

# Mixing times via super-fast coupling

Robert M. Burton \*

Yevgeniy Kovchegov †

## Abstract

For the probabilistic model of shuffling by random transpositions we provide a coupling construction with the expected coupling time of order  $Cn \log(n)$ , where  $C$  is a moderate constant. We enlarge the methodology of coupling by including intuitive non-Markovian coupling rules. We discuss why a typical Markovian coupling is not always sufficient for resolving mixing time questions.

## 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 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

---

\*Department of Mathematics, Oregon State University, Corvallis, OR 97331-4605, USA [bob@oregonstate.edu](mailto:bob@oregonstate.edu)

†Department of Mathematics, Oregon State University, Corvallis, OR 97331-4605, USA [kovchegy@math.oregonstate.edu](mailto:kovchegy@math.oregonstate.edu)

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 i.i.d. random pairs  $\langle \mathbf{a}, \mathbf{b} \rangle$  which are uniformly distributed on  $\mathcal{S}^2$ . Each pair  $\langle \mathbf{a}, \mathbf{b} \rangle$  generates an independent transposition, 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 configurations of cards. This gives a well defined Markov chain. It will sometimes be useful to Poissonify the Markov chain to continuous time with i.i.d. exponentially distributed interarrival times.

A coupling argument would require the construction of a joint distribution for  $X_t$  and  $Y_t$  on the product of sample spaces  $S_n \times S_n$ , where the distribution of  $X_0$ ,  $dist(X_0)$  is the point mass distribution on  $S_n$ , giving the identity configuration probability 1. This joint distribution must have correct marginal distributions, and the random permutations  $X_t$  and  $Y_t$  must 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  $X_t$  is a Markov Chain on  $S_n$ . The chain is said to be **group invariant** if  $dist(\gamma X_{t+1} | X_t = \alpha)$  is equal to the  $dist(X_{t+1} | X_t = \gamma^{-1}\alpha)$  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  $\alpha$ . The set of group elements with a given cycle structure form a conjugacy class in the group,  $S_n$ . For example, if  $\alpha$  is a transposition, i.e. a 2-cycle, then regardless of what  $\alpha$  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 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  $\sigma$ -fields so that the present and past are measurable with respect to their corresponding  $\sigma$ -fields, and that the  $\sigma$ -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  $X_t$  and  $Y_t$  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  $X_t$  and  $Y_t$  are converging in the total variation distance as  $t$  becomes large.

Diaconis and Shahshahani [9] define a finite state Markov chain  $X_t$  with invariant probability  $U$  to be **strong uniform mixing** if there is a stopping time  $T$ , so that  $P[T = t, X_t = \alpha]$  is independent of the group element  $\alpha$ . Because of this and the invariance of the uniform distribution on the group  $G$ , for any  $\alpha$  we have  $P[X_t = \alpha | t \geq T] = 1/|G|$ , 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,  $\|U - \text{dist}(X_t)\|_{TV} \leq P[T > t]$ , 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  $\beta$ -mixing [20]).

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

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  $X_t$  is **tree weak Bernoulli** if there is a coupling  $\{(X'_t, X''_t) : t \in \mathbb{Z}\}$  such that (i)  $\{X'_t\}$  and  $\{X''_t\}$  have the same distribution as  $\{X_t\}$ , (ii) the past of  $X''_t$ ,  $\{X''_t : t = 0, -1, -2, \dots\}$  is independent of  $\{X'_t\}$ , (iii) there is a random a.s. finite time  $T$  so that  $t \geq T$  implies that  $X'_t = X''_t$ , and (iv) the coupling respects the tree structure of future sample paths, so if  $X'_t$  and  $X''_t$  are coupled at time  $t > 0$  then each of their descendants (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  $\{X_t\}_{t \in \{1, 2, \dots, M\}}$  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.

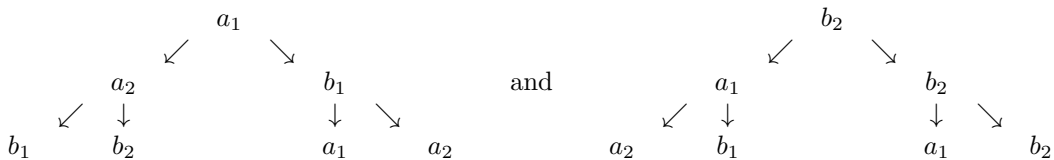
In any given situation, it is possible, in principle, to find the optimal coupling because the total variation norm is achievable. However, the size of the state space and dependence within the process often make this impractical.

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 a path-wise property and requires 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 as given in the following transition probability matrix

	$a_1$	$a_2$	$b_1$	$b_2$
$a_1$	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  $Y_0 = b_2$ . The 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

$X_0 = 123$ , 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 1s. The maximum total variation distance between any starting point and any other starting point is  $(21 - 12)/81 = 1/9$ . If we raise  $M$  to the power  $2t$  we get  $M^{2t} = (9I + 12N)^t = a_t I + b_t N$ , where  $a_t = 9^t$ . This means that after  $2t$  iterations the total variation norm between any two distinct starting places is  $1/3^{2t}$ , so the mixing rate is geometric with error of  $\frac{1}{2 \cdot 3^{2t}}$ . 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  $X_t$  and the uniform  $U$  as

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

So for all  $t$  we have

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

If we use the  $Q \times P$  association, picking a face value and then picking a location to swap cards with, and if  $X_0 = 123$ , then the first transposition will have nine equally likely outcomes  $X_1$ , three of which are configuration 123, and there are two chances for each of three configurations, 132, 213, and 321. These are shown in the chart below. In keeping with group invariance, each turn there 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  $X_0 = 132$  using group invariance, we get the following table:

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

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: an example

In this short section we present an example of a coupling that utilizes the same approach as the coupling construction for the shuffling by random transpositions that will be described in section 4. 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  $X_t$  changes to  $1 - X_t$  with rate 10, independently of the second coordinate  $Y_t$ . 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  $X'_t$ . Coupling the second coordinate is similarly 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

We consider an  $n$  card deck. A random transposition shuffle is generated by making two independent uniform choices of cards, and interchanging them. We assume this being a continuous time process, where the transpositions happen one at a time with exponential waiting times of rate one in between. For each pair of distinct cards  $\boxed{\mathbf{a}}$  and  $\boxed{\mathbf{b}}$  there are two ways to order them,  $\langle \boxed{\mathbf{a}}, \boxed{\mathbf{b}} \rangle$  and  $\langle \boxed{\mathbf{b}}, \boxed{\mathbf{a}} \rangle$ , each with equal probability. Thus, each transposition  $\langle \boxed{\mathbf{a}}, \boxed{\mathbf{b}} \rangle$  occurs with rate  $\frac{2}{n^2}$ , i.e. the two identical transpositions  $\langle \boxed{\mathbf{a}}, \boxed{\mathbf{b}} \rangle$  and  $\langle \boxed{\mathbf{b}}, \boxed{\mathbf{a}} \rangle$  will each happen with rate  $\frac{1}{n^2}$ . The choice of two identical cards happens with probability  $\frac{1}{n}$ , and the unnecessary transposition  $\langle \boxed{\mathbf{a}}, \boxed{\mathbf{a}} \rangle$  is therefore assigned the 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] a classical coupling approach was shown to produce the upper

bound of order  $O(n^2)$ , while [17] lists showing  $O(n \log n)$  mixing time via *coupling* construction as an open problem. In this section we produce such coupling construction.

Throughout this section we will use the following 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>	- 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

#### 4.1 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  $i$  in both the top and the bottom processes,  $A_t$  and  $B_t$ , was applied at each iteration. 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 upper bound can be achieved by slightly modifying the coupling rules. On each iteration, we randomly select a card  $\boxed{\mathbf{a}}$ . If  $\boxed{\mathbf{a}}$  is not coupled, then we apply transposition  $\ll \boxed{\mathbf{a}}, i \gg$  for a randomly selected location  $i$ . If  $\boxed{\mathbf{a}}$  is one of the coupled cards, we apply transposition  $\ll \boxed{\mathbf{a}}, \boxed{\mathbf{b}} \gg$ , transposing  $\boxed{\mathbf{a}}$  with a randomly selected card  $\boxed{\mathbf{b}}$ , in both processes,  $A_t$  and  $B_t$ .

This latter, slightly modified 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 in spirit to the one used by Euler in computing the number of permutations of  $n$  elements with all elements displaced.

For the rest of the section, transpositions  $\ll \boxed{\mathbf{a}}, i \gg$  will be called *label-to-location*, while transpositions  $\ll \boxed{\mathbf{a}}, \boxed{\mathbf{b}} \gg$  will be called *label-to-label*.

Now, without loss of generality, we would like to state that, in the last coupling construction, whenever a coupled card, for example  $\boxed{\mathbf{a}}$ , is selected, and thus a transposition  $\ll \boxed{\mathbf{a}}, \boxed{\mathbf{b}} \gg$  with another card  $\boxed{\mathbf{b}}$  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. In general, for any permutation  $\sigma$  of face values on the cards, the coupled process  $\begin{pmatrix} A_t \\ B_t \end{pmatrix}$  is isomorphic to  $\begin{pmatrix} \sigma A_t \\ \sigma B_t \end{pmatrix}$  under what we will call *the group invariance* of the coupled process. The point of mentioning the group invariance property is to stress our right to suppress label-to-label transpositions in the coupled process whenever necessary.

## 4.2 Improving the coupling

The basic strategy for the type of coupling constructions described in this paper is to condition on a key  $\sigma$ -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 legitimate 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.

In order to illustrate the forthcoming coupling construction and the notations used therein let us consider the following example.

**Example.** Consider a deck of 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  $\ll \boxed{2}, i_1 \gg$  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 ( $\ll \boxed{4}, \boxed{3} \gg_A$ ), getting

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

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

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

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  $\ll \boxed{4}, \boxed{1} \gg$  and  $\ll \boxed{4}, \boxed{2} \gg$  getting

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

respectively, which we suppress noting. The rates being the same for each one of  $\ll \boxed{4}, \boxed{3} \gg_A$ ,  $\ll \boxed{4}, \boxed{3} \gg_B$ ,  $\ll \boxed{4}, \boxed{1} \gg$  and  $\ll \boxed{4}, \boxed{2} \gg$ .

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

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

$$\ll \boxed{4}, \boxed{3} \gg_A \quad \text{followed by} \quad \ll \boxed{2}, i_1 \gg$$

or

$$\ll \boxed{4}, \boxed{3} \gg_B \quad \text{followed by} \quad \ll \boxed{2}, i_1 \gg$$

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.

#### 4.2.1 Two discrepancies and one association map ( $d = 2, k = 1$ )

We will start with the case of two discrepancies,  $d_1$  and  $d_2$  as illustrated below, in (1), when the coupling speed must be the slowest, and introduce the first improvement to the classical coupling construction.

$$\begin{array}{cccccccccccc} A_t : & \dots & \boxed{4} & \boxed{6} & \boxed{b} & \boxed{9} & \boxed{a} & \boxed{8} & \boxed{a1} & \boxed{2} & \dots \\ B_t : & \dots & \boxed{4} & \boxed{6} & \boxed{a} & \boxed{9} & \boxed{b} & \boxed{8} & \boxed{a1} & \boxed{2} & \dots \end{array} \quad (1)$$

$$\begin{array}{ccccccc} & & & \uparrow & \uparrow & \uparrow & \\ & & & d_2 & d_1 & i_1 & \end{array}$$

If we let the uncoupled cards  $\boxed{a}$  and  $\boxed{b}$  do the label-to-location transpositions, and the coupled cards do the label-to-label transpositions, where the transpositions are synchronized for the top and the bottom decks,  $A_t$  and  $B_t$ . Then, because of the group invariance property that was mentioned earlier in the paper, the above configuration will not change until either of the uncoupled cards  $\boxed{a}$  or  $\boxed{b}$  transposes with one of the discrepancy locations  $d_1$  or  $d_2$ , simultaneously in both the top and the bottom processes. The label-to-label transpositions of the uncoupled cards can be suppressed citing group invariance. 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. Next, we introduce the first modification to the coupling construction.

We can adjust the coupling as follows. We take one of the two uncoupled cards, for example  $\boxed{a}$ . Randomly select a site  $i_1$ , and a random exponential time  $t_1$ , and condition on transposition  $\ll \boxed{a}, i_1 \gg$  happening at time  $t_1$ . All with respect to group invariance. If  $i_1 = d_1$  or  $d_2$ , the discrepancies disappear, and the coupling is completed by letting all coupled cards do label-to-label transpositions that we suppress to notice, citing the group invariance. We want to utilize  $\ll \boxed{a}, i_1 \gg$  if  $i_1 \neq d_1$  or  $d_2$ .

So we condition on the label-to-location transposition  $\ll \boxed{a}, i_1 \gg$  occurring at time  $t_1$ . If  $i_1 \neq d_1$  or  $d_2$ , the following construction will help us set the transition rates for the coupling in the time interval  $[0, t_1]$ . The first step is to associate locations in the top and the bottom decks as follows:

- (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  $d_1/d_2$

Therefore, at time  $t = 0$ , the decks are aligned accordingly:

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

The above diagram reads as follows. The card  $\boxed{\mathbf{b}}$  in the upper deck is located at the site  $d_2$ , while the card  $\boxed{\mathbf{a1}}$  in the lower deck is located at the site  $i_1$ . Similarly, the card  $\boxed{\mathbf{a1}}$  in the upper deck is located at the site  $i_1$ , while the card  $\boxed{\mathbf{b}}$  in the lower deck is located at the site  $d_1$ . Last, the card  $\boxed{\mathbf{a}}$  in the upper deck is located at the site  $d_1$ , while the card  $\boxed{\mathbf{a}}$  in the lower deck is located at the site  $d_2$ . So the above configuration is exactly the same as in (1).

What is different is that the transposition  $\ll \boxed{\mathbf{a}}, i_1 \gg$  in (1) is equivalent to the relabeling of the sites in (2) described in the following diagram:

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

The site association (2) will be called the association map with respect to transposition  $\ll \boxed{\mathbf{a}}, i_1 \gg$  and jump time  $t_1$ . 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 set the rates for the new coupling process in the time interval  $[0, t_1]$ :

- Rates for  $\boxed{\mathbf{a}}$ : the first transposition  $\ll \boxed{\mathbf{a}}, i_1 \gg$  occurs at time  $t_1$ .
- Rates for  $\boxed{\mathbf{a1}}$ :  $\boxed{\mathbf{a1}}$  does the label-to-location transpositions, where locations are defined by the association map (2). In other words, transpositions  $\ll \boxed{\mathbf{a1}}, i \gg$  ( $i \neq i_1, d_1, d_2$ ),  $\ll \boxed{\mathbf{a1}}, i_1/d_1 \gg$ ,  $\ll \boxed{\mathbf{a1}}, d_2/i_1 \gg$  and  $\ll \boxed{\mathbf{a1}}, d_1/d_2 \gg$  occur with the usual rate of  $\frac{1}{n^2}$ .

Transposition  $\ll \boxed{\mathbf{a1}}, i_1/d_1 \gg$  should be interpreted as simultaneous occurrence of transpositions  $\langle \boxed{\mathbf{a1}}, i_1 \rangle_A$  and  $\langle \boxed{\mathbf{a1}}, d_1 \rangle_B$  in the top and the bottom decks. Similarly  $\ll \boxed{\mathbf{a1}}, d_2/i_1 \gg$  should be interpreted as simultaneous occurrence of transpositions  $\langle \boxed{\mathbf{a1}}, d_2 \rangle_A$  and  $\langle \boxed{\mathbf{a1}}, i_1 \rangle_B$  in the top and the bottom decks. Transposition  $\ll \boxed{\mathbf{a1}}, d_1/d_2 \gg = \ll \boxed{\mathbf{a1}}, \boxed{\mathbf{a}} \gg$  is label-to-label and therefore should be suppressed.

- Rates for  $\boxed{\mathbf{b}}$ : transpositions  $\ll \boxed{\mathbf{b}}, d_1 \gg$  and  $\ll \boxed{\mathbf{b}}, d_2 \gg$ ,  $\langle \boxed{\mathbf{b}}, \boxed{\mathbf{a1}} \rangle_A$ ,  $\langle \boxed{\mathbf{b}}, \boxed{\mathbf{a1}} \rangle_B$  and  $\ll \boxed{\mathbf{b}}, i \gg$  ( $i \neq i_1, d_1, d_2$ ) occur independently with rate  $\frac{1}{n^2}$  each.
- The rest of the cards do label-to-label jumps simultaneously in the top and the bottom process, which we suppress according to group invariance.

We need to add one more condition. We observe that any of the following six transpositions combined with the occurrence of transposition  $\ll \boxed{\mathbf{a}}, i_1 \gg$  at time  $t_1$  will lead to the discrepancies cancelation. The

six transpositions are  $\ll \boxed{\mathbf{a1}}, i_1/d_1 \gg$ ,  $\ll \boxed{\mathbf{a1}}, d_2/i_1 \gg$ ,  $\ll \boxed{\mathbf{b}}, d_1 \gg$ ,  $\ll \boxed{\mathbf{b}}, d_1 \gg$ ,  $\langle \boxed{\mathbf{b}}, \boxed{\mathbf{a1}} \rangle_A$ , and  $\langle \boxed{\mathbf{b}}, \boxed{\mathbf{a1}} \rangle_B$ .

Therefore, if one of the above mentioned six transpositions occurs before time  $t_1$ , from that time until time  $t_1$ , all the cards other than  $\boxed{\mathbf{a}}$  will only be allowed to do label-to-label jumps.

Now, since we classified (and therefore suppressed)  $\ll \boxed{\mathbf{a1}}, d_1/d_2 \gg$  as a label-to-label transposition, the jump time  $t_2$  for the card  $\boxed{\mathbf{a1}}$  is exponential with rate  $\frac{n-1}{n^2}$ . Now,  $P[t_2 < t_1] = \frac{\frac{n-1}{n^2}}{1 + \frac{n-1}{n^2}} = \frac{n-1}{2n-1}$ , and conditioning on  $t_2 < t_1$ , either  $\ll \boxed{\mathbf{a1}}, i_1/d_1 \gg$  or  $\ll \boxed{\mathbf{a1}}, d_2/i_1 \gg$  will occur with probability  $\frac{2}{n-1}$ . Once again observe that both will result in the discrepancy cancellation at time  $t_1$ .

Let  $t_b$  denote 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. Then  $t_b$  is exponential with rate  $\frac{4}{n^2}$ . Recall that if  $t_b < t_1$  then the discrepancies will cancel out at time  $t_1$ .

Therefore the probability of canceling the discrepancy by time  $t_1$  is

$$P[i_1 = d_1 \text{ or } d_2] + \left(1 - \frac{2}{n}\right) \cdot P[t_2 < t_1] \cdot \frac{2}{n-1} + \left(1 - \frac{2}{n} \left(2 - \frac{1}{n-1}\right)\right) P[t_b < t_1] \approx \frac{8}{n},$$

where  $P[i_1 = d_1 \text{ or } d_2] + \left(1 - \frac{2}{n}\right) \cdot P[t_2 < t_1] \cdot \frac{2}{n-1} = \frac{2}{n} + \left(1 - \frac{2}{n}\right) \cdot \frac{n-1}{2n-1} \cdot \frac{2}{n-1} = \frac{2}{n} \left(2 - \frac{1}{n-1}\right)$  is the probability that the discrepancy is canceled by  $\ll \boxed{\mathbf{a}}, i_1 \gg$  when  $i = i_1$ , or as the result of one of the two transpositions,  $\ll \boxed{\mathbf{a1}}, i_1/d_1 \gg$  or  $\ll \boxed{\mathbf{a1}}, d_2/i_1 \gg$ , combined with  $\ll \boxed{\mathbf{a}}, i_1 \gg$ . Now, 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 discrepancies can be canceled out with one of the four transpositions,  $\ll \boxed{\mathbf{b}}, d_1 \gg$ ,  $\ll \boxed{\mathbf{b}}, d_1 \gg$ ,  $\langle \boxed{\mathbf{b}}, \boxed{\mathbf{a1}} \rangle_A$  or  $\langle \boxed{\mathbf{b}}, \boxed{\mathbf{a1}} \rangle_B$ , combined with  $\ll \boxed{\mathbf{a}}, i_1 \gg$ .

So, with probability  $\approx \frac{8}{n}$ , the discrepancies will 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  $\frac{n^2}{4}$ . However we can do better by using more than one association map at a time.

**Example.** Next, we illustrate how a cancellation of discrepancies on an association map (2) implies discrepancies' cancellation at time  $t_1$  on (1). Consider the following case. If  $t_2 < t_1$ , and if  $\ll \boxed{\mathbf{a1}}, d_2/i_1 \gg$  occurs at time  $t_2$ , 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 (1) will be as follows. From

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

to

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

at time  $t_2$ , and to

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

at time  $t_1$ .

#### 4.2.2 Two discrepancies and $\varepsilon n$ association maps ( $d = 2, k = \lfloor \varepsilon n \rfloor$ )

Next we develop a calculus of association maps to show the coupling construction for the case of two discrepancies with coupling time of order  $O(n \log n)$ .

Since the rate of  $\frac{n^2}{8}$  is still not good enough, we need to revise and enhance the coupling construction by introducing chains of association maps. The first part is similar to the preceding subsection. Namely, we begin by randomly selecting a site  $i_1$ , and a random exponential time  $t_1$ , and conditioning on transposition  $\ll \boxed{a}, i_1 \gg$  happening at time  $t_1$ . See (1).

In order to set the rates for  $\boxed{a1}$  in the coupling process, we considered the following association map

$$\begin{array}{cccccccccccc} A_t : & \dots & \boxed{a2} & \boxed{6} & \boxed{b} & \boxed{9} & \boxed{a1} & \boxed{8} & \boxed{a} & \boxed{2} & \dots \\ B_t : & \dots & \boxed{a2} & \boxed{6} & \boxed{a1} & \boxed{9} & \boxed{b} & \boxed{8} & \boxed{a} & \boxed{2} & \dots \\ & & \uparrow & & \uparrow & & \uparrow & & \uparrow & & \\ & & i_2 & & d_2^* & & d_1^* & & i_1^* & & \end{array} \quad (3)$$

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  $i_2 \in \{1, 2, \dots, d_1^*, \dots, d_2^*, \dots, n\} \setminus \{i_1^*\}$  on the association map and a random time  $t_2$ , distributed exponentially  $\frac{1}{n} \cdot (1 - \frac{1}{n})$ . We condition on the transposition  $\ll \boxed{a1}, i_2 \gg$  happening at the time  $t_2$ . The group invariance allows us to suppress  $\ll \boxed{a1}, i_1^* \gg$  as  $\ll \boxed{a1}, i_1^* \gg = \ll \boxed{a1}, \boxed{a} \gg$  is label-to-label.

On the new scheme, if  $i_2 = d_1^*$  or  $d_2^*$ , the discrepancies cancel out, and the process couples after time  $\tau_2 = \max\{t_1, t_2\}$ . See the example in the preceding subsection.

Now, if  $i_2 \neq d_1^*$  or  $d_2^*$ , we will construct one more association map in order to set the transition rates for the card  $\boxed{\mathbf{a}2}$  located at  $i_2$ . We pair the site  $i_2$  in  $A_t$  with site  $d_1^*$  in  $B_t$  and call it  $i_2/d_1^*$ ; pair the site  $d_2^*$  in  $A_t$  with site  $i_2$  in  $B_t$  and call it  $d_2^*/i_2$ ; and finally, pair the site  $d_1^*$  in  $A_t$  with site  $d_2^*$  in  $B_t$  and call it  $d_1^*/d_2^*$ . After pairing anew the locations on top with the locations at the bottom, the association map (3) is realigned as

$$\begin{array}{cccccccccccc} A_t : & \dots & \boxed{\mathbf{a}1} & \boxed{6} & \boxed{\mathbf{b}} & \boxed{9} & \boxed{\mathbf{a}2} & \boxed{8} & \boxed{\mathbf{a}} & \boxed{2} & \dots \\ B_t : & \dots & \boxed{\mathbf{a}1} & \boxed{6} & \boxed{\mathbf{a}2} & \boxed{9} & \boxed{\mathbf{b}} & \boxed{8} & \boxed{\mathbf{a}} & \boxed{2} & \dots \\ & & \uparrow & & \uparrow & & \uparrow & & \uparrow & & \\ & & d_1^*/d_2^* & & d_2^*/i_2 & & i_2/d_1^* & & i_1^* & & \end{array}$$

In other word, we construct the second association map

$$\begin{array}{cccccccccccc} A_t : & \dots & \boxed{\mathbf{a}1} & \boxed{6} & \boxed{\mathbf{b}} & \boxed{9} & \boxed{\mathbf{a}2} & \boxed{8} & \boxed{\mathbf{a}} & \boxed{\mathbf{a}3} & \dots \\ B_t : & \dots & \boxed{\mathbf{a}1} & \boxed{6} & \boxed{\mathbf{a}2} & \boxed{9} & \boxed{\mathbf{b}} & \boxed{8} & \boxed{\mathbf{a}} & \boxed{\mathbf{a}3} & \dots \\ & & \uparrow & & \uparrow & & \uparrow & & \uparrow & \uparrow & \\ & & i_2^{**} & & d_2^{**} & & d_1^{**} & & i_1^* & i_3 & \end{array} \quad (4)$$

where

- (a)  $d_1^{**}$  denotes  $i_2/d_1^*$  before  $t_2$ , and  $d_1^*$  after  $t_2$ ,
- (b)  $d_2^{**}$  denotes  $d_2^*/i_2$  before  $t_2$ , and  $d_2^*$  after  $t_2$ ,
- (c)  $i_2^{**}$  denotes  $d_1^*/d_2^*$  before  $t_2$ , and  $i_2$  after  $t_2$ .

Next, we pick a random location  $i_3 \in \{1, 2, \dots, d_1^{**}, \dots, d_2^{**}, \dots, n\} \setminus \{i_1^*, i_2^{**}\}$  on the association map (4). We pick a random waiting time  $t_3$ , distributed exponentially with the parameter  $\frac{1}{n} \cdot (1 - \frac{2}{n})$ , and condition on transposition  $\ll \boxed{\mathbf{a}2}, i_3 \gg$  on the association map (4) at time  $t_3$ . Once again the group invariance allows us to suppress  $\ll \boxed{\mathbf{a}2}, i_1^* \gg$  and  $\ll \boxed{\mathbf{a}2}, i_2^{**} \gg$  as  $\ll \boxed{\mathbf{a}2}, i_1^* \gg = \ll \boxed{\mathbf{a}2}, \boxed{\mathbf{a}} \gg$  and  $\ll \boxed{\mathbf{a}2}, i_2^{**} \gg = \ll \boxed{\mathbf{a}2}, \boxed{\mathbf{a}1} \gg$  are label-to-label.

If  $i_3 = d_1^{**}$  or  $d_2^{**}$ , the discrepancies cancel out on the association map (4), 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 in order to set up the transition rates for the card  $\boxed{\mathbf{a}3}$  located at  $i_3$ . 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 construct a chain of at most  $k = \lfloor \varepsilon n \rfloor$  association maps like that, where  $\varepsilon \in (0, 1)$  is fixed, thus setting the transition rates for the cards  $\boxed{\mathbf{a}}, \boxed{\mathbf{a}1}, \boxed{\mathbf{a}2}, \dots, \boxed{\mathbf{a}k}$ . As for all the cards different (w.r.t. group invariance) from  $\boxed{\mathbf{a}}, \boxed{\mathbf{a}1}, \boxed{\mathbf{a}2}, \dots, \boxed{\mathbf{a}k}$  and  $\boxed{\mathbf{b}}$ , all their jumps are set as label-to-label transpositions, which we suppress.

Let  $T_a$  be the first time one of the transpositions

$$\ll \boxed{\mathbf{a}}, d_1 \gg \text{ and } \ll \boxed{\mathbf{a}}, d_2 \gg, \quad \text{and } \ll \boxed{\mathbf{a}j}, d_1^{j*} \gg \text{ and } \ll \boxed{\mathbf{a}j}, d_2^{j*} \gg \text{ for all } j = 1, \dots, k$$

occurs. Then  $T_a$  is exponential with rate  $\frac{2(k+1)}{n^2}$ . Observe that each of the above transpositions leads to the discrepancies cancellation on of the  $k$  association maps, or in the original association of sites.

Lastly, we have to set the rates for  $\boxed{\mathbf{b}}$ . We let the card  $\boxed{\mathbf{b}}$  make label-to-label jumps  $\ll \boxed{\mathbf{b}}, \boxed{\mathbf{c}} \gg$  simultaneously in the top in the bottom decks with rate  $\frac{1}{n^2}$  whenever the card  $\boxed{\mathbf{c}}$  is **not**  $\boxed{\mathbf{a}}$ , or  $\boxed{\mathbf{a}1}$ , or  $\boxed{\mathbf{a}2}, \dots, \boxed{\mathbf{a}k}$ . As for transposing card  $\boxed{\mathbf{b}}$  with  $\boxed{\mathbf{a}}, \boxed{\mathbf{a}1}, \boxed{\mathbf{a}2}, \dots, \boxed{\mathbf{a}k}$ , we let the transpositions

$$\langle \boxed{\mathbf{b}}, \boxed{\mathbf{a}} \rangle_A, \langle \boxed{\mathbf{b}}, \boxed{\mathbf{a}} \rangle_B, \langle \boxed{\mathbf{b}}, \boxed{\mathbf{a}1} \rangle_A, \langle \boxed{\mathbf{b}}, \boxed{\mathbf{a}1} \rangle_B, \dots, \langle \boxed{\mathbf{b}}, \boxed{\mathbf{a}k} \rangle_A, \langle \boxed{\mathbf{b}}, \boxed{\mathbf{a}k} \rangle_B$$

occur independently in either the top or the bottom decks with the standard rate of  $\frac{1}{n^2}$ , up until the two discrepancies get canceled.

If  $T_b$  is the first time one of the label-to-label jumps

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

occurs, then  $T_b$  must be exponential with rate  $\frac{2(k+1)}{n^2}$ . Each one of these jumps leads to the discrepancies cancelation on one of the  $k$  association maps, or in the original association of sites. Thus the average time for the discrepancies cancelation on at least one of the association maps will be bounded above by

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

Now, the discrepancies cancelation on anyone of the  $k$  association map leads to cancelation for the original association of locations, and coupling by the time  $\tau_k = \max\{t_1, t_2, \dots, t_k\}$  when each association map expires. Here for each  $j \in \{1, 2, \dots, k\}$ ,

$$t_j \text{ is exponential with rate } \frac{1}{n} \cdot \left(1 - \frac{j-1}{n}\right),$$

and 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  $\frac{n}{1-\varepsilon} \log(\varepsilon n)$  units of time.

Thus the upper bound of  $\frac{n}{1-\varepsilon} \log(\varepsilon n) + \frac{n}{4\varepsilon}$  on the expected coupling time  $E[T_c]$  is obtained in the case of two discrepancies.

### 4.3 The general case of $d \geq 2$ discrepancies

In this subsection we consider the situation when there are  $d \geq 2$  discrepancies. For each of the  $d$  discrepancy cards  $\boxed{\mathbf{b}_m}$  ( $m = 1, \dots, d$ ), we condition on the first label-to-location jump  $\ll \boxed{\mathbf{b}_m}, i_{m,1} \gg$ , and use this to set the transition rates for the cards  $\boxed{\mathbf{a}_{m,1}}$  occupying the locations  $i_{m,1}$  at time zero, all with respect to the group invariance. We do this by creating corresponding association maps. Next, we condition on the jumps  $\ll \boxed{\mathbf{a}_{m,1}}, i_{m,2} \gg$ , where  $i_{m,2}$  ( $m = 1, \dots, d$ ) are locations on the corresponding association maps. We use this in order to set the rates for the cards  $\boxed{\mathbf{a}_{m,2}}$  occupying the locations  $i_{m,2}$  at time zero. Thus for each  $m$  we create a chain of conditionings  $\boxed{\mathbf{b}_m} \rightarrow \boxed{\mathbf{a}_{m,1}} \rightarrow \boxed{\mathbf{a}_{m,2}} \rightarrow \dots \rightarrow \boxed{\mathbf{a}_{m,j}} \rightarrow \dots$ . We let the rest of the cards do label-to-label jumps, simultaneously in the top and the bottom decks.

Note on the notations: in this subsection, we will not use the  $*$  symbol in the notations of discrepancies on the association maps.

In the picture example below,  $d_1, d_2, \dots, d_d$  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 general case of  $d \geq 2$  discrepancies, for each discrepancy card  $\boxed{\mathbf{b}_1}, \boxed{\mathbf{b}_2}, \dots, \boxed{\mathbf{b}_d}$  we construct a chain of association maps via conditioning, adding new association maps at expiration times. Each chain

will determine the rates for  $\boxed{b_m}, \boxed{a_{m,1}}, \boxed{a_{m,2}}, \dots$ . A transposition  $\ll \boxed{a_{m,j}}, d_i \gg$  will cancel discrepancies on the  $j$ th association map, in the  $m$ th chain of association maps. For each  $m$  from 1 to  $d$ , a transposition  $\ll \boxed{b_m}, i_{m1} \gg$  is set to take place at time  $t_{m,1}$ , and conditioning on this, an association map is created. Here again,  $i_{2,1} \neq i_{1,1}, i_{3,1} \neq i_{1,1}, i_{2,1}$  etc., and therefore  $t_{1,1}, t_{2,1}, \dots$  are exponential random variables with respective rates  $\frac{1}{n}, \frac{n-1}{n^2}, \frac{n-2}{n^2}, \dots$ . Transposition  $\ll \boxed{a_{m,1}}, i_{m,2} \gg$ , where  $i_{m,2}$  ( $i_{m,2} \neq i_{1,1}, \dots, i_{d,1}, i_{2,1}, \dots, i_{(m-1),2}$ ) is a location with respect the corresponding association map, is conditioned to take place at time  $t_{m2}$ , and a new association map is created. 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 conditionings below  $\kappa n$  at all times, then each expiration time  $t_{mj}$  will be assigned a Poisson exponential rate of at most  $\frac{1-\kappa}{n}$ .

Observe that in the  $d = 2$  case considered in the preceding subsection, the main point of the construction was that without the adjustment, only the discrepancy cards could cancel the discrepancies. There, using the conditionings and association maps we adjusted the coupling to enable discrepancy cancellations via the jumps by the non-discrepancy cards  $\boxed{a_1}, \boxed{a_2}, \dots, \boxed{a_k}$  to discrepancy locations on the corresponding association maps. Here too, we want the non-discrepancy cards to cancel the discrepancies. We condition on the jumps of  $\boxed{b_1}, \boxed{b_2}, \dots, \boxed{b_d}$  to make  $\boxed{a_{1,1}}, \boxed{a_{2,1}}, \dots, \boxed{a_{d,1}}$  the new discrepancy cards on the corresponding association maps, enabling them to cancel the discrepancies by jumping to a discrepancy location.

$$\begin{array}{cccccccccccc}
 A_t : & \dots & \boxed{4} & \boxed{6} & \boxed{a_{2,1}} & \boxed{9} & \boxed{a_{8,1}} & \boxed{a_{3,1}} & \boxed{a_{4,1}} & \boxed{a_{1,1}} & \dots \\
 B_t : & \dots & \boxed{4} & \boxed{6} & \boxed{a_{8,1}} & \boxed{9} & \boxed{a_{1,1}} & \boxed{a_{4,1}} & \boxed{a_{2,1}} & \boxed{a_{3,1}} & \dots \\
 & & & & \uparrow & & \uparrow & \uparrow & \uparrow & \uparrow & \\
 & & & & d_1 & & d_2 & d_3 & d_4 & d_5 & 
 \end{array}$$

We condition on the jumps of  $\boxed{a_{1,1}}, \boxed{a_{2,1}}, \dots, \boxed{a_{d,1}}$  to make  $\boxed{a_{1,2}}, \boxed{a_{2,2}}, \dots, \boxed{a_{d,2}}$  the discrepancy cards on the new association maps, and so on. Once a discrepancy is canceled out on an association map, the card that was coupled will not be used in the consequent rounds of conditionings.

Therefore, the average time for the discrepancy cancellation over all  $k$  association maps is

$$E[T_a] \leq \frac{n^2}{kd} \leq \frac{n}{\varepsilon d}$$

Since we are dealing with random transpositions, the construction can be adjusted to couple twice as fast. So without loss of generality we expect the first discrepancy cancellation time to be  $\leq \frac{n}{2\varepsilon d}$ .

When a discrepancy is canceled, after an average of  $\frac{n}{2\varepsilon d}$  units of time, we will reuse all of the  $d-1$  chains of association maps that were not responsible for the cancellation. So at most,  $\lceil \frac{\varepsilon n}{d} \rceil$  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  $\approx \frac{\varepsilon n}{d}$ .

If we impose an upper bound  $\kappa n$  ( $\varepsilon \leq \kappa < 1$ ) 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

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

time steps 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

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

Thus we obtained a  $Cn \log n$  upper bound on coupling time, where after taking optimal  $\varepsilon$  and  $\kappa$ , constant  $C$  becomes fairly small. Namely  $C < 6$ . Observe that in the above constructed upper bound, the term  $\frac{1}{2\varepsilon}n \log n$  is responsible for the discrepancy cancelations on the association maps. From the information theory perspective this points at the  $\frac{1}{2}n \log n$  cut-off for the mixing time.

## 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.