(b) *We have* $0 < \gamma < 2 - \sqrt{2}$ *and for all* $t \geq 0$ *it holds that*

$$\left|\mathbb{E}\left[\Phi(X_{t+1})|X_t\right] - (1-\gamma)\Phi(X_t)\right| \leq \delta.$$

(c) *We also have that* $\mathbb{E}[(\Delta\Phi_t)^2|X_t] \leq R$, *where* $\Delta\Phi_t := \Phi(X_{t+1}) - \Phi(X_t)$.

Then for $t = \frac{1}{\gamma_\star}\log(\Phi(x_0)) - \frac{1}{2\gamma_\star}\log(\frac{48(\delta\|\Phi\|_\infty + R)}{\gamma\varepsilon})$, *we have*

$$\|P_{x_0}^t - \mu\|_{\text{T.V.}} \geq 1 - \varepsilon,$$

where $\gamma_\star := -\log(1-\gamma)$.

PROOF. Let $\varepsilon > 0$. By iterating the condition (b), we get that

$$\mathbb{E}_{x_0}\left[\Phi(X_t)\right] \geq (1-\gamma)^t\Phi(x_0) - \delta/\gamma. \tag{2.1}$$

To control the variance, we notice the inequality that

$$\begin{aligned}\mathbb{E}\left[\left(\Phi(X_{t+1})\right)^2|X_t\right] &= \left(\Phi(X_t)\right)^2 + 2\Phi(X_t)\mathbb{E}[\Delta\Phi_t|X_t] + \mathbb{E}\left[(\Delta\Phi_t)^2|X_t\right] \\ &\leq (1-2\gamma)\Phi(X_t)^2 + \left(\delta\|\Phi\|_\infty + R\right).\end{aligned} \tag{2.2}$$

Iterating (2.2), we have that

$$\mathbb{E}_{x_0}\left[\left(\Phi(X_t)\right)^2\right] \leq (1-2\gamma)^t\Phi(x_0)^2 + \frac{\delta\|\Phi\|_\infty + R}{2\gamma}. \tag{2.3}$$

Using (2.1), this implies that

$$\operatorname{Var}\left(\Phi(X_t)|X_0 = x_0\right) \leq \frac{\delta\|\Phi\|_\infty + R}{2\gamma} + \frac{2\delta\|\Phi\|_\infty}{\gamma} \leq \frac{3(\delta\|\Phi\|_\infty + R)}{\gamma}. \tag{2.4}$$

Letting t go to infinity, we also get the same bound for $\operatorname{Var}(\Phi)$ under the stationary distribution.

Let $t = \frac{1}{\gamma_\star}\log(\Phi(x_0)) - \frac{1}{2\gamma_\star}\log(\frac{48(\delta\|\Phi\|_\infty + R)}{\gamma\varepsilon})$ and consider the event

$$A = \left\{x \in \Omega_n : \Phi(x) < \frac{1}{2}\mathbb{E}_{x_0}\left[\Phi(X_t)\right]\right\}.$$

Then by Chebychev's inequality combined with (2.1) and (2.4), we have that

$$\mathbb{P}_{x_0}(X_t \in A) \leq \frac{12(\delta\|\Phi\|_\infty + R)/\gamma}{(1-\gamma)^{2t}\Phi(x_0)^2 - 2\delta\|\Phi\|_\infty/\gamma} \leq \frac{\varepsilon}{2}. \tag{2.5}$$

Similarly with respect to the stationary measure, we obtain that

$$\mathbb{P}_\mu(X \in A) \geq 1 - \frac{\varepsilon}{2}. \tag{2.6}$$

Combining (2.5) and (2.6), we deduce that

$$\|P_{x_0}^t - \mu\|_{\text{T.V.}} \geq \left|P_{x_0}^t(A) - \mu(A)\right| \geq 1 - \varepsilon. \qquad \square$$

3. The lower bound. In [17], the lower bound on the mixing time for the random AT shuffle is obtained by analyzing the *height function representation* of the chain. In this case, one can compute the exact eigenvalues and eigenfunctions of the transition of height functions.

On the other hand, the main difficulty of investigating the CAT shuffle is that we cannot precisely calculate such eigenvalues and eigenfunctions since the update locations are not given randomly. However, we can still overcome this obstacle by using the objects which approximately behave like eigenfunctions with small enough errors.

In Section 3.1, we introduce the *height function representation* of the CAT shuffle and describe its first and the second moment estimates, based on the aforementioned idea of "approximate eigenfunctions." Then Section 3.2 is devoted to proving Theorem 1, part (a) using the ingredients obtained in Section 3.1 and Lemma 2.

3.1. *The moment estimates.* Let $\sigma_0 := \mathrm{id} \in S_n$ be the starting state of the CAT shuffle and let σ_s denote the deck at time s. For each $t \in \mathbb{N}$, the *height function* $h_t : [n] \to \mathbb{R}$ of (σ_s) is defined as

$$h_t(x) := \sum_{z=1}^{x} \mathbf{1}\{\sigma_{(n-1)t}(z) \le \lfloor n/2 \rfloor\} - \frac{\lfloor n/2 \rfloor}{n} x. \tag{3.1}$$

Let $\mathcal{F}_t$ denote the sigma-algebra for the shuffling until time $(n-1)t$. Our goal in this subsection is to obtain the first and the second moment estimates on the following quantity Φ_t:

$$\Phi_t := \sum_{x=1}^{n-1} h_t(x) \sin\left(\frac{\pi x}{n}\right). \tag{3.2}$$

We begin with the first moment estimate of Φ_t. The following lemma is proven similarly as Lemma 17, and the proof can be found in Section A.3.

LEMMA 3. *Let Φ_t, $\mathcal{F}_t$ defined as above. For any $t \in \mathbb{N}$, we have*

$$\left|\mathbb{E}[\Phi_{t+1} | \mathcal{F}_t] - (1-\gamma)\Phi_t\right| \le \frac{3\pi}{4n}, \tag{3.3}$$

where $\gamma := \pi^2/n^2 - O(n^{-4})$.

REMARK 4. Although we cannot have a more precise form such as $\mathbb{E}[\Phi_{t+1} | \mathcal{F}_t] = (1-\gamma)\Phi_t$ as [17], Lemma 2 says that the estimate of Lemma 3 is sufficient to get a lower bound.

Our next goal is to bound the second moment of Φ_t. One convenient way of doing this is to look at $\Delta\Phi_t := \Phi_{t+1} - \Phi_t$, similar to what is done in [17].

LEMMA 5. *There exists an absolute constant $C > 0$ such that for any $t \in \mathbb{N}$,*

$$\mathbb{E}\big[(\Delta\Phi_t)^2 | \mathcal{F}_t\big] \le Cn \log n.$$

PROOF. For each $a \in [n]$, let $q_t(a)$ denote the position of the card a at time $(n-1)t$, that is, $q_t(a) := \sigma^{-1}_{(n-1)t}(a)$. Observe that we can write $h_t(x)$ in terms of $q_t(a)$ in the following way:

$$h_t(x) = \sum_{a=1}^{\lfloor n/2 \rfloor} \mathbf{1}_{\{q_t(a) \le x\}} - \frac{x}{n}\left\lfloor \frac{n}{2} \right\rfloor. \tag{3.4}$$

Therefore, $\Delta\Phi_t = \Phi_{t+1} - \Phi_t$ becomes

$$\Delta\Phi_t = \sum_{a=1}^{\lfloor n/2 \rfloor} \left\{ \sum_{x=1}^{n-1} (\mathbf{1}_{\{q_{t+1}(a) \le x\}} - \mathbf{1}_{\{q_t(a) \le x\}}) \sin\left(\frac{\pi x}{n}\right) \right\} = \sum_{a=1}^{\lfloor n/2 \rfloor} \psi_t(a),$$

where we define $\psi_t(a)$ by

$$\psi_t(a) := \sum_{x=q_{t+1}(a)}^{n-1} \sin\left(\frac{\pi x}{n}\right) - \sum_{x=q_t(a)}^{n-1} \sin\left(\frac{\pi x}{n}\right).$$

We begin with estimating $\mathbb{E}[\psi_t(a)^2 \mid \mathcal{F}_t]$. Let $\overrightarrow{\mathbb{E}}$ (resp. $\overleftarrow{\mathbb{E}}$) denote the conditional expectation given the event that we explore the deck from position 1 to n (resp. n to 1) over the time period of $(n-1)t+1$ to $(n-1)(t+1)$. In other words, if $\mathbf{c}_t \in \{1, n\}$ is the random variable that denotes the starting position of exploration at time $(n-1)t$, then $\overrightarrow{\mathbb{E}}[\cdot \mid \mathcal{F}_t] = \mathbb{E}[\cdot \mid \mathcal{F}_t, \mathbf{c}_t = 1]$. Recall that $q_{t+1}(a) - q_t(a)$ follows the distribution (4.2, 4.3). Letting j count the displacement of card a, we have that for $2 \le q_t(a) \le n-1$,

$$\begin{aligned}
&\overrightarrow{\mathbb{E}}\big[\psi_t(a)^2 \mid \mathcal{F}_t\big] \\
&\quad \le \frac{1}{2}\sin^2\left(\frac{\pi(q_t(a)-1)}{n}\right) + \sum_{k=1}^{\infty} \frac{1}{2^{k+2}} \left\{ \sum_{j=0}^{k-1} \sin\left(\frac{\pi(q_t(a)+j)}{n}\right) \right\}^2 \\
&\quad \le \frac{\pi^2}{2n^2}\left\{ (q_t(a)-1)^2 + \sum_{k=1}^{\infty} \frac{1}{2^{k+1}} \left(kq_t(a) + \frac{k(k-1)}{2} \right)^2 \right\} \le C_1,
\end{aligned} \tag{3.5}$$

for some absolute constant $C_1 > 0$, using the fact that $\sin\theta \le \theta$ and $q_t(a) \le n$. We can conduct a similar calculation for the cases $q_t(a) = 1, n$ as well as for $\overleftarrow{\mathbb{E}}[\psi_t(a)^2 \mid \mathcal{F}_t]$ and obtain that for all a,

$$\mathbb{E}\big[\psi_t(a)^2 \mid \mathcal{F}_t\big] \le C_1. \tag{3.6}$$

We turn our attention to estimating the correlation and show that $|\mathbb{E}[\psi_t(a)\psi_t(b) \mid \mathcal{F}_t]| = O(\frac{1}{n})$ for a, b which are far apart from each other. In particular, let us assume that both $q_t(a) \geq 2$ and $q_t(a) + 4\log n \leq q_t(b) \leq n-1$ hold true. Define A to be the event that

$$A := \{q_{t+1}(a) - q_t(a) \leq 4\log n - 2\}.$$

Then $q_{t+1}(a)$ and $q_{t+1}(b)$ are conditionally independent given $\mathcal{F}_t$ and the event

$$\{\mathbf{c}_t = 1\} \cap A.$$

Therefore, we can express $\overrightarrow{\mathbb{E}}[\psi_t(a)\psi_t(b) \mid \mathcal{F}_t]$ by

$$\begin{aligned}\overrightarrow{\mathbb{E}}\big[\psi_t(a)\psi_t(b) \mid \mathcal{F}_t\big] &= \overrightarrow{\mathbb{P}}(A)\overrightarrow{\mathbb{E}}\big[\psi_t(a) \mid A, \mathcal{F}_t\big]\overrightarrow{\mathbb{E}}\big[\psi_t(b) \mid A, \mathcal{F}_t\big] \\ &\quad + \overrightarrow{\mathbb{E}}\big[\psi_t(a)\psi_t(b)\mathbf{1}_{A^c} \mid \mathcal{F}_t\big].\end{aligned} \tag{3.7}$$

Since $\mathbb{P}(A^c) \leq n^{-4}$, Hölder's inequality implies that

$$\begin{aligned}\overrightarrow{\mathbb{E}}\big[\psi_t(a)\psi_t(b)\mathbf{1}_{A^c} \mid \mathcal{F}_t\big] &\leq \overrightarrow{\mathbb{E}}\big[\psi_t(a)^4 \mid \mathcal{F}_t\big]^{\frac{1}{4}}\overrightarrow{\mathbb{E}}\big[\psi_t(b)^4 \mid \mathcal{F}_t\big]^{\frac{1}{4}}\overrightarrow{\mathbb{P}}\big(A^c\big)^{\frac{1}{2}} \\ &\leq \frac{C_2}{n^2},\end{aligned} \tag{3.8}$$

by noting that the fourth moment of $\psi_t(a)$ conditioned on $\mathcal{F}_t$ can be estimated in the same way as (3.5). On the other hand, we have

$$\begin{aligned}&\big|\overrightarrow{\mathbb{E}}\big[\psi_t(a)\mathbf{1}_A \mid \mathcal{F}_t\big]\big| \\ &\quad = \left|\frac{1}{2}\sin\left(\frac{\pi(q_t(a)-1)}{n}\right) - \sum_{k=1}^{\lfloor 4\log n\rfloor - 2}\frac{1}{2^{k+2}}\sum_{j=0}^{k-1}\sin\left(\frac{\pi(q_t(a)+j)}{n}\right)\right| \\ &\quad \leq \left|\frac{1}{2}\sin\left(\frac{\pi(q_t(a)-1)}{n}\right) - \sum_{k=1}^{\infty}\frac{1}{2^{k+2}}\sum_{j=0}^{k-1}\sin\left(\frac{\pi(q_t(a)+j)}{n}\right)\right| + \frac{1}{n^3}.\end{aligned}$$

Using $|\sin(x+\delta) - \sin(x)| \leq \delta$ to control the right-hand side, we obtain that

$$\big|\overrightarrow{\mathbb{E}}\big[\psi_t(a)\mathbf{1}_A \mid \mathcal{F}_t\big]\big| \leq \frac{\pi}{2n} + \sum_{k=1}^{\infty}\frac{1}{2^{k+2}}\sum_{j=0}^{k-1}\frac{j\pi}{n} + \frac{1}{n^3} \leq \frac{C_3'}{n}, \tag{3.9}$$

for an absolute constant $C_3' > 0$. Similar computations can be done for $\overleftarrow{\mathbb{E}}$. Since $\overrightarrow{\mathbb{P}}(A) \geq 1 - n^{-4}$ and $|\overrightarrow{\mathbb{E}}[\psi_t(b) \mid \mathcal{F}_t]| \leq C_1$, we deduce by combining (3.7–3.9) that

$$\big|\mathbb{E}\big[\psi_t(a)\psi_t(b) \mid \mathcal{F}_t\big]\big| \leq \frac{C_3}{n}, \tag{3.10}$$

for some absolute constant $C_3 > 0$.

Let $Q \subset [\lfloor n/2 \rfloor]^2$ be defined as

$$Q := \{(a,b) \in [\lfloor n/2 \rfloor]^2 : 2 \le q_t(a), q_t(b) \le n-1, |q_t(a) - q_t(b)| \le 4\log n\}.$$

We also denote $Q^c := [\lfloor n/2 \rfloor]^2 \setminus Q$. Then we can estimate $\mathbb{E}[(\Delta\Phi_t)^2|\mathcal{F}_t]$ using the inequalities (3.6) and (3.10) as follows:

$$\begin{aligned}
\mathbb{E}[(\Delta\Phi_t)^2|\mathcal{F}_t] &= \sum_{a,b=1}^{\lfloor n/2 \rfloor} \mathbb{E}[\psi_t(a)\psi_t(b)|\mathcal{F}_t] \\
&\le \sum_{(a,b)\in Q^c} \mathbb{E}[\psi_t(a)\psi_t(b)|\mathcal{F}_t] \\
&\quad + \sum_{(a,b)\in Q} \mathbb{E}[\psi_t(a)^2|\mathcal{F}_t]^{\frac{1}{2}} \mathbb{E}[\psi_t(b)^2|\mathcal{F}_t]^{\frac{1}{2}} \\
&\le \frac{n^2}{4} \cdot \frac{C_3}{n} + 4n\log n \cdot C_1 \le Cn\log n,
\end{aligned}$$

for an absolute constant $C > 0$. $\square$

3.2. *Proof of Theorem* 1, *part* (*a*). In this section, we conclude the proof of Theorem 1, part (a). Lemma 3 says that

$$|\mathbb{E}[\Phi_{t+1}|\mathcal{F}_t] - (1-\gamma)\Phi_t| \le \frac{3\pi}{4n}, \tag{3.11}$$

where $\gamma = \pi^2/n^2 - O(n^{-4})$. Moreover, Lemma 5 gives us that

$$\mathbb{E}[(\Delta\Phi_t)^2|\mathcal{F}_t] \le Cn\log n. \tag{3.12}$$

Also, by the definition of Φ_t, when $t = 0$ it satisfies that

$$\Phi_0 = \sum_{x=1}^{n-1} \frac{1}{2}\{x \wedge (n-x)\} \sin\left(\frac{\pi x}{n}\right) \ge 2 \sum_{x=\frac{n}{4}}^{\frac{n}{2}} \frac{n}{4} \sin\left(\frac{\pi}{4}\right) \ge \frac{n^2}{8\sqrt{2}}. \tag{3.13}$$

Define $\Phi : S_n \to \mathbb{R}$ to be

$$\Phi(\sigma) = \sum_{x=1}^{n} h(\sigma, x) \sin\left(\frac{\pi x}{n}\right),$$

where $h(\sigma, \cdot)$ is the height function of σ defined in (3.1). We also have that $\Phi(\sigma_{(n-1)t}) = \Phi_t$. Plugging Φ into Lemma 2, equations (3.11), (3.12) and (3.13) imply that

$$t_{\text{mix}}(1-\varepsilon) \ge \frac{n^3}{2\pi^2}\log n - \frac{n^3}{2\pi^2}\log\left(\frac{c\log n}{\varepsilon}\right),$$

where $c > 0$ is a universal constant.

4. Following one card. Throughout this section, we label our deck of n-cards by $[n_0] := \{0, 1, \dots, n-1\}$. Our state space is the symmetric group on $[n_0]$, which is denoted by S_{n_0}. For $a \in [n_0]$, let $q_t(a) = \sigma^{-1}_{(n-1)t}(a)$ denote the position of card a at time $(n-1)t$.

Let $T_a := \min\{t : q_t(a) = n-1\}$ be the first (scaled) time that the card a in the deck reaches at the right end. Our goal in this subsection is to prove the following lemma on T_a.

LEMMA 6. *Let a be an arbitrary element of $[n_0]$, and define $q_t(a)$, T_a as above. For any CAT shuffle (σ_t) and any $\delta > 0$, there exist $N_\delta, \theta_\delta > 0$ independent of n such that for all $n \geq N_\delta$ and $\theta_\delta \leq \theta \leq n$, we have*

$$\mathbb{P}\left(T_a > \frac{\theta n^2}{\pi^2}\right) \leq \left(1 + O\left(\theta n^{-2}\right)\right) e^{-\frac{(1-\delta)}{4}\theta}. \tag{4.1}$$

In order to prove Lemma 6, we analyze the process $q_t(a)$ by coupling it with another random walk that we may have a better control over. From now on, we focus on the process $\{q_t(a)\}_{t\in\mathbb{N}}$, regarding each exploration of the whole line as a single step. Let X be a random variable on $\mathbb{Z}$ with the following probability distribution:

- For all $k \in \mathbb{Z}$, $\mathbb{P}(X = k) = 2^{-(|k|+3)} + 2^{-(3-|k|)}\mathbf{1}_{\{|k|\leq 1\}}$.

Note that X has mean 0 and variance 2. Let X_i be i.i.d. copies of X, and define S_t to be

$$S_t := \sum_{i=1}^{t} X_i.$$

LEMMA 7. *For all $a \in [n_0]$, there is a coupling between $\{q_t(a)\}_{t\in\mathbb{N}}$ and $\{X_i\}_{i\in\mathbb{N}}$ such that for all $t \geq 0$, on the event $\{T_a > t\}$ we have*

$$q_t(a) \geq \widehat{S}^a_t := S_t - \left(\min\{S_s : s \leq t\} \wedge \left(-q_0(a)\right)\right).$$

REMARK 8. $\widehat{S}^a_t$ is obtained by pushing $S_t + q_0(a)$ above as little as possible while making it stay nonnegative. Figure 2 describes an example of its sample path.

PROOF OF LEMMA 7. We first notice that the distribution of $q_t(a) - q_{t-1}(a)$ is very similar to that of X, as drawn in Figure 1. Given that $0 < x := q_{t-1}(a) < n-1$, one can see that:

- For $-x+1 \leq k \leq n-x-2$,

$$\mathbb{P}\left(q_t(a) = x + k\right) = 2^{-(|k|+3)} + 2^{-(3-|k|)}\mathbf{1}_{\{|k|\leq 1\}} = \mathbb{P}(X = k); \tag{4.2}$$

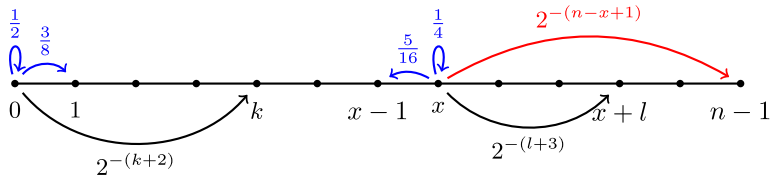

FIG. 1. *Jump probabilities of the process* $\{q_t(a)\}$.

- For $k \in \{-x, n-x-1\}$, $\mathbb{P}(q_t(a) = x+k) = 2^{-(|k|+2)} + 2^{-(3-|k|)}\mathbf{1}_{\{|k|\le 1\}}$.

If $x = 0$, then

- For $0 \le k \le n-2$,

$$\mathbb{P}\big(q_t(a) = k\big) = 2^{-(k+2)} + \frac{1}{4}\mathbf{1}_{\{k\le 1\}} \ge \mathbb{P}(X = k); \tag{4.3}$$

- For $k = n-1$, $\mathbb{P}(q_t(a) = k) = 2^{-n}$.

Notice that if $0 < x := q_{t-1}(a) < n-1$, we have $\mathbb{P}(q_t(a) = 0) = \mathbb{P}(X \le -x)$. Combined with (4.2), this implies that when $0 < x < n-1$, the laws of $q_t(a)$ and X_t can be coupled so that

$$q_t(a) - x = X_t \vee (-x). \tag{4.4}$$

Similarly when $x = 0$, we have $\mathbb{P}(q_t(a) = 0) \le \mathbb{P}(X \le 0)$, and hence with (4.3) gives us that we have a coupling of $q_t(a)$ and X_t that satisfies

$$q_t(a) - x \ge X_t \vee (-x). \tag{4.5}$$

As mentioned in Remark 8, $\widehat{S}_t^A$ is the process obtained by pushing $S_t + q_0(a)$ above to 0 whenever it hits a negative point. Therefore, under the aforementioned

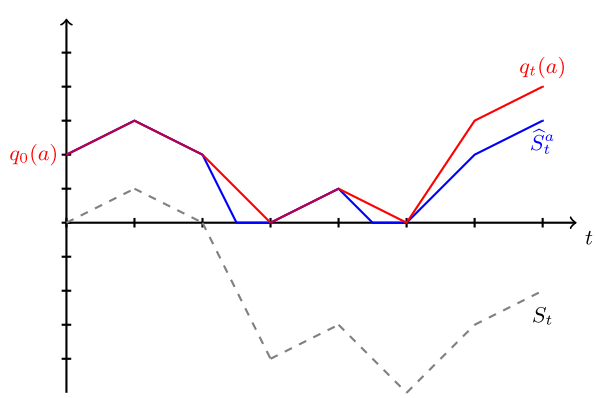

FIG. 2. *Sample paths of* $q_t(a)$ *and* $\widehat{S}_t^a := S_t - (\min_{s\le t} S_s \wedge (-q_0(a)))$, *with* $q_0(a) = 2$.

coupling (4.4, 4.5), $q_t(a)$ and $\widehat{S}_t^a$ satisfies

$$q_t(a) \geq \widehat{S}_t. \qquad \square$$

Let us define the stopping time τ_n^x such that

$$\tau_n^x := \min\Big\{t : S_t - \Big(\min_{s\leq t} S_s \wedge (-x)\Big) \geq n\Big\}.$$

Due to Lemma 7, it suffices to prove the corresponding inequality for τ_n^x as (4.1) for arbitrary x. Since $S_t - (\min_{s\leq t} S_s \wedge (-x))$ is increasing in x, it is enough to look at τ_n^0. Consider the extension S_t to noninteger t's by setting (t, S_t) to be the point on the linear segment connecting (r, S_r) and $(r+1, S_{r+1})$ for $r = \lfloor t \rfloor$. Since the increments of S_t for integer times are i.i.d. with mean 0 and variance 2, Donsker's theorem (see, e.g., [4]) directly implies the following.

PROPOSITION 9. *Let B_t and R_t denote the standard Brownian motion and the standard reflected Brownian motion, respectively. Then, as $m \to \infty$, we have the following weak convergence when viewed as measures on $C[0,\infty)$, the space of continuous functions on* $[0,\infty)$:

$$\frac{S_{m^2 t}}{\sqrt{2}m} \longrightarrow B_t;$$

$$\frac{S_{m^2 t} - \min_{s\leq m^2 t} S_s}{\sqrt{2}m} \longrightarrow R_t.$$

PROOF. The first equation is a restatement of Donsker's theorem. For the second part, define $\Psi : C[0,\infty) \to C[0,\infty)$ as $\Psi(f)(t) = f(t) - \min_{s\leq t}\{f(s)\}$. Since $\Psi(B_t) = R_t$ in law and Ψ is continuous with respect to the supremum norm topology, the second convergence follows from the first one. $\square$

Define $\tau^R := \min\{t > 0 : R_t \geq 1/\sqrt{2}\}$. As an immediate consequence of Proposition 9, we have the following corollary.

COROLLARY 10. *For any constant $\theta > 0$, we have*

$$\lim_{n\to\infty} \mathbb{P}(\tau_n^0 > \theta n^2) = \mathbb{P}(\tau^R > \theta) = \mathbb{P}(\tau^{|B|} > \theta), \tag{4.6}$$

where $\tau^{|B|} := \min\{t > 0 : |B_t| \geq 1/\sqrt{2}\}$.

PROOF. The first equality is obvious by Proposition 9. The second follows by the fact that $(R_t)_{t\geq 0} = (|B_t|)_{t\geq 0}$ in law (see, e.g., Chapter 3.6 of [5]). $\square$

From now on, we choose to look at $\tau^{|B|}$ instead of τ^R. Let $\widetilde{S}_m$ denote the simple random walk so that the increments are i.i.d. $2\,\mathrm{Ber}(\frac{1}{2}) - 1$ and $\widetilde{S}_0 = 0$. Let $\widetilde{\tau}^n := \min\{m : \widetilde{S}_m \notin (-n/\sqrt{2}, n/\sqrt{2})\}$. Then the following lemma is based on the same spirit as Lemma 9 of [17]. We postpone its proof to Section A.2.

LEMMA 11. *There exists a constant $C > 0$ that satisfies*

$$\mathbb{P}(\widetilde{\tau}^n > \theta n^2) < C(1 + O(\theta n^{-2})) \exp\left(-\frac{\pi^2\theta}{4}\right)$$

for all $\theta > 0$ (θ may depend on n).

The following corollary is a consequence of Corollary 10 and Lemma 11.

COROLLARY 12. *For any $\delta > 0$, there exist constants $\theta_0 = \theta_0(\delta)$ and $N = N(\delta, \theta_0) > 0$ such that for all $n \geq N$, we have*

$$\mathbb{P}(\tau_n^0 > \theta_0 n^2) \leq (1 + O(\theta_0 n^{-2})) \exp\left(-\frac{(1-\delta)}{4}\pi^2\theta_0\right).$$

PROOF. Let $\delta > 0$ be given. By Lemma 11, we can pick a large $\theta_0 = \theta_0(\delta)$ such that for all constants $\theta \geq \theta_0 - \delta$ not depending on n,

$$\mathbb{P}(\widetilde{\tau}^n > \theta n^2) \leq (1 + O(\theta n^{-2})) \exp\left(-\frac{(1-\delta/2)}{4}\pi^2\theta\right).$$

Then Donsker's theorem implies that $\mathbb{P}(\tau^{|B|} > \theta_0) \leq \exp(-(1-\delta/2)\pi^2\theta_0)$, since $\widetilde{S}_{m^2 t}/m$ converges to (B_t) as in Proposition 9. Noting that τ^R and $\tau^{|B|}$ share the same law, we use (4.6) to deduce that there exists $N = N(\delta, \theta_0)$ such that for all $n > N$,

$$\mathbb{P}(\tau_n^0 > \theta_0 n^2) \leq \mathbb{P}(\tau^R > \theta_0 - \delta) \leq (1 + O(\theta_0 n^{-2})) \exp\left(-\frac{(1-\delta)}{4}\pi^2\theta_0\right),$$

which is the desired inequality. $\square$

PROOF OF LEMMA 6. As observed in Lemma 7, we can couple the two processes $q_t(a)$ and S_t such that $T_a \leq \tau_n^0$, as a single increment in S_t corresponds to $n-1$ steps of swapping in the CAT shuffle. Therefore, Corollary 12 implies that

$$\mathbb{P}(T_a > \theta_0 n^2) \leq (1 + O(\theta_0 n^{-2})) \exp\left(-\frac{(1-\delta/2)}{4}\pi^2\theta_0\right) \tag{4.7}$$

for some constant $\theta_0 > 0$ depending on δ. For any $\theta > \theta_0$ and $n > N_\delta$, we have

$$\begin{aligned}\mathbb{P}(T_a > \theta n^2) &\leq (1 + O(\theta n^{-2}))) \exp\left(-\frac{(1-\delta/2)}{4}\pi^2\theta_0\left\lfloor\frac{\theta}{\theta_0}\right\rfloor\right)\\ &\leq (1 + O(\theta n^{-2})) \exp\left(-\frac{(1-\delta)}{4}\pi^2\theta\right),\end{aligned}$$

where the first inequality is obtained by iterating (4.7) for $\lfloor\theta/\theta_0\rfloor$-times and by the fact that $\theta \leq n$, and the last inequality holds for all $\theta_\delta < \theta \leq n$ where θ_δ is a large constant. $\square$

5. The upper bound. In [6], Lacoin derives the sharp upper bound on the mixing time of the random AT shuffle by introducing a monotone coupling of the model and implementing the censoring inequality in a clever way. He defines the function $\widetilde{\sigma}_t$ which can be understood as the "height" of σ_t to build up a monotone framework of the system. It turns out that a similar argument is also applicable to the CAT shuffle along with some appropriate adjustments.

However, in the CAT shuffle, the major difficulty of adopting this argument comes from understanding the decay of the height $\widetilde{\sigma}_t$. In [6], exponential decay of $\widetilde{\sigma}_t$ is obtained using algebraic properties of the model based on its eigenvalues and eigenfunctions. Since this approach seems impossible in the current context, we rely on the ideas developed in Section 4 to deduce the same property for the CAT shuffle.

In Section 5.1, we introduce a monotone coupling for the CAT shuffle. In such a monotone system, we can take advantage of the censoring inequality, which essentially says that if we ignore some updates (swaps) in a CAT shuffle, then the distance from equilibrium of the resulting chain is greater than that of the original one. In Section 5.2, we conclude the proof of Theorem 1 based on the tools from the previous sections. Finally, we prove the decay estimate on $\widetilde{\sigma}_t$ in Section A.1.

5.1. *Monotone coupling*. In this subsection, we introduce a monotone coupling for the CAT shuffle following the argument of Lacoin [6] and Wilson [17]. Via the monotone coupling, we derive the censoring inequality which will be crucial in Section 5.2.

Let S_n be the group of permutations on $[n] = \{1, \dots, n\}$. For each $\sigma \in S_n$, we define the function $\widetilde{\sigma} : [n] \times [n] \to \mathbb{R}$ as follows:

$$\widetilde{\sigma}(x, y) := \sum_{z=1}^{x} \mathbf{1}_{\{\sigma(z) \le y\}} - \frac{xy}{n}. \tag{5.1}$$

Subtracting xy/n is introduced in order to set the average $\mu(\widetilde{\sigma}(x, y))$ to be 0, where μ denotes the uniform measure on S_n. Throughout this section, we use the following partial order on S_n based on the function $\widetilde{\sigma}$:

- For $\sigma, \sigma' \in S_n$, $\sigma \ge \sigma'$ if and only if $\widetilde{\sigma}(x, y) \ge \widetilde{\sigma}'(x, y)$ for all $x, y \in [n]$.

Under this ordering, one can observe that the identity element (denoted id) is maximal, and the permutation that maps x to $n + 1 - x$ is minimal.

DEFINITION 13 (Monotone coupling). Let $\{U_t : t \in \mathbb{N}\}$ be the family of i.i.d. $\mathrm{Ber}(\frac{1}{2})$ random variables, and let $\{\mathbf{c}_i : i \in \mathbb{N}\}$ be the family of i.i.d. $\mathrm{Unif}\{1, n\}$ random variables (i.e., $\mathbf{c}_i = 1$ or n, each with probability half) that is independent from U_t's. Using $\mathbf{c}_i$'s and U_t's, we define the updates of the CAT shuffle as follows:

(1) At time $(n-1)i + 1$ for each $i = 0, 1, \dots$, we begin the exploration starting from position $\mathbf{c}_i$. That is, if for instance $\mathbf{c}_i = 1$, then during the time interval from $(n-1)i + 1$ to $(n-1)(i+1)$, we explore the deck from left to right.

(2) Suppose that $(x, x+1)$ is the edge we are about to swap or not at time $t+1$.

- If either $U_t = 0$ and $\sigma_t(x) < \sigma_t(x+1)$ or $U_t = 1$ and $\sigma_t(x) > \sigma_t(x+1)$, then we swap the edge $(x, x+1)$, hence obtaining $\sigma_{t+1}(x) = \sigma_t(x+1)$ and $\sigma_{t+1}(x+1) = \sigma_t(x)$.
- In other cases, we do nothing.

In other words, if $U_t = 0$ we reverse-sort the cards at positions $x, x+1$, whereas if $U_t = 1$ we sort the cards at x, $x+1$. One can easily check that this update rule exhibits the same transition matrix as the CAT shuffle. The following proposition describes a significant advantage of this coupling, namely the preservation of monotonicity. For a proof, we refer to [6].

PROPOSITION 14 ([6], Proposition 3.1). *Let $\xi, \xi' \in S_n$ and let σ_t^{ξ} (resp., $\sigma_t^{\xi'}$) denote the CAT shuffle starting from ξ (resp., ξ') coupled by the aforementioned update rules. If $\xi \geq \xi'$, then we have*

$$\sigma_t^{\xi} \geq \sigma_t^{\xi'} \quad \textit{for all } t \geq 0.$$

DEFINITION 15. A probability distribution ν on S_n is called *increasing* if $\nu(\sigma) \geq \nu(\sigma')$ holds for all $\sigma, \sigma' \in S_n$ such that $\sigma \geq \sigma'$.

One property of the adjacent transposition shuffle is that it preserves the monotonicity of measures. This fact is formalized in the following lemma, whose proof can be found in [6].

LEMMA 16 ([6], Proposition A.1). *Let ν be an increasing probability measure on S_n. For any $x \in [n-1]$, let σ^x be the resulting state of σ after performing an update at edge $(x, x+1)$, that is, either swap the labels $\sigma(x)$, $\sigma(x+1)$ with probability half or stay fixed otherwise. Let ν^x denote the distribution of σ^x when $\sigma \sim \nu$. Then, ν^x is increasing.*

Furthermore, we introduce two additional tools which will be used in the next subsection: decay estimate of $\widetilde{\sigma}_t$ and the censoring inequality.

For any fixed $x, y \in [n]$, the average of $\widetilde{\sigma}(x, y)$ over μ is 0. We are interested in decay speed of the expected value of $\widetilde{\sigma}_t(x, y)$, which can be described as the following lemma.

LEMMA 17. *Let (σ_t) denote the CAT shuffle on $[n]$ that starts from an arbitrary initial state and let $\delta > 0$ be arbitrary. Then there exist $N_\delta, \theta_\delta > 0$ independent of n such that for any $n \geq N_\delta$, $x, y \in [n]$, and $\theta_\delta n^3 < t \leq n^4$, we*

have

$$\left|\mathbb{E}[\widetilde{\sigma}_t(x,y)]\right| \le n\big(1+O(tn^{-5})\big)\exp\left(-(1-\delta)\frac{\pi^2}{n^3}t\right).$$

REMARK 18. The random AT shuffle version of Lemma 17 is discussed in [6], Lemma 4.1. In the random AT shuffle, this is proven by a direct computation of the eigenvalues and eigenvectors of the simple random walk. In the present context, such method seems extremely difficult to be applied because the model is more complicated. Instead, we choose an alternative approach, based on the ideas similar to Lemmas 6 and 11. Due to its technicality, we defer the proof of Lemma 17 to Section A.1.

To conclude this section, we introduce the censoring inequality for the CAT shuffle. In [10], Peres and Winkler proved the censoring inequality for the Glauber dynamics on monotone spin systems. The message of this inequality is that ignoring updates can only slow down the mixing. In [6], Lacoin extended the inequality to the random AT shuffle. It turns out that in the CAT shuffle, the censoring inequality is still true.

To formalize, a *censoring scheme* is a function $\mathcal{C}:\mathbb{N}\to\mathcal{P}([n-1])$ that is interpreted as follows:

- At each time t, the edge $(x,x+1)$ that we are about to update is ignored if and only if $x\in\mathcal{C}(t)$.

Let P_ν^t be the probability distribution of the CAT shuffle at time t with initial distribution ν, and let $P_{\nu,\mathcal{C}}^t$ denote the distribution of the CAT shuffle at time t which has performed the censoring dynamics according to $\mathcal{C}$ while started from the same distribution ν. Intuitively, the updates are the triggers that carry the chain to its equilibrium, so one might guess that the censored dynamics is further away from the equilibrium than the original one. This intuition turns out to be true for the CAT shuffle due to monotonicity of the system, as long as we start from an "increasing" initial distribution ν. Lacoin [6] and [10] describe this phenomenon as follows.

PROPOSITION 19 (The censoring inequality). *Let ν be an increasing probability distribution on S_n. For any censoring scheme $\mathcal{C}:\mathbb{N}\to\mathcal{P}([n-1])$ and any $t\ge 0$, we have*

$$\|P_\nu^t-\mu\|_{\mathrm{T.V.}} \le \|P_{\nu,\mathcal{C}}^t-\mu\|_{\mathrm{T.V.}}. \tag{5.2}$$

In particular, (5.2) *holds for the starting distribution $\nu=\delta_{\mathrm{id}}$, the point mass at the identity.*

We omit the proof of the censoring inequality. A proof can be found either in [6] or in [10].

5.2. *Proof of Theorem* 1, *part* (*b*). In this subsection, we prove the second part of Theorem 1. Implementing the ingredients we obtained in the previous sections, the proof follows similarly as in the case of random AT shuffle [6].

To this end, we first explain the projection of measures on S_n which serves as a pretty tool to understand the mixing clearly. After that, we describe the main ideas of the proof. Some of the details will be presented in the Appendix.

Let K be a fixed integer and define $x_i := \lfloor \frac{in}{K} \rfloor$ for all $i = 0, 1, \dots, n$. Following the notation in [6], we define the functions $\widehat{\sigma}$ and $\bar{\sigma}$ for each $\sigma \in S_n$ and the sets $\widehat{S}_n$, $\bar{S}_n$ by

$$\widehat{\sigma} : [n] \times [K] \to \mathbb{R}, \qquad \widehat{\sigma}(x, j) := \widetilde{\sigma}(x, x_j), \qquad \widehat{S}_n := \{\widehat{\sigma} : \sigma \in S_n\};$$
$$\bar{\sigma} : [K] \times [K] \to \mathbb{R}, \qquad \bar{\sigma}(i, j) := \widetilde{\sigma}(x_i, x_j), \qquad \bar{S}_n := \{\bar{\sigma} : \sigma \in S_n\}.$$

That is, we are intentionally forgetting information from $\widetilde{\sigma}$ by projecting it to a smaller domain. For a probability measure ν on S_n, we similarly define the measure $\widehat{\nu}$ (resp., $\bar{\nu}$) on $\widehat{S}_n$ (resp., $\bar{S}_n$) by

$$\widehat{\nu}(\widehat{\sigma}) := \sum_{\xi : \widehat{\xi} = \widehat{\sigma}} \nu(\xi); \qquad \bar{\nu}(\bar{\sigma}) := \sum_{\xi : \bar{\xi} = \bar{\sigma}} \nu(\xi).$$

Furthermore, we introduce one more notation which is closely related to the projection $\widehat{\nu}$. Let T_n be the subset of S_n defined as

$$T_n := \big\{\sigma \in S_n : \sigma\big(\{x_{i-1} + 1, \dots, x_i\}\big) = \{x_{i-1} + 1, \dots, x_i\} \text{ for all } i \in [K]\big\}.$$

It is clear that $|T_n| = \prod_{i=1}^{K} (\Delta x_i)!$, where $\Delta x_i := x_i - x_{i-1}$. For a probability measure ν on S_n, the probability measure ν^u is defined by

$$\nu^u(\sigma) := \frac{1}{|T_n|} \sum_{\tau \in T_n} \nu(\tau \circ \sigma).$$

Therefore, ν^u becomes an invariant measure under composing an element of T_n. In other words, it is locally uniformized in the sense that permuting the label $\sigma(x) \in (x_{i-1}, x_i]$ within the same interval $(x_{i-1}, x_i]$ does not affect its probability. In addition, note that for any $\sigma \in S_n$ and $\tau \in T_n$, $\widehat{\sigma} = \widehat{\tau \circ \sigma}$. Based on this observation, we can deduce a connection between $\widehat{\nu}$ and ν^u.

LEMMA 20 ([6], Lemma 4.3). *Let μ denote the uniform measure on S_n. For any probability measure ν on S_n,*

$$\|\widehat{\nu} - \widehat{\mu}\|_{\text{T.V.}} = \|\nu^u - \mu\|_{\text{T.V.}}.$$

PROOF. The lemma readily follows from the above observation. Since ν^u is constant on $\{\sigma : \widehat{\sigma} = \widehat{\xi}\}$ for each fixed $\widehat{\xi} \in \widehat{S}_n$, we have

$$\begin{aligned}\sum_{\sigma} |\nu^u(\sigma) - \mu(\sigma)| &= \sum_{\widehat{\xi} \in \widehat{S}_n} \Big| \sum_{\sigma : \widehat{\sigma} = \widehat{\xi}} \big(\nu^u(\sigma) - \mu(\sigma)\big) \Big| \\ &= \sum_{\widehat{\xi} \in \widehat{S}_n} \Big| \sum_{\sigma : \widehat{\sigma} = \widehat{\xi}} \big(\nu(\sigma) - \mu(\sigma)\big) \Big| = \sum_{\widehat{\xi} \in \widehat{S}_n} |\widehat{\nu}(\widehat{\xi}) - \widehat{\mu}(\widehat{\xi})|. \qquad \square\end{aligned}$$

In order to establish the main theorem, we will introduce a censoring scheme $\mathcal{C}$, and show that the censored dynamics indeed mixes in the desired time, and hence implying the mixing of the original chain by the censoring inequality. We follow [6] for the construction of $\mathcal{C}$, while the proofs for each step rely on different ingredients to fit with the CAT shuffle.

Let $\eta > 0$ be a small fixed constant, set $K := \lfloor \eta^{-1} \rfloor$ and let $x_i := \lfloor in/K \rfloor$ as before. Define the censoring scheme $\mathcal{C} : \mathbb{N} \to \mathcal{P}([n-1])$ by

$$\mathcal{C}(t) = \begin{cases} \{x_i : i \in [K-1]\}, & \text{if } t \in [0, t_1] \cup [t_2, t_3]; \\ \varnothing & \text{if } t \in (t_1, t_2), \end{cases}$$

where the times t_1, t_2 and t_3 are given by

$$t_1 := \left(\frac{\eta}{3}\right) \frac{n^3}{2\pi^2} \log n; \qquad t_2 := \left(1 + \frac{2\eta}{3}\right) \frac{n^3}{2\pi^2} \log n; \qquad t_3 := (1+\eta) \frac{n^3}{2\pi^2} \log n.$$

In other words, in the first and the third steps, we ignore the updates happening at edges $(x_i, x_i + 1)$ for all $i \in [K-1]$, while running the chain without censoring in the second phase. Thus, in the first and the third steps, the chain operates separately at each interval $(x_{i-1}, x_i]$, while being dependent from each other since they share the directions of exploration $\{\mathbf{c}_l\}$. What happens in the censored shuffle can intuitively be described as follows (also see Figure 3):

(1) At time t_1, the cards in the same interval $(x_{i-1}, x_i]$ are distributed nearly uniformly, hence becoming indistinguishable. Thus, we can label all cards in $(x_{i-1}, x_i]$ by the *same index* i.

(2) At time t_2, cards with different indices get mixed, and each interval $(x_{i-1}, x_i]$ contains approximately equal number of cards of index j for all j. However, the locations within $(x_{i-1}, x_i]$ of the cards of different indices might not be uniform.

(3) After time t_3, within each interval $(x_{i-1}, x_i]$, the placement of cards of different indices become almost uniform.

Let us denote the uniform measure by μ as before, and define

$$\nu_t := P^t_{\mathrm{id},\mathcal{C}},$$

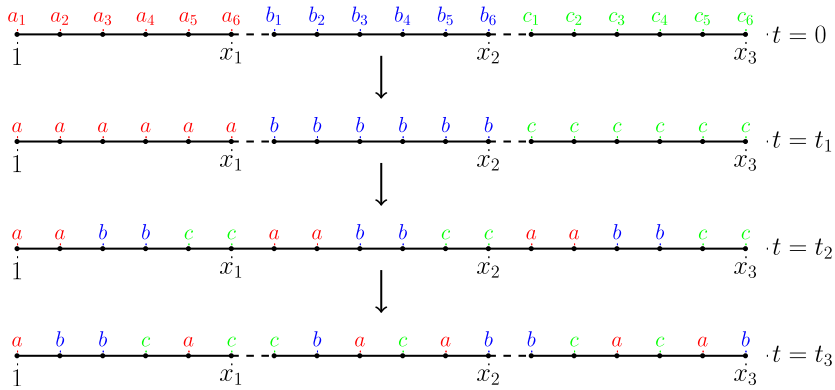

FIG. 3. *An illustration of mixing divided into three steps. Dashed edges indicate the censored updates during the first and the third phases.*

the probability distribution of the censored CAT shuffle at time t under the censoring scheme $\mathcal{C}$ which started from the initial state id. Then by Lemma 20,

$$\|\nu_t - \mu\|_{\text{T.V.}} \le \left\|\nu_t - \nu_t^u\right\|_{\text{T.V.}} + \|\widehat{\nu}_t - \widehat{\mu}\|_{\text{T.V.}}. \tag{5.3}$$

Having (5.3) in mind, we will establish mixing in terms of $\|\nu_t - \nu_t^u\|$ and $\|\widehat{\nu}_t - \widehat{\mu}\|$ as follows.

PROPOSITION 21. *For any given $\eta, \varepsilon > 0$, the following holds for all large enough n and all $t > t_1$:*

$$\left\|\nu_t - \nu_t^u\right\|_{T.V.} \le \varepsilon/3.$$

PROPOSITION 22. *For any given $\eta, \varepsilon > 0$, the following holds for all large enough n:*

$$\|\widehat{\nu}_{t_3} - \widehat{\mu}\|_{T.V.} \le 2\varepsilon/3.$$

PROOF OF THEOREM 1, PART (B) FROM PROPOSITIONS 21–22. The censoring inequality and equation (5.3) implies that

$$\left\|P_{\text{id}}^{t_3} - \mu\right\|_{\text{T.V.}} \le \|\nu_{t_3} - \mu\|_{\text{T.V.}} \le \left\|\nu_{t_3} - \nu_{t_3}^u\right\|_{\text{T.V.}} + \|\widehat{\nu}_{t_3} - \widehat{\mu}\|_{\text{T.V.}} \le \varepsilon.$$

Therefore, the mixing time $t_{\text{mix}}(\varepsilon)$ of the CAT shuffle satisfies

$$t_{\text{mix}}(\varepsilon) \le t_3 = (1+\eta)\frac{n^3}{2\pi^2}\log n,$$

where $\eta > 0$ can be taken arbitrarily small as n tends to infinity. $\square$

PROOF OF PROPOSITION 21. Our approach will be essentially the same as Lemma 6. Let σ_t be the state at time t under performing the censoring of $\mathcal{C}$ with

initial condition $\sigma_0 = \mathrm{id}$. Since the cards can only move within each intervals $\{x_{i-1}+1, \dots, x_i\}$ for $i \in [K-1]$ until time t_1, we have $\sigma_t \in T_n$ for all $t \le t_1$. This implies that

$$\nu_t^u = \mathbf{1}_{\mathrm{id}}^u,$$

where $\mathbf{1}_{\mathrm{id}}$ is the point mass at id. Moreover, $\mathbf{1}_{\mathrm{id}}^u$ is the stationary distribution of our chain until time t_1. Therefore, by the coupling inequality,

$$\|\nu_{t_1} - \nu_{t_1}^u\|_{\mathrm{T.V.}} = \|\nu_{t_1} - \mathbf{1}_{\mathrm{id}}^u\|_{\mathrm{T.V.}} \le \max_{\tau,\tau' \in T_n} \mathbb{P}\big(\sigma_{t_1}^\tau \ne \sigma_{t_1}^{\tau'}\big), \tag{5.4}$$

where σ_t^τ denotes the censored chain with initial condition $\sigma_0^\tau = \tau$. Note that the inequality holds for any coupling $(\sigma_t^\tau, \sigma_t^{\tau'})$. Let $\tau, \tau' \in T_n$ be arbitrary and for each $a \in [n]$, define

$$\widetilde{T}_a := \min\big\{t \ge 0 : \big(\sigma_t^\tau\big)^{-1}(a) = \big(\sigma_t^{\tau'}\big)^{-1}(a)\big\}$$

to be the coupling time of the card a in both decks. In order to estimate the decay of the coupling time between σ_t^τ and $\sigma_t^{\tau'}$, we adopt the coupling which differs from the monotone coupling in Section 5.1. This can be described as follows:

At time $(n-1)s$ for each $s = 0, 1, 2, \dots$, we choose the same orientation of exploration in both decks. At time t, let $(x, x+1)$ denote the edge that we are about to swap or not.

(1) If $\sigma_t(x) = \sigma_t'(x+1)$ or $\sigma_t(x+1) = \sigma_t'(x)$ then we do opposite moves. In other words, we pick either σ_t or σ_t' uniformly at random and swap the cards at positions $x, x+1$ of the chosen one while leaving the other fixed.

(2) Otherwise, we do identical moves; we either transpose the cards at $x, x+1$ for both σ_t and σ_t' or do nothing for both of them, each with probability $1/2$.

This rule ensures that once a specific card is in the same position in both decks, then it will remain matched forever. Thus, if $a \in (x_{i-1}, x_i]$ and $\tau^{-1}(a) \le (\tau')^{-1}(a)$, then $\widetilde{T}_a$ is bounded by the hitting time T_a defined as

$$T_a := \min\big\{t \ge 0 : \big(\sigma_t^\tau\big)^{-1}(a) = x_i\big\}.$$

Therefore, we are in the identical situation as Lemma 6, except that the length of the interval which the process $(\sigma_t^\tau)^{-1}(a)$ can move around is now $\Delta x_i \le \lfloor \eta n \rfloor + 1$. Therefore, Lemma 6 gives that

$$\mathbb{P}(T_a > t_1) \le \big(1 + O\big(n^{-1}\big)\big) \exp\Big(-\frac{\pi^2}{5\eta^2 n^3} t_1\Big),$$

and by a union bound over all $a \in [n]$ we obtain that

$$\mathbb{P}\big(\sigma_{t_1}^\tau \ne \sigma_{t_1}^{\tau'}\big) \le \sum_{a=1}^n \mathbb{P}(T_a > t_1) \le n \exp\Big(-\frac{\pi^2}{15\eta} \log n\Big) \le \varepsilon/3,$$

for all large enough n. Combining with (5.4) implies the desired result. $\square$

REMARK 23. In [6], where they study the random AT shuffle, the censored shuffle under the same censoring scheme $\mathcal{C}$ during time $[0, t_1]$ simplifies to the product chain of K copies of independent random AT shuffle on $(x_{i-1}, x_i]$. Thus, they prove Proposition 21 using this fact without introducing the above coupling.

In order to prove Proposition 22, we need the following observation.

PROPOSITION 24. *For any given $\eta, \varepsilon > 0$, the following holds for all large enough n:*

$$\|\bar{\nu}_{t_2} - \bar{\mu}\|_{T.V.} \le \varepsilon/3.$$

PROOF. We define the function $h : S_n \to \mathbb{R}$ to be

$$h(\sigma) := \sum_{i,j=1}^{K-1} \bar{\sigma}(i, j).$$

Then for any increasing probability measure ν on S_n, we have the following lemma from [6] which tells us how the expected value $\nu(h)$ controls the distance $\|\nu - \mu\|_{\mathrm{T.V.}}$ from the uniform measure.

LEMMA 25 ([6], Lemma 5.5). *Let ν be an increasing probability measure on S_n. For all $\varepsilon > 0$, there exists a constant $\gamma(K, \varepsilon) > 0$ such that for all sufficiently large n,*

$$\nu(h) \le \gamma\sqrt{n} \quad \textit{implies} \quad \|\bar{\nu} - \bar{\mu}\|_{\mathrm{T.V.}} \le \varepsilon/3.$$

Lemma 25 stems from the observation that if $\sigma \sim \mu$, then $n^{-1/2}\bar{\sigma}(i, j)$ converges to a Gaussian distribution as n tends to infinity. Due to this fact, one can show that if $\nu(\bar{\sigma}(i, j))$ is less than a small constant times $\sqrt{n}$, then the distance between $\bar{\mu}_{i,j}$ and $\bar{\nu}_{i,j}$ is accordingly small, where $\bar{\nu}_{i,j}$ (resp., $\bar{\mu}_{i,j}$) denotes the distribution of $\bar{\sigma}(i, j)$ under $\sigma \sim \nu$ (resp., $\sigma \sim \mu$). The function h combines the information for all i, j.

Due to Lemma 17, $\nu_{t_2}(h)$ can be bounded by $\gamma\sqrt{n}$, and hence we can apply Lemma 25 to obtain the desired inequality. Letting $\delta = \eta/7$ in Lemma 17, we have

$$\nu_{t_2}(h) \le P_{\mathrm{id}}^{t_2 - t_1}(h) \le n(K-1)^2 \exp\left(-\left(1 - \frac{\eta}{6}\right)\frac{\pi^2}{n^3}(t_2 - t_1)\right) \le \gamma\sqrt{n},$$

where the last inequality holds for any fixed $\gamma > 0$ when n is large enough. Moreover, since $\mathbf{1}_{\mathrm{id}}$ is increasing, Lemma 16 implies that ν_{t_2} is also increasing. Therefore, Lemma 25 tells us that

$$\|\bar{\nu}_{t_2} - \bar{\mu}\|_{\mathrm{T.V.}} \le \varepsilon/3. \qquad \square$$

Now we conclude the proof of Proposition 22. We again rely on the ideas in the proof of Proposition 21 and then have the following.

PROOF OF PROPOSITION 22. Let σ_t be the state of the censored CAT shuffle at time t. Due to our censoring scheme, we have

$$\sigma_t\big(\{x_{i-1}+1,\dots,x_i\}\big)=\sigma_{t_2}\big(\{x_{i-1}+1,\dots,x_i\}\big)\quad\text{for all } i\in[K],\, t\in[t_2,t_3].$$

Therefore, the stationary distribution $\mu_{\sigma_{t_2}}$ for the chain during time $t\in[t_2,t_3]$ can be written as

$$\mu_{\sigma_{t_2}}(\cdot):=\mu\big(\cdot\mid\sigma\big(\{x_{i-1}+1,\dots,x_i\}\big)=\sigma_{t_2}\big(\{x_{i-1}+1,\dots,x_i\}\big),\forall i\in[K]\big).$$

(Note the difference between $\mu_{\sigma_{t_2}}$ and $\mathbf{1}^{u}_{\sigma_{t_2}}$; the former uniformizes over the positions $x\in(x_{i-1},x_i]$ while the latter uniformizes over the labels $\sigma(x)\in(x_{i-1},x_i]$.) Thus, the same coupling argument in Proposition 21 implies that

$$\|\nu_{t_3}(\cdot\mid\sigma_{t_2})-\mu_{\sigma_{t_2}}\|_{\text{T.V.}}\le\varepsilon/3, \tag{5.5}$$

where $\nu_{t_3}(\cdot\mid\sigma_{t_2})$ denotes the probability distribution of σ_{t_3} given that it was at state σ_{t_2} at time t_2. For arbitrary $\xi\in\bar{S}_n$, we average the inequality (5.5) on the event $\{\bar\sigma=\xi\}$ to obtain that

$$\begin{aligned}&\sum_{\sigma_{t_2}:\bar\sigma_{t_2}=\xi}\nu_{t_2}(\sigma_{t_2}|\bar\sigma_{t_2}=\xi)\|\nu_{t_3}(\cdot\mid\sigma_{t_2})-\mu_{\sigma_{t_2}}\|_{\text{T.V.}}\\&\qquad\ge\|\nu_{t_3}(\cdot\mid\bar\sigma_{t_2}=\xi)-\mu(\cdot\mid\bar\sigma=\xi)\|_{\text{T.V.}}.\end{aligned}$$

Thus, by taking projections and using (5.5) we have

$$\|\widehat{\nu}_{t_3}(\cdot\mid\bar\sigma_{t_2}=\xi)-\widehat{\mu}(\cdot\mid\bar\sigma=\xi)\|_{\text{T.V.}}\le\varepsilon/3. \tag{5.6}$$

Following the computation in Proposition 5.3 of [6], this implies that

$$\begin{aligned}\sum_{\widehat\sigma\in\widehat{S}_n}|\widehat{\nu}_{t_3}(\widehat\sigma)-\widehat{\mu}(\widehat\sigma)|&\le\sum_{\xi\in\bar{S}_n}\sum_{\widehat\sigma:\bar\sigma=\xi}|\widehat{\nu}_{t_3}(\widehat\sigma)-\widehat{\mu}(\widehat\sigma)|\\&\le\sum_{\xi\in\bar{S}_n}\sum_{\widehat\sigma:\bar\sigma=\xi}\big(\bar\nu_{t_2}(\xi)|\widehat{\nu}_{t_3}(\widehat\sigma\mid\bar\sigma_{t_2}=\xi)-\widehat{\mu}(\widehat\sigma\mid\bar\sigma=\xi)|\\&\qquad+\bar\mu(\widehat\sigma\mid\widehat\sigma=\xi)|\bar\nu_{t_2}(\xi)-\bar\mu(\xi)|\big)\\&\le\frac{2\varepsilon}{3}+\frac{2\varepsilon}{3}\le\frac{4\varepsilon}{3},\end{aligned}$$

where the last line follows from (5.6) and Proposition 24. □

6. Application to the systematic simple exclusion process. In this section, we study the systematic simple exclusion process using the techniques developed from the previous chapters. We show that the mixing time of this process satisfies a similar bound as Theorem 1.

The systematic simple exclusion process can be understood as a projection of the CAT shuffle. To define the model, consider we have a length $(n-1)$ path on $\{1,\dots,n\}$ and locate $k \le n$ particles at vertices, with each vertex being occupied by at most one particle. We introduce the dynamics similar to the CAT shuffle: At the beginning, we pick either 1 or n uniformly at random. If 1 is chosen, then at time $t \in \{1,\dots,n-1\}$ we update the edge $(t, t+1)$, meaning that we either swap the possessions of the endpoints of the edge or leave it stay fixed, each with probability $\frac{1}{2}$. If n is chosen, we explore in the opposite direction. After updating all $(n-1)$ edges, we again choose a random initial location out of $\{1, n\}$ and continue the systematic updates starting from the chosen point. In other words, it is the projection of the CAT shuffle which regards k cards as particles and the rest as empty sites.

Using the argument from previous sections, we have the following mixing time bound for the systematic simple exclusion process.

THEOREM 26. *Consider the systematic simple exclusion process on the line $\{1,\dots,n\}$ with $k(n)$ particles such that both k and $n-k$ tends to infinity as $n \to \infty$. Let $k' := \min\{k, n-k\}$. Then for any $\varepsilon > 0$, we have*

(a) $t_{\mathrm{mix}}(1-\varepsilon) \ge \frac{n^3}{2\pi^2}\log k' - \frac{n^3}{2\pi^2}\log(\frac{c\log k'}{\varepsilon})$, *where c is a universal constant.*
(b) $t_{\mathrm{mix}}(\varepsilon) \le (1+o(1))\frac{4n^3}{\pi^2}\log k'$.

REMARK 27. We conjecture that the lower bound of Theorem 26 is sharp, that is, the systematic simple exclusion process should exhibit cutoff at $t_{\mathrm{mix}}(\varepsilon) = (1+o(1))\frac{n^3}{2\pi^2}\log k'$. The main difficulty of improving (b) of Theorem 26 stems from the deterministic aspects of the update rule. For instance, in [6] where cutoff for the simple exclusion process is established, the problem can be reduced to analyzing simple random walks. However, in the systematic case, the increments of the random walks corresponding to those derived in [6] are heavily correlated, which makes it more difficult to study.

PROOF. We can assume that $k \le \frac{n}{2}$, since in the other case we can swap the roles of empty sites and particles. Let $\Omega_{n,k} := \{\xi \in \{0,1\}^n : \sum_{x=1}^n \xi(x) = k\}$ be the state space of the chain, where $\xi(x) = 1$ (resp., $\xi(x) = 0$) indicates that position x is occupied (resp., empty).

To prove part (a), We consider the following height function for each $\xi \in \Omega_{n,k}$:

$$g_\xi(x) := \sum_{z=1}^{x} \xi(z) - \frac{xk}{n}. \tag{6.1}$$

Using the height function, define

$$\Psi(\xi) := \sum_{x=1}^{n} g_\xi(x) \sin\left(\frac{\pi x}{n}\right). \tag{6.2}$$

We additionally define $\wedge_t$ to be the state at time t of the systematic simple exclusion process with initial condition that has particles in the first k positions (i.e., $\wedge(x) = \mathbf{1}_{\{x \le k\}}$ for all x), and let

$$\Psi_t := \Psi(\wedge_{(n-1)t}). \tag{6.3}$$

Then the following is a straightforward generalization of Lemmas 3 and 5.

LEMMA 28. *Let Ψ_t be defined as (6.3). For any $t \in \mathbb{N}$, we have*

(a) $|\mathbb{E}[\Psi_{t+1}|\mathcal{F}_t] - (1-\gamma)\Psi_t| \le \frac{3\pi}{4n}$, *where* $\gamma := \pi^2/n^2 - O(n^{-4})$.
(b) $\mathbb{E}[(\Delta\Psi_t)^2|\mathcal{F}_t] \le Ck\log k$, *where $C > 0$ is a universal constant.*

We can then conclude the proof of the lower bound by following the approach of Section 3.2, based on Lemmas 2 and 28.

To prove the upper bound, we consider two copies ξ_t^1, ξ_t^2 of systematic simple exclusion processes with different initial configurations, and estimate their coupling time using Lemma 6. To be specific, we first label the k particles arbitrarily in both chains, and consider the coupling introduced in the proof of Proposition 21. For each i, the coupling time of the ith particle in ξ_t^1 and ξ_t^2 is bounded by the hitting time of the left particle reaching at the right end of the deck. Therefore, if we call the latter quantity T_i, then for any $\varepsilon, \delta > 0$ and $t = (1+\delta)\frac{4n^3}{\pi^2}\log k$, Lemma 6 implies that

$$\max_{\xi_0^1, \xi_0^2 \in \Omega_n} \mathbb{P}(\xi_t^1 \ne \xi_t^2) \le \sum_{i=1}^{k} \mathbb{P}(T_i > t) \le \varepsilon,$$

for all sufficiently large n. $\square$

APPENDIX

A.1. The decay estimate: Proof of Lemma 17.

LEMMA 17. *Let (σ_t) denote the CAT shuffle starts from an arbitrary initial state and let $\delta > 0$ be arbitrary. Then there exist $N_\delta, \theta_\delta > 0$ such that for any*

$x, y \in [n]$, $n \geq N_\delta$ and $t > \theta_\delta n^3$ satisfying $t = O(n^4)$, we have

$$|\mathbb{E}[\widetilde{\sigma}_t(x, y)]| \leq n\big(1 + O(tn^{-5})\big) \exp\Big(-(1-\delta)\frac{\pi^2}{n^3}t\Big). \tag{A.1}$$

PROOF. Let $y \in [n]$ be given and let t be of the form $t = (n-1)i$ for $i \in \mathbb{N}$. Assume that $y < n$ (otherwise we have nothing to prove) and set $\Delta := n - 1$. We analyze the expected difference between $\widetilde{\sigma}$ at time $t + \Delta$ and t given the information $\mathcal{F}_t$ until time t. Recall Definition 13, where we defined the random variables U_s, $\mathbf{c}_i$ and the update rules using them. Notice that between time t and $t + \Delta$, $\widetilde{\sigma}_s(x, y)$ can only be changed when updating the edge $(x, x+1)$. Also, when update is performed at edge $(x, x+1)$ at time s, it goes up by 1 if $\sigma_s(x) > y \geq \sigma_s(x+1)$ and $U_s = 1$, whereas it moves down by 1 if $\sigma_s(x) \leq y < \sigma_s(x+1)$ and $U_s = 0$.

Set $v_t(x) := \widetilde{\sigma}_t(x, y)$. We compute $\mathbb{E}[v_{t+\Delta}(x) - v_t(x) \mid \mathcal{F}_t]$ based on the above properties of $\widetilde{\sigma}$, by considering the cases $\mathbf{c}_i = 1$ and $\mathbf{c}_i = n$ separately. If $2 \leq x \leq n-2$, we have

$$\begin{aligned}
&\mathbb{E}\big[v_{t+\Delta}(x) - v_t(x) \mid \mathcal{F}_t, \mathbf{c}_i = 1\big] \\
&\quad = \sum_{k=0}^{x-2} \frac{1}{2^{k+2}} \big(\mathbf{1}_{\{\sigma_t(x+1) \leq y < \sigma_t(x-k)\}} - \mathbf{1}_{\{\sigma_t(x+1) > y \geq \sigma_t(x-k)\}}\big) \\
&\qquad + \frac{1}{2^x}\big(\mathbf{1}_{\{\sigma_t(x+1) \leq y < \sigma_t(1)\}} - \mathbf{1}_{\{\sigma_t(x+1) > y \geq \sigma_t(1)\}}\big), \\
&\mathbb{E}\big[v_{t+\Delta}(x) - v_t(x) \mid \mathcal{F}_t, \mathbf{c}_i = n\big] \\
&\quad = \sum_{k=0}^{n-x-2} \frac{1}{2^{k+2}} \big(\mathbf{1}_{\{\sigma_t(x+1+k) \leq y < \sigma_t(x)\}} - \mathbf{1}_{\{\sigma_t(x+1+k) > y \geq \sigma_t(x)\}}\big) \\
&\qquad + \frac{1}{2^{n-x}}\big(\mathbf{1}_{\{\sigma_t(n) \leq y < \sigma_t(x)\}} - \mathbf{1}_{\{\sigma_t(n) > y \geq \sigma_t(x)\}}\big).
\end{aligned} \tag{A.2}$$

Notice the following relation between the indicators:

$$\begin{aligned}
&\mathbf{1}_{\{\sigma(x_1) \leq y < \sigma(x_2)\}} - \mathbf{1}_{\{\sigma(x_1) > y \geq \sigma(x_2)\}} \\
&\quad = \mathbf{1}_{\{\sigma(x_1) \leq y\}} - \mathbf{1}_{\{\sigma(x_1), \sigma(x_2) \leq y\}} = \mathbf{1}_{\{\sigma(x_2) \leq y\}} + \mathbf{1}_{\{\sigma(x_1), \sigma(x_2) \leq y\}} \\
&\quad = \mathbf{1}_{\{\sigma(x_1) \leq y\}} - \mathbf{1}_{\{\sigma(x_2) \leq y\}} \\
&\quad = \widetilde{\sigma}(x_1, y) - \widetilde{\sigma}(x_1 - 1, y) - \widetilde{\sigma}(x_2, y) + \widetilde{\sigma}(x_2 - 1, y),
\end{aligned}$$

where we define $\widetilde{\sigma}(0, y) := 0$. This property implies that

$$
\begin{aligned}
&\mathbb{E}\big[v_{t+\Delta}(x) - v_t(x) \mid \mathcal{F}_t\big] \\
&\quad = \sum_{k=0}^{x-2} \frac{1}{2^{k+3}}\big\{v_t(x+1) - v_t(x) - v_t(x-k) + v_t(x-k-1)\big\} \\
&\qquad + \frac{1}{2^{x+1}}\big\{v_t(x+1) - v_t(x) - v_t(1)\big\} \\
&\qquad + \sum_{k=0}^{n-x-2} \frac{1}{2^{k+3}}\big\{v_t(x+1+k) - v_t(x+k) - v_t(x) + v_t(x-1)\big\} \\
&\qquad + \frac{1}{2^{n-x+1}}\big\{-v_t(n-1) - v_t(x) + v_t(x-1)\big\}.
\end{aligned} \tag{A.3}
$$

Letting $\bar{v}_t(x) = \mathbb{E}[v_t(x)]$, taking expectations on both sides of (A.3) and rearranging the terms in the right-hand side, we have that for each $2 \le x \le n-2$,

$$
\bar{v}_{t+\Delta}(x) = \sum_{k=-1}^{n-x-2} \frac{\bar{v}_t(x+k)}{2^{k+3}} + \sum_{k=-1}^{x-2} \frac{\bar{v}_t(x-k)}{2^{k+3}}. \tag{A.4}
$$

Similar calculations for $x = 1$ and $x = n-1$ yield that

$$
\begin{aligned}
\bar{v}_{t+\Delta}(1) &= \frac{1}{8}\bar{v}_t(1) + \frac{1}{4}\bar{v}_t(2) + \sum_{k=2}^{n-2} \frac{\bar{v}_t(k)}{2^{k+2}}; \\
\bar{v}_{t+\Delta}(n-1) &= \frac{1}{8}\bar{v}_t(n-1) + \frac{1}{4}\bar{v}_t(n-2) + \sum_{k=2}^{n-2} \frac{\bar{v}_t(n-k)}{2^{k+2}}.
\end{aligned} \tag{A.5}
$$

Due to the monotonicity of $\widetilde{\sigma}(x, y)$ in terms of σ, it suffices to prove the desired inequality (A.1) for the initial condition $\sigma_0 = \mathrm{id}$ which is the maximal case. The minimal case with initial state $\sigma^-(z) = n + 1 - z$ is also included in the maximal case; the only differences are the sign and taking $\widetilde{\sigma}^-(\cdot, n-y)$ instead of $\widetilde{\sigma}(\cdot, y)$.

Thus, let us assume that $\sigma_0 = \mathrm{id}$. In order to establish the main inequality (A.1), we will introduce $u_s : [n] \to \mathbb{R}$ which satisfies $u_s(x) \ge \bar{v}_{\Delta s}(x)$ and the bound

$$
\|u_s\|_\infty \le n\big(1 + O\big(sn^{-4}\big)\big)\exp\bigg(-(1-\delta)\frac{\pi^2}{n^2}s\bigg).
$$

Let $u_0(x) := \bar{v}_0(x)$, $u_s(0) = u_s(n) = 0$ and define $u_{s+1}(x)$ to follow (A.4) so that

$$
u_{s+1}(x) = \sum_{k=-1}^{n-x-1} \frac{u_s(x+k)}{2^{k+3}} + \sum_{k=-1}^{x-1} \frac{u_s(x-k)}{2^{k+3}}, \tag{A.6}
$$

for each $x \in [n]$ and $j \in \mathbb{N}$. Note the difference between $\bar{v}_s$ and u_s as $\bar{v}_{s+1}(x)$ satisfies (A.4) only for $2 \le x \le n-2$. Since $u_0 = \bar{v}_0$ is positive and the coefficients in (A.6) are at least as large as those in (A.4) and (A.5), we have $u_s \ge \bar{v}_s$ for all s.

Furthermore, we define $d_s : [n] \to \mathbb{R}$ by

$$d_s(x) := u_s(x) - u_s(x-1). \tag{A.7}$$

We analyze d_s instead of u_s since it has a tractable initial condition. Indeed, note that $\|d_0\|_\infty \le 1$, which is much smaller compared to $\|u_0\|_\infty \asymp n$. Also, note the obvious inequality that

$$\|u_s\|_\infty \le \|d_s\|_1. \tag{A.8}$$

Based on (A.6), we compute the transition rule of d_s as follows:

$$d_{s+1}(x) = \sum_{k=-1}^{n-x} \frac{d_s(x+k)}{2^{k+3}} + \sum_{k=-1}^{x-1} \frac{d_s(x-k)}{2^{k+3}}. \tag{A.9}$$

Therefore, one can observe that the equation (A.9) is equivalent to the transition rule of the random walk on $\mathbb{Z}$ that has i.i.d. increments $X_j \sim X$ with

$$\mathbb{P}(X = k) = \frac{1}{2^{|k|+3}} + \frac{1}{2^{-|k|+3}} \mathbf{1}_{\{|k| \le 1\}},$$

and that dies out when reaching outside of $[n]$.

Let $S_m^x := x + \sum_{j=1}^m X_j$ be the symmetric random walk on $\mathbb{Z}$ with i.i.d. $X_j \sim X$ that starts at x, and let $\widehat{\tau}_n^x := \min\{m \ge 0 : S_m^x \notin [n]\}$ be its first exit time from $[n]$. For each $l \in [n]$, let $d_s^{(l)}$ denote the vector such that $d_0^{(l)} = \mathbf{1}_{\{l\}}$ and follows the transition rule (A.9). Then we have

$$\|d_s^{(l)}\|_1 = \sum_{x=1}^n d_s^{(l)}(x) = \mathbb{P}(\widehat{\tau}_n^l > s). \tag{A.10}$$

Thus, our goal is to bound the probability $\mathbb{P}(\widehat{\tau}_n^l > s)$, which can be done similarly as Lemmas 6 and 11. Since $\mathrm{Var}(X) = 2$, Donsker's theorem implies that for any $\delta > 0$, there exists N_δ such that

$$\mathbb{P}(\widehat{\tau}_n^{zn} > \theta n^2) \le \mathbb{P}(\widehat{\tau}_B^{z/\sqrt{2}} > \theta - \delta)$$

for all $n \ge N_\delta$, where $\widehat{\tau}_B^z$ is the first exit time from $[0, 1/\sqrt{2}]$ of the standard Brownian motion with initial position z. Notice that we already have computed the probability in the right-hand side in Lemma 11 and Corollary 12. According to these results, we obtain that for any constant $\theta > 0$,

$$\mathbb{P}(\widehat{\tau}_B^z > \theta) \le C \exp(-\pi^2 \theta),$$

for some absolute constant $C > 0$. (Although Lemma 11 is proven for the walk that starts at the midpoint of the given interval, generalization to the arbitrary starting location is straightforward.) Repeating the argument done in Corollary 12 and

Lemma 6, we deduce the following: For any $\delta > 0$, there exist $\theta_\delta > 0$ and $N_\delta > 0$ such that for all $\theta \geq \theta_\delta$, $n \geq N_\delta$ and $z \in [n]$,

$$\mathbb{P}(\tilde{\tau}_n^z > \theta n^2) \leq (1 + O(\theta n^{-2})) \exp(-(1-\delta)\pi^2\theta).$$

The original vector d_s can be written as

$$d_s = \sum_{l=1}^{n} d_0(l) \cdot d_s^{(l)}.$$

Since $\|d_0\|_\infty \leq 1$, we have

$$\|d_s\|_1 \leq \sum_{l=1}^{n} |d_0(l)| \|d_s^{(l)}\|_1 \leq n(1 + O(sn^{-4})) \exp\left(-(1-\delta)\frac{\pi^2}{n^2}s\right),$$

for all large $n > N_\delta$ and $\theta_\delta n^2 < s \leq n^3$. Therefore, we deduce the desired result by (A.8). $\square$

A.2. Proof of Lemma 11.

LEMMA 11 1. *Let $\bar{\tau}^n$ be the first time that the simple random walk on $\mathbb{Z}$ starting at the origin hits $\pm n$. There exists a constant $C > 0$ that satisfies $\mathbb{P}(\bar{\tau}^n > \theta n^2) < C(1 + O(\theta n^{-2})) \exp(-\pi^2\theta/8)$ for all $\theta > 0$ (θ may depend on n).*

REMARK 29. Lemma 11 is originally stated in terms of the hitting time at $\pm n/\sqrt{2}$. Here, we presented an equivalent statement regarding the hitting time at $\pm n$.

PROOF. Let $\widetilde{S}_m$ denote the simple random walk on $\mathbb{Z}$ that starts at the origin. By the definition of $\bar{\tau}^n$, it suffices to show the desired inequality for τ_+^n, where τ_+^n is the first time that $S_m^+ := |\widetilde{S}_m|$ hits n.

Let (Z_m) be the random walk on $\{0, 1, \dots, n-1\}$ that has the same jump rate as S_m^+ on $\{0, 1, \dots, n-2\}$, and that at $(n-1)$ jumps to $(n-2)$ with probability $1/2$ and stays fixed otherwise. Then one can notice that $\mathbb{P}(\tau_+^n > \theta n^2)$ is equal to the survival probability of $Z_{\theta n^2}$. We focus on computing the latter quantity.

Denote the transition matrix of (Z_m) by M_n and note that the matrix M_n is symmetric. For each $j = 0, 1, \dots, n-1$, let f_j be the n-dimensional vector defined by

$$f_j(x) = \cos\left(\frac{(2j+1)\pi x}{2n}\right) \quad \text{for all } x \in \{0, 1, \dots, n-1\}.$$

Observe that f_j's are the eigenvectors of M_n, particularly since $\cos(\frac{(2j+1)\pi x}{2n})$ becomes zero at $x = n$. The corresponding eigenvalues are given by

$$\lambda_j = \cos\left(\frac{(2j+1)\pi}{2n}\right) \quad \text{for all } j \in \{0, 1, \dots, n-1\}.$$

It is also straightforward to check that f_j's are orthogonal:

$$\sum_{x=0}^{n-1} f_j(x) f_k(x) = \frac{1}{2}\sum_{x=0}^{n-1} \cos\left(\frac{(j+k+1)\pi x}{n}\right) + \cos\left(\frac{(j-k)\pi x}{n}\right),$$

and the right-hand side is nonzero if and only if $j = k$. Therefore, $\{f_j\}$ forms an orthogonal basis of the space of n-dimensional vectors. Let δ_0 be the point mass at the origin. Elementary calculation yields that

$$\begin{aligned}
\mathbb{P}(\tau_+^n > t) &= \sum_{x=0}^{n-1} M_n^t \delta_0(x) = \sum_{x=0}^{n-1}\sum_{j=0}^{n-1} \frac{\delta_0 \cdot f_j}{f_j \cdot f_j} \lambda_j^t f_j(x) \\
&= \sum_{j=0}^{n-1} \frac{2}{n+1} \cos^t\left(\frac{(2j+1)\pi}{2n}\right) \sum_{x=0}^{n-1} \cos\left(\frac{(2j+1)\pi x}{2n}\right) \qquad \text{(A.11)} \\
&\le 4 \sum_{j=0}^{\lfloor n/2 \rfloor} \cos^t\left(\frac{(2j+1)\pi}{2n}\right),
\end{aligned}$$

where in the second line we used the identity $f_j \cdot f_j = \frac{n+1}{2}$. If we consider the line passing $(0, \cos 0)$ and $(\alpha, \cos\alpha)$ for $\alpha = \pi/2n$, it lies above $(z, \cos z)$ for $z \in [\pi/2n, \pi/2]$. Thus, we can bound λ_j^t by

$$\begin{aligned}
\cos^t\left(\frac{(2j+1)\pi}{2n}\right) &\le \left\{1 - (2j+1)\left(1 - \cos\left(\frac{\pi}{2n}\right)\right)\right\}^t \\
&\le \exp\left\{-t(2j+1)\left(1 - \cos\left(\frac{\pi}{2n}\right)\right)\right\}.
\end{aligned}$$

Hence, summation over $j = 0, \dots, n-1$ yields that

$$\begin{aligned}
\sum_{j=0}^{\lfloor n/2 \rfloor} \cos^t\left(\frac{(2j+1)\pi}{2n}\right) &\le \frac{\exp(-(1-\cos(\frac{\pi}{2n}))t)}{1 - \exp(-(1-\cos(\frac{\pi}{2n}))t)} \\
&\le \left(1 + O\left(\frac{t}{n^4}\right)\right) \frac{\exp(-\pi^2 t/8n^2)}{1 - \exp(-\pi^2 t/8n^2)}.
\end{aligned}$$

Therefore, combining with (A.11), we obtain that

$$\begin{aligned}
\mathbb{P}(\tau_+^n > \theta n^2) &\le \min\left\{1, (4 + O(\theta n^{-2})) \frac{\exp(-\pi^2\theta/8)}{1 - \exp(-\pi^2\theta/8)}\right\} \\
&\le (5 + O(\theta n^{-2})) \exp\left(-\frac{\pi^2}{8}\theta\right),
\end{aligned}$$

which is the desired result with $C = 5$. $\square$

A.3. Proof of Lemma 3.

LEMMA 3. *Let Φ_t defined as (3.2). For any $t \in \mathbb{N}$, we have*

$$\left|\mathbb{E}[\Phi_{t+1}|\mathcal{F}_t] - (1-\gamma)\Phi_t\right| \le \frac{4\pi}{3n},$$

where $\gamma := \pi^2/n^2 - O(n^{-4})$.

PROOF. According to the computations in (A.4, A.5), we have

$$\begin{aligned}
&\mathbb{E}[\Phi_{t+1}|\mathcal{F}_t] \\
&\quad = \sum_{x=2}^{n-2}\left[\sum_{k=-1}^{n-x-2}\frac{h_t(x+k)}{2^{k+3}} + \sum_{k=-1}^{x-2}\frac{h_t(x-k)}{2^{k+3}}\right]\sin\frac{\pi x}{n} \\
&\quad\quad + \left[\left(\sum_{k=0}^{n-3}\frac{h_t(1+k)}{2^{k+3}}\right) + \frac{1}{4}h_t(2)\right]\sin\left(\frac{\pi}{n}\right) \\
&\quad\quad + \left[\left(\sum_{k=0}^{n-3}\frac{h_t(n-1-k)}{2^{k+3}}\right) + \frac{1}{4}h_t(n-2)\right]\sin\left(\frac{\pi(n-1)}{n}\right).
\end{aligned} \tag{A.12}$$

By rearranging the right-hand side, we obtain that

$$\begin{aligned}
&\mathbb{E}[\Phi_{t+1}|\mathcal{F}_t] \\
&\quad = \sum_{y=2}^{n-2}\left[\sum_{k=-1}^{n-y-1}\frac{h_t(y)}{2^{k+3}}\sin\left(\frac{\pi(y+k)}{n}\right) + \sum_{k=-1}^{y-1}\frac{h_t(y)}{2^{k+3}}\sin\left(\frac{\pi(y-k)}{n}\right)\right] \\
&\quad\quad + \big(h(1)+h(n-1)\big)\left(\frac{1}{4}\sin\left(\frac{2\pi}{n}\right) + \frac{1}{8}\sin\left(\frac{\pi}{n}\right)\right) \\
&\quad = \left[\sum_{k=-1}^{\infty}\frac{\cos(\frac{\pi k}{n})}{2^{k+2}}\right]\Phi_t + \sum_{y=2}^{n-2}h(y)\left\{\sum_{k=1}^{\infty}\frac{\sin(\frac{\pi k}{n})}{2^{n-y+3+k}} + \sum_{k=1}^{\infty}\frac{\sin(\frac{\pi k}{n})}{2^{y+3+k}}\right\} \\
&\quad\quad - \big(h(1)+h(n-1)\big)\left[\sum_{k=1}^{\infty}\frac{3\sin(\frac{\pi k}{n})}{2^{k+4}}\right].
\end{aligned} \tag{A.13}$$

Noting that

$$\sum_{k=-1}^{\infty}\frac{\cos(\frac{\pi k}{n})}{2^{k+3}} = 1 - \frac{\pi^2}{n^2} + O\left(\frac{1}{n^4}\right) = 1-\gamma$$

as well as that $|h_t(x)| \le \frac{1}{2}x \wedge (n-x)$, we can deduce from (A.13) that

$$\begin{aligned} &\left|\mathbb{E}[\Phi_{t+1}|\mathcal{F}_t] - (1-\gamma)\Phi_t\right| \\ &\quad \le \sum_{y=2}^{\infty} y \left\{ \sum_{k=1}^{\infty} \frac{\pi}{n} \frac{k}{2^{y+3+k}} \right\} + \sum_{k=1}^{\infty} \frac{3k}{2^{k+4}} \frac{\pi}{n} = \frac{3\pi}{4n}. \end{aligned} \tag{A.14}$$

□

Acknowledgments. We are grateful to Yuval Peres and Allan Sly for their helpful comments.

REFERENCES

[1] ALDOUS, D. and DIACONIS, P. (1986). Shuffling cards and stopping times. *Amer. Math. Monthly* **93** 333–348. MR0841111

[2] ANGEL, O. and HOLROYD, A. E. (2018). Perfect shuffling by lazy swaps. *Electron. Commun. Probab.* **23** Paper No. 47, 11. MR3841408

[3] BLANCA, A., CAPUTO, P., SINCLAIR, A. and VIGODA, E. (2018). Spatial mixing and non-local Markov chains. In *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms* 1965–1980. SIAM, Philadelphia, PA. MR3775916

[4] DURRETT, R. (1996). *Probability: Theory and Examples*, 2nd ed. Duxbury Press, Belmont, CA. MR1609153

[5] KARATZAS, I. and SHREVE, S. E. (1998). *Brownian Motion and Stochastic Calculus*, 2nd ed. *Graduate Texts in Mathematics* **113**. Springer, New York.

[6] LACOIN, H. (2016). Mixing time and cutoff for the adjacent transposition shuffle and the simple exclusion. *Ann. Probab.* **44** 1426–1487. MR3474475

[7] MIRONOV, I. (2002). (Not so) random shuffles of RC4. In *Advances in Cryptology—CRYPTO* 2002. *Lecture Notes in Computer Science* **2442** 304–319. Springer, Berlin. MR2054828

[8] MORRIS, B., NING, W. and PERES, Y. (2014). Mixing time of the card-cyclic-to-random shuffle. *Ann. Appl. Probab.* **24** 1835–1849. MR3226165

[9] MOSSEL, E., PERES, Y. and SINCLAIR, A. (2004). Shuffling by semi-random transpositions. In *Proceeding FOCS* '04 *Proceedings of the* 45*th Annual IEEE Symposium on Foundations of Computer Science* 572–581.

[10] PERES, Y. and WINKLER, P. (2013). Can extra updates delay mixing? *Comm. Math. Phys.* **323** 1007–1016. MR3106501

[11] PINSKY, R. G. (2015). Probabilistic and combinatorial aspects of the card-cyclic to random insertion shuffle. *Random Structures Algorithms* **46** 362–390. MR3302902

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

[13] SALOFF-COSTE, L. and ZÚÑIGA, J. (2009). Merging for time inhomogeneous finite Markov chains. I. Singular values and stability. *Electron. J. Probab.* **14** 1456–1494. MR2519527

[14] SALOFF-COSTE, L. and ZÚÑIGA, J. (2010). Time inhomogeneous Markov chains with wave-like behavior. *Ann. Appl. Probab.* **20** 1831–1853. MR2724422

[15] SALOFF-COSTE, L. and ZÚÑIGA, J. (2011). Merging and stability for time inhomogeneous finite Markov chains. In *Surveys in Stochastic Processes. EMS Ser. Congr. Rep.* 127–151. Eur. Math. Soc., Zürich. MR2883857

[16] THORP, E. O. (1965). Classroom Notes: Attempt at a Strongest Vector Topology. *Amer. Math. Monthly* **72** 532–533. MR1533262

[17] WILSON, D. B. (2004). Mixing times of Lozenge tiling and card shuffling Markov chains. *Ann. Appl. Probab.* **14** 274–325. MR2023023

DEPARTMENT OF MATHEMATICS
PRINCETON UNIVERSITY
FINE HALL, WASHINGTON ROAD
PRINCETON, NEW JERSEY 08544-1000
USA
E-MAIL: dhnam@princeton.edu
exn@princeton.edu