

Mixing of general biased adjacent transposition chains

Reza Gheissari*, Holden Lee[†], and Eric Vigoda[‡]

November 4, 2025

Abstract

We analyze the general biased adjacent transposition shuffle process, which is a well-studied Markov chain on the symmetric group S_n . In each step, an adjacent pair of elements i and j are chosen, and then i is placed ahead of j with probability p_{ij} . This Markov chain arises in the study of self-organizing lists in theoretical computer science, and has close connections to exclusion processes from statistical physics and probability theory. Fill (2003) conjectured that for general p_{ij} satisfying $p_{ij} \geq 1/2$ for all $i < j$ and a simple monotonicity condition, the mixing time is polynomial. We prove that for any fixed $\varepsilon > 0$, as long as $p_{ij} > 1/2 + \varepsilon$ for all $i < j$, the mixing time is $\Theta(n^2)$ and exhibits pre-cutoff. Our key technical result is a form of spatial mixing for the general biased transposition chain after a suitable burn-in period. In order to use this for a mixing time bound, we adapt multiscale arguments for mixing times from the setting of spin systems to the symmetric group.

1 Introduction

We study a natural Markov chain on biased permutations which is motivated by the study of self-organizing data structures, and is a natural generalization of fundamental stochastic processes in probability theory and statistical mechanics. The Markov chain is defined on the set of $n!$ permutations weighted by given pairwise probabilities indicating the preference for the pairwise ordering.

For integer $n \geq 1$, let $[n] = \{1, \dots, n\}$, and let S_n denote the symmetric group; thus, S_n denotes the set of $n!$ permutations of the n particles $\{1, \dots, n\}$. For a permutation $\sigma \in S_n$ and an index $i \in [n]$, let $\sigma(i)$ denote the particle in position i in σ , and let $\sigma^{-1}(i)$ denote the position in σ of the i -th particle. We denote an adjacent transposition as follows: for a permutation $\sigma \in S_n$ and an index $i \in [n-1]$, let $\sigma \circ (i, i+1)$ denote the permutation obtained by swapping the particles at positions i and $i+1$.

*Department of Mathematics, Northwestern University. Email: gheissari@northwestern.edu

[†]Department of Applied Mathematics and Statistics, Johns Hopkins University. Email: hlee283@jhu.edu

[‡]Department of Computer Science, University of California, Santa Barbara. Email: vigoda@ucsb.edu

We consider the adjacent transposition Markov chain, denoted $\mathcal{M}_{\text{AT}}$, on S_n weighted by given parameters $p_{i,j}$ for all $i < j$, where $0 \leq p_{i,j} \leq 1$. Given the input probabilities $p_{i,j}$ for all $i < j$, then for all $j > i$ set $p_{j,i} = 1 - p_{i,j}$, and let $\vec{p} = (p_{i,j})_{1 \leq i,j \leq n}$ denote the collection of probabilities. From a state $\sigma_t \in S_n$, the transitions $\sigma_t \rightarrow \sigma_{t+1}$ of the Markov chain $\mathcal{M}_{\text{AT}}$ are the following:

- Choose an index $i \in [n-1]$ uniformly at random.
- Let

$$\sigma_{t+1} = \begin{cases} \sigma_t \circ (i, i+1) & \text{with probability } p_{\sigma_t(i+1), \sigma_t(i)} \\ \sigma_t & \text{with probability } p_{\sigma_t(i), \sigma_t(i+1)} \end{cases}.$$

The Markov chain $\mathcal{M}_{\text{AT}}$ is time-reversible with unique stationary distribution μ given by

$$\mu(\sigma) = \mu_{\vec{p}, [n]}(\sigma) = \frac{1}{Z} \prod_{1 \leq i < j \leq n} p_{\sigma(i), \sigma(j)}, \quad (1)$$

where the partition function is defined as $Z = \sum_{\sigma \in S_n} \prod_{i < j} p_{\sigma(i), \sigma(j)}$. It is an open problem to give an efficient algorithm under general conditions on $(p_{i,j})_{1 \leq i,j \leq n}$ for approximate sampling from this distribution, or by the standard reduction between counting and sampling, for approximation of the partition function. One case where this is known is when $p_{i,j} = \frac{w_i}{w_i + w_j}$ with $w_1 \leq w_2 \leq \dots \leq w_n$. In this case, the partition function is a permanent of a matrix, which is efficiently approximable [JS89, JSV04]. However, this relies on specific structure on $p_{i,j}$. One motivation for studying $\mathcal{M}_{\text{AT}}$ is that it provides a generic algorithm that does not rely on specific structure on $p_{i,j}$.

The efficiency of using a Markov chain for approximate sampling is quantified by its mixing time. To be precise, for an ergodic Markov chain (X_t) on state space Ω with transition matrix P and unique stationary distribution μ , the *mixing time* (to $\delta < 1/2$) is

$$T_{\text{mix}}(\delta) = \min_t \max_{X_0 \in \Omega} \|P^t(X_0, \cdot) - \mu\|_{\text{TV}} \leq \delta, \quad (2)$$

where $\|\mu - \nu\|_{\text{TV}} = \frac{1}{2} \|\mu - \nu\|_1$ is the total-variation distance between probability measures. When there is no δ argument, we use the convention $T_{\text{mix}} = T_{\text{mix}}(1/4)$.

A well-studied special case is when $p_{ij} = q$ for all $i < j$, for some constant q . The chain exhibits different behavior for $q = 1/2$ and $q \neq 1/2$. When $q = 1/2$, then the stationary distribution μ is uniform over all $n!$ permutations and the Markov chain $\mathcal{M}_{\text{AT}}$ is the unbiased adjacent transposition chain, which is well-studied in probability theory; see [Ald83, DSC93]. In this setting, Wilson [Wil03] proved that the mixing time to δ is $\Theta(n^3 \log n)$, and Lacoïn [Lac16] established the cutoff phenomenon, identifying the constant in front of $n^3 \log n$ and that showing it is independent of δ . The constant bias case, when $q \in (\frac{1}{2}, 1)$, is closely related to the asymmetric simple exclusion process (ASEP) from interacting particle systems (see the original paper of [Spi70] and the book of Liggett [Lig85]), with deep connections to statistical mechanics and the KPZ universality class (see e.g., [KRBN10] and [Cor16]). In

this setting, mixing is faster: Benjamini, Berger, Hoffman, and Mossel [BBHM05] proved that the mixing time is $\Theta(n^2)$, and Labbe and Lacoïn [LL19] established cutoff.

An additional special case of particular interest corresponds to the problem of linear extensions of a partial order. In this setting, some subset of pairs have $p_{ij} = 1$ (and thus require $\sigma^{-1}(i) < \sigma^{-1}(j)$) and all other pairs have $p_{ij} = 1/2$ (and hence are unbiased). For this problem, Bubley and Dyer [BD99] proved that the mixing time is $O(n^3)$.

The general version of the Markov chain $\mathcal{M}_{\text{AT}}$ has been studied in the context of self-organizing lists, where the frequency that elements are accessed is according to an underlying distribution, and elements are moved towards the front of the list when accessed to minimize the expected access time. In that context, the chain $\mathcal{M}_{\text{AT}}$ corresponds to the Move-Ahead-One (MA1) list update algorithm, and is important in the probabilistic analysis of its stationary behavior; we refer the interested reader to Hendricks [Hen76] and Hester and Hirschberg [HH85] for an introduction to this area, and to Fill [Fil03] for further connections.

With no conditions on $\vec{p}$, it is easy to construct slow-mixing examples, and therefore the following two conditions on the bias parameters $\vec{p} = (p_{i,j})_{1 \leq i < j \leq n}$ have become important for analysis the general setting:

1. ε -positive bias: For all $1 \leq i < j \leq n$, $\frac{p_{i,j}}{p_{j,i}} \geq 1 + \varepsilon$.
2. Monotonicity: For all $i < j$, $p_{i,j} \leq p_{i,j+1}$ and for all $i + 1 < j$, $p_{i,j} \geq p_{i+1,j}$.

A well-known and long-standing conjecture of Fill [Fil03] is that the Markov chain $\mathcal{M}_{\text{AT}}$ has polynomial mixing time whenever $\vec{p}$ satisfies 0-positive bias and monotonicity. Monotonicity is necessary: Bhakta, Miracle, Randall and Streib [BMRS22] provided an example where the mixing time is exponential in n and where $\vec{p}$ satisfies the 0-positive bias condition but not the monotonicity condition.

In the positive direction, progress towards obtaining polynomial mixing time bounds for positively biased chains have relied on various special conditions on $\vec{p}$ (either assuming some algebraic relations, or some block structure across which the p_{ij} 's are constant). Bhakta et al. [BMRS22] proved polynomial mixing time in two restricted settings: in the first case, for $i < j$, $p_{i,j}$ only depends on i ; in the second setting, the probabilities $p_{i,j}$ are defined based on a binary tree.

Haddadan and Winkler [HW17] studied a simpler chain known as the gladiator chain where each particle has a strength $s(i)$ and the transposition of particles i and j is a fixed function of $s(i)$ and $s(j)$, and proved polynomial mixing when there are $k = 3$ different available particle strengths. Miracle and Streib [MS18] considered the setting where the n particles are partitioned into k classes $\mathcal{C}_1, \dots, \mathcal{C}_k$ and the particles within the same class are indistinguishable; they proved mixing time of $O(n^{2k+6})$. Finally, Miracle, Streib, and Streib [MSS25] also considered the setting of k classes, and proved that if the inter-class biases satisfy ε -positive bias and all classes have size $|\mathcal{C}_i| \geq C(\varepsilon) \log n$ for some constant $C(\varepsilon)$, then the mixing time of $\mathcal{M}_{\text{AT}}$ is $O(n^9)$ (independent of k).

We prove the following general result: for every $\varepsilon > 0$ and any $\vec{p}$ which satisfies ε -positive bias, the process $\mathcal{M}_{\text{AT}}$ has mixing time of $\Theta(n^2)$, where the constant in the upper bound

depends on ε . We do not require monotonicity, and note that our result does not contradict the slow mixing example of [BMRS12, Section 6] without monotonicity as they have $\varepsilon \approx 1/n$.

Theorem 1.1: For every $\varepsilon > 0$ and any fixed $\delta > 0$, there exists $C(\varepsilon)$ such that for any $(p_{i,j})_{1 \leq i < j \leq n}$ which satisfies ε -positive bias, the adjacent transposition Markov chain $\mathcal{M}_{\text{AT}}$ has mixing time¹

$$n^2 \leq T_{\text{mix}}(\delta) \leq C(\varepsilon)n^2 + \tilde{O}(n \log(1/\delta)).$$

In particular, the family has *pre-cutoff* (see [LP17, Chapter 18]): there exists $C(\varepsilon)$ such that for any sequence of $\mathcal{M}_{\text{AT}}^{(n)}$ with ε -positively biased $\vec{p}^{(n)}$, for any fixed $\delta \in (0, \frac{1}{2})$,

$$\limsup_{n \rightarrow \infty} \frac{T_{\text{mix}}(\delta)}{T_{\text{mix}}(1-\delta)} \leq C(\varepsilon) < \infty.$$

Note this implies that the spectral gap of the chain is at least $\Omega(C(\varepsilon)^{-1}n^{-2})$, due to a standard inequality between mixing and relaxation time [LP17, Theorem 12.5]. It is natural to wonder whether the limit of T_{mix}/n^2 can be pinned down, thereby showing cutoff. However, among ε -positively biased $\vec{p}$, the mixing time can truly differ by order n^2 . In particular, the totally asymmetric case where $p_{i,j} = 1$ for all $i < j$ saturates the lower bound of Theorem 1.1, and the constant-bias case $p_{i,j} = q$ with $q/(1-q) = 1 + \varepsilon$ for all $i < j$ saturates the upper bound with our constant $C(\varepsilon)$. Therefore, any refinement into a cutoff result would require assuming a further consistency relation between the sequences $\vec{p}^{(n)}$.

Remark 1.2. For fixed $\varepsilon > 0$ and n sufficiently large (in a way that depends on ε), the constant $C(\varepsilon)$ in Theorem 1.1 behaves like $C(\varepsilon) = O(1/\varepsilon)$. On the other hand, for all ε, n we give an upper bound of $e^{\tilde{O}_\varepsilon(1/\varepsilon^2)} \text{poly}(n)$ where $\tilde{O}_\varepsilon$ hides poly-logarithmic factors in $1/\varepsilon$ (see Proposition 5.1). There is a lower bound of $e^{\Omega(1/\varepsilon)}$ that can be constructed as follows: for $n_0 = \Theta(1/\varepsilon)$, use the slow mixing example of [BMRS12, Section 6] on $1 \leq i, j \leq n_0$ and for all i, j that are not both in $[n_0]$, set $p_{ij} = \mathbf{1}_{i < j}$. Closing this gap would be interesting. We also note that with the added assumption of monotonicity, one would expect a $\text{poly}(1/\varepsilon)$ dependence on ε to hold for all ε, n , consistent with the conjecture of [Fil03].

2 Proof overview

In many of the previous results where polynomial bounds are known under special settings on the bias parameters $\vec{p}$, there is some underlying algebraic structure (e.g., bijections to Markov chains on up-down walks or Dyck paths) that is used as an initial ingredient into the mixing time analysis. As a notable example, in the homogenous case of $p_{i,j} = q$ for all $i < j$, the process which tracks the locations (but not the labels) of the first k particles is Markovian, and is exactly a k -particle ASEP. With general biased $\vec{p}$, we lose these exact connections and we instead import tools from spin systems to our analysis. In particular,

¹Here we are using $\tilde{O}$ to mean up to polylogarithmic factors in n ; the implicit constant here may, and in fact does, depend on ε . However, the constant in the $\tilde{O}$ is independent of δ .

our overall strategy is to do a multi-scale reduction of the mixing time for $\mathcal{M}_{\text{AT}}$ on $[n]$ to that on $[\alpha n]$ for $\alpha < 1$ and recurse.

The typical way such reductions are achieved in spin system analysis is through decay of correlations between the configuration σ on far apart regions on $[n]$. In our setting, decay of correlations does *not* hold for general pinnings of the configuration, but does hold for a family of typical ones which we call ℓ -localized: roughly speaking, this means that for all i , particle i is within distance ℓ of location i .

Our strategy is captured by the following three stages:

1. Burn-in: After $O(n^2)$ steps, regardless of the initialization, the configuration is $O(\log n)$ -localized, and stays so for polynomially many steps.
2. Polynomial bound by spatial recursion: We show that within $O(\log n)$ -localized configurations, there is near-exponential decay of correlations; this property is then used together with a multi-scale block dynamics recursion to obtain $n^{O(1)}$ mixing time.
3. Boosting to quadratic mixing time: We use a second block dynamics reduction within the set of $O(\log n)$ -localized configurations to boost the polynomial mixing time to obtain mixing in just $O(n(\log n)^{O(1)})$ steps beyond the initial $O(n^2)$ length burn-in.

We describe each of these steps in more detail in the remainder of this section. In our proof overview, we will state lemmas in the form needed for establishing the polynomial upper bound. As we will describe when outlining the boosting argument, all of these steps are needed conditional on a more general form of ℓ -localization for the boosting to optimal quadratic mixing time. Therefore, throughout the proof sketch, we will accompany lemmas with pointers to later in the paper for their more general, conditional versions.

2.1 Burn-in to localized configurations

Since we assume positive bias, the most likely configuration, known as the ground state, is the permutation with the particles in order $1, \dots, n$. We define the following set of configurations, which we refer to as ℓ -localized, abbreviated ℓ -Loc, in which no particle deviates by more than ℓ positions from its ground state location.

Definition 2.1. *For an integer $\ell \geq 0$, we say a permutation $\sigma \in S_n$ is ℓ -localized if the following holds:*

$$\sigma \in \ell\text{-Loc} := \bigcap_{1 \leq k \leq n} \{|\sigma^{-1}(k) - k| \leq \ell\}.$$

We will show that in the stationary distribution μ , the property ℓ_0 -Loc holds with high probability (at least $1 - 1/\text{poly}(n)$) for $\ell_0 = C_0 \log n$.

Lemma 2.2. *There exists a constant $C_0 = C_0(\varepsilon) = O(1/\varepsilon)$ such that the following holds:*

$$\mu((C_0 \log n)\text{-Loc}) \geq 1 - n^{-10}.$$

Lemma 2.2 is proved in Section 3. The key step in the proof of Lemma 2.2 is to show that for every $1 \leq k \leq n$, under the stationary distribution, the distance of particle k from location k is stochastically dominated by an appropriate geometric random variable. This follows from first showing that the identity of the particle in location 1 is stochastically dominated by a geometric random variable (Lemma 3.6) and then applying this fact repeatedly after conditioning. Lemma 3.6 is shown by a many-to-one map between configurations that have none of the first k particles in location 1, to those that do, noting that moving all these particles r spaces to the left increases the probability weight by a factor of at least $(1 + \varepsilon)^{kr}$. This step is in the spirit of Peierls arguments from statistical physics.

While the above results show useful properties of the stationary distribution, an additional challenge is showing that the dynamics quickly attains (and maintains) the typical stationary properties; this is often referred to as the burn-in phase of the dynamics. We bound the burn-in time in the following result.

Lemma 2.3. *For every ε , there exists $C_0(\varepsilon) > 0$ such that for $\ell_0 = C_0(\varepsilon) \log n$ and any initial $\sigma_0 \in S_n$, for any $t \geq C_0(\varepsilon) \cdot n^2$,*

$$\mathbb{P}(\sigma_t \in \ell_0\text{-Loc}) \geq 1 - n^{-10}.$$

A more general version of Lemma 2.3 is stated in Lemma 4.1, which is proved in Section 4. We also obtain a burn-in estimate of the form of Lemma 2.3 for a block dynamics variant of $\mathcal{M}_{\text{AT}}$, which we define in the following subsection; see Lemma 4.2.

To prove Lemma 2.3, we use a stochastic domination relation between (k -particle projections of the) $\mathcal{M}_{\text{AT}}$ and the asymmetric exclusion process (ASEP). In ASEP, there are k indistinguishable particles which lie at k distinct locations on $[n]$, with a given bias parameter $q \in (1/2, 1)$. In each step, the ASEP chooses a random location $i \in [n - 1]$, and if exactly one the positions i or $i + 1$ contains a particle, then with probability q the particle is placed ahead at position i , and with probability $1 - q$, it is placed behind in position $i + 1$; thus, there is a bias $q > 1/2$ for moving the particles to the left. For the $\mathcal{M}_{\text{AT}}$ chain with general $\vec{p}$, the process $(\mathbf{1}\{\sigma_t(i) \leq k\})_i$ is not Markov, but it is stochastically “to the left” of a k -particle ASEP with appropriate parameter q . These projections, together with known mixing time bounds of the ASEP, yield the bound on the burn-in phase as stated in Lemma 2.3; see Section 4 for further details.

2.2 Spatial mixing within the localized set

The role played by the localized set of configurations is that within that set of configurations, the distribution μ satisfies a certain near-exponential decay of correlations. To formalize this, we introduce the following set notations. Throughout the paper, when we write an interval $[\ell, m]$ for integers ℓ, m , we mean the set $\{\ell, \ell + 1, \dots, m\} \subseteq [n]$.

Definition 2.4. *For an integer $r \geq 0$ and integers $i, j \geq 1$, let $A = A_{i,j} = [i + 1, n - j]$ and let A_r denote the subset of A which is distance at least r from the boundary. Formally, for $A = [i + 1, n - j]$, define $A_r = [i + 1 + r, n - j - r]$. Moreover, let $A^c = [n] \setminus A$ denote the complement of the region A .*

Our spatial mixing property says that conditioned on the property ℓ -Loc, if we consider two different configurations $\eta, \bar{\eta}$ on A^c (compatible with ℓ -Loc), then the influence of $\eta, \bar{\eta}$ on the configuration on A_r decays (almost) exponentially in r , the distance from A_r to A^c . In the following, we use the notation $\sigma(S)$ to denote the restriction of σ to S ; i.e., we consider the image as an ordered set.

Lemma 2.5. *There exists a constant $C' = C'(\varepsilon) > 0$ such that for any integers $\ell \geq 1$, $i, j \geq 1$, and $0 \leq r \leq (j - i - \ell - 1)/2$, the following holds for any pair $\eta, \bar{\eta} \in \ell$ -Loc:*

$$\|\mu(\sigma(A_r) \in \cdot \mid \sigma(A^c) = \eta, \sigma \in \ell\text{-Loc}) - \mu(\bar{\sigma}(A_r) \in \cdot \mid \bar{\sigma}(A^c) = \bar{\eta}, \bar{\sigma} \in \ell\text{-Loc})\|_{\text{TV}} \leq \exp(-r/C'\ell).$$

This property is the analogue of the well-studied notion of spatial mixing from analysis of spin system dynamics [Mar04], where usually the spatial mixing holds for *all* configurations on A^c . In our setting, however, that stronger statement can be seen to obviously fail: take $A = \{1, \dots, n/2\}$ and suppose that η assigns particles $\{1, \dots, n/2\}$ to A^c while $\bar{\eta}$ assigns particles $\{n/2+1, \dots, n\}$ to A^c ; then the available alphabets on A will be completely disjoint, and there is no opportunity for σ and $\bar{\sigma}$ to agree anywhere in A .

The proof of Lemma 2.5 goes by directly constructing a coupling $\sigma, \bar{\sigma}$ to agree on A_r with high probability. We will prove a generalization of Lemma 2.5 stated in Lemma 3.3 in Section 3.4. The coupling relies on finding what we call disconnecting positions in $A \setminus A_r$.

Definition 2.6. *We say a position k is disconnecting if $\sigma([k]) = [k]$ as unordered sets. In words, the first k particles are contained (in some order) in the first k positions of σ .*

We prove that a disconnecting point is likely for n sufficiently large (as a function of ε).

Lemma 2.7. *For every ε , there exists $c(\varepsilon) > 0$ and $C(\varepsilon) > 0$ such that the following holds for all $n \geq C(\varepsilon)$, all $\eta \in \ell$ -Loc, and all $k \in A_\ell = [i + 1 + \ell, n - j - \ell]$:*

$$\mu(k \text{ disconnecting for } \sigma \mid \sigma(A^c) = \eta, \sigma \in \ell\text{-Loc}) \geq c(\varepsilon). \quad (3)$$

To prove Lemma 2.7, we utilize the stochastic domination properties established in the proof of Lemma 2.2. A more general version of Lemma 2.7 is stated in Lemma 3.9 and proved in Section 3.4.

Now note that if two configurations $\sigma([k]), \bar{\sigma}([k])$ are such that k is a disconnecting point for both (though the $[k]$ particles may be in different orders in them), then the distribution induced on $[k]^c$ by each of them is the same, and hence the identity coupling couples them to agree on $[k]^c$. In every segment of length ℓ in $A \setminus A_r$, by Lemma 2.7, there is a new attempt at having a disconnecting point in each of σ and $\bar{\sigma}$, with constant probability of success. Therefore, one can use a revealing procedure to prove the spatial mixing property in Lemma 2.5.

2.3 Recursive argument to prove polynomial mixing

We utilize the above spatial mixing and burn-in results to establish polynomial mixing time.

Theorem 2.8: There exists $C > 0$ such that for all $\varepsilon > 0$, there exists $C'(\varepsilon)$, such that the mixing time of $\mathcal{M}_{\text{AT}}$ satisfies

$$T_{\text{mix}} \leq C'(\varepsilon) \cdot n^C.$$

A more refined version of Theorem 2.8 is stated in Proposition 5.1, which includes the quantitative dependence on ε in the mixing time.

To obtain Theorem 2.8, we utilize a recursive argument. As is common in the analysis of Glauber dynamics for spin systems like the Ising model on lattice graphs, the recursive argument uses comparison with a suitable block dynamics.

We consider the following block dynamics which we denote as $\mathcal{M}_{\text{WE}}$. Define the blocks

$$B_W = \{1, \dots, \lceil 2n/3 \rceil\} \quad B_E = \{\lfloor n/3 \rfloor, \dots, n\}.$$

The transitions $\hat{\sigma}_t \rightarrow \hat{\sigma}_{t+1}$ operate as follows. From $\hat{\sigma}_t$, randomly select $B \in \{B_W, B_E\}$, set $\hat{\sigma}_{t+1}(B^c) = \hat{\sigma}_t(B^c)$ and sample $\hat{\sigma}_{t+1}(B)$ from μ conditional on $\hat{\sigma}_{t+1}(B^c) = \hat{\sigma}_t(B^c)$.

We utilize the following decomposition result, which is standard in the setting of spin systems. For a reversible, ergodic Markov chain $\mathcal{M}$ on state space Ω with transition matrix P , let $\lambda_1 = 1 > \lambda_2 \geq \lambda_3 \geq \dots \geq \lambda_{|\Omega|} > -1$ denote the eigenvalues of P . Let $\gamma(\mathcal{M}) = 1 - \lambda_2$ denote the *spectral gap* of $\mathcal{M}$. We state here the following simple form of the decomposition result; see Proposition 5.3 for a more general statement.

$$\gamma(\mathcal{M}_{\text{AT}}) \geq \frac{1}{2} \cdot \gamma(\mathcal{M}_{\text{WE}}) \cdot \min_{B \in \{B_W, B_E\}} \min_{\eta} \gamma(\mathcal{M}_{\text{AT}}(B^\eta)), \quad (4)$$

where η is a fixed configuration on B^c and $\mathcal{M}_{\text{AT}}(B^\eta)$ is the chain $\mathcal{M}_{\text{AT}}$ operating on B (so n is rescaled to $\lceil 2n/3 \rceil$) conditional on the fixed configuration η on B^c .

The high-level idea is to lower bound $\gamma(\mathcal{M}_{\text{WE}})$, and then treat $\gamma(\mathcal{M}_{\text{AT}}(B^\eta))$ as (up to relabeling of the particles) an adjacent transposition shuffle on a smaller scale $\lceil \lceil 2n/3 \rceil \rceil$ instead of $[n]$, with its own ε -positively biased vector $\vec{p}$.

Lemma 2.9. *There exists a constant $C_{\text{block}} > 0$ (independent of ε) such that*

$$\gamma(\mathcal{M}_{\text{WE}})^{-1} \leq C_{\text{block}}.$$

A more general version of Lemma 2.9 is stated in Lemma 5.4, which is proved in Section 5. The proof goes by showing that block dynamics first burns in after $O(1)$ steps per (a variant of) Lemma 2.3, then once it is burnt-in, the spatial mixing estimates of Lemma 2.5 are applicable. Namely, if a block update occurs at time t on B_E and two block dynamics chains we are trying to couple, $\sigma_t, \bar{\sigma}_t$, are both burnt-in, then conditional on their values on $B_W \setminus B_E$, their marginals on $B_E \setminus B_W$ are coupled to agree with high probability. If the next update occurs on B_W then the identity coupling will get them to agree everywhere.

With the main ingredients in place, we provide the proof of Theorem 2.8.

Proof of Theorem 2.8. Let $C' = C'(\varepsilon)$ and C be sufficiently large constants, where C' depends on ε and C is independent of ε . We will prove inductively that for all n , the

maximum (over all ε -positively biased $\vec{p}$) inverse spectral gap of $\mathcal{M}_{\text{AT}}$ on $[n]$ is bounded by $C'n^C$.

Let n_0 be a sufficiently large constant as a function of ε . Then the inverse gap on a segment of length less than or equal to n_0 is at most some constant $C'(\varepsilon)$. (This can be seen to be bounded by a constant that is uniform over ε -positively biased $\vec{p}$ by using the bound on the coupling time that comes from, in every step, making the choice that puts the adjacent particles in their preferred order.) This provides the base case.

Now assume the inductive statement holds for all $N < n$, and we will prove it holds for n . By applying (4), one obtains the following upper bound on the inverse spectral gap:

$$2C_{\text{block}} \max_{B \in B_W, B_E} \max_{\eta} \gamma(\mathcal{M}_{\text{AT}}(B^\eta))^{-1}.$$

For any fixed B and configuration η on B^c , by relabeling the remaining particles (see Lemma 3.5 for details on this relabeling), $\mathcal{M}_{\text{AT}}(B^\eta)$ is equivalent to another adjacent transposition chain with ε -positively biased probabilities, on a segment now of length $2n/3$. Thus by the inductive assumption, we get the following upper bound on the inverse spectral gap:

$$2C_{\text{block}} \cdot C'(\varepsilon) \cdot (2/3)^C n^C.$$

As long as C is large enough that $2C_{\text{block}}(2/3)^C < 1$, this evidently satisfies the desired bound for n , concluding the proof. $\square$

2.4 Boosting to quadratic mixing time

We conclude the proof sketch section by describing how to boost a polynomial mixing time bound for $\mathcal{M}_{\text{AT}}$ to a quadratic one.

By Lemma 2.3, we know that for a polynomial stretch of times, $t \in [n^2, n^4]$, say, the chain $\mathcal{M}_{\text{AT}}$ is in $\ell_0\text{-Loc}$ for $\ell_0 = C_0(\varepsilon) \log n$. We will show that restricted to the burn-in set (i.e., rejecting any $\mathcal{M}_{\text{AT}}$ updates that take the dynamics outside $\ell_0\text{-Loc}$), the mixing time is near-linear $\tilde{O}(n)$, from which the upper bound of Theorem 1.1 would follow.

For showing that the restricted adjacent transposition chain mixes quickly, we utilize another block dynamics reduction. Namely, we construct the following blocks

$$B_0 = \bigcup_{i \geq 0} (6i(\log n)^3, (6i+4)(\log n)^3] \quad B_1 = \bigcup_{i \geq 0} ((6i+3)(\log n)^3, (6i+7)(\log n)^3]$$

where each block itself consists of $O(n/(\log n)^3)$ many disjoint segments of length $4(\log n)^3$. Furthermore, note that the distance between $B_0 \setminus B_1$ and $B_1 \setminus B_0$ is $(\log n)^3$ as B_0 and B_1 intersect on stretches of length $(\log n)^3$ at the ends of each constituent segment.

We apply the analogue of the bound of (4) to this block dynamics. The idea of the boosting is that restricted to $\ell_0\text{-Loc}$, the inverse spectral gap of the block dynamics is $O(1)$ because the configuration on $B_0 \setminus B_1$ can be decoupled from the marginal on $B_1 \setminus B_0$ using the spatial mixing estimate Lemma 2.5 on each constituent $(\log n)^3$ -length segment of B_1 , and union bounding.

At the same time, we wish to bound the inverse spectral gap of the adjacent transposition on $B \in \{B_0, B_1\}$ by reasoning that

- restricted to ℓ_0 -Loc, the individual constituent segments evolve independently, and
- the mixing time on each $(\log n)^3$ -length segment is bounded by $(\log n)^{O(1)}$ by application of Theorem 2.8 to that smaller scale.

With these two ingredients, due to tensorization of the spectral gap for product chains, we would get a bound of $O(n(\log n)^{O(1)})$ for the inverse gap of $\mathcal{M}_{\text{AT}}(B^n)$ for $B \in \{B_0, B_1\}$.

There is a subtlety here, however, that the restriction to ℓ_0 -Loc (needed for the configurations on the constituent segments of B_i to be independent) forces the block dynamics to be performed conditional on the configuration after each resampling to remain in ℓ_0 -Loc. In turn, the inverse gap bound on $B \in \{B_0, B_1\}$ needs to be for the restricted adjacent transposition shuffle that rejects any update that takes the full configuration out of ℓ_0 -Loc. This is ultimately what forces us to actually prove all the preceding lemmas and Theorem 2.8 in more general forms, conditioned on certain generalizations of the ℓ -Loc events, as alluded to at the beginning of the proof sketch.

3 Spatial mixing for the stationary distribution

In this section, we study the stationary distribution μ associated to the ε -positively biased adjacent transposition shuffle. Our main aim in the section is to show that for $O(\log n)$ -localized configurations, there is (near) exponential decay of correlations in the sense of Lemma 2.5. As hinted at in the end of the proof sketch, we actually need these stationary estimates not only under μ but also under $\mu(\cdot \mid \vec{\ell}\text{-Loc})$ where the event $\vec{\ell}\text{-Loc}$ constrains each particle i to be a distance $\ell_i \geq 0$ away from its ground state location i . Namely, the following generalizes Definition 2.1.

Definition 3.1. For a vector $\vec{\ell} = (\ell_i^-, \ell_i^+)_{i \in [n]} \in \mathbb{R}^{2n}$ with $\ell_i^\pm \geq 0$, we say σ is $\vec{\ell}$ -localized if

$$\sigma \in \vec{\ell}\text{-Loc} := \bigcap_{1 \leq k \leq n} \{\sigma^{-1}(k) - k \in [-\ell_k^-, \ell_k^+]\}.$$

We let $\ell_{\max}^- = \max_k \ell_k^-$, $\ell_{\max}^+ = \max_k \ell_k^+$, and $\ell_{\max} = \max\{\ell_{\max}^-, \ell_{\max}^+\}$.

Note the earlier definition of ℓ -Loc in Definition 2.1 is a special case of $\vec{\ell}$ -Loc with the constant vector $\vec{\ell} = (\ell, \dots, \ell)$. We use the same terminology for a partial (injective) assignment $\sigma : A \rightarrow [n]$ where $A \subseteq [n]$, where the intersection is now only over $k \in \sigma(A)$.

In order to ensure connectivity (under moves of $\mathcal{M}_{\text{AT}}$) of the set of $\vec{\ell}$ -localized permutations, we restrict attention to admissible localizations, defined below.

Definition 3.2. A vector $\vec{\ell} \in \mathbb{R}^{2n}$ is n -admissible if

$$j - \ell_j^- \leq k - \ell_k^- \quad \text{and} \quad j + \ell_j^+ \leq k + \ell_k^+ \quad \forall j < k.$$

We let $\mathcal{L}_n$ denote the set of n -admissible vectors.

The following is the general form of our spatial mixing lemma, where the spatial mixing given $(C_0 \log n)$ -**Loc** holds even if you condition further on an admissible $\vec{\ell}$ -**Loc**. Establishing this estimate will be the main purpose of this section. The following result is a generalization of Lemma 2.5.

Lemma 3.3. *There exists a constant $C'(\varepsilon) = e^{O(1/\varepsilon)}$ such that the following holds. Consider two permutations $\eta, \bar{\eta} \in \vec{\ell}$ -**Loc** and let $A = [i+1, n-j]$, $A^c = [n] \setminus A$, and $A_r = [i+1+r, n-j-r]$. Then for $r \geq \ell_{\max}$, $r \leq \frac{n-j-i-\ell_{\max}-1}{2}$,*

$$\begin{aligned} \left\| \mu \left(\sigma(A_r) \in \cdot \mid \sigma(A^c) = \eta, \sigma \in \vec{\ell}\text{-Loc} \right) - \mu \left(\bar{\sigma}(A_r) \in \cdot \mid \bar{\sigma}(A^c) = \bar{\eta}, \bar{\sigma} \in \vec{\ell}\text{-Loc} \right) \right\|_{\text{TV}} \\ \leq \exp(-r/C'(\varepsilon)\ell_{\max}). \end{aligned}$$

If $i = 0$, the above holds with $A_r = [1, n-j-r]$, and if $j = 0$, the above holds with $A_r = [i+1, n]$.

Note that after a burn-in to ℓ_0 -**Loc** for $\ell_0 = C_0 \log n$, this implies exponential decay of correlations of the configuration beyond distances $O(\log n)$.

3.1 Self-reducibility of the set of localized configurations

Before working towards the proof of Lemma 2.2, we prove some basic properties about the connectedness and self-reducibility of the Markov chain when restricted to $\vec{\ell}$ -**Loc** for $\vec{\ell}$ admissible.

Claim 3.4. *If $\sigma \in \vec{\ell}$ -**Loc**, and σ' swaps (any, not necessarily adjacent) two particles that are out of order in σ , then $\sigma' \in \vec{\ell}$ -**Loc**. In particular, the adjacent transposition chain restricted to an $\vec{\ell}$ -**Loc** for $\vec{\ell}$ admissible is still ergodic.*

Proof. Suppose that configuration σ has particle i at location v and particle j at location $v+1$ for $i > j$ so that they are “out of order” and swapping the locations of particles i and j . Suppose by way of contradiction that after the update, $v+1-i \notin [-\ell_i^-, \ell_i^+]$ or $v-j \notin [-\ell_j^-, \ell_j^+]$, but before the update $v-i \in [-\ell_i^-, \ell_i^+]$ and $v+1-j \in [-\ell_j^-, \ell_j^+]$.

In the first case, $v+1-i \notin [-\ell_i^-, \ell_i^+]$ while $v-i \in [-\ell_i^-, \ell_i^+]$, then necessarily $v = i + \ell_i^+$ and $v+1 > i + \ell_i^+$. But then, since $i > j$ one also has $v+1 > j + \ell_j^+$ by admissibility of $\vec{\ell}$, which gives a contradiction.

Symmetrically, if $v-j \notin [-\ell_j^-, \ell_j^+]$ while $v+1-j \in [-\ell_j^-, \ell_j^+]$ then necessarily $v+1 = j - \ell_j^-$ so $v < j - \ell_j^-$ but then $i > j$ and admissibility of $\vec{\ell}$ implies $v < i - \ell_i^-$ which is a contradiction.

The ergodicity follows from the fact that from any configuration, there is a path of adjacent transpositions that are swapping particles that are out of order, that takes it to the fully ordered configuration $(1, 2, \dots, n)$. $\square$

We will need the fact that μ conditioned on the values of $\sigma(1), \dots, \sigma(i)$ is also a distribution with ε -positive bias, after relabeling $[n] \setminus \{\sigma(1), \dots, \sigma(i)\}$. Moreover, $\vec{\ell}$ -locality induces some $(n-i)$ -admissible $\vec{\ell}'$ -locality on $[n] \setminus [i]$. This is formalized below.

Lemma 3.5. Let $A = [i+1, n-j]$ and $A^c = [n] \setminus A = [i] \cup [n-j+1, n]$. Let $\vec{\ell}$ be n -admissible. For $b : A^c \rightarrow [n]$ that is $\vec{\ell}$ -localized, define the relabeling map $r_b : [n-i-j] \rightarrow [n] \setminus b(A^c)$ as the unique strictly increasing (one-to-one) map between those sets. Given $\sigma : A \rightarrow [n]$ (such that $\sigma \oplus b \in S_n^2$), define the restriction $R_b\sigma \in S_{n-i-j}$ by

$$R_b\sigma(x) = r_b^{-1}(\sigma(i+x)) \quad \text{for } x \in [n-i-j].$$

Then the following hold.

1. (Labels preserved (up to translation) away from boundary) Let $\sigma \in S_n$ and $b = \sigma(A^c)$. Then $r_b(k) = k+i$ for $k \in [1 + \ell_{\max}^-, n-i-j - \ell_{\max}^+]$.
2. (Locality maintained) Let $\sigma \in S_n$. $\sigma \in \vec{\ell}\text{-Loc}$ if and only if letting $b = \sigma(A^c)$, b is $\vec{\ell}$ -localized and $R_b\sigma \in \vec{\ell}'\text{-Loc}$, where $\vec{\ell}' = R_b\vec{\ell}$ is defined by

$$\begin{aligned} \ell_k'^- &= \min\{\ell_{r_b(k)}^- + k + i - r_b(k), k - 1\} \\ \ell_k'^+ &= \min\{\ell_{r_b(k)}^+ + r_b(k) - k - i, n - i - j - k\}. \end{aligned}$$

Moreover, $\vec{\ell}'$ is $n-i-j$ -admissible with $\ell_{\max}^{\pm} \leq \ell_{\max}^{\pm}$, respectively.

3. (Measure on restriction) Let μ be the probability measure on S_n given by $p_{k,k'}$, and let $\mu|_b$ be the measure conditioned on $\{\sigma \in S_n : \sigma(A^c) = b\}$. Then $(R_b)_*\mu|_b^3$ is the probability measure on S_{n-i-j} given by (1) with $\vec{q}$ given by $q_{k,k'} = p_{r_b(k), r_b(k')}$. If $\vec{p}$ satisfies ε -positive bias, then $\vec{q}$ satisfies ε -positive bias.

Proof. We go through the proof item by item. Note that in the definition of $\vec{\ell}\text{-Loc}$, the following is equivalent:

$$\bigcap_{1 \leq k \leq n} \{\sigma^{-1}(k) - k \in [-\ell_k^-, \ell_k^+]\} = \bigcap_{1 \leq k \leq n} \{\sigma(k) - k \in [-\ell_{\sigma(k)}^+, \ell_{\sigma(k)}^-]\}.$$

1. We only need to consider when the given interval is non-empty. If $k \in [i]$, then because b is $\vec{\ell}$ -localized, $b(k) \leq k + \ell_{b(k)}^- \leq i + \ell_{\max}^-$. If $k \in [n-j+1, n]$, then $b(k) \geq k - \ell_{b(k)}^+ \geq n-j+1 - \ell_{\max}^+$. Therefore, we can write the set of particles not assigned by b as

$$\begin{aligned} [n] \setminus b(A^c) &= ([i + \ell_{\max}^-] \setminus b([i])) \cup [i + \ell_{\max}^- + 1, n - j - \ell_{\max}^+] \\ &\quad \cup ([n - j + 1 - \ell_{\max}^+, n] \setminus b([n - j + 1, n])). \end{aligned}$$

We must have that r_b maps $[\ell_{\max}^-]$, $[\ell_{\max}^- + 1, n-i-j - \ell_{\max}^+]$, $[n-i-j - \ell_{\max}^+ + 1, n-i-j]$ to these three sets, respectively. Considering the middle set gives the conclusion.

² $\sigma \oplus b$ denotes the function that takes values $\sigma(x)$ when $x \in A$ and $b(x)$ when $x \in A^c$.

³We use the notation $f_*\mu$ to denote the pushforward of the measure μ by the function f .

2. Note that $\sigma \in \vec{\ell}\text{-Loc}$ if and only if b is $\vec{\ell}$ -localized and for all $k \in [n - i - j]$,

$$\sigma^{-1}(r_b(k)) \in [r_b(k) - \ell_{r_b(k)}^-, r_b(k) + \ell_{r_b(k)}^+]. \quad (5)$$

The condition that $R_b\sigma \in \vec{\ell}'\text{-Loc}$ is that

$$(R_b\sigma)^{-1}(k) - k \in [-\ell_k'^-, \ell_k'^+]$$

or equivalently that

$$\sigma^{-1}(r_b(k)) \in [i + k - \ell_k'^-, i + k + \ell_k'^+]. \quad (6)$$

Note that if not for the min operation, $\vec{\ell}'$ was chosen exactly so that (5) and (6) coincide. For admissibility, note that these intervals are monotonically moving to the right (since $\vec{\ell}$ is admissible) and r_b is increasing. Finally, note that taking the min simply truncates the intervals to $[1, n]$, and this does not change admissibility.

It remains to check that $\ell_{\max}^{\pm} \leq \ell_{\max}^{\pm}$. For that it is sufficient to show

$$\begin{aligned} \forall k \geq \ell_{\max}^- + 1, & \quad \ell_{r_b(k)}^- + k + i - r_b(k) \leq \ell_{\max}^-, \\ \forall k \leq n - i - j - \ell_{\max}^+, & \quad \ell_{r_b(k)}^+ + r_b(k) - k - i \leq \ell_{\max}^+. \end{aligned}$$

In light of part 1, r_b maps $[\ell_{\max}^-]$ to $[i + \ell_{\max}^-] \setminus b([i])$. Because r_b is defined on an interval of integers, $r_b(x) - x$ is non-decreasing. Therefore, $r_b(k) \geq k + i$ in the first case, showing the first statement. Similarly, r_b maps $[n - i - j - \ell_{\max}^+ + 1, n - i - j]$ to $[n - j + 1 - \ell_{\max}^+, n] \setminus b([n - j + 1, n])$, so $r_b(k) \leq k + i$ in the second case, showing the second statement.

3. Suppose $\sigma' = R_b\sigma$. Then

$$\begin{aligned} (R_b)_* \mu_{|b}(\sigma') &= \mu_{|b}(\sigma) \propto \prod_{i < x < y \leq n-j} p_{\sigma(x), \sigma(y)} \\ &= \prod_{i < x < y \leq n-j} p_{r_b(\sigma'(x-i)), r_b(\sigma'(y-i))} = \prod_{1 \leq x < y \leq n-i-j} q_{\sigma'(x), \sigma'(y)}. \end{aligned}$$

The fact that ε -positive bias is maintained follows from r_b being increasing. $\square$

3.2 Bounds on deviations of particle locations

We begin by showing a geometric decay bound on the label of the particle at location 1.

Lemma 3.6. *Suppose $\vec{p}$ is ε -positively biased and $\vec{\ell} \in \mathcal{L}_n$. For every $k \geq 1$,*

$$\mu(\sigma(1) \notin [k] \mid \sigma \in \vec{\ell}\text{-Loc}) \leq \frac{1}{(1 + \varepsilon)^k}.$$

Hence, the particle label at location 1 is stochastically dominated by a $\text{Geom}(\frac{\varepsilon}{1+\varepsilon})$ random variable.

Proof. Consider the following function $\Phi : S_n \rightarrow S_n$ which maps any σ such that $\sigma(1) \notin [k]$ to a σ' such that $\sigma'(1) \in [k]$ as follows: if the first particle in $[k]$ is at position j , then move all particles in $[k]$ $j - 1$ spaces to the left. Formally, if the first k particles are in positions $j = i_1 < \dots < i_k$, then

$$\Phi(\sigma) = \sigma \circ \bigcirc_{\ell=0}^{j-2} ((i_1 - \ell - 1, i_1 - \ell) \circ \dots \circ (i_k - \ell - 1, i_k - \ell)).$$

(where the last index in the composition is the right-most one). We claim that if $\sigma \in \vec{\ell}\text{-Loc}$, then $\Phi(\sigma) \in \vec{\ell}\text{-Loc}$. This is because each transposition switches the order of two particles which are out of order: the particle on the left is $> k$, and the particle on the right is $\leq k$.

Observe that if σ and $\sigma(i) > \sigma(i+1)$, then considering the only factors which do not cancel out in $\prod_{1 \leq i < j \leq n} p_{\sigma(i)} p_{\sigma(j)}$,

$$\frac{\mu(\sigma \circ (i, i+1))}{\mu(\sigma)} \geq \frac{p_{\sigma(i+1), \sigma(i)}}{p_{\sigma(i), \sigma(i+1)}} \geq 1 + \varepsilon. \quad (7)$$

Now using the fact that $\Phi^{-1}(\sigma)$ for $\sigma : \sigma(1) \in [k]$ partition $\{\sigma : \sigma(1) \notin [k]\}$,

$$\frac{\mu(\sigma(1) \notin [k] \mid \sigma \in \vec{\ell}\text{-Loc})}{\mu(\sigma(1) \in [k] \mid \sigma \in \vec{\ell}\text{-Loc})} = \frac{\sum_{\sigma(1) \in [k], \sigma \in \vec{\ell}\text{-Loc}} \mu(\Phi^{-1}(\sigma) \cap \vec{\ell}\text{-Loc})}{\sum_{\sigma(1) \in [k], \sigma \in \vec{\ell}\text{-Loc}} \mu(\sigma)} \leq \max_{\sigma \in \vec{\ell}\text{-Loc}} \frac{\mu(\Phi^{-1}(\sigma))}{\mu(\sigma)}.$$

Using (7) and the fact that in each of those transpositions, we are switching the order of two particles which are out of order, we can bound the right-hand side by

$$\begin{aligned} \frac{\mu(\Phi^{-1}(\sigma))}{\mu(\sigma)} &= \frac{\sum_{j \geq 1} \mu\left(\sigma \circ \bigcirc_{\ell=0}^{j-2} ((i_1 - \ell - 1, i_1 - \ell) \circ \dots \circ (i_k - \ell - 1, i_k - \ell))\right)}{\mu(\sigma)} \\ &\leq \sum_{j \geq 1} \frac{1}{(1 + \varepsilon)^{jk}}. \end{aligned}$$

Therefore (using that $\mu(A)r \geq \mu(A^c)$ implies $\mu(A) \geq 1/(1+r)$),

$$\mu(\sigma(1) \notin [k] \mid \sigma \in \vec{\ell}\text{-Loc}) \leq 1 - \left(1 + \sum_{j \geq 1} \frac{1}{(1 + \varepsilon)^{jk}}\right)^{-1} = \frac{1}{(1 + \varepsilon)^k},$$

as claimed. □

We now prove that configurations from μ are $C_0 \log n\text{-Loc}$ with high probability. Though this proof would hold even conditioned on $\vec{\ell}\text{-Loc}$ for admissible $\vec{\ell}$, we only ever use it and therefore prove it unconditionally; the conditional version would follow mutatis mutandis.

For this purpose, we begin with a corollary of Lemma 3.6 which inverts that lemma to show that the position of the first particle with index less than or equal to k is stochastically below a geometric random variable.

Corollary 3.7. *Suppose $\vec{p}$ is ε -positively biased and $\vec{\ell} \in \mathcal{L}_n$. For $k \geq 1$, let*

$$X = \min\{s : \sigma(s) \leq k\}.$$

Then, the law of X under $\mu(\cdot \mid \vec{\ell}\text{-Loc})$ is stochastically below a geometric random variable with success probability $1 - (1 + \varepsilon)^{-k}$

Proof. Consider the event that $X > s$. This is equivalent to $\{\sigma(1) \notin [k]\} \cap \dots \cap \{\sigma([s]) \notin [k]\}$. Thus,

$$\mu(X > s \mid \vec{\ell}\text{-Loc}) = \prod_{i=1}^s \mu(\sigma(i) \notin [k] \mid \sigma(1) \notin [k], \dots, \sigma(i-1) \notin [k], \vec{\ell}\text{-Loc}).$$

By increasing the conditioning to fully reveal the particle assignments at locations $1, \dots, i-1$, then the law on the remainder of the locations $[i, n]$ is an ε -positively biased instance of the same problem, with all particles $[k]$ still available. Thus, by Lemma 3.5, each term above is equivalent to some $\mu'(\sigma(1) \notin [k] \mid \vec{\ell}'\text{-Loc})$ for some μ' corresponding to a ε -positively biased sequence $\vec{p}'$, and an $n-i$ -admissible $\vec{\ell}'$. In particular, by Lemma 3.6,

$$\mu(X > s \mid \vec{\ell}\text{-Loc}) \leq \frac{1}{(1 + \varepsilon)^{ks}},$$

which implies the claimed stochastic domination. $\square$

3.3 Typical configurations are logarithmically localized

We can now prove Lemma 2.2.

Proof of Lemma 2.2. We use a coupon collecting argument to bound the location r where $\sigma(r) = k$, showing that $r = k + O\left(\frac{1}{\varepsilon} \log\left(\frac{n}{\varepsilon}\right)\right)$ with probability at least $1 - n^{-C}$ for arbitrarily large C . Given k , let $X_0 = 0$ and $X_1, X_2, \dots, X_k$ be such that

$$X_s = \min\{r > X_{s-1} : \sigma(r) \leq k\}.$$

Note that $\sigma^{-1}(k) \leq X_k$. We claim that conditional on $\sigma([X_{s-1}])$, the random variable $Y_s = X_s - X_{s-1}$ is stochastically dominated by a geometric random variable with success probability $1 - \frac{1}{(1+\varepsilon)^{k+1-s}}$. Indeed, revealing $\sigma([X_{s-1}])$, the remaining particles to be assigned to $[X_{s-1}]^c$ have $k+1-s$ many of the letters of $[k]$. Relabeling that remainder by Lemma 3.5 and applying Corollary 3.7, we get the stochastic domination for Y_s .

We then use the following tail bound on sums of independent geometric random variables.

Lemma 3.8. *Let $Y_i \sim \text{Geometric}(a_i)$ be independent for $i \in [k]$. Then letting $S = \sum_{i=1}^k \frac{1}{a_i}$, for some universal constant C ,*

$$\mathbb{P}\left(\sum_{i=1}^k Y_i \geq S + u\right) \leq \exp\left(-C \min\left\{\frac{u^2}{\sum_{i=1}^k \frac{1-a_i}{a_i^2}}, u \min_i a_i\right\}\right).$$

Proof. The moment generating function of $Y \sim \text{Geometric}(a)$ is $\mathbb{E}[e^{tX}] = \frac{ae^t}{1-e^t(1-a)}$. We have by Taylor expansion that for $t < \frac{a}{2}$,

$$\mathbb{E}e^{tX - \frac{t}{a}} = e^{t(1-\frac{1}{a}) - \log(1-(e^t-1)(\frac{1-a}{a}))} = e^{O\left(t^2\left(\frac{1}{a}-1+\left(\frac{1}{a}-1\right)^2\right)\right)} = e^{O\left(t^2\left(\frac{1-a}{a^2}\right)\right)},$$

where the constant in the big- O is uniform in $(a_i)_i$ and k . Hence,

$$\mathbb{P}\left(\sum_{i=1}^k X_i \geq S + u\right) \leq \frac{\mathbb{E}e^{t\sum_{i=1}^k X_i}}{e^{t(S+u)}} \leq e^{O\left(t^2\sum_{i=1}^k \frac{1-a_i}{a_i^2}\right) - tu}.$$

Optimizing over $t \leq \frac{1}{2} \min a_i$ gives the lemma. $\square$

We can now complete the proof of the high probability properties of the stationary distribution stated in Lemma 2.2.

In our case, for $a_i = 1 - \frac{1}{(1+\varepsilon)^i}$ for $1 \leq i \leq k$, splitting into the cases $i \leq \log \frac{1}{\varepsilon}$ and $i > \log \frac{1}{\varepsilon}$, we get that the contribution to $\sum_{i=1}^k \frac{1-a_i}{a_i^2}$ is $O\left(\frac{1}{\varepsilon^2}\right)$ and $O\left(\frac{1}{\varepsilon}\right)$, respectively. Note

$$\begin{aligned} \sum_{i=1}^k \frac{1}{a_i} &= \sum_{i=1}^{\lfloor 1/\varepsilon \rfloor} \frac{1}{a_i} + \sum_{i=\lfloor 1/\varepsilon \rfloor+1}^k \frac{1}{a_i} = \sum_{i=1}^{\lfloor 1/\varepsilon \rfloor} O\left(\frac{1}{i\varepsilon}\right) + \sum_{i=\lfloor 1/\varepsilon \rfloor+1}^k \left[1 + O\left(\left(\frac{1}{1+\varepsilon}\right)^{i-\lfloor 1/\varepsilon \rfloor}\right)\right] \\ &= O\left(\frac{1}{\varepsilon} \log\left(\frac{1}{\varepsilon}\right)\right) + \left(k + O\left(\frac{1}{\varepsilon}\right)\right). \end{aligned}$$

Hence, by Lemma 3.8, there is C so that for any $0 < \delta < 1$,

$$\mu\left(\sigma^{-1}(k) \geq k + \frac{C}{\varepsilon} \log\left(\frac{1}{\varepsilon\delta}\right)\right) \leq \frac{\delta}{2}.$$

For $k' = k - \frac{C}{\varepsilon} \log\left(\frac{1}{\varepsilon\delta}\right)$ for appropriate C , by a complimentary stochastic domination argument using Lemma 3.5 and Corollary 3.7, we have

$$\begin{aligned} \mu(\sigma^{-1}(k) > k') &\geq \prod_{i=1}^{k'} \mu(\sigma(i) \in [k-1] \mid \sigma(1) \in [k-1], \dots, \sigma(i-1) \in [k-1]) \\ &\geq \prod_{i=1}^{k'} \left(1 - \frac{1}{(1+\varepsilon)^{k-i}}\right) \geq 1 - \sum_{i=1}^{k'} \frac{1}{(1+\varepsilon)^{k-i}} \geq 1 - \sum_{i=k-k'}^{\infty} \frac{1}{(1+\varepsilon)^i} \geq 1 - \frac{\delta}{2}. \end{aligned}$$

By a union bound,

$$\mu\left(|\sigma^{-1}(k) - k| \geq \frac{C}{\varepsilon} \log\left(\frac{n}{\varepsilon\delta}\right)\right) \leq \delta.$$

Taking δ to be an arbitrarily small polynomial in n concludes the proof of Lemma 2.2. $\square$

3.4 Spatial mixing for localized configurations

It remains to use a coupling argument to establish Lemma 3.3. As described in the proof sketch, this will rely on the notion of disconnecting points which isolate influence from the configuration on one side of the point from the other. Recall from Definition 2.6, for a permutation $\sigma \in S_n$ and a position $k \in [n]$, we refer to position k as *disconnecting* if $\sigma([k]) = [k]$ as unordered sets.

The following lemma shows that disconnecting points are common in intervals of length at least $n_0 = e^{O(1/\varepsilon)}$. A simpler form of the following lemma was stated in Lemma 2.7.

Lemma 3.9. *The following hold for all ε -positively biased and all $\vec{\ell} \in \mathcal{L}_n$.*

1. *For any $k \in [n]$, $\mu(k \text{ disconnecting for } \sigma \mid \sigma \in \vec{\ell}\text{-Loc}) \geq e^{-O(1/\varepsilon)}$.*
2. *Let $A^c = [i] \cup [n - j + 1, n]$. For $k \in [i + \ell_{\max}^-, n - j - \ell_{\max}^+]$ and any $b : A^c \rightarrow [n]$ that is the restriction of some $\sigma \in \vec{\ell}\text{-Loc}$,*

$$\mu(k \text{ disconnecting for } \sigma \mid \sigma \in \vec{\ell}\text{-Loc}, \sigma(A^c) = b) \geq e^{-O(1/\varepsilon)}.$$

Proof. We prove the lemma item by item.

1. The probability we are interested in can be written as

$$\prod_{i=1}^k \mu(\sigma(i) \leq k \mid \sigma \in \vec{\ell}\text{-Loc}; \sigma(1) \leq k, \dots, \sigma(i-1) \leq k).$$

By revealing $\sigma(1), \dots, \sigma(i-1)$ and applying Lemma 3.5, each of the terms in the product, Lemma 3.6 are lower bounded by $1 - \frac{1}{(1+\varepsilon)^{k+1-i}}$, so that

$$\mu(k \text{ disconnecting for } \sigma \mid \sigma \in \vec{\ell}\text{-Loc}) \geq \prod_{i=1}^k \left(1 - \frac{1}{(1+\varepsilon)^{k+1-i}}\right) = e^{-O(1/\varepsilon)}.$$

2. Using the notation of Lemma 3.5, by that lemma $(R_b)_* \mu$ is a measure on biased permutations that satisfies ε -positive bias and monotonicity. Applying item 1, if we let $\sigma' = R_b \sigma$, we obtain

$$((R_b)_* \mu)(k - i \text{ disconnecting for } \sigma' \mid \sigma' \in R_b \vec{\ell}\text{-Loc}) \geq e^{-O(1/\varepsilon)}.$$

It remains to show that if $k - i$ is disconnecting for $\sigma' = R_b \sigma$ and $\sigma(A^c) = b$, then k is disconnecting for σ . To show this claim, note that we have $\sigma([i]) = b([i]) \subseteq [i + \ell_{\max}^-] \subseteq [k]$ and by $k - i$ being disconnecting for σ' ,

$$[k - i] \supseteq (\sigma')^{-1}([k - i]) = \sigma^{-1}(r_b([k - i])) - i,$$

so $\sigma^{-1}(r_b([k - i])) \subseteq [k]$. Note $r_b([k - i]) \subseteq [k]$ because $k \leq n - j - \ell_{\max}^+$. Finally, note that $r_b([k - i])$ and $b([i])$ are disjoint subsets of $[k]$ and the union has k elements, so the union equals $[k]$, and we obtain $\sigma^{-1}([k]) = k$. $\square$

We can now use disconnecting points to prove the spatial mixing lemma.

Proof of Lemma 3.3. We construct a coupling of the distributions. The idea is to sample independently and look for common disconnecting points between the two configurations, revealing from the outside-inwards. On the one hand, we have an independent attempt every $\ell_{\max}$ indices, and on the other upon encountering a pair of common disconnecting points, the marginal distribution on the interval between them is identical.

Formally, we proceed as follows. First, by rounding down r and adjusting the constant as necessary, it suffices to consider $r = r'\ell_{\max}$. Consider independently sequentially sampling,

$$\begin{aligned}\sigma(i+k) \mid \sigma(A^c) = \eta, \sigma([i+1, i+k-1]), \sigma \in \vec{\ell}\text{-Loc}, \\ \bar{\sigma}(i+k) \mid \bar{\sigma}(A^c) = \bar{\eta}, \bar{\sigma}([i+1, i+k-1]), \bar{\sigma} \in \vec{\ell}\text{-Loc},\end{aligned}$$

until we arrive at a k such that $k = k'\ell_{\max}$, $k' \leq r'$ such that $i+k$ is disconnecting for both σ and $\bar{\sigma}$ (and declare failure otherwise). Note that this event is measurable with respect to the sigma-algebras generated by $\sigma([i+k]), \bar{\sigma}([i+k])$. Conditioned on $\sigma(A^c) = \eta$, $\sigma(i+1, i+k'\ell_{\max})$, and $\sigma \in \vec{\ell}\text{-Loc}$, the probability that $i+(k'+1)\ell_{\max}$ is disconnecting is at least $\rho := e^{-O(1/\varepsilon)}$ by item (2) of Lemma 3.9. Thus, the probability that we fail to find such a disconnecting point for both σ and $\bar{\sigma}$ is $(1-\rho^2)^{k'} = e^{-e^{-O(1/\varepsilon)}r/\ell_{\max}}$.

Supposing that $i+k$ where $k = k'\ell_{\max}$ is disconnecting for both σ and $\bar{\sigma}$, we now independently sequentially sample in reverse,

$$\begin{aligned}\sigma(n-j+1-m) \mid \sigma(A^c) = \eta, \sigma([i+1, i+k] \cup [n-j+2-m, n-j]), \sigma \in \vec{\ell}\text{-Loc} \\ \bar{\sigma}(n-j+1-m) \mid \bar{\sigma}(A^c) = \bar{\eta}, \bar{\sigma}([i+1, i+k] \cup [n-j+2-m, n-j]), \bar{\sigma} \in \vec{\ell}\text{-Loc}.\end{aligned}$$

Define m be reverse-disconnecting for σ if $\sigma([m, n]) = [m, n]$ as an unordered set. The same proof as in Lemma 3.9 and as above shows that the probability we fail to find a reverse-disconnecting point $n-j+1-m$ where $m = m'\ell_{\max}$, $m' \leq r'$ for both σ and $\bar{\sigma}$ with probability $e^{-e^{-O(1/\varepsilon)}r/\ell_{\max}}$. Note by the upper bound assumption on r , there is a buffer of size at least $\ell_{\max}$ as necessary for Lemma 3.9.

If we have found both a disconnecting point $i+k$ and reverse-disconnecting point $n-j+1-m$, then we have $[i+k] = \sigma([i+k]) = \bar{\sigma}([i+k])$, and $[n-j+1-m, n] = \sigma([n-j+1-m, n]) = \bar{\sigma}([n-j+1-m, n])$. Letting $B = [n] \setminus ([k] \cup [n-j+1-m, n])$, we note $B \supseteq A_r$ and the distribution of $\sigma(B)$, $\bar{\sigma}(B)$ conditioned on the set values (and $\sigma, \bar{\sigma} \in \vec{\ell}\text{-Loc}$) is the same.

Finally, note that if either $i = 0$ or $j = 0$, then there are no boundary conditions on that side, and it suffices to find a disconnecting point on the opposite side. $\square$

4 Burn-in via dominating exclusion processes

Our aim in this section is to use dominating ASEP processes to establish the following burn-in type of estimates to get to $\ell_0\text{-Loc}$ where $\ell_0 = C_0(\varepsilon) \log n$.

Our main result in this section is that for $\vec{\ell} \in \mathcal{L}_n$, the adjacent transposition chain $\mathcal{M}_{\text{AT}}$ restricted to $\vec{\ell}\text{-Loc}$ for $\vec{\ell} \in \mathcal{L}_n$ reaches $\ell_0\text{-Loc}$ and stays there for arbitrary long polynomial timescales. In particular, this holds for the unrestricted dynamics.

Lemma 4.1. *Consider $\mathcal{M}_{\text{AT}}$ with parameters $\vec{p}$ satisfying ε -positive bias. There exist $C_0(\varepsilon) = O(1/\varepsilon)$ and n_0 depending on ε such that for all $n \geq n_0$ and any $\vec{\ell} \in \mathcal{L}_n$, $\mathcal{M}_{\text{AT}}$ restricted to $\vec{\ell}\text{-Loc}$ is such that for $t \geq C_0(\varepsilon)n^2$ the chain is in $(C_0(\varepsilon)\log n)\text{-Loc}$ except with probability $e^{-\Omega(1/\varepsilon)}n^{-10}$.*

Note, Lemma 4.1 is a more general form of Lemma 2.3. We also need a form of this burn-in statement for a block version of the dynamics where in each time-step, $B \in \{B_W, B_E\}$ is chosen randomly, and the configuration $\sigma(B)$ is resampled conditionally on $\sigma(B^c)$.

Lemma 4.2. *There exists $C_1 > 0$, and for any ε , there exist $C_0(\varepsilon) = O(1/\varepsilon)$ and $n_0 = O(\log(1/\varepsilon)/\varepsilon)$ such that the following holds for all $\vec{p}$ satisfying ε -positive bias and $\vec{\ell} \in \mathcal{L}_n$. The block dynamics restricted to $\vec{\ell}\text{-Loc}$ is such that for any $n \geq n_0$ and any $t \geq C_1 \log n$, the chain is in $(C_0(\varepsilon)\log n)\text{-Loc}$ except with probability $e^{-\Omega(1/\varepsilon)}n^{-10}$.*

We prove these burn-in style results by stochastic comparisons of various particle counts to the Asymmetric Exclusion Process (ASEP), for which the mixing time and convergence behavior are well understood.

4.1 The asymmetric exclusion process

The k -particle asymmetric exclusion process (ASEP) is a Markov chain on the state space of assignments $\binom{[n]}{k} := \{Y \in \{0, 1\}^n : \sum_{i=1}^n Y(i) = k\}$, where the value 1 at a site means a particle is present, and a 0 means the site is empty.

We describe the process in discrete time. At each step, the ASEP chain with parameter $q \in (\frac{1}{2}, 1)$ picks an edge (pair of adjacent sites) $(i, i+1)$ uniformly at random, and exchanges the states on the sites with probabilities

$$P((0, 1) \rightarrow (1, 0)) = q \quad \text{and} \quad P((1, 0) \rightarrow (0, 1)) = 1 - q.$$

(Notice that if both sides are occupied or both sides are empty, the interchange does nothing so the probability does not matter.) We denote the stationary distribution by $\nu_{[n]}^{(k)}$. For an extensive background on the asymmetric exclusion process, see Section III of [Lig85]. Throughout this section, we will take q given by the relation

$$\frac{q}{1-q} = 1 + \varepsilon, \tag{8}$$

for which the constant- ε -positively-biased adjacent transposition shuffle is directly related to the q -ASEP.

4.2 Coupling the biased adjacent transposition shuffle and ASEP

In the below, we make some observations on monotonicity for the ASEP process itself, and monotonicity between the projected $\mathcal{M}_{\text{AT}}$ (see Definition 4.5) and a k -particle ASEP. We define a partial order on $\binom{[n]}{k}$ where $\eta \leq \eta'$ (read: η is to the left of η') if for every $r \leq n$,

η has more particles on $[r]$ than η' does. Such monotonicity properties are standard in the literature (see e.g., Section 23 of [LP17]); we include proofs for our specific setting for completeness. We begin with an observation about the potential ways such a monotonicity can be violated in one step.

Observation 4.3. *Suppose $Y, Y' \in \binom{[n]}{k}$ with $Y \leq Y'$. If they both (possibly) make a swap on an edge $e = (i, i+1)$ to generate Y_1, Y'_1 , then $Y_1 \leq Y'_1$ unless at least one of the following occurred:*

1. $(Y(i), Y(i+1)) = (1, 0)$ and $(Y'(i), Y'(i+1)) \in \{(1, 0), (0, 1)\}$ and $(Y_1(i), Y_1(i+1)) = (0, 1)$ while $(Y'_1(i), Y'_1(i+1)) = (1, 0)$.
2. $(Y(i), Y(i+1)) \in \{(0, 1), (1, 0)\}$ and $(Y'(i), Y'(i+1)) = (0, 1)$ and $(Y_1(i), Y_1(i+1)) = (0, 1)$ while $(Y'_1(i), Y'_1(i+1)) = (1, 0)$.

Proof. We consider the immediate cases first. If both Y and Y' are $(0, 0)$ or $(1, 1)$ or don't execute the swap at $(i, i+1)$ (regardless of their configuration), then evidently the ordering cannot change. Likewise, if they both are $(1, 0)$ and both swap, or both are $(0, 1)$ and both swap, the monotonicity cannot change. Finally, if Y is making a $(0, 1) \rightarrow (1, 0)$ swap, or Y' is making a $(1, 0) \rightarrow (0, 1)$ swap, then the ordering is only improving.

We now consider the remaining cases that are not in items 1–2 in the observation:

1. Y is making a $(1, 0) \rightarrow (0, 1)$ swap, while Y' is $(0, 0)$;
2. Y is making a $(1, 0) \rightarrow (0, 1)$ swap while Y' is $(1, 1)$;
3. Y' is making a $(0, 1) \rightarrow (1, 0)$ swap while Y is $(0, 0)$;
4. Y' is making a $(0, 1) \rightarrow (1, 0)$ swap while Y is $(1, 1)$.

For these cases, let $N_i = |\{j \leq i : Y(j) = 1\}|$ and define N'_i analogously. The inequality $Y \leq Y'$ is equivalent to $N_i \geq N'_i$ for all $i \in [n]$.

In the first case, since $Y \leq Y'$, one has $N_{i-1} \geq N'_{i-1}$ while $N_i = N_{i-1} + 1$ and $N'_i = N'_{i-1}$, so $N_i \geq N'_i + 1$ and so decrementing N_i while incrementing N_{i+1} and leaving N'_i, N'_{i+1} unchanged retains the ordering.

In the second case, since $Y \leq Y'$, one has $N_{i+1} \geq N'_{i+1}$, while $N_{i+1} = N_i$ and $N'_{i+1} = N'_i + 1$, so $N_i \geq N'_i + 1$ and decrementing N_i while incrementing N_{i+1} and leaving N'_i, N'_{i+1} unchanged retains the ordering.

In the third case, since $Y \leq Y'$, one has $N_{i+1} \geq N'_{i+1}$ but $N_{i+1} = N_i$ while $N'_{i+1} = N'_i + 1$, so $N_i \geq N'_i + 1$; thus incrementing N'_i while decrementing N'_{i+1} retains the ordering.

In the fourth case, since $Y \leq Y'$, one has $N_{i-1} \geq N'_{i-1}$ but $N_i = N_{i-1} + 1$ while $N'_i = N'_{i-1}$, so $N_i \geq N'_i + 1$; thus incrementing N'_i while decrementing N'_{i+1} retains the ordering. $\square$

The following lemma states that the ASEP process is monotone in this partial order.

Lemma 4.4. *If Y_t and $\tilde{Y}_t$ are k -particle ASEP's with $Y_0 \leq \tilde{Y}_0$, then there exists a coupling such that for all $t \geq 0$, one has $Y_t \leq \tilde{Y}_t$.*

Proof. Consider the following one-step coupling. If Y_s, Y'_s have k particles each and $Y_s \leq Y'_s$, then to generate Y_{s+1}, Y'_{s+1} select the same edge $e = (i, i+1)$ to update, and use a single uniform random variable $U \sim \text{Unif}[0, 1]$ to generate (Y_{s+1}, Y'_{s+1}) as follows: if either one's state is $(1, 0)$ or $(0, 1)$, set it equal to $(1, 0)$ if and only if $U \leq q$.

By Observation 4.3, we only need to consider the possibilities listed in items 1–2 therein as ways that monotonicity could be violated by this update. But the definition of the coupling ensures that if both $(Y_s(i), Y_s(i+1))$ and $(Y'_s(i), Y'_s(i+1))$ are in $\{(0, 1), (1, 0)\}$ then at time $s+1$ they agree on locations $i, i+1$ and so neither case occurs. $\square$

The next lemma says that projections of an ε -positively biased adjacent transposition shuffle are stochastically to the left of the q -ASEP with $q/(1-q) = 1 + \varepsilon$. To properly define this stochastic comparison, we define the following projections which track the locations of the first k particles.

Definition 4.5. For each $0 < k < n$, define the function $\eta_k : S_n \rightarrow \binom{[n]}{k}$ which takes a configuration $\sigma \in S_n$ and outputs $\eta_k(i) = \mathbf{1}\{\sigma(i) \leq k\}$.

Lemma 4.6. If $\vec{p}$ is ε -positively biased, then the chain $\sigma_t \sim \mathcal{M}_{\text{AT}}$ on S_n is such that for each k , the process $\eta_k(\sigma_t)$ is stochastically to the left of the q -ASEP process with parameter $q/(1-q) = 1 + \varepsilon$, initialized from $\eta_k(\sigma_0)$.

Proof. We couple $(\sigma_t)_{t \geq 0}$ to $n-1$ different ASEP process $Y_t^{(1)}, \dots, Y_t^{(n)}$ where $Y_t^{(k)}$ is a k -particle ASEP initialized from $Y_0^{(k)} = \eta_k(\sigma_0)$. The coupling operates by choosing the same edge to update at each time step, and using the same uniform random variable $U_{t+1} \sim \text{Unif}[0, 1]$ to generate the update. The latter statement means that we update as follows: suppose $\sigma_t(i) < \sigma_t(i+1)$ and $(Y_{t+1}^{(k)}(i), Y_{t+1}^{(k)}(i+1)) \in \{(0, 1), (1, 0)\}$; then

$$\begin{aligned} (\sigma_{t+1}(i), \sigma_{t+1}(i+1)) &= (\sigma_t(i), \sigma_t(i+1)) && \text{if and only if } U_{t+1} \leq p_{\sigma_t(i), \sigma_t(i+1)}, \\ (Y_{t+1}^{(k)}(i), Y_{t+1}^{(k)}(i+1)) &= (1, 0) && \text{if and only if } U_{t+1} \leq q. \end{aligned}$$

If $\sigma_t(i) > \sigma_t(i+1)$ instead, then set

$$(\sigma_{t+1}(i), \sigma_{t+1}(i+1)) = (\sigma_t(i+1), \sigma_t(i)) \quad \text{if and only if } U_{t+1} \leq p_{\sigma_t(i+1), \sigma_t(i)}.$$

In this manner, one has the property that if $(\eta^{(k)}(\sigma_t)(i), \eta^{(k)}(\sigma_t)(i+1))$ and $(Y_t^{(k)}(i), Y_t^{(k)}(i+1))$ are both in $\{(0, 1), (1, 0)\}$, then

- if the former is $(1, 0)$, since $p_{\sigma_t(i), \sigma_t(i+1)} \geq q$ by ε -positive bias,

$$(Y_{t+1}^{(k)}(i), Y_{t+1}^{(k)}(i+1)) = (1, 0) \implies (\eta^{(k)}(\sigma_{t+1})(i), \eta^{(k)}(\sigma_{t+1})(i+1)) = (1, 0);$$

- if the former is $(0, 1)$, since $p_{\sigma_t(i+1), \sigma_t(i)} \geq q$ by ε -positive bias,

$$(\eta^{(k)}(\sigma_{t+1})(i), \eta^{(k)}(\sigma_{t+1})(i+1)) = (0, 1) \implies (Y_{t+1}^{(k)}(i), Y_{t+1}^{(k)}(i+1)) = (0, 1).$$

By Observation 4.3 and the fact that the monotonicity held at time t , the above two items imply it also holds at time $t + 1$. $\square$

Taking $t \rightarrow \infty$ in the above lemma, we also deduce the domination relation for the stationary distributions of the first k coordinates in the shuffle and k -particles ASEP.

Corollary 4.7. *Suppose $\vec{p}$ is ε -positively biased. If $\sigma \sim \mu$, then for every $k \leq n$, one has $\eta_k(\sigma) \preceq Y_\infty^{(k)}$ where $Y_\infty^{(k)}$ is drawn from the q -ASEP stationary distribution with k particles.*

The same inequalities hold for the restricted dynamics where we restrict to $\vec{\ell}$ -Loc for a vector $\vec{\ell}$ in $\mathcal{L}_n$. Indeed, that will be a direct consequence of the ASEP's own monotonicity.

Lemma 4.8. *Suppose $\vec{\ell}$ -Loc for $\vec{\ell} \in \mathcal{L}_n$. If σ_t is the adjacent transposition chain restricted to $\vec{\ell}$ -Loc, for any k the processes $\eta_k(\sigma_t)$ are stochastically to the left of the k -particle ASEP $Y_t^{(k)}$ started from $\eta_k(\sigma_0)$. It also holds that $\eta_k(\sigma) \preceq Y^{(k)}$ where σ is drawn from the restricted stationary distribution and $Y^{(k)}$ is drawn from the stationary distribution of q -ASEP.*

Proof. Consider the same coupling as in the proof of Lemma 4.6, except that rejections override the coupling. By Claim 3.4, the only moves that get rejected are ones where $\sigma_t(v) = i$ and $\sigma_t(v+1) = j$ for $i < j$. In the coupling to the k -particle ASEP, $(Y_t^{(k)}(v), Y_t^{(k)}(v+1))$ has to be either $(0, 0)$, $(1, 1)$, or $(1, 0)$. This means that a move $(0, 1) \rightarrow (1, 0)$ is never rejected. By the monotonicity of the ASEP process per Lemma 4.4, this implies that the restricted dynamics $\eta_k(\sigma_t)$ stays stochastically below the ASEP $Y_t^{(k)}$. $\square$

4.3 Analysis of the ASEP

The domination of the biased adjacent transposition by n q -ASEP processes implies that if the coupled ASEP processes have mixed, their stationary distributions can be used to control the locations of the particles in the shuffle. The paper [BBHM05] established that the discrete-time mixing time of q -ASEP is $O(n^2)$, and [LL19] proved cutoff.

Lemma 4.9 ([BBHM05], with [LL19] for the ε dependencies). *The mixing time of the k -particle $(\frac{1}{2} + \varepsilon')$ -ASEP process is at most $\frac{n^2}{\varepsilon'}(1 + o(1))$ for every $k \leq n$, asymptotically as $n \rightarrow \infty$, and its inverse spectral gap is $O(\frac{n}{\varepsilon'})$ for all n .⁴*

Given the stochastic comparisons of Lemma 4.8 and the mixing time bound of Lemma 4.9, we can use the stationary distribution of the ASEP to bound the behavior of the biased adjacent transposition shuffle using stationary estimates of the ASEP for certain monotone events. The stationary distribution of the q -ASEP is carefully characterized in [Lig85]; we need the following tail bound on the right-most particle.

⁴Note this ε' is such that $\frac{\frac{1}{2} + \varepsilon'}{\frac{1}{2} - \varepsilon'} = 1 + \varepsilon$. When $\varepsilon, \varepsilon'$ are small, they differ by constant factors.

Lemma 4.10. *Let $Y^{(k)}$ be drawn from the k -particle q -ASEP stationary distribution with $q > \frac{1}{2}$ as in (8) ($\frac{q}{1-q} = 1 + \varepsilon$). Then for every $k \leq n$, for $r \geq C_q$,*

$$\nu_{[n]}^{(k)}(\max\{v : Y^{(k)}(v) = 1\} \geq k + r) \leq e^{-\varepsilon' r/4}$$

when $r \geq \frac{4}{\varepsilon'} \log\left(\frac{2}{\varepsilon'}\right)$.

Proof. Consider the distribution $\bar{\mu}$ on S_n with

$$p_{i,j} = \begin{cases} q, & i \leq k < j \\ 1, & i, j \in \{1, \dots, k\} \text{ or } i, j \in \{k+1, \dots, n\} \end{cases}.$$

Recall the projection $\eta_k : S_n \rightarrow \binom{[n]}{k}$ defined in Definition 4.5. Because the particles $\{1, \dots, k\}$ must be in order and $\{k+1, \dots, n\}$ must also be in order, the preimage of each $Y \in \binom{[n]}{k}$ is a single point, and

$$(\eta_k)_* \bar{\mu}(Y) = q^{|\{1 \leq i < j \leq n : Y(i)=0, Y(j)=1\}|} (1-q)^{|\{1 \leq i < j \leq n : Y(i)=1, Y(j)=0\}|}.$$

It is easy to check that this distribution satisfies detailed balance with respect to the ASEP moves, and hence this is the stationary distribution of ASEP.

Now letting ε be as in (8), by Lemma 3.6 applied to the probability conditioned on specific values of $\sigma^{-1}(1), \dots, \sigma^{-1}(i-1)$ and then averaging,

$$\bar{\mu}(\sigma^{-1}(i) < k + r \mid \sigma^{-1}(1), \dots, \sigma^{-1}(i-1) < k + r) \geq 1 - \left(\frac{1}{1+\varepsilon}\right)^{k+r-i},$$

noting that we are asking for the probability that $\sigma^{-1}(i)$ is one of the first $k+r-i$ available indices. Therefore,

$$\begin{aligned} \bar{\mu}(\sigma^{-1}(1), \dots, \sigma^{-1}(k) < k + r) &\geq \prod_{i=1}^k \left(1 - \left(\frac{1}{1+\varepsilon}\right)^{k+r-i}\right) \geq \prod_{i=r}^{\infty} \left(1 - \left(\frac{1}{1+\varepsilon}\right)^i\right) \\ &\geq 1 - \sum_{i=r}^{\infty} \left(\frac{1}{1+\varepsilon}\right)^i = 1 - \left(\frac{1}{1+\varepsilon}\right)^r \frac{1+\varepsilon}{\varepsilon} \end{aligned}$$

Hence

$$\begin{aligned} \mathbb{P}(\max\{v : Y^{(k)}(v) = 1\} \geq k + r) &\leq \left(\frac{1}{1+\varepsilon'}\right)^r \frac{1+\varepsilon'}{\varepsilon'} \\ &\leq e^{-\varepsilon' r/4} \cdot e^{-\varepsilon' r/4} \frac{1+\varepsilon'}{\varepsilon'} \leq e^{-\varepsilon' r/4}, \end{aligned}$$

where $\varepsilon' = \min\{\varepsilon, 1\}$ and $r \geq \frac{4}{\varepsilon'} \log\left(\frac{2}{\varepsilon'}\right)$. □

Since we also want a burn-in estimate for block dynamics, not just the adjacent transposition chain, we also require a mixing time estimate for q -ASEP block dynamics.

We define a block-update dynamics for the ASEP, and argue that the mixing time of this block dynamics of the ASEP is order 1 and moreover, that its inverse spectral gap is order 1.

Consider the block dynamics with the following blocks (West and East):

$$B_W = \{1, \dots, \lceil 2n/3 \rceil\} \quad B_E = \{\lfloor n/3 \rfloor, \dots, n\}.$$

In each step, randomly pick $B \in \{B_W, B_E\}$ to update, and when block B is selected to update, the configuration on B is replaced by a stationary sample of the q -ASEP on B conditional on the configuration on B^c . Note that this distribution is simply that of a q -ASEP on B with the number of particles equal to k minus the number on B^c .

Lemma 4.11. *Consider the block dynamics for q -ASEP on an interval of length n with $k \leq n$ particles. There exists a constant C_1 independent of ε such that for all $n \geq C_1 \log(1/\varepsilon)/\varepsilon$, the mixing time and inverse spectral gap of the block dynamics is at most C_1 .*

Proof. We first consider the case $k \leq n/2$. By Lemma 4.10 and the lower bound on n , the stationary distribution of k -particle q -ASEP on $[n]$ and k -particle q -ASEP on B_W it suffices to couple the chain $Y_t^{(k)}$ from any initialization $Y_0^{(k)}$ to that stationary distribution $\nu_{[\lceil 2n/3 \rceil]}^{(k)}$.

We claim that so long as $n \geq \frac{C \log(1/\varepsilon)}{\varepsilon}$ for a uniform constant C , then the mixing time of the block dynamics is an ε -independent constant.

1. If the East block is updated first, then by Lemma 4.10 and the lower bound on n , with probability at least 0.99, the number of particles k_E that remain in $B_E \setminus B_W$ is at most $n/2 - n/3 + \delta n < n/(5.5)$ (for $\delta = 1/100$).
2. If the West block is updated next, then by Lemma 4.10 and the lower bound on n , with probability at least 0.99 at most $n/2 - k_E - n/3 + \delta n$ will remain in $B_E \cap B_W$ and therefore at most $n/(5.5) + \delta n < n/5$ many particles will remain in B_E .
3. If the East block is updated next, by Lemma 4.10 and the lower bound on n , with probability at least 0.99, all particles will be in B_W ;
4. If the West block is updated next, then the distribution will be exactly equal to the stationary q -ASEP distribution conditioned on all particles being in B_W .

The above bounds the mixing time on a segment of length $n \geq (C \log(1/\varepsilon))/\varepsilon$ for a uniform constant C by the minimum time by which with probability ≥ 0.99 , we get the above four consecutive updates. This therefore gives a uniform constant bound on the block mixing time (time 100 should for instance suffice).

For $k \geq n/2$, we argue by symmetry: consider the mapping which flips 0 and 1, so that the particles are following q -ASEP where the bias is to the right instead of to the left, and where the number of particles is at most $n/2$. The argument then proceeds exactly symmetrically to the above. $\square$

4.4 Proofs of burn-in consequences

We are now able to combine the results of the previous two subsections to deduce the burn-in results for the adjacent transposition and block chains for the biased adjacent transposition shuffle.

Proof of Lemma 4.1. By Lemma 4.8, the k th particle of σ_t is to the left of the rightmost particle of the k -particle ASEP, $Y_t^{(k)}$. By the mixing time and inverse spectral gap bounds of Lemma 4.9, for each k , for large enough n , for all $t \geq T_1 := C(\varepsilon)n^2 + rn \log n$ for $C(\varepsilon) = O(\frac{1}{\varepsilon})$ and constant r , one has

$$\max_{y_0} \|\mathbb{P}_{y_0}(Y_t^{(k)} \in \cdot) - \nu^{(k)}\|_{\text{TV}} \leq n^{-12}.$$

In particular, combined with Lemma 4.10 and ensuring $n \geq n_0 = \Omega(\frac{\log(1/\varepsilon)}{\varepsilon})$ and $\ell = \Omega(\frac{\log(1/\varepsilon)}{\varepsilon})$, we have for all $t \geq T_1$ that

$$\mathbb{P}(\exists k : \max\{v : Y_t^{(k)}(v) = 1\} \geq k + \ell) \leq n(n^{-12} + e^{-\varepsilon\ell/4}).$$

By the stochastic domination between $Y_t^{(k)}$ and σ_t , this implies

$$\mathbb{P}(\exists k : \sigma_t^{-1}(k) \geq k + \ell_0) \leq n(n^{-12} + e^{-c_\varepsilon\ell_0})$$

which is at most $n^{-10}/2$ for $\ell_0 = C'(\varepsilon) \log n$ with a large enough constant $C'(\varepsilon) = O(\frac{1}{\varepsilon})$. To localize in the other direction, consider ASEP where holes are swapped with particles, and carry out the identical argument symmetrically to also obtain

$$\mathbb{P}(\exists k : \sigma_t^{-1}(k) \leq k - \ell_0) \leq \frac{1}{2}n^{-10}.$$

Together with a union bound, these two inequalities imply the lemma. $\square$

Proof of Lemma 4.2. By the stochastic domination of the stationary distributions from Lemma 4.8, the projections of the block dynamics via $\eta^{(k)}$ are each stochastically dominated by the k -particle ASEP block dynamics.

By the spectral gap bound of Lemma 4.11, there is a constant C_1 not depending on ε and $n_0(\varepsilon) = O(\log(1/\varepsilon)/\varepsilon)$ such that as long as $n \geq n_0(\varepsilon)$, after $C_1 \log n$ many steps of block dynamics, for any $k \leq n$, the TV-distance of the ASEP block dynamics is within n^{-12} from its stationary distribution. Following the same reasoning as for the single site dynamics, that implies that

$$\mathbb{P}(\exists k : \sigma_t^{-1}(k) \notin [k - \ell_0, k + \ell_0]) \leq 2n(n^{-12} + e^{-\varepsilon\ell_0/4}) \leq n^{-10},$$

implying the claimed burn-in time. $\square$

5 Polynomial mixing time

In this section we prove a more detailed version of the polynomial mixing time bound stated in Theorem 2.8.

Proposition 5.1. *There exists $C > 0$ such that for all $\varepsilon > 0$, there exists $C'(\varepsilon) = e^{\tilde{O}(1/\varepsilon^2)}$, such that if $\vec{\ell} \in \mathcal{L}_n$, then $\mathcal{M}_{\text{AT}}$ restricted to $\vec{\ell}\text{-Loc}$ has mixing time at most $C'(\varepsilon)n^C$.*

As a corollary, by taking $\vec{\ell} = (-\infty, \infty)_i$, which is evidently admissible, this gives polynomial mixing of the original unrestricted chain as stated in Theorem 2.8.

The proof of Proposition 5.1 combines the burn-in estimates of Section 4 with the spatial mixing estimates amongst burnt-in configurations from Section 3 to show that the adjacent transposition chain restricted to any admissible $\vec{\ell}\text{-Loc}$ has polynomial mixing time.

5.1 Block dynamics

The block dynamics is a standard tool used in spin system dynamics for upper bounding mixing times using spatial self-similar structure of the process. It is typically phrased in the context of spin system dynamics.

Let us first define what we mean by block dynamics for the biased adjacent transposition shuffle dynamics. In this subsection we are using N to denote an arbitrary length of the 1D chain.

Definition 5.2. *Let $\mathcal{B} = \{B_1, \dots, B_\ell\}$ denote any collection of subsets of $[N]$ where $\bigcup_i B_i = [N]$. The heat-bath block dynamics is the discrete-time Markov chain which, from a configuration X_t , chooses a block B_i with probability $|B_i|/(\sum_i |B_i|)$ from $\mathcal{B}$, and resamples the entire configuration $X_t(B_i)$ according to the stationary distribution conditioned on $X_t(B_i^c)$.*

In the context of the biased adjacent transposition shuffle and its stationary distribution μ , as long as the edge-sets of the blocks cover the edges of $[N]$, then the block dynamics is ergodic and reversible with respect to μ . Ergodicity follows since the block dynamics can simulate any adjacent transposition, and reversibility is straightforward from the definition of heat-bath.

The block dynamics are important for analyzing spin systems because they give us the ability to upper bound the inverse gap of the adjacent transposition chain on $[N]$ by the worst inverse gap of the adjacent transposition chain on a block, times the inverse gap of the block dynamics chain. More precisely, the following decomposition result is standard in the analysis of the Glauber dynamics for spin systems, see e.g., [Mar04, Proposition 3.4]. Our setting does not exactly fit into these spin system setting because two vertices update their assignments simultaneously, but the proof is essentially the same. We include a proof for completeness.

Proposition 5.3. *Let $\chi(\mathcal{B}) = \max_{j \in [N]} |\{i : B_i \ni j\}|$ denote the maximum coverage of a position in $[N]$. For any collection of blocks $\mathcal{B}$, we have the following:*

$$\gamma(\mathcal{M}_{\text{AT}}([N])) \geq \chi(\mathcal{B})^{-1} \cdot \gamma(\mathcal{M}_{\mathcal{B}}) \cdot \min_{\eta, i} \gamma(\mathcal{M}_{\text{AT}}(B_i^\eta)),$$

where $\gamma(\mathcal{M}_{\text{AT}}([N]))$ is the spectral gap of $\mathcal{M}_{\text{AT}}$ on $[N]$, $\gamma(\mathcal{M}_{\mathcal{B}})$ is the spectral gap of the block dynamics defined by $\mathcal{B}$, and $\gamma(\mathcal{M}_{\text{AT}}(B_i^\eta))$ is the spectral gap of $\mathcal{M}_{\text{AT}}$ restricted to the set B_i with configuration η on $B_i^c = [N] \setminus B_i$.

Proof. By definition of the block dynamics transition rates and their Dirichlet form,

$$\mathcal{E}_{\mathcal{B}}(f, f) \leq \frac{1}{2} \sum_{\sigma, \sigma'} \mu(\sigma) (f(\sigma) - f(\sigma'))^2 \frac{1}{|\mathcal{B}|} \sum_i \frac{|B_i|}{\sum_j |B_j|} \mu_{B_i}^\sigma(\sigma') \mathbf{1}\{\sigma(B_i^c) = \sigma'(B_i^c)\}$$

where $\mu_{B_i}^\sigma$ denotes the distribution on B_i conditional on equaling σ on B_i^c . By the Markov property for the distribution, $\mu(\sigma) = \mu(\sigma(B_i^c)) \mu_{B_i}^\sigma(\sigma(B_i))$, so we can rewrite this sum as

$$\begin{aligned} \frac{1}{2} \frac{1}{|\mathcal{B}|} \sum_i \sum_{\sigma(B_i^c)} \mu(\sigma(B_i^c)) \sum_{\sigma(B_i), \sigma'(B_i)} \frac{|B_i|}{\sum_j |B_j|} \mu_{B_i}^\sigma(\sigma(B_i)) \mu_{B_i}^{\sigma'}(\sigma'(B_i)) (f(\sigma) - f(\sigma'))^2 \\ = \frac{1}{|\mathcal{B}|} \sum_i \frac{|B_i|}{\sum_j |B_j|} \sum_{\sigma(B_i^c)} \mu(\sigma(B_i^c)) \text{Var}_{B_i}^\sigma(f). \end{aligned}$$

Plugging in the spectral gap of B_i with boundary conditions, we get

$$\mathcal{E}_{\mathcal{B}}(f, f) \leq \frac{1}{|\mathcal{B}|} \sum_i \frac{|B_i|}{\sum_j |B_j|} \sum_{\sigma(B_i^c)} \mu(\sigma(B_i^c)) \frac{\mathcal{E}_{B_i}^\sigma(f, f)}{\gamma(\mathcal{M}_{\text{AT}}(B_i^\sigma))}. \quad (9)$$

At the same time, for every value of $\mu(\sigma(B_i^c))$,

$$\mathcal{E}_{B_i}^\sigma(f, f) = \frac{1}{2} \sum_{\sigma'(B_i)} \mu_{B_i}^\sigma(\sigma'(B_i)) \frac{1}{|B_i|} \sum_{(v, v+1) \in E(B_i)} p_{\sigma(v), \sigma(v+1)} (f(\sigma \oplus \sigma') - f(\sigma \oplus \sigma' \circ (v, v+1)))^2,$$

where $E(B)$ denotes the set of edges (pairs of adjacent positions) in B and $(\sigma \oplus \sigma')(x) = \begin{cases} \sigma(x), & x \notin B_i \\ \sigma'(x), & x \in B_i \end{cases}$. Plugging into (9), we overcount edges $(v, v+1)$ at most $\chi = \chi(\mathcal{B})$ many times, and thus using the lower bound $\sum_i |B_i| \geq N$, we can bound

$$\begin{aligned} \mathcal{E}_{\mathcal{B}}(f, f) &\leq \chi \left(\frac{1}{N} \sum_{\sigma} \mu(\sigma) \sum_{v \in [N-1]} p_{\sigma(v), \sigma(v+1)} (f(\sigma) - f(\sigma \circ (v, v+1)))^2 \right) \cdot \max_{i, \sigma(B_i^c)} \gamma^{-1}(\mathcal{M}_{\text{AT}}(B_i^\sigma)) \\ &= \chi \cdot \mathcal{E}(f, f) \cdot \max_{i, \sigma(B_i^c)} \gamma^{-1}(\mathcal{M}_{\text{AT}}(B_i^\sigma)). \end{aligned}$$

Using $\text{Var}(f) \leq \mathcal{E}_{\mathcal{B}}(f, f) / \gamma(P_{\mathcal{B}})$, we get

$$\frac{\text{Var}(f)}{\mathcal{E}(f, f)} \leq \chi \cdot \gamma^{-1}(\mathcal{M}_{\mathcal{B}}) \cdot \max_{i, \sigma(B_i^c)} \gamma^{-1}(\mathcal{M}_{\text{AT}}(B_i^\sigma)),$$

which gives the claimed bound on $\gamma(\mathcal{M}_{\text{AT}})$ (after inverting everything). $\square$

Remark 5.1: Just as the adjacent transposition chain could be restricted to $\vec{\ell}\text{-Loc}$ per Definition 2.1, the block dynamics can be restricted to $\vec{\ell}\text{-Loc}$ by rejecting any update that takes it out of the set. This restricted block dynamics remains ergodic (as it can simulate any move of the restricted adjacent transposition chain, which was ergodic by Claim 3.4), and reversible with respect to the conditional stationary distribution $\mu(\cdot \mid \vec{\ell}\text{-Loc})$.

The same proof of Proposition 5.3 holds in that context, since this is just a change of the reference measure with respect to which we are making heat-bath updates, and the conditioned measure still satisfies the consistency relation $\mu(\sigma \mid \vec{\ell}\text{-Loc}) = \mu(\sigma(B_i^c) \mid \vec{\ell}\text{-Loc})\mu_{B_i}^\sigma(\sigma(B_i) \mid \vec{\ell}\text{-Loc})$ used in the proof.

5.2 Spatial recursion

We now describe the main spatial recursion that gives us the polynomial mixing time bound. Consider the block dynamics $\mathcal{M}_{\text{WE}}$ with the following blocks:

$$B_W = \{1, \dots, \lceil 2N/3 \rceil\}, \quad B_E = \{\lfloor N/3 \rfloor, \dots, N\}.$$

We first establish fast mixing of the block dynamics. The following result is a generalization of Lemma 2.9.

Lemma 5.4. *There exist constants $C_{\text{block}} > 0$ (independent of ε), $C'_{\text{block}}(\varepsilon) = e^{O(1/\varepsilon)}$, $n_0 = O(\log(1/\varepsilon)/\varepsilon)$ and $n_1 = e^{O(1/\varepsilon)}$, such that*

- *for all $N \geq n_1$, the inverse gap of the block dynamics restricted to $\vec{\ell}\text{-Loc}$ for $\vec{\ell} \in \mathcal{L}_N$ is at most C_{block} ;*
- *for all $N \in [n_0, n_1]$, the inverse gap of the block dynamics restricted to $\vec{\ell}\text{-Loc}$ for $\vec{\ell} \in \mathcal{L}_N$ is at most $C'_{\text{block}}(\varepsilon)$.*

Proof. To show the block dynamics has $C(\varepsilon)$ spectral gap for some ε , it is sufficient to show that in time $O(C(\varepsilon) \log N)$, the distance to stationarity is at most N^{-8} . This is because the exponential decay rate to stationarity is bounded by the spectral gap: see [LP17, Lemma 20.11].

By Lemma 4.2, the restricted block dynamics $\hat{X}_t$ burns in to $\ell_0\text{-Loc}$ where $\ell_0 = C_0(\varepsilon) \log N$ for $C_0(\varepsilon) = O(1/\varepsilon)$: there exists C_1 such that for any $t \geq C_1 \log N$,

$$\mathbb{P}(\hat{X}_t \notin \ell_0\text{-Loc}) \leq N^{-10}.$$

Let $C''(\varepsilon) = \Theta(1)$ for $N \geq n_1$, and $C''(\varepsilon) = e^{\Theta(1/\varepsilon)}$ for $N \in [n_0, n_1]$ (to match the expressions for C_{block} and $C'_{\text{block}}(\varepsilon)$). Then by a union bound, the above holds for all times between $C''(\varepsilon) \log N$ and $2C''(\varepsilon) \log N$:

$$\mathbb{P}(\exists t \in [C''(\varepsilon) \log N, 2C''(\varepsilon) \log N] : \hat{X}_t \notin \ell_0\text{-Loc}) \leq N^{-9}.$$

We now bound the coupling time of two restricted block dynamics chains $\hat{X}_t, \hat{X}'_t$ from two different initializations. We use the same choices of blocks for the two, and work on the event that both of them are in $\ell_0\text{-Loc}$ for all $t \in [C''(\varepsilon) \log N, 2C''(\varepsilon) \log N]$.

Now consider the event that there is an East block update, immediately followed by a West block update. This happens at least $\delta^{-1} \log N$ many times in the time-frame $[C'(\varepsilon) \log N, 2C'(\varepsilon) \log N]$ with probability at least $1 - N^{-10}$, as long as the ε -independent constant in $C'(\varepsilon)$ is large enough (as a function of δ). On that event, intersected with the above event that $\hat{X}_t, \hat{X}'_t \in \ell_0\text{-Loc}$ for all such times,

- When B_E updates, by Lemma 3.3, as long as $r \geq 10C_0(\varepsilon) \log N$, there exists a coupling of the resampling step such that with probability at least $1 - e^{-r/(C''(\varepsilon)C_0(\varepsilon) \log N)}$ the two configurations agree fully on $B_E \setminus B_W$, where $C''(\varepsilon) = \exp(O(\frac{1}{\varepsilon}))$.
 - For $N \geq n_1$, with appropriate $n_1 = e^{O(1/\varepsilon)}$, this probability is $\Omega(1)$.
 - For $n_0 \leq N \leq n_1$, with appropriate choice of $n_0 = O(\log(1/\varepsilon)/\varepsilon)$, we have $10C_0(\varepsilon) \log N \leq N \leq n_1$, so this probability is $1 - \exp(-\exp(-O(1/\varepsilon))) = \Omega(\exp(-O(1/\varepsilon)))$.
- On that event, when B_W updates in the next step, the induced distributions by the two chains on B_W given the configurations on $B_E \setminus B_W$ are identical and therefore they are coupled to equal using the identity coupling.

Hence for $N \geq n_1$, with probability at least $1 - N^{-9}$, the two chains couple after time $O(\log N)$. For $n_0 \leq N \leq n_1$, with probability at least $1 - N^{-9}$, the two chains couple after time $\exp(O(1/\varepsilon)) \log N$. Choose constants such that these are bounded by $C'(\varepsilon) \log N$. This bounds the mixing time to N^{-9} and hence finishes the proof. $\square$

5.3 Proof of the polynomial bound

We are now in position to recurse using Lemma 5.4 and apply a crude bound for the base case bound to deduce Proposition 5.1.

Proof of Proposition 5.1. We make the following more precise statement which we prove by induction: For all $\mathcal{M}_{\text{AT}}$ on $[N]$ with ε -positively biased $\vec{p}$, and restricted to any $\vec{\ell}^{(N)}\text{-Loc}$ for $\vec{\ell}^{(N)} \in \mathcal{L}_N$, the spectral gap of $\mathcal{M}_{\text{AT}}$ is bounded as follows, where C is ε -independent and $C'(\varepsilon) = e^{O(1/\varepsilon^2)}$. Here, $n_0 = O(\log(1/\varepsilon)/\varepsilon)$ and $n_1 = e^{O(1/\varepsilon)}$.

1. For $N \geq n_0$, $\gamma(\mathcal{M}_{\text{AT}})^{-1} \leq C'(\varepsilon)N^{1/\varepsilon+C}$.
2. For $N \geq n_1$, $\gamma(\mathcal{M}_{\text{AT}})^{-1} \leq C'(\varepsilon)N^C$.

Base case. As in previous lemmas, let $n_0 = O(\log(1/\varepsilon)/\varepsilon)$. Then as base cases, for all $N \leq 2n_0$, we give a crude bound on the mixing time. We claim that we can couple any two initializations in time $[0, n_0^2]$ with probability at least $\exp(-\tilde{\Omega}(n_0^2))$. Indeed, suppose that the first $N(N-1)$ many updates are exactly N ordered sweeps from right-to-left (this has probability at least $(\frac{1}{N})^{N^2} = e^{-\tilde{O}(n_0^2)}$). Furthermore, suppose that every time an edge is selected to be updated, it swaps the particles at its end points if and only if they are out of order. This has probability at least $\frac{1}{2}$ in each step, and therefore $(\frac{1}{2})^{N^2}$, conditional on

the sequence of updates. The resulting configuration at time N^2 will be exactly the ground state $(1, 2, \dots, N)$ in both chains, and they will be thereafter coupled. Since this coupling can be reattempted every N^2 many steps, we get a coupling time, and therefore mixing time bound, of $O(N^2)N^{N^2}2^{N^2} = e^{\tilde{O}(n_0^2)}$.

Inductive step. By Proposition 5.3, we can bound the inverse spectral gap in terms of the inverse spectral gap of the block dynamics $\mathcal{M}_B$ and of the restriction:

$$\gamma(\mathcal{M}_{\text{AT}})^{-1} \leq 2\gamma(\mathcal{M}_B)^{-1} \max_{\xi} \max_{B \in \{B_W, B_E\}} \gamma(\mathcal{M}_{\text{AT}}(B^\xi))^{-1},$$

where the 2 comes from χ , and where ξ is a “boundary condition” meaning a configuration on $[N] \setminus B$. Since B^ξ is itself a segment of length $\lceil 2N/3 \rceil$ with some arbitrary boundary conditions and alphabet available to it, and with ε still serving as a lower bound on the bias, we can apply Lemma 3.5 to relabel it so that it is equivalent to an ε -positively biased adjacent transposition chain on $\lceil 2N/3 \rceil$ restricted to some $\vec{\ell}' \in \mathcal{L}_{\lceil 2N/3 \rceil}$. Letting G_N be the maximum possible value of $\gamma(\mathcal{M}_{\text{AT}})^{-1}$ for ε -positively biased $\vec{p}$ and restricted to any $\vec{\ell}^{(N)}$ -Loc for $\vec{\ell}^{(N)} \in \mathcal{L}_N$. Then substituting in the bounds from Lemma 5.4,

$$G_N \leq \begin{cases} 2C'_{\text{block}}(\varepsilon)G_{\lceil 2N/3 \rceil}, & N \geq n_0. \\ 2C_{\text{block}}G_{\lceil 2N/3 \rceil}, & N \geq n_1, \end{cases}$$

Now the inductive statement holds for $N \in [n_0, 2n_0]$ by the base case analysis. For $N \geq 2n_0$, the induction step is

$$G_N \leq 2C'_{\text{block}}(\varepsilon)C'(\varepsilon) \lceil 2N/3 \rceil^{C_2/\varepsilon+C} \leq C'(\varepsilon) \left(\frac{N}{\lceil 2N/3 \rceil} \right)^{C_2/\varepsilon+C} \lceil 2N/3 \rceil^{C_2/\varepsilon+C} \leq C'(\varepsilon)N^{C_2/\varepsilon+C},$$

for appropriate choice of C_2, C . This gives that the inductive statement holds for $N \in [n_1, 2n_1]$ possibly for a different $C'(\varepsilon)$, still of the same order, by absorbing $(2n_1)^{1/\varepsilon+C}$ into the constant. For $N \geq 2n_1$, the induction step is

$$G_N \leq 2C_{\text{block}}C'(\varepsilon) \lceil 2N/3 \rceil^C \leq C'(\varepsilon) \left(\frac{N}{\lceil 2N/3 \rceil} \right)^C \lceil 2N/3 \rceil^C \leq C'(\varepsilon)N^C,$$

for appropriate C independent of ε . □

6 Boosting to sharp mixing time bound

In this section, we boost the polynomial bound of Proposition 5.1 to the optimal $O(n^2)$ mixing time, and prove Theorem 1.1.

6.1 Fast mixing within the localized set

Because of Lemma 4.1, we know that in $C'(\varepsilon)n^2 + O(n \log n)$ time, the adjacent transposition process is burnt-in and gets to a localized configuration, and stays there for large polynomial times. We then need to show that the mixing time restricted to $\ell_0\text{-Loc}$ is $o(n^2)$.

Consider the restricted Markov chain $\bar{\sigma}_t$ which rejects any updates that take the configuration outside $\ell_0\text{-Loc}$ for $\ell_0 = C_0(\varepsilon) \log n$.

Lemma 6.1. *There exist $C(\varepsilon) = e^{O(1/\varepsilon^2)}$, $n_0(\varepsilon) = O(\log(1/\varepsilon)/\varepsilon)$, $C_0(\varepsilon) = O(1/\varepsilon)$ and a constant $C_1 > 0$ independent of ε such that the following holds. For all $n \geq n_0(\varepsilon)$, for $\ell_0 = C_0(\varepsilon) \log n$, the $\ell_0\text{-Loc}$ restricted chain $\bar{\sigma}_t$ has mixing time at most $C(\varepsilon)n(\log n)^{C_1}$.*

Proof. We prove this utilizing a block dynamics restricted to the set $\ell_0\text{-Loc}$ which we will now denote by $\bar{\Omega}$. Let $M = e^{\Theta(1/\varepsilon)}(\log n)^3$ with appropriate constants to be specified implicitly later in the proof. Let $\hat{X}_t$ be the restricted block dynamics with the following two blocks:

1. $B_0 = \bigcup_{i \geq 0} (6iM, (6i+4)M]$;
2. $B_1 = \bigcup_{i \geq 0} ((6i+3)M, (6i+7)M]$.

Observe that except at the ends, for the i -th segment of B_0 , its left-most and right-most stretches of length M are subsets of B_1 .

Recall from Lemma 4.2 that the restricted block dynamics means that when one of these blocks, say B_0 is updated, the configuration is resampled on B_0 conditioned on the configuration on $B_1 \setminus B_0$ as well as on remaining in $\bar{\Omega}$.

By Proposition 5.3, we need to bound the inverse gap of the restricted block dynamics, and the inverse gap of the restricted adjacent transposition shuffle within a block.

Block dynamics spectral gap. We establish an (ε -independent) $O(1)$ bound on the mixing time of the restricted block dynamics, which implies the same bound on the inverse spectral gap. Consider any two initial states σ, σ' both in $\ell_0\text{-Loc}$. Let T be the first time for an update of B_0 followed by an update of B_1 at time $T+1$. For any $\delta > 0$, observe that $T \leq C(\delta) = O(1)$ with probability at least $1 - \delta$.

First observe that when updating B_0 (at time T), the update within each segment $(6iM, (6i+4)M]$ is independent of each other. By Lemma 3.3, because M is greater than the required buffer of ℓ_0 , for any i , for any fixed configurations σ, σ' on $(6iM, (6i+4)M]^c$, we can couple so that $\sigma([(6i+1)M, (6i+3)M]) \equiv \sigma'([(6i+1)M, (6i+3)M])$ with probability $\geq 1 - \exp(-M/C'(\varepsilon)\ell_0)$, where $C'(\varepsilon) = e^{O(1/\varepsilon)}$. Therefore, by a union bound over i we have the following:

$$\mathbb{P}(\sigma_{T+1}(B_0 \setminus B_1) \not\equiv \sigma'_{T+1}(B_0 \setminus B_1)) \leq n \cdot e^{-M/C'(\varepsilon)\ell_0} = e^{-C''(\log n)^2}$$

for some absolute constant C'' . On the next update at time $T+1$, which is on B_1 , assuming the above coupling event occurred (i.e., they agree on $B_0 \setminus B_1$) then the two chains have the same configuration on the complement of B_1 and thus we can use the identity coupling to

obtain $\sigma_{T+2}(B_1) = \sigma'_{T+2}(B_1)$ and hence $\sigma_{T+2} = \sigma'_{T+2}$. This bounds the mixing time of the block dynamics by an ε -independent constant C .

Spectral gap in a block. We first observe that restricted to $\ell_0\text{-Loc}$, the possible particles that could be in a constituent segment of B_0 and the possible particles that could be in another constituent segment of B_0 are disjoint. Therefore, the $\ell_0\text{-Loc}$ -restricted measure on B_0 is a product distribution over its distinct segments. The adjacent transposition shuffle on B_0 is also a product chain across the disjoint segments, and therefore the spectral gap of B_0 tensorizes to give n times the minimal spectral gap across the segments $((6iM, (6i+4)M])_i$ with worst-case boundary conditions in $\ell_0\text{-Loc}$, and restricted to $\ell_0\text{-Loc}$ inside the segment.

By Lemma 3.5, up to a relabeling of the particle labels, this is equivalent to an ε -positively biased adjacent transposition chain $\mathcal{M}_{\text{AT}}([4M])$ on $[4M]$ restricted to some $\vec{\ell}^{(4M)} \in \mathcal{L}_{4M}$. In particular, the bound of Proposition 5.1 can be applied to bound its mixing time and inverse spectral gap by $C'(\varepsilon)(4M)^{C_2} = C'(\varepsilon)(e^{O(1/\varepsilon)} \log^3 n)^{C_2}$ for a C_2 that is ε -independent.

Applying Proposition 5.3 we get the bound on the inverse spectral gap of the $\ell_0\text{-Loc}$ -restricted adjacent transposition shuffle $\bar{\sigma}_t$ by

$$\gamma^{-1} \leq 2C'(\varepsilon)C_1n(\log n)^{3C_2},$$

where the factor 2 is from χ and the factor n comes from the tensorization in discrete time of the gap of B_0, B_1 . $\square$

6.2 Proof of main theorem

We can now combine Lemma 6.1 with the burn-in result for $\mathcal{M}_{\text{AT}}$ from Lemma 4.1, and a simple lower bound, to deduce our main result.

Proof of Theorem 1.1. *Upper bound.* Let $(\bar{\sigma}_t)_{t \geq 0}$ be the adjacent transposition shuffle restricted to $\ell_0\text{-Loc}$. Let $\hat{\sigma}_t$ be the chain that follows the unrestricted $\mathcal{M}_{\text{AT}}$ chain σ_t for time $T_0 := C_\varepsilon n^2$, and then starting from $\hat{\sigma}_{T_0} = \sigma_{T_0}$, follows the restricted chain $\hat{\sigma}_{T_0+s} = \bar{\sigma}_s$ with $\bar{\sigma}_0 = \hat{\sigma}_{T_0}$, coupled so that $\hat{\sigma}_{T_0+s}$ agrees with σ_{T_0+s} on the event that $(\sigma_{T_0+r})_{r \leq s}$ is entirely in $\ell_0\text{-Loc}$.

By this coupling, for $t = T_0 + s$ and $s \leq n^3 - C_\varepsilon n^2$, we get for any event $A \subseteq S_n$,

$$\max_{x_0} \mathbb{P}_{x_0}(\sigma_t \in A) \leq \max_{x_0} \mathbb{P}_{x_0} \left(\bigcup_{t \in [T_0, n^3]} \{\sigma_t \notin \ell_0\text{-Loc}\} \right) + \max_{y_0 \in \ell_0\text{-Loc}} \mathbb{P}_{y_0}(\bar{\sigma}_s \in A).$$

By Lemma 6.1, if $s \geq C(\varepsilon)n(\log n)^{C_1+1}$ for $C(\varepsilon) = e^{O(1/\varepsilon^2)}$, $C_1 = O(1/\varepsilon)$, then the second term is at most

$$\mu(A \mid \ell_0\text{-Loc}) + n^{-10} \leq \mu(A) + 3n^{-10},$$

where we use Lemma 2.2 to remove the conditioning. By Lemma 4.1 and a union bound over $t \in [T_0, n^3]$, the first term is at most n^{-7} , yielding

$$\max_{x_0} \mathbb{P}_{x_0}(\sigma_t \in A) \leq \mu(A) + 4n^{-7}.$$

The same bound holds for A^c . Hence for any δ , for large enough n , we achieve total variation distance of δ in time $T_0 + C(\varepsilon)n(\log n)^{C_1+1} \log(1/\delta)$, giving the upper bounds for the mixing time as well as for pre-cutoff.

Lower bound. For the lower bound, we use a test statistic, which is the location of particle 1. Consider the initialization that is fully inverted, $\sigma_0 = (n, n-1, \dots, 1)$. The location of particle 1, i.e., $\sigma^{-1}(1)$ is stochastically to the right of the process which flips a $\text{Ber}\left(\frac{1}{n-1}\right)$ coin to decide whether or not to move, and if it comes up heads, then it deterministically moves to the left. It is straightforward to see by a Chernoff bound that for any $\eta > 0$, this latter process takes time at least $(1-\eta)n^2$ to have $\Omega(1)$ probability of particle 1 being in the interval $[\sqrt{n}]$. Therefore, by the stochastic domination,

$$\mathbb{P}(\sigma_{(1-\eta)n^2}^{-1}(1) \in [\sqrt{n}]) = o(1).$$

On the other hand, by Lemma 2.2, the event $\sigma^{-1}(1) \in [\sqrt{n}]$ has probability $1 - o(1)$ under stationarity as long as $\frac{\sqrt{n}}{\log n} \gg 1/\varepsilon$. This implies for every $\eta > 0$, a lower bound of $T_{\text{mix}}(1-\delta) \geq (1-\eta)n^2$ for all $\delta > 0$ and all $n = \tilde{\Omega}(1/\varepsilon^2)$. $\square$

Acknowledgments

This work originated from conversations at the Rocky Mountain Summer Workshop in Algorithms, Probability, and Combinatorics at Colorado State University. The research of R.G. is supported in part by NSF CAREER grant 2440509 and NSF DMS grant 2246780. The research of E.V. is supported by NSF grant CCF-2147094.

References

- [Ald83] David Aldous. Random walks on finite groups and rapidly mixing Markov chains. In *Séminaire de Probabilités XVII 1981/1982*, pages 243–297. Springer-Verlag, 1983.
- [BBHM05] Itai Benjamini, Noam Berger, Christopher Hoffman, and Elchanan Mossel. Mixing times of the biased card shuffling and the asymmetric exclusion process. *Transactions of the American Mathematical Society*, 357(8):3013–3029, 2005.
- [BD99] Russ Bubley and Martin Dyer. Faster random generation of linear extensions. *Discrete Mathematics*, 201(1-3):81–88, 1999.
- [BMRS12] Prateek Bhakta, Sarah Miracle, Dana Randall, and Amanda Pascoe Streib. Mixing times of Markov chains for self-organizing lists and biased permutations. *arXiv:1204.3239*, 2012.
- [BMRS22] Prateek Bhakta, Sarah Miracle, Dana Randall, and Amanda Pascoe Streib. Mixing times of Markov chains for self-organizing lists and biased permutations. *Random Structures & Algorithms*, 61(4):638–665, 2022.

- [Cor16] Ivan Corwin. Kardar-Parisi-Zhang universality. *Notices of the AMS*, 63(3):230–239, 2016.
- [DSC93] P. Diaconis and L. Saloff-Coste. Comparison techniques for random walks on finite groups. *Ann. Probab.*, 21:2131–2156, 1993.
- [Fil03] James Allen Fill. An interesting spectral gap problem, from Jim Fill. *arXiv:2508.12557*, 2003. Notes from 2003, posted to arXiv in 2025.
- [Hen76] W. J. Hendricks. An account of self-organizing systems. *SIAM Journal on Computing*, 5(4):715–723, 1976.
- [HH85] J. H. Hester and D. S. Hirschberg. Self-organizing linear search. *ACM Comp. Surveys*, 19:295–311, 1985.
- [HW17] Shahrzad Haddadan and Peter Winkler. Mixing time for some adjacent transposition Markov chains. In *Proceedings of the 34th Symposium on Theoretical Aspects of Computer Science (STACS)*, 2017.
- [JS89] Mark Jerrum and Alistair Sinclair. Approximating the permanent. *SIAM journal on computing*, 18(6):1149–1178, 1989.
- [JSV04] Mark Jerrum, Alistair Sinclair, and Eric Vigoda. A polynomial-time approximation algorithm for the permanent of a matrix with nonnegative entries. *Journal of the ACM (JACM)*, 51(4):671–697, 2004.
- [KRBN10] Pavel L Krapivsky, Sidney Redner, and Eli Ben-Naim. *A kinetic view of statistical physics*. Cambridge University Press, 2010.
- [Lac16] Hubert Lacoin. Mixing time and cutoff for the adjacent transposition shuffle and the simple exclusion. *The Annals of Probability*, 44(2):1426 – 1487, 2016.
- [Lig85] Thomas M Liggett. *Interacting particle systems*, volume 2. Springer, 1985.
- [LL19] Cyril Labbé and Hubert Lacoin. Cutoff phenomenon for the asymmetric simple exclusion process and the biased card shuffling. *The Annals of Probability*, 47(3):1541 – 1586, 2019.
- [LP17] David A. Levin and Yuval Peres. *Markov chains and mixing times*. American Mathematical Society, 2017.
- [Mar04] Fabio Martinelli. Lectures on Glauber dynamics for discrete spin models. In *Lectures on probability theory and statistics: Ecole d’été de probabilités de Saint-Flour XXVII-1997*, pages 93–191. Springer, 2004.
- [MS18] Sarah Miracle and Amanda Pascoe Streib. Rapid mixing of k -class biased permutations. In *LATIN*, volume 10807 of *Lecture Notes in Computer Science*, pages 820–834, 2018.

- [MSS25] Sarah Miracle, Amanda Pascoe Streib, and Noah Streib. Iterated decomposition of biased permutations via new bounds on the spectral gap of Markov chains. *Theory of Computing*, 21(3):1–41, 2025.
- [Spi70] Frank Spitzer. Interaction of Markov processes. *Advances in Mathematics*, 5(2):246–290, 1970.
- [Wil03] David Wilson. Mixing time of the Rudvalis shuffle. *Electronic Communications in Probability*, 8:77–85, 2003.