# Mixing times for random $k$-cycles and coalescence-fragmentation chains

Nathanaël Berestycki* Oded Schramm† Ofer Zeitouni‡

January 10, 2010

**Dedicated to the memory of ODED SCHRAMM**

**Abstract**

Let $\mathcal{S}_n$ be the permutation group on $n$ elements, and consider a random walk on $\mathcal{S}_n$ whose step distribution is uniform on $k$-cycles. We prove a well-known conjecture that the mixing time of this process is $(1/k)n \log n$, with threshold of width linear in $n$. Our proofs are elementary and purely probabilistic, and do not appeal to the representation theory of $\mathcal{S}_n$.

*CMS - Wilberforce Rd., Cambridge University, Cambridge CB3 0WB, United Kingdom. Email: N.Berestycki@statslab.cam.ac.uk, http://www.statslab.cam.ac.uk/∼beresty/. The work of this author was partially supported by EPSRC grant EP/GO55068/1

†December 10, 1961 – September 1, 2008.

‡Faculty of Mathematics, Weizmann Institute, POB 26, Rehovot 76100, Israel and School of Mathematics, University of Minnesota, 206 Church St. SE, Minneapolis, MN 55455, USA. Email: zeitouni@math.umn.edu, http://www.wisdom.weizmann.ac.il/∼zeitouni/. The work of this author was partially supported by NSF grant DMS-0804133 and by a grant from the Israel Science Foundation

# 1 Introduction

## 1.1 Main result

Let $\mathcal{S}_n$ be the group of permutations of $\{1, \ldots, n\}$. Any permutation $\sigma \in \mathcal{S}_n$ has a unique cycle decomposition, which partitions the set $\{1, \ldots, n\}$ into orbits under the natural action of $\sigma$. The cycle structure of $\sigma$ is the integer partition of $n$ associated with this set partition, in other words, the ordered sizes of the cycles (blocks of the partition) ranked in decreasing size. It is customary not to include the fixed points of $\sigma$ in this structure. For instance, the permutation

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 4 & 2 & 6 & 7 & 3 & 5 & 1 \end{pmatrix}$$

has 3 cycles, (1 4 7)(2)(3 6 5), so its cycle structure is $(3, 3)$ (and one fixed point which doesn't appear in this structure). A conjugacy class $\Gamma \subset \mathcal{S}_n$ is the set of permutations having a given cycle structure. Let $|\Gamma|$ denote the support of $\Gamma$, that is, the number of non fixed-points of any permutation $\sigma \in \Gamma$. It is well-known and easy to see that if $|\Gamma|$ is even, then $\Gamma$ generates $\mathcal{S}_n$, while if $|\Gamma| > 2$ is odd, then $\Gamma$ generates the alternate group $\mathcal{A}_n$ of even permutations. Let $(\pi_t, t \geq 0)$ be the continuous-time random walk associated with $(\mathcal{S}_n, \Gamma)$. That is, let $\gamma_1, \gamma_2, \ldots$ be a sequence of i.i.d. elements uniformly distributed on $\Gamma$, and let $(N_t, t \geq 0)$ be an independent Poisson process with rate 1, then we take:

$$\pi_t = \gamma_1 \circ \ldots \circ \gamma_{N_t} \tag{1}$$

where the dot product $\gamma \circ \gamma'$ indicates the composition of the permutations $\gamma$ and $\gamma'$. $(\pi_t, t \geq 0)$ is a Markov chain on $\mathcal{S}_n$ which converges to the uniform distribution $\mu$ on $\mathcal{S}_n$ when $|\Gamma|$ is even, and to the uniform distribution on $\mathcal{A}_n$ when $|\Gamma| > 2$ is odd. In any case we shall write $\mu$ for that limiting distribution. We shall be interested in the mixing properties of this process as $n \to \infty$, as measured in terms of the total variation distance. Let $p_t(\cdot)$ be the distribution of $\pi_t$ on $\mathcal{S}_n$, and let $\mu$ be the invariant distribution of the chain. Let

$$d(t) = \|p_t(\cdot) - \mu\| = \frac{1}{2} \sum_{\sigma \in \mathcal{S}_n} |p_t(\sigma) - \mu(\sigma)|.$$

$d(t)$ is the total variation distance between the state of the chain at time $t$ and its limiting distribution $\mu$. (Below, we will also use the notation $\|X - Y\|$ where $X$ and $Y$ are collections of random variables with laws $p_X, p_Y$ to mean $\|p_X - p_Y\|$.)

The main goal of this paper is to prove that the chain exhibits a sharp cutoff, in the sense that $d(t)$ drops abruptly from its maximal value 1 to its minimal value

0 around a certain time $t_{\text{mix}}$, called the mixing time of the chain. (See [7] or [11] for a general introduction to mixing times). Note that if $\Gamma$ is a fixed conjugacy class of $\mathcal{S}_n$, and $m > n$, $\Gamma$ can also be considered a conjugacy class of $\mathcal{S}_m$ by simply adding $m - n$ fixed points to any permutation $\sigma \in \Gamma$. With this in mind, our theorem states the following:

**Theorem 1.** *Let $k \geq 2$ be an integer, and let $\Gamma_k$ be the conjugacy class of $\mathcal{S}_n$ corresponding to $k$-cycles. The continuous time random walk $(\pi_t, t \geq 0)$ associated with $(\mathcal{S}_n, \Gamma_k)$ has a cutoff at time $t_{\text{mix}} := (1/k) n \log n$, in the sense that for any $\varepsilon > 0$ , there exist $N_{\varepsilon,k}, C_{\varepsilon,k} > 0$ large enough so that for all $n \geq N_{\varepsilon,k}$,*

$$d\left(t_{\text{mix}} - C_{\varepsilon,k} n\right) > 1 - \varepsilon \tag{2}$$

$$d\left(t_{\text{mix}} + C_{\varepsilon,k} n)\right) < \varepsilon. \tag{3}$$

As explained in Section 1.2 below, this result solves a well-known conjecture formulated by several people over the course of the years.

**Remark 2.** *Theorem 1 can be extended, without a significant change in the proofs, to cover the case of general fixed conjugacy classes $\Gamma$, with $k = |\Gamma| > 2$ independent of $n$. In order to alleviate notation, we present here only the proof for $k$-cycles. A more delicate question, that we do not investigate, is what growth of $k = k(n)$ is allowed so that Theorem 1 would still be true in the form*

$$d\left(t_{\text{mix}}(1 - \delta)\right) > 1 - \varepsilon \tag{4}$$

$$d\left(t_{\text{mix}}(1 + \delta)\right) < \varepsilon. \tag{5}$$

*The lower bound in* (4) *is easy. For the upper bound in* (5)*, due to the birthday problem, the case $k = o(\sqrt{n})$ should be fairly similar to the arguments we develop below, with adaptations at several places, e.g. at the argument following* (31)*; we have not checked the details. Things are likely to become more delicate when $k$ is of order $\sqrt{n}$ or larger. Yet, we conjecture that* (5) *holds as long as $k = o(n)$.*

## 1.2 Background

This problem has a rather long history, which we now sketch. Mixing times of Markov chain were studied independently by Aldous [1] and by Diaconis and Shahshahani [8] at around the same time, in the early 80's. The paper [8], in particular, establishes the existence of what became known as the *cutoff phenomenon* for the composition of random transpositions. Random transpositions is perhaps the simplest example of a random walk on $\mathcal{S}_n$ and is a particular case of the walks covered in this paper, arising when the conjugacy class $\Gamma$ contains exactly all transpositions. The authors of [8] obtained a version of Theorem 1 for this particular case (with explicit choices of $C_{2,\varepsilon}$ for a given $\varepsilon$). As is the case here, the hard part of the result is the upper-bound (3). Remarkably, their solution involved a connection with the representation theory of $\mathcal{S}_n$, and uses rather delicate estimates on so-called character ratios.

Soon afterwards, a flurry of paper tried to generalize the results of [8] in the direction we are tackling in this paper, that is, when the step distribution is uniform over a fixed conjugacy class $\Gamma$. However, the estimates on character ratios that are needed become harder and harder as $|\Gamma|$ increases. Flatto, Odlyzko and Wales [9], building on earlier work of Vershik and Kerov [22], obtained finer estimates on character ratios and were able to show that mixing must occur before $(1/2)n \log n$ for $|\Gamma|$ fixed - thus giving another proof of the Diaconis-Shahshahani result when $|\Gamma| = 2$. (Although this doesn't appear explicitly in the paper [9], this is recounted in the book of Diaconis [7, p.44].) Improving further the estimates on character ratios, Roichman [16, 17] was able to prove a weak version of Theorem 1, where it is shown that $d(t)$ is small if $t > C\, t_{\text{mix}}$ for some large enough $C > 0$. In his result, $|\Gamma|$ is allowed to grow to infinity as fast as $(1-\delta)n$ for any $\delta > 0$. To our knowledge, it is in the paper [16] that Theorem 1 first formally appears as a conjecture, although no doubt that it had been privately made before. (The lower-bound for random transpositions, which is based on counting the number of fixed points in $\pi_t$, works equally well in this context and provides the conjectured correct answer in all cases.) Lulov [12] dedicated his Ph.D. thesis to the problem, and Lulov and Pak [13] obtained a partial proof of the conjecture of Roichman, in the case where $|\Gamma|$ is very large, i.e., greater than $n/2$. More recently, Roussel [18] and [19] made some progress in the small $|\Gamma|$ case, working out the character ratios estimates to treat the case where $|\Gamma| \leq 6$. Saloff-Coste, in his survey article [15, Section 9.3], discusses the sort of difficulties that arise in these computations, and states the conjecture again. A summary of the results discussed above is also given. See also [20, Page 381], where work in progress of Schlage-Puchta that overlaps the result in Theorem 1 is mentioned.

## 1.3 Structure of the proof

To prove Theorem 1, it suffices to look at the cycle structure of $\pi_t$ and check that if $N_t(i)$ is the number of cycles of $\pi_t$ of size $i$ for every $i \geq 1$, and if $t \geq t_{\text{mix}} + C_{k,\varepsilon} n$ then the total variation distance between $(N_t(i))_{1\leq i\leq n}$ and $(N(i))_{1\leq i\leq n}$ is close to 0, where $(N(i))_{1\leq i\leq n}$ is the cycle distribution of a random permutation sampled from $\mu$. We thus study the dynamics of the cycle distribution of $\pi_t$, which we view as a certain coagulation-fragmentation chain. Using ideas from Schramm [21], it can be shown large cycles are at equilibrium much before $t_{\text{mix}}$, i.e. at a time of order $O(n)$. Very informally speaking, the idea of the proof is the following. We focus for a moment on the case $k = 2$ of random transpositions, which is the easiest to explain. The process $(\pi_t, t \geq 0)$ may be compared to an Erdős-Reny random graph process $(G_t, t \geq 0)$ where random edges are added to the graph at rate 1, in such a way that the cycles of the permutation are subsets of the connected components of $G_t$. Schramm's result from [21] then says that, if $t = cn$ with $c > 1/2$ (so that $G_t$ has a giant component), then the macroscopic cycles within the giant component have relaxed to equilibrium. By an old result of Erdős and Renyi, it takes time $t = t_{\text{mix}} + C_{k,\varepsilon} n$ for $G_t$ to be connected with probability greater than $1 - \varepsilon$. By this point the giant component encompasses every vertex and thus, extrapolating Schramm's result to this time, the macroscopic cycles of $\pi_t$ have the correct distribution at this point. A separate and somewhat more technical argument is needed to deal with small cycles.

More formally, the proof of Theorem 1 thus proceeds in two main steps. In the first step, presented in Section 2 and culminating in Proposition 19, we show that after time $t_{\text{mix}} + c_{\varepsilon,k} n$, the distribution of *small cycles* is close (in variation distance) to the invariant measure, where a *small cycle* means that it is smaller than a suitably chosen threshold approximately equal to $n^{7/8}$. This is achieved by combining a queueing-system argument (whereby initial discrepancies are cleared by time slightly larger than $t_{\text{mix}}$ and equilibrium is achieved) with a-priori rough estimates on the decay of mass in small cycles (Section 2.1). In the second step, contained in Section 3, a variant of Schramm's coupling from [21] is presented, which allows to couple the chain after time $t_{\text{mix}} + c_{\varepsilon,n}$ to a chain started from equilibrium, within time of order $n^{5/8} \ln n$, if all small cycles agree initially.

# 2 Small cycles

In this section we prove the following proposition. Let $(N_i(t))_{1 \leq i \leq n}$ be the number of cycles of size $i$ of the permutation $\pi_t$, where $(\pi_t, t \geq 0)$ evolves according to random $k$-cycles (where $k \geq 2$), but does not necessarily start at the identity permutation. Let $(Z_i)_{i=1}^n$ denote independent Poisson random variable with mean $1/i$.

Fix $0 < \chi < 1$ and let $K = K(n)$ be the closest dyadic integer to $\lfloor n^\chi \rfloor$. We think of cycles smaller than $K$ as being small, and big otherwise. Let $I_j = \{i \in \mathbb{Z} : i \in [2^j, 2^{j+1})\}$, $L_j = |I_j| = 2^j$ and

$$M_j(t) = \sum_{i \in I_j} N_i(t)\,. \tag{6}$$

Introduce the stopping time

$$\tau = \inf\left\{t \geq 0 : \exists\, 0 \leq j \leq \log_2 K + k, M_j(t) > (\log n)^2/2\right\}. \tag{7}$$

Therefore, prior to $\tau$, the total number of small cycles in each dyadic strip $[2^j, 2^{j+1})$ ($j \leq k + \log_2 K$) never exceeds $(\log n)^2/2$.

**Proposition 3.** *Suppose that*

$$\mathbb{P}(\tau < n \log n) \longrightarrow 0 \tag{8}$$

*as $n \to \infty$, and that initially,*

$$M_j(0) \leq D \log(j+2) \tag{9}$$

*for all $0 \leq j \leq \log_2 \log n$, for some $D > 0$ independent of $j$ or $n$. Then for any sequence $t = t(n)$ such that $t(n)/n \to \infty$ as $n \to \infty$ and $t(n) \leq n \log n$,*

$$\left\|(N_i(t))_{i=1}^K - (Z_i)_{i=1}^K\right\| \longrightarrow 0.$$

In particular, under the assumptions of Proposition 3, for any $\varepsilon > 0$ there is a $c_{\varepsilon,k} > 0$ such that for all $n$ large,

$$\left\|(N_i(c_{\varepsilon,k} n))_{i=1}^K - (Z_i)_{i=1}^K\right\| < \varepsilon.$$

*Proof.* The proof of this proposition relies on the analysis of the dynamics of the small cycles, where each step of the dynamics corresponds to an application of a $k$-cycle, by viewing it as a coagulation-fragmentation process. To start with, note that every $k$-cycle may decomposed as a product of $k-1$ transpositions

$$c = (x_k, \ldots, x_1) = (x_k, x_{k-1}) \ldots (x_2, x_1).$$

Thus the application of a $k$-cycle may be decomposed into the application of $k-1$ transpositions: namely, applying $c$ is the same as first applying the transposition $(x_1, x_2)$ followed by $(x_2, x_3)$, and so on until $(x_{k-1}, x_k)$. Whenever one of those transpositions is applied, say $(a, b)$, this can yield either a fragmentation or a coagulation, depending on whether $a$ and $b$ are in the same cycle or not at this time. If they are, say if $b = \sigma^i(a)$ (where $i \geq 1$ and $\sigma$ denotes the permutation at this time), then the cycle $C$ containing $a$ and $b$ splits into $(a, \ldots, \sigma^{i-1}(a))$ and everything else, that is, $(b, \ldots, \sigma^{|C|-i}(b))$. If they are in different cycles $C$ and $C'$ then the two cycles merge.

To track the evolution of cycles, we color the cycles with different colors (blue, red or black) according (roughly) to the following rules. The blue cycles will be the large ones, and the small ones consist of red and black. Essentially, red cycles are those which undergo a "normal" evolution, while the black ones are those which have experienced some kind of error. By "normal evolution", we mean the following: in a given step, one (and only one) small cycle is generated by fragmentation of a blue cycle. It is the only small cycle that is involved in this step. In a later step of the random walk, this cycle coagulates with a large cycle and thus becomes large again. If at any point of this story, something unexpected happens (e.g., this cycle gets fragmented instead of coagulating with a large cycle, or coagulates with another small cycle) we will color it black. In addition, we introduce ghost cycles to compensate for this sort of errors.

We now describe this procedure more precisely. We start by coloring every cycle of the permutation $\sigma(t)$ which is larger than $K$ blue. We denote by $\theta(t)$ the fraction of mass contained in blue cycles, i.e.,

$$\theta(t) = \frac{1}{n} \sum_{i=K+1}^{n} iN_i(t). \tag{10}$$

Note that by definition of $\tau$,

$$1 - \frac{K}{n}(\log n)^2 \leq \theta(t) \leq 1 \tag{11}$$

for all $t \leq \tau$.

We now color the particles which are smaller than $K$ either red or black according to the following dynamics. Suppose we are applying a certain $k$-cycle $c = (x_k, \ldots, x_1)$, which we write as a product of $k-1$ transpositions:

$$c = (x_k, \ldots, x_1) = (x_k, x_{k-1}) \ldots (x_2, x_1) \tag{12}$$

(note that we require that $x_i \neq x_j$ for $i \neq j$).

*Red cycles.* Assume that a blue cycle is fragmented and one of the pieces is small, and that this transposition is the first one in the application of the $k$-cycle $(x_1, \ldots, x_k)$ to generate a small cycle by fragmentation of a blue cycle. In that case (and only in that case), we color it red. Red cycles may depart through coagulation or fragmentation. A coagulation with a blue cycle, if it is the first in the step and no small cycles were created in this step prior to it, will be called *lawful*. Any other departure will be called *unlawful*. If a blue cycle breaks up in a way that would create a red cycle and both cycles created are small (which may happen if the size of the cycle is between $K$ and $2K$), then we color only one red, and the other black, in an arbitrary way, say randomly. Note that only blue cycles whose size at the beginning of the step is comprised between $K$ and $2^kK$

may generate black cycles in this fashion. The resulting black cycles are then of size greater than $K2^{-k}$.

*Black cycles.* Black cycles are created in one of two ways. First, any red cycle that departs in an unlawful fashion and stays small becomes black. Further, if the transposition $(a, b)$ is not the first transposition in this step to create a small cycle from a blue cycle, or if it is but a previous transposition in the step involved a small cycle, then the small cycle(s) created is colored black. Now, assume that $(a, b)$ involves only cycles which are smaller than $K$: this may be a fragmentation producing two new cycles, or a merging of two cycles producing one new cycle. In this case, we color the new cycle(s) black, no matter what the initial color of the cycles, except if this operation is a coagulation *and* the size of this new cycle exceeds $K$, in which case it is colored blue again. Thus, black cycles are created through either coagulations of small parts or fragmentation of either small or large parts, but black cycles disappear only through coagulation.

We aim to analyze the dynamics of the red and black system, and the idea is that the dynamics of this system is essentially dominated by that of the red cycles, where the occurrence of black cycles is an error that we aim to control.

*Ghosts.* Let $R_i(t), B_i(t)$ be the number of red and black cycles respectively of size $i$ at time $t$. It will be helpful to introduce another type of cycles, called ghost cycles, which are nonexisting cycles which we add for counting purposes: the point is that we do not want to touch more than one red cycle in any given step. Thus, for any red cycle departing in an unlawful way, we compensate it by creating a ghost cycle of the same size. For instance, suppose two red cycles $C_1$ and $C_2$ coagulate (this could form a blue or a black cycle). Then we leave in the place of $C_1$ and $C_2$ two ghost cycles $C_1'$ and $C_2'$ of sizes identical to $C_1$ and $C_2$.

An exception to this rule is that if, during a step, a transposition creates a small red cycle by fragmentation of a blue cycle, and later within the same step this red cycle either is immediately fragmented again in the next transposition or coagulates with another red or black cycle and remains small, then it becomes black as above but we do not leave a ghost in its place.

Finally, we also declare that every ghost cycle of size $i$ is killed independently of anything else at an instantaneous rate which is precisely given by $i\mu(t)$, where $\mu(t)$ is a random nonnegative number (depending on the state of the system at time $t$) which will be defined below in (17) and corresponds to the rate of lawful departures of red cycles.

Let $G_i(t)$ denote the number of ghost cycles of size $i$ at time $t$, and let $Y_i = R_i + G_i$, which counts the number of red and ghost cycles of size $i$. Our goal is twofold. On the one hand, we want to show that $(Y_i(t))_{i=1}^K$ is close in total variation distance to $(Z_i)_{i=1}^K$, and secondly, that at time $t = t(n)$ the probability that there is any black cycle or a ghost cycle converges to 0 as $n \to \infty$.

**Remark 4.** *Note that with our definitions, at each step at most one red cycle can be created, and at most one red cycle can disappear without being compensated by the creation of a ghost. Furthermore these two events cannot occur in the same step.*

**Lemma 5.** *Assume (8) as well as (9), and let $t = t(n)$ be as in Proposition 3. Then*

$$\left\| (Y_i(t))_{i=1}^K - (Z_i)_{i=1}^K \right\| \longrightarrow 0.$$

*Proof.* The idea is to observe that $Y_i$ has approximately the following dynamics:

$$\begin{cases} \text{rate: } (x \to x+1) & = \lambda \text{ if } x \geq 0, \\ \text{rate: } (x \to x-1) & = ix\mu \text{ if } x \geq 1. \end{cases}$$

and that $\lambda = \mu = k/n + o(1/n)$, so that $(Y_i)$ is approximately a system of $M/M/\infty$ queues where the arrival rate is $k/n$ and the departure rate of every customer is $ik/n$. The equilibrium distribution of $(Y_i)$ is thus approximately Poisson with parameter the ratio of the two rates, i.e. $1/i$. The number of initial customers in the queues is by the assumption (8) small enough so that by time $t(n)$ they are all gone and thus the queue has reached equilibrium.

We now make this heuristics precise. To increase $Y_i$ by 1, i.e. create a red cycle, one needs to specify the $j$-th transposition, $1 \leq j \leq k-1$, of the $k$-cycle at which it is created. The first point $x_1$ of the $k$-cycle must fall wherever in a blue cycle (which has probability $\theta$). Say that $x_1 \in C_1$, with $C_1$ a blue cycle. In order to create a cycle of size exactly $i$ at this transposition, the second point $x_2$ must fall at either of exactly *two* places within $C_1$: either $\sigma^i(x_1)$ or $\sigma^{-i}(x_1)$. However, note that if $x_2 = \sigma^{-i}(x_1)$ and $|c| = k \geq 3$, then the next transposition is guaranteed to involve the newly formed cycle, either to reabsorb it in the blue cycles, or to turn into a black cycle through coalescence with another small cycle or fragmentation. Either way, this newly formed cycle does not eventually lead to an increase in $Y_i$ since by our conventions, we do not leave a ghost in its place. On the other hand, if $x_2 = \sigma^i(x_1)$ then the newly formed red cycle will stay on as a red or a ghost cycle in the next transpositions of the application of the cycle $c$. Whether it stays as a ghost or a red cycle does not change the value of $Y_i$, and therefore, this event (which has probability $\theta'/n$ with $\theta' = (n\theta - 1)/(n-1)$, because repetitions of indices in a cycle are not allowed) leads to a net increase of $Y_i$ by 1. This is true for all of the first $k-2$ transpositions of the $k$-cycle $c$ (with the appropriate change of the factor $n_1$), but not for the last one, where both $x_k = \sigma^i(x_{k-1})$ and $x_k = \sigma^{-i}(x_{k-1})$ will create a red cycle of size $i$. It follows from this analysis that the total rate $\lambda(t)$ at which $Y_i$ increases by 1 satisfies

$$\lambda(t) \le \lambda^+ = \frac{k-2}{n-k+1} + \frac{2}{n-k+1} = \frac{k}{n-k+1}. \tag{13}$$

To get a lower-bound, observe that for $t \le \tau$, $\theta(t) \ge 1 - K(\log n)^2/n$ at the beginning of the step. When a $k$-cycle is applied and we decompose it into $k-1$ elementary transpositions, the value $\theta(t)$ for each of the transpositions may take different successive values which we denote by $\theta(t,j), j = 1, \ldots, k-1$. However, note that at each such transposition, $\theta$ can only change by at most $\pm 2K/n$. Thus it is also the case that for all $1 \le j \le k-1$, $\theta(t,j) \ge 1 - 2(k-1)K(\log n)^2/n$. Therefore, the probability that a fragmentation of a blue cycle does not create any small cycle is also bounded below by

$$1 - 2(k-1)K(\log n)^2/n - 2K(\log n)^2/n = 1 - 2kK(\log n)^2/n =: \theta_-(t).$$

It thus follows that the total rate $\lambda(t)$ is bounded below by

$$\lambda(t) \ge \theta_-^{k-1}\left(\frac{2}{n} + \frac{k-2}{n}\right) \ge \frac{k}{n}\left(1 - 8k\frac{K(\log n)^2}{n}\right) =: \lambda^- \tag{14}$$

Of course, by this we mean that the $Y_i(t)$ are nonnegative jump processes whose jumps are of size $\pm 1$, and that if $\mathcal{F}_t$ is the filtration generated by the entire process up to time $t$, then

$$\lim_{h \to 0^+} \frac{\mathbb{P}(Y_i(t+h) = x+1 | \mathcal{F}_t, Y_i(t) = x)}{h} = \lambda(t); \text{ and } \lambda^- \le \lambda(t) \le \lambda^+ \tag{15}$$

almost surely on the event $\{t \le \tau\}$. As for negative jumps, we have that for $x \ge 1$,

$$\lim_{h \to 0^+} \frac{\mathbb{P}(Y_i(t+h) = x-1 | \mathcal{F}_t, Y_i(t) = x)}{h} = ix\mu(t), \tag{16}$$

where $\mu(t)$ depends on the partition and satisfies the estimates

$$\mu^- \le \mu(t) \le \mu^+ \tag{17}$$

where

$$\mu^- := \frac{k}{n}\left(1 - 8k\frac{K(\log n)^2}{n}\right) \text{ and } \mu^+ = \frac{k}{n-k}. \tag{18}$$

The reason for this is as follows. To decrease $Y_i$ by 1 by decreasing $R_i$, note that the only way to get rid of a red cycle without creating a ghost is to coagulate it with a blue cycle at the $j$-th transposition, $1 \le j \le k-1$, with no other transpositions creating small cycles. The probability of this event is bounded above by $ik/(n-k)$ and, with $\theta_-$ as above, bounded below by

$$\frac{i\theta}{n}\theta_-^{k-2} + \theta\frac{i}{n-1}\theta_-^{k-2} + \theta\theta_-\frac{i}{n-2}\theta_-^{k-3} + \cdots + \theta\theta_-^{k-2}\frac{i}{n-k+1} \ge \frac{ik}{n}\theta_-^{k-1}.$$

Therefore, if in addition ghosts are each killed independently with rate $\mu(t)$ as above, then (16) holds. More generally, if $1 \le m \le K$ and $i_1 < \ldots < i_m \le K$ are pairwise distinct integers, then we may consider the vector $(Y_{i_1}(t), \ldots, Y_{i_m}(t))$. If its current state is $x = (x_1, \ldots, x_m)$ then it may make transitions to $x' = (x_1', \ldots, x_m')$ where the two vectors $x$ and $x'$ differ by exactly one coordinate (say the $j^{\text{th}}$ one) and $x_j - x_j' = \pm 1$ (since only one queue $Y_i$ can change at any time step, thanks to our coloring rules). Also, writing $Y(t)$ for the vector $(Y_{i_1}(t), \ldots, Y_{i_m}(t))$, we find:

$$\lim_{h \to 0^+} \frac{\mathbb{P}(Y(t+h) = x' | \mathcal{F}_t, Y = x)}{h} = \begin{cases} \lambda(t) & \text{if } x_j' = x_j + 1 \\ i_j x_j \mu(t) & \text{if } x_j' = x_j - 1. \end{cases}$$

These observations show that we can compare $\{(Y_i(t \wedge \tau)_{1 \le i \le K}, t \ge 0\}$ to a system of independent Markov queues $\{(Y_i^+(t \wedge \tau))_{1 \le i \le K}, t \ge 0\}$ with respect to a common filtration $\mathcal{F}_t$, with no simultaneous jumps almost surely, and such that the arrival rate of each $Y_i$ is $\lambda^+$ and the departure rate of each client in $Y_i$ is $i\mu^-$. We may also define a system of queues $(Y_i^-)_{1 \le i \le K}$ by accepting every new client of $Y_i^+$ with probability $\lambda^-/\lambda^+$ and rejecting it otherwise. Subsequently, each accepted client tries to depart at a rate $\mu^+ - \mu^-$, or when it departs in $Y_i^+$, whichever comes first. Then one can construct all three processes $(Y_i^-)_{1 \le i \le K}$, $(Y_i)_{1 \le i \le K}$ and $(Y_i^+)_{1 \le i \le K}$ on a common probability space in such a way that $Y_i^-(t) \le Y_i(t) \le Y_i^+(t)$ for all $t \le \tau$.

Note that if $(Z_i^+)_{1 \le i \le K}$ denote independent Poisson random variables with mean $\lambda^+/(i\mu^-)$, then $(Z_i^+)_{1 \le i \le K}$ forms an invariant distribution for the system $(Y_i^+(t), t \ge 0)_{1 \le i \le K}$. Let $(Z_i^+(t), t \ge 0)_{1 \le i \le K}$ denote the system of Markov queues $Y_i^+$ started from its equilibrium distribution $(Z_i^+)_{1 \le i \le K}$. Then $(Y_i^+(t))_{1 \le i \le K}$ and $(Z_i^+(t))_{1 \le i \le K}$ can be coupled as usual by taking each coordinate to be equal after the first time that they coincide. In particular, once all the initial customers of $Y_i^+$ and of $Z_i^+(t)$ have departed (let us call $\tau'$ this time), then the two processes $(Y_i^+)_{1 \le i \le K}$ and $(Z_i^+)_{1 \le i \le K}$ are identical.

We now check that this happens before $t = t(n)$ with high probability. It is an easy exercise to check this for $Z_i^+(t)$ so we focus on $Y_i^+(t)$. To see this, note that by (9), there are no more than $D\log(j+2)$ customers in every strip $[2^j, 2^{j+1})$ initially if $j \le \log_2 \log n$. Moreover, each customer departs with rate at least

$2^{j-1}/n$ when in this strip. Thus the time $\tau_j'$ it takes for all initial customers of $Y^+$ in strip $[2^j, 2^{j+1})$ to depart is dominated by $(n/2^{j-1})\max_{1\le q\le D\log(j+2)} E_q$, where $(E_q)_{q\ge 1}$ is a collection of i.i.d. standard exponential random variables. Hence

$$\mathbb{E}(\tau_j') \le \frac{n}{2^{j-2}}(\log_2 D + \log_2\log(j+4)).$$

For larger strips we use the crude and obvious bound $M_j(0)\le n$ if $j \ge \log_2\log n$. Moreover, each customer departs at rate $n/2^{j-1}$ with $j = \lfloor \log_2 \log n\rfloor$. Thus, in distribution,

$$\tau_j \preceq \frac{n}{2^{j-2}} \max_{1\le q\le n} E_q$$

so that $\mathbb{E}(\tau_j) \le n\log n/2^{j-1}$. Since we obviously have $\tau' \le \sum_{j=0}^{\log K+k}\tau_j'$, we conclude

$$\mathbb{E}(\tau') \le \sum_{j=0}^{\log_2\log n} \frac{n}{2^{j-2}}(\log D + \log_2\log(j+4)) + \sum_{j\ge \log_2\log n}\frac{n\log n}{2^{j-1}} \le a(D)n,$$

where $a(D)<\infty$ depends solely on $D$. By Markov's inequality and since $t(n)/n\to\infty$, we conclude that $\tau'\le t$ with high probability. We now claim that $(Y_i^-(t))_{1\le i\le K} = (Y_i^+(t))_{1\le i\le K}$ with high probability. To see this, we note that at equilibrium $\mathbb{E}(Z_i^+) = \lambda^+/(i\mu^-)\le 2/i$. Therefore,

$$\begin{aligned}\mathbb{P}(Y_i^+(t)\ne Y_i^-(t)\text{ for some } 1\le i\le K) &\le \mathbb{E}(\sum_{i=1}^K Y_i^+(t) - Y_i^-(t); \tau' < t) + \mathbb{P}(\tau'>t)\\ &\le \sum_{i=1}^K \frac{2}{i}\{(1-\frac{\lambda^-}{\lambda^+}) + (1-\frac{\mu^-}{\mu^+})\} + \mathbb{P}(\tau'>t)\\ &\le 16(k-1)\frac{K(\log n)^3}{n} + \mathbb{P}(\tau'>t).\end{aligned}$$

Since we have already checked that $\mathbb{P}(\tau'>t)\to 0$ as $n\to\infty$, this shows that on the event $\{\tau'\le t\le\tau\}$ and $\{Y_i^+(t) = Y_i^-(t)\text{ for all } 1\le i\le K\}$, (an event of probability asymptotically one), then $(Y_i(t))_{1\le i\le K}$ can be coupled to $(Z_i^+(t))_{1\le i\le K}$ which has the same law as $(Z_i^+)_{1\le i\le K}$. Thus

$$\left\|(Y_i)_{i=1}^K - (Z_i^+)_{i=1}^K\right\| \longrightarrow 0 \tag{19}$$

as $n\to\infty$. On the other hand, we claim that

$$\left\|(Z_i)_{i=1}^K - (Z_i^+)_{i=1}^K\right\| \longrightarrow 0$$

also. Indeed, it is easy to see and well known that for $\alpha, \beta > 0$

$$\|\mathrm{Po}(\alpha) - \mathrm{Po}(\beta)\| \leq 1 - \exp(-|\alpha - \beta|) \leq |\alpha - \beta|.$$

Since the coordinates of $Z_i$ and $Z_i^+$ are both independent Poisson random variables but with different parameters, we find that

$$\begin{aligned}
\left\|(Z_i)_{i=1}^K - (Z_i^+)_{i=1}^K\right\| &\leq \sum_{i=1}^K \frac{\lambda^+}{i\mu^-} - \frac{1}{i} \\
&\leq \sum_{i=1}^K \frac{1}{i}\left(\frac{1}{1 - 2(k-1)K(\log n)^2/n} - 1\right) \\
&\leq \frac{4(k-1)K(\log n)^3}{n} \longrightarrow 0
\end{aligned}$$

as $N \to \infty$. By the triangle inequality and (19), this concludes the proof of Lemma 5. □

**Lemma 6.** *Let $t = t(n)$ be as in Proposition 3. Then, with probability tending to 1 as $n \to \infty$, $B_i(t) = 0$ for all $1 \leq i < K2^{-k}$.*

*Proof.* Let us consider black cycles in scale $j$, that is, whose size $i$ satisfies $2^j \leq i < 2^{j+1}$ with $j \leq \log_2 K - k$. By assumption (8), before time $t$ the total mass of small cycles never exceeds $2K \log n$ with high probability. Thus the rate at which a black cycle in scale $j$ is generated by fragmentation of a red cycle (or from another black cycle) is at most

$$\lambda_j^{B,1} = (k-1)\frac{2K \log n}{n}\frac{2^{j+1}}{n}.$$

Black cycles can also be generated directly by fragmenting a blue cycle and immediately afterwards fragmenting it again in one of the next transpositions in the step. The rate at which this occurs is smaller than

$$\lambda_j^{B,2} = (k-1)\frac{K}{n}\frac{2^{j+1}}{n}.$$

(Since $i \leq K2^{-k}$, this is the only way black cycles can arise in this range. The case of a fragmentation of a blue cycle whose size is in the interval $[K, 2^k K)$ in such a way that several parts are smaller than $K$ is treated in the following Lemma 7.)

This combined rate is therefore smaller than $\lambda_j^B = 2\lambda_j^{B,1}$. Note however that it may be the case that several black cycles are produced in one step, however this number may not exceed $2k$. On the other hand, every black cycle departs at a rate which is at least

$$\mu_j^B = \frac{\theta}{n} 2^j \geq \frac{2^{j-1}}{n}$$

since $\theta \geq 1/2$ for $t \leq \tau$, say. (Note that when two back cycles coalesce, the new black cycle has an even greater departure rate than either piece before the coalescence, so ignoring these events can only increase stochastically the total number of black cycles). Thus we see that the number of black cycles in this scale is dominated by a Markov chain $(\beta_j(s), s \geq 0)$ where the rate of jumps from $x$ to $x + 2k$ is $\lambda_j^B$ and the rate of jumps from $x$ to $x - 1$ is $\mu_j^B$. Speeding up time by $n/2^{j-1}$, $\beta_j$ becomes a Markov chain $\beta_j'$ whose rates are respectively $\lambda_j'^B = 4(k-1)K \log n/n$ and 1. We are interested in

$$\mathbb{P}(\beta_j(t) > 0) = \mathbb{P}(\beta_j'(t') > 0); \text{ where } t' = t2^{j-1}/n.$$

Note that when there is a jump of size $2k$ (i.e., when $2k$ individuals are born) the time it takes for them to all die in this new time-scale is a random variable $E$ which has the same distribution as $E = \max_{1 \leq j \leq 2k} E_j$ where $(E_j)_{1 \leq j}$ are i.i.d. standard exponential random variables. Decomposing on possible birth times of individuals, and noting that $\mathbb{P}(E > x) \leq 2ke^{-x}$ by a simple union bound, we see that

$$\begin{aligned}
\mathbb{P}(\beta_j'(t') > 0) &= \int_0^{t'} \lambda_j'^B \mathbb{P}(E > t' - s) ds \\
&\leq \frac{4(k-1)K \log n}{n} \int_0^\infty \mathbb{P}(E > x) dx \leq \frac{8k^2 K \log n}{n}.
\end{aligned}$$

There are $\log_2 K$ possible scales to sum on, so by a union bound the probability that there is any black cycle at time $t$ is smaller or equal to $k^2(\log n)^3/n \to_{n\to\infty} 0$. $\square$

**Lemma 7.** *Let $t = t(n)$ be as in Proposition 3. Then, with probability tending to 1 as $n \to \infty$, $B_i(t) = 0$ for all $K2^{-k} \leq i < K$.*

*Proof.* The creation of black cycles by unlawful departures of red cycles is handled as in Lemma 6. Thus, we need only show that with high probability, there is no black cycle in this range at time $t$ which was generated by the fragmentation of a blue cycle in the range $[K, 2^k K)$. By assumption, there are no more than $\log n$ such cycles initially, and they are all touched by a time stochastically dominated by $\tau'$. Even if they all fragment, they each generate no more than $k - 1$ black cycles. Thus by a time stochastically dominated by $\tau_1 + \ldots + \tau_k$ they are all departed, where the $\tau_i$ are i.i.d. with same distribution as $\tau'$. Since $\mathbb{E}(\tau') = o(t)$, we conclude by Markov's inequality that none of them may remain at time $t$ with high probability. We now use analogous (but rougher) estimates to Lemma 5 to show that the number of cycles in the strip $[K, 2^k K)$ is basically of order 1: the rate at which a cycle in this strip is generated by fragmentation is smaller than $k2^{k+1}K/n$, and by coagulation this rate is smaller than $k^2(2K\log n/n)^2$ prior to time $\tau$ since the total mass of small cycle never exceeds $2K\log n$. The departure rates exceeds $(K/n)\times(1/2)$ since the total mass of cycles greater than $2^kK$ must be greater than $n/2$ prior to time $\tau$. Thus, discarding the initial cycles of size $i \in [K, 2^kK)$ the number of such cycles $N_K$ is dominated by a birth and death chain of birth rate $2^{k+2}kK/n$ and death rate $K/(2n)$. Speeding up time by $K/n$, this is a birth death chain of birth death rate $2^{k+2}k$ and $(1/2)$ respectively. The equilibrium of this chain is Poisson with mean $2^{k+3}k < \infty$, and thus we conclude $\mathbb{E}(N_K(s)) \le 2^{k+3}k$ for all $s > 0$. Now, every cycle of this size may be in fact fragmented with probability at most $kK/n$ the next time it is touched. Thus, in any time-interval $[s_1, s_2] \subset [0, t]$, the expected number of black cycles that are generated from such fragmentations is at most $2^{k+3}k^3K^2(s_2-s_1)/n^2$. Now, if such a black cycle is generated at time $s \le t$, it coagulates with blue cycles up to time $t$ with rate at least $K/(4n)$ when $t \le \tau$, so the probability it survives up to time $t$ is at most $\exp(-(t-s)K/(4n))$. It follows that the expected number of such black cycles at time $t$ is at most

$$\frac{2^{k+3}k^3K^2}{n^2}\int_0^t du \exp(-\frac{K}{4n}u) \le \frac{2^{k+5}k^3K}{n} \longrightarrow 0$$

as $N \to \infty$. □

The case of ghost particles is treated as follows.

**Lemma 8.** *Let $t = t(n)$ be as in Proposition 3. Then, with probability tending to 1 as $n \to \infty$, $G_i(t) = 0$ for all $1 \le i \le K$.*

*Proof.* Suppose a red cycle is created, and consider what happens to it the next time it is touched. With probability at least $\theta^{k-2}$ this will be to coagulate with a blue cycle with no other small cycle being touched in that step, in which case this cycle is not transformed into a ghost. However, in other cases it might become a ghost. It follows that any given cycle in $Y_i$ is in fact a ghost with probability at most

$$\frac{1-\theta^{k-2}}{\theta^{k-2}} \le (k-2)\frac{K(\log n)^2}{n}.$$

It follows that (using the notations from Lemma 5):

$$\begin{aligned}
\mathbb{P}(G_i(t) > 0 \text{ for some } i) &\le \sum_{i=1}^{K} \mathbb{E}(G_i(t); \tau' < t) + \mathbb{P}(\tau' > t) \\
&\le \mathbb{P}(\tau' > t) + \sum_{i=1}^{K} \frac{2}{i} \frac{(k-2)K(\log n)^2}{n} \\
&\le \mathbb{P}(\tau' > t) + 2(k-2)\frac{K(\log n)^3}{n}
\end{aligned}$$

which tends to 0 as $N \to \infty$. This completes the proof of Lemma 8. □

**Completion of the proof of Proposition 3:** Since $N_i(t) = Y_i - G_i + B_i$, we get the proposition by combining Lemmas 5, 6, 7 and 8. □

## 2.1 Verification of (8) and (9)

In order for Proposition 3 to be useful, we need to show that the assumptions (8) and (9) actually hold with large enough probability. This will be accomplished in Proposition 12 and 17 below.

Recall the variable $M_j$, see (6), and let

$$\mathcal{A}_j^s = \left\{ \max_{t \in [sn \log\log n, n \log n]} M_j(t) < n 2^{-j} / (\log n)^3 \right\}.$$

Recall that $K$ is the dyadic integer closest to $[n^\chi]$.

We begin with the following lemma. Its proof is a warm-up to the subsequent analysis.

**Lemma 9.** *Let*

$$\mathcal{A}_\chi = \cap_{j=0}^{\log_2 K + k} \mathcal{A}_j^6 .$$

*Then,*

$$\mathbb{P}(\mathcal{A}_\chi^{\complement}) \underset{n\to\infty}{\longrightarrow} 0 .$$

*Proof.* It is convenient to reformulate the cycle chain as a chain that at independent exponential times (with parameter $k$), makes a random transposition, where the $\ell$th transposition is chosen uniformly at random (if $\ell - 1$ is an integer multiple of $k$), or uniformly among those transpositions that involve the ending point of the previous transposition and that would result with a legitimate $k$-cycle (that is, no repetitions are allowed) if $\ell - 1$ is not an integer multiple of $k$.

We begin with $j = 0$. Note that $M_0(t)$ decreases by 1 with rate $kM_0(t)n^{-1}$ and increases, at most by 2, with rate bounded above by $k(1 - M_0(t)/n)n^{-1}$. In particular, by time $n \log n$, the number of increase events is dominated by twice a Poisson variable of parameter $k \log n$. Thus, with probability bounded below by $1 - e^{-(\log n)^2}$, at most $2(\log n)^2$ parts of size 1 have been born. On this event, $M_0(t) \leq 2(\log n)^2 + \tilde{M}_0(t)$ where $\tilde{M}_0(t)$ is a process with death only at rate $k\tilde{M}_0(t)/n$. In particular, the time of the $n - n/2(\log n)^3$th death in $\tilde{M}_0(t)$ is distributed like the random variable

$$Z_0 := \sum_{i=0}^{n-n/2(\log n)^3} \mathcal{E}_i$$

where the $\mathcal{E}_i$ are independent exponential random variables of parameter $k(n-i)$. It follows that $EZ_0 \sim 3n \log\log n/k$ and the Chebycheff bound gives, with $\zeta > 0$,

$$\begin{aligned}
\mathbb{P}(Z_0 > 2EZ_0) &\leq \mathbb{E}(e^{\zeta Z_0})e^{-2\zeta \mathbb{E}Z_0} \\
&\leq e^{-\sum_{i=0}^{n-n/2(\log n)^3} \log(1-\zeta/k(n-i))} e^{-6\zeta n \log\log n/k} \leq c^{-1}e^{-cn/(\log n)^3} ,
\end{aligned}$$

for an appropriate constant $c$, by choosing $\zeta = kn/2(\log n)^3$. We thus conclude that

$$\mathbb{P}((\mathcal{A}_0^{6/k})^{\complement}) \leq 2e^{-(\log n)^2} .$$

We continue on the event $\mathcal{A}_0^{6/k}$. We consider the process $\bar{M}_1(t) = M_1(t - 6n \log\log n/k)$. By definition $\bar{M}_1(0) \leq n/2$. The difference in the analysis of $\bar{M}_1(t)$ and $M_0(t)$ lies in the fact that now, $\bar{M}_1(t)$ may increase due to a merge of two parts of size 1 and the departure rate is now bounded below by $2k\bar{M}_1(t)n^{-1}$. Note that by time $n \log n$, the total number of arrivals due to merge of parts of size 1 has mean bounded by $n \log n \cdot k(1/(\log n)^3)^2 < kn/(\log n)^6$. Repeating the analysis concerning $M_0$, we conclude similarly that

$$\mathbb{P}((\mathcal{A}_1^{6/k+3/k})^{\complement} | \mathcal{A}_0^{6/k}) \leq 2e^{-(\log n)^2} .$$

The analysis concerning $M_j(t)$ proceeds with one important difference. Let $s_j = 6\sum_{i=0}^{j} 2^{-i}/k$, $T_j = s_j n \log\log n$ and set $\bar{M}_j(t) = M_j(t - T_{j-1})$. Now, $\bar{M}_j(t)$ can increase due to the merge of a part of size $[2^{j-1}n, 2^j n)$ with a part of size smaller than $2^j n$. On $\cap_{i=0}^{j-1} \mathcal{A}_i^{s_i}$, this has rate bounded above by

$$k\frac{1}{(\log n)^3} \cdot \frac{j}{(\log n)^3} \leq k\frac{1}{(\log n)^5} .$$

One cannot bound brutally the total number of such arrivals. However, note that the rate of departures $D_t$ is bounded below by $k2^j[\bar{M}_j(t) - 1]_+(1 - 1/(\log n)^2)/n$

(because the total mass below $2^j$ at times $t \in [T_j, n\log n]$ is, on $\cap_{i=0}^{j-1}\mathcal{A}_i^{s_i}$, bounded above by $jn/(\log n)^3 < n/(\log n)^2$). Thus, when $\bar{M}_j(t) > n2^{-j-1}/(\log n)^3$, the rate of departure $D_t >> k\frac{1}{(\log n)^5}$. Analyzing this simple birth-death chain, one concludes that

$$\mathbb{P}((\mathcal{A}_j^{s_j})^{\complement} | \cap_{i=0}^{j-1} \mathcal{A}_i^{s_i}) \le 2e^{-(\log n)^2} .$$

Since $T_j < 12n \log\log n/k \le 6n\log\log n$, this completes the proof. $\square$

An important corollary is the following control on the total mass of large parts.

**Corollary 10.** *Let* $m_\chi(t) = \sum_{i>n^\chi} N_i(t)$. *Then,*

$$\lim_{n\to\infty} \mathbb{P}\left(\min_{t\in[6n\log\log n, n\log n]} m_\chi(t) < n(1-\frac{1}{(\log n)^2})\right) = 0 .$$

The next step is the following.

**Lemma 11.** *Set* $\mathcal{B}_j = \{\max_{t\in[k^{-1}n(\log n - \log\log n - 1), n\log n)} M_j(t) \le (\log n)^2/2\}$. *Then,*

$$\lim_{n\to\infty} \mathbb{P}\left(\cup_{j=0}^{\frac{1}{4}\log_2(\log n)} \mathcal{B}_j^{\complement}\right) = 0 .$$

The proof of Lemma 11, while conceptually simple, requires the introduction of some machinery and thus deferred to the end of this subsection. Equipped with Lemma 11, we can complete the proof of the following proposition.

**Proposition 12.** *With notation as above,*

$$\lim_{n\to\infty} \mathbb{P}\left(\max_{t\in[k^{-1}n(\log n - \log\log n), n\log n]} \max_{j=0}^{\log_2 K + k} M_j(t) > (\log n)^2/2\right) = 0 .$$

*Proof.* Let $R = R(n) = \frac{1}{4}\log_2(\log n)$. Because of Lemma 11, it is enough to consider $M_j(t)$ for $j > R$.

We begin by considering $M_{R+1}(t)$. Let $B_R$ denote the intersection of $\cap_{j=0}^{R}\mathcal{B}_j$ with the complement of the event inside the probability in Corollary 10. On the event $B_R$, for $t > k^{-1}n[\log n - \log\log n - 1] := T_R$, the rate of arrivals due to merging of parts smaller than $2^R$ is bounded above by $k(2^R(\log(n))^2/n)^2$. The rate of arrivals due to parts larger than $2^R$ is bounded above by $k(2^R/n)$. Thus, the total rate of arrival is bounded above by $k2^{R+1}/n$. The rate of departure on the other hand is, due to Corollary 10, bounded below by $kM_{R+1}(t)2^R/n \cdot (1 - 1/(\log n)^2)$. Thus, for $M_{R+1}(t) > \log n/2$, the difference between the departure rate and the arrival rate is bounded below by $kM_{R+1}(t)2^R/2n$. By definition, $M_{R+1}(T_R) \le n2^{-R}$. Define $T_{R+1} = T_R + n2^{-R}$. Let $C_{R+1} = \{\max_{t\in[T_{R+1}, n\log n]} M_{R+1}(t) < \log n\}$. Then,

$$\mathbb{P}(C_{R+1}^{\complement} | B_R) \le e^{-(\log n)^2} .$$

Let $B_{R+1} = B_R \cap C_{R+1}$.

One proceeds by induction. Letting $T_{R+j} = T_{R+j-1} + n2^{-R-j+1}$, and $C_{R+j} = \{\max_{t\in[T_{R+j},n\log n]} M_{R+j}(t) < \log n\}$, we obtain from the same analysis that for $j = 1, \ldots, \log(K) + k$,

$$\mathbb{P}(C^{\complement}_{R+j+1}|B_{R+j}) \leq e^{-(\log n)^2} .$$

Thus, $\mathbb{P}(B^{\complement}_{R+\log(K)+k}) \leq \mathbb{P}(B^{\complement}_R) + (\log n)e^{-(\log n)^2} \to_{n\to\infty} 0$, while $T_{R+\log(K)+k} \leq k^{-1}n[\log n - \log\log n - 1 + 2^{-R}\sum_j 2^{-j}]$. This completes the proof, since $2^R = (\log n)^{1/4}$. □

## 2.2 Proof of Lemma 11

While a proof could be given in the spirit of the proof of Lemma 9, we prefer to present a conceptually simple proof based on comparison with the random $k$-regular hypergraph. This coupling is analogue to the usual coupling with an Erdős-Renyi random graph, see, e.g., [5] and [21]. Toward this end, we need the following definitions.

**Definition 13.** *A* $k$-regular hypergraph *is a pair $G = (V, H)$ where $V$ is a (finite) collection of vertices and $H$ is a collection of subsets of $V$ of size $k$. The* random *hypergraph $G_k(n, p)$ is defined as the hypergraph consisting of $V = \{1, \ldots, n\}$, with each subset $h$ of $V$ with $|h| = k$ taken independently to belong to $G_k(n, p)$ with probability $p$.*

Let $G_t$ denote the random $k$-hypergraph obtained by taking $V = \{1, \ldots, n\}$ and taking $H$ to consist of the $k$-hyperedges corresponding to the $k$-cycles $\gamma_1, \ldots, \gamma_{N_t}$ of the random walk $\pi_t$. It is immediate to check that $G_t$ is distributed like $G_k(n, p_t)$ with

$$p_t = 1 - \exp\left(-\frac{t}{\binom{n}{k}}\right) \sim \frac{k!t}{n^k} .$$

**Definition 14.** *A* $k$-hypertree *with $h$ hyperedges in a $k$-regular hypergraph $G$ is a connected component of $G$ with $i = (k-1)h + 1$ vertices.*

(Pictorially, a $k$-hypertree corresponds to a standard tree with hyperedges, where any two hyperedges have at most one vertex in common.) $k$-hypertrees can be easily enumerated, as in the following, which is Lemma 1 of [10].

**Lemma 15.** *The number of $k$-hypertrees with $i$ (labeled) vertices is*

$$\frac{[(k-1)h]!i^{h-1}}{h!((k-1)!)^h}, h \geq 0, \tag{20}$$

*where $h$ is the number of hyperedges and thus $i = (k-1)h + 1$.*

The next lemma controls the number of $k$-hypertrees with a prescribed number of edges in $G_t$.

**Lemma 16.** *Let*

$$\mathcal{D}_{t,h} = \{\# \text{ of } k\text{-hypertrees with } h \text{ hyperedges in } G_t \text{ is not larger than } (\log n)^{1.1}\} .$$

*Then,*

$$\mathbb{P}\left(\cap_{t>(n/k)[\log n - \log\log n - 1]} \cap_{h=0}^{\log n} \mathcal{D}_{t,h}\right) \to_{n\to\infty} 1 . \tag{21}$$

*Proof.* Let $t_0 = k^{-1}n[\log n - \log\log n - 1]$ By monotonicity, it is enough to check that

$$\mathbb{P}\left(\cap_{h=0}^{\log n}\mathcal{D}_{t_0,h}\right) \to_{n\to\infty} 1 . \tag{22}$$

Note that, with $i = (k-1)h + 1$, and adopting as a convention $h \log h = 0$ when $h = 0$,

$$\begin{aligned}
\mathbb{P}\left(\cup_{h=0}^{\log n}\mathcal{D}^{\complement}_{t_0,h}\right) &\le \sum_{h=0}^{\log n} \mathbb{P}\left(\mathcal{D}^{\complement}_{t_0,h}\right) \\
&\le \sum_{h=0}^{\log n} \frac{\mathbb{E}(\# \text{ of } k\text{-hypertrees with } h \text{ hyperedges in } G_{t_0})}{(\log n)^{1.1}} \\
&\le \frac{1}{(\log n)^{1.1}} \sum_{h=0}^{\log n} \binom{n}{r} \frac{((k-1)h)! i^{h-1}}{h!((k-1)!)^h} p_{t_0}^h (1-p_{t_0})^{\binom{i}{k}-h+i\binom{n-i}{k-1}} \\
&\le C_k \sum_{h=0}^{\log n} (\log n)^{i+h-1.1} e^{-(k-1)h(\log n - \log h)} \underset{n\to\infty}{\longrightarrow} 0 . 
\end{aligned} \tag{23}$$

(Indeed recall that $i\binom{n-i}{k-1}$ is the number of $k$-hyperedges connecting a subset of $k$ vertices to its complement in $V$, while $\binom{r}{k}$ is the number of $k$-hyperedges inside a complete $k$-hypergraph on $r$ vertices of a given spanning $k$-hypertree.) $\square$

We can now provide the

*Proof of Lemma 11.* At time $t$, $N_i(t)$ consists of cycles that have been obtained from the coagulation of cycles that have never fragmented during the evolution by time $t$, denoted $N_i^c(t)$, and of cycles that have been obtained from cycles that have fragmented and created a part of size less than $i$, denoted $N_i^f(t)$. Note that $N_i^f(t)$ is dominated above by the number of $k$-hypertrees with $s$ edges in $G_t$, which by Lemma 16 is bounded above by $(\log n)^{1.1}$ with high probability for all $i < \log n$. On the other hand, the rate of creation by fragmentation of cycles of size $i$ is bounded above by $4k/n$, and hence by time $n \log n$, with probability approaching 1 no more than $(\log n)^{1.1}$ cycles of size $i$ have been created, for all $i \leq \log n$. We thus conclude that with probability tending to 1, we have, with $t_0 = k^{-1} n[\log n - \log\log n - 1]$,

$$\max_{i \leq (\log n)^{1/4}} \max_{t \in [t_0, n \log n]} N_i^f(t) \leq (\log n)^{1.35}.$$

This yields the lemma, since for $j \leq \frac{1}{4} \log_2(\log n)$,

$$M_j(t) \leq (\log n)^{1/4} \max_{i \leq (\log n)^{1/4}} N_i(t).$$

□

## 2.3 Proof of (9)

We now prove that at time $t_{\mathrm{mix}} = (1/k) n \ln n$, the assumption (9) is satisfied, with high probability.

**Proposition 17.** *For every $\varepsilon > 0$ there exist $D = D(\varepsilon) > 0$ and $n_0 = n_0(\varepsilon)$ such that for $n > n_0$,*

$$\mathbb{P}\left(M_j\left(t_{\mathrm{mix}}\right) \leq D \log(2+j), j = 0, 1, \ldots, \log_2 \log n + k\right) \geq (1 - \varepsilon).$$

*Proof.* Consider first the time $u = \frac{1}{k}(n \log n - n \log\log n)$.

**Lemma 18.** *With probability approaching 1 as $n \to \infty$, we have $M_j(u) \leq 2^{j+4} \log n$ for all $0 \leq j \leq \log_2 n$.*

*Proof of Lemma 18.* As in the proof of Lemma 11, split $M_j(t)$ into two components $M_j^f(t)$ and $M_j^c(t)$. Note that the rate at which a fragment of size less than $2^{j+1}$ is produced is smaller than $2^{j+2}k/n$, so for any $w \leq (1/k) n \log n$, $M_j^f(w) \leq \mathrm{Poisson}(2^{j+2} \log n)$. The probability that such a Poisson random variable is more than twice its expectation is (by standard large deviation bounds) smaller than $n^{-\alpha}$ for some $\alpha > 0$, so summing over $\log_2 \log n$ values of $j$ we easily obtain that with high probability, $M_j^f(u) \leq 2^{j+3} \log n$ for all $0 \leq j \leq \log_2 \log n$.

It remains to show that $M_j^c(u) \leq \log n$ for all $0 \leq j \leq \log_2 \log n$ with high probability. To deal with this part, note that if $T_h$ denotes the number of hypertrees with $h$ hyperedges in $G_u$, then $N_i^c(u) \leq T_h$ where $i = 1 + h(k-1)$ is the number of vertices. Reasoning as in (23), we compute after simplifications (recalling that $u = (1/k)(n\log n - n\log\log n)$ and $i = 1 + h(k-1)$), for $h \geq 0$:

$$\begin{aligned}\mathbb{E}(T_h) &= \binom{n}{i}\frac{(i-1)!i^{h-1}}{h!((k-1)!)^h}p_u^h(1-p_u)^{\binom{i}{k}-h+i\binom{n-i}{k-1}} \\ &\leq \frac{n(\log n)^h}{h!i}(1-p_u)^{i\binom{n-i}{k-1}} \leq \frac{n^{1-i}(\log n)^{1+hk}}{h!i}. \end{aligned} \tag{24}$$

Thus summing over $i = 2$ to $i = \lceil \log n \rceil$, we conclude by Markov's inequality that $M_j^c(u) = 0$ for all $1 \leq j \leq \log_2 \log n$ with high probability. For $i = 1$ or $h = 0$, we get from (24)

$$\mathbb{E}(T_0) \leq \log n.$$

Computing the variance is easy: writing $T_0 = \sum_{v\in V} \mathbf{1}_{\{v \text{ is isolated}\}}$, we get

$$\mathrm{var}(T_0) \leq \mathbb{E}(T_0) + \sum_{v\neq w} \mathrm{cov}(\mathbf{1}_{\{v \text{ is isolated}\}}, \mathbf{1}_{\{w \text{ is isolated}\}}).$$

But note that

$$\mathbb{P}(v \text{ is isolated}, w \text{ is isolated}) = \frac{\mathbb{P}(v \text{ is isolated})^2}{1-p_u},$$

so

$$\mathrm{var}(T_0) \leq \mathbb{E}(T_0) + \mathbb{E}(T_0)^2\left(\frac{1}{1-p_u} - 1\right) \leq \mathbb{E}(T_h) + o(1).$$

Thus by Chebyshev's inequality, $\mathbb{P}(M_0^c(u) > 2\log n) \to 0$ as $n \to \infty$. This proves the lemma. $\square$

With this lemma we now complete the proof of Proposition 17. We compare $(M_j(t), t \geq u)$ to independent queues as follows. By Proposition 12, on an event of high probability, during the interval $[u, t_{\text{mix}}]$ the rate at which some two cycles of size smaller than $\log n$ coagulate is smaller than $O((\log n)^6/n^2)$, so the probability that this happens during this interval of time is $o(1)$. Likewise, the rate at which some cluster smaller than $\log n$ will fragment is at most $k(\log n)^4/n^2$, so the probability that this happens during the interval $[u, t_{\text{mix}}]$ is $o(1)$. Now, apart if we reject any $k$-cycle that would create such a transition, the only possible transition for $M_j$ are increases by 1 (through the fragmentation of a component larger than $2\log n$) and decrease by 1 (through coagulation with cycle larger than $\log n$). The respective rates of these transitions is, as in (13), at most $2^j\lambda^+ = 2^j k/(n-k)$, and at least $\nu = 2^j(k/n)(1-(\log n)^3/n))$ as in (18). This can be compared to a queue where both the departure rate and the arrival rate are equal to $\lambda^+$, say $\bar{M}_j(t)$. The difference between $M_j(t)$ and $\bar{M}_j(t)$ is that some of the customers having left in $\bar{M}_j(t)$ might not have left yet in $M_j(t)$. Excluding the initial customers, a total of Poisson$(2^j \log\log n)$ customers arrive in the queue $\bar{M}_j(t)$ during the interval $[u, t_{\text{mix}}]$, so the probability that any one of those customers has not yet left by time $t_{\text{mix}}$ in $M_j(t)$ given that it did leave in $\bar{M}_j(t)$ is no more than $\lambda^+/\nu - 1 = O((\log n)^3/n)$, where the constants implicit in $O(\cdot)$ do not depend on $j$ or $n$. Thus with probability greater than $1 - O(2^j \log\log n(\log n)^3/n)$, there is no difference between $M_j(t_{\text{mix}})$ and $\bar{M}_j(t_{\text{mix}})$. Moreover,

$$\bar{M}_j(t_{\text{mix}}) \preceq \text{Poisson}(1) + R_j, \tag{25}$$

where $R_j$ is the total number of initial customers customers that haven't departed yet by time $t_{\text{mix}}$. Using Lemma 18,

$$R_j \preceq \frac{1}{\lambda^+} \max_{1 \le q \le 2^{j+4}\log n} E_q \tag{26}$$

where $(E_q, q \ge 1)$ is a collection of i.i.d. standard exponential random variables. Using the independence of the queues $\bar{M}_j(t)$, in combination with (25) and (26) as well as standard large deviations for Poisson random variables, the proposition follows immediately. $\square$

## 2.4 Conclusion - small cycles

Combining Propositions 3 and 12, and using the notation introduced in the beginning of this section, we have proved the following. Fix $\varepsilon > 0$. Then there is a $c_{\varepsilon,k} > 0$ such that with $t = t(n) = k^{-1} n \log n + c_{\varepsilon,k} n$, and all large $n$,

$$\left\| (N_i(t))_{i=1}^K - (Z_i)_{i=1}^K \right\| < \varepsilon.$$

We now deduce the following:

**Proposition 19.** *Fix $\varepsilon > 0$. Then there is a $c_{\varepsilon,k} > 0$ such that with $t = t(n) = k^{-1} n \log n + c_{\varepsilon,k} n$, and all large $n$,*

$$\left\| (N_i(t))_{i=1}^K - (N_i)_{i=1}^K \right\| < \varepsilon, \tag{27}$$

*where $(N_i)_{1 \le i \le n}$ is the cycle distribution of a random permutation sampled according to the invariant distribution $\mu$.*

*Proof.* By (27)), and the triangle inequality, all that is needed is to show that

$$\left\| (Z_i)_{i=1}^K - (N_i)_{i=1}^K \right\| \to 0. \tag{28}$$

Whenever $k$ is even, and thus $\mu$ is uniform on $\mathcal{S}_n$, (28) is a classical result of Diaconis–Pitman and of Barbour, with explicit upper bound of $4K/n$, see [4] or the discussions around [3, Theorem 2] and [2, Theorem 4.18].

In case $k$ is odd, $\mu$ is uniform on $\mathcal{A}_n$. A sample $\gamma$ from $\mu$ can be obtained from a sample $\gamma'$ of the uniform measure on $\mathcal{S}_n$ using the following procedure. If $\gamma'$ is even, take $\gamma = \gamma'$, otherwise let $\gamma = \pi \circ \gamma'$ where $\pi$ is some fixed transposition (say (12)). The probability that the collection of small cycles in $\gamma$ differs from the corresponding one in $\gamma'$ is bounded above by $4K/n \to 0$, which completes the proof. $\square$

# 3 Large cycles and Schramm's coupling

Fix $\varepsilon > 0$ and $\chi \in (7/8, 1)$. Recall that $K$ is the closets dyadic integer to $\lfloor n^\chi \rfloor$ and that a cycle is called small if its size is smaller than $K$. For $n$ large, let $t = t(n) = k^{-1} n \log n + c_{\varepsilon,k} n$. We know by the previous section, see Proposition 19, that at this time, for $n$ large, the distribution of the small cycles of the permutation $\pi_t$ is arbitrarily close (variational distance smaller than $\varepsilon$) to that of a (uniformly chosen) random permutation $\pi'$. Therefore we can find a coupling of $\pi_t$ and $\pi'$ in such a way that:

$$\mathbb{P}(\text{the small cycles of } \pi := \pi_t \text{ and } \pi' \text{ are identical}) \geq 1 - \varepsilon. \tag{29}$$

We can now provide the

*Proof of Theorem 1.* We will construct an evolution of $\pi'$, denoted $\pi'_s$, that follows the random $k$-cycle dynamic (and hence, $\pi'_s$ has cycle structure whose law coincides with the law of the cycle structure of a uniformly chosen permutation, at all times). The idea is that with small cycles being the hardest to mix, coupling $\pi_{t+s}$ and $\pi'_s$ will now take very little time. To prove this, we describe a modified version of the Schramm coupling introduced in [21], which has the additional property that it is difficult to create small unmatched pieces.

To describe this coupling, we will need some notations from [21]. Let $\Omega_n$ be the set of discrete partitions of unity:

$$\Omega_n = \{(x_1 \geq \ldots \geq x_n) : x_i \in \{0/n, \ldots, n/n\} \text{ for all } 1 \leq i \leq n, \text{ and } \sum_{i=1}^{n} x_i = 1\}.$$

We identify the cycle count of $\pi_t$ with a vector $Y_t \in \Omega_n$. We thus want to describe a coupling between two processes $Y_t$ and $Z_t$ taking their values in $\Omega_n$ and started from some arbitrary initial states. The coupling will be described by a joint Markovian evolution of $(Y_t, Z_t)$.

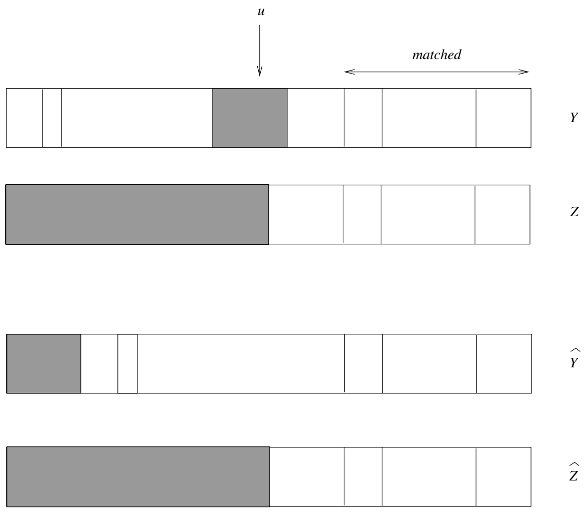


Figure 1: First step of the coupling. A point $\tilde{u}$ is uniformly chosen on $(0, 1)$ and picks a part in $Y$ and $Z$, which are then rearranged into $\hat{Y}, \hat{Z}$.

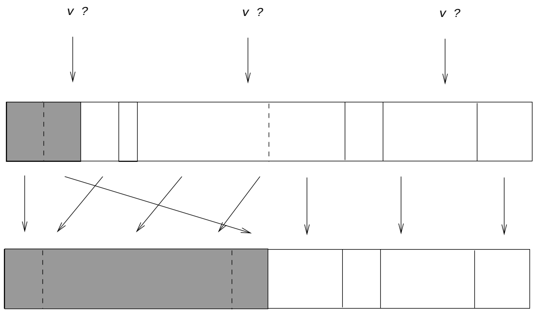


Figure 2: A second point $\tilde{v}$ is chosen uniformly in (0,1) and serves as a second size-biased pick for $\hat{Y}$. $\tilde{v}$ is mapped to $\tilde{v}' = \Phi(\tilde{v})$ which gives a second size-biased pick for $\hat{Z}$.

We now begin by describing the construction of a random transposition. For $x \in (0,1)$, let $\{x\}_n$ denote the smallest element of $\{1/n, \ldots, n/n\}$ not smaller than $x$. Let $\tilde{u}, \tilde{v}$ be two random points uniformly distributed in $(0,1)$, set $u = \{\tilde{u}\}_n, v = \{\tilde{v}\}_n$ and condition them so that $u \neq v$. Note that $u, v$ are both uniformly distributed on $\{1/n, \ldots, n/n\}$. If we focus for one moment on the marginal evolution of $(Y_t)$, then applying one transposition to $Y_t$ can be realized by associating to $Y_t \in \Omega_n$ a tiling of the semi-open interval $(0,1]$ where each tile is equally semi-open and there is exactly one tile for each nonzero coordinate of $Y_t$. (The order in which those tiles are put down may be chosen arbitrarily and does not matter for the moment.) If $u$ and $v$ fall in different tiles then we merge the two tiles together and get a new element of $\Omega_n$ by sorting in decreasing order the size of the tiles. If $u$ and $v$ fall in the same tile then we use the location of $v$ to split that tile into two parts: one that is to the left of $v$, and one that is to its right (we keep the same semi-open convention for every tile). This procedure works because, conditionally on falling in the same tile $C$ as $u$, then $v$ is equally likely to be on any point of $C \cap \{1/n, \ldots, n/n\}$ distinct from $v$, which is the same fragmenting rule as explained at the beginning of the proof of Proposition 3.

We now explain how to construct one step of the joint evolution. If $Y, Z \in \Omega^n$ are two unit discrete partitions, then we can differentiate between the entries that are matched and those that are unmatched; two entries from $Y$ and $Z$ are matched if they are of identical size. Our goal will be to create as many matched parts as possible. Let $Q$ be the total mass of the unmatched parts. When putting down the tilings associated with $Y$ and $Z$ we will do so in such a way that all matched parts are at the right of the interval $(0,1]$ and the unmatched parts occupy the left part of the interval, as in Figure 1. If $u$ falls into the matched parts, we do not change the coupling beyond that described in [21]: that is, if $v$ falls in the same component as $u$ we make the same fragmentation in both copies, while otherwise we make the corresponding coalescence. The difference occurs if $u$ falls in the unmatched parts. Let $y$ and $z$ be the respective components of $Y$ and $Z$ where $u$ falls, and let $\hat{Y}, \hat{Z}$ be the reordering of $Y, Z$ in which these components have been put to the left of the interval $(0,1]$. Let $a = |y|$ and let $b = |z|$ be the respective lengths of the pieces selected with $u$, and assume without loss of generality that $a < b$. Further rearrange, if needed, $y$ and $z$ so that after the rearrangement, $|u| = 1/n$. Because $v \neq u$, necessarily $v > 1/n$ (and is uniformly distributed on the set $\{2/n, \ldots, n/n\}$). The point $v$ designates a size-biased sample from the partition $\hat{Y}$ and we will construct another point $v'$, which will also be uniformly distributed on $\{2/n, \ldots, n/n\}$, to similarly select a size-biased sample from $\hat{Z}$. However, while in the coupling of [21] one takes $v = v'$, here we do not take them equal and apply to $v$ a measure-preserving map $\Phi$, defined as follows. Define the function

$$\Phi(x) = \begin{cases} x & \text{if } x > b \text{ or if } 1/n \le x \le \gamma_n + 1/n, \\ x - \gamma_n & \text{if } a < x \le b, \\ x + b - a & \text{if } \gamma_n + 1/n < x \le a, \end{cases} \tag{30}$$

where $\gamma_n := \{(a - 1/n)/2\}_n$. See Figure 2 for description of $\Phi$. Note that $\Phi$ is a measure-preserving map and hence $\tilde{v}' := \Phi(\tilde{v})$ is uniformly distributed on $(0, 1)$. Define $v' = \{\tilde{v}'\}_n$. With $u, v$ and $v'$ selected, the rest of the algorithm is unchanged, i.e., we make the corresponding coagulations and fragmentations.

This coupling has a number of remarkable properties which we summarize below. Essentially, the total number of unmatched entries can only decrease, and furthermore it is very difficult to create small unmatched entries, as the smallest unmatched entry can only become smaller by a factor of at most 2.

**Lemma 20.** *Let $U$ be the size of the smallest unmatched entry in two partitions $Y, Z \in \Omega^n$, and let $Y', Z'$ be the corresponding partitions after one transposition of the coupling, and let $U'$ be the size of the smallest unmatched entry in $Y', Z'$. Assume that $2^j \le U < 2^{j+1}$ for some $j \ge 0$. Then it is always the case that $U' \ge U - \{U/2\}_n$, and moreover,*

$$\mathbb{P}(U' \le 2^j) \le 2^{j+2}/n.$$

*Finally, the number of unmatched parts may only decrease.*

**Remark 21.** *Since $U' \ge U - \{U/2\}_n$, in particular $U' \ge 2^{j-1}/n$.*

*Proof.* That the number of unmatched entries can only decrease is similar to the proof of Lemma 3.1 in [21]. (In fact it is simpler here, since that lemma requires looking at the total number of unmatched entries of size greater than $\varepsilon$. Since in our discrete setup no entry can be smaller than $\varepsilon = 1/n$ we do not have to take this precaution). Recall that $M_j$ is the total number of parts in the range $[2^j, 2^{j+1})$. The only case that $U$ can decrease is if there is a fragmentation of an unmatched entry, since matched entries must fragment in exactly the same way. Now, note that the coupling is such that when an unmatched entry is selected and is fragmented, then all subsequent pieces are either greater or equal to $a - \{a/2\}_n$ (where $a$ is the size of the smaller of the two selected unmatched entries), or are matched. Moreover, for such a fragmentation to occur, one must select the lowest unmatched entry (this has probability at most $M_j 2^{j+1}/n$, since there may be several unmatched entries with size $U$), and then fragment it, which has probability at most $2^{j+1}/n$, and thus $\mathbb{P}(U' < U) \le 4M_j 4^j/n^2$. Since $M_j 2^j \le n$, this completes the proof. $\square$

We have described the basic step of a (random) transposition in the coupling. The step corresponding to a random $k$-cycle $\gamma = (\gamma_1, \gamma_2, \ldots, \gamma_k)$ is obtained by taking $u_1 = \gamma_1$, generating $v, v'$ as in the coupling above (corresponding to the choice of $\gamma_2$), rearranging and taking $u_2$ to correspond to the location of $v, v'$ after the rearrangement, drawing new $v, v'$ (corresponding to $\gamma_3$) and so on. In doing so, we are disregarding the constraint that no repetitions are present in $\gamma$. However, as it turns out, we will be interested in an evolution lasting at most

$$\Delta := n^{5/8} \log n, \tag{31}$$

and the expected number of times that a violation of this constraint occurs during this time is bounded by $2\Delta k^2/n$, which converges to 0 as $n \to \infty$. Hence, we can in what follows disregard this violation of the constraint.

Now, start with two configurations $Y_0, Z_0$ such that $Z_0$ is the element of $\Omega_n$ associated with a random uniform permutation. Assume also that initially, the small parts of $Y_0$ and $Z_0$ (i.e., those that are smaller than $K$, the closest dyadic integer to $\lfloor n^\chi \rfloor$), are exactly identical, and that they have the same parity. As we will now see, at time $\Delta$, $\pi_{t+\Delta}$ and $\pi'_\Delta$ will be coupled. Note also that, since initially all the parts that are smaller than $K$ are matched, the initial number of unmatched entries cannot exceed $n/K \leq n^{1/8}$, and this may only decrease with time by Lemma 20.

**Lemma 22.** *In the next $\Delta$ units of time, the random permutation $\pi'_s$ never has more than $2n^{-1/4}(\log n)^2$ mass in parts smaller than $n^{3/4}$, with high probability.*

*Proof.* The proof is the same as that of Proposition 12, only simpler because the initial number of small clusters is within the required range. We omit further details. [This can also be seen by computing the probability that a given uniform permutation $\pi'$ has more than $n^{-1/4} \log n$ mass in parts smaller than $n^{3/4}$, and summing over Poisson($\Delta$) steps.] $\square$

**Lemma 23.** *In the next $\Delta$ units of time, every unmatched part is greater than or equal to $n^{-1/4}/2$, with high probability.*

*Proof.* Recall that the total number of unmatched parts can never increase. Suppose the smallest unmatched part at time $s$ is of scale $j$ (i.e., of size in $[2^j, 2^{j+1})$), and let $j = U(s)$ be this scale. Then, when touching this part, the smallest scale it could go to is $j-1$, by the properties of the coupling (see Lemma 20). This happens with probability at most $2^{j+2}/n$. On the other hand, with the complementary probability, this part experiences a coagulation. And with reasonable probability, what it coagulates with is larger than itself, so that it will jump to scale $j+1$

or larger. To compute this probability, note that since this is the smallest unmatched part, all smaller parts are matched and thus have a total mass controlled by Lemma 22. In particular, on an event of high probability, this mass is at most $q := 2n^{-1/4}(\log n)^2$. It follows that with probability at least $1-q$, the part jumps to scale at least $j+1$, and with probability at most $p := 2^j/n < n^{-1/4} < q$, to scale $j-1$. Now, when this part jumps to scale at least $j+1$, this does not necessarily mean that the *smallest* unmatched part is in scale at least $j+1$, since there may be several small unmatched parts in scale $j$. However, there can never be more than $2n^{1/8}$ such parts. If an unmatched piece in scale $j$ is touched, we declare a success if it moves to scale $j+1$ (which has probability at least $1-q$, given that it is touched) and a failure if it goes to scale $j-1$ or stays at scale $j$ (which has probability at most $q$). If $2n^{1/8}$ successes occur before any failure occurs at scale $j$, then we know that no unmatched cycle can exist at scale smaller than $j$. But the number of trials before a first failure occurs at scale $j$ with $2^j \le n^{1/4}$ is stochastically dominated below by a geometric random variable with parameter $q$. The probability that this happens before the $2n^{1/8}$th trial is, by standard large deviation bounds, at most $e^{-cq^{-1}} \le e^{-cn^{1/5}}$. The number of times any unmatched piece is touched at scale $j$ is by a very crude bound at most $2\Delta$ with high probability, so there cannot be any more trials than this. Since $2\Delta e^{-cn^{1/5}} \to 0$, we deduce that with high probability,

$$U(s) \ge U(0), \quad s \le \Delta.$$

Since every unmatched part is initially greater than $n^{-1/4}$, this implies the statement of the lemma. □

*End of the proof of Theorem 1.* We now are going to prove that, after $\Delta = n^{5/8}\log n$ steps, there are no more unmatched parts with high probability. The basic idea is that, on the one hand, the number of unmatched parts may never increase, and on the other hand, it does decrease frequently enough. Since each unmatched part is greater than $n^{-1/4}/2$ during this time, any given pair of unmatched parts is merging at rate roughly $n^{-1/2}$. There are initially no more than $2n^{1/8}$ unmatched parts, so after $n^{5/8}\log n = \Delta$ steps, no more unmatched part remains with high probability.

To be precise, assume that there are $L$ unmatched parts. Let $T_L$ be the time to decrease the number of unmatched parts from $L$ to $L-2$. Observe that, for parity reasons ($\pi$ and $\pi'$ must have the same parity of number of parts at all times), $L$ is always even. Note also that $L=2$ is impossible, so $L$ is at least 4. Assume to start with that both copies have at least 2 unmatched parts. Then, at rate greater than $n^{-1/4}/2$ we pick an unmatched part in the first point $u_1$ for the $k$-cycle. Since there are at least 2 unmatched parts in each copy, let $R$ be the interval of $(0,1)$

corresponding to a second unmatched part in the copy that contains the larger of the two selected ones. Then $|R| > n^{-1/4}/2$, and moreover when $v$ falls in $R$, we are guaranteed that a coagulation is going to occur in both copies. We interpret this event as a success, and declare every other possibility a failure. Hence if $G$ is a geometric random variable with success probability $n^{-1/4}/2$, and $(X_j)_{j=1}^{\infty}$ are i.i.d. exponentials with mean $2n^{1/4}$, the total amount of time before a success occurs is dominated by $\sum_{j=1}^{G} X_j$.

If however one copy (say $\pi$) has only one unmatched part, then one first has to break that component, which takes at most an exponential random variable with rate $n^{-1/2}/4$. Note that the other copy must have had at least 3 unmatched parts, so after breaking the big one, both copies have now at least two unmatched copies and we are back to the preceding case. It follows from this analysis that in any case, $T_L$ is dominated by

$$T_L \preceq Y + \sum_{j=1}^{G} X_j$$

and so $\mathbb{E}(T_L) \leq 4n^{1/2} + 4n^{1/2} = 8n^{1/2}$. Now, let

$$\tau_L = T_L + T_{L-2} + \ldots + T_4$$

and let $T = \tau_{2n^{1/8}}$. Then $T$ is the time to get rid of all unmatched parts. We obtain from the above $\mathbb{E}(T) \leq 16n^{5/8}$. By Markov's inequality, it follows that $T < n^{5/8} \log n = \Delta$ with high probability. This concludes the proof of Theorem 1. $\square$

**Acknowledgments** N.B. is grateful to the Weizmann Institute, Microsoft Research's Theory Group and the Technion for their invitations in June 2008, August 2008 and April 2009, respectively. Some of this research was carried out during these visits.

# References

[1] D. Aldous (1983). Random walks on finite groups and rapidly mixing Markov chains. *Séminaire de Probabilités XVII. Lecture Notes in Math.* 986, 243–297. Springer, New-York.

[2] R. Arratia, A. Barbour and S. Tavaré (2003). *Logarithmic Combinatorial Structures: a probabilistic approach.* EMS Monographs in Mathematics. European Mathematical Society (EMS), Zürich.

[3] R. Arratia and S. Tavaré (1992). The cycle structure of random permutations. *Ann. Probab.*, 20, 1567–1591.

[4] A. Barbour (1990) Comments on "Poisson approximations and the Chen-Stein method", *Statistical Science*, 5, 425–427.

[5] N. Berestycki and R. Durrett (2006). A phase transition in the random transposition random walk. *Probab. Theor. Rel. Fields.* 136, 203–233.

[6] R. M. Burton and Y. Kovchegov (2006). Mixing times via super-fast coupling. Preprint, arXiv:math/0609568

[7] P. Diaconis (1988). *Group representations in probability and statistics.* Lecture Notes - Monograph Series, Vol. 11, Institute of Mathematical Statistics.

[8] P. Diaconis and M. Shahshahani (1981). Generating a random permutation with random transpositions. *Z. Wahrsch. Verw. Geb.* 57, 159-179.

[9] L. Flatto, A.M. Odlyzko, and D.B. Wales (1985). Random shuffles and group representations. *Ann. Probab.* 13, 154–178.

[10] M. Karoński and T. Luczak (1997). The number of connected sparsely edged uniform hypergraphs. *Discrete Math.* 171, 153–167.

[11] D. Levin, Y. Peres and E. Wilmer. *Markov chains and mixing times.* American Mathematical Society, 2009.

[12] N. Lulov (1996). Random walks on symmetric groups generated by conjugacy classes, Ph.D. Thesis, Harvard University.

[13] N. Lulov and I. Pak (2002). Rapidly mixing random walks and bounds on characters of the symmetric group, *Journal of Algebraic Combinatorics*, vol. 16, 151-163.

[14] P. Matthews (1988). A strong uniform time for random transpositions. *J. Theoret. Probab.*, 1, 411-423.

[15] L. Saloff-Coste (2003). Random Walks on Finite Groups. In: H. Kesten, ed. *Probability on Discrete Structures*, Encyclopaedia of Mathematical Sciences (110), Springer.

[16] Y. Roichman (1996). Upper bound on the characters of the symmetric groups. *Inventiones Mathematicae* 125, 451–485.

[17] Y. Roichman (1996). Characters of the symmetric group: formulas, estimates, and applications. In: *Emerging applications of number theory,* Ed: D. A. Hejhal, J. Friedman, M. C. Gutzwiller, A. M. Odlyzko

[18] S. Roussel (1999). Marches aléatoires sur e groupe symétrique. Thèse de doctorat, Toulouse.

[19] S. Roussel (2000). Phénomène de cutoff pour certaines marches aléatoires sur le groupe syymétrique. *Colloquium Math.*, 86, 111–135.

[20] L. Saloff-Coste and J. Zúñiga (2008). Refined estimates for some basic random walks on the symmetric and alternating groups. *Alea*, 4, 359-392.

[21] O. Schramm (2005). Composition of random transpositions. *Israel J. Math.*, 147, 221-244.

[22] A. Vershik and S. Kerov (1981). Asymptotic theory of characters on the symmetric group. *Functional Analysis*, 15, 246–255.