

Mixing times via super-fast coupling.

Robert M. Burton
Department of Mathematics,
Oregon State University
Corvallis, OR
97331-4605, USA
bob@oregonstate.edu

Yevgeniy Kovchegov
Department of Mathematics,
Oregon State University
Corvallis, OR
97331-4605, USA
kovchegy@math.oregonstate.edu

Abstract

We provide a coupling proof that the transposition shuffle on a deck of n cards is mixing of rate $Cn \log n$ with a moderate constant, C . This rate was determined by Diaconis and Shahshahani, but the question of a natural probabilistic coupling proof has been missing, and questions of its existence have been raised. The proof, and indeed any proof, requires that we enlarge the methodology of coupling to include intuitive but non-adapted coupling rules, because a typical Markovian coupling is incapable of resolving finer questions of rates.

1 Introduction.

Random shufflings of a deck of n distinct cards are well-studied objects, and a frequent metaphor describing a class of Markov chains invariant with respect to the symmetric group, S_n . Here the focus is on transposition shuffling, one of the simplest shuffles, defined by uniformly sampling the deck twice with replacement, and then interchanging the positions of these cards, if they are different.

Clearly, as it is always the case for mixing finite state Markov chains, the distribution of the ordering of the deck converges in total variation to the invariant measure, which by symmetry is uniform on the permutation group, S_n . It is the rate of mixing which presently holds our interest, as well as the coupling methods by which one might attempt to show good upper bounds on this rate.

This is a well-defined problem in probability theory, and one would expect that a coupling argument would be the instrument of first choice. Indeed, there is such an approach, given in the online notes of Aldous and Fill [2] (also see [1]). This method gives a rate of $O(n^2)$, and will be discussed later in the article. Unfortunately, but necessarily, this rate is not the optimal rate [we expect, $O(n \log n)$, which was proven by Diaconis and Shahshahani [9] using methods from the relatively rarified mathematical residential district of representation theory]. This gap is apparent, and somewhat long-standing; indeed, Peres has listed the problem of showing the $O(n \log n)$ as the rate of uniform mixing, using a coupling approach,

as one of a number of interesting open problems (see [17]). The problem is also mentioned in Saloff-Coste [19] and other important publications. Here we solve this problem. Moreover, we deconstruct different ways of looking at this kind of problem, and enlarge the intuitions that might guide us in coupling.

We would like to mention that Matthews [15] gives a purely probabilistic proof of the $O(n \log n)$ bound using strong stationary times. Diaconis and Saloff-Coste [8] also treat a host of other transposition problems, and a survey of the many appearances of random transpositions can be found in Diaconis [7]. Among the publications related to this research, we would like to list the recent work of Berestycki and Durrett [3] on random transposition, and the paper of Hayes and Vigoda on a non-Markovian coupling for graph coloring problems.

Coupling comes from at least two fields. Of course, one is probability theory where, typically, the intent of the proof is to give an intuition of the phenomena and why it happens. For this reason, most proofs are Markovian and depend solely on information available to the present moment of the evolution of a Markov process or process derived from a Markov process such as a weak Bernoulli process. Then one could, for example, use coupling to show strong versions of the central limit theorem (see for example, [4]). Another field that coupling developed within is ergodic theory. Here, the problems had complicated forms of dependence and the coupling was non-intuitive, using the marriage lemma to their existence. See [16] and [10]. The result that any two weakly Bernoulli processes with the same entropy are isomorphic in a measure-theoretic translation invariant way was proven in such a way. Indeed, the choice of coupling depends essentially on the distinction between quantitative and qualitative. To show the ideas one would prefer Markovian coupling, even though this kind of coupling cannot be best possible and to get precise rates one would use non-Markovian coupling because no other kind would work. The history of research on non-Markovian coupling techniques and maximal coupling of Markov chains goes back to the publications of D.Griffeath, J.Pitman, and S.Goldstein in the 70's (see [11], [18], and [12]). K.Burdzy and W.S.Kendall have studied the efficient Markovian couplings in [5].

1.1 Preliminary details.

Suppose that the number of cards n is fixed and define $\mathcal{S} = \{0, 1, \dots, n-1\}$, and suppose \mathcal{S}^2 is the state space of a sequence of iid random variables $\{\tau_t : t \in \mathbb{Z}\}$ which are uniformly distributed on \mathcal{S}^2 . We will also assume, or construct, a bijection from \mathcal{S}^2 to the set of mechanisms creating a sequence of transpositions (or identity) on \mathcal{S} . Each τ_t thus produces independent uniform transpositions, although there is a probability of $\frac{1}{n}$ that there will be no change in the ordering of the deck. Given an initial probability distribution on S_n , we let the law of the distribution on S_n be actualized by a random permutation X_t describing the images of cards $0, 1, \dots, n-1$ in order. This gives a well defined Markov Chain. It will sometimes be useful to Poissonify the Markov Chain to continuous time with iid exponentially distributed interarrival times. In other words, τ_t occurs at a time $t_m = T_1 + T_2 + \dots + T_m$ where T_m are iid mean one exponentially distributed random variables. We denote the law of the totality of this setup by \mathcal{P} , and if λ is the distribution of the initial configuration we denote the law of this conditional distribution by $\mathcal{P}|\lambda$.

A coupling argument would require the construction of a joint distribution on the product

of \mathcal{P} and $\tilde{\mathcal{P}} = (\mathcal{P}|\delta_{id})$, where δ_{id} denotes the point mass distribution on S_n , giving the identity probability 1. This joint distribution must have marginal distributions that agree with \mathcal{P} and $\tilde{\mathcal{P}}$ and so that the random permutations X_t and \tilde{X}_t agree from some time T onward. Notice that the shuffle is invariant with respect to the symmetric group in the following sense:

Definition. Suppose that X_m is a Markov Chain on S_n . The chain is said to be **group invariant** if $dist(\gamma X_{m+1}|X_m = \alpha)$ is equal to the $dist(X_{m+1}|X_m = \gamma^{-1}\alpha)$ for all $\gamma, \alpha \in S_n$.

This is a homogeneity condition similar to that of independent increments, and says that the shuffle is independent of the values or printed labels on the cards. It implies that the distribution of cycle structures of X_1 given $X_0 = \alpha$ depends only on the cycle structure of α . The set of group elements with a given cycle structure form a conjugacy class in the group, S_n . For example, if α is a transposition, i.e. a 2-cycle, then regardless of what α is specifically it has the same coupling and transition structure as any other transposition. Since the identity transitions to each of the $\binom{n}{2}$ transpositions with equal probability $\frac{2}{n^2}$, and remains at the identity with probability $\frac{1}{n}$, we could analyze this process as a Markov chain on the conjugacy classes, i.e. classes of permutations with the same cycle structure. Further, this random walk on the cyclic decompositions is biased toward intermediate central weights of the permutations (where the weight is equal to the sum of the cycle lengths minus the number of cycles, and equal to the minimum number of transpositions required to reduce the permutation to the identity). We could have used this approach to analyze recurrence rates. The number of distinct cycle structures is equal to the partition function on the integers $P(n)$ which grows exponentially with rate $\sqrt{\frac{4n}{3}}$ much slower than the rate of $n \log(n)$ that $n!$ grows with.

There are many ways to associate an element $\langle \mathbf{a}, \mathbf{b} \rangle$ in \mathcal{S}^2 with a transposition. Each card has a value or label printed on it, and each card has a location: the number of cards above it on the deck plus 1. Let Q be the set of values or labels, and let P be the set of positions of the cards. If $\langle \mathbf{a}, \mathbf{b} \rangle$ is a member of $P \times P$ then we associate $\langle \mathbf{a}, \mathbf{b} \rangle$ with interchanging the locations of a card at location \mathbf{a} and a card at location \mathbf{b} . If $\langle \mathbf{a}, \mathbf{b} \rangle$ is a member of $Q \times P$ (in which case we will use $\langle \boxed{\mathbf{a}}, \mathbf{b} \rangle$ notation), then we associate $\langle \boxed{\mathbf{a}}, \mathbf{b} \rangle$ with taking card $\boxed{\mathbf{a}}$ and placing it at location \mathbf{b} , and then using the card that was in location \mathbf{b} to replace $\boxed{\mathbf{a}}$ in its original location. When we begin to couple, we will be following the motion of two decks. It is natural to apply the same operation to each one of the two decks. For example, if we did this with $Q \times P$ association for evolution, we could do the same transposition $\langle \boxed{\mathbf{a}}, \mathbf{b} \rangle$ for both to eventually obtain coupling in $O(n^2)$, as in Aldous and Fill [2]. We however have much more flexibility than that with a bijective map from $Q \times P$ to $Q \times P$, called an **association mapping** to relate the coupled moves of each deck. An association mapping tells us how to couple the immediate descendants of two group elements that are coupled. Recall that we must preserve relations, like siblings and cousins in the tree.

Most coupling arguments are made up of present and past measurable constituents; that

is the coupling method is adapted to an increasing sequence of σ -fields so that the present and past are measurable with respect to their corresponding σ -fields, and that the σ -fields have enough extra randomness to perform independent experiments, subdivide atoms, and so on. In some ways this tendency toward adaptive coupling is historical, and in some ways it is natural to follow one's intuition, and then make it rigorous. The first coupling arguments most people see is a passive coupling in which two Markov chains are allowed to go their own ways independently of each other until they happen to obtain the same state at the same time. From this random time onward they are coupled together. It is also true that if we had perfect knowledge of the situation, we would be able to make the optimal coupling at any time. Unfortunately, the numbers are usually too large and the relationships too complex. The point here is that non-adapted coupling can be natural, intuitive, and have great power in solving problems. Indeed, as we shall see, we can even couple so that we anticipate the future and prepare for it, while maintaining the only essential ingredient of coupling, that of having perfect (or near perfect) distributions on the marginal processes. Moreover, it is possible to have couplings made up of surgically cut and pasted pieces of sample path.

1.2 Strong Uniform Mixing, Weak Bernoulli, and Coupling.

If X_t and Y_t are stochastic processes, then a coupling of X_t and Y_t is a joint probability distribution on the product, (X_t, Y_t) so that the marginal distributions on X and Y agree with original distributions. We say that $X_t(\omega)$ and $Y_t(\omega)$ are coupled at the random time T if for $t \geq T$ it is the case that $X_t = Y_t$. If T is finite a.s., then this argument shows that the distributions of $\{X_t, t > k\}$ and $\{Y_t, t > k\}$ are converging in the total variation distance as k becomes large, and hence have convergent distributions.

Diaconis and Shahshahani [9] define a finite state Markov chain X with invariant probability U to be **strong uniform mixing** if there is a stopping time T , so that $P[T = k, X_k = \alpha]$ is independent of the group element α . Because of this and the invariance of the uniform distribution on the group G , for any α we have $P[X_k = \alpha | k \geq T] = 1/|G|$, as the invariant measure is uniform on G . Regardless of the distribution of X_0 , we have the following bound on the total variation norm, $\|U - \text{dist}(X_k)\|_{TV} \leq P[T > k]$, because this is the only part of the probability space where the total variation is not forced to be 0. The rate of mixing is carried by the distribution of the stopping time. Coupling arises because the total variation norm is achieved by joining the distributions together in a probability preserving way. The process is Markov, so it makes sense for the definition to be independent of the initial state. Coupling for general processes is usually connected to weak Bernoulli, which has a rich history (a.k.a. absolutely regular and β -mixing [20]).

Definition. A finite-valued stochastic process X_m is **weak Bernoulli** if there is a coupling $\{(X'_m, X''_m) : m \in \mathbb{Z}\}$ such that (i) $\{X'\}$ and $\{X''\}$ have the same distribution as $\{X\}$, (ii) the past of X'' , $\{X''_m : m = 0, -1, -2, \dots\}$ is independent of $\{X'\}$, and (iii) there is a random variable T so that $m \geq T$ implies that $X'_m = X''_m$.

In this case, the future becomes independent of past values in a strong way. The pathwise coupling version of weak Bernoulli is

Definition. A finitely valued stochastic process X_m is **tree weak Bernoulli** if there is a coupling $\{(X'_m, X''_m) : m \in \mathbb{Z}\}$ such that (i) $\{X'\}$ and $\{X''\}$ have the same distribution as $\{X\}$, (ii) the past of X'' , $\{X''_m : m = 0, -1, -2, \dots\}$ is independent of $\{X'\}$, (iii) there is a random time a.s. finite time T so that $m \geq T$ implies that $X'_m = X''_m$, and (iv) the coupling respects the tree structure of future sample paths, so if X' and X'' are coupled at time $m > 0$ then each of their descendents (or successors) are also coupled together, i.e. the coupling is given by a tree automorphism of the branching future paths.

The terminology comes from Hoffman and Rudolph [14], in which they used tree very weak Bernoulli to study isomorphisms of 1 to p endomorphisms. The conditional distributions of $\{X_m, m \in [1, 2, \dots, M]\}$ live on the set of labeled trees of length M . We raise the definition because many, if not most, coupling methods have this property of being tree consistent.

When applying tree weak Bernoulli or tree coupling, it is important to consider the events defined by the finitely valued random variables to be atomless, e.g. a subset of the unit interval. Achieving optimal total variation norm typically requires subdividing these events into events of smaller but arbitrary probability.

In any given situation, it is possible, in principle, to find the optimal coupling because the total variation norm is achievable. However, the size of the state space and dependence within the process often make this impractical.

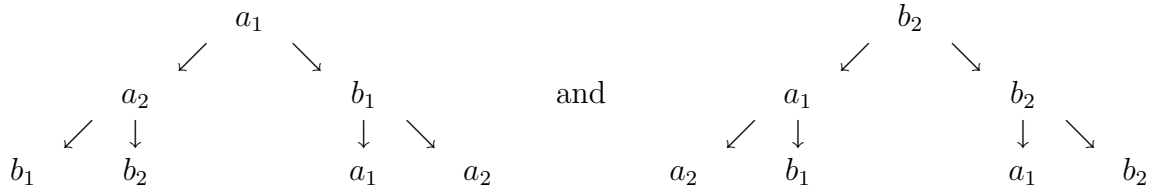
We developed methods to speed up the coupling time while maintaining the distributions on the marginal processes. The methods rely on intuition and insight, and were usually dynamical. By dynamical, we mean that the coupling at a given time depended upon the process up to that time, and, perhaps, some external sources of independent randomness. In other words, the coupling was adapted to the pair of processes as they were being constructed. Unlike uniform strong mixing, tree weak Bernoulli is pathwise and needs a probability-preserving path-wise isomorphism between the future trees of possibilities of the processes.

To illustrate the difference between uniform strong mixing and tree weak Bernoulli suppose we have a Markov Chain with state space a_1, a_2, b_1, b_2 , and suppose the only allowable transitions, each with conditional probability one-half, are:

$$a_1 \rightarrow a_2, \quad a_1 \rightarrow b_1, \quad a_2 \rightarrow b_1, \quad a_2 \rightarrow b_2, \quad b_1 \rightarrow a_1, \quad b_1 \rightarrow a_2, \quad b_2 \rightarrow a_1, \quad b_2 \rightarrow b_2.$$

	a_1	a_2	b_1	b_2
a_1	0	0	1/2	1/2
a_2	1/2	0	1/2	0
b_1	1/2	1/2	0	0
b_2	0	1/2	0	1/2

The matrix is doubly stochastic, so the invariant probability is uniform on each state. Suppose we have two Markov chains with this law, and initial states $X_0 = a_1$, and $Y_0 = b_2$. Possible forward paths are



The distribution of X_2 and Y_2 are equal and at equilibrium, $\|dist(X_2) - dist(Y_2)\| = \|dist(X_2) = U\| = 0$ while $\|dist(X_1) - dist(Y_1)\| = 1$ so $X_1 \neq Y_1 a.s.$ On the other hand, any tree coupling of (X_0, X_1, X_2) with (Y_0, Y_1, Y_2) has $P[X_2 = Y_2] = \frac{1}{2}$.

Similarly, the tree weak Bernoulli, or tree coupling coefficients of a process, may be different than the uniformly strong mixing coefficients, although the coupling distance is never less than the total variation distance.

2 Transposition shuffling for the case $n = 3$.

The models we deal with here are invariant random walks on a group in which each step has the same probability, although the case that different steps will go to the same location is allowed and expected. The set up is a Markov Chain that is itself a hidden Markov Chain with the same number of equally likely outcomes at any stage.

We begin with the group S_3 . We will make some comparisons between two processes, (X_t) where $X_0 = (123)$, the identity and, Y_t . Consider the adjacency matrix, M ,

	123	231	312	132	213	321
123	3	0	0	2	2	2
231	0	3	0	2	2	2
312	0	0	3	2	2	2
132	2	2	2	3	0	0
213	2	2	2	0	3	0
321	2	2	2	0	0	3

If M is the matrix shown then the $(1/9)M$ is the stochastic matrix that represents the dynamics of the chain.

The square of M is $M^2 = 9I + 12N$ where N is the 6 by 6 matrix consisting of all 1's. The maximum total variation distance between any starting point and any other starting point $(2112)/81 = 9/3^4 = 1/9$. If we raise M to the $2m$ power we get $M^{2m} = (9 * I + 12 * N)^m = a_m I + b_m N$, and continuing this calculation shows that the difference between the (1,1) entry and the (1,2) entry is $9^m = 3^{2m}$. This means that after $2m$ iterations the total variation norm between any two distinct starting places is $1/3^{2m}$, so the mixing rate is geometric with error of $\frac{1}{2 \cdot 3^{2m}}$, and the mixing rate is $1/3$. The $1/2$ comes from reducing the total variation distance between distinct starting places to that of one starting permutation and the uniform distribution.

Since every odd power of M is the product of an even power of M with M , we can similarly compute the total variation distance in this case. An analogous calculation gives the total

variation distance between X_m and the uniform U as

$$\|dist(X_m) - U\|_{TV} = \begin{cases} \frac{5}{2} \cdot 3^{-m} & m \text{ odd} , \\ \frac{1}{2} \cdot 3^{-m} & m \text{ even} . \end{cases}$$

So for all m we have

$$\frac{1}{2} \cdot 3^{-m} \leq \|dist(X_m) - U\|_{TV} \leq \frac{5}{2} \cdot 3^{-m}$$

Using the $Q \times P$ association (picking a face value and then picking a location to swap cards with) there are 9 equal possibilities. If $X_0 = 123$, the identity then X_1 has nine equal outcomes, three of which are 123, and there are two chances for each of three transpositions, 132, 213, and 321. These are shown in the chart below. In keeping with group invariance, the possible successors of any permutation are three chances to remain in its previous state and two chances for each of two group elements of opposite parity.

$Q \setminus P$	1	2	3
1	123	213	321
2	213	123	132
3	321	132	123

On the other hand, if we wish to couple with the path arising from 132 using group invariance, we get the following table:

$Q \setminus P$	1	2	3
1	132	312	231
2	312	132	123
3	231	123	132

Looking at the diagram, we see that four out of nine outcomes can be perfectly coupled, and that the remaining line up as transpositions of each other. Actually, we can do a bit better by coupling two that differ by a 3-cycle and then four that couple as transpositions. Specifically, couple two 123s with the corresponding 123s, the two 132s with the corresponding 132s, the remaining 123 with 312, and 213 with 132, and then the remaining three couple in any way with the corresponding remaining 3 outcomes. After drawing some infinite trees, and calculating many geometric series one arrives at the optimal tree coupling which is $m3^{-m}$. Note that this coupling is strictly larger, leaving open the possibility that tree coupling may be unable to achieve the optimal mixing rates. As an aside, note that the trajectories of the process are given by a two dimensional substitution system given by iterating the three by three diagrams above.

3 Coupling to the future

Consider a continuous-time process (X_t, Y_t) on $S = \{0, 1\} \times \{0, 1, 2\}$ with generator

	(0,0)	(0,1)	(0,2)	(1,0)	(1,1)	(1,2)
(0,0)	-12	0	0	10	1	1
(0,1)	0	-12	0	1	10	1
(0,2)	0	0	-12	1	1	10
(1,0)	10	1	1	-12	0	0
(1,1)	1	10	1	0	-12	0
(1,2)	1	1	10	0	0	-12

Here the first coordinate flips $X_t \rightarrow 1 - X_t$ with rate 10, while the second coordinate switches $Y_t \rightarrow (Y_t - 1) \bmod 3$ or $Y_t \rightarrow (Y_t + 1) \bmod 3$ with rate 1 each. However every time the second coordinate changes, the first coordinate must also change.

We want to find a fast coupling for the above process. Coupling the first coordinate is simple: wait with rate 20, then assign with equal probabilities of $\frac{1}{2}$ either 0 or 1 to both X_t and X'_t . Similarly, coupling the second coordinate is also simple: wait with rate 3 before assigning any one of the three values 0, 1, or 2 to both Y_t and Y'_t . However coupling both coordinates simultaneously for the processes (X_t, Y_t) and (X'_t, Y'_t) may create the following complication: let T_1 be the exponential r.v. with parameter 20, and T_2 be the exponential r.v. with parameter 3. With large probability, the first coordinates will couple before the second, i.e. $T_1 < T_2$. In two out of three cases, either Y_t or Y'_t will not change at T_2 . Therefore in two thirds of the cases, when we couple the second coordinate, we simultaneously decouple the first. Thus we will need extra time for the first coordinates to couple again.

Coupling to the future works in the following way. We start by generating T_1 and T_2 . If $T_1 < T_2$, we will start by deciding the value for $Y_{T_2} = Y'_{T_2}$. If it is different than both Y_0 and Y'_0 , then we need to randomly generate $X_{T_1} = X'_{T_1}$, and we are done, the process is coupled at T_2 . However, if $T_1 < T_2$ and $Y_{T_2} = Y'_{T_2}$ matches either Y_0 or Y'_0 , we will need to pair X_t and X'_t differently. Namely we will need to couple X_{T_1} and $1 - X'_{T_1}$, and keep $X_t = 1 - X'_t$ paired until time T_2 . Therefore at T_2 , when the second coordinates Y_t and Y'_t couple, one of the first coordinates must also flip, thus coupling the two processes, (X_t, Y_t) and (X'_t, Y'_t) . In any situation, the coupling time will be $\max\{T_1, T_2\}$, while before it was larger.

Looking into the future of Y_t and Y'_t , we decided whether to pair X_t with X'_t or X_t with $1 - X'_t$. We will call this pairing an *association map*.

4 Super-fast coupling.

Consider an n -card deck. Recall that the random transposition shuffle occurs by making two independent uniform choices of cards, and interchanging them. We assume time is continuous, so the transpositions happen one at a time with exponential times of rate one in between, i.e. for each pair of cards $\boxed{\mathbf{a}}$ and $\boxed{\mathbf{b}}$, transposition $\langle \boxed{\mathbf{a}}, \boxed{\mathbf{b}} \rangle$ has rate $\frac{2}{n^2}$, i.e. the two identical transpositions $\langle \boxed{\mathbf{a}}, \boxed{\mathbf{b}} \rangle$ and $\langle \boxed{\mathbf{b}}, \boxed{\mathbf{a}} \rangle$ happen with rate $\frac{1}{n^2}$ each. Transposition $\langle \boxed{\mathbf{a}}, \boxed{\mathbf{a}} \rangle$ has rate $\frac{1}{n^2}$. Diaconis and Shahshahani used group representation

methods to show that the mixing time for this shuffling process is $O(n \log n)$ (see [9] and [6]). In [2] the possible coupling approach was shown to produce upper bound = $O(n^2)$, while [17] lists showing $O(n \log n)$ mixing time via *coupling* construction as an open problem. Here we solve it.

4.1 Notations and vocabulary.

Notations and Vocabulary	
$\langle \cdot, \cdot \rangle$	- transpositions in the card shuffling process
$\langle \boxed{\mathbf{a}}, \cdot \rangle$	- transpositions initiated by card $\boxed{\mathbf{a}}$
A_t	- the top shuffling process
B_t	- the bottom shuffling process
$\begin{pmatrix} A_t \\ B_t \end{pmatrix}$	- the coupled process
$\langle \cdot, \cdot \rangle_A$	- transpositions in the top shuffling process A_t
$\langle \cdot, \cdot \rangle_B$	- transpositions in the bottom shuffling process B_t
$\ll \cdot, \cdot \gg$	- simultaneous transpositions in the coupled process $\begin{pmatrix} A_t \\ B_t \end{pmatrix}$
<i>association map</i>	- hidden association between positions/locations in the top process and positions/locations in the bottom process that will be used to establish the rates for the coupled process

4.2 Label-to-location and label-to-label transpositions.

One of the possible coupling constructions was described in [2]. There, at each step, a card $\boxed{\mathbf{a}}$ and a location i were selected at random, and the transposition $\ll \boxed{\mathbf{a}}, i \gg$ that moves card $\boxed{\mathbf{a}}$ to location i was applied in both processes, A_t and B_t . That is we used $Q \times P$ association for both decks. Clearly, this coupling slows down significantly when the number of discrepancies is small enough, thus producing an upper bound of order $O(n^2)$, instead of $O(n \log n)$. An equivalent result can be achieved by first randomly selecting a card $\boxed{\mathbf{a}}$, and then applying transposition $\ll \boxed{\mathbf{a}}, i \gg$ with location i selected at random, *if $\boxed{\mathbf{a}}$ is not coupled*, and applying transposition $\ll \boxed{\mathbf{a}}, \boxed{\mathbf{b}} \gg$ of cards $\boxed{\mathbf{a}}$ and $\boxed{\mathbf{b}}$ in both processes, for a randomly selected card $\boxed{\mathbf{b}}$, *if $\boxed{\mathbf{a}}$ is coupled*. i.e. in the situation when $\boxed{\mathbf{a}}$ is at the same location for both processes, A_t and B_t . The second coupling is important, as it can be improved to match the correct $O(n \log n)$ order for mixing time. The improvement comes in the form of a combinatorial trick similar to the one used by Euler in computing the number of permutations of n elements with all elements displaced.

From here on transpositions $\ll \boxed{\mathbf{a}}, \boxed{\mathbf{b}} \gg$ will be called **label-to-label** transpositions.

4.3 Group invariance.

Without loss of generality, we would like to state that, in our coupling construction, whenever a coupled card, for example $\boxed{\mathbf{b}}$, is selected, and thus a transposition $\ll \boxed{\mathbf{b}}, \boxed{\mathbf{a}} \gg$ with another card $\boxed{\mathbf{a}}$ is to be applied, we actually need not do this transposition, as the resulting

discrepancies will be the same (with different card values, of course) as before the transposition. This is because they have the same cycle structure, and so are conjugate. Moreover, if the card $\boxed{\mathbf{a}}$ is not coupled, and is waiting for the transposition $\ll \boxed{\mathbf{a}}, i \gg$, then after $\ll \boxed{\mathbf{b}}, \boxed{\mathbf{a}} \gg$ the card $\boxed{\mathbf{b}}$ will be uncoupled, and must wait for $\ll \boxed{\mathbf{b}}, i \gg$ with the same random i , as i has not been used yet. If we ignore transposition $\ll \boxed{\mathbf{b}}, \boxed{\mathbf{a}} \gg$, then $\boxed{\mathbf{a}}$ will continue waiting for $\ll \boxed{\mathbf{a}}, i \gg$ as if no changes have happened. If σ is a permutation of face values on the cards, then from our point of view, the coupled process $\begin{pmatrix} A_t \\ B_t \end{pmatrix}$ is equivalent $\begin{pmatrix} \sigma A_t \\ \sigma B_t \end{pmatrix}$ because of group invariance of the process. Also, if $\boxed{\mathbf{a}}$ is uncoupled, and $\ll \boxed{\mathbf{a}}, i \gg$ is to occur, where site i is occupied by the same $\boxed{\mathbf{b}}$ in both processes, no relabeling is necessary (relabeling is optional). If we do not relabel, the card $\boxed{\mathbf{a}}$ will have to reselect a random new i , and wait for $\ll \boxed{\mathbf{a}}, i \gg$ to happen. The key point is that the situation is invariant under label-to-label transpositions.

4.4 Improving the coupling.

The basic strategy of the coupling described in this paper is to condition on a key σ -field in the future and then to use this future information to arrange the intermittent events so that the process is set up for a successful coupling event. This is completely legal as long as we take care to have the marginal stochastic processes maintain the correct finite-dimensional distributions. If this is followed, we are able to modify the joint distributions of the processes however we like to obtain our goal.

Example. To illustrate the forthcoming construction (and our notation), suppose we have four cards that are paired so they have two discrepancies ($d = 2$) at time $t_0 = 0$.

$$\begin{array}{rcccc} \text{deck } A : & \boxed{1} & \boxed{2} & \boxed{3} & \boxed{4} \\ \text{deck } B : & \boxed{1} & \boxed{3} & \boxed{2} & \boxed{4} \\ \text{location :} & 1 & 2 & 3 & 4 \end{array}$$

We pick a random uniform location $i_1 \in \{1, 2, 3, 4\}$ and an exponential time t_1 .

Conditioned on the event of card $\boxed{2}$ jumping to i_1 (i.e. $\ll \boxed{2}, i_1 \gg$) at time t_1 , we provide the following coupling rules for time $t \in [0, t_1]$.

Case I: $i_1 = 2$ or 3 .

Here cards $\boxed{1}$, $\boxed{3}$ and $\boxed{4}$ do label-to-label jumps only, and each of these jumps leads to card setups that are isomorphic to the original setup, up to relabeling the cards. Because of this we suppress noting any change at all. To illustrate this, a label-to-label jump $\ll \boxed{1}, \boxed{3} \gg$ will take

$$\begin{array}{rcccc} \text{deck } A : & \boxed{1} & \boxed{2} & \boxed{3} & \boxed{4} \\ \text{deck } B : & \boxed{1} & \boxed{3} & \boxed{2} & \boxed{4} \end{array} \quad \text{to} \quad \begin{array}{rcccc} \text{deck } A : & \boxed{3} & \boxed{2} & \boxed{1} & \boxed{4} \\ \text{deck } B : & \boxed{3} & \boxed{1} & \boxed{2} & \boxed{4} \end{array}$$

This latter set up is equivalent to the former setup, so this case leads to no change. At the end, card $\boxed{2}$ does the label-to-location jump $\langle\langle \boxed{2}, i_1 \rangle\rangle$ at time t_1 , canceling the discrepancies.

Case II: $i_1 = 1$ or 4 , one of the non-discrepancy locations. Again it suffices to only consider $i_1 = 4$. Here are the rates for this case:

- Cards $\boxed{1}$ and $\boxed{3}$ do label-to-label jumps, and as in Case I, no real change occurs and we again suppress any notational changes.
- On the other hand if we pick card $\boxed{4}$, then we couple both decks together as follows: either we interchange cards $\boxed{4}$ and $\boxed{3}$ on the top ($\langle \boxed{4}, \boxed{3} \rangle_A$), getting

$$\begin{array}{l} \text{deck } A : \quad \boxed{1} \quad \boxed{2} \quad \boxed{4} \quad \boxed{3} \\ \text{deck } B : \quad \boxed{1} \quad \boxed{3} \quad \boxed{2} \quad \boxed{4} \end{array}$$

The other possibility is we interchange $\boxed{4}$ and $\boxed{3}$ on the bottom ($\langle \boxed{4}, \boxed{3} \rangle_B$) to get

$$\begin{array}{l} \text{deck } A : \quad \boxed{1} \quad \boxed{2} \quad \boxed{3} \quad \boxed{4} \\ \text{deck } B : \quad \boxed{1} \quad \boxed{4} \quad \boxed{2} \quad \boxed{3} \end{array}$$

Once one of the above two transpositions occurs, card $\boxed{4}$ joins $\boxed{1}$ and $\boxed{3}$ as one of the cards that does label-to-label jumps only which will be unnoticed for the rest of the time in $[0, t_1]$.

The other two options for card $\boxed{4}$ are label-to-label transpositions $\langle\langle \boxed{4}, \boxed{1} \rangle\rangle$ and $\langle\langle \boxed{4}, \boxed{2} \rangle\rangle$ getting

$$\begin{array}{l} \text{deck } A : \quad \boxed{4} \quad \boxed{2} \quad \boxed{3} \quad \boxed{1} \\ \text{deck } B : \quad \boxed{4} \quad \boxed{3} \quad \boxed{2} \quad \boxed{1} \end{array} \quad \text{and} \quad \begin{array}{l} \text{deck } A : \quad \boxed{1} \quad \boxed{4} \quad \boxed{3} \quad \boxed{2} \\ \text{deck } B : \quad \boxed{1} \quad \boxed{3} \quad \boxed{4} \quad \boxed{2} \end{array}$$

respectively, which we suppress noting. The rates being the same for each one of $\langle \boxed{4}, \boxed{3} \rangle_A$, $\langle \boxed{4}, \boxed{3} \rangle_B$, $\langle\langle \boxed{4}, \boxed{1} \rangle\rangle$ and $\langle\langle \boxed{4}, \boxed{2} \rangle\rangle$.

- Card $\boxed{2}$ does the label to location jump $\langle\langle \boxed{2}, i_1 \rangle\rangle$ at time t_1 .

The point is that in Case II, either of the following transposition sequences

$$\langle \boxed{4}, \boxed{3} \rangle_A \quad \text{followed by} \quad \langle\langle \boxed{2}, i_1 \rangle\rangle$$

or

$$\langle \boxed{4}, \boxed{3} \rangle_B \quad \text{followed by} \quad \langle\langle \boxed{2}, i_1 \rangle\rangle$$

would lead to the discrepancies' cancelation.

If at time t_1 the discrepancies are not canceled out, start anew with a new random i_1 and exponential t_1 . There are two possibilities at these trials: either we end up with no discrepancies or we end up in the same boat as before and we try again. Here we set the

coupling rules by conditioning only on one event, $\ll \boxed{2}, i_1 \gg$. If we condition on more than one upcoming events (say $\ll \boxed{4}, \boxed{3} \gg_A$ in Case II), by conditioning inside the conditioning, we can increase the probability of coincidence, thus producing a faster coupling time. Later in the paper we will deal with conditioning on a chain of events.

The above example is a simplified version of the coupling construction to follow.

4.4.1 Two discrepancies and one association map ($d = 2, k = 1$).

We will start with *the case of two discrepancies* ($d = 2$): that is, when the coupling speed is slowest. As previously implied, there are only two discrepancies, at sites d_1 and d_2 :

$$\begin{array}{cccccccccccc}
 A_t: & \dots & \boxed{4} & \boxed{6} & \boxed{\mathbf{b}} & \boxed{9} & \boxed{\mathbf{a}} & \boxed{8} & \boxed{\mathbf{a1}} & \boxed{2} & \dots \\
 B_t: & \dots & \boxed{4} & \boxed{6} & \boxed{\mathbf{a}} & \boxed{9} & \boxed{\mathbf{b}} & \boxed{8} & \boxed{\mathbf{a1}} & \boxed{2} & \dots \\
 & & & & \uparrow & & \uparrow & & \uparrow & & \\
 & & & & d_2 & & d_1 & & i_1 & &
 \end{array}$$

Due to the “group invariance” that was mentioned before, the above picture does not change until either $\boxed{\mathbf{a}}$ or $\boxed{\mathbf{b}}$ jumps to either d_1 or d_2 , in both the top and the bottom processes. The rest of the transpositions are “label-to-label”, and as such, will not change the picture. If we do not adjust this coupling, the waiting time to cancel the two discrepancies will average $\frac{n^2}{4}$, which is too large, so we are going to modify the coupling construction.

Here is how we adjust the coupling. We take one of the two uncoupled cards, for example $\boxed{\mathbf{a}}$. We pick a random site i_1 , and condition on transposition $\ll \boxed{\mathbf{a}}, i_1 \gg$, with respect to group invariance, happening at a random exponential time t_1 . If $i_1 = d_1$ or d_2 , the discrepancies disappear, and the coupling is complete.

Now, if $i_1 \neq d_1$ or d_2 , we will associate sites as if the transposition $\ll \boxed{\mathbf{a}}, i_1 \gg$ has already occurred. That is, up until time t_1 ,

- (a) we pair the site i_1 in A_t with the site d_1 in B_t , and call it i_1/d_1 ,
- (b) we pair the site d_2 in A_t with the site i_1 in B_t , and call it d_2/i_1
- (c) we pair the site d_1 in A_t with the site d_2 in B_t , and call it d_1/d_2

So we will get the following association map with respect to transposition $\ll \boxed{\mathbf{a}}, i_1 \gg$ and jump time t_1 :

$$\begin{array}{cccccccccccc}
 A_t: & \dots & \boxed{4} & \boxed{6} & \boxed{\mathbf{b}} & \boxed{9} & \boxed{\mathbf{a1}} & \boxed{8} & \boxed{\mathbf{a}} & \boxed{2} & \dots \\
 B_t: & \dots & \boxed{4} & \boxed{6} & \boxed{\mathbf{a1}} & \boxed{9} & \boxed{\mathbf{b}} & \boxed{8} & \boxed{\mathbf{a}} & \boxed{2} & \dots \\
 & & & & \uparrow & & \uparrow & & \uparrow & & \\
 & & & & d_2/i_1 & & i_1/d_1 & & d_1/d_2 & &
 \end{array} \tag{1}$$

Here the transposition $\ll \boxed{\mathbf{a}}, i_1 \gg$ at time t_1 will correspond to relabeling the three sites via the following association map rule

$$\boxed{\begin{array}{l} d_1/d_2 \longrightarrow i_1 \\ i_1/d_1 \longrightarrow d_1 \\ d_2/i_1 \longrightarrow d_2 \end{array}}$$

From our new perspective, time t_1 is the time when the location names change according to the above rule. We will say that the association map expires at t_1 .

We use the above association map to determine the *rates* for the new coupling process in the time interval $(0, t_1]$:

- *Rates for $\boxed{\mathbf{a}}$* : the first jump $\ll \boxed{\mathbf{a}}, i_1 \gg$ occurs at time t_1 .
- *Rates for $\boxed{\mathbf{a1}}$* : $\boxed{\mathbf{a1}}$ does label-to-location jumps, where locations are defined by the association map (1). Transpositions $\ll \boxed{\mathbf{a1}}, i \gg$ occur for all $i \neq i_1, d_1, d_2$, with the usual rate of $\frac{1}{n^2}$. Transpositions $\ll \boxed{\mathbf{a1}}, i_1/d_1 \gg$ and $\ll \boxed{\mathbf{a1}}, d_2/i_1 \gg$ occur with rate $\frac{1}{n^2}$ as well. Transposition $\ll \boxed{\mathbf{a1}}, d_1/d_2 \gg$ is label-to-label, hence we do not count it, and do not relabel.
- *Rates for $\boxed{\mathbf{b}}$* : transpositions $\ll \boxed{\mathbf{b}}, d_1 \gg$ and $\ll \boxed{\mathbf{b}}, d_1 \gg$ of the coupled process, as well as transpositions $\langle \boxed{\mathbf{b}}, \boxed{\mathbf{a1}} \rangle_A$ of the top process A_t , and $\langle \boxed{\mathbf{b}}, \boxed{\mathbf{a1}} \rangle_B$ of the top process B_t , occur independently with rate $\frac{1}{n^2}$ each. The rest of the transpositions initiated by $\boxed{\mathbf{b}}$ are label-to-label transpositions of the coupled process, for which we do not need to relabel. After one of the six transpositions that cancel discrepancies on the association map (i.e. $\ll \boxed{\mathbf{a1}}, i_1/d_1 \gg$, $\ll \boxed{\mathbf{a1}}, d_2/i_1 \gg$, $\ll \boxed{\mathbf{b}}, d_1 \gg$, $\ll \boxed{\mathbf{b}}, d_1 \gg$, $\langle \boxed{\mathbf{b}}, \boxed{\mathbf{a1}} \rangle_A$, and $\langle \boxed{\mathbf{b}}, \boxed{\mathbf{a1}} \rangle_B$) occurs, $\boxed{\mathbf{b}}$ will only be allowed to do label-to-label jumps.
- *For the rest of the cards*, we apply label-to-label transpositions of the coupled process.

Since we classified (and therefore excluded) $\ll \boxed{\mathbf{a1}}, d_1/d_2 \gg$ as a label-to-label transposition, the jump time t_2 for the card $\boxed{\mathbf{a1}}$ is *exponential* $(\frac{1}{n} \cdot \frac{n-1}{n})$. Now, $t_2 < t_1$ with probability $\frac{\frac{n-1}{n}}{1+\frac{n-1}{n}} = \frac{n-1}{2n-1}$. In that case, transpositions $\ll \boxed{\mathbf{a1}}, i_1/d_1 \gg$ and $\ll \boxed{\mathbf{a1}}, d_2/i_1 \gg$ cancel the discrepancies on the association map (1). That happens with probability $\frac{2}{n-1}$. Thus, the probability of canceling the discrepancy by time t_1 is

$$\frac{2}{n} \left(2 - \frac{1}{n-1} \right) + P[t_b < t_1] \left(1 - \frac{2}{n} \left(2 - \frac{1}{n-1} \right) \right) \approx \frac{8}{n},$$

where $\frac{2}{n} \left(2 - \frac{1}{n-1} \right) = \frac{2}{n} + \left(1 - \frac{2}{n} \right) \cdot \frac{n-1}{2n-1} \cdot \frac{2}{n-1}$ is the probability that the discrepancy is canceled as the result of the jump initiated by $\boxed{\mathbf{a}}$ at t_1 or $\boxed{\mathbf{a1}}$ at $t_2 < t_1$, and if this does not happen, then with probability $P[t_b < t_1] = \frac{\frac{4}{n^2}}{\frac{1}{n} + \frac{4}{n^2}} \approx \frac{4}{n}$, the discrepancy is canceled with the jump of $\boxed{\mathbf{b}}$. Here t_b is the first time one of the four transpositions $\ll \boxed{\mathbf{b}}, d_1 \gg$, $\ll \boxed{\mathbf{b}}, d_1 \gg$, $\langle \boxed{\mathbf{b}}, \boxed{\mathbf{a1}} \rangle_A$ and $\langle \boxed{\mathbf{b}}, \boxed{\mathbf{a1}} \rangle_B$ occurs, so t_b is *exponential* $(\frac{4}{n^2})$.

So, with probability $\approx \frac{8}{n}$, the discrepancies cancel by time t_1 . If they do not cancel, we repeat the association trick, thus coupling the two discrepancies in approximately $\frac{n^2}{8}$ steps on average, instead of $\frac{n^2}{4}$. We only used one association map at a time, however.

Example. We will illustrate how a cancellation of discrepancies on an association map implies discrepancies' cancelation at time t_1 . Consider the case when $\ll \mathbf{a1}, d_2/i_1 \gg$ occurs at time $t_2 < t_1$. Then we will observe the following dynamics on the association map: The configurations will evolve from

$$\begin{array}{cccccccc}
 A_t : & \dots & \boxed{4} & \boxed{6} & \boxed{\mathbf{b}} & \boxed{9} & \boxed{\mathbf{a1}} & \boxed{8} & \boxed{\mathbf{a}} & \boxed{2} & \dots \\
 B_t : & \dots & \boxed{4} & \boxed{6} & \boxed{\mathbf{a1}} & \boxed{9} & \boxed{\mathbf{b}} & \boxed{8} & \boxed{\mathbf{a}} & \boxed{2} & \dots \\
 & & & & \uparrow & & \uparrow & & \uparrow & & \\
 & & & & d_2/i_1 & & i_1/d_1 & & d_1/d_2 & &
 \end{array}$$

to

$$\begin{array}{cccccccc}
 A_t : & \dots & \boxed{4} & \boxed{6} & \boxed{\mathbf{a1}} & \boxed{9} & \boxed{\mathbf{b}} & \boxed{8} & \boxed{\mathbf{a}} & \boxed{2} & \dots \\
 B_t : & \dots & \boxed{4} & \boxed{6} & \boxed{\mathbf{a1}} & \boxed{9} & \boxed{\mathbf{b}} & \boxed{8} & \boxed{\mathbf{a}} & \boxed{2} & \dots \\
 & & & & \uparrow & & \uparrow & & \uparrow & & \\
 & & & & d_2/i_1 & & i_1/d_1 & & d_1/d_2 & &
 \end{array}$$

at time t_2 , and

$$\begin{array}{cccccccc}
 A_t : & \dots & \boxed{4} & \boxed{6} & \boxed{\mathbf{a1}} & \boxed{9} & \boxed{\mathbf{b}} & \boxed{8} & \boxed{\mathbf{a}} & \boxed{2} & \dots \\
 B_t : & \dots & \boxed{4} & \boxed{6} & \boxed{\mathbf{a1}} & \boxed{9} & \boxed{\mathbf{b}} & \boxed{8} & \boxed{\mathbf{a}} & \boxed{2} & \dots \\
 & & & & \uparrow & & \uparrow & & \uparrow & & \\
 & & & & d_2 & & d_1 & & i_1 & &
 \end{array}$$

at time t_1 .

The above were the transformations one would see on the association map. The corresponding evolutions of the decks with respect to the original site associations will be as follows. From

$$\begin{array}{cccccccc}
 A_t : & \dots & \boxed{4} & \boxed{6} & \boxed{\mathbf{b}} & \boxed{9} & \boxed{\mathbf{a}} & \boxed{8} & \boxed{\mathbf{a1}} & \boxed{2} & \dots \\
 B_t : & \dots & \boxed{4} & \boxed{6} & \boxed{\mathbf{a}} & \boxed{9} & \boxed{\mathbf{b}} & \boxed{8} & \boxed{\mathbf{a1}} & \boxed{2} & \dots \\
 & & & & \uparrow & & \uparrow & & \uparrow & & \\
 & & & & d_2 & & d_1 & & i_1 & &
 \end{array}$$

to

$$\begin{array}{cccccccc}
 A_t : & \dots & \boxed{4} & \boxed{6} & \boxed{\mathbf{a1}} & \boxed{9} & \boxed{\mathbf{a}} & \boxed{8} & \boxed{\mathbf{b}} & \boxed{2} & \dots \\
 B_t : & \dots & \boxed{4} & \boxed{6} & \boxed{\mathbf{a}} & \boxed{9} & \boxed{\mathbf{b}} & \boxed{8} & \boxed{\mathbf{a1}} & \boxed{2} & \dots \\
 & & & & \uparrow & & \uparrow & & \uparrow & & \\
 & & & & d_2 & & d_1 & & i_1 & &
 \end{array}$$

at time t_2 , and to

$$\begin{array}{cccccccc}
 A_t : & \dots & \boxed{4} & \boxed{6} & \boxed{\mathbf{a1}} & \boxed{9} & \boxed{\mathbf{b}} & \boxed{8} & \boxed{\mathbf{a}} & \boxed{2} & \dots \\
 B_t : & \dots & \boxed{4} & \boxed{6} & \boxed{\mathbf{a1}} & \boxed{9} & \boxed{\mathbf{b}} & \boxed{8} & \boxed{\mathbf{a}} & \boxed{2} & \dots \\
 & & & & \uparrow & & \uparrow & & \uparrow & & \\
 & & & & d_2 & & d_1 & & i_1 & &
 \end{array}$$

at time t_1 .

4.4.2 Two discrepancies and εn association maps ($d = 2, k = \lfloor \varepsilon n \rfloor$).

In the preceding subsection, in order to enhance the coupling process, we considered the following association map

$$\begin{array}{cccccccccccc}
A_t : & \dots & \boxed{\mathbf{a2}} & \boxed{6} & \boxed{\mathbf{b}} & \boxed{9} & \boxed{\mathbf{a1}} & \boxed{8} & \boxed{\mathbf{a}} & \boxed{2} & \dots \\
B_t : & \dots & \boxed{\mathbf{a2}} & \boxed{6} & \boxed{\mathbf{a1}} & \boxed{9} & \boxed{\mathbf{b}} & \boxed{8} & \boxed{\mathbf{a}} & \boxed{2} & \dots \\
& & \uparrow & & \uparrow & & \uparrow & & \uparrow & & \\
& & i_2 & & d_2^* & & d_1^* & & i_1^* & &
\end{array} \tag{2}$$

where

- (a) d_1^* denotes i_1/d_1 before t_1 , and d_1 after t_1 ,
- (b) d_2^* denotes d_2/i_1 before t_1 , and d_2 after t_1 ,
- (c) i_1^* denotes d_1/d_2 before t_1 , and i_1 after t_1 .

On the above association map, we again have two discrepancies, and we adjust the coupling rules so that to cancel these new discrepancies. We do this by selecting a random location $i_2 \in \{1, 2, \dots, n\} \setminus \{i_1^*\}$ on the association map and a random *exponential* $(\frac{1}{n} \cdot [1 - \frac{1}{n}])$ time t_2 . We condition on the transposition $\ll \boxed{\mathbf{a1}}, i_2 \gg$ happening at time t_2 .

On the new scheme, if $i_2 = d_1^*$ or d_2^* , the discrepancies cancel out, and the process couples after time $\tau_2 = \max\{t_1, t_2\}$. Now, if $i_2 \neq d_1^*$ or d_2^* , we will construct one more association map, paring the site i_2 in A_t with site d_1^* in B_t and calling it i_2/d_1^* ; paring the site d_2^* in A_t with site i_2 in B_t and calling it d_2^*/i_2 ; and finally, paring the site d_1^* in A_t and site d_2^* in B_t and calling it d_1^*/d_2^* . After pairings, the association map (2) becomes

$$\begin{array}{cccccccccccc}
A_t : & \dots & \boxed{\mathbf{a1}} & \boxed{6} & \boxed{\mathbf{b}} & \boxed{9} & \boxed{\mathbf{a2}} & \boxed{8} & \boxed{\mathbf{a}} & \boxed{2} & \dots \\
B_t : & \dots & \boxed{\mathbf{a1}} & \boxed{6} & \boxed{\mathbf{a2}} & \boxed{9} & \boxed{\mathbf{b}} & \boxed{8} & \boxed{\mathbf{a}} & \boxed{2} & \dots \\
& & \uparrow & & \uparrow & & \uparrow & & \uparrow & & \\
& & d_1^*/d_2^* & & d_2^*/i_2 & & i_2/d_1^* & & i_1^* & &
\end{array}$$

We construct the second association map

$$\begin{array}{cccccccccccc}
A_t : & \dots & \boxed{\mathbf{a1}} & \boxed{6} & \boxed{\mathbf{b}} & \boxed{9} & \boxed{\mathbf{a2}} & \boxed{8} & \boxed{\mathbf{a}} & \boxed{\mathbf{a3}} & \dots \\
B_t : & \dots & \boxed{\mathbf{a1}} & \boxed{6} & \boxed{\mathbf{a2}} & \boxed{9} & \boxed{\mathbf{b}} & \boxed{8} & \boxed{\mathbf{a}} & \boxed{\mathbf{a3}} & \dots \\
& & \uparrow & & \uparrow & & \uparrow & & \uparrow & \uparrow & \\
& & i_2^{**} & & d_2^{**} & & d_1^{**} & & i_1^* & i_3 &
\end{array} \tag{3}$$

where d_1^{**} denotes i_2/d_1^* before t_2 , and d_1^* after t_2 , d_2^{**} denotes d_2^*/i_2 before t_2 , and d_2^* after t_2 , and i_2^{**} denotes d_1^*/d_2^* before t_2 , and i_2 after t_2 .

Next, we pick a random location $i_3 \in \{1, 2, \dots, n\} \setminus \{i_1^*, i_2\}$ on the association map (3). We pick a random *exponential* $(\frac{1}{n} \cdot [1 - \frac{2}{n}])$ waiting time t_3 and condition on transposition $\ll \boxed{\mathbf{a2}}, i_3 \gg$ on the association map (3) at time t_3 . If $i_3 = d_1^{**}$ or d_2^{**} , the discrepancies cancel out, and the process couples after time $\tau_3 = \max\{t_1, t_2, t_3\}$. If $i_3 \neq d_1^{**}$ or d_2^{**} , we will construct the third association map. Proceeding inductively, after j iterations, we either

cancel the discrepancies with probability of at least $\frac{2}{n-j+1}$, or construct one more association map.

We will construct a chain of at most $k = \lfloor \varepsilon n \rfloor$ association maps like that, where $\varepsilon \in (0, 1)$ is fixed. *Observe:* For all cards other than $\boxed{\mathbf{a}}$, $\boxed{\mathbf{a1}}$, $\boxed{\mathbf{a2}}, \dots, \boxed{\mathbf{ak}}$ and $\boxed{\mathbf{b}}$ w.r.t. group invariance, all jumps are label-to-label transpositions. For cards $\boxed{\mathbf{a}}$, $\boxed{\mathbf{a1}}$, $\boxed{\mathbf{a2}}, \dots, \boxed{\mathbf{ak}}$, each $\boxed{\mathbf{aj}}$ does label-to-location jumps with respect to the j -th association map.

For the rest of the paper we will not use the $*$ in the notations of discrepancies on the association maps (e.g. we will write d_1 instead of d_1^* when talking of the discrepancies on the second association map). Let T_a be the first time one of the following transpositions occurs on the corresponding association map: $\ll \boxed{\mathbf{a}}, d_1 \gg$, $\ll \boxed{\mathbf{a}}, d_2 \gg$, $\ll \boxed{\mathbf{aj}}, d_1 \gg$ and $\ll \boxed{\mathbf{aj}}, d_2 \gg$ for all $j = 1, \dots, k$.

The $k = \lfloor \varepsilon n \rfloor$ association maps also define the *rates for* $\boxed{\mathbf{b}}$: for each $j \in \{1, \dots, k\}$, we will cancel discrepancies with transpositions $\langle d_2, \boxed{\mathbf{aj}} \rangle_A$ and $\langle d_1, \boxed{\mathbf{aj}} \rangle_B$ occurring independently until a discrepancy is canceled, all with respect to j -th association map. The k association maps will, on average, expire in less than $n \log(\varepsilon n)$ units of time. In other words, card $\boxed{\mathbf{b}}$ (w.r.t. group invariance) makes jumps that are label-to-label transpositions $\ll \boxed{\mathbf{b}}, \mathbf{c} \gg$ if the card \mathbf{c} is **not** $\boxed{\mathbf{a}}$, or $\boxed{\mathbf{a1}}$, or $\boxed{\mathbf{a2}}, \dots, \boxed{\mathbf{ak}}$. For $\boxed{\mathbf{a}}$, $\boxed{\mathbf{a1}}$, $\boxed{\mathbf{a2}}, \dots, \boxed{\mathbf{ak}}$, transpositions

$$\langle \boxed{\mathbf{b}}, \boxed{\mathbf{a}} \rangle_A, \langle \boxed{\mathbf{b}}, \boxed{\mathbf{a}} \rangle_B, \langle \boxed{\mathbf{b}}, \boxed{\mathbf{a1}} \rangle_A, \langle \boxed{\mathbf{b}}, \boxed{\mathbf{a1}} \rangle_B, \dots, \langle \boxed{\mathbf{b}}, \boxed{\mathbf{aj}} \rangle_A, \langle \boxed{\mathbf{b}}, \boxed{\mathbf{aj}} \rangle_B$$

(with respect to j -th association map) occur independently until a discrepancy is canceled.

Recall $k = \lfloor \varepsilon n \rfloor$. We let T_b be the first time of a jump from one of the discrepancy locations, d_1 or d_2 , to some $\boxed{\mathbf{aj}}$ in the top, or the bottom, or both processes. Then T_b is *exponential* $(\frac{2(k+1)}{n^2})$.

So, the average time for the discrepancy cancellation over all k association maps is $T_d = \min\{T_a, T_b\}$ with

$$E[T_d] = \frac{n^2}{4(k+1)} \leq \frac{n}{4\varepsilon}.$$

If we cancel the discrepancy on one of the k association map, then the coupling completes at time $\tau_k = \max\{t_1, t_2, \dots, t_k\}$, where for each $j \in \{1, 2, \dots, k\}$,

$$t_j \text{ is } \textit{exponential}(\frac{1}{n} \cdot [1 - \frac{j-1}{n}]).$$

For n large enough,

$$E[\tau_k] \leq \frac{1}{\frac{1}{n} \cdot [1 - \frac{k-1}{n}]} \log k \leq \frac{n}{1-\varepsilon} \log(\varepsilon n).$$

The k association maps will, on average, expire in less than $\frac{n}{1-\varepsilon} \log(\varepsilon n) + \frac{n}{4\varepsilon}$ units of time.

4.5 The case of $d \geq 2$ discrepancies and $k = \lfloor \varepsilon n \rfloor$ association maps.

Let d_1, d_2, \dots, d_d denote all the discrepancies, and $\boxed{\mathbf{b}_1}, \boxed{\mathbf{b}_2}, \dots, \boxed{\mathbf{b}_d}$ denote all the discrepancy cards.

$$\begin{array}{cccccccccccc}
 A_t : & \dots & \boxed{4} & \boxed{6} & \boxed{\mathbf{b}_2} & \boxed{9} & \boxed{\mathbf{b}_8} & \boxed{\mathbf{b}_3} & \boxed{\mathbf{b}_4} & \boxed{\mathbf{b}_1} & \dots \\
 B_t : & \dots & \boxed{4} & \boxed{6} & \boxed{\mathbf{b}_8} & \boxed{9} & \boxed{\mathbf{b}_1} & \boxed{\mathbf{b}_4} & \boxed{\mathbf{b}_2} & \boxed{\mathbf{b}_3} & \dots \\
 & & & & \uparrow & & \uparrow & \uparrow & \uparrow & \uparrow & \\
 & & & & d_1 & & d_2 & d_3 & d_4 & d_5 &
 \end{array}$$

In the case of $d \geq 2$ discrepancies, each discrepancy card $\boxed{\mathbf{b}_1}, \boxed{\mathbf{b}_2}, \dots, \boxed{\mathbf{b}_d}$ will have a chain of $\approx \frac{\varepsilon n}{d}$ association maps attached to it, adding new association maps at expiration times. See the preceding paragraphs for the construction of chains of association maps. Each chain will determine the rates for $\boxed{\mathbf{b}_m}, \boxed{\mathbf{a}_{m1}}, \boxed{\mathbf{a}_{m2}}, \dots$. A transposition $\ll \boxed{\mathbf{a}_{mj}}, d_i \gg$ will cancel discrepancies on the j -th association map, in the m -th chain of association maps. So, for each m from 1 to d , a transposition $\ll \boxed{\mathbf{b}_m}, i_{m1} \gg$ is expected to take place at time t_{m1} , and an association map is created. Transposition $\ll \boxed{\mathbf{a}_{m1}}, i_{m2} \gg$, where i_{m2} ($i_{m2} \neq i_{m1}$) is a location with respect the association map, is expected to take place at time t_{m2} , and a new association map is created, so as we continue, creating, for each $m \in \{1, \dots, d\}$ a **chain** of $\lfloor \frac{\varepsilon n}{d} \rfloor$ or $\lfloor \frac{\varepsilon n}{d} \rfloor + 1$ association maps. Since all locations i_{mj} corresponding to the $k = \lfloor \varepsilon n \rfloor$ association maps are different, for a given fixed $\kappa \in (\varepsilon, 1)$, if we keep the number of association maps below κn at all times, then each expiration time t_{mj} will have a Poisson exponential rate below $\frac{1-\kappa}{n}$.

Consider for instance the j -th association map with respect to the m -th chain of association maps, $\boxed{\mathbf{b}_m} \rightarrow \boxed{\mathbf{a}_{m1}} \rightarrow \boxed{\mathbf{a}_{m2}} \rightarrow \dots \rightarrow \boxed{\mathbf{a}_{mj}} \rightarrow \dots$ (here j is between 1 and $\frac{\varepsilon n}{d}$):

$$\begin{array}{cccccccccccc}
 A_t : & \dots & \boxed{4} & \boxed{6} & \boxed{\mathbf{b}_2} & \boxed{9} & \boxed{\mathbf{a}_{mj}} & \boxed{\mathbf{b}_3} & \boxed{\mathbf{b}_4} & \boxed{\mathbf{b}_1} & \dots \\
 B_t : & \dots & \boxed{4} & \boxed{6} & \boxed{\mathbf{a}_{mj}} & \boxed{9} & \boxed{\mathbf{b}_1} & \boxed{\mathbf{b}_4} & \boxed{\mathbf{b}_2} & \boxed{\mathbf{b}_3} & \dots \\
 & & & & \uparrow & & \uparrow & \uparrow & \uparrow & \uparrow & \\
 & & & & d_1 & & d_2 & d_3 & d_4 & d_5 &
 \end{array}$$

Here, each $\boxed{\mathbf{a}_{mj}}$ does label-to-location jumps according to the above j -th association map. *Each time an association map expires, we construct a new one.* So each time we have $k = \lfloor \varepsilon n \rfloor$ association maps. Thus, the time T_a of cancellation using one of the following transpositions

$$\left\{ \ll \boxed{\mathbf{a}_{mj}}, d_i \gg : (i, m, j) \in \{1, \dots, d\} \times \{1, \dots, d\} \times \{1, \dots, \lfloor \frac{\varepsilon n}{d} \rfloor\} \right\}$$

or $\ll \boxed{\mathbf{b}_m}, d_i \gg$ is *exponential* $(\frac{d(k+1)}{n^2})$.

We will also cancel discrepancies with $\ll d_3, \boxed{\mathbf{a}_{mj}} \gg, \ll d_4, \boxed{\mathbf{a}_{mj}} \gg, \dots, \ll d_d, \boxed{\mathbf{a}_{mj}} \gg$ and two transpositions $\langle d_1, \boxed{\mathbf{a}_{mj}} \rangle_A$ and $\langle d_2, \boxed{\mathbf{a}_{mj}} \rangle_B$ occurring independently until a

discrepancy is canceled. All with respect to j -th association map in the chain of association maps that starts with $\boxed{\mathbf{b}_m}$.

This will not disturb the rates of $\boxed{\mathbf{b}_1}, \boxed{\mathbf{b}_2}, \dots, \boxed{\mathbf{b}_d}$ (excluding $\boxed{\mathbf{b}_m}$), because $\boxed{\mathbf{a}_{mj}}$ will not occupy any of the discrepancy locations until the time of discrepancy cancellation.

We let T_b again be the first time of a jump from one of the discrepancy locations d_i to some $\boxed{\mathbf{a}_{mj}}$ in the top, or the bottom, or both processes. Again, T_b is *exponential* $(\frac{d(k+1)}{n^2})$.

Therefore, the average time for the discrepancy cancellation over all k association maps is $T_d = \min\{T_a, T_b\}$ with

$$E[T_d] \leq \frac{n^2}{2(k+1)d} \leq \frac{n}{2\varepsilon d}.$$

When a discrepancy is cancelled, after $\frac{n}{2\varepsilon d}$ units of time (on average), we will reuse $d-1$ chains of association maps. So at most, only $\varepsilon n/d$ association maps will not be reused.

Every time we cancel one of d discrepancies, we are left with an average of $\frac{1}{2} \cdot \frac{\varepsilon n}{d}$ non-reusable association maps in the corresponding chain as chain of association maps has length $\approx \frac{\varepsilon n}{d}$. There the discrepancy is canceled on one of them, and all the association maps built on top of it can be forgotten, while the average number of $\frac{1}{2} \cdot \frac{\varepsilon n}{d}$ of the remaining association maps cannot be reused to cancel more discrepancies.

If we impose an upper bound κn ($\varepsilon \leq \kappa < 1$) on the number of association maps allowed at any one time, then the average time of discrepancy cancellation will be delayed by at most

$$\frac{n}{1-\kappa} \cdot \log \left(1 + \frac{\varepsilon}{(\kappa-\varepsilon)d} \right) \leq \frac{\varepsilon}{(1-\kappa)(\kappa-\varepsilon)} \cdot \frac{n}{d}.$$

Thus it will take less than

$$\left[\frac{1}{2\varepsilon} + \frac{\varepsilon}{(1-\kappa)(\kappa-\varepsilon)} \right] \cdot n \log n$$

to cancel every discrepancy, each w.r.t. some association map.

After the discrepancies cancel with respect to the association maps, it will take an average of about $\frac{n}{1-\kappa} \log(\kappa n)$ units of time for all remaining association maps to expire. Thus the upper bound on the expectation of coupling time will be

$$\left[\frac{1}{2\varepsilon} + \frac{\kappa}{(1-\kappa)(\kappa-\varepsilon)} \right] \cdot n \log n \quad \text{for any } 0 < \varepsilon < \kappa < 1.$$

We obtained a $Cn \log n$ upper bound on coupling time with optimal $C < 6$. We do, however, believe that improving the above construction to produce $\frac{1}{2\varepsilon} n \log n$ upper bound is achievable.

4.5.1 Computations.

Let the time interval it takes to cancel one of the d discrepancies on one of the $k = \lfloor \varepsilon n \rfloor$ association maps be denoted by I_d . Then $|I_d| = T_d$, where $E[T_d] = \frac{n}{2\varepsilon d}$. We want to impose an upper bound κn ($\varepsilon \leq \kappa < 1$) on the number of association maps allowed at any one

time, then each expiration time t_{mj} for the association map corresponding to $\boxed{\mathbf{a}_{mj}}$ will have Poisson exponential rate $\geq \frac{1-\kappa}{n}$. Let Q_d denote the number of non-reusable association maps at the start of I_d time interval. Then $Q_d < (\kappa - \varepsilon)n$. Let us bound an expectation of the length of time interval J_d (with the same starting point as I_d) requiring enough of Q_d non-reusable association maps to expire, so that there are less than $(\kappa - \varepsilon)n - \frac{\varepsilon n}{d}$ of them left. We let

$$k_* = \max\{0, \lfloor Q_d - (\kappa - \varepsilon - \varepsilon/d)n \rfloor + 1\}$$

be the number of expirations of Q_d non-reusable association maps to occur within J_d time interval. Then by coupon collector argument, the interval's length $|J_d|$ is dominated by the sum of k_* independent exponential random variables with parameters

$$Q_d \cdot \frac{1-\kappa}{n}, (Q_d-1) \cdot \frac{1-\kappa}{n}, \dots, (Q_d-k_*+1) \cdot \frac{1-\kappa}{n}.$$

Thus

$$\begin{aligned} E[|J_d|] &= \frac{n}{1-\kappa} \cdot \left(\sum_{q=Q_d-k_*+1}^{Q_d} \frac{1}{q} \right) \leq \frac{n}{1-\kappa} \cdot \log \left(\frac{Q_d}{Q_d-k_*} \right) \\ &\leq \frac{n}{1-\kappa} \cdot \log \left(1 + \frac{\varepsilon}{(\kappa-\varepsilon)d} \right) \leq \frac{\varepsilon}{(1-\kappa)(\kappa-\varepsilon)} \cdot \frac{n}{d} \end{aligned}$$

So only after $\max\{|I_d|, |J_d|\}$ units of time will we have the discrepancy canceled on one of the association maps, and the total number of non-reusable association maps smaller than $(\kappa - \varepsilon)n$. Only then will we be able to start the new round I_{d-1} (it can also be I_{d-2} or I_{d-3} if more than one discrepancy is cancelled at once) of discrepancy cancelations with respect to association maps. Here

$$E[\max\{|I_d|, |J_d|\}] \leq E[|I_d|] + E[|J_d|] \leq \left[\frac{1}{2\varepsilon} + \frac{\varepsilon}{(1-\kappa)(\kappa-\varepsilon)} \right] \cdot \frac{n}{d}.$$

It will take an average of

$$\sum_d E[\max\{|I_d|, |J_d|\}] \leq \left[\frac{1}{2\varepsilon} + \frac{\varepsilon}{(1-\kappa)(\kappa-\varepsilon)} \right] \cdot n \log n$$

units of time in order to cancel all the discrepancies with respect to the association maps while keeping the number of non-reusable association maps under $(1 - \kappa)n$.

References

- [1] D.Aldous, *Random Walks on Finite Groups and Rapidly Mixing Markov Chains* Seminaire de Probabilites XVII, Lecture Notes in Math. **986** (1983) 243–297
- [2] D.Aldous and J.A.Fill, *Reversible Markov Chains and Random Walks on Graphs* <http://www.stat.berkeley.edu/users/aldous/>

- [3] N.Berestycki and R.Durrett, *A phase transition in the random transposition random walk* Probability Theory and Related Fields **136** (2006), no.2, 203-233
- [4] S.Borovkova, R.M.Burton and H.Dehting, *Limit theorems for functionals of mixing processes with applications to U-statistics and dimension estimation* Trans. Amer. Math. Soc. **353** (2001), no.11, 4261-4318
- [5] K.Burdzy and W.S.Kendall, *Efficient Markovian Couplings: Examples and Counterexamples* Ann. Appl. Prob. **10** (2000), no.2, 362-409
- [6] P.Diaconis, *Group Representations in Probability and Statistics* Institute of Mathematical Statistics, Hayward CA (1973)
- [7] P.Diaconis, *Random Walk on Groups: Geometry and Character Theory* Groups St. Andrews, 2001. Neuman et al, ed. Cambridge: Cambridge University Press (2003)
- [8] P.Diaconis and L.Saloff-Coste, *Comparison Theorems for Reversible Markov Chains* Ann. Prob. **21** (1993), no.4, 2131-2156
- [9] P.Diaconis and M.Shahshahani, *Generating a random permutation with random transpositions* Z. Wahrsch. Verw. Gebiete **57** (1981), 159-179
- [10] N. A.Friedman and D.S.Ornstein, *On the isomorphism of weak Bernoulli transformation* Adv. Math **5** (1970), 365-394
- [11] D.Griffeath, *A Maximal Coupling for Markov Chains* Z. Wahrsch. Verw. Gebiete **31** (1975), 95-106
- [12] S.Goldstein, *Maximal Coupling* Z. Wahrsch. Verw. Gebiete **46** (1979), 193-204
- [13] T. Hayes and E. Vigoda, *A Non-Markovian Coupling for Randomly Sampling Colorings*. Proceedings of the 44th Annual IEEE Symposium on Foundations of Computer Science (2003), 618-627
- [14] C.Hoffman and D.Rudolph, *Uniform Endomorphisms which are isomorphic to a Bernoulli shift* Annals of Mathematics **156** (2002), no.1, 79-101
- [15] P.Matthews, *A strong uniform time for random transpositions* Journal of Theoretical Probability **1** (1988), no.4, 411-423
- [16] D.S.Ornstein, *Bernoulli shifts with the same entropy are isomorphic* Adv. Math **4** (1970), 337-352
- [17] Y.Peres, *Mixing for Markov Chains and Spin Systems*. <http://www.stat.berkeley.edu/users/peres/ubc.pdf>
- [18] J.W.Pitman, *On Coupling of Markov Chains* Z. Wahrsch. Verw. Gebiete **35** (1976), 315-322

-
- [19] L.Saloff-Coste, *Random Walks on Finite Groups* Probability on Discrete Structures, H.Kesten (Editor), Springer (2004), 263-346
- [20] V.A.Volkonskii and Yu. A. Rozanov, *Some limit theorem for random functions*. Theory of Probability and Its Applications (4) (1959), 178-197.