

# Cutoff for the Burnside process on the hypercube.

Evita Nestoridi <sup>\*</sup>      Allan Sly <sup>†</sup>

## Abstract

We study the Burnside process for the action of  $S_n$  on the binary hypercube  $\{0,1\}^n$ . We prove worst-case total variation cutoff and identify a Gaussian limit profile.

## 1 Introduction

Jerrum introduced the Burnside process to sample orbits of a finite group action nearly uniformly [16]. It implements Burnside's lemma as a two-step chain. If a finite group  $G$  acts on a finite set  $X$ , write

$$G_x := \{g \in G : gx = x\}, \quad X_g := \{y \in X : gy = y\}.$$

From  $x$ , the chain chooses  $g$  uniformly from  $G_x$ , then chooses the next state uniformly from  $X_g$ . Thus

$$P(x, y) = \frac{1}{|G_x|} \sum_{g \in G_x \cap G_y} \frac{1}{|X_g|}.$$

The chain is reversible with stationary measure  $\pi(x) = 1/(Z|O_x|)$ , where  $O_x$  is the orbit of  $x$  and  $Z$  is the number of orbits.

---

<sup>\*</sup>Department of Mathematics, Stony Brook University, NY 11794. E-mail: [evrydiki.nestoridi@stonybrook.edu](mailto:evrydiki.nestoridi@stonybrook.edu). Supported by the Simons Foundation Travel Support for Mathematicians – MPSTSM00007955 and the NSF grant DMS-2450510.

<sup>†</sup>Department of Mathematics, Princeton University, NJ 08544. [allansly@princeton.edu](mailto:allansly@princeton.edu). Supported by a Simons Investigator Grant.

We consider  $X_n = \{0, 1\}^n$  under coordinate permutations by  $S_n$ . The orbits are indexed by the number of ones, so for  $x \in X_n$ , we set  $|x| = \sum_i x_i$ , and

$$\pi_n(x) = \frac{1}{(n+1)\binom{n}{|x|}}.$$

One step has a useful cycle description: choose a uniform permutation that preserves the zero and one coordinates of  $x$ , independently label each cycle by a fair bit, and give every coordinate in that cycle its label.

For probability measures  $\mu, \nu$  on a finite set, write

$$\|\mu - \nu\|_{\text{TV}} := \frac{1}{2} \sum_{x \in X} |\mu(x) - \nu(x)|.$$

Let  $P_n$  be the transition kernel of the hypercube chain, and write  $(X_t)$  for the chain. Define

$$d_n^x(t) := \|P_n^t(x, \cdot) - \pi_n\|_{\text{TV}}, \quad d_n(t) := \max_{x \in X_n} d_n^x(t).$$

**Definition 1.** Let  $(X_n, P_n, \pi_n)$  be a sequence of finite irreducible Markov chains. Let  $d_n$  denote the corresponding worst-case total variation distance. The sequence exhibits cutoff at time  $t_n$  with window  $w_n$  if  $w_n = o(t_n)$  and

$$\lim_{c \rightarrow \infty} \liminf_{n \rightarrow \infty} d_n(\lfloor t_n - cw_n \rfloor) = 1, \quad \lim_{c \rightarrow \infty} \limsup_{n \rightarrow \infty} d_n(\lfloor t_n + cw_n \rfloor) = 0.$$

Our main theorem states that the Burnside process exhibits cutoff at  $t_n := \frac{\log n}{\log 16}$ , with a window of order  $O(\sqrt{\log n})$ . Moreover, we determine that the limit profile is Gaussian.

**Theorem 2.** The Burnside process on  $\{0, 1\}^n$  exhibits total variation cutoff at time  $t_n$  with window  $a_n$ . Moreover, for every  $u \in \mathbb{R}$ ,

$$\lim_{n \rightarrow \infty} d_n(\lfloor t_n + ua_n \rfloor) = 1 - \Phi(u),$$

where  $\Phi$  is the standard normal distribution function and

$$a_n := \frac{\pi}{4\sqrt{6}(\log 2)^{3/2}} \sqrt{\log n}.$$

Jerrum proved that  $O(\sqrt{n})$  steps suffice for total variation mixing for the Burnside process on the hypercube, and Aldous improved this to  $O(\log n)$  [16, 1]. In contrast, Goldberg and Jerrum constructed permutation groups for which Burnside-process mixing is exponential in the degree [13].

Diaconis studied the hypercube chain as the Bose–Einstein model of placing  $n$  indistinguishable balls in two boxes and proved that its lumped chain mixes in a bounded number of steps, uniformly in  $n$  [4]. Diaconis and Zhong used Hahn-polynomial diagonalization to obtain sharp bounds from special states; from the all-zero state the distance after  $\ell$  steps is of order  $4^{-\ell}$  [8]. Diaconis, Lin, and Ram showed that arbitrary starts mix in total variation in  $O(\log n)$  steps, whereas most starts require order  $n/\log n$  in  $\ell^2$  distance; their diagonalization uses binomial-coefficient-valued eigenvectors [6]. Feng and Paguyo studied a twisted Bose–Einstein chain arising from the twisted Burnside process, proving mixing bounds with applications to sampling Catalan structures [11]. We identify the worst-case total variation cutoff and its Gaussian profile.

Beyond the hypercube, the process has been studied via auxiliary variables and for commuting chains, set partitions, partitions and contingency tables, Pólya trees, orbit counting, lifted inference, flag varieties, and Sylow double cosets; a dual process has also recently been developed [2, 21, 19, 5, 3, 9, 14, 7, 15, 10].

## 1.1 Proof strategy

This subsection outlines the two time scales in the proof and the coupling mechanism that produces the cutoff profile.

Symmetry of the model means it suffices to analyse a two dimensional Markov chain  $(N_t, M_t)$ . The magnetization

$$N_t = \sum_{i=1}^n X_t(i),$$

is itself a Markov chain and mixes in  $O(1)$  steps. The second  $M_t$  counts the coordinates equal to 1 in both  $X_t$  and the initial configuration. We will study the centered overlap  $K_t = M_t - N_t N_0/n$  which mixes in time  $O(\log n)$  and determines the cutoff time. Conditional on the two selected totals in the next step,

$$\mathbb{E}[K_{t+1} \mid \mathcal{G}_t] = K_t(Z_t - Z'_t),$$

where  $\mathcal{G}_t$  is defined in Subsection 3.2 and  $Z_t, Z'_t$  in Section 4. Thus the drift is governed by

$$\Gamma_t = K_0 \prod_{j=0}^{t-1} (Z_j - Z'_j).$$

The logarithms of these factors are asymptotically i.i.d. with negative mean. Their central limit theorem determines when  $|\Gamma_t|$  reaches the  $\sqrt{n}$ -scale and yields the Gaussian profile. Once two copies have centered overlaps of this order, their one-step laws are built from nearby hypergeometric distributions. At a positive fraction of bulk times these laws have overlap bounded away from zero, so repeated maximal couplings give a geometric failure bound.

**Probabilistic asymptotic notation.** For deterministic  $b_n > 0$ ,  $Y_n = o_{\mathbb{P}}(b_n)$  means  $Y_n/b_n \rightarrow 0$  in probability, while  $Y_n = O_{\mathbb{P}}(b_n)$  means that  $(Y_n/b_n)_{n \geq 1}$  is tight. We reserve deterministic  $o(\cdot)$  and  $O(\cdot)$  for non-random quantities.

## 1.2 The cutoff constants

The constants in the cutoff location and window come from the limiting one-step contraction of the centered overlap. In the continuous model the factor  $Z_t - Z'_t$  is replaced by  $R - R'$ ; the mean and variance of  $\log |R - R'|$  determine the two cutoff scales. We now define  $R$  and compute these constants explicitly.

**Definition 3.** *Let*

$$R = \sum_{j=1}^{\infty} \eta_j \xi_j, \quad \xi_j = U_j \prod_{i < j} (1 - U_i),$$

where  $(U_j)_{j \geq 1}$  are i.i.d. uniform random variables on  $[0, 1]$ , and  $(\eta_j)_{j \geq 1}$  are i.i.d. Bernoulli(1/2) random variables, independent of the  $U_j$ 's. Let  $R'$  be an independent copy of  $R$ , and set

$$\lambda := -\mathbb{E}[\log |R - R'|], \quad \sigma^2 := \text{Var}(\log |R - R'|).$$

The cutoff scales satisfy  $t_n = \log n / (2\lambda)$  and  $a_n = (\sigma/\lambda) \sqrt{\log n / (2\lambda)}$ .

The weights  $(\xi_j)$  have the GEM(1) law. Coloring them independently by Bernoulli(1/2) labels gives the mass assigned to the atom 1 in the standard Dirichlet-process paintbox [12, 22, 20], and hence

$$R \sim \text{Beta}(1/2, 1/2).$$

Equivalently,  $R = (1 + \cos \Theta)/2$  for  $\Theta \sim \text{Unif}[0, \pi]$ . For independent  $\Theta, \Theta'$ , the identity  $\log |R - R'| = \log |\cos \Theta - \cos \Theta'| - \log 2$ , together with the Fourier series for  $\log(2|\sin x|)$ , gives

$$\mathbb{E} \log |R - R'| = -2 \log 2, \quad \text{Var}(\log |R - R'|) = \frac{\pi^2}{6}.$$

Thus  $\lambda = 2 \log 2$ ,  $\sigma^2 = \pi^2/6$ , and the displayed values of  $t_n$  and  $a_n$  follow.

## 2 Mixing of the Magnetization Chain

This section proves that the magnetization mixes in a bounded number of steps, uniformly in  $n$ . We write  $\mathbb{P}_k$  for probabilities when the magnetization chain starts from  $N_0 = k$ .

Let  $\rho$  be uniform in  $S_m$ , label its cycles independently by fair bits, and let  $J_m$  be the total size of the cycles labeled 1, with  $J_0 = 0$ . Set

$$\alpha_m := \mathbb{E}_{\rho \in S_m} [2^{-\#\text{cyc}(\rho)}], \quad \alpha_0 := 1.$$

Then, for  $0 \leq q \leq m$ ,

$$\mathbb{P}(J_m = q) = \alpha_q \alpha_{m-q}.$$

Indeed, for a fixed  $q$ -subset  $A$ , the probability that  $A$  is the selected set is  $\binom{m}{q}^{-1} \alpha_q \alpha_{m-q}$ ; summing over  $A$  gives the formula. The cycle-index identity and Stirling's formula give

$$\alpha_m = \frac{(1/2)(3/2) \cdots (m-1/2)}{m!} = \frac{\binom{2m}{m}}{4^m} \asymp \frac{1}{\sqrt{m+1}}, \quad m \geq 1.$$

In particular,  $\alpha_m \geq c_0/\sqrt{m+1}$  for a universal  $c_0 > 0$ .

**Lemma 4.** *There is a universal  $\delta > 0$  such that, for every  $k \in \{0, \dots, n\}$  and every integer  $x \in [n/10, 2n/5]$ ,*

$$\mathbb{P}_k(N_1 = x) \geq \frac{\delta}{n}.$$

*Proof.* Set  $m = k$  and  $m' = n - k$ . The two level sets contribute independent variables with laws  $J_m$  and  $J_{m'}$ , so

$$\mathbb{P}_k(N_1 = x) = \sum_{q \in I_x} \mathbb{P}(J_m = q) \mathbb{P}(J_{m'} = x - q),$$

where

$$I_x := \{q \in \mathbb{Z} : 0 \leq q \leq m, 0 \leq x - q \leq m'\}.$$

By the lower bound on  $\alpha$ ,

$$\mathbb{P}(J_m = q) \geq \frac{c_0^2}{m+1}, \quad \mathbb{P}(J_{m'} = x - q) \geq \frac{c_0^2}{m'+1}.$$

Writing  $s = \min\{m, m'\}$ , a direct interval count gives

$$|I_x| = 1 + \min\{m, m', x, n - x\}.$$

For  $x \in [n/10, 2n/5]$ , this is at least  $(s+1)/5$ , while  $(m+1)(m'+1) \leq (n+1)(s+1)$ . Hence

$$\mathbb{P}_k(N_1 = x) \geq \frac{c_0^4 |I_x|}{(m+1)(m'+1)} \geq \frac{c_0^4}{5(n+1)} \geq \frac{c_0^4}{10n},$$

which proves the claim.  $\square$

**Lemma 5.** *Let*

$$d_N(t) := \max_{0 \leq k \leq n} \|\mathbb{P}_k(N_t \in \cdot) - \text{Unif}\{0, \dots, n\}\|_{\text{TV}}.$$

*Then*

$$d_N(t) \leq e^{-c_{\text{mag}} t},$$

*for a universal  $c_{\text{mag}} > 0$ .*

*Proof.* Let

$$\bar{d}(1) = \max_{0 \leq k, k' \leq n} \left( 1 - \sum_{z=0}^n \mathbb{P}_k(N_1 = z) \wedge \mathbb{P}_{k'}(N_1 = z) \right).$$

For all sufficiently large  $n$ , the interval  $\mathbb{Z} \cap [n/10, 2n/5]$  contains at least  $n/5$  points, so Lemma 4 gives  $\bar{d}(1) \leq 1 - \delta/5$ . For each of the finitely many

remaining  $n$ , every one-step probability is positive. Taking the minimum of their overlaps and  $\delta/5$  gives a universal  $\delta' > 0$  such that  $\bar{d}(1) \leq 1 - \delta'$ . Dobrushin contraction, or Lemmas 4.10 and 4.11 of [18], now yields

$$d_N(t) \leq (1 - \delta')^t = e^{-c_{\text{mag}} t}, \quad c_{\text{mag}} := -\log(1 - \delta') > 0.$$

□

**Corollary 6.** *Let  $B \subseteq \{0, \dots, n\}$ , let  $0 \leq t_0 \leq t$  be integers, and let  $c_{\text{mag}}$  be as in Lemma 5. Uniformly over the initial state,*

$$\mathbb{P}(\exists j \in \{t_0, \dots, t\} : N_j \in B) \leq \frac{(t - t_0 + 1)|B|}{n + 1} + e^{-c_{\text{mag}} t_0} \frac{e^{c_{\text{mag}}}}{e^{c_{\text{mag}}} - 1}.$$

*Proof.* By the union bound,

$$\begin{aligned} \mathbb{P}(\exists j \in \{t_0, \dots, t\} : N_j \in B) &\leq \sum_{j=t_0}^t \mathbb{P}(N_j \in B) \\ &\leq \sum_{j=t_0}^t \left( \frac{|B|}{n + 1} + d_N(j) \right) \\ &\leq \frac{(t - t_0 + 1)|B|}{n + 1} + \sum_{j=t_0}^t e^{-c_{\text{mag}} j}. \end{aligned}$$

Evaluating the geometric series completes the proof.

□

**Corollary 7.** *For every  $\eta \in (0, 1/2)$  and every  $\varepsilon > 0$ , there exists an integer  $r_{\eta, \varepsilon} \geq 1$ , independent of  $n$ , such that from every starting state,*

$$\mathbb{P}(N_{r_{\eta, \varepsilon}} \in [\eta n, (1 - \eta)n]) \geq 1 - 2\eta - \varepsilon$$

*for all sufficiently large  $n$ .*

*Proof.* Under the stationary law of the magnetization chain,  $N$  is uniform on  $\{0, \dots, n\}$ , and the mass outside  $[\eta n, (1 - \eta)n]$  is at most  $2\eta + O(n^{-1})$ . Choose  $r_{\eta, \varepsilon}$  so large that  $e^{-c_{\text{mag}} r_{\eta, \varepsilon}} \leq \varepsilon/2$ . Lemma 5, followed by taking  $n$  large enough, gives the claim. □

On this event, restart from  $X_{r_{\eta, \varepsilon}}$  and use it as the new reference configuration. Then  $M_0 = N_0$ , and

$$K_0 := N_0 \left( 1 - \frac{N_0}{n} \right) \geq \eta(1 - \eta)n.$$

### 3 Dependence on the Initial Configuration

This section isolates the chain's dependence on a fixed reference configuration. We reduce the problem to a two-dimensional projection and derive the centered-overlap recursion and its hypergeometric fluctuations.

Fix a reference configuration  $x_0 \in X_n$  and set  $N_0 := \sum_{i=1}^n x_0(i)$ . All quantities in this section are defined relative to  $x_0$ .

#### 3.1 Reduction to a two-dimensional chain

This subsection shows that the two overlap counts  $(M_t, S_t)$  retain exactly the information needed to measure convergence to stationarity.

For  $x \in X_n$ , define

$$M(x) := \sum_{i=1}^n x(i)x_0(i), \quad S(x) := \sum_{i=1}^n (1-x(i))(1-x_0(i)).$$

Thus  $M$  and  $S$  count common ones and common zeros, respectively. Set

$$M_t := M(X_t), \quad S_t := S(X_t).$$

Since the number of ones of  $X_t$  on the zero coordinates of  $x_0$  is  $n - N_0 - S_t$ ,

$$N_t = M_t + n - N_0 - S_t.$$

**Lemma 8.** *Let  $(X_t)$  start from  $x_0$ , and let  $\pi'_n$  be the law of  $(M, S)$  under  $\pi_n$ . Then  $(M_t, S_t)$  is Markov and*

$$\|P_n^t(x_0, \cdot) - \pi_n\|_{\text{TV}} = \|\mathbb{P}_{x_0}((M_t, S_t) \in \cdot) - \pi'_n\|_{\text{TV}}.$$

*Proof.* The same symmetry argument used below shows that the projected process  $(M_t, S_t)$  is Markov: two configurations with the same values of  $(M, S)$  differ by an element of the subgroup preserving  $x_0$ , and the Burnside transition kernel is equivariant under this subgroup.

Since  $(M_t, S_t)$  is a projection of  $X_t$ ,

$$\|P_n^t(x_0, \cdot) - \pi_n\|_{\text{TV}} \geq \|\mathbb{P}_{x_0}((M_t, S_t) \in \cdot) - \pi'_n\|_{\text{TV}}.$$

For the reverse inequality, define

$$X_{m,s} := \{x \in X_n : (M(x), S(x)) = (m, s)\}.$$

Let  $H$  be the subgroup of coordinate permutations preserving the sets  $\{i : x_0(i) = 1\}$  and  $\{i : x_0(i) = 0\}$ . The Burnside transition kernel is equivariant under coordinate permutations, and every element of  $H$  fixes  $x_0$ . Hence the law of  $X_t$ , started from  $x_0$ , is  $H$ -invariant. The group  $H$  acts transitively on each nonempty  $X_{m,s}$ , so conditional on  $\{(M_t, S_t) = (m, s)\}$ , the law of  $X_t$  is uniform on  $X_{m,s}$ .

The same is true for  $\pi_n$  conditioned on  $\{(M, S) = (m, s)\}$ , because  $\pi_n(x)$  depends only on  $|x|$ , which is constant on  $X_{m,s}$ . Moreover,

$$|X_{m,s}| = \binom{N_0}{m} \binom{n - N_0}{s}.$$

Therefore, for each  $x \in X_n$ ,

$$\begin{aligned} |P_n^t(x_0, x) - \pi_n(x)| &= \left| \sum_{\substack{m,s: \\ X_{m,s} \neq \emptyset}} \frac{\mathbf{1}_{X_{m,s}}(x)}{\binom{N_0}{m} \binom{n - N_0}{s}} (\mathbb{P}_{x_0}((M_t, S_t) = (m, s)) - \pi'_n(m, s)) \right| \\ &\leq \sum_{\substack{m,s: \\ X_{m,s} \neq \emptyset}} \frac{\mathbf{1}_{X_{m,s}}(x)}{\binom{N_0}{m} \binom{n - N_0}{s}} |\mathbb{P}_{x_0}((M_t, S_t) = (m, s)) - \pi'_n(m, s)|. \end{aligned}$$

Summing over  $x \in X_n$  gives the reverse inequality and completes the proof.  $\square$

### 3.2 A centered overlap variable

This subsection introduces the centered overlap  $K_t$  and derives its exact one-step conditional drift.

**Definition 9.**

$$K_t := M_t - \frac{N_t N_0}{n}.$$

*This measures the deviation of the overlap with  $x_0$  from its natural mean under uniform sampling with fixed magnetization.*

For the step from  $t$  to  $t + 1$ , let

$$E_t^{(1)} := \{i : X_t(i) = X_{t+1}(i) = 1\}, \quad E_t^{(0)} := \{i : X_t(i) = 0, X_{t+1}(i) = 1\},$$

and define the selected totals

$$Q_{t+1} := |E_t^{(1)}|, \quad \Lambda_{t+1} := |E_t^{(0)}|.$$

Then  $N_{t+1} = Q_{t+1} + \Lambda_{t+1}$ . Define

$$\mathcal{F}_t := \sigma(X_0, \dots, X_t), \quad \mathcal{G}_t := \sigma(\mathcal{F}_t, Q_{t+1}, \Lambda_{t+1}).$$

Thus  $\mathcal{G}_t$  reveals the history through time  $t$  and the two selected totals. We set  $Q_{t+1}/N_t = 0$  when  $N_t = 0$ , and  $\Lambda_{t+1}/(n - N_t) = 0$  when  $N_t = n$ .

**Lemma 10.** *Conditional on  $\mathcal{G}_t$ , the sets  $E_t^{(1)}$  and  $E_t^{(0)}$  are independent and uniform among, respectively, the  $Q_{t+1}$ -subsets of  $\{i : X_t(i) = 1\}$  and the  $\Lambda_{t+1}$ -subsets of  $\{i : X_t(i) = 0\}$ . Consequently,*

$$\mathbb{E} \left[ \sum_{i \in E_t^{(1)}} x_0(i) \middle| \mathcal{G}_t \right] = \frac{M_t Q_{t+1}}{N_t}, \quad \mathbb{E} \left[ \sum_{i \in E_t^{(0)}} x_0(i) \middle| \mathcal{G}_t \right] = \frac{(N_0 - M_t) \Lambda_{t+1}}{n - N_t},$$

with the boundary convention stated above.

*Proof.* Conditional on  $\mathcal{F}_t$ , the two level sets use independent randomness, and each selected set has a law invariant under every permutation of its level set. Conditioning on the two cardinalities preserves independence and invariance. Since the symmetric group acts transitively on subsets of a fixed size, the two conditional laws are uniform. Each site in the first level set is therefore selected with probability  $Q_{t+1}/N_t$ , and exactly  $M_t$  of those sites satisfy  $x_0(i) = 1$ . This gives the first expectation; the second is identical. The boundary cases follow from the stated convention.  $\square$

**Corollary 11.** *We have*

$$\mathbb{E}[K_{t+1} \mid \mathcal{G}_t] = K_t \left( \frac{Q_{t+1}}{N_t} - \frac{\Lambda_{t+1}}{n - N_t} \right).$$

*Proof.* Since  $N_{t+1} = Q_{t+1} + \Lambda_{t+1}$ ,

$$K_{t+1} = \sum_{i \in E_t^{(1)}} x_0(i) + \sum_{i \in E_t^{(0)}} x_0(i) - \frac{(Q_{t+1} + \Lambda_{t+1})N_0}{n}.$$

Taking conditional expectation and applying Lemma 10 gives

$$\mathbb{E}[K_{t+1} \mid \mathcal{G}_t] = \frac{M_t Q_{t+1}}{N_t} + \frac{(N_0 - M_t) \Lambda_{t+1}}{n - N_t} - \frac{(Q_{t+1} + \Lambda_{t+1})N_0}{n}.$$

Substituting  $M_t = K_t + N_t N_0/n$  gives the result, including the boundary cases under our convention.  $\square$

### 3.3 A hypergeometric decomposition

This subsection describes the one-step fluctuations of the overlap as a sum of two independent hypergeometric variables.

Write  $\text{Hyp}(N, m, r)$  for the law of the number of marked items in a uniform  $r$ -subset of a population of size  $N$  containing  $m$  marked items. Decompose the next overlap according to the two current level sets:

$$M_{t+1}^{(1)} := \sum_{i=1}^n X_{t+1}(i)X_t(i)x_0(i), \quad M_{t+1}^{(0)} := \sum_{i=1}^n X_{t+1}(i)(1 - X_t(i))x_0(i).$$

Then

$$M_{t+1} = M_{t+1}^{(1)} + M_{t+1}^{(0)}.$$

Conditional on  $\mathcal{G}_t$ , Lemma 10 gives independent laws

$$M_{t+1}^{(1)} \mid \mathcal{G}_t \sim \text{Hyp}(N_t, M_t, Q_{t+1}), \quad M_{t+1}^{(0)} \mid \mathcal{G}_t \sim \text{Hyp}(n - N_t, N_0 - M_t, \Lambda_{t+1}).$$

Empty populations give the point mass at zero. In particular,

$$\mathbb{E}[M_{t+1}^{(1)} \mid \mathcal{G}_t] = \frac{M_t Q_{t+1}}{N_t}, \quad \mathbb{E}[M_{t+1}^{(0)} \mid \mathcal{G}_t] = \frac{(N_0 - M_t) \Lambda_{t+1}}{n - N_t}.$$

With the same boundary convention, the Hoeffding–Serfling inequality and the fact that both sample sizes are at most  $n$  give, for  $b \in \{0, 1\}$  and  $a > 0$ ,

$$\mathbb{P}\left(\left|M_{t+1}^{(b)} - \mathbb{E}[M_{t+1}^{(b)} \mid \mathcal{G}_t]\right| \geq a\sqrt{n} \mid \mathcal{G}_t\right) \leq 2e^{-2a^2}.$$

**Corollary 12.** *For every  $a > 0$ ,*

$$\mathbb{P}\left(\left|K_{t+1} - \mathbb{E}[K_{t+1} \mid \mathcal{G}_t]\right| \geq 2a\sqrt{n} \mid \mathcal{G}_t\right) \leq 4e^{-2a^2}.$$

*Proof.* The term  $N_{t+1}N_0/n$  is  $\mathcal{G}_t$ -measurable. Thus the centered fluctuation of  $K_{t+1}$  is the sum of the two centered fluctuations above, and the result follows by a union bound.  $\square$

## 4 Studying the Conditional Expectation

This section compares the difference of selected fractions in Corollary 11 with the stick-breaking factor  $R - R'$  from Definition 3. We first establish estimates for the continuous model, then couple it to the discrete factors and transfer the resulting logarithmic barriers.

## 4.1 The continuous stick-breaking model

This subsection records tail and small-ball estimates for the limiting stick-breaking variables used in the discrete comparison. We use the variables  $(\xi_j)$  and  $R$  from Definition 3. Then  $\sum_{j \geq 1} \xi_j = 1$  almost surely.

**Lemma 13.** *For every  $j \geq 1$ ,*

1.  $\mathbb{P}(\xi_j < 2^{-10j}) \leq 2^{-4j}$ ,
2.  $\mathbb{P}(\xi_j > e^{-j/2}) \leq e^{-\gamma j}$ , where  $\gamma := \log(3/2) - 1/4 > 0$ ,
3.  $\mathbb{E}[\xi_j] = 2^{-j}$ .

*Proof.* Because  $U_j, 1 - U_1, \dots, 1 - U_{j-1}$  are independent uniforms,

$$\xi_j = U_j \prod_{i=1}^{j-1} (1 - U_i) \stackrel{d}{=} \prod_{i=1}^j U_i,$$

so

$$-\log \xi_j \stackrel{d}{=} S_j := \sum_{i=1}^j \log(1/U_i).$$

The summands in  $S_j$  are i.i.d. exponential random variables of mean one. Since  $\mathbb{E}[e^{S_j/2}] = 2^j$  and  $\mathbb{E}[e^{-S_j/2}] = (2/3)^j$ , Markov's inequality gives

$$\mathbb{P}(\xi_j < 2^{-10j}) \leq 2^{-5j} \mathbb{E}[e^{S_j/2}] = 2^{-4j},$$

and

$$\mathbb{P}(\xi_j > e^{-j/2}) \leq e^{j/4} \mathbb{E}[e^{-S_j/2}] = e^{-\gamma j}.$$

Finally, independence gives

$$\mathbb{E}[\xi_j] = \prod_{i=1}^j \mathbb{E}[U_i] = 2^{-j}.$$

□

**Lemma 14.** *Let  $m = \lfloor (\log N)^2 \rfloor$  for  $N \geq 2$ . Then*

$$\mathbb{P}\left(\sum_{j=1}^m \xi_j \leq 1 - \frac{1}{N}\right) \leq N 2^{-m}.$$

*In particular,*

$$\mathbb{P}\left(\sum_{j=1}^{\lfloor (\log N)^2 \rfloor} \xi_j \leq 1 - \frac{1}{N}\right) \leq 2N \exp(-(\log 2)(\log N)^2).$$

*Proof.* For  $T_m := \sum_{j>m} \xi_j$ , Lemma 13 gives

$$\mathbb{E}[T_m] = \sum_{j>m} \mathbb{E}[\xi_j] = 2^{-m},$$

so Markov's inequality gives  $\mathbb{P}(T_m \geq N^{-1}) \leq N 2^{-m}$ . The second bound follows from  $2^{-\lfloor y \rfloor} \leq 2e^{-(\log 2)y}$ .  $\square$

**Lemma 15.** *There exists a universal constant  $C > 0$  such that*

$$\mathbb{P}(R < x) \leq Cx^{1/10}, \quad 0 < x \leq 1.$$

*Proof.* By the arcsine identification in Subsection 1.2,

$$R \sim \text{Beta}(1/2, 1/2), \quad f_R(r) = \frac{1}{\pi \sqrt{r(1-r)}} \mathbf{1}_{(0,1)}(r).$$

For  $0 < x \leq 1/2$ ,

$$\mathbb{P}(R < x) \leq C \int_0^x r^{-1/2} dr \leq Cx^{1/2} \leq Cx^{1/10}.$$

The remaining values of  $x$  are covered by increasing  $C$ .  $\square$

**Lemma 16.** *Let  $R'$  be an independent copy of  $R$ . Then there exists a constant  $C > 0$  such that*

$$\mathbb{P}(|R - R'| < x) \leq Cx^{1/20}, \quad 0 < x \leq 1.$$

*In particular,*

$$\mathbb{E}[|\log |R - R'||^2] < \infty.$$

*Proof.* It is enough to consider  $0 < x \leq 1/4$ , since the bound is trivial for larger  $x$ . Set  $\delta = x^{1/2}$ . By Lemma 15 and the symmetry  $R \stackrel{d}{=} 1 - R$ ,

$$\mathbb{P}(R \notin [\delta, 1 - \delta]) \leq C\delta^{1/10} = Cx^{1/20}.$$

On the event  $R \in [\delta, 1 - \delta]$ , the interval  $[R - x, R + x] \cap [0, 1]$  lies in  $[\delta/2, 1 - \delta/2]$ . The arcsine density is bounded by  $C\delta^{-1/2}$  on this interval, and therefore

$$\mathbb{P}(|R - R'| < x \mid R) \leq Cx\delta^{-1/2} = Cx^{3/4}.$$

Combining the two estimates gives

$$\mathbb{P}(|R - R'| < x) \leq Cx^{1/20}.$$

Thus  $\mathbb{P}(-\log |R - R'| > u) \leq Ce^{-u/20}$ ; integrating this tail proves the final assertion.  $\square$

## 4.2 Products in the continuous model

This subsection establishes laws of large numbers and barrier estimates for products of the limiting contraction factors. Let  $(R_j, R'_j)_{j \geq 0}$  be i.i.d. copies of  $(R, R')$ , and set

$$Y_j := \log |R_j - R'_j|, \quad \tilde{L}_{i,t} := Y_{t-i}, \quad 0 \leq i \leq t.$$

**Lemma 17.** *We have*

$$\frac{1}{t} \sum_{j=0}^{t-1} \log |R_j - R'_j| \longrightarrow -2 \log 2 \quad \text{almost surely as } t \rightarrow \infty.$$

*Proof.* Lemma 16 gives integrability, while Subsection 1.2 gives the mean. The statement follows from the strong law of large numbers.  $\square$

For  $a \in (0, 2 \log 2)$ , define

$$M_a := \sup_{k \geq 0} \left( ak + \sum_{i=0}^k Y_i \right).$$

For  $w > 0$ , also define

$$\mathcal{E}_{t,w}(a) := \bigcap_{k=0}^t \left\{ \sum_{i=0}^k \tilde{L}_{i,t} \leq \log w - ak \right\}.$$

**Lemma 18.** Fix  $a \in (0, 2 \log 2)$ . Then  $M_a < \infty$  almost surely. Moreover, for every  $\varepsilon > 0$ , there exists  $w_0 = w_0(a, \varepsilon)$  such that

$$\mathbb{P}(\mathcal{E}_{t,w}(a)) \geq 1 - \varepsilon \quad \text{for all } t \geq 1 \text{ and } w \geq w_0.$$

*Proof.* By Lemma 17,

$$\frac{1}{k} \left( ak + \sum_{i=0}^k Y_i \right) \longrightarrow a - 2 \log 2 < 0 \quad \text{almost surely.}$$

Consequently,  $ak + \sum_{i=0}^k Y_i \rightarrow -\infty$  almost surely. In particular, the sequence is bounded above, so  $M_a < \infty$  almost surely.

The reversed vector  $(Y_t, \dots, Y_0)$  has the same law as  $(Y_0, \dots, Y_t)$ . Hence, for every  $t$ ,

$$\mathbb{P}(\mathcal{E}_{t,w}(a)) \geq \mathbb{P}(M_a \leq \log w).$$

Since  $M_a < \infty$  almost surely, the right-hand side is at least  $1 - \varepsilon$  once  $w$  is sufficiently large.  $\square$

For the rest of the paper, set

$$m_n := \lceil \log \log n \rceil.$$

For  $a \in (0, 2 \log 2)$ ,  $t \geq m_n$ , and  $w \geq 1$ , define

$$\mathcal{E}_{t,w}^-(a) := \bigcap_{k=m_n}^t \left\{ \sum_{i=0}^k \tilde{L}_{i,t} \leq \log w - ak \right\}.$$

**Corollary 19.** Fix  $a \in (0, 2 \log 2)$ . Uniformly for  $t \geq m_n$  and  $w \geq 1$ ,

$$\mathbb{P}(\mathcal{E}_{t,w}^-(a)) = 1 - o(1).$$

*Proof.* Define

$$T_a := \inf \left\{ m \geq 0 : \sum_{i=0}^k Y_i \leq -ak \text{ for every } k \geq m \right\}.$$

By the strong law,

$$ak + \sum_{i=0}^k Y_i \longrightarrow -\infty \quad \text{almost surely,}$$

so  $T_a < \infty$  almost surely. For every  $t \geq m_n$ , reversal invariance of  $(Y_0, \dots, Y_t)$  gives

$$\begin{aligned} \mathbb{P}(\mathcal{E}_{t,1}^-(a)) &= \mathbb{P}\left(\sum_{i=0}^k Y_i \leq -ak \text{ for every } m_n \leq k \leq t\right) \\ &\geq \mathbb{P}(T_a \leq m_n). \end{aligned}$$

Since  $m_n \rightarrow \infty$  and  $T_a < \infty$  almost surely, the last probability tends to one. This lower bound is independent of  $t$ , so the convergence is uniform for  $t \geq m_n$ . Finally,  $\mathcal{E}_{t,w}^-(a)$  is increasing in  $w$ , which gives the asserted uniformity for all  $w \geq 1$ .  $\square$

### 4.3 The discrete model

This subsection couples the selected fractions in the finite chain to their continuous stick-breaking limits.

We work on  $1 \leq N_t \leq n-1$ ; later, the magnetization estimates control the excluded boundary events. By Corollary 11,

$$\mathbb{E}[K_{t+1} \mid \mathcal{G}_t] = K_t(Z_t - Z'_t),$$

where

$$Z_t := \frac{Q_{t+1}}{N_t} = \sum_{j \geq 1} \zeta_j^t \frac{|C_j^t|}{N_t}, \quad \text{and} \quad Z'_t := \frac{\Lambda_{t+1}}{n - N_t} = \sum_{j \geq 1} (\zeta_j^t)' \frac{|(C_j^t)'|}{n - N_t},$$

with the same boundary convention as above. Here the cycles  $C_j^t$  are the cycles of the permutation acting on the 1-coordinates at time  $t$ , and  $(C_j^t)'$  are the cycles of the 0-coordinates;  $\zeta_j^t$  and  $(\zeta_j^t)'$  are their respective labels. Define the corresponding linear drift term by

$$\Gamma_t := K_0 \prod_{j=0}^{t-1} (Z_j - Z'_j). \tag{1}$$

Conditional on  $N_t = N$ , order the cycles by repeatedly exposing the cycle containing the least unused coordinate. Let

$$D_{t,0} := N, \quad D_{t,\ell} := D_{t,\ell-1} - U_{t,\ell},$$

where, given  $D_{t,\ell-1}$ , the random variable  $U_{t,\ell}$  is uniform on  $\{1, \dots, D_{t,\ell-1}\}$  if  $D_{t,\ell-1} > 0$ , and  $U_{t,\ell} = 0$  if  $D_{t,\ell-1} = 0$ . Set

$$V_{t,\ell} := \frac{U_{t,\ell}}{D_{t,\ell-1}} \quad \text{if } D_{t,\ell-1} > 0, \quad V_{t,\ell} := 0 \quad \text{if } D_{t,\ell-1} = 0.$$

Then

$$D_{t,\ell} = D_{t,\ell-1}(1 - V_{t,\ell}),$$

and

$$|C_\ell^t| = U_{t,\ell}, \quad \frac{|C_\ell^t|}{N_t} = \frac{D_{t,\ell-1}V_{t,\ell}}{N_t}.$$

**Lemma 20.** *For every  $\ell \geq 0$ ,*

$$\mathbb{E}[D_{t,\ell} \mid N_t] \leq 2^{-\ell} N_t.$$

*Proof.* The claim is immediate if  $D_{t,\ell-1} = 0$ . Otherwise,

$$\mathbb{E}[U_{t,\ell} \mid D_{t,\ell-1}] = \frac{D_{t,\ell-1} + 1}{2}.$$

Hence

$$\mathbb{E}[D_{t,\ell} \mid D_{t,\ell-1}] = D_{t,\ell-1} - \frac{D_{t,\ell-1} + 1}{2} \leq \frac{D_{t,\ell-1}}{2}.$$

Iteration proves the claim. □

For  $m \geq 1$  and  $N_t > 0$ , define

$$Z_t^{(m)} := \frac{1}{N_t} \sum_{\ell=1}^m \zeta_\ell^t D_{t,\ell-1} V_{t,\ell}.$$

**Lemma 21.** *Let  $m_N = \max\{1, \lfloor (\log N)^2 \rfloor\}$ . Then, conditional on  $N_t = N$ ,*

$$\mathbb{P}\left(Z_t - Z_t^{(m_N)} \geq \frac{1}{N} \mid N_t = N\right) \leq N 2^{-m_N}.$$

*Proof.* After  $m_N$  cycles,  $D_{t,m_N}$  coordinates remain. Hence

$$0 \leq Z_t - Z_t^{(m_N)} \leq \frac{D_{t,m_N}}{N}.$$

The claim follows from Markov's inequality and Lemma 20. □

We next couple the discrete stick-breaking variables with their continuous analogues.

Let  $(W_{t,\ell})_{\ell \geq 1}$  be i.i.d. uniform random variables on  $[0, 1]$ , coupled so that

$$V_{t,\ell} = \frac{\lceil W_{t,\ell} D_{t,\ell-1} \rceil}{D_{t,\ell-1}} \quad \text{if } D_{t,\ell-1} > 0, \quad V_{t,\ell} = 0 \quad \text{if } D_{t,\ell-1} = 0.$$

When  $D_{t,\ell-1} > 0$ ,

$$W_{t,\ell} \leq V_{t,\ell} \leq W_{t,\ell} + \frac{1}{D_{t,\ell-1}}. \quad (2)$$

Define

$$\tilde{D}_{t,\ell} := N_t \prod_{r=1}^{\ell} (1 - W_{t,r}), \quad \tilde{Z}_t^{(m)} := \frac{1}{N_t} \sum_{\ell=1}^m \zeta_{\ell}^t \tilde{D}_{t,\ell-1} W_{t,\ell}.$$

**Lemma 22.** *For every  $\ell \geq 0$ ,*

$$\tilde{D}_{t,\ell} - \ell \leq D_{t,\ell} \leq \tilde{D}_{t,\ell}.$$

*Consequently,*

$$\left| \tilde{D}_{t,\ell-1} W_{t,\ell} - D_{t,\ell-1} V_{t,\ell} \right| \leq \ell,$$

*and therefore for every  $m \geq 1$ ,*

$$|Z_t^{(m)} - \tilde{Z}_t^{(m)}| \leq \frac{1}{N_t} \sum_{\ell=1}^m \ell \leq \frac{m(m+1)}{2N_t}.$$

*In particular, for  $m = \lfloor (\log n)^2 \rfloor$ ,*

$$|Z_t^{(m)} - \tilde{Z}_t^{(m)}| \leq \frac{\log^4 n}{N_t}.$$

*Proof.* Almost surely,

$$D_{t,\ell} = \lfloor (1 - W_{t,\ell}) D_{t,\ell-1} \rfloor.$$

The first bounds follow by induction, using  $x - 1 \leq \lfloor x \rfloor \leq x$ . If  $D_{t,\ell-1} = 0$ , the induction hypothesis gives  $\tilde{D}_{t,\ell-1} \leq \ell - 1$ , which also proves the second bound below. Otherwise, (2) and the first bounds give

$$\begin{aligned} \left| \tilde{D}_{t,\ell-1} W_{t,\ell} - D_{t,\ell-1} V_{t,\ell} \right| &\leq W_{t,\ell} |\tilde{D}_{t,\ell-1} - D_{t,\ell-1}| + D_{t,\ell-1} |W_{t,\ell} - V_{t,\ell}| \\ &\leq (\ell - 1) + 1 = \ell. \end{aligned}$$

Summing proves the remaining assertions.  $\square$

Realize the variables  $R_t$  from Subsection 4.2 in this coupling by

$$R_t := \sum_{\ell=1}^{\infty} \zeta_{\ell}^t W_{t,\ell} \prod_{r < \ell} (1 - W_{t,r}), \quad R_t^{(m)} := \sum_{\ell=1}^m \zeta_{\ell}^t W_{t,\ell} \prod_{r < \ell} (1 - W_{t,r}).$$

Extend each label sequence by fresh fair coins and take all  $W$ -arrays, label sequences, and primed copies independent across times and level sets. Then  $(R_t, R'_t)$  are i.i.d. copies of  $(R, R')$ , and  $R_t^{(m)}$  has the same law as  $\sum_{\ell=1}^m \eta_{\ell} \xi_{\ell}$ .

For  $M \geq 1$ , set

$$\epsilon_M := \frac{\log^4 M + 2}{M}.$$

**Lemma 23.** *There exists a universal constant  $C > 0$  such that the following holds. Let  $m_N = \max\{1, \lfloor (\log N)^2 \rfloor\}$ . Conditional on  $N_t = N$ ,*

$$\mathbb{P}(|Z_t - R_t| > \epsilon_N \mid N_t = N) \leq CN \exp(-(\log 2)(\log N)^2).$$

*Proof.* Let  $m = m_N$ . Except on an event of probability at most  $2N2^{-m}$ , Lemmas 21 and 14 give

$$Z_t - Z_t^{(m)} \leq N^{-1}, \quad R_t - R_t^{(m)} \leq N^{-1}.$$

Also  $\tilde{Z}_t^{(m)} = R_t^{(m)}$ , and Lemma 22 gives

$$|Z_t^{(m)} - R_t^{(m)}| \leq \frac{m(m+1)}{2N}.$$

For  $N \geq 3$ , the sum of these three errors is at most  $\epsilon_N$ . Since  $2^{-m} \leq 2 \exp\{-(\log 2)(\log N)^2\}$ , the result follows; the cases  $N = 1, 2$  are absorbed by increasing  $C$ .  $\square$

We now deduce a small-ball bound for the discrete factors.

**Lemma 24.** *There exists a universal constant  $C > 0$  such that for every  $N \in \{1, \dots, n-1\}$  and every  $x > 0$ ,*

$$\begin{aligned} \mathbb{P}(|Z_t - Z'_t| < x \mid N_t = N) &\leq C \left( x + \frac{\log^4 N + 2}{N} + \frac{\log^4(n - N) + 2}{n - N} \right)^{1/20} \\ &\quad + CN \exp(-(\log 2)(\log N)^2) \\ &\quad + C(n - N) \exp(-(\log 2)(\log(n - N))^2). \end{aligned}$$

*Proof.* Use Lemma 23 to couple  $Z_t, Z'_t$  to the independent variables  $R_t, R'_t$ . Outside the two coupling failures,

$$|Z_t - Z'_t| < x \implies |R_t - R'_t| < x + \epsilon_N + \epsilon_{n-N}.$$

Lemmas 16 and 23 now give the claim, with the trivial bound used if the argument of the small-ball estimate exceeds one.  $\square$

#### 4.4 Comparison of logarithmic sums

This subsection transfers the continuous logarithmic estimates to the discrete contraction factors over the time intervals used later.

**Definition 25.** For  $0 \leq j \leq t$ , let

$$\mathcal{H}_j := \{\sqrt{n} \leq N_j \leq n - \sqrt{n}\}, \quad \mathcal{H}_{t,i} := \bigcap_{j=i}^t \mathcal{H}_j.$$

Also set

$$L_{i,t} := \log |Z_{t-i} - Z'_{t-i}|, \quad 0 \leq i \leq t,$$

with  $\log 0 = -\infty$ .

On  $\mathcal{H}_{t,i}$ , both denominators in Lemma 24 are at least  $\sqrt{n}$ .

**Lemma 26.** Let  $\log \log n \leq i < t \leq (\log n)^2$ . Define

$$\Delta_{i,t} := \max_{0 \leq k \leq t-i} \left| \sum_{r=0}^k (L_{r,t} - \tilde{L}_{r,t}) \right|.$$

If a summand is infinite, set  $\Delta_{i,t} = +\infty$ . Then for all sufficiently large  $n$ ,

$$\begin{aligned} \mathbb{P}(\Delta_{i,t} > 3(\log n)^{-2}) &\leq 3C(t-i+1) \left( n^{-1/5} + \frac{4 \log^4 n}{\sqrt{n}} \right)^{1/20} \\ &\quad + Cn(t-i+1) \exp\left(-\frac{\log 2}{4}(\log n)^2\right) \\ &\quad + \frac{3(t-i+1)}{\sqrt{n}} + e^{-c_{\text{mag}} i} \frac{e^{c_{\text{mag}}}}{e^{c_{\text{mag}}} - 1}. \end{aligned}$$

In particular, if  $\log \log n \leq i < t \leq (\log n)^2$ , then

$$\Delta_{i,t} = o_{\mathbb{P}}(1).$$

*Proof.* For each  $j \in [i, t]$ , on  $\mathcal{H}_j$  both level-set sizes lie in  $[\sqrt{n}, n]$ . In particular,  $\epsilon_{N_j} \vee \epsilon_{n-N_j} \leq 2 \log^4 n / \sqrt{n}$  for large  $n$ . Thus Lemma 24 gives, for every  $u > 0$ ,

$$\mathbb{P}(|Z_j - Z'_j| < u, \mathcal{H}_j) \leq 3C \left( u + \frac{4 \log^4 n}{\sqrt{n}} \right)^{1/20} + Cn \exp\left(-\frac{\log 2}{4}(\log n)^2\right).$$

Define

$$\mathcal{C}_{t,i} := \bigcap_{j=i}^t \{|Z_j - Z'_j| > n^{-1/5}\}.$$

Let

$$c_n := \frac{2 \log^4 n}{\sqrt{n}}$$

and define

$$\mathcal{B}_{t,i} := \bigcap_{j=i}^t \{|Z_j - R_j| \leq c_n, |Z'_j - R'_j| \leq c_n\}.$$

The preceding estimate, Lemma 23, and a union bound control  $\mathcal{C}_{t,i}^c \cap \mathcal{H}_{t,i}$  and  $\mathcal{B}_{t,i}^c \cap \mathcal{H}_{t,i}$  by the first two terms in the stated bound. Corollary 6, applied to the two boundary intervals of total size at most  $3\sqrt{n}$ , controls  $\mathcal{H}_{t,i}^c$  by the last two terms.

On  $\mathcal{H}_{t,i} \cap \mathcal{C}_{t,i} \cap \mathcal{B}_{t,i}$ , for every  $j \in [i, t]$ ,

$$||R_j - R'_j| - |Z_j - Z'_j|| \leq 2c_n,$$

and since  $|Z_j - Z'_j| > n^{-1/5}$ ,

$$\left| \frac{|R_j - R'_j|}{|Z_j - Z'_j|} - 1 \right| \leq 4n^{-3/10} \log^4 n.$$

For large  $n$ , this is at most  $(\log n)^{-4}$ , so

$$\left| \log \frac{|R_j - R'_j|}{|Z_j - Z'_j|} \right| \leq 2(\log n)^{-4}.$$

Consequently,

$$\Delta_{i,t} \leq 2(t - i + 1)(\log n)^{-4}.$$

If  $t - i + 1 \leq (\log n)^2$ , then

$$\Delta_{i,t} \leq 2(\log n)^{-2}.$$

Combining this deterministic estimate with the three exceptional-event bounds proves the display. Its right-hand side tends to zero uniformly over the stated range of  $i, t$ , proving the final assertion.  $\square$

For the limit profile computations, we also need a forward version of this comparison after the initial burn-in.

**Lemma 27.** *Fix  $C > 0$ , and let*

$$\theta_n := \lfloor C \log n \rfloor.$$

*Then, uniformly over the initial state,*

$$\max_{m_n \leq s \leq \theta_n} \left| \sum_{j=m_n}^{s-1} \log |Z_j - Z'_j| - \sum_{j=m_n}^{s-1} \log |R_j - R'_j| \right| = o_{\mathbb{P}}(1).$$

*Proof.* Define the logarithmic comparison error

$$\text{Err}_j := \log |Z_j - Z'_j| - \log |R_j - R'_j|.$$

For large  $n$ , Lemma 26, applied with  $i = m_n$  and  $t = \theta_n - 1$ , gives  $\Delta_{m_n, \theta_n - 1} = o_{\mathbb{P}}(1)$ . On the event that this quantity is finite, reversing the order of summation yields

$$\max_{m_n \leq q \leq \theta_n - 1} \left| \sum_{j=q}^{\theta_n - 1} \text{Err}_j \right| = o_{\mathbb{P}}(1).$$

For  $m_n \leq s \leq \theta_n$ ,

$$\sum_{j=m_n}^{s-1} \text{Err}_j = \sum_{j=m_n}^{\theta_n - 1} \text{Err}_j - \sum_{j=s}^{\theta_n - 1} \text{Err}_j,$$

with an empty second sum when  $s = \theta_n$ . Therefore

$$\max_{m_n \leq s \leq \theta_n} \left| \sum_{j=m_n}^{s-1} \text{Err}_j \right| \leq 2 \max_{m_n \leq q \leq \theta_n - 1} \left| \sum_{j=q}^{\theta_n - 1} \text{Err}_j \right| = o_{\mathbb{P}}(1),$$

as required.  $\square$

**Definition 28.** For  $a > 0$ ,  $t \geq m_n$ , and  $w \geq 1$ , set

$$\mathcal{D}_{t,w}^-(a) := \left\{ \max_{m_n \leq k \leq t} \left( ak + \sum_{i=0}^k L_{i,t} \right) \leq \log w \right\},$$

$$\mathcal{J}_{t,w}(a) := \left\{ \max_{1 \leq k \leq m_n} \left( ak + \sum_{i=0}^k L_{i,t} \right) \leq \log w \right\}, \quad \mathcal{I}_{t,w}(a) := \mathcal{D}_{t,w}^-(a) \cap \mathcal{J}_{t,w}(a).$$

**Lemma 29.** Fix  $c \in (0, \log 2)$ . If

$$2m_n \leq t \leq (\log n)^2,$$

then, for every  $w \geq 1$  and all sufficiently large  $n$ ,

$$\mathbb{P}(\mathcal{D}_{t,w}^-(c)) = 1 - o(1).$$

*Proof.* Choose  $c_\star$  so that

$$2c < c_\star < 2 \log 2.$$

Set

$$\Delta_t := \max_{0 \leq k \leq t - m_n} \left| \sum_{i=0}^k (L_{i,t} - \tilde{L}_{i,t}) \right|.$$

Let

$$\Omega_n := \mathcal{E}_{t,1}^-(c_\star) \cap \{\Delta_t \leq 3(\log n)^{-2}\}.$$

Corollary 19 and Lemma 26, applied with  $i = m_n$ , give

$$\mathbb{P}(\Omega_n) = 1 - o(1).$$

On  $\Omega_n$ , for  $m_n \leq k \leq t - m_n$ ,

$$\sum_{i=0}^k L_{i,t} \leq \sum_{i=0}^k \tilde{L}_{i,t} + 3(\log n)^{-2} \leq -c_\star k + 3(\log n)^{-2}.$$

For large  $n$ , this is at most  $-ck$ .

If  $t - m_n < k \leq t$ , all logarithmic increments are non-positive because  $|Z_j - Z'_j| \leq 1$ . Hence

$$\sum_{i=0}^k L_{i,t} \leq \sum_{i=0}^{t-m_n} L_{i,t} \leq -c_\star(t - m_n) + 3(\log n)^{-2}.$$

Because  $t \geq 2m_n$  and  $c_\star > 2c$ ,

$$c_\star(t - m_n) - ck \geq (c_\star/2 - c)t \longrightarrow \infty.$$

Thus  $\mathcal{D}_{t,1}^-(c)$  holds on  $\Omega_n$  for large  $n$ . Monotonicity in  $w$  proves the claim.  $\square$

**Corollary 30.** *Fix  $c \in (0, \log 2)$ . Then*

$$\sup_{w \geq 1} \sup_{2m_n \leq t \leq (\log n)^2} \mathbb{P}(\mathcal{D}_{t,w}^-(c)^c) = o(1).$$

*Proof.* Corollary 19 is uniform in  $t$ , and the explicit bound in Lemma 26 is uniform when  $i = m_n$  and  $t \leq (\log n)^2$ . Hence the two exceptional probabilities in the preceding proof are uniform in the displayed range. The deterministic argument there is also uniform, and monotonicity reduces the supremum over  $w$  to  $w = 1$ .  $\square$

For the remainder of the paper, keep  $c \in (0, \log 2)$  fixed.

**Lemma 31.** *Fix  $c \in (0, \log 2)$  and  $\varepsilon > 0$ . Assume that*

$$2m_n \leq t \leq (\log n)^2.$$

*Then there exists  $w_0 = w_0(c, \varepsilon) > 0$  such that, for every  $w \geq w_0$ ,*

$$\mathbb{P}(\mathcal{J}_{t,w}(c)) \geq 1 - \varepsilon$$

*for all sufficiently large  $n$ .*

*Proof.* By Lemma 18,  $M_c < \infty$  almost surely. Choose  $A = A(c, \varepsilon)$  so that

$$\mathbb{P}(M_c \leq A) \geq 1 - \frac{\varepsilon}{2}.$$

By reversal invariance,

$$\mathbb{P}\left(\max_{1 \leq k \leq m_n} \left(ck + \sum_{i=0}^k \tilde{L}_{i,t}\right) \leq A\right) \geq 1 - \frac{\varepsilon}{2}.$$

Lemma 26, applied with  $i = t - m_n$ , gives

$$\Delta_{t-m_n,t} = o_{\mathbb{P}}(1).$$

Thus, for all sufficiently large  $n$ , uniformly over the stated range of  $t$ ,

$$\mathbb{P}(\Delta_{t-m_n, t} \leq 1) \geq 1 - \frac{\varepsilon}{2}.$$

On the intersection of these two events, for every  $1 \leq k \leq m_n$ ,

$$ck + \sum_{i=0}^k L_{i,t} \leq ck + \sum_{i=0}^k \tilde{L}_{i,t} + \left| \sum_{i=0}^k (L_{i,t} - \tilde{L}_{i,t}) \right| \leq A + 1.$$

Taking  $w_0 = e^{A+1}$ , this implies  $\mathcal{J}_{t,w}(c)$  for every  $w \geq w_0$ , and the union bound completes the proof.  $\square$

**Lemma 32.** Fix  $c \in (0, \log 2)$  and  $\varepsilon \in (0, 1/2)$ , and let  $M := \lfloor (\log \log n)^2 \rfloor$ . Assume

$$2m_n \leq t \quad \text{and} \quad t + M \leq (\log n)^2.$$

There exists  $w_0 = w_0(c, \varepsilon) > 0$  such that, for every  $w \geq w_0$ ,

$$\mathbb{P} \left( \frac{1}{M} \sum_{s=t}^{t+M} \mathbf{1}_{\mathcal{J}_{s,w}(c)} \geq 1 - 2\varepsilon \right) = 1 - o(1).$$

*Proof.* Define the continuous analogue

$$\mathcal{J}_{s,w}^{\text{cont}}(c) := \left\{ \max_{1 \leq k \leq m_n} \left( ck + \sum_{i=0}^k \tilde{L}_{i,s} \right) \leq \log w \right\}.$$

Choose  $w_1 = w_1(c, \varepsilon) \geq 1$  so that  $\mathbb{P}(\mathbf{M}_c \leq \log w_1) \geq 1 - \varepsilon/2$ . Reversal invariance gives

$$\mathbb{P}(\mathcal{J}_{s,w_1}^{\text{cont}}(c)) \geq 1 - \frac{\varepsilon}{2}$$

for every  $s$ . If  $w \geq ew_1$ , then  $\mathcal{J}_{s,w/e}^{\text{cont}}(c)$  has probability at least  $1 - \varepsilon/2$ .

These events are independent when  $|s - s'| > m_n$ , because they depend on disjoint sets of the i.i.d. increments  $(Y_j)$ . Hence

$$\text{Var} \left( \sum_{s=t}^{t+M} \mathbf{1}_{\mathcal{J}_{s,w/e}^{\text{cont}}(c)} \right) \leq CMm_n = o(M^2).$$

Chebyshev's inequality gives

$$\mathbb{P} \left( \sum_{s=t}^{t+M} \mathbf{1}_{\mathcal{J}_{s,w/e}^{\text{cont}}(c)} < (1 - \varepsilon) M \right) = o(1).$$

The explicit bound in Lemma 26, summed over  $s \in [t, t + M]$  with  $i = s - m_n$ , gives

$$\mathbb{P} \left( \max_{t \leq s \leq t+M} \Delta_{s-m_n, s} > 1 \right) = o(1).$$

Here  $s \leq t + M \leq (\log n)^2$  and  $M = O((\log \log n)^2)$ . On the complementary event,

$$\mathcal{J}_{s, w/e}^{\text{cont}}(c) \subseteq \mathcal{J}_{s, w}(c) \quad \text{for every } s \in [t, t + M].$$

Consequently, after enlarging  $w_0$  so that  $w_0 \geq ew_1$ ,

$$\mathbb{P} \left( \frac{1}{M} \sum_{s=t}^{t+M} \mathbf{1}_{\mathcal{J}_{s, w}(c)} \geq 1 - 2\varepsilon \right) = 1 - o(1),$$

which proves the claim.  $\square$

The long- and short-barrier estimates give the following consequence.

**Corollary 33.** *Fix  $c \in (0, \log 2)$ , and let  $M := \lfloor (\log \log n)^2 \rfloor$ . There exists  $w_0 = w_0(c) > 0$  such that, for every  $w \geq w_0$  and every integer  $r$  satisfying*

$$2m_n \leq r \quad \text{and} \quad r + M \leq (\log n)^2,$$

*we have*

$$\mathbb{P} \left( \sum_{s=r}^{r+M} \mathbf{1}_{\mathcal{I}_{s, w}(c)} \geq \frac{4}{5}M \right) = 1 - o(1).$$

*The  $o(1)$  is uniform over the initial state,  $w \geq w_0$ , and all admissible  $r$ .*

*Proof.* Apply Lemma 32 with  $\varepsilon = 1/20$ , and let  $w_0$  be the resulting constant. Fix  $w \geq w_0$ .

With probability  $1 - o(1)$ ,

$$\sum_{s=r}^{r+M} \mathbf{1}_{\mathcal{J}_{s, w}(c)} \geq \frac{9}{10}M.$$

By Corollary 30,

$$\sup_{r \leq s \leq r+M} \mathbb{P}(\mathcal{D}_{s, w}^-(c)^c) = o(1).$$

Consequently, if

$$F := \sum_{s=r}^{r+M} \mathbf{1}_{\mathcal{D}_{s, w}^-(c)^c},$$

then  $\mathbb{E}F = o(M)$ , and Markov's inequality gives  $\mathbb{P}(F > M/10) = o(1)$ . Since

$$\sum_{s=r}^{r+M} \mathbf{1}_{\mathcal{I}_{s,w}(c)} \geq \sum_{s=r}^{r+M} \mathbf{1}_{\mathcal{J}_{s,w}(c)} - F,$$

the claimed  $4M/5$  bound follows.  $\square$

## 5 The Coupling

This section decomposes the centered overlap into the linear drift  $\Gamma_t$  and accumulated one-step fluctuations. We then identify the time at which the drift reaches the  $\sqrt{n}$ -scale and transfer its Gaussian hitting-time asymptotic to the discrete chain.

**Definition 34.** For  $t \geq 1$ , define the centered one-step fluctuation

$$A_t := K_t - \mathbb{E}[K_t \mid \mathcal{G}_{t-1}],$$

and, for  $0 \leq j \leq t$ , set

$$B_{j,t} := \prod_{r=j}^{t-1} (Z_r - Z'_r),$$

where an empty product equals one. Thus  $B_{t,t} = 1$  and  $K_0 B_{0,t} = \Gamma_t$ .

Because  $Q_t$  and  $\Lambda_t$  are  $\mathcal{F}_t$ -measurable,  $\mathcal{G}_{t-1} \subseteq \mathcal{G}_t$ ; hence  $(A_t)$  is a martingale-difference sequence for  $(\mathcal{G}_t)$ .

Corollary 11 gives

$$K_t = (Z_{t-1} - Z'_{t-1})K_{t-1} + A_t.$$

**Lemma 35.** For every  $t \geq 1$ ,

$$K_t = K_0 B_{0,t} + \sum_{j=1}^t A_j B_{j,t}.$$

*Proof.* Iterate the preceding recursion.  $\square$

The product barriers from Section 4 control the coefficients in this decomposition. Whenever  $\mathcal{I}_{t-1,w}(c)$  is defined, for  $0 \leq j \leq t$  and  $w \geq 1$ ,

$$\mathcal{I}_{t-1,w}(c) \implies |B_{j,t}| \leq e^c w e^{-c(t-j)}.$$

Indeed, for  $j \leq t-2$ , take  $k = t-1-j$  in the definition of  $\mathcal{I}_{t-1,w}(c)$ ; the cases  $j = t-1, t$  follow from  $|Z_{t-1} - Z'_{t-1}| \leq 1$  and  $B_{t,t} = 1$ .

## 5.1 The $\sqrt{n}$ -scale

Lemma 35 reduces the entrance problem to the decay of the linear drift and the accumulated fluctuations. We first locate the entrance time of the linear drift.

The increments  $Y_j$  from Subsection 4.2 satisfy

$$\mu := \mathbb{E}[Y_0] = -\lambda = -2 \log 2, \quad \text{Var}(Y_0) = \sigma^2 = \frac{\pi^2}{6}.$$

Thus  $a_n \asymp \sqrt{\log n}$  and  $a_n = o(t_n)$ .

**Definition 36.** For fixed  $w > 0$ , let

$$\tau_w := \inf \left\{ t \in \mathbb{Z}_{\geq 0} : |K_0| \prod_{j=0}^{t-1} |R_j - R'_j| \leq w\sqrt{n} \right\},$$

$$\hat{\tau}_w := \inf \{ t \in \mathbb{Z}_{\geq 0} : |\Gamma_t| \leq w\sqrt{n} \}.$$

**Lemma 37.** Fix  $0 < c_- \leq c_+ < \infty$  and  $w > 0$ . For every  $u \in \mathbb{R}$  and every deterministic sequence  $u_n \rightarrow u$ , uniformly over deterministic  $K_0$  satisfying  $c_-n \leq |K_0| \leq c_+n$ ,

$$\mathbb{P}(\tau_w > \lfloor t_n + u_n a_n \rfloor) \longrightarrow \mathbb{P}(G > u),$$

where  $G \sim N(0, 1)$ .

*Proof.* Let

$$t = \lfloor t_n + u_n a_n \rfloor, \quad \Sigma_t := \sum_{j=0}^{t-1} Y_j, \quad b_n := -\log \frac{|K_0|}{w\sqrt{n}}.$$

Since  $Y_j \leq 0$ , the continuous product is nonincreasing and, for all large  $n$ ,  $\{\tau_w > t\} = \{\Sigma_t > b_n\}$ . Uniformly over the stated values of  $K_0$ ,

$$b_n = -\frac{1}{2} \log n + O(1), \quad \mathbb{E}\Sigma_t = -\lambda t,$$

and hence, using  $\lambda a_n = \sigma\sqrt{t_n}$ ,

$$\frac{b_n - \mathbb{E}\Sigma_t}{\sigma\sqrt{t}} = u + o(1).$$

The central limit theorem for  $\Sigma_t$  now proves the claim. The bounded  $O(1)$  term above also makes the convergence uniform in  $K_0$ .  $\square$

Section 4 compares the discrete and continuous logarithmic products from time  $m_n$  onward; the next lemma controls the preceding times.

Let  $\rho_M$  be uniform in  $S_M$ , write its cycles as  $C_1, \dots, C_{L_M}$ , and let  $\zeta_1, \dots, \zeta_{L_M}$  be independent Bernoulli(1/2) variables, independent of  $\rho_M$ . Define the selected fraction

$$V_M := \frac{1}{M} \sum_{\ell=1}^{L_M} \zeta_\ell |C_\ell|.$$

**Lemma 38.** *Uniformly for  $M \geq n/2$  and  $0 < x \leq 1$ ,*

$$\begin{aligned} \mathbb{P}(V_M < x) \vee \mathbb{P}(V_M > 1 - x) &\leq C \left( x + \frac{\log^4 M + 2}{M} \right)^{1/10} \\ &\quad + CM \exp \left( -(\log 2)(\log M)^2 \right). \end{aligned}$$

*Proof.* The law of  $V_M$  is that of  $Z_t$  conditional on  $N_t = M$ . Lemma 23 couples  $V_M$  to  $R$  so that

$$\mathbb{P}(|V_M - R| > \epsilon_M) \leq CM \exp \left( -(\log 2)(\log M)^2 \right).$$

On the complement of this coupling failure,  $V_M < x$  implies  $R < x + \epsilon_M$ . Lemma 15 therefore gives

$$\mathbb{P}(V_M < x) \leq C(x + \epsilon_M)^{1/10} + CM \exp \left( -(\log 2)(\log M)^2 \right),$$

with the trivial bound used if  $x + \epsilon_M > 1$ . Complementing every cycle label preserves the joint law and replaces  $V_M$  by  $1 - V_M$ , so the same argument proves the upper-tail bound.  $\square$

**Lemma 39.** *Fix  $c_- \in (0, 1/4)$ . Assume that the reference configuration is the initial state  $X_0 = x_0$ , so that  $K_0 = N_0(1 - N_0/n)$ , and that  $K_0 \geq c_- n$ . Uniformly over such reference configurations,*

$$\sum_{j=0}^{m_n-1} \log |Z_j - Z'_j| = o_{\mathbb{P}}(a_n).$$

*Proof.* The lower bound on  $K_0$  keeps  $N_0/n$  in a fixed compact subinterval of  $(0, 1)$ , so  $\mathcal{H}_0$  holds for large  $n$ . For each  $j < m_n$ , let

$$M_j := \max\{N_j, n - N_j\} \geq n/2.$$

Conditional on  $\mathcal{F}_j$ , the selected fraction from this macroscopic level set has law  $V_{M_j}$ . If  $\mathcal{H}_{j+1}$  fails, then either  $V_{M_j} < 2/\sqrt{n}$  or  $V_{M_j} > 1 - 2/\sqrt{n}$ . Thus Lemma 38 gives, uniformly for  $j < m_n$ ,

$$\mathbb{P}(\mathcal{H}_{j+1}^c \mid \mathcal{F}_j) \leq Cn^{-1/20} + o(n^{-1/20}).$$

Consequently, a union bound yields

$$\mathbb{P}\left(\bigcup_{j=0}^{m_n} \mathcal{H}_j^c\right) = o(1).$$

Fix  $\varepsilon > 0$ , and set

$$\delta_n := \exp(-\varepsilon a_n/m_n).$$

On  $\mathcal{H}_j$ , Lemma 24 gives, uniformly for  $0 \leq j < m_n$ ,

$$\begin{aligned} \mathbb{P}(|Z_j - Z'_j| < \delta_n, \mathcal{H}_j) &\leq C \left( \delta_n + \frac{4 \log^4 n}{\sqrt{n}} \right)^{1/20} \\ &\quad + Cn \exp\left(-\frac{\log 2}{4}(\log n)^2\right) = o(m_n^{-1}). \end{aligned}$$

After another union bound, with probability  $1 - o(1)$ , all  $m_n$  factors satisfy  $|Z_j - Z'_j| \geq \delta_n$ . Since their moduli are at most one,

$$0 \geq \sum_{j=0}^{m_n-1} \log |Z_j - Z'_j| \geq -\varepsilon a_n.$$

Since  $\varepsilon > 0$  is arbitrary, the claim follows.  $\square$

**Lemma 40.** Fix  $0 < c_- < 1/4$  and  $w > 0$ . For every  $u \in \mathbb{R}$  and every deterministic sequence  $u_n \rightarrow u$ ,

$$\sup_{x_0 \in X_n: K_0(x_0) \geq c_- n} |\mathbb{P}_{x_0}(\widehat{\tau}_w > \lfloor t_n + u_n a_n \rfloor) - \mathbb{P}(G > u)| \longrightarrow 0,$$

where  $G \sim N(0, 1)$ .

*Proof.* Since  $(u_n)$  is bounded, choose  $C > 0$  so that, for all large  $n$ ,

$$\lfloor t_n + u_n a_n \rfloor \leq C \log n.$$

By Lemma 27,

$$\max_{m_n \leq s \leq C \log n} \left| \sum_{j=m_n}^{s-1} \log |Z_j - Z'_j| - \sum_{j=m_n}^{s-1} \log |R_j - R'_j| \right| = o_{\mathbb{P}}(1),$$

uniformly over the reference configurations in the statement. Set

$$t = \lfloor t_n + u_n a_n \rfloor,$$

and define

$$\widehat{\Sigma}_t := \sum_{j=0}^{t-1} \log |Z_j - Z'_j|, \quad \Sigma_t := \sum_{j=0}^{t-1} Y_j.$$

Since  $m_n = o(a_n)$  and  $m_n \sigma^2 = o(a_n^2)$ , Chebyshev's inequality gives

$$\sum_{j=0}^{m_n-1} Y_j = o_{\mathbb{P}}(a_n).$$

Together with Lemma 39 and the forward comparison, this yields

$$\widehat{\Sigma}_t = \Sigma_t + o_{\mathbb{P}}(a_n)$$

uniformly. Since  $\sigma\sqrt{t} \sim \lambda a_n$ , Slutsky's theorem gives

$$\frac{\widehat{\Sigma}_t - \mu t}{\sigma\sqrt{t}} \Rightarrow G.$$

Finally,  $|\Gamma_s|$  is nonincreasing, so, for all large  $n$ ,  $\{\widehat{\tau}_w > t\}$  is the event

$$\widehat{\Sigma}_t > -\log \frac{|K_0|}{w\sqrt{n}}.$$

Since  $K_0 = N_0(1 - N_0/n) \leq n/4$ , the normalized threshold tends to  $u$ , uniformly in  $x_0$ , by the calculation in Lemma 37 with  $c_+ = 1/4$ . This proves the stated uniform convergence.  $\square$

Thus  $|\Gamma_t|$  enters the  $\sqrt{n}$ -scale window at time  $t_n + O(a_n)$ , with Gaussian fluctuations.

## 6 The Final Coupling Step

We now complete the upper-bound coupling argument. Fix the reference configuration  $x_0$ , and hence the value  $N_0$ . We couple two copies of the projected chain,  $(M_t, S_t)$  and  $(M'_t, S'_t)$ , associated with this same reference configuration. Lemma 8 shows that this projected chain captures exactly the mixing behavior of  $X_t$ .

**Definition 41.** Let  $K_t$  be the centered overlap from Definition 9 and  $\Gamma_t$  the drift product in (1). Set

$$h_n := \lceil \log \log \log n \rceil, \quad \ell_n := \lfloor (\log \log n)^2 \rfloor.$$

For  $w > 0$ , let

$$\mathcal{K}_s(w) := \{|K_s| \leq 2w\sqrt{n}\}, \quad \mathcal{R}_s(w) := \{|\Gamma_s| \leq w\sqrt{n}\},$$

and suppress  $w$  when it is fixed. Finally, set

$$\mathcal{M}_n := \{N_s = N'_s \text{ for every } s \geq h_n\}.$$

Both  $h_n$  and  $\ell_n$  are  $o(a_n)$ , and all block endpoints below are integer times.

Lemma 40 shows that at times of order  $t_n$ ,  $|\Gamma_t|$  reaches the  $\sqrt{n}$ -scale. Since  $|Z_j - Z'_j| \leq 1$ , the event  $\mathcal{R}_t$  persists at all later times. Over the following block, we show that  $\mathcal{K}_s$  holds for a positive fraction of times. At simultaneous bulk times for the two copies, their one-step laws come from nearby hypergeometric laws and can be maximally coupled with a uniformly positive success probability. Repeated attempts then give a geometric failure bound.

By Lemma 5, after  $h_n$  preliminary steps the two magnetization chains can be coupled with failure probability  $\mathbb{P}(\mathcal{M}_n^c) = O(e^{-c_{\text{mag}} h_n}) = o(1)$ . Once they agree, common cycle lengths and labels keep the magnetizations equal. Thus, under this construction,  $\mathcal{M}_n = \{N_{h_n} = N'_{h_n}\}$ , an event known at time  $h_n$ . On  $\mathcal{M}_n$ , we use common cycle-length and label data and couple the remaining randomness arbitrarily except at the designated maximal-coupling attempts. After the projected chains meet, we use identical randomness; this preserves both marginals. On  $\mathcal{M}_n^c$ , complete the joint construction in any marginally correct way. Its probability is absorbed into subsequent  $o(1)$  terms.

**Proposition 42.** Fix  $c \in (0, \log 2)$ , and let  $r$  be an integer satisfying

$$2m_n + 1 \leq r \quad \text{and} \quad r + \ell_n \leq (\log n)^2.$$

There exist constants  $v_0 = v_0(c) > 0$  and  $L_0 = L_0(c) > 0$  such that, for every  $v \geq v_0$  and every  $w \geq L_0 v$ ,

$$\mathbb{P} \left( \mathcal{R}_r \cap \left\{ \sum_{s=r}^{r+\ell_n} \mathbf{1}_{\mathcal{I}_{s-1,v}(c) \cap \mathcal{K}_s} < \frac{19\ell_n}{30} \right\} \right) = o(1).$$

The  $o(1)$  is uniform over the initial state and all admissible  $r, v, w$ .

*Proof.* Choose  $v_0 \geq 1$  large enough so that Corollary 33 applies. By that corollary, applied to the shifted block  $[r-1, r+\ell_n-1]$ , with probability  $1 - o(1)$ ,

$$\sum_{s=r}^{r+\ell_n} \mathbf{1}_{\mathcal{I}_{s-1,v}(c)} \geq \frac{4}{5} \ell_n.$$

Let

$$\text{Bad}_r := \sum_{s=r}^{r+\ell_n} \mathbf{1}_{\mathcal{I}_{s-1,v}(c) \cap \mathcal{R}_s \cap \mathcal{K}_s^c}.$$

By Lemma 35, on  $\mathcal{I}_{s-1,v}(c) \cap \mathcal{R}_s$ ,

$$K_s = K_0 B_{0,s} + \sum_{j=1}^s A_j B_{j,s}, \quad |K_0 B_{0,s}| = |\Gamma_s| \leq w\sqrt{n} \quad \text{and} \quad |B_{j,s}| \leq e^c v e^{-c(s-j)}.$$

Since  $\sum_{j=1}^s e^{-c(s-j)/2} \leq (1 - e^{-c/2})^{-1}$ , failure of  $\mathcal{K}_s$  implies, for some  $j \leq s$ ,

$$|A_j| \geq \alpha \frac{w}{v} e^{c(s-j)/2} \sqrt{n}, \quad \alpha := e^{-c}(1 - e^{-c/2}).$$

Consequently,

$$\text{Charge}_j^{(r)} := \sum_{q=\max\{0, r-j\}}^{r+\ell_n-j} \mathbf{1}_{\{|A_j| \geq \alpha(w/v)e^{cq/2}\sqrt{n}\}} \quad \text{satisfies} \quad \text{Bad}_r \leq \sum_{j=1}^{r+\ell_n} \text{Charge}_j^{(r)}.$$

By Corollary 12, there are constants  $C_0, c_0 > 0$  such that, uniformly in the past,

$$\mathbb{P}(|A_j| \geq \alpha(w/v)e^{cq/2}\sqrt{n} \mid \mathcal{G}_{j-1}) \leq C_0 \exp\{-c_0(w/v)^2 e^{cq}\}.$$

Writing  $q_j := \max\{0, r - j\}$ , the event  $\{\text{Charge}_j^{(r)} \geq m\}$  forces the threshold event with  $q = q_j + m - 1$ . Summing the preceding bound and using the tail-sum formula therefore gives

$$\begin{aligned} \sum_{j=1}^{r+\ell_n} \mathbb{E}[\text{Charge}_j^{(r)} \mid \mathcal{G}_{j-1}] &\leq C_1 \ell_n e^{-c_0(w/v)^2} + C_2, \\ \sum_{j=1}^{r+\ell_n} \mathbb{E}[(\text{Charge}_j^{(r)})^2 \mid \mathcal{G}_{j-1}] &\leq C_3 \ell_n e^{-c_0(w/v)^2} + C_4. \end{aligned}$$

Here the  $j \geq r$  terms give the multiples of  $\ell_n$ , while the  $j < r$  terms form summable tails.

Since  $\text{Charge}_j^{(r)}$  is  $\mathcal{G}_j$ -measurable and  $(\mathcal{G}_j)$  is the filtration identified in Section 5, the partial sums of  $\text{Charge}_j^{(r)} - \mathbb{E}[\text{Charge}_j^{(r)} \mid \mathcal{G}_{j-1}]$  form a martingale. Choose  $L_0 \geq 1$  so that  $C_1 e^{-c_0 L_0^2} \leq 1/24$ . For large  $n$ , the first moment bound is at most  $\ell_n/12$ ; the second and Chebyshev's inequality then give

$$\mathbb{P}\left(\text{Bad}_r > \frac{\ell_n}{6}\right) \leq \mathbb{P}\left(\sum_{j=1}^{r+\ell_n} \text{Charge}_j^{(r)} > \frac{\ell_n}{6}\right) = o(1).$$

On  $\mathcal{R}_r$ , all later  $\mathcal{R}_s$  occur. Outside the two exceptional events above,

$$\sum_{s=r}^{r+\ell_n} \mathbf{1}_{\mathcal{I}_{s-1,v}(c) \cap \mathcal{K}_s} \geq \frac{4}{5} \ell_n - \frac{1}{6} \ell_n = \frac{19}{30} \ell_n,$$

which proves the proposition.  $\square$

From now on, fix constants  $v \geq v_0$  and  $w \geq L_0 v$ , where  $v_0, L_0$  are from Proposition 42, independently of  $n$ .

**Definition 43.** Set

$$K'_s := M'_s - \frac{N'_s N_0}{n}.$$

Let  $\overline{Q}_{s+1}$  and  $\overline{\Lambda}_{s+1}$  be the selected totals from its 1- and 0-level sets, and set

$$\overline{Z}_s := \frac{\overline{Q}_{s+1}}{N'_s}, \quad \overline{Z}'_s := \frac{\overline{\Lambda}_{s+1}}{n - N'_s},$$

with the same boundary conventions as for  $Z_s, Z'_s$ . Define

$$\bar{\Gamma}_t := K'_0 \prod_{j=0}^{t-1} (\bar{Z}_j - \bar{Z}'_j).$$

$$\mathcal{K}'_s := \{|K'_s| \leq 2w\sqrt{n}\}, \quad \bar{\mathcal{R}}_s := \{|\bar{\Gamma}_s| \leq w\sqrt{n}\},$$

and let  $\bar{\mathcal{I}}_{s,v}(c)$  be the event obtained from  $\mathcal{I}_{s,v}(c)$  by replacing  $Z_j, Z'_j$  with  $\bar{Z}_j, \bar{Z}'_j$ . Call  $s$  a simultaneous good time if  $\mathcal{I}_{s-1,v}(c) \cap \mathcal{K}_s$  and  $\bar{\mathcal{I}}_{s-1,v}(c) \cap \mathcal{K}'_s$  both occur.

**Lemma 44.** *Let  $r$  be an integer satisfying*

$$2m_n + 1 \leq r \quad \text{and} \quad r + \ell_n \leq (\log n)^2.$$

*Then*

$$\mathbb{P} \left( \mathcal{R}_r \cap \bar{\mathcal{R}}_r \cap \left\{ \sum_{s=r}^{r+\ell_n} \mathbf{1}_{\{s \text{ is a simultaneous good time}\}} < \frac{\ell_n}{4} \right\} \right) = o(1).$$

*The  $o(1)$  is uniform over the two initial states and all admissible  $r$ .*

*Proof.* By Proposition 42, applied to each copy, outside an event of probability  $o(1)$  each has at least  $19\ell_n/30$  good times on its respective entrance event. On  $\mathcal{R}_r \cap \bar{\mathcal{R}}_r$ , their intersection therefore has size at least

$$\frac{19\ell_n}{30} + \frac{19\ell_n}{30} - (\ell_n + 1) = \frac{4\ell_n}{15} - 1.$$

This is at least  $\ell_n/4$  for large  $n$ . □

List the first copy's cycles in an auxiliary random order independent of their coordinate supports, and let

$$\mathcal{C}_s^{\text{com}} := \sigma((|C_j|, \zeta_j)_{j \leq k}, (|C'_j|, \zeta'_j)_{j \leq \ell}).$$

This reveals its labeled cycle lengths, but not their supports, and makes  $Q_{s+1}, \Lambda_{s+1}$  measurable. Whenever  $N_s = N'_s$ , the two copies have identically distributed ordered cycle-length vectors, so use the same lengths and labels for the second copy. In particular,  $\bar{Q}_{s+1} = Q_{s+1}$  and  $\bar{\Lambda}_{s+1} = \Lambda_{s+1}$ . When the magnetizations differ, sample the second copy's data separately.

**Definition 45.** A simultaneous good time  $s$  is  $\eta$ -usable if  $N_s = N'_s$  and

$$\eta n \leq N_s \leq (1 - \eta)n, \quad \frac{Q_{s+1}}{N_s}, \frac{\Lambda_{s+1}}{n - N_s} \in [\eta, 1 - \eta].$$

Set

$$\mathcal{U}_{r, \ell_n}^{(\eta)} := \left\{ \sum_{s=r}^{r+\ell_n} \mathbf{1}_{\{s \text{ is } \eta\text{-usable simultaneous good}\}} \geq \frac{\ell_n}{12} \right\}.$$

**Lemma 46.** There exists  $\eta_0 \in (0, 1/4)$  such that, for every  $0 < \eta \leq \eta_0$ , the following holds. Let  $r$  be an integer satisfying

$$2m_n + 1 \leq r \quad \text{and} \quad r + \ell_n \leq (\log n)^2.$$

Then

$$\mathbb{P}(\mathcal{R}_r \cap \overline{\mathcal{R}}_r \cap (\mathcal{U}_{r, \ell_n}^{(\eta)})^c) = o(1).$$

For each fixed  $\eta$ , the  $o(1)$  is uniform over the two initial states and all admissible  $r$ .

*Proof.* Since  $\mathbb{P}(\mathcal{M}_n^c) = o(1)$ , it suffices to control discards on  $\mathcal{M}_n$ . First, without conditioning on  $\mathcal{M}_n$ , let

$$\text{Bd}_\eta := \{0, \dots, \lfloor \eta n \rfloor\} \cup \{n - \lfloor \eta n \rfloor, \dots, n\}.$$

For  $r \leq s \leq t \leq r + \ell_n$ , the Markov property and Lemma 5 give

$$|\text{Cov}(\mathbf{1}_{\text{Bd}_\eta}(N_s), \mathbf{1}_{\text{Bd}_\eta}(N_t))| \leq C e^{-c_{\text{mag}}(t-s)}.$$

Indeed, the transition law from any state is within  $e^{-c_{\text{mag}}(t-s)}$  of stationarity after  $t - s$  steps. The sum below therefore has variance  $O(\ell_n)$ . Its normalized mean is at most  $C\eta + o(1)$ , because the stationary magnetization is uniform and  $r \rightarrow \infty$ . Hence

$$\frac{1}{\ell_n} \sum_{s=r}^{r+\ell_n} \mathbf{1}_{\{N_s \in \text{Bd}_\eta\}} \leq C\eta + o_{\mathbb{P}}(1).$$

Let

$$\mathcal{P}_s := \sigma(((M_j, S_j), (M'_j, S'_j))_{j \leq s}, (Q_{j+1}, \Lambda_{j+1}, \overline{Q}_{j+1}, \overline{\Lambda}_{j+1})_{0 \leq j < s}).$$

For large  $n$ ,  $h_n < r$ , so  $\mathcal{M}_n$  and simultaneous goodness are  $\mathcal{P}_s$ -measurable. On  $\mathcal{M}_n$ , conditional on  $\mathcal{P}_s$ , the fresh totals are selected fractions from level sets of sizes  $N_s$  and  $n - N_s$ . Lemmas 23 and 15, together with the symmetry  $R \stackrel{d}{=} 1 - R$ , therefore give  $\rho(\eta) := C\eta^{1/10}$  such that, uniformly when  $N_s \notin \text{Bd}_\eta$ ,

$$\mathbb{P}\left(\frac{Q_{s+1}}{N_s} \notin [\eta, 1 - \eta] \text{ or } \frac{\Lambda_{s+1}}{n - N_s} \notin [\eta, 1 - \eta] \mid \mathcal{P}_s\right) \leq \rho(\eta) + o(1).$$

Hence, for

$$\text{Disc}_s := \mathbf{1}_{\mathcal{M}_n} \mathbf{1}_{\{s \text{ is simultaneous good but not } \eta\text{-usable}\}},$$

$$\mathbb{E}[\text{Disc}_s \mid \mathcal{P}_s] \leq \mathbf{1}_{\{N_s \in \text{Bd}_\eta\}} + (\rho(\eta) + o(1)) \mathbf{1}_{\{s \text{ is simultaneous good}\}}.$$

The variable  $\text{Disc}_s$  is  $\mathcal{P}_{s+1}$ -measurable, so its centered values are bounded martingale differences. Combining their  $o_{\mathbb{P}}(1)$  average with the boundary estimate gives

$$\frac{\mathbf{1}_{\mathcal{M}_n}}{\ell_n} \sum_{s=r}^{r+\ell_n} \mathbf{1}_{\{s \text{ is simultaneous good but not } \eta\text{-usable}\}} \leq C\eta + \rho(\eta) + o_{\mathbb{P}}(1).$$

Choose  $\eta_0 < 1/4$  so small that  $C\eta + \rho(\eta) < 1/24$  for every  $0 < \eta \leq \eta_0$ . With probability  $1 - o(1)$ , at most  $\ell_n/12$  simultaneous good times are discarded on  $\mathcal{M}_n$ . On  $\mathcal{R}_r \cap \overline{\mathcal{R}_r}$ , Lemma 44 supplies at least  $\ell_n/4$  simultaneous good times, so  $\mathcal{U}_{r, \ell_n}^{(\eta)}$  holds.  $\square$

The hypergeometric estimate used at each coupling attempt is proved in Appendix A.

**Lemma 47.** *Fix  $A \geq 0$  and  $\eta \in (0, 1/2)$ . There exists a constant  $\delta_{\text{hyp}} = \delta_{\text{hyp}}(A, \eta) < 1$  such that the following holds for all sufficiently large  $n$ . Suppose*

$$N \in [\eta n, n], \quad \frac{m}{N}, \frac{m'}{N}, \frac{r}{N} \in [\eta, 1 - \eta],$$

and

$$|m - m'| \leq A\sqrt{n}.$$

Then

$$\|\text{Hyp}(N, m, r) - \text{Hyp}(N, m', r)\|_{\text{TV}} \leq \delta_{\text{hyp}}.$$

*The same conclusion holds for products of two hypergeometric laws whenever both factors satisfy these bulk assumptions, after possibly increasing  $\delta_{\text{hyp}}$ , but still with  $\delta_{\text{hyp}} < 1$ .*

**Lemma 48.** Fix  $\eta \in (0, 1/4)$  and assume  $N_0/n \in [2\eta, 1 - 2\eta]$ . There is  $\delta_{\text{hyp}} = \delta_{\text{hyp}}(4w, \eta) < 1$  such that, whenever  $t$  is  $\eta$ -usable,

$$\left\| \mathcal{L}((M_{t+1}, S_{t+1}) \mid M_t, S_t, \mathcal{C}_t^{\text{com}}) - \mathcal{L}((M'_{t+1}, S'_{t+1}) \mid M'_t, S'_t, \mathcal{C}_t^{\text{com}}) \right\|_{\text{TV}} \leq \delta_{\text{hyp}}.$$

Thus, conditional on the current projected states and  $\mathcal{C}_t^{\text{com}}$ , a maximal coupling succeeds with probability at least  $1 - \delta_{\text{hyp}}$ .

*Proof.* At a simultaneous good time,  $|K_t|, |K'_t| \leq 2w\sqrt{n}$ . Since  $N_t = N'_t$ ,

$$|M_t - M'_t| = |K_t - K'_t| \leq 4w\sqrt{n},$$

and the same bound holds for  $|(N_0 - M_t) - (N_0 - M'_t)|$ . Usability gives  $N_t, n - N_t \geq \eta n$  and

$$\frac{Q_{t+1}}{N_t}, \frac{\Lambda_{t+1}}{n - N_t} \in [\eta, 1 - \eta].$$

Moreover,

$$\frac{M_t}{N_t} = \frac{N_0}{n} + \frac{K_t}{N_t}, \quad \frac{N_0 - M_t}{n - N_t} = \frac{N_0}{n} - \frac{K_t}{n - N_t},$$

and likewise for the primed copy. Since  $N_0/n \in [2\eta, 1 - 2\eta]$ , all four success proportions belong to  $[\eta, 1 - \eta]$  for sufficiently large  $n$ . Thus both pairs of hypergeometric factors satisfy Lemma 47 with  $A = 4w$ .

By the hypergeometric decomposition in Subsection 3.3, conditional on  $\mathcal{C}_t^{\text{com}}$  and the current projected state,  $M_{t+1} = M_{t+1}^{(1)} + M_{t+1}^{(0)}$ , where independently

$$M_{t+1}^{(1)} \sim \text{Hyp}(N_t, M_t, Q_{t+1}) \quad \text{and} \quad M_{t+1}^{(0)} \sim \text{Hyp}(n - N_t, N_0 - M_t, \Lambda_{t+1}).$$

This follows from Lemma 10: conditional on the ordered block sizes and labels, their supports are uniform ordered partitions, so the union of the 1-labeled blocks is a uniform subset of the prescribed size. Also  $N_{t+1} = Q_{t+1} + \Lambda_{t+1}$  and  $S_{t+1} = M_{t+1} + n - N_0 - N_{t+1}$ , so the projected state is a deterministic function of the pair. The product estimate in Lemma 47 and contraction of total variation under a deterministic map give the bound  $\delta_{\text{hyp}} < 1$ .  $\square$

**Proposition 49.** *Under the assumptions of Lemma 48, suppose maximal couplings are attempted at every  $\eta$ -usable simultaneous good time in  $[r, r + \ell_n]$ . Then, for every integer  $q \geq 1$ ,*

$$\mathbb{P} \left( \begin{array}{c} (M_{r+\ell_n+1}, S_{r+\ell_n+1}) \neq (M'_{r+\ell_n+1}, S'_{r+\ell_n+1}), \\ \sum_{s=r}^{r+\ell_n} \mathbf{1}_{\{s \text{ is } \eta\text{-usable simultaneous good}\}} \geq q \end{array} \right) \leq \delta_{\text{hyp}}^q.$$

*Proof.* Let  $r_i$  be the time of the  $i$ -th coupling attempt, with  $r_i = \infty$  if fewer than  $i$  attempts occur. Usability is decided from the current states and  $\mathcal{C}_{r_i}^{\text{com}}$ , before the cycle supports are sampled. Thus, on  $\{r_i \leq r + \ell_n\}$ , provided the chains are not yet coupled, Lemma 48 bounds the conditional failure probability at  $r_i$  by  $\delta_{\text{hyp}}$ . After a success, run the chains together. Induction shows that the probability of failing the first  $q$  attempts while  $r_q \leq r + \ell_n$  is at most  $\delta_{\text{hyp}}^q$ . The event in the display is contained in this event, proving the claim.  $\square$

**Proposition 50.** *Let  $r$  be an integer satisfying*

$$2m_n + 1 \leq r \quad \text{and} \quad r + \ell_n \leq (\log n)^2.$$

*Let  $\eta_0$  be the constant from Lemma 46. Fix  $0 < \eta \leq \eta_0$ , and assume the bulk condition  $N_0/n \in [2\eta, 1 - 2\eta]$ . Under the coupling constructed above, uniformly over reference configurations satisfying this bulk condition,*

$$\mathbb{P} \left( \mathcal{R}_r \cap \overline{\mathcal{R}}_r \cap \left\{ (M_{r+\ell_n+1}, S_{r+\ell_n+1}) \neq (M'_{r+\ell_n+1}, S'_{r+\ell_n+1}) \right\} \right) = o(1).$$

*Proof.* By Lemma 46, on  $\mathcal{R}_r \cap \overline{\mathcal{R}}_r$ , the event  $\mathcal{U}_{r, \ell_n}^{(\eta)}$  fails with probability  $o(1)$ . On  $\mathcal{U}_{r, \ell_n}^{(\eta)}$ , Proposition 49 with  $q = \lfloor \ell_n/12 \rfloor$  bounds the failure probability by  $\delta_{\text{hyp}}^{\lfloor \ell_n/12 \rfloor} = o(1)$ , because  $\delta_{\text{hyp}}(4w, \eta) < 1$  is fixed.  $\square$

## 7 The Limit Profile

This section combines the Gaussian entrance-time asymptotic for  $|\Gamma_t|$  with the final coupling. Before entrance, the centered overlap distinguishes the chain from stationarity; after entrance, the coupling from Section 6 gives the matching upper bound.

**Lemma 51.** *For every  $u \in \mathbb{R}$  and every deterministic sequence  $u_n \rightarrow u$ ,*

$$\liminf_{n \rightarrow \infty} d_n(\lfloor t_n + u_n a_n \rfloor) \geq \mathbb{P}(G > u).$$

*Proof.* Start from a deterministic configuration  $x_0$  with  $N_0 = \lfloor n/2 \rfloor$ , so  $K_0 = \Theta(n)$ , and set

$$t = \lfloor t_n + u_n a_n \rfloor.$$

Fix  $v, W \geq 1$ . On  $\mathcal{I}_{t-1,v}(c)$ , the decomposition in Lemma 35 and the product-barrier bound give  $|B_{j,t}| \leq e^c v e^{-c(t-j)}$ . Arguing as in Proposition 42 and using Corollary 12,

$$\mathbb{P}\left(|K_t - \Gamma_t| > \frac{W}{2}\sqrt{n}, \mathcal{I}_{t-1,v}(c)\right) \leq \varepsilon(W/v),$$

where  $\varepsilon(a) \rightarrow 0$  as  $a \rightarrow \infty$ , uniformly in  $n$ . Indeed, the event on the left forces  $|A_j| \geq c_1(W/v)e^{c(t-j)/2}\sqrt{n}$  for some  $j$ , and the resulting tail probabilities are summable in  $t - j$ . Since  $\hat{\tau}_W > t$  implies  $|\Gamma_t| > W\sqrt{n}$ ,

$$\mathbb{P}_{x_0}\left(|K_t| > \frac{W}{2}\sqrt{n}\right) \geq \mathbb{P}(\hat{\tau}_W > t) - \mathbb{P}(\mathcal{I}_{t-1,v}(c)^c) - \varepsilon(W/v).$$

Under stationarity, conditional on the magnetization,  $M$  is hypergeometric with mean  $NN_0/n$  and variance  $O(n)$ . Hence

$$\pi_n\left(|K| > \frac{W}{2}\sqrt{n}\right) \leq \frac{C}{W^2}.$$

Testing total variation with this event and applying Lemma 40 gives

$$\liminf_{n \rightarrow \infty} d_n(t) \geq \mathbb{P}(G > u) - \limsup_{n \rightarrow \infty} \mathbb{P}(\mathcal{I}_{t-1,v}(c)^c) - \varepsilon(W/v) - \frac{C}{W^2}.$$

By Lemmas 29 and 31, the limsup involving  $\mathcal{I}_{t-1,v}(c)^c$  tends to zero as  $v \rightarrow \infty$ . After taking  $n \rightarrow \infty$ , let  $v \rightarrow \infty$  and choose  $W = W(v)$  so that  $W/v \rightarrow \infty$ . All three error terms then vanish, proving the lemma.  $\square$

**Lemma 52.** *For every  $u \in \mathbb{R}$  and every deterministic sequence  $u_n \rightarrow u$ ,*

$$\limsup_{n \rightarrow \infty} d_n(\lfloor t_n + u_n a_n \rfloor) \leq \mathbb{P}(G > u).$$

*Proof.* Fix  $\varepsilon > 0$ . Choose  $v \geq v_0$ , and choose  $w \geq L_0 v$ , as in Section 6. Let  $\eta_0 > 0$  be the constant from Lemma 46. Choose  $0 < \eta \leq \min\{\eta_0, 1/8\}$  small enough that  $4\eta \leq \varepsilon/2$ . Let

$$T_n := \lfloor t_n + u_n a_n \rfloor.$$

By Corollary 7, applied with bulk parameter  $2\eta$  and error  $\varepsilon/2$ , let  $r_* := r_{2\eta, \varepsilon/2} = O_{\eta, \varepsilon}(1)$ . From every starting state, with probability at least  $1 - \varepsilon$ ,  $X_{r_*}$  belongs to

$$\text{Bulk}_{n, \eta} := \{y \in X_n : 2\eta n \leq |y| \leq (1 - 2\eta)n\}.$$

Indeed, Corollary 7 gives probability at least  $1 - 4\eta - \varepsilon/2 \geq 1 - \varepsilon$ . The Markov property and convexity of total variation therefore give

$$d_n(T_n) \leq \varepsilon + \sup_{y \in \text{Bulk}_{n, \eta}} d_n^y(T_n - r_*).$$

Set

$$s_n := T_n - r_*, \quad r_n := s_n - \ell_n - 1.$$

Both are  $t_n + (u + o(1))a_n$ , and  $[r_n, r_n + \ell_n]$  satisfies the range assumptions of Proposition 50. Fix  $y \in \text{Bulk}_{n, \eta}$ , use  $y$  as the reference configuration, and couple the chain from  $y$  to a stationary copy as in Section 6. On  $\{\hat{\tau}_w \leq r_n\}$ , monotonicity gives  $\mathcal{R}_{r_n}$ . For the stationary copy,  $|\bar{\Gamma}_{r_n}| \leq |K'_0|$ , and the conditional hypergeometric variance bound gives

$$\mathbb{P}(\bar{\mathcal{R}}_{r_n}^c) \leq \frac{C}{w^2}.$$

Lemma 8, the coupling inequality, and Proposition 50 therefore yield, uniformly in  $y \in \text{Bulk}_{n, \eta}$ ,

$$d_n^y(s_n) \leq \mathbb{P}_y(\hat{\tau}_w > r_n) + \frac{C}{w^2} + o(1).$$

For such  $y$ , writing  $p = |y|/n$ ,

$$2\eta(1 - 2\eta)n \leq K_0 = np(1 - p) \leq \frac{n}{4}.$$

The uniform form of Lemma 40, applied at  $r_n = t_n + (u + o(1))a_n$ , now gives

$$\limsup_{n \rightarrow \infty} d_n(T_n) \leq \mathbb{P}(G > u) + \frac{C}{w^2} + \varepsilon.$$

Finally let  $w \rightarrow \infty$ , keeping  $v$  fixed and  $w \geq L_0 v$ , and then let  $\varepsilon \downarrow 0$ . This gives

$$\limsup_{n \rightarrow \infty} d_n(\lfloor t_n + u_n a_n \rfloor) \leq \mathbb{P}(G > u).$$

□

Since  $\mathbb{P}(G > u) = 1 - \Phi(u)$ , the two lemmas give the stated profile. Letting  $u \rightarrow -\infty$  and  $u \rightarrow \infty$  gives cutoff with window  $a_n = o(t_n)$ , completing the proof of Theorem 2.

## A A hypergeometric total variation estimate

This appendix proves the hypergeometric overlap estimate used in the final coupling. A uniform local normal approximation gives a common interval on which the two laws have probabilities of order  $n^{-1/2}$ , yielding a positive overlap independent of  $n$ .

*Proof of Lemma 47.* Let

$$H \sim \text{Hyp}(N, m, r), \quad H' \sim \text{Hyp}(N, m', r),$$

with means  $\mu = rm/N, \mu' = rm'/N$  and variances  $v, v'$ . The bulk assumptions give, for some  $\kappa = \kappa(\eta) > 0$ ,

$$\kappa n \leq v, v' \leq \kappa^{-1} n, \quad |\mu - \mu'| \leq A\sqrt{n}.$$

If  $L = \max\{0, m + r - N\}$  and  $U = \min\{m, r\}$ , then

$$\mu - L \geq \eta^2 N, \quad U - \mu \geq \eta^2 N,$$

and the same support bounds hold for  $H'$ .

Let  $\phi_{\mu, v}$  denote the density of  $N(\mu, v)$ . The local normal expansion of Lahiri–Chatterjee–Maiti [17, Lemma 3.1] uses

$$\tilde{v} = \frac{N-1}{N} v \asymp n, \quad a_{k,r} = \frac{k - \mu}{\tilde{v}} = O(n^{-1/2}) \quad \text{when } |k - \mu| \leq C\sqrt{n}.$$

Its explicit remainder is therefore  $O(n^{-1/2})$ , uniformly in the bulk parameters. Exponentiating and replacing  $\tilde{v}$  by  $v$  gives, for every fixed  $C$ ,

$$\sup_{|k - \mu| \leq C\sqrt{n}} |\mathbb{P}(H = k) - \phi_{\mu, v}(k)| = o(n^{-1/2}),$$

and analogously for  $H'$ .

Set  $I_n = [\mu, \mu + \sqrt{n}]$ . The support bounds and  $|\mu - \mu'| \leq A\sqrt{n}$  place every integer in  $I_n$  in both supports for large  $n$ . On this interval,

$$|x - \mu| \leq \sqrt{n}, \quad |x - \mu'| \leq (A + 1)\sqrt{n},$$

so, for some  $c_0 = c_0(A, \eta) > 0$ ,

$$\phi_{\mu, v}(x) \wedge \phi_{\mu', v'}(x) \geq \frac{c_0}{\sqrt{n}}.$$

The local approximation therefore gives

$$\mathbb{P}(H = k) \wedge \mathbb{P}(H' = k) \geq \frac{c_0}{2\sqrt{n}}$$

for every integer  $k \in I_n$ . Since  $I_n$  contains at least  $\sqrt{n}/2$  integers,

$$\sum_{k \in \mathbb{Z}} \min\{\mathbb{P}(H = k), \mathbb{P}(H' = k)\} \geq \beta(A, \eta)$$

for some  $\beta(A, \eta) > 0$ . Hence

$$\|\mathcal{L}(H) - \mathcal{L}(H')\|_{\text{TV}} \leq 1 - \beta(A, \eta) < 1.$$

This proves the single hypergeometric estimate, with  $\delta_{\text{hyp}} = 1 - \beta(A, \eta)$ .

Finally,

$$1 - \|\rho_1 \otimes \rho_2 - \rho'_1 \otimes \rho'_2\|_{\text{TV}} \geq (1 - \|\rho_1 - \rho'_1\|_{\text{TV}})(1 - \|\rho_2 - \rho'_2\|_{\text{TV}}).$$

Thus the product overlap is at least  $\beta(A, \eta)^2$ , proving the product version with  $\delta_{\text{hyp}} = 1 - \beta(A, \eta)^2 < 1$ .  $\square$

## Acknowledgements

We thank Persi Diaconis for useful discussions.

## References

- [1] David Aldous and James Allen Fill. Reversible markov chains and random walks on graphs, 2002. Unfinished monograph, recompiled 2014.
- [2] Hans C. Andersen and Persi Diaconis. Hit and run as a unifying device. *Journal de la Société Française de Statistique*, 148(4):5–28, 2007.
- [3] Laurent Bartholdi and Persi Diaconis. An algorithm for uniform generation of unlabeled trees (Polya trees), with an extension of Cayley’s formula, 2024.
- [4] Persi Diaconis. Analysis of a Bose–Einstein markov chain. *Annales de l’Institut Henri Poincaré, Probabilités et Statistiques*, 41(3):409–418, 2005.
- [5] Persi Diaconis and Michael Howes. Random sampling of contingency tables and partitions: Two practical examples of the Burnside process. *Statistics and Computing*, 35, 2025.
- [6] Persi Diaconis, Andrew Lin, and Arun Ram. A curiously slowly mixing markov chain, 2025.
- [7] Persi Diaconis and Calder Morton-Ferguson. Markov chains on Weyl groups from the geometry of the flag variety, 2025.
- [8] Persi Diaconis and Chenyang Zhong. Hahn polynomials and the Burnside process. *The Ramanujan Journal*, 61(2):567–595, 2023.
- [9] Persi Diaconis and Chenyang Zhong. Counting the number of group orbits by marrying the Burnside process with importance sampling, 2026.
- [10] Ivan Z. Feng. The dual Burnside process. *Journal of Theoretical Probability*, 39:78, 2026.
- [11] Ivan Z. Feng and J. E. Paguyo. Analysis of a twisted Bose–Einstein markov chain with applications to sampling Catalan structures, 2026.
- [12] Thomas S. Ferguson. A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, 1(2):209–230, 1973.

- [13] Leslie Ann Goldberg and Mark Jerrum. The “Burnside Process” converges slowly. *Combinatorics, Probability and Computing*, 11(1):21–34, 2002.
- [14] Steven Holtzen, Todd Millstein, and Guy Van den Broeck. Generating and sampling orbits for lifted probabilistic inference. In *Proceedings of the 35th Uncertainty in Artificial Intelligence Conference*, volume 115 of *Proceedings of Machine Learning Research*, pages 985–994. PMLR, 2020.
- [15] Michael Howes. Limit profiles and cutoff for the Burnside process on Sylow double cosets, 2025.
- [16] Mark Jerrum. Uniform sampling modulo a group of symmetries using markov chain simulation. In Joel Friedman, editor, *Expanding Graphs*, volume 10 of *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, pages 37–47. American Mathematical Society, Providence, RI, 1993.
- [17] S. N. Lahiri, A. Chatterjee, and T. Maiti. Normal approximation to the hypergeometric distribution in nonstandard cases and a sub-gaussian berry–esseen theorem. *Journal of Statistical Planning and Inference*, 137(11):3570–3590, 2007.
- [18] David A. Levin and Yuval Peres. *Markov chains and mixing times*. American Mathematical Society, Providence, RI, 2017. Second edition of [MR2466937], With contributions by Elizabeth L. Wilmer, With a chapter on “Coupling from the past” by James G. Propp and David B. Wilson.
- [19] J. E. Paguyo. Mixing times of a Burnside process Markov chain on set partitions. *Advances in Applied Mathematics*, 173:102982, 2026.
- [20] Jim Pitman. *Combinatorial Stochastic Processes*, volume 1875 of *Lecture Notes in Mathematics*. Springer, Berlin, 2006.
- [21] John Rahmani. Mixing times for the commuting chain on CA groups. *Journal of Theoretical Probability*, 35(1):457–483, 2022.
- [22] Jayaram Sethuraman. A constructive definition of Dirichlet priors. *Statistica Sinica*, 4(2):639–650, 1994.