

LARGE DEVIATIONS FOR THE INTERCHANGE PROCESS ON THE INTERVAL AND INCOMPRESSIBLE FLOWS

MICHAŁ KOTOWSKI AND BÁLINT VIRÁG



Abstract. We use the framework of permuton processes to show that large deviations of the interchange process are controlled by the Dirichlet energy. This establishes a rigorous connection between processes of permutations and one-dimensional incompressible Euler equations. While our large deviation upper bound is valid in general, the lower bound applies to processes corresponding to incompressible flows, studied in this context by Brenier. These results imply the Archimedean limit for relaxed sorting networks and allow us to asymptotically count such networks.

Contents

1	Introduction	1357
2	Preliminaries	1363
2.1	Permutons and stochastic processes.	1363
2.2	Euler equations and generalized incompressible flows.	1368
2.3	Proof outline and structure of the paper.	1370
3	ODEs and Generalized Solutions to Euler Equations	1372
4	The Biased Interchange Process and Stationarity	1383
5	Law of Large Numbers	1387
6	One Block Estimate	1398
7	Large Deviation Lower Bound	1406
8	Large Deviation Upper Bound	1411
9	Asymptotics of Relaxed Sorting networks	1422
	References	1425

1 Introduction

In this paper we investigate the large deviation principle for a model of random permutations called the one-dimensional *interchange process*. The process can be roughly described as follows. We put N particles, labelled from 1 to N , on a line $\{1, \dots, N\}$ and at each time step perform the following procedure: an edge is chosen at random and adjacent particles are swapped. By comparing the particles' initial positions with their positions after given time t we obtain a random permutation from the symmetric group \mathcal{S}_N on N elements.

The interchange process on the interval (whose discrete time analog is known as the adjacent transposition shuffle) and on more general graphs has attracted considerable attention in probability theory, for example with regard to the analysis of mixing times. It is natural to ask whether, after proper rescaling and as $N \rightarrow \infty$, the permutations obtained in the interchange process converge in distribution to an appropriately defined limiting process.

Such limits have been recently studied ([HKM+13, RVV19]) under the name of permutons and permuton processes. These notions have been inspired by the theory of graph limits ([Lov12]), where the analogous notion of a graphon as a limit of dense graphs appears. A *permuton* is a Borel probability measure on $[0, 1]^2$ with uniform marginals on each coordinate. A sequence of permutations $\sigma^N \in \mathcal{S}_N$ is said to converge to a permuton μ as $N \rightarrow \infty$ if the corresponding empirical measures

$$\frac{1}{N} \sum_{i=1}^N \delta_{\left(\frac{i}{N}, \frac{\sigma^N(i)}{N}\right)}$$

converge weakly to μ . A *permuton process* is a stochastic process $X = (X_t, 0 \leq t \leq T)$ taking values in $[0, 1]$, with continuous sample paths and having uniform marginals at each time $t \in [0, T]$. A permutation-valued path, such as a sample from the interchange process, is said to converge to X if the trajectory of a randomly chosen particle converges in distribution to X .

Depending on the time scale considered, one observes different asymptotic structure in the permutations arising from the interchange process. If the average number of all swaps is greater than $\sim N^3 \log N$, the process will be close to its stationary distribution ([Ald83, Lac16]), which is the uniform distribution on \mathcal{S}_N . For $\sim N^3$ swaps each particle has displacement of order N and the whole process converges, in the sense of permuton processes, to a Brownian motion on $[0, 1]$ ([RV17]).

Here we will be interested in yet shorter time scales, corresponding to $\sim N^{2+\varepsilon}$ swaps for fixed $\varepsilon \in (0, 1)$. In this scaling each particle has displacement $\ll N$, so the resulting permutations will be close to the identity permutation. Nevertheless, in the spirit of large deviation theory one can still ask questions about rare events, for example “what is the probability that starting from the identity permutation we are close to a fixed permuton after time t ?” or, more generally, “what is the probability that the interchange process behaves like a given permuton process X ?”. We expect such probabilities to decay exponentially in N^γ for some $\gamma > 0$, with the decay rate given by a *rate function* on the space of permuton processes.

The large deviation principle we obtain in this paper can be informally summarized as follows: for a class of permuton processes solving a natural energy minimization problem, the probability $\mathbb{P}(A)$ that the interchange process is close in distribution to a process X satisfies asymptotically

$$\frac{1}{N^\gamma} \log \mathbb{P}(A) \approx -I(X), \tag{1}$$

where $\gamma = 2 - \varepsilon$ and $I(X)$ is the *energy* of X , defined as the expected Dirichlet energy of a path sampled from X . Apart from a purely probabilistic interest, the result is relevant to two other seemingly unrelated subjects, namely the study of Euler equations in fluid dynamics and the study of sorting networks in combinatorics.

Let us first state the energy minimization problem in question, which is as follows – given a permuton μ , find

$$\inf_{(X_0, X_T) \sim \mu} I(X), \quad (2)$$

where the infimum is over all permuton processes X such that (X_0, X_T) has distribution μ . As it happens, such energy-minimizing processes have been considered in fluid dynamics in the study of *incompressible Euler equations*, under the name of *generalized incompressible flows*. This connection is discussed in more detail in Sect. 2.2. Very roughly speaking, Euler equations in a domain $D \subseteq \mathbb{R}^d$ describe motion of fluid particles whose trajectories satisfy the equation

$$x''(t) = -\nabla p(t, x) \quad (3)$$

for some function p called the *pressure*. The incompressibility constraint means that the flow defined by the equation has to be volume-preserving. Classical, smooth solutions to Euler equations correspond to flows which are diffeomorphisms of D . Generalized incompressible flows are a stochastic variant of such solutions in which each particle can choose its initial velocity independently from a given probability distribution.

It turns out that, under additional regularity assumptions, such generalized solutions to Euler Eq. (3) for $D = [0, 1]$ correspond exactly to permuton processes solving the energy minimization problem (2) for some permuton μ . Our large deviation result (1) is valid precisely for such energy-minimizing permuton processes (again, under certain regularity assumptions).

As it happens, the original motivation for our work came from a different direction, namely from the study of *sorting networks* in combinatorics. This connection is explained in more detail below. Using our large deviation principle (1), we are able to prove novel results on a variant of the model we call *relaxed sorting networks*. Thus the large deviation principle presented in this paper provides a rather unexpected link between problems in combinatorics (sorting networks) and fluid dynamics (incompressible Euler equations), along with a quite general framework for analyzing permuton processes which we hope will find further applications.

Main results. Let us now state our main results more formally, still with complete definitions and discussion of assumptions deferred until Sects. 2.1 and 3. Let $\mathcal{D} = \mathcal{D}([0, T], [0, 1])$ be the space of càdlàg paths from $[0, T]$ to $[0, 1]$ and let $\mathcal{M}(\mathcal{D})$ be the space of Borel probability measures on \mathcal{D} . Let $\mathcal{P} \subseteq \mathcal{M}(\mathcal{D})$ denote the space of permuton processes and their approximations by permutation-valued processes. For $\pi \in \mathcal{M}(\mathcal{D})$ by $I(\pi)$ we will denote the expected Dirichlet energy of the process X whose distribution is π .

Let η^N denote the interchange process in continuous time on the interval $\{1, \dots, N\}$, speeded up by N^α for some $\alpha \in (1, 2)$. Let $\gamma = 3 - \alpha$. We have the following large deviation principle

Theorem A (Large deviation lower bound). *Let \mathbb{P}^N be the law of the interchange process η^N and let $\mu^{\eta^N} \in \mathcal{M}(\mathcal{D})$ be the empirical distribution of its trajectories. Let π be a permuton process which is a generalized solution to Euler Eq. (19). Provided π satisfies Assumptions (3.1), for any open set $\mathcal{O} \subseteq \mathcal{P}$ such that $\pi \in \mathcal{O}$ we have*

$$\liminf_{N \rightarrow \infty} N^{-\gamma} \log \mathbb{P}^N \left(\mu^{\eta^N} \in \mathcal{O} \right) \geq -I(\pi).$$

Theorem B (Large deviation upper bound). *Let \mathbb{P}^N be the law of the interchange process η^N and let $\mu^{\eta^N} \in \mathcal{M}(\mathcal{D})$ be the empirical distribution of its trajectories. For any closed set $\mathcal{C} \subseteq \mathcal{P}$ we have*

$$\limsup_{N \rightarrow \infty} N^{-\gamma} \log \mathbb{P}^N \left(\mu^{\eta^N} \in \mathcal{C} \right) \leq - \inf_{\pi \in \mathcal{C}} I(\pi).$$

The results are referred to as respectively Theorems 7.3 and 8.4 in the following sections. Here the large deviation upper bound is valid for *all* permuton processes, without any additional assumptions. On the other hand, in the proof of the lower bound we exploit rather heavily the special structure possessed by generalized solutions to Euler equations. We expect the lower bound to hold for arbitrary permuton processes as well, since one can locally approximate any permuton process by energy minimizers. However, for our techniques to apply one would need to understand in more detail regularity of the associated velocity distributions and pressure functions, which falls outside the scope of our work.

The reader may notice that the rate function, which is the energy $I(\pi)$, is similar to the one appearing in the analysis of large deviations for independent random walks. In fact, the crux of our proofs lies in proving that particles in the interchange process and its perturbations are in a certain sense almost independent. The main techniques used here come from the field of interacting particle systems. A comprehensive introduction to the subject can be found in [KL99]. The novelty in our approach is in applying tools usually used to study hydrodynamic limits to a setting which is in some respects more involved, since the limiting objects we consider, permuton processes, are stochastic processes instead of deterministic objects like solutions of PDEs appearing, for example, for exclusion processes.

Sorting networks and the sine curve process. The large deviation bounds can be applied to obtain results on a model related to sorting networks. A *sorting network* on N elements is a sequence of $M = \binom{N}{2}$ transpositions $(\tau_1, \tau_2, \dots, \tau_M)$ such that each τ_i is a transposition of adjacent elements and $\tau_M \circ \dots \circ \tau_1 = \text{rev}_N$, where $\text{rev}_N = (N \dots 21)$ denotes the *reverse permutation*. It is easy to see that any sequence of adjacent transpositions giving the reverse permutation must have length at least $\binom{N}{2}$, hence sorting networks can be thought of as shortest paths

joining the identity permutation and the reverse permutation in the Cayley graph of \mathcal{S}_N generated by adjacent transpositions.

A *random sorting network* is obtained by sampling a sorting network uniformly at random among all sorting networks on N elements. Let us work in continuous time, assuming each transposition τ_i happens at time $\frac{i}{M+1}$. It was conjectured in [AHRV07] and recently proved in [Dau22] that the trajectory of a randomly chosen particle in a random sorting network has a remarkable limiting behavior as $N \rightarrow \infty$, namely it converges in the sense of permuton processes to a deterministic limit, which is the sine curve process described below.

Here it will be more natural to consider the square $[-1, 1]^2$ and processes with values in $[-1, 1]$ instead of $[0, 1]$ (with the obvious changes in the notion of a permuton and a permuton process which we leave implicit). The *Archimedean law* is the measure on $[-1, 1]^2$ obtained by projecting the normalized surface area of a 2-dimensional half-sphere to the plane or, equivalently, the measure supported inside the unit disk $\{x^2 + y^2 \leq 1\}$ whose density is given by $1/(2\pi\sqrt{1-x^2-y^2}) dx dy$. Observe that thanks to the well-known plank property each strip $[a, b] \times [-1, 1]$ has measure proportional to $b - a$, hence the Archimedean law defines a permuton.

The *sine curve process* is the permuton process $\mathcal{A} = (\mathcal{A}_t, 0 \leq t \leq 1)$ with the following distribution – we sample (X, Y) from the Archimedean law and then follow the path

$$\mathcal{A}_t = X \cos \pi t + Y \sin \pi t.$$

One can directly check that \mathcal{A}_t has uniform distribution on $[-1, 1]$ at each time t , hence \mathcal{A}_t indeed defines a permuton process. Observe that $(\mathcal{A}_0, \mathcal{A}_0) = (X, X)$ and $(\mathcal{A}_0, \mathcal{A}_1) = (X, -X)$, thus the sine curve process defines a path between the identity permuton and the *reverse permuton*.

An equivalent way of describing the sine curve process consists of choosing a pair (R, θ) at random, where the angle θ is uniform on $[0, 2\pi]$ and R has density $r/2\pi\sqrt{1-r^2} dr$ on $[0, 1]$, and following the path $\mathcal{A}_t = R \cos(\pi t + \theta)$. Thus the trajectories of this process are sine curves with random initial phase and amplitude – the path of a random particle is determined by its initial position X and velocity V , given by $(X, V) = (R \cos \theta, -\pi R \sin \theta)$.

Recall now the energy minimization problem (18). The sine curve process is the unique minimizer of energy among all permuton processes joining the identity to the reverse permuton ([Bre89], see also [RVV19]), with the minimal energy equal to $I(\mathcal{A}) = \frac{\pi^2}{6}$. It is one of the few examples where the solution to the problem (18) can be explicitly calculated for a target permuton μ . It also seems to play a special role in constructing generalized incompressible flows which are non-unique solutions to the energy minimization problem in dimensions greater than one, see, e.g., [BFS09].

The sine curve process is a generalized solution to Euler equations with the pressure function $p(x) = \frac{x^2}{2}$, which unsurprisingly leads to each particle satisfying the harmonic oscillator equation $x'' = -x$. The reader may check that the sine curve

process satisfies the Assumptions (3.1) (with the velocity distribution being time-independent), thus providing a non-trivial and explicit example for which our large deviation bounds hold. To the best of our knowledge the connection between sorting networks on the one hand and Euler equations on the other hand was first observed in the literature in [Dau22].

Let us now describe the results on relaxed sorting networks. Fix $\delta > 0$ and $N \geq 1$. We define a δ -relaxed sorting network of length M on N elements to be a sequence of M adjacent transpositions (τ_1, \dots, τ_M) such that the permutation $\sigma_M = \tau_M \circ \dots \circ \tau_1$ is δ -close to the reverse permutation $\text{rev} = (N \dots 2 1)$ in the Wasserstein distance on the space $\mathcal{M}([0, 1]^2)$ of Borel probability measures on $[0, 1]^2$ (see Sect. 2.1 for the definition). For fixed $\kappa \in (0, 1)$ we define a random δ -relaxed sorting network on N elements by choosing M from a Poisson distribution with mean $\lfloor \frac{1}{2}N^{1+\kappa}(N - 1) \rfloor$ and then sampling a δ -relaxed sorting network of length M uniformly at random.

Our first result is that the analog of the sorting network conjecture holds for relaxed sorting networks, that is, in a random relaxed sorting network the trajectory of a random particle is with high probability close in distribution to the sine curve process. Precisely, we have the following

Theorem 1.1. *Fix $\kappa \in (0, 1)$ and let π_δ^N denote the empirical distribution of the permutation process (as defined in 5) associated to a random δ -relaxed sorting network on N elements. Let $\pi_{\mathcal{A}}$ denote the distribution of the sine curve process. Given any $\varepsilon > 0$ we have for all sufficiently small $\delta > 0$*

$$\lim_{N \rightarrow \infty} \mathbb{P}^N (\pi_\delta^N \in B(\pi_{\mathcal{A}}, \varepsilon)) = 1,$$

where $B(\pi_{\mathcal{A}}, \varepsilon)$ is the ε -ball in the Wasserstein distance on \mathcal{P} .

Here for consistency of notation we assume that the sine curve process is rescaled so that it is supported on $[0, 1]$ rather than $[-1, 1]$.

The second result is more combinatorial and concerns the problem of enumerating sorting networks. A remarkable formula due to Stanley ([Sta84]) says that the number of all sorting networks on N elements is equal to

$$\frac{\binom{N}{2}!}{1^{N-1}3^{N-2} \dots (2N - 3)^1},$$

which is asymptotic to $\exp \left\{ \frac{N^2}{2} \log N + \left(\frac{1}{4} - \log 2 \right) N^2 + O(N \log N) \right\}$.

For relaxed sorting networks we have the following asymptotic estimate

Theorem 1.2. *For any $\kappa \in (0, 1)$ let $\mathcal{S}_{\kappa, \delta}^N$ be the number of δ -relaxed sorting networks on N elements of length $M = \lfloor \frac{1}{2}N^{1+\kappa}(N - 1) \rfloor$. We have*

$$\mathcal{S}_{\kappa, \delta}^N = \exp \left\{ \frac{1}{2} N^{1+\kappa} (N - 1) \log(N - 1) - \left(\frac{\pi^2}{6} + \varepsilon_\delta^N \right) N^{2-\kappa} \right\},$$

where ε_δ^N satisfies $\lim_{\delta \rightarrow 0} \lim_{N \rightarrow \infty} \varepsilon_\delta^N = 0$.

The asymptotics is analogous to that of Stanley's formula – the first term in the exponent corresponds simply to the number of all paths of required length, and, crucially, the factor $\frac{\pi^2}{6}$ corresponds to the energy of the sine curve process.

The proofs of Theorems 1.1 and 1.2 are given in Sect. 9. It would be an interesting problem to obtain analogous results for relaxed sorting networks reaching *exactly* the reverse permutation, not only being δ -close in the permutation topology. This case is not covered by the results of this paper, since the set of permutation processes reaching exactly the reverse permutation is not open, hence the lower bound of Theorem A does not apply.

2 Preliminaries

2.1 Permutons and stochastic processes. Permutons. Consider the space $\mathcal{M}([0, 1]^2)$ of all Borel probability measures on the unit square $[0, 1]^2$, endowed with the weak topology. A *permuton* is a probability measure $\mu \in \mathcal{M}([0, 1]^2)$ with uniform marginals. In other words, μ is the joint distribution of a pair of random variables (X, Y) , with X, Y taking values in $[0, 1]$ and having marginal distribution $X, Y \sim \mathcal{U}[0, 1]$. We will sometimes call the pair (X, Y) itself a permuton if there is no risk of ambiguity. A few simple examples of permutons are the *identity permuton* (X, X) , the *uniform permuton* (the distribution of two independent copies of X , which is the uniform measure on the square) or the *reverse permuton* $(X, 1 - X)$.

Permutons can be thought of as continuous limits of permutations in the following sense. Let \mathcal{S}_N be the symmetric group on N elements and let $\sigma \in \mathcal{S}_N$. We associate to σ its *empirical measure*

$$\mu_\sigma = \frac{1}{N} \sum_{i=1}^N \delta_{\left(\frac{i}{N}, \frac{\sigma(i)}{N}\right)}, \quad (4)$$

which is an element of $\mathcal{M}([0, 1]^2)$. By a slight abuse of terminology we will sometimes identify σ with μ_σ . Since every such measure has uniform marginals on $\left\{\frac{1}{N}, \frac{2}{N}, \dots, 1\right\}$, it is not difficult to see that if a sequence of empirical measures converges weakly, the limiting measure will be a permuton. Conversely, every permuton can be realized as a limit of finite permutations, in the sense of weak convergence of empirical measures (see [HKM+13]). We will consider $\mathcal{M}([0, 1]^2)$ endowed with the Wasserstein distance corresponding to the Euclidean metric on $[0, 1]^2$, under which the distance of measures μ and ν is given by

$$d_W(\mu, \nu) = \inf_{\{(X, Y), (X', Y')\}} \mathbb{E} \left[\sqrt{(X - X')^2 + (Y - Y')^2} \right],$$

where the infimum is over all couplings of (X, Y) and (X', Y') such that $(X, Y) \sim \mu$, $(X', Y') \sim \nu$.

The path space \mathcal{D} and stochastic processes. A natural setting for analyzing trajectories of particles in random permutation sequences is to consider $\mathcal{D} = \mathcal{D}([0, T], [0, 1])$, the space of all càdlàg paths from $[0, T]$ to $[0, 1]$. We endow it with

the standard Skorokhod topology, metrized by a metric ρ under which \mathcal{D} is separable and complete. By $\mathcal{M}(\mathcal{D})$ we will denote the space of all Borel probability measures on \mathcal{D} , endowed with the weak topology. It will be convenient to metrize $\mathcal{M}(\mathcal{D})$ by the Wasserstein distance, under which the distance between measures μ and ν is given by

$$d_{\mathcal{W}}(\mu, \nu) = \inf_{(X, Y)} \mathbb{E} [\rho(X, Y)],$$

where the infimum is over all couplings (X, Y) such that $X \sim \mu$, $Y \sim \nu$. We will also make use of the Wasserstein distance associated to the supremum norm, given by

$$d_{\mathcal{W}}^{sup}(\mu, \nu) = \inf_{(X, Y)} \mathbb{E} [\|X - Y\|_{sup}],$$

where $\|\cdot\|_{sup}$ is the supremum norm on \mathcal{D} and again the infimum is over all couplings (X, Y) as above.

Given two times $0 \leq s \leq t \leq T$ and a stochastic process $X = (X_t, 0 \leq t \leq T)$ with distribution $\mu \in \mathcal{M}(\mathcal{D})$, by $\mu_{s,t} \in \mathcal{M}([0, 1]^2)$ we will denote the distribution of the marginal (X_s, X_t) . Note that the projection $\mu \mapsto \mu_{s,t}$ is continuous as a map from $\mathcal{M}(\mathcal{D})$ to $\mathcal{M}([0, 1]^2)$ as long as paths $X \sim \mu$ sampled from μ have almost surely no jumps at times s and t . We will sometimes implicitly identify the stochastic process with its distribution when there is no risk of misunderstanding.

Permutation processes and permuton processes. Consider a permutation-valued path $\eta^N = (\eta_t^N, 0 \leq t \leq T)$, with η_t^N taking values in the symmetric group \mathcal{S}_N . We will always assume that η^N is càdlàg as a map from $[0, T]$ to \mathcal{S}_N . Let $\eta^N(i) = (\eta_t^N(i), 0 \leq t \leq T)$ be the trajectory of i under η^N and let $X^{\eta^N}(i) = \frac{1}{N}\eta^N(i)$ be the rescaled trajectory. We define the empirical measure

$$\mu^{\eta^N} = \frac{1}{N} \sum_{i=1}^N \delta_{X^{\eta^N}(i)}, \quad (5)$$

where $\delta_{X^{\eta^N}(i)}$ is the delta measure concentrated on the trajectory $X^{\eta^N}(i)$.

The associated *permutation process* $X^{\eta^N} = (X_t^{\eta^N}, 0 \leq t \leq T)$ is obtained by choosing $i = 1, \dots, N$ uniformly at random and following the path $X^{\eta^N}(i)$. In other words, X^{η^N} is a random path with values in $[0, 1]$ whose distribution is $\mu^{\eta^N} \in \mathcal{M}(\mathcal{D})$. If η^N is fixed, the only randomness here comes from the random choice of the particle i . Note that at each time t the marginal distribution of $X_t^{\eta^N}$ is uniform on $\{\frac{1}{N}, \frac{2}{N}, \dots, 1\}$.

A *permuton process* is a stochastic process $X = (X_t, 0 \leq t \leq T)$ taking values in $[0, 1]$, with continuous sample paths and such that for every $t \in [0, T]$ the marginal X_t is uniformly distributed on $[0, 1]$. The name is justified by observing that if π is the distribution of X , then for any fixed $s, t \in [0, T]$ the joint distribution

$\pi_{s,t} \in \mathcal{M}([0, 1]^2)$ of (X_s, X_t) defines a permuton. As explained in the next subsection, permuton processes arise naturally as limits of permutation processes defined above.

Since every permutation process has marginals uniform on $\{\frac{1}{N}, \frac{2}{N}, \dots, 1\}$, we will call it an *approximate permuton process*. By \mathcal{P} we will denote the space of all permuton processes and approximate permuton processes, treated as a subspace of $\mathcal{M}(\mathcal{D})$ (with the same topology and the metric $d_{\mathcal{W}}$).

Random permutation and permuton processes. A *random permuton process* is a permuton process chosen from some probability distribution on the space of all permuton processes, i.e., a random variable X , defined for a probability space Ω , such that $X(\omega)$ is a permuton process for $\omega \in \Omega$. By identifying the random variable with its distribution we can also think of a random permuton process as a random element of $\mathcal{M}(\mathcal{P})$. In this setting, with weak topology on $\mathcal{M}(\mathcal{P})$, one can consider convergence in distribution of random permuton processes X_n to a (possibly also random) permuton process X .

One can prove (see [RVV19]) that if a sequence of random permutation processes X^{η^N} converges in distribution, then the limit is a permuton process (in general also random). Of particular interest will be sequences of random permutation-valued paths η^N (coming for example from the interchange process) such that the corresponding permutation processes X^{η^N} converge in distribution to a *deterministic* permuton process (for example the sine curve process described below).

For any random permuton process X we define its associated *random particle process* $\bar{X} = \mathbb{E}_\omega X(\omega)$, which is a process with a deterministic distribution, obtained by first sampling a permuton process $X(\omega)$ and then sampling a random path according to $X(\omega)$.

To elucidate the difference between random and deterministic permuton processes, consider a random permuton process X and its associated random particle process \bar{X} . If we sample an outcome $X(\omega)$ and then a path from $X(\omega)$, then obviously the distribution of paths will be the same as for \bar{X} . However, consider now sampling an outcome $X(\omega)$ and then sampling independently two paths from $X(\omega)$. The distribution of a pair of paths obtained in this way will not in general be the same as the distribution of two independent copies sampled from \bar{X} , since the paths might be correlated within the outcome $X(\omega)$. The following general lemma will be useful later for showing that limits of certain random permutation processes are in fact deterministic ([RV17, Lemma 3]):

LEMMA 2.1. *Let K be a compact metric space and let μ be a random probability measure on K , i.e., a random variable with values in $\mathcal{M}(K)$. Let X and Y be two independent samples from an outcome of μ and let Z be a sample from an outcome of an independent copy of μ . If (X, Y) , as a K^2 -valued random variable, has the same distribution as (X, Z) , then μ is in fact deterministic, i.e., there exists $\nu \in \mathcal{M}(K)$ such that $\mu = \nu$ almost surely.*

Energy. Here we introduce several related notions of energy for paths, permutations, permutons and permuton processes.

Given a path $\gamma : [0, T] \rightarrow [0, 1]$ and a finite partition $\Pi = \{0 = t_0 < t_1 < \dots < t_k = T\}$ we define the *energy of γ with respect to Π* as

$$\mathcal{E}^\Pi(\gamma) = \frac{1}{2} \sum_{i=1}^k \frac{|\gamma(t_i) - \gamma(t_{i-1})|^2}{t_i - t_{i-1}}, \quad (6)$$

and the *energy of γ* as

$$\mathcal{E}(\gamma) = \sup_{\Pi} \mathcal{E}^\Pi(\gamma), \quad (7)$$

where the supremum is over all finite partitions $\Pi = \{0 = t_0 < t_1 < \dots < t_k = T\}$. For a path which is not absolutely continuous the supremum is equal to $+\infty$. If a path γ is differentiable, its energy is equal to

$$\frac{1}{2} \int_0^T \dot{\gamma}(s)^2 ds.$$

For a permutation $\sigma \in \mathcal{S}_N$ we define its energy as

$$I(\sigma) = \frac{1}{2} \left(\frac{1}{N} \sum_{i=1}^N \left(\frac{\sigma(i) - i}{N} \right)^2 \right). \quad (8)$$

Likewise, for a permuton $\mu \in \mathcal{M}([0, 1]^2)$ its energy is defined by

$$I(\mu) = \frac{1}{2} \mathbb{E}|X - Y|^2, \quad (9)$$

where the pair (X, Y) has distribution μ . If $\mu = \mu_\sigma$ is the empirical measure of a permutation $\sigma \in \mathcal{S}_N$, defined by (4), then we have $I(\mu_\sigma) = I(\sigma)$. Note also that $I = I(\mu)$ is a continuous function of μ in the weak topology on $\mathcal{M}([0, 1]^2)$.

Finally, we define the energy of a permuton process π as

$$I(\pi) = \mathbb{E}_{\gamma \sim \pi} \mathcal{E}(\gamma), \quad (10)$$

where the expectation is over paths γ sampled from π . We can extend this definition to any process $\pi \in \mathcal{M}(\mathcal{D})$ by adopting the convention that $I(\pi) = +\infty$ if paths sampled from π are not absolutely continuous almost surely. The function I will turn out to correspond to the rate function in large deviation bounds for random permuton process. It can be checked that I is lower semicontinuous (in the weak topology on \mathcal{P}) and its level sets $\{\pi \in \mathcal{P} : I(\pi) \leq C\}$ are compact.

We will also use the notation

$$I^\Pi(\pi) = \mathbb{E}_{\gamma \sim \pi} \mathcal{E}^\Pi(\gamma) \quad (11)$$

to denote the approximation of energy of π associated to the finite partition Π . The following lemma will be useful in characterizing the large deviation rate function in terms of these approximations

LEMMA 2.2. *For any process $\pi \in \mathcal{M}(\mathcal{D})$ we have*

$$I(\pi) = \sup_{\Pi} I^{\Pi}(\pi),$$

where the supremum is taken over all finite partitions $\Pi = \{0 = t_0 < t_1 < \dots < t_k = T\}$.

Proof. Let $\Pi_n = \{0 < \frac{1}{2^n} < \frac{2}{2^n} < \dots < 1\}$, $n = 0, 1, 2, \dots$, be the sequence of dyadic partitions of $[0, 1]$. It is elementary to show that if a path γ is continuous, then $\mathcal{E}(\gamma) = \lim_{n \rightarrow \infty} \mathcal{E}^{\Pi_n}(\gamma)$. Note that if Π' is a refinement of Π , then we have $\mathcal{E}^{\Pi}(\gamma) \leq \mathcal{E}^{\Pi'}(\gamma)$, thus $\mathcal{E}^{\Pi_n}(\gamma) \rightarrow \mathcal{E}(\gamma)$ monotonically as $n \rightarrow \infty$. Now we apply the monotone convergence theorem to get the same convergence for the expectations $\mathbb{E}_{\gamma \sim \pi} \mathcal{E}^{\Pi_n}(\gamma)$. \square

The interchange process. The *interchange process* on the interval $\{1, \dots, N\}$ is a Markov process in continuous time defined in the following way. Consider particles labelled from 1 to N on a line with N vertices. Each edge has an independent exponential clock that rings at rate 1. Whenever a clock rings, the particles at the endpoints of the corresponding edge swap places. By comparing the initial position of each particle with its position after time t we obtain a random permutation of $\{1, \dots, N\}$.

Formally, we define the state space of the process as consisting of permutations $\eta \in \mathcal{S}_N$, with the notation $\eta = (x_1, \dots, x_N)$ indicating that the particle with label i is at the position x_i , or in other words, $x_i = \eta(i)$. The dynamics is given by the generator

$$(\mathcal{L}f)(\eta) = \frac{1}{2} N^\alpha \sum_{x=1}^{N-1} (f(\eta^{x,x+1}) - f(\eta)), \quad (12)$$

where $\eta^{x,x+1}$ is the configuration η with particles at locations x and $x+1$ swapped and $\alpha \in (1, 2)$ is a fixed parameter (introduced so that we will be able to consider the limit $N \rightarrow \infty$). Since we will also be considering variants of this process with modified rates, we will often refer to the process with generator \mathcal{L} as the *unbiased interchange process*.

The interchange process defines a probability distribution on permutation-valued paths $\eta^N = (\eta_t^N, 0 \leq t \leq T)$ for any $T \geq 0$. Consider now the permutation process X^{η^N} associated to η^N , that is, sample η^N according to the interchange process, pick a particle uniformly at random and follow its trajectory in η^N . The distribution μ^{η^N} of X^{η^N} , defined by (5), is then a random element of $\mathcal{M}(\mathcal{D})$.

The position of a random particle in the interchange process will be distributed as the stationary simple random walk (in continuous time) on the line $\{1, \dots, N\}$. If we look at timescales much shorter than N^2 , typically each particle will have distance $o(N)$ from its origin, so the permutation obtained at time t such that $tN^\alpha \ll N^2$ will be close (in the sense of permutations) to the identity permutation. As mentioned in the introduction, we will be interested in large deviation bounds for rare events such as seeing a nontrivial permutation after a short time.

2.2 Euler equations and generalized incompressible flows. Let us now discuss the connection to fluid dynamics and incompressible flows (the discussion here follows [AF09] and [BFS09]). The Euler equations describe the motion of an incompressible fluid in a domain $D \subseteq \mathbb{R}^d$ in terms of its *velocity field* $u(t, x)$, which is assumed to be divergence-free. The evolution of u is given in terms of the *pressure field* p

$$\begin{cases} \partial_t u + (u \cdot \nabla)u = -\nabla p & \text{in } [0, T] \times D, \\ \operatorname{div} u = 0 & \text{in } [0, T] \times D, \\ u \cdot n = 0 & \text{on } [0, T] \times \partial D, \end{cases}$$

where the second equation encodes the incompressibility constraint and the third equation means that u is parallel to the boundary ∂D .

Assuming u is smooth, the trajectory $g(t, x)$ of a fluid particle initially at position x is obtained by solving the equation

$$\begin{cases} \dot{g}(t, x) = u(t, g(t, x)), \\ g(0, x) = x. \end{cases}$$

Since u is assumed to be divergence-free, the flow map $\Phi_g^t : D \rightarrow D$ given by $\Phi_g^t(x) = g(t, x)$ is a measure-preserving diffeomorphism of D for each $t \in [0, T]$. This means that $(\Phi_g^t)_* \mu_D = \mu_D$, where from now on by f_* we denote the pushforward map on measures, associated to f , and μ_D is the Lebesgue measure inside D . Denoting by $\operatorname{SDiff}(D)$ the space of all measure-preserving diffeomorphisms of D , we can rewrite the Euler equations in terms of g

$$\begin{cases} \ddot{g}(t, x) = -\nabla p(t, g(t, x)) & \text{in } [0, T] \times D, \\ g(0, x) = x & \text{in } D, \\ g(t, \cdot) \in \operatorname{SDiff}(D) & \text{for each } t \in [0, T]. \end{cases} \tag{13}$$

Arnold proposed an interpretation according to which the equation above can be viewed as a geodesic equation on $\operatorname{SDiff}(D)$. Thus one can look for solutions to (13) by considering the variational problem

$$\text{minimize } \frac{1}{2} \int_0^T \int_D |\dot{g}(t, x)|^2 d\mu_D(x) dt \tag{14}$$

among all paths $g(t, \cdot) : [0, T] \rightarrow \operatorname{SDiff}(D)$ such that $g(0, \cdot) = f$, $g(T, \cdot) = h$ for some prescribed $f, h \in \operatorname{SDiff}(D)$ (by right invariance without loss of generality f can be assumed to be the identity). The pressure p then arises as a Lagrange multiplier coming from the incompressibility constraint.

Shnirelman proved ([Shn87]) that in dimensions $d \geq 3$ the infimum in this minimization problem is not attained in general and in dimension $d = 2$ there exist

diffeomorphisms $h = g(T, \cdot)$ which cannot be connected to the identity map by a path with finite action. This motivated Brenier ([Bre89]) to consider the following relaxation of this problem. With $C(D)$ denoting the space of continuous paths from $[0, T]$ to D and $\mathcal{M}(C(D))$ the set of probability measures on $C(D)$, the variational problem is

$$\text{minimize} \quad \int_{C(D)} \left(\frac{1}{2} \int_0^T |\dot{\gamma}(t)|^2 dt \right) d\pi(\gamma) \quad (15)$$

over all $\pi \in \mathcal{M}(C(D))$ satisfying the constraints

$$\begin{cases} \pi_{0,T} = (id, h)_* \mu_D, \\ \pi_t = \mu_D \quad \text{for each } t \in [0, T], \end{cases} \quad (16)$$

where $\pi_{0,T}$, π_t denote the marginals of π at times respectively $0, T$ and at time t .

Following Brenier, a probability measure $\pi \in \mathcal{M}(C(D))$ satisfying constraints (16) is called a *generalized incompressible flow* between the identity id and h . To see that indeed (15) is a relaxation of (14), note that any sufficiently regular path $g(t, \cdot) : [0, T] \rightarrow \text{SDiff}(D)$, for example corresponding to a solution of (13), induces a generalized incompressible flow given by $\pi = (\Phi_g)_* \mu_D$, where as before $\Phi_g(x) = g(\cdot, x)$. As evidenced by the sine curve process mentioned in the introduction, the converse is false – trajectories of particles sampled from a generalized flow can cross each other or split at a later time when starting from the same position, which is not possible for classical, smooth flows. We refer the reader to [Bre08] for an interesting discussion of physical relevance of this phenomenon.

The problem admits a natural further relaxation in which the target map is “non-deterministic”, in the sense that we have $\pi_{0,T} = \mu$ with μ being an arbitrary probability measure supported on $D \times D$ and having uniform marginals on each coordinate, not necessarily of the form $\mu = (id, h)_* \mu_D$ for some map h . From now on whenever we refer to problem (15) or generalized incompressible flows we will be always considering this more general variant.

The connection between the generalized problem (15) and the original Euler equations (13) is provided by a theorem due to Ambrosio and Figalli ([AF09]), with earlier weaker results by Brenier ([Bre99]). Roughly speaking, they showed that given a measure μ with uniform marginals there exists a pressure function $p(t, x)$ such that the following holds – one can replace the problem of minimizing the functional (15) over incompressible flows satisfying $\pi_{0,T} = \mu$ by an easier problem in which the incompressibility constraint is dropped, provided one adds to the functional a Lagrange multiplier given by p . We refer the reader to [AF09, Section 6] for a precise formulation and further results on regularity of p .

In particular, if π is optimal for (15) and the corresponding pressure p is smooth enough, their result implies that almost every path γ sampled from π minimizes the

functional

$$\gamma \mapsto \int_0^T \left(\frac{1}{2} |\dot{\gamma}(t)|^2 - p(t, \gamma(t)) \right) dt. \quad (17)$$

In that case the equation $\ddot{g}(t, x) = -\nabla p(t, g(t, x))$ from (13) is nothing but the Euler-Lagrange equation for extremal points of the functional (17). We can therefore, at least under some regularity assumptions on p , think of generalized incompressible flows as solutions to (13) in which instead of having a diffeomorphism we assume random initial conditions for each particle.

From now on let us restrict the discussion to $D = [0, 1]$, which will be most directly relevant to the results of this paper. In this case the original problem (14) is somewhat uninteresting, since the only measure-preserving diffeomorphisms of $[0, 1]$ are $f(x) = x$ and $f(x) = 1 - x$. However, the relaxed problem (15) is non-trivial and indeed for the target map $h(x) = 1 - x$ and $T = 1$ the unique optimal solution is given by the sine curve process.

In this setting, the reader may recognize that generalized incompressible flows are in fact the same objects as permuton processes. The term *measure-preserving plans* is used in [AF09] for what we call permutons. The functional minimized in (15) is the energy $I(\pi)$ of a permuton process, defined in (10). In this language the optimization problem we are interested in can be rephrased as follows:

$$\text{find } \inf_{\substack{\pi \in \mathcal{P} \\ \pi_{0,T} = \mu}} I(\pi), \quad (18)$$

where the infimum is over all permuton processes $\pi \in \mathcal{P}$ satisfying $\pi_{0,T} = \mu$ for a given permuton $\mu \in \mathcal{M}([0, 1]^2)$.

Generalized solutions to Euler equations. We will say that a permuton process π is a *generalized solution to Euler equations* if there exists a function $p : [0, T] \times [0, 1] \rightarrow \mathbb{R}$, differentiable in the second variable, such that almost every path $x : [0, T] \rightarrow [0, 1]$ sampled from π satisfies the equation

$$\begin{cases} x'(t) = v(t), \\ v'(t) = -\partial_x p(t, x(t)), \end{cases} \quad (19)$$

for $t \in [0, T]$. This is of course equivalent to $x''(t) = -\partial_x p(t, x(t))$.

By the remarks above, if π minimizes the energy in (18) and the associated pressure p is smooth enough, then π is always a generalized solution to Euler equations. However, this is only a necessary condition – for a discussion of corresponding sufficient conditions see [BFS09].

2.3 Proof outline and structure of the paper. Let us now give a brief outline of the proof strategy for Theorems A and B. For the lower bound, given a process X we construct a perturbation of the interchange process (defined by introducing asymmetric jump rates based on 19) for which a law of large numbers

holds, namely, the distribution of the path of a random particle converges to a deterministic limit (which is the distribution of X). The large deviation principle is then proved by estimating the Radon–Nikodym derivative between the biased process and the original one.

The key property which makes this construction possible is that the process X satisfies a second order ODE given by (19), so its trajectories are fully specified by the particle’s position and velocity (the latter chosen initially from a mean zero distribution). The biased process is then constructed by endowing each particle with an additional parameter keeping track of its velocity, but we perform an additional change variables, working instead of velocity with a variable we call *color*. The advantage of this is that the uniform distribution of colors is stationary when the jump rates are properly chosen, which will greatly facilitate the analysis. An additional technical difficulty arises if the velocity distribution of X is time-dependent or not regular enough near the boundary, in which case we first approximate X by a process with a sufficiently regular and piecewise time-homogeneous velocity distribution.

To prove the law of large numbers we need to show that in the biased interchange process particles’ trajectories behave approximately like independent samples from X . This requires proving that their velocities remain uncorrelated when averaged over time and is accomplished by means of a local mixing result called the *one block estimate*. It is here that we rely on stationarity of the uniform distribution of colors in the biased process and the fact that X has velocity zero on average.

The strategy for proving the upper bound is somewhat simpler. We consider a family of exponential martingales similar to the one employed in analyzing independent random walks and use the one block estimate to show that the particles’ velocities are typically nonnegatively correlated. This enables us to prove the large deviation upper bound for compact sets and the extension to closed sets is done by proving exponential tightness.

Structure of the paper. The rest of the paper is structured as follows. In Sect. 3 we introduce the change of variables needed to define the process with colors and prove the approximation result for X mentioned above (Proposition 3.7). In Sect. 4 we define the biased interchange process and derive the conditions on its rates which guarantee stationarity. Section 5 contains the proof of the law of large numbers for the biased interchange process (Theorem 5.1). In Sect. 6 we prove two variants of the one block estimate – one needed for the large deviation upper bound (Lemma 6.2) and a more involved one needed for the proof of the law of large numbers (Lemma 5.4). In Sect. 7 these pieces are then used to prove the large deviation lower bound (Theorem 7.3). Section 8 is devoted to the proof of the large deviation upper bound (Theorem 8.4) and is independent of the previous sections (apart from the use of Lemma 6.2). Finally, in Sect. 9 we prove Theorem 1.1 and Theorem 1.2 on relaxed sorting networks.

3 ODEs and Generalized Solutions to Euler Equations

Regularity assumptions and properties of generalized solutions. Suppose π is a generalized solution to Euler Eq. (19) and let X be a process with distribution π . For the proof of the large deviation lower bound we will need to impose additional regularity assumptions on π . For $t \in [0, T]$ let μ_t denote the joint distribution of $(x(t), x'(t))$ when x is sampled according to π . In particular, μ_0 is the joint distribution of the initial conditions of the ODE (19). If $\Phi^{t,s}(x, v)$ denotes the solution $x(s)$ of (19) satisfying $(x(t), v(t)) = (x, v)$, then $\mu_t = \Phi_*^{0,t} \mu_0$.

We will assume that each μ_t has a density $\rho_t(x, v)$ with respect to the Lebesgue measure on $[0, 1] \times \mathbb{R}$. For $x \in [0, 1]$ and $t \in [0, T]$ let $\mu_{t,x}$ denote the conditional distribution of v , given x , at time t . In addition we assume that for $x = 0$ or 1 the distribution $\mu_{t,x}$ is a delta mass at 0 , as otherwise the process X cannot stay confined to $[0, 1]$ and have mean velocity zero everywhere (see the discussion of incompressibility below).

Let $F_{t,x}$ denote the cumulative distribution function of $\mu_{t,x}$ and let $V_t(x, \cdot) : [0, 1] \rightarrow \mathbb{R}$ be the quantile function of $\mu_{t,x}$, defined for $x \in [0, 1]$ and $\phi \in (0, 1]$ by

$$V_t(x, \phi) = \inf \{v \in \mathbb{R} \mid F_{t,x}(v) \geq \phi\}$$

and $V_t(x, 0) = \inf \{v \in \mathbb{R} \mid F_{t,x}(v) > 0\}$. In particular for $x = 0, 1$ we have $V_t(x, \phi) = 0$.

Assumption 3.1. Throughout the paper, we will assume that for a generalized solution to Euler equations π the following properties are satisfied

- (1) the pressure function $(t, x) \mapsto p(t, x)$ in (19) is measurable in t and differentiable in x , with the derivative $\partial_x p(t, x)$ Lipschitz continuous in x (with the Lipschitz constant uniform in t);
- (2) there exists a compact set $K \subseteq [0, 1] \times \mathbb{R}$ such that for each $t \in [0, T]$ the density ρ_t is supported in K ;
- (3) for $t \in [0, T], x \in [0, 1]$ the support of $\mu_{t,x}$ is a connected interval in \mathbb{R} ;
- (4) the density ρ_t is continuously differentiable in t, x and v for each $t \in [0, T]$ and x, v in the interior of the support of ρ_t .

Let us comment on the relevance of these assumptions. Assumption (1) will guarantee uniqueness of solutions to (19). Assumption (2) implies that the velocity of a particle moving along a path sampled from π stays uniformly bounded in time. Assumption (3) implies that for any $x \in (0, 1)$ and $\phi \in [0, 1]$ we have $F_{t,x}(V_t(x, \phi)) = \phi$, i.e., $V_t(x, \cdot)$ is the inverse function of $F_{t,x}$. Assumptions (3) and (4) imply that $V_t(x, \phi)$ is a continuous function of t, x, ϕ and it is continuously differentiable in all variables for $x \in (0, 1)$.

Note that for $V_t(x, \phi)$ to be differentiable at $\phi = 0, 1$, the distribution function $F_{t,x}$ necessarily has to be non-differentiable at corresponding v such that $F_{t,x}(v) = \phi$. This is why we can require the density ρ_t to be smooth only in the interior of its support and not at the boundary.

From now on we assume that π is a fixed generalized solution to Euler equations, satisfying Assumptions (3.1). Almost every path $x : [0, T] \rightarrow [0, 1]$ sampled from π satisfies the ODE

$$\begin{cases} x'(t) = v(t), \\ v'(t) = -\partial_x p(t, x(t)). \end{cases} \quad (20)$$

Note that since π is a permuton process, each measure μ_t satisfies the *incompressibility* condition, meaning that its projection onto the first coordinate is equal to the uniform measure on $[0, 1]$. This is equivalent to the property that for any test function $f : [0, 1] \rightarrow \mathbb{R}$ we have

$$\int f(x) d\mu_t(x, v) = \int_0^1 f(x) dx.$$

An important consequence of the incompressibility assumption is that under μ_t the velocity has mean zero at each x , that is, we have the following

LEMMA 3.2. *For any $t \in [0, T]$ and $x \in [0, 1]$ we have*

$$\int v d\mu_{t,x}(v) = 0.$$

Proof. Consider any test function $f : [0, 1] \rightarrow \mathbb{R}$ and write

$$\int f(x) d\mu_{t+s}(x, v) = \int f(x) d(\Phi_*^{t,t+s} \mu_t)(x, v) = \int f(\Phi^{t,t+s}(x, v)) d\mu_t(x, v).$$

By incompressibility the integral above is always equal to $\int_0^1 f(x) dx$, in particular does not depend on time. On the other hand its derivative with respect to s is

$$\begin{aligned} \frac{d}{ds} \int f(x) d\mu_{t+s}(x, v) &= \frac{d}{ds} \int f(\Phi^{t,t+s}(x, v)) d\mu_t(x, v) \\ &= \int f'(\Phi^{t,t+s}(x, v)) \frac{d\Phi^{t,t+s}}{ds}(x, v) d\mu_t(x, v). \end{aligned}$$

Since $\Phi^{t,t+s}(x, v)|_{s=0} = x$ and $\frac{d\Phi^{t,t+s}}{ds}(x, v)|_{s=0} = v$, by evaluating the derivative at $s = 0$ we arrive at $\int f'(x)v d\mu_t(x, v) = 0$. Since $\int g(x, v) d\mu_t(x, v) = \int g(x, v) d\mu_{t,x}(v) dx$ for any measurable g and f was an arbitrary test function, the claim of the lemma holds for almost every x . Since we have assumed that μ_t has a continuous density, the claim in fact holds for all x , which ends the proof. \square

We will also make use of an explicit evolution equation that the densities ρ_t have to satisfy. This is the content of the following lemma.

LEMMA 3.3. *For any $t \in [0, T]$ and x, v in the interior of the support of ρ_t we have*

$$\frac{\partial \rho_t}{\partial t}(x, v) = -v \frac{\partial \rho_t}{\partial x}(x, v) + \partial_x p(t, x) \frac{\partial \rho_t}{\partial v}(x, v).$$

Proof. Let $f : [0, 1] \times \mathbb{R} \rightarrow \mathbb{R}$ be any test function and consider the integral

$$I_{t+s} = \int f(x, v) d\mu_{t+s}(x, v).$$

On the one hand, its derivative with respect to s is equal to

$$\begin{aligned} \frac{d}{ds} I_{t+s} &= \frac{d}{ds} \int f(x, v) d\mu_{t+s}(x, v) \\ &= \frac{d}{ds} \int f \left(\Phi^{t,t+s}(x, v), \frac{d\Phi^{t,t+s}}{ds}(x, v) \right) \rho_t(x, v) dx dv \\ &= \int \left[\frac{\partial f}{\partial x} \left(\Phi^{t,t+s}(x, v), \frac{d\Phi^{t,t+s}}{ds}(x, v) \right) \frac{d\Phi^{t,t+s}}{ds}(x, v) \right. \\ &\quad \left. + \frac{\partial f}{\partial v} \left(\Phi^{t,t+s}(x, v), \frac{d\Phi^{t,t+s}}{ds}(x, v) \right) \frac{d^2\Phi^{t,t+s}}{ds^2}(x, v) \right] \rho_t(x, v) dx dv. \end{aligned}$$

Since $\Phi^{t,t+s}(x, v)$ is a solution to (20), we have $\frac{d\Phi^{t,t+s}}{ds}(x, v)|_{s=0} = v$ and $\frac{d^2\Phi^{t,t+s}}{ds^2}(x, v)|_{s=0} = -\partial_x p(t, x)$, which gives us

$$\frac{d}{ds} I_{t+s} \Big|_{s=0} = \int \left(\frac{\partial f}{\partial x}(x, v)v - \frac{\partial f}{\partial v}(x, v)\partial_x p(t, x) \right) \rho_t(x, v) dx dv.$$

Performing integration by parts with respect to x for the first term and with respect to v for the second term gives (noting that f has compact support so the boundary terms vanish)

$$\frac{d}{ds} I_{t+s} \Big|_{s=0} = - \int f(x, v)v \frac{\partial \rho_t}{\partial x}(x, v) dx dv + \int f(x, v)\partial_x p(t, x) \frac{\partial \rho_t}{\partial v}(x, v) dx dv.$$

On the other hand, we have

$$\begin{aligned} \frac{d}{ds} I_{t+s} &= \frac{d}{ds} \int f(x, v) d\mu_{t+s}(x, v) = \frac{d}{ds} \int f(x, v)\rho_{t+s}(x, v) dx dv \\ &= \int f(x, v) \frac{\partial \rho_{t+s}}{\partial s}(x, v) dx dv, \end{aligned}$$

so

$$\frac{d}{ds} I_{t+s} \Big|_{s=0} = \int f(x, v) \frac{\partial \rho_t}{\partial t}(x, v) dx dv$$

and thus

$$\int f(x, v) \left(-v \frac{\partial \rho_t}{\partial x}(x, v) + \partial_x p(t, x) \frac{\partial \rho_t}{\partial v}(x, v) - \frac{\partial \rho_t}{\partial t}(x, v) \right) dx dv.$$

Since the test function f was arbitrary, the equation from the statement of the lemma must hold for every t, x, v as assumed. □

The colored trajectory process. Let $X = (X_t, 0 \leq t \leq T)$ be the permuton process with distribution π . For the large deviation lower bound we will need to construct a suitable interacting particle system in which the behavior of a random particle mimics that of the permuton process X . A crucial ingredient will be a property analogous to Lemma 3.2, i.e., having velocity distribution whose mean is locally zero. Instead of working with velocity v , whose distribution $\rho_t(x, v)$ at a given site x may change in time, it will be more convenient to perform a change variables and use another variable ϕ , which we call *color*, whose distribution will be invariant in time.

Recall that under Assumptions (3.1) the distribution function $F_{t,x}(\cdot)$ and the quantile function $V_t(x, \cdot)$ are related by

$$\begin{cases} F_{t,x}(V_t(x, \phi)) = \phi, \\ V_t(x, F_{t,x}(v)) = v, \end{cases} \quad (21)$$

for any $t \in [0, T]$, $x \in (0, 1)$, $\phi \in [0, 1]$, $v \in \text{supp } \mu_{t,x}$.

The reason for introducing the variable ϕ is the following elementary property – if ϕ is sampled from the uniform distribution on $[0, 1]$, then $V_t(x, \phi)$ is distributed according to $\mu_{t,x}$. Thus instead of working with (x, v) variables in the ODE (20), where the distribution of v evolves in time, we can set up an ODE for x and ϕ such that the joint distribution of (x, ϕ) will be uniform on $[0, 1]^2$ at each time. The velocity v and its distribution can then be recovered via the equation $v = V_t(x, \phi)$.

Let $(x(t), v(t))$ be a solution to (20) such that $x(t) \neq 0, 1$ and let

$$\phi(t) = F_{t,x(t)}(v(t)).$$

Let us derive the ODE that $(x(t), \phi(t))$ satisfies. Since $(x(t), v(t))$ is a solution of (20), we have

$$\begin{aligned} \phi'(t) &= \frac{\partial F_{t,x(t)}}{\partial t}(v(t)) + \frac{\partial F_{t,x(t)}}{\partial x}(v(t))x'(t) + \frac{\partial F_{t,x(t)}}{\partial v}(v(t))v'(t) \\ &= \frac{\partial F_{t,x(t)}}{\partial t}(v(t)) + \frac{\partial F_{t,x(t)}}{\partial x}(v(t))v(t) + \rho_t(x(t), v(t))[-\partial_x p(t, x(t))]. \end{aligned}$$

Lemma 3.3 implies that

$$\frac{\partial F_{t,x(t)}}{\partial t}(x(t), v(t)) = - \int_{-\infty}^{v(t)} w \frac{\partial \rho_t}{\partial x}(x(t), w) + [\partial_x p(t, x(t))] \rho_t(x(t), v(t)),$$

which gives

$$\phi'(t) = \frac{\partial F_{t,x(t)}}{\partial x}(v(t))v(t) - \int_{-\infty}^{v(t)} w \frac{\partial \rho_t}{\partial x}(x(t), w)$$

and upon integrating by parts in the last integral we obtain

$$\phi'(t) = \int_{-\infty}^{v(t)} \frac{\partial F_{t,x(t)}}{\partial x}(x(t), w) dw. \tag{22}$$

Now, differentiating (21) with respect to x and ϕ gives

$$\begin{cases} \frac{\partial F_{t,x}}{\partial x}(V_t(x, \phi)) + \rho_t(x, \phi) \frac{\partial V_t}{\partial x}(x, \phi) = 0, \\ \rho_t(x, \phi) \frac{\partial V_t}{\partial \phi}(x, \phi) = 1. \end{cases}$$

Also by (21) we have $v(t) = V_t(x(t), \phi(t))$, so a change of variables $w = V_t(x(t), \psi)$ in (22) yields

$$\phi'(t) = R_t(x(t), \phi(t)),$$

where $R_t(x, \phi) = - \int_0^\phi \frac{\partial V_t}{\partial x}(x, \psi) d\psi$.

Thus we have shown that $(x(t), \phi(t))$ satisfies the ODE

$$\begin{cases} x'(t) = V_t(x(t), \phi(t)), \\ \phi'(t) = R_t(x(t), \phi(t)). \end{cases} \tag{23}$$

If $x(t) \neq 0, 1$, this equation is equivalent to (20), i.e., $(x(t), \phi(t))$ is a solution of (23) with initial conditions $(x(0), \phi(0)) = (x_0, \phi_0)$ if and only if $(x(t), v(t))$ is a solution of (20) with initial conditions $(x(0), v(0)) = (x_0, V_0(x_0, \phi_0))$. We also note that Lemma 3.2 expressed in terms of (x, ϕ) variables states that for each $t \in [0, T]$ and $x \in [0, 1]$ we have

$$\int_0^1 V_t(x, \psi) d\psi = 0. \tag{24}$$

From now on we work exclusively with (23). We will need to make two approximations necessary for the interacting particle system analysis later on. One is necessitated by the fact that the function $V_t(x, \phi)$ might not be smooth with respect to x at the boundaries $x = 0, 1$ (this happens, for example, for the sine curve process). We will therefore replace the function by its smooth approximation in a β -neighborhood of the boundary and in the end take $\beta \rightarrow 0$. The other approximation consists in dividing the time interval $[0, T]$ into intervals of length δ and approximating $V_t(x, \phi)$ for given x, ϕ with a piecewise-constant function of t . This will enable us to give a simple stationarity condition for the corresponding interacting particle system and in the end take $\delta \rightarrow 0$ as well.

Let $\beta \in (0, \frac{1}{4})$ and let $V_t^\beta(x, \phi)$ be a function with the following properties

- (a) $V_t^\beta(x, \phi)$ is continuously differentiable for every $t \in [0, T]$, $x \in [0, 1]$, $\phi \in [0, 1]$,
- (b) $V_t^\beta(x, \phi) = V_t(x, \phi)$ for $x \in [\beta, 1 - \beta]$ and $V_t^\beta(0, \phi) = V_t^\beta(1, \phi) = 0$,

- (c) for each $x \in [0, 1]$ we have $\int_0^1 V_t^\beta(x, \psi) d\psi = 0$,
- (d) $|V_t^\beta(x, \phi)| \leq |V_t(x, \phi)| + 1$,
- (e) we have $\lim_{\beta \rightarrow 0} \int_0^1 \int_0^1 |V_t^\beta(x, \phi)|^2 dx d\phi = \int_0^1 \int_0^1 |V_t(x, \phi)|^2 dx d\phi$.

The existence of such a function V_t^β is proved at the end of this section. By $(x^\beta(t), \phi^\beta(t))$ we will denote the solution to the ODE

$$\begin{cases} x'(t) = V_t^\beta(x(t), \phi(t)), \\ \phi'(t) = R_t^\beta(x(t), \phi(t)). \end{cases} \quad (25)$$

Take any $\delta > 0$ (to simplify notation we will assume that T is an integer multiple of δ , this will not influence the argument in any substantial way) and consider a partition $0 = t_0 < t_1 < \dots < t_M = T$ of $[0, T]$ into $M = \frac{T}{\delta}$ intervals of length δ , with $t_k = k\delta$. Let $V^{\beta, \delta}(t, x, \phi)$ be the piecewise-constant in time approximation of $V_t^\beta(x, \phi)$, defined by

$$V^{\beta, \delta}(t, x, \phi) = V_{t_k}^\beta(x, \phi) \quad \text{for } t \in [t_k, t_{k+1}), \quad k = 0, 1, \dots, M-1. \quad (26)$$

We can now define the piecewise-stationary process which will be our main tool in subsequent arguments. Consider the ODE

$$\begin{cases} y'(t) = V^{\beta, \delta}(t, y(t), \phi(t)), \\ \phi'(t) = R^{\beta, \delta}(t, y(t), \phi(t)), \end{cases} \quad (27)$$

where

$$R^{\beta, \delta}(t, y, \phi) = - \int_0^\phi \frac{\partial V^{\beta, \delta}}{\partial y}(t, y, \psi) d\psi.$$

Solutions to (27) exist and are unique as usual for any initial conditions, provided we interpret $(y'(t), \phi'(t))$ above as right-handed derivatives at $t = 0, t_1, t_2, \dots, t_{M-1}$ (we adopt this convention from now on).

Let $P^{\beta, \delta} = ((X_t^{\beta, \delta}, \Phi_t^{\beta, \delta}), 0 \leq t \leq T)$ be the stochastic process with values in $[0, 1]^2$ with the following distribution: choose $(X_0^{\beta, \delta}, \Phi_0^{\beta, \delta})$ uniformly at random from $[0, 1]^2$ and then take $(X_t^{\beta, \delta}, \Phi_t^{\beta, \delta}) = (y(t), \phi(t))$, where (y, ϕ) is the solution of the system (27) with initial conditions given by $(y(0), \phi(0)) = (X_0^{\beta, \delta}, \Phi_0^{\beta, \delta})$. We will call this process the *colored trajectory process* associated to (27).

We also define the process $P^\beta = ((X_t^\beta, \Phi_t^\beta), 0 \leq t \leq T)$, which is obtained in the same way as $P^{\beta, \delta}$ except that we follow solutions to (25) instead of (27), i.e., make no piecewise approximation in time of V_t^β .

The key property of the process $P^{\beta, \delta}$ is the following

LEMMA 3.4. *For each $t \in [0, T]$ the distribution of $(X_t^{\beta, \delta}, \Phi_t^{\beta, \delta})$ is uniform on $[0, 1]^2$.*

Proof. First we show that the process stays confined to $[0, 1]^2$. Because of uniqueness of solutions to (27) it is enough to show that if a solution starts in the interior of $[0, 1]^2$, it never reaches the boundary, or, equivalently, that if a solution is at the boundary at some t , it is actually at the boundary for all $s \in [0, T]$. If $y(t) = 0$ or 1 for any t , then $y'(t) = 0$, since $V^{\beta, \delta}(t, 0, \phi) = V^{\beta, \delta}(t, 1, \phi) = 0$ for any ϕ . By uniqueness of solutions we then have $y(t) \equiv 0$ or 1. If $\phi(t) = 0$ for any t , then $R^{\beta, \delta}(t, y, 0) = 0$ regardless of y , so as before $\phi'(t) = 0$ and $\phi(t) \equiv 0$. Finally, if $\phi(t) = 1$, then using the property (c) of the function $S_t^\beta(x, \phi)$ we have

$$R^{\beta, \delta}(t, y, 1) = - \int_0^1 \frac{\partial V^{\beta, \delta}}{\partial y}(t, y, \psi) d\psi = - \frac{\partial}{\partial y} \int_0^1 V^{\beta, \delta}(t, y, \psi) d\psi = 0,$$

so as before $\phi'(t) = 0$ and $\phi(t) \equiv 1$.

Now we observe that the form of $V^{\beta, \delta}$ and $R^{\beta, \delta}$ in (27) implies that the vector field $(V^{\beta, \delta}(t, \cdot, \cdot), R^{\beta, \delta}(t, \cdot, \cdot))$ is divergence-free at each t , so by Liouville’s theorem the uniform measure on $[0, 1]^2$ is invariant for the corresponding flow map. \square

In particular, the process $X^{\beta, \delta} = (X_t^{\beta, \delta}, 0 \leq t \leq T)$ is a permuton process. Crucially, we can couple it to the process X in a natural way. Consider (x_0, ϕ_0) chosen uniformly at random from $[0, 1]^2$ and take $(x(0), v(0)) = (x_0, V_0(x_0, \phi_0))$, resp. $(y(0), \phi(0)) = (x_0, \phi_0)$, as initial conditions for (23), resp. (27). By definition of $V_0(x, \phi)$, the pair $(x(0), v(0))$ has distribution given by μ_0 , so indeed the pair of solutions $(x(t), y(t))$ corresponding to the initial conditions above defines a coupling of X and $X^{\beta, \delta}$. From now on X and $X^{\beta, \delta}$ are always assumed to be coupled in this way.

It is readily seen that the statements above also hold for P^β instead of $P^{\beta, \delta}$, hence with a slight abuse of notation we can allow $\delta = 0$ and write $P^{\beta, 0} = P^\beta$, $X^{\beta, 0} = X^\beta$ etc.

Our goal in the remainder of this section is to show that, as $\beta, \delta \rightarrow 0$, the processes X and $X^{\beta, \delta}$ typically stay close to each other and have approximately the same Dirichlet energy, so in the probabilistic part of the arguments it will be enough to work with the process $(X^{\beta, \delta}, \Phi^{\beta, \delta})$, which is more convenient thanks to piecewise stationarity.

First we prove a simple lemma, showing that X^β is unlikely to ever be close to the boundary (so that approximation of X with X^β is meaningful as $\beta \rightarrow 0$).

LEMMA 3.5. *Let \mathbb{P} denote the law of the process X^β . Let*

$$B^\beta = \left\{ \exists t \in [0, T] X_t^\beta \notin [\beta, 1 - \beta] \right\}.$$

We have

$$\mathbb{P}(B^\beta) \xrightarrow{\beta \rightarrow 0} 0.$$

Proof. We will prove that $X_t^\beta \notin [0, \beta]$ with high probability as $\beta \rightarrow 0$ (the proof for $[1 - \beta, 1]$ is analogous). Suppose that y is a solution of (27) with initial condition $y(0) \notin [0, 2\beta]$ and that $y(t) \in [0, \beta]$ for some $t \in [0, T]$. Then there exists a time interval $[s, s']$ such that $y(s) = 2\beta$, $y(s') = \beta$ and $y(u) \in [\beta, 2\beta]$ for every $u \in [s, s']$. Without loss of generality we can assume that $[s, s'] \subseteq [t_k, t_{k+1})$ for some k (the other case is easily dealt with by further subdividing $[\beta, 2\beta]$ into two equal subintervals and repeating the argument for each of them). By the mean value theorem

$$|y(s) - y(s')| = (s' - s)y'(w)$$

for some $w \in [s, s']$. For $x \in [\beta, 2\beta]$ we have $V^\beta(w, x, \phi) = V_{t_k}(x, \phi)$, so $y'(w) = V_{t_k}(w, y(w), \phi(w))$. Since $|y(w)| \leq 2\beta$ and $V_{t_k}(x, \phi)$ is continuous at $x = 0$, we have $|y'(w)| \leq f(\beta)$ for some function f (depending only on V) satisfying $\lim_{\beta \rightarrow 0} f(\beta) = 0$.

As $|y(s) - y(s')| = \beta$, altogether this implies that $s' - s \geq \frac{\beta}{f(\beta)}$, i.e., if the process X^β starts outside $[0, 2\beta]$, it has to spend time at least $\frac{\beta}{f(\beta)}$ before it reaches $[0, \beta]$. Thus

$$\int_0^T \mathbb{1}_{\{X_s^\beta \in [0, \beta]\}} ds \geq \frac{\beta}{f(\beta)} \mathbb{1}_{\{\exists t \in [0, T] X_t^\beta \in [0, \beta]\}} \mathbb{1}_{\{X_0^\beta \notin [0, 2\beta]\}}.$$

Taking expectation yields

$$\mathbb{E} \int_0^T \mathbb{1}_{\{X_s^\beta \in [0, \beta]\}} ds \geq \frac{\beta}{f(\beta)} \mathbb{P} \left(\left\{ \exists t \in [0, T] X_t^\beta \in [0, \beta] \right\} \cap \left\{ X_0^\beta \notin [0, 2\beta] \right\} \right).$$

Since X^β is a permuton process, X_s^β has uniform distribution for each s , which gives

$$\mathbb{E} \int_0^T \mathbb{1}_{\{X_s^\beta \in [0, \beta]\}} ds = \int_0^T \mathbb{E} \mathbb{1}_{\{X_s^\beta \in [0, \beta]\}} ds = \int_0^T \mathbb{P} \left(X_s^\beta \in [0, \beta] \right) ds = T\beta.$$

Together with the inequality above this implies

$$T\beta \geq \frac{\beta}{f(\beta)} \left(\mathbb{P} \left(\exists t \in [0, T] X_t^\beta \in [0, \beta] \right) - \mathbb{P}(X_0^\beta \in [0, 2\beta]) \right).$$

Since X_0^β has uniform distribution, we have $\mathbb{P}(X_0^\beta \in [0, 2\beta]) = 2\beta$. Thus

$$\mathbb{P} \left(\exists t \in [0, T] X_t^\beta \in [0, \beta] \right) \leq 2\beta + Tf(\beta).$$

Since $f(\beta) \rightarrow 0$ as $\beta \rightarrow 0$, the claim is proved. \square

PROPOSITION 3.6. Fix $\beta \in (0, \frac{1}{4})$ and $(x_0, \phi_0) \in [0, 1]^2$. Let $(x^\beta(t), \phi^\beta(t))$, resp. $(x^{\beta,\delta}(t), \phi^{\beta,\delta}(t))$, be the solution to (25), resp. (27), with initial conditions (x_0, ϕ_0) . We have

$$\begin{aligned} \sup_{t \in [0, T]} |x^{\beta,\delta}(t) - x^\beta(t)| &\xrightarrow{\delta \rightarrow 0} 0, \\ \sup_{t \in [0, T]} |\phi^{\beta,\delta}(t) - \phi^\beta(t)| &\xrightarrow{\delta \rightarrow 0} 0. \end{aligned}$$

Proof. The statement follows from continuous dependence of solutions to an ODE on parameters, see e.g., [CL55, Theorem 4.2]. Denoting $V^{\beta,0}(t, y, \phi) = V_t^\beta(y, \phi)$, $R^{\beta,0}(t, y, \phi) = R_t^\beta(y, \phi)$, we only need to check that for $f(t, y, \phi, \delta) = V^{\beta,\delta}(t, y, \phi)$, $g(t, y, \phi, \delta) = R^{\beta,\delta}(t, y, \phi)$ we have

- (1) $f(\cdot, y, \phi, \delta)$ and $g(\cdot, y, \phi, \delta)$ are measurable on $[0, T]$,
- (2) for any fixed $t \in [0, T]$ and $\delta > 0$ $f(t, \cdot, \cdot, \delta)$ and $g(t, \cdot, \cdot, \delta)$ are continuous in (y, ϕ) ,
- (3) for any fixed $t \in [0, T]$ $f(t, \cdot, \cdot, \cdot)$ and $g(t, \cdot, \cdot, \cdot)$ are continuous in (y, ϕ, δ) at $\delta = 0$,
- (4) $f(t, y, \phi, \delta)$, $g(t, y, \phi, \delta)$ are uniformly bounded.

Properties 1), 2) and 4) follow directly from our regularity assumptions about $V^{\beta,\delta}(t, y, \phi)$ (in case of $R^{\beta,\delta}(t, y, \phi)$ we use continuity of $\frac{\partial V^{\beta,\delta}}{\partial y}(t, y, \phi)$). Property 3) follows from pointwise convergence $f(t, y, \phi, \delta) \xrightarrow{\delta \rightarrow 0} f(t, y, \phi, 0)$ and equicontinuity of $\{f(t, y, \phi, \delta)\}_{\delta \geq 0}$ in (y, ϕ) , which in turn follows from uniform continuity of $V_t^\beta(y, \phi)$ in t, y and ϕ . The argument for $g(t, y, \phi, \delta)$ is analogous (again, using uniform continuity of $\frac{\partial V^{\beta,\delta}}{\partial y}(t, y, \phi)$). \square

Now we can prove the main result of this section, which states that the trajectories of the process X and its energy can be approximated by those of the process $X^{\beta,\delta}$.

PROPOSITION 3.7. Let $\pi \in \mathcal{M}(\mathcal{D})$ be the distribution of the process X and let $\pi^{\beta,\delta} \in \mathcal{M}(\mathcal{D})$ be the distribution of the process $X^{\beta,\delta}$. Then we have

$$\lim_{\beta \rightarrow 0} \lim_{\delta \rightarrow 0} d_{\mathcal{W}}^{sup}(\pi, \pi^{\beta,\delta}) = 0,$$

where $d_{\mathcal{W}}^{sup}$ is the Wasserstein distance associated to the supremum norm on \mathcal{D} . Furthermore,

$$\lim_{\beta \rightarrow 0} \lim_{\delta \rightarrow 0} I(\pi^{\beta,\delta}) = I(\pi),$$

where $I(\mu)$ is the energy of the process μ defined in (10).

Proof. For the first convergence it is enough to show that $\mathbb{E} \|X - X^{\beta, \delta}\|_{sup} \rightarrow 0$ in the coupling between X and $X^{\beta, \delta}$ considered before. We have

$$\mathbb{E} \|X - X^{\beta, \delta}\|_{sup} \leq \mathbb{E} \|X - X^{\beta}\|_{sup} + \mathbb{E} \|X^{\beta} - X^{\beta, \delta}\|_{sup}.$$

Let B^{β} be the event from the statement of Lemma 3.5. Since the supremum norm is bounded by 1, we have

$$\mathbb{E} \|X - X^{\beta}\|_{sup} \leq \mathbb{P}(B^{\beta}) + \mathbb{E} \left[\|X - X^{\beta}\|_{sup} \mathbb{1}_{(B^{\beta})^c} \right].$$

By Lemma 3.5 the first term is $o(1)$ as $\beta \rightarrow 0$. Since $V_t^{\beta}(x, \phi) = V_t(x, \phi)$ if $x \in [\beta, 1 - \beta]$, on the event $(B^{\beta})^c$ we have $X^{\beta} = X$, so the second term above is equal to 0. As for $\mathbb{E} \|X^{\beta} - X^{\beta, \delta}\|_{sup}$, by Proposition 3.6 for fixed $\beta > 0$ we have with probability one $\|X^{\beta} - X^{\beta, \delta}\|_{sup} \rightarrow 0$ as $\delta \rightarrow 0$, which together with the estimate on $\mathbb{E} \|X - X^{\beta}\|_{sup}$ proves the first claim of the theorem.

As for the energy, let π^{β} denote the distribution of the process X^{β} , with X , X^{β} and $X^{\beta, \delta}$ coupled as before. Since

$$|I(\pi) - I(\pi^{\beta, \delta})| \leq |I(\pi) - I(\pi^{\beta})| + |I(\pi^{\beta}) - I(\pi^{\beta, \delta})|$$

it is enough to show that $\lim_{\delta \rightarrow 0} I(\pi^{\beta, \delta}) = I(\pi^{\beta})$ and $\lim_{\beta \rightarrow 0} I(\pi^{\beta}) = I(\pi)$. We have

$$I(\pi^{\beta, \delta}) = \mathbb{E} \int_0^T |\dot{X}^{\beta, \delta}(t)|^2 dt = \mathbb{E} \int_0^T V^{\beta, \delta}(t, X^{\beta, \delta}(t), \Phi^{\beta, \delta}(t))^2 dt.$$

For fixed $t \in [0, T]$ by Lemma 3.4 $(X^{\beta, \delta}(t), \Phi^{\beta, \delta}(t))$ has uniform distribution on $[0, 1] \times [0, 1]$ and moving the expectation inside the integral we obtain

$$I(\pi^{\beta, \delta}) = \int_0^T \mathbb{E} \left[V^{\beta, \delta}(t, X^{\beta, \delta}(t), \Phi^{\beta, \delta}(t))^2 \right] dt = \int_0^T \left(\int_0^1 \int_0^1 V^{\beta, \delta}(t, x, \phi)^2 dx d\phi \right) dt.$$

The analogous formula is valid for $I(\pi)$ as well. Now, for fixed $\beta > 0$ we have $V^{\beta, \delta}(t, x, \phi) \xrightarrow{\delta \rightarrow 0} V_t^{\beta}(x, \phi)$ and $V^{\beta, \delta}(t, x, \phi)$ is uniformly bounded in t, x and ϕ , independently of δ , which by dominated convergence implies the convergence of the integrals above as well. Thus $\lim_{\delta \rightarrow 0} I(\pi^{\beta, \delta}) = I(\pi^{\beta})$. The convergence $\lim_{\beta \rightarrow 0} I(\pi^{\beta}) = I(\pi)$ follows directly from properties (d) and (e) of $V_t^{\beta}(x, \phi)$ and dominated convergence. \square

Construction of V_t^β . We will construct the desired modification of $V_t(x, \phi)$ for $x \in [0, \beta]$, the construction for $x \in [1 - \beta, 1]$ is analogous. Fix $t \in [0, T]$. Let

$$L_t(x, \phi) = \frac{\partial V_t}{\partial x}(\beta, \phi)(x - \beta) + V_t(\beta, \phi).$$

Consider $\beta' < \beta/2$ to be fixed later and let f be a smooth approximation of a step function which has values in $[0, 1]$, is equal to 0 on $[0, \beta - 2\beta']$, equal to 1 on $[\beta' - \beta, \beta]$ and is increasing on $[\beta - 2\beta', \beta - \beta']$. In particular we have $f(0) = 0$, $f(\beta) = 1$ and $f'(\beta) = 0$.

Let us now take

$$\tilde{V}_t(x, \phi) = f(x)L_t(x, \phi)$$

and

$$V_t^\beta(x, \phi) = \begin{cases} \tilde{V}_t(x, \phi) & \text{for } x \in [0, \beta], \\ V_t(x, \phi) & \text{otherwise.} \end{cases}$$

We will check that $V_t^\beta(x, \phi)$ indeed satisfies the desired properties.

Let us first check that the property (c) is satisfied for $x \in [0, \beta]$. We have

$$\begin{aligned} \int_0^1 \tilde{V}_t(x, \psi) \, d\psi &= \int_0^1 f(x)L_t(x, \psi) \, d\psi = f(x) \int_0^1 \left(\frac{\partial V_t}{\partial x}(\beta, \psi)(x - \beta) + V_t(\beta, \psi) \right) \, d\psi \\ &= f(x)(x - \beta) \int_0^1 \frac{\partial V_t}{\partial x}(\beta, \psi) \, d\psi + f(x) \int_0^1 V_t(\beta, \psi) \, d\psi \\ &= f(x)(x - \beta) \frac{d}{dx} \Big|_{x=\beta} \left(\int_0^1 V_t(x, \psi) \, d\psi \right) + f(x) \int_0^1 V_t(\beta, \psi) \, d\psi = 0, \end{aligned}$$

thanks to (24).

Property (b) follows directly from $f(0) = 0$. As for (a), for $x \in [0, \beta)$ continuous differentiability of $V_t^\beta(x, \phi)$ follows from continuous differentiability of $f(x)$. At $x = \beta$ we have

$$\tilde{V}_t(\beta, \phi) = f(\beta)L_t(\beta, \phi) = f(\beta)V_t(\beta, \phi)$$

and $f(\beta) = 1$, so $\tilde{V}_t(x, \phi)$ is continuous at $x = \beta$. Likewise,

$$\frac{\partial \tilde{V}_t}{\partial x}(x, \phi) = f'(x)L_t(x, \phi) + f(x) \frac{\partial L_t}{\partial x}(x, \phi) = f'(x)L_t(x, \phi) + f(x) \frac{\partial V_t}{\partial x}(x, \phi).$$

Since $f(\beta) = 1$ and $f'(\beta) = 0$, we have $\frac{\partial \tilde{V}_t}{\partial x}(\beta, \phi) = \frac{\partial V_t}{\partial x}(\beta, \phi)$. As the functions in the formula above are continuously differentiable at $x = \beta$, $V_t^\beta(x, \phi)$ is continuously differentiable at $x = \beta$ as well.

To see that property (d) is satisfied, we note that by continuity of $V_t(x, \phi)$ and $\frac{\partial V_t}{\partial x}(x, \phi)$ for $x \neq 0, 1$ we can take β' in the definition of $f(x)$ above to be arbitrarily small (depending on V_t , $\frac{\partial V_t}{\partial x}$ and β) so that on $[\beta - 2\beta', \beta]$ the function $\tilde{V}_t(x, \phi)$ is less than $|V_t(\beta, \phi)| + 1$ in absolute value. Since on $[0, \beta - 2\beta']$ we have $V_t^\beta(x, \phi) = 0$, the desired bound on $|V_t^\beta(x, \phi)|$ follows.

Finally, to prove that property (e) holds it is enough to show that

$$\int_0^1 \int_0^\beta |V_t^\beta(x, \phi)|^2 dx d\phi \rightarrow 0$$

as $\beta \rightarrow 0$, since $V_t^\beta(x, \phi) = V_t(x, \phi)$ for $x \notin [0, \beta]$. The claim follows immediately from property (d), since the integrand is bounded independently of β .

4 The Biased Interchange Process and Stationarity

The biased interchange process. For the sake of proving a large deviation lower bound, we will need to perturb the interchange process to obtain dynamics which typically exhibits (otherwise rare) behavior of a fixed permutation process. Let us introduce the *biased interchange process*. Its configuration space E consists of sequences $\eta = ((x_i, \phi_i))_{i=1}^N$, where as before (x_1, \dots, x_N) is a permutation of $\{1, \dots, N\}$ and ϕ_i has N possible values, $1, \dots, N$. Here x_i will be the *position* of the particle with label i and ϕ_i will be its *color*.

By a slight abuse of notation we will write $\eta^{-1}(x)$ to denote the label (number) of the particle at position x in configuration η (so that $\eta^{-1}(x_i) = i$). For a position x we will often write ϕ_x as a shorthand for $\phi_{\eta^{-1}(x)}$ (the positions will be always denoted by x or y and labels by i , so there is no risk of ambiguity). In this way we can treat any configuration η as a function which assigns to each site x a pair $(\eta^{-1}(x), \phi_x)$, the label and the color of the particle present at x .

The configuration at time t will be denoted by η_t^N (or simply η_t), and likewise by $x_i(\eta_t^N)$ and $\phi_i(\eta_t^N)$ we denote the position and the color of the particle number i at time t . We will use notation $X_i(\eta_t^N) = \frac{1}{N}x_i(\eta_t^N)$, $\Phi_i(\eta_t^N) = \frac{1}{N}\phi_i(\eta_t^N)$ for the rescaled positions and colors. By the same convention as above $\Phi_x(\eta_t^N)$ will denote the rescaled color of the particle at site x at time t .

Let $\varepsilon = N^{1-\alpha}$, with the same $\alpha \in (1, 2)$ as in (12). Suppose we are given functions $v, r : [0, T] \times \{1, \dots, N\} \times \{1, \dots, N\}$. The dynamics of the corresponding biased interchange process is defined by the (time-inhomogeneous) generator

$$\begin{aligned}
 (\tilde{\mathcal{L}}_t f)(\eta) &= \frac{1}{2} N^\alpha \sum_{x=1}^{N-1} (1 + \varepsilon [v(t, x, \phi_x(\eta)) - v(t, x + 1, \phi_{x+1}(\eta))]) (f(\eta^{x,x+1}) - f(\eta)) \\
 &\quad + \frac{1}{2} N^\alpha \sum_{x=1}^N [(1 + \varepsilon r(t, x, \phi_x(\eta))) (f(\eta^{x,+}) - f(\eta)) + (1 - \varepsilon r(t, x, \phi_x(\eta))) (f(\eta^{x,-}) - f(\eta))].
 \end{aligned}
 \tag{28}$$

Here $\eta^{x,x+1}$ is the configuration η with particles at locations x and $x+1$ swapped, and $\eta^{y,\pm}$ is the configuration η with ϕ_y changed by ± 1 (with the convention that $\eta^{y,+} = \eta^y$ if $\phi_y = N$ and likewise $\eta^{y,-} = \eta^y$ if $\phi_y = 1$). We will often use the abbreviated notation $v_x(t, \eta) = v(t, x, \phi_x(\eta))$ (with the convention $v_0(t, \eta) = v_{N+1}(t, \eta) = 0$).

In other words, at each time neighboring particles make a swap at rate close to 1, with bias proportional to the difference of their *velocities* $v(t, x, \phi_x)$, and each particle independently changes its color by ± 1 , also at rate close to 1 with bias proportional to $\pm r(t, x, \phi_x)$. The parameter ε has been chosen so that we expect particles to have displacement of order N at macroscopic times.

Since the interchange process is a pure jump Markov process, for each particle its rescaled position $X_i(\eta^N)$ and color $\Phi_i(\eta^N)$ will be càdlàg paths from $[0, T]$ to $[0, 1]$ and thus elements of \mathcal{D} . In the same way we can consider the joint trajectory $P_i(\eta^N) = (X_i(\eta^N), \Phi_i(\eta^N))$ as an element of $\tilde{\mathcal{D}} = \mathcal{D}([0, T], [0, 1]^2)$, the space of càdlàg paths from $[0, T]$ to $[0, 1]^2$ (equipped with the Skorokhod topology). By $\mathcal{M}(\tilde{\mathcal{D}})$ we will denote the space of Borel probability measures on $\tilde{\mathcal{D}}$, endowed with the weak topology, and by a slight abuse of notation the corresponding Wasserstein distance will be denoted by $d_{\mathcal{W}}$, as for $\mathcal{M}(\mathcal{D})$.

If η^N is the trajectory of the biased interchange process, then by analogy with the permutation process X^{η^N} we can define the *colored permutation process* $P^{\eta^N} = (X^{\eta^N}, \Phi^{\eta^N})$, obtained by choosing a particle i at random and following the path $(X_i(\eta_t^N), \Phi_i(\eta_t^N))$. Thus we keep track both of the position and the color of a random particle. Since η^N is random, the distribution ν^{η^N} of P^{η^N} , given by

$$\nu^{\eta^N} = \frac{1}{N} \sum_{i=1}^N \delta_{P_i^{\eta^N}},$$

is a random element of $\mathcal{M}(\tilde{\mathcal{D}})$.

Stationarity conditions. Let us now connect the discussion of the interchange process with deterministic permuton processes and generalized solutions to Euler equations considered in Sect 3. Recall the colored trajectory process $P^{\beta,\delta} = (X^{\beta,\delta}, \Phi^{\beta,\delta})$ defined in Sect. 3. From now on we consider $\beta \in (0, \frac{1}{4})$ and $\delta > 0$ to be fixed and we suppress them in the notation, writing $X = X^{\beta,\delta}$, $\Phi = \Phi^{\beta,\delta}$, $V(t, x, \phi) = V^{\beta,\delta}(t, x, \phi)$, $R(t, x, \phi) = R^{\beta,\delta}(t, x, \phi)$. Note that this should not be confused with the actual generalized solution to Euler equations, which was also denoted by X , but does not appear in this and the following sections except in Theorem 7.3.

Our goal is to set up a biased interchange process so that typically trajectories of particles will behave like trajectories of the process X . We would also like to

preserve the stationarity of the uniform distribution of colors, which will greatly facilitate parts of the argument. To find the correct rates $v(t, x, \phi)$ and $r(t, x, \phi)$ in (28), recall that by definition the trajectories of the colored trajectory process $P = (X, \Phi)$ satisfy the equation

$$\begin{cases} \frac{dX}{dt}(t) = V(t, X(t), \Phi(t)), \\ \frac{d\Phi}{dt}(t) = R(t, X(t), \Phi(t)), \end{cases} \quad (29)$$

with the functions V and R satisfying

$$\begin{cases} V(t, X, \Phi) = \frac{\partial F}{\partial \Phi}(t, X, \Phi), \\ R(t, X, \Phi) = -\frac{\partial F}{\partial X}(t, X, \Phi) \end{cases} \quad (30)$$

for $F(t, X, \Phi) = \int_0^\Phi V(t, X, \psi) d\psi$. Note that $F(t, X, 0) = 0$ and $F(t, X, 1) = 0$, where the latter equality follows from property (c) of $V_t^\beta(x, \phi)$ (and thus of $V = V^{\beta, \delta}$).

It is clear that v and r should be chosen so that approximately we have $v(t, x, \phi) \approx V\left(t, \frac{x}{N}, \frac{\phi}{N}\right)$, $r(t, x, \phi) \approx R\left(t, \frac{x}{N}, \frac{\phi}{N}\right)$. To analyze the stationarity condition, consider the uniform distribution on configurations of the biased interchange process, i.e., a distribution in which the labelling of particles is a uniformly random permutation and each particle has a uniformly random color, chosen independently from $\{1, \dots, N\}$ for each of them. We want to find a condition on rates $\tilde{v}(t, x, \phi)$ and $r(t, x, \phi)$ such that this measure will be invariant for the dynamics of $\tilde{\mathcal{L}}_t$.

Note that since $V(t, X, \Phi)$, $R(t, X, \Phi)$ are piecewise-constant as functions of t , the dynamics induced by $\tilde{\mathcal{L}}_t$ is time-homogeneous on each interval $[t_k, t_{k+1})$ from the definition (26) of V . Thus the stationarity condition for the uniform measure is that for each state (i.e., each configuration η) the sums of outgoing and incoming jump rates have to be equal. We write down this condition as follows. For any given configuration η , with particle at location x having color $\phi_x = \phi_x(\eta)$, there are the following possible outgoing jumps:

- for some $x \in \{1, \dots, N-1\}$ the particles at locations x and $x+1$ swap, at rate $1 + \varepsilon [v(t, x, \phi_x) - v(t, x+1, \phi_{x+1})]$;
- for some $x \in \{1, \dots, N\}$ the particle at x changes its color from ϕ_x to $\phi_x \pm 1$, at rate $1 \pm \varepsilon r(t, x, \phi_x)$;

and incoming jumps:

- for some $x \in \{1, \dots, N-1\}$ the particles at locations x and $x+1$ swap, at rate $1 + \varepsilon [v(t, x, \phi_{x+1}) - v(t, x+1, \phi_x)]$;
- for some $x \in \{1, \dots, N\}$ the particle at x changes its color from $\phi_x \pm 1$ to ϕ_x , at rate $1 \mp \varepsilon r(t, x, \phi_x \pm 1)$.

Thus the condition on the sums of jump rates is

$$\begin{aligned} & \sum_{x=1}^{N-1} (v(t, x, \phi_x) - v(t, x + 1, \phi_{x+1})) \\ &= \sum_{x=1}^{N-1} (v(t, x, \phi_{x+1}) - v(t, x + 1, \phi_x)) + \sum_{x=1}^N (r(t, x, \phi_x - 1) - r(t, x, \phi_x + 1)), \end{aligned}$$

where we adopt the convention $r(t, x, 0) = r(t, x, N + 1) = 0$. This implies

$$\begin{aligned} & \sum_{x=2}^{N-1} (v(t, x - 1, \phi_x) - v(t, x + 1, \phi_x) + r(t, x, \phi_x - 1) - r(t, x, \phi_x + 1)) \\ &+ v(t, N - 1, \phi_N) + v(t, N, \phi_N) - v(t, 1, \phi_1) - v(t, 2, \phi_1) \\ &+ [r(t, 1, \phi_1 - 1) - r(t, 1, \phi_1 + 1)] + [r(t, N, \phi_N - 1) - r(t, N, \phi_N + 1)] = 0. \end{aligned}$$

Since we would like this equation to be satisfied for any configuration, regardless of the choice of ϕ_x for each x , we want each term in the sum and each of the boundary terms to vanish. This gives us a set of equations

$$\begin{cases} v(t, 1, \phi) + v(t, 2, \phi) = r(t, 1, \phi - 1) - r(t, 1, \phi + 1), \\ v(t, x + 1, \phi) - v(t, x - 1, \phi) = r(t, x, \phi - 1) - r(t, x, \phi + 1), \quad x = 2, \dots, N - 1, \\ v(t, N - 1, \phi) + v(t, N, \phi) = r(t, N, \phi + 1) - r(t, N, \phi - 1), \end{cases} \tag{31}$$

which have to be satisfied for every $\phi = 1, \dots, N$.

Let us consider the function $f(t, x, \phi)$ defined for $x \in \{0, \dots, N + 1\}$, $\phi \in \{1, \dots, N\}$ by

$$f(t, x, \phi) = \begin{cases} F\left(t, \frac{x}{N}, \frac{\phi}{N+1}\right), & x = 2, \dots, N - 1, \\ 0, & x = 0, 1, N, N + 1, \end{cases}$$

where F is the function appearing in (30). It is straightforward to check that the rates given by

$$\begin{cases} v(t, x, \phi) = \frac{N}{2} (f(t, x, \phi - 1) - f(t, x, \phi + 1)), \\ r(t, x, \phi) = \frac{N}{2} (f(t, x + 1, \phi) - f(x - 1, \phi)), \end{cases} \tag{32}$$

solve the equations for stationarity, given by (31), for any $x, \phi \in \{1, \dots, N\}$.

Note that with this choice of rates we have for any $x, \phi \in \{1, \dots, N\}$

$$\begin{cases} v(t, x, \phi) = V\left(t, \frac{x}{N}, \frac{\phi}{N}\right) + O\left(\frac{1}{N}\right), \\ r(t, x, \phi) = R\left(t, \frac{x}{N}, \frac{\phi}{N}\right) + O\left(\frac{1}{N}\right), \end{cases} \tag{33}$$

uniformly in x , ϕ and t , because of smoothness of $F(t, X, \Phi)$ in X and Φ variables. In particular the rates v and r are uniformly bounded for all N .

From now on we will always assume that the biased interchange process has rates $v(t, x, \phi)$ and $r(t, x, \phi)$ given by (32) and is started from the uniform distribution (which by the discussion above is stationary). The properties of v and r which will be relevant to our analysis is that they are bounded, approximately equal to some smooth functions V , R , that the corresponding dynamics has the uniform measure as the stationary distribution and, crucially, that in stationarity the velocities are independent and mean zero. This last property, which should be thought of as the particle system analog of Lemma 3.2, is conveniently summarized in the following proposition.

PROPOSITION 4.1. *Let ϕ_x , $x = 1, \dots, N$, be independent and uniformly distributed on $\{1, \dots, N\}$. Then for each $t \in [0, T]$ the random variables $v(t, x, \phi_x)$, $x = 1, \dots, N$, are independent and for each x we have*

$$\mathbb{E} v(t, x, \phi_x) = 0.$$

Proof. Under the uniform distribution of ϕ_x we have

$$\mathbb{E} v(t, x, \phi_x) = \frac{1}{N} \sum_{\phi=1}^N v(t, x, \phi),$$

which by definition of v is equal to

$$\frac{1}{N} \sum_{\phi=1}^N \frac{N}{2} (f(t, x, \phi - 1) - f(t, x, \phi + 1)) = \frac{1}{2} \left(F \left(t, \frac{x}{N}, 0 \right) - F \left(t, \frac{x}{N}, 1 \right) \right).$$

Recalling the definition of F below (30), the right-hand side is equal to 0. \square

5 Law of Large Numbers

Throughout this section $\tilde{\mathbb{P}}^N$ will denote the probability law of the biased interchange process on N particles, started in stationarity, associated to the equation (29) (with all the assumptions from the previous section). To simplify notation we will usually write $\eta = \eta^N$. Whenever we use $o(\cdot)$ or $O(\cdot)$ asymptotic notation the implicit constants will depend only on the rates v , r and possibly on T .

Let $P = (X, \Phi)$ be the colored trajectory process associated to the Eq. (29) and let P^{η^N} be the colored permutation process defined in Sect. 4. Let us denote the distributions of P and P^{η^N} respectively by ν and ν^{η^N} , with $\nu, \nu^{\eta^N} \in \mathcal{M}(\tilde{\mathcal{D}})$. We will prove the following theorem

Theorem 5.1. *Let η^N be the trajectory of the biased interchange process. The measures ν^{η^N} converge in distribution, as random elements of $\mathcal{M}(\tilde{\mathcal{D}})$, to the deterministic measure ν as $N \rightarrow \infty$.*

In other words, the random processes P^{η^N} converge in distribution to the process P whose distribution is deterministic. The theorem above can be thought of as a law of large numbers for random permutation processes and it will be useful for establishing the large deviation lower bound.

REMARK 5.2. Since the limiting measure ν is deterministic and supported on continuous trajectories, Theorem 5.1 implies that the convergence $\nu^{\eta^N} \rightarrow \nu$ in fact holds in a stronger sense, namely in probability when $\mathcal{M}(\tilde{\mathcal{D}})$ is endowed with the Wasserstein distance $d_{\mathcal{W}}^{sup}$ associated to the supremum norm on $\tilde{\mathcal{D}}$.

To prove Theorem 5.1, we will show that typically trajectories of most particles approximately follow the same ODE (29) as trajectories of the limiting process. In other words, if a given particle is at site x , it should locally move according to its velocity $v(t, x, \phi_x)$. However, because of swaps between particles the actual jump rates of the particle will be influenced by velocities of its neighbors. Nevertheless, since velocity at each site has mean 0 in stationarity, we will be able to show that the contribution from velocities of the particle's neighbors cancels out when averaged over time – this will be the content of the *one block estimate* proved in the next section.

Note that to prove that the random processes converge indeed to a deterministic process, it is not enough to look only at single path distributions, as explained in Sect. 2.1. Nevertheless, we will show that in the interchange process typically any two particles (in fact almost all of them) behave like independent random walks, which by Lemma 2.1 will be enough to establish a deterministic limit.

Throughout this and the following sections we will make extensive use of martingales associated to Markov processes (see [KL99] for a comprehensive treatment of such techniques applied to interacting particle systems). For any Markov process with generator \mathcal{L} and a bounded function $F : E \rightarrow \mathbb{R}$, where E is the configuration space of the process, the following processes are mean zero martingales ([KL99, Lemma A1.5.1])

$$M_t = F(\eta_t) - F(\eta_0) - \int_0^t \mathcal{L}F(\eta_s) ds, \quad (34)$$

$$N_t = M_t^2 - \int_0^t (\mathcal{L}F(\eta_s)^2 - 2F(\eta_s)\mathcal{L}F(\eta_s)) ds. \quad (35)$$

Furthermore, for any F as above the following process is a mean one positive martingale (see discussion following [KL99, Lemma A1.7.1])

$$\mathbb{M}_t = \exp \left\{ F(\eta_t) - F(\eta_0) - \int_0^t e^{-F(\eta_s)} \mathcal{L}e^{F(\eta_s)} ds \right\}. \quad (36)$$

In the following sections we will also consider the case when F is not necessarily bounded, in which case M_t , N_t , \mathbb{M}_t are only local martingales.

Our first goal is to prove that with high probability almost all particles move according to their local velocity $v(t, x_i, \phi_i)$. Recall that

$$X_i(\eta_t) = \frac{1}{N}x_i(\eta_t), \quad \Phi_i(\eta_t) = \frac{1}{N}\phi_i(\eta_t)$$

are respectively the rescaled position and color of the particle with label i . Our first goal is to prove the following

PROPOSITION 5.3. *For any fixed $t \in [0, T]$ and $\varepsilon > 0$ we have in the biased interchange process*

$$\begin{aligned} \tilde{\mathbb{P}}^N \left(\frac{1}{N} \sum_{i=1}^N \left| X_i(\eta_t) - X_i(\eta_0) - \int_0^t v(s, x_i(\eta_s), \phi_i(\eta_s)) ds \right| > \varepsilon \right) &\rightarrow 0, \\ \tilde{\mathbb{P}}^N \left(\frac{1}{N} \sum_{i=1}^N \left| \Phi_i(\eta_t) - \Phi_i(\eta_0) - \int_0^t r(s, x_i(\eta_s), \phi_i(\eta_s)) ds \right| > \varepsilon \right) &\rightarrow 0, \end{aligned}$$

as $N \rightarrow \infty$.

As a starting point let us rewrite $X_i(\eta_t)$ in a more useful form. Recall from (28) that $\tilde{\mathcal{L}}$ denotes the generator of the biased interchange process. By the formula (34) applied to $F(\eta_s) = X_i(\eta_s)$ we have

$$X_i(\eta_t) - X_i(\eta_0) = M_t^i + \int_0^t \tilde{\mathcal{L}}X_i(\eta_s) ds,$$

where M_t^i is a mean zero martingale with respect to $\tilde{\mathbb{P}}^N$. Recall that $v_x(t, \eta) = v(t, x, \phi_x(\eta))$ denotes the velocity of the particle at site x in configuration η at time t . For simplicity we will also write $v_{x_i}(t, \eta) = v(t, x_i(\eta), \phi_i(\eta))$ for the velocity of the particle with label i . We have

$$\begin{aligned} \tilde{\mathcal{L}}X_i(\eta_s) &= \frac{1}{N}\tilde{\mathcal{L}}(x_i(\eta_s)) \\ &= \frac{1}{2}N^{\alpha-1} \sum_{x=1}^{N-1} (1 + \varepsilon [v_x(s, \eta_s) - v_{x+1}(s, \eta_s)]) (x_i(\eta_s^{x, x+1}) - x_i(\eta_s)) \\ &= \frac{1}{2}N^{\alpha-1}\varepsilon \left[- [v_{x_i-1}(s, \eta_s) - v_{x_i}(s, \eta_s)] + [v_{x_i}(s, \eta_s) - v_{x_i+1}(s, \eta_s)] \right] \\ &= \frac{1}{2}(2v_{x_i}(s, \eta_s) - v_{x_i-1}(s, \eta_s) - v_{x_i+1}(s, \eta_s)), \end{aligned}$$

since the position of the particle i changes by ± 1 depending on whether it makes a swap with its left or right neighbor.

Thus we obtain

$$X_i(\eta_t) - X_i(\eta_0) = M_t^i + \int_0^t v_{x_i}(s, \eta_s) ds + \frac{1}{2} \int_0^t (v_{x_{i-1}}(s, \eta_s) + v_{x_{i+1}}(s, \eta_s)) ds,$$

or in other words

$$X_i(\eta_t) - X_i(\eta_0) - \int_0^t v_{x_i}(s, \eta_s) ds = M_t^i + \frac{1}{2} \int_0^t (v_{x_{i-1}}(s, \eta_s) + v_{x_{i+1}}(s, \eta_s)) ds. \tag{37}$$

For the sake of proving the first part of Proposition 5.3 it will be enough to show that

$$\frac{1}{N} \sum_{i=1}^N \mathbb{E} \left(X_i(\eta_t) - X_i(\eta_0) - \int_0^t v_{x_i}(\eta_s) ds \right)^2 \rightarrow 0 \tag{38}$$

as $N \rightarrow \infty$. First we prove that for most particles the martingale term M_t^i will be small with high probability. Let us define

$$Q_s^i = \tilde{\mathcal{L}}X_i(\eta_s)^2 - 2X_i(\eta_s)\tilde{\mathcal{L}}X_i(\eta_s).$$

By the martingale formula (35) we have that

$$N_t^i = (M_t^i)^2 - \int_0^t Q_s^i ds \tag{39}$$

is a mean zero martingale. A quick calculation gives

$$\begin{aligned} \tilde{\mathcal{L}}X_i(\eta_s)^2 &= \frac{1}{2} \left[(v_{x_{i-1}}(s, \eta_s) - v_{x_i}(s, \eta_s)) \left(\frac{-2x_i(\eta_s) + 1}{N} \right) + (v_{x_i}(s, \eta_s) \right. \\ &\quad \left. - v_{x_{i+1}}(s, \eta_s)) \left(\frac{2x_i(\eta_s) + 1}{N} \right) \right] + N^{\alpha-2} \end{aligned}$$

and

$$2X_i(\eta_s)\tilde{\mathcal{L}}X_i(\eta_s) = \frac{x_i(\eta_s)}{N} (2v_{x_i}(s, \eta_s) - v_{x_{i-1}}(s, \eta_s) - v_{x_{i+1}}(s, \eta_s)),$$

so these two quantities are the same up to terms of order $o(1)$. Thus $Q_s^i = o(1)$ (uniformly in s and i) and, since $\mathbb{E}N_t^i = 0$, we obtain from (39) that $\mathbb{E}(M_t^i)^2 = o(1)$ as well.

Incidentally, a similar calculation (only simpler, since it does not involve correlations between adjacent particles) and the martingale argument gives us that for $\Phi_i(\eta_t) = \frac{1}{N}\phi_i(\eta_t)$ we have

$$\Phi_i(\eta_t) - \Phi_i(\eta_0) - \int_0^t r(s, x_i(\eta_s), \phi_i(\eta_s)) ds = o(1)$$

for any fixed particle i . This proves the second part of Proposition 5.3.

Recalling (37) and (38), to finish the proof of the first part of Proposition 5.3 we only need to show that

$$\frac{1}{N} \sum_{i=1}^N \mathbb{E} (Y_i^t)^2 \rightarrow 0$$

as $N \rightarrow \infty$, where

$$Y_i^t = \int_0^t (v_{x_{i-1}}(s, \eta_s) + v_{x_{i+1}}(s, \eta_s)) ds.$$

Recall from (26) that $V^{\beta, \delta}(s, x, \phi)$ was defined in terms of a partition $0 = t_0 < t_1 < \dots < t_M = T$. We would like to take advantage of the fact that on each interval the dynamics of the biased interchange process is time-homogeneous. Suppose that $t \in [t_l, t_{l+1})$ for some $l \leq M - 1$ and let us write

$$Y_i^t = \sum_{k=0}^{l-1} \int_{t_k}^{t_{k+1}} (v_{x_{i-1}}(s, \eta_s) + v_{x_{i+1}}(s, \eta_s)) ds + \int_{t_l}^t (v_{x_{i-1}}(s, \eta_s) + v_{x_{i+1}}(s, \eta_s)) ds.$$

For any $t \geq 0$ let

$$Y_i^{t,k} = \int_{t_k}^{t_k+t} (v_{x_{i-1}}(s, \eta_s) + v_{x_{i+1}}(s, \eta_s)) ds.$$

Since M is fixed, it is enough to show that for any fixed $k \leq M - 1$ and $t \in [0, t_{k+1} - t_k]$ we have

$$\frac{1}{N} \sum_{i=1}^N \mathbb{E} (Y_i^{t,k})^2 \rightarrow 0$$

as $N \rightarrow \infty$.

To keep the notation simple we will prove the desired statement just for $k = 0$, with the general case being exactly analogous. Recall that $t_0 = 0$. By definition of the piecewise-constant in time approximation of $V^{\beta, \delta}$, for $s \in [0, t_1]$ we have $v_x(s, \eta_s) = v_x(0, \eta_s)$. Let us define $v_x(\eta) = v_x(0, \eta)$. Fix any $t \in [0, t_1]$ and let us look at

$$(Y_i^{t,0})^2 = \left(\int_0^t (v_{x_{i-1}}(\eta_s) + v_{x_{i+1}}(\eta_s)) ds \right)^2.$$

We will have four cross-terms here, it is enough to show that each of them is small in expectation. The argument will be similar in all cases, so we will only present the proof for one of them. Let us focus on

$$\mathbb{E} \left[\left(\int_0^t v_{x_{i-1}}(\eta_s) ds \right) \left(\int_0^t v_{x_{i-1}}(\eta_s) ds \right) \right] = \mathbb{E} \int_0^t \int_0^t v_{x_{i-1}}(\eta_{u_1}) v_{x_{i-1}}(\eta_{u_2}) du_1 du_2.$$

For each particle i we are looking at the correlation of the velocity of its left neighbor at time u_1 with the velocity of its left neighbor at time u_2 . By averaging over particles $i = 1, \dots, N$ and using the symmetry between u_1 and u_2 we can write the contribution to the second moment of $Y_i^{t,0}$ as

$$\begin{aligned} & \frac{2}{N} \sum_{i=1}^N \mathbb{E} \int_0^t \int_{u_1}^t v_{x_{i-1}}(\eta_{u_1}) v_{x_{i-1}}(\eta_{u_2}) du_2 du_1 \\ &= 2 \int_0^t du_1 \left(\frac{1}{N} \sum_{i=1}^N \mathbb{E} \int_{u_1}^t v_{x_{i-1}}(\eta_{u_1}) v_{x_{i-1}}(\eta_{u_2}) du_2 \right). \end{aligned}$$

Since the rates v are bounded, it is enough to show that for each fixed $u_1 \in [0, t]$ the expression inside the bracket is close to 0 as $N \rightarrow \infty$. Let us look at

$$\frac{1}{N} \sum_{i=1}^N \mathbb{E} \int_{u_1}^t v_{x_{i-1}}(\eta_{u_1}) v_{x_{i-1}}(\eta_{u_2}) du_2.$$

Since the average here depends only on the configuration at time u_1 and its evolution from that point on (and not otherwise on the trajectory of the process before time u_1), by stationarity of the biased interchange process it will be the same as

$$\frac{1}{N} \sum_{i=1}^N \mathbb{E} \int_0^{t-u_1} v_{x_{i-1}}(\eta_0) v_{x_{i-1}}(\eta_s) ds, \tag{40}$$

since the dynamics of the process is time-homogeneous on $[0, t_1)$.

Thus we have to prove that for a random particle the velocity of its initial left neighbor is uncorrelated (when averaged over time) with the velocity of its current left neighbor. Let us introduce the following setup – we can rewrite the average above in terms of a sum over sites (for $y = x_i(\eta_s)$) instead over particles

$$\frac{1}{N} \sum_{y=1}^N \mathbb{E} \int_0^{t-u_1} v_{x_{\eta_s^{-1}(y)}-1}(\eta_0) v_{y-1}(\eta_s) ds \tag{41}$$

To analyze this average we introduce the following extension of the biased interchange process. Consider the extended configuration space \tilde{E} consisting of sequences

$((x_i, \phi_i, L_i))_{i=1}^N$, with $L_i \in \{1, \dots, N\}$. Here each particle, in addition to its color ϕ_i , also has an additional color L_i in which we keep information about the velocity of its left neighbor at time 0, that is

$$L_i = v_{x_i(\eta_0)-1}(\eta_0).$$

The dynamics is given by the same generator (28) as before, i.e., labels (together with their corresponding colors ϕ_i and L_i) are exchanged by swaps of adjacent particles, each ϕ_i has its own evolution and L_i does not evolve. For a site x let $L_x(\eta)$ be the additional color at site x in configuration η , i.e., $L_x(\eta) = L_{\eta^{-1}(x)}$. We can now treat η as a function which assigns to each site x a triple $(\eta^{-1}(x), \phi_x, L_x)$ or simply a pair (ϕ_x, L_x) , since we are not interested in particles' labels at this point, only in the distribution of colors.

In this setup the average (41) can be written as

$$\frac{1}{N} \sum_{y=1}^N \mathbb{E} \int_0^{t-u_1} f_y(\eta_s) ds, \quad (42)$$

where $f_y(\eta) = L_y(\eta)v_{y-1}(\eta)$. Let

$$\Lambda_{x,l} = \{x-l, x-l+1, \dots, x+l\},$$

denote a box of size l around x (with the convention that the box is truncated if the endpoints $x-l$ or $x+l$ exceed 1 or N , but this will not influence the argument in any substantial way) and let $\widehat{\mu}_{x,l}^\eta$ be the empirical distribution of colors in $\Lambda_{x,l}$ in configuration η , given for any (L, ϕ) by

$$\widehat{\mu}_{x,l}^\eta(L, \phi) = \frac{1}{|\Lambda_{x,l}|} \#\{z \in \Lambda_{x,l} \mid (L_z(\eta), \phi_z(\eta)) = (L, \phi)\}.$$

Consider the associated i.i.d. distribution on configurations restricted to $\Lambda_{x,l}$, given by

$$\mu_{x,l}^\eta \left((L_y, \phi_y)_{y=x-l}^{x+l} \right) = \prod_{y=x-l}^{x+l} \widehat{\mu}_{x,l}^\eta(L_y, \phi_y).$$

In other words, under the measure $\mu_{x,l}^\eta$ the probability of seeing a color pair (L, ϕ) at site $y \in \Lambda_{x,l}$ is proportional to the number of sites in $\Lambda_{x,l}$ with the color pair (L, ϕ) , independently for each site.

The *superexponential one block estimate* says that on an event of high probability we can replace $f_y(\eta_s)$ in the time average (42) by its average $\mathbb{E}_{\mu_{y,l}^{\eta_s}}(f)$ with respect to the local i.i.d. distribution over a sufficiently large box. In other words, due to local mixing the distribution of colors in a microscopic box can be approximated by an i.i.d. distribution for large l .

LEMMA 5.4. *Let $U_{x,l}(\eta) = |f_x(\eta) - \mathbb{E}_{\mu_{x,l}^\eta}(f)|$. For any $t \in [0, t_1]$ and $\delta > 0$ we have*

$$\limsup_{l \rightarrow \infty} \limsup_{N \rightarrow \infty} N^{-\gamma} \log \tilde{\mathbb{P}}^N \left(\int_0^t \frac{1}{N} \sum_{x=1}^N U_{x,l}(\eta_s) ds > \delta \right) = -\infty,$$

where $\gamma = 3 - \alpha$.

The lemma is proved in the next section. Let us see how it enables us to finish the proof of Proposition 5.3. By the one block estimate, in (42) we can replace

$$\frac{1}{N} \sum_{y=1}^N \int_0^{t-u_1} f_y(\eta_s) ds$$

by

$$\frac{1}{N} \sum_{y=1}^N \int_0^{t-u_1} \mathbb{E}_{\mu_{y,l}^{\eta_s}} f_y(\eta_s) ds, \tag{43}$$

with the difference going to 0 in expectation as first $N \rightarrow \infty$ and then $l \rightarrow \infty$, so we only need to show that the latter expression goes to 0 in the same limit.

Observe that in $f_y(\eta) = L_y(\eta)v_{y-1}(\eta) = L_y(\eta)v(y-1, \phi_{y-1}(\eta))$ the colors ϕ_{y-1} and L_y depend on different sites, so they are independent under $\mu_{y,l}^{\eta_s}$, since the measure is product. Thus in the average above we can simply write

$$\mathbb{E}_{\mu_{y,l}^{\eta_s}} f_y(\eta_s) = \mathbb{E}_{\mu_{y,l}^{\eta_s}} [L_y(\eta)v_{y-1}(\eta)] = \left(\mathbb{E}_{\sigma \sim \mu_{y,l}^{\eta_s}} L_y(\sigma) \right) \left(\mathbb{E}_{\sigma \sim \mu_{y,l}^{\eta_s}} v_{y-1}(\sigma) \right),$$

where by a slight abuse of notation we have denoted by σ the local configuration of colors in a box $\Lambda_{y,l}$ and considered L_y, v_{y-1} as functions of σ . The average (43) now becomes

$$\frac{1}{N} \sum_{y=1}^N \int_0^{t-u_1} \left(\mathbb{E}_{\sigma \sim \mu_{y,l}^{\eta_s}} L_y(\sigma) \right) \left(\mathbb{E}_{\sigma \sim \mu_{y,l}^{\eta_s}} v_{y-1}(\sigma) \right) ds.$$

Since the distribution of η_s in the biased interchange process without the additional colors L_i is stationary, the distribution of the average $\mathbb{E}_{\sigma \sim \mu_{y,l}^{\eta_s}} v_{y-1}(\sigma)$ does not depend on s . So we only need to show that $\mathbb{E}_{\sigma \sim \mu_{y,l}^{\eta_0}} v_{y-1}(\sigma)$ is small, since L_y is bounded.

Recall that in stationarity ϕ_y has uniform distribution, so for any y the expectation of $v_{y-1}(\sigma) = v(0, y-1, \phi_{y-1}(\sigma))$ with respect to $\mu_{y,l}^{\eta_0}$ is simply equal to

$$\frac{1}{2l+1} \sum_{j=1}^{2l+1} v(0, y-1, \phi_j),$$

where ϕ_j are independent and uniformly distributed on $\{1, \dots, N\}$. As for each x the random variables $v(0, x, \phi_j)$ are independent, bounded and have mean 0 (see Proposition 4.1), an easy application of Hoeffding's inequality gives that for fixed y the sum above goes to 0 in probability as $l \rightarrow \infty$. This finishes the proof of Proposition 5.3.

We can now prove the law of large numbers.

Proof of Theorem 5.1. Consider the random particle process $\bar{P}^N = (\bar{X}^N, \bar{\Phi}^N)$, obtained by first sampling $\eta = \eta^N$ and then following the trajectory $P_i(\eta_t) = (X_i(\eta_t), \Phi_i(\eta_t))$ of a randomly chosen particle i . We will first show that the (deterministic) distribution $\bar{\nu}^N$ converges to ν , the distribution of P (in the metric $d_{\mathcal{W}}^{sup}$).

Let us start by proving that the estimate from Proposition 5.3 holds not only at each time t , but also with the supremum over all times $t \leq T$ under the sum over particles. Consider the process (A^N, B^N) defined as

$$A_t^N = X_i(\eta_t) - X_i(\eta_0) - \int_0^t v(s, x_i(\eta_s), \phi_i(\eta_s)) ds,$$

$$B_t^N = \Phi_i(\eta_t) - \Phi_i(\eta_0) - \int_0^t r(s, x_i(\eta_s), \phi_i(\eta_s)) ds,$$

where i is a random particle and $\eta = \eta^N$ comes from the biased interchange process. Proposition 5.3 implies that all finite-dimensional marginals of (A^N, B^N) converge to 0. To obtain convergence to 0 for the whole process in the supremum norm we only need to check tightness in the Skorokhod topology (which will imply convergence in the supremum norm, since the limiting process is continuous). We will use the following stopping time criterion ([KL99, Proposition 4.1.6]). Let Y^N be a family of stochastic processes with sample paths in $\tilde{\mathcal{D}}$ such that for each time $t \in [0, T]$ the marginal distribution of Y_t^N is tight. If for every $\varepsilon > 0$ we have

$$\lim_{\gamma \rightarrow 0} \limsup_{N \rightarrow \infty} \sup_{\substack{\tau \\ \theta \leq \gamma}} \mathbb{P}(\|Y_{\tau+\theta}^N - Y_{\tau}^N\| > \varepsilon) = 0, \quad (44)$$

where the supremum is over all stopping times τ bounded by T , then the family Y^N is tight. Here $\|\cdot\|$ denotes the Euclidean distance on $[0, 1]^2$ and for simplicity we write $\tau + \theta$ instead of $(\tau + \theta) \wedge T$. Let τ be any stopping time bounded by T . We have from formula 37

$$A_{\tau+\theta}^N - A_{\tau}^N = M_{\tau+\theta}^i - M_{\tau}^i - \frac{1}{2} \int_{\tau}^{\tau+\theta} [v(s, x_i(\eta_s) - 1, \phi_{x_i(\eta_s)-1}(\eta_s)) + v(s, x_i(\eta_s) + 1, \phi_{x_i(\eta_s)+1}(\eta_s))] ds.$$

Since $v(\cdot, \cdot, \cdot)$ is bounded, the integral is bounded by $C\theta$ for some constant $C > 0$, regardless of τ , so goes to 0 as $\theta \rightarrow 0$ (deterministically and for every i). Thus it only

remains to bound the martingale term. As τ is a stopping time, by formula (39) we have for each i

$$\mathbb{E} \left[(M_{\tau+\theta}^i)^2 - (M_\tau^i)^2 \right] = \mathbb{E} \int_\tau^{\tau+\theta} Q_s ds.$$

As in the calculation of $\mathbb{E}(M_t^i)^2$ following (39) we have that for fixed θ the right hand side is $o(1)$ as $N \rightarrow \infty$. Since M_t^i is bounded, we obtain $\mathbb{E} |M_{\tau+\theta}^i - M_\tau^i| \rightarrow 0$ as $N \rightarrow \infty$, for any θ and i (independently of τ). The calculation for B^N is analogous.

This shows that the family (A^N, B^N) satisfies the tightness criterion (44). In particular it converges to 0 in the supremum norm as $N \rightarrow \infty$. Thus for any $\varepsilon > 0$ we have

$$\tilde{\mathbb{P}}^N \left(\frac{1}{N} \sum_{i=1}^N \sup_{0 \leq t \leq T} \left| X_i(\eta_t) - X_i(\eta_0) - \int_0^t v(s, x_i(\eta_s), \phi_i(\eta_s)) ds \right| > \varepsilon \right) \rightarrow 0, \tag{45}$$

$$\tilde{\mathbb{P}}^N \left(\frac{1}{N} \sum_{i=1}^N \sup_{0 \leq t \leq T} \left| \Phi_i(\eta_t) - \Phi_i(\eta_0) - \int_0^t r(s, x_i(\eta_s), \phi_i(\eta_s)) ds \right| > \varepsilon \right) \rightarrow 0, \tag{46}$$

as $N \rightarrow \infty$.

Now we can prove that $\bar{\nu}^N$ converges to ν . Recalling the definition of the Wasserstein distance $d_{\mathcal{W}}^{sup}$, it is enough to construct for each N a coupling (\bar{P}^N, P) such that

$$\mathbb{E} \| \bar{P}^N - P \|_{sup} \rightarrow 0$$

as $N \rightarrow \infty$.

Let us couple these two processes in the following way: first we let $\bar{P}^N = ((\bar{X}_t^N, \bar{\Phi}_t^N), 0 \leq t \leq T)$ be a path sampled according to $\bar{\nu}^N$, starting at $(\bar{X}_0^N, \bar{\Phi}_0^N)$ (whose distribution is uniform on $\{\frac{1}{N}, \dots, 1\} \times \{\frac{1}{N}, \dots, 1\}$). We then take $P(t) = (X(t), \Phi(t))$ to be the solution of the ODE (29) started from an initial condition $(X(0), \Phi(0))$ chosen uniformly at random from $[\bar{X}_0^N - \frac{1}{N}, \bar{X}_0^N] \times [\bar{\Phi}_0^N - \frac{1}{N}, \bar{\Phi}_0^N]$ (so the two processes start close to each other). Because the initial condition is distributed uniformly on $[0, 1]^2$, the path $P = (P(t), 0 \leq t \leq T)$ will be distributed according to ν .

Since $P(t) = (X(t), \Phi(t))$ is the solution of (29), we have at each time $t \leq T$

$$X(t) - X(0) = \int_0^t V(s, X(s), \Phi(s)) ds,$$

$$\Phi(t) - \Phi(0) = \int_0^t R(s, X(s), \Phi(s)) ds.$$

Bounds (45), (46) imply that for all times $t \leq T$ we have

$$\begin{aligned}\bar{X}^N(t) - \bar{X}^N(0) &= \int_0^t v(s, N\bar{X}^N(s), N\bar{\Phi}^N(s)) ds + \varepsilon_t^1, \\ \bar{\Phi}^N(t) - \bar{\Phi}^N(0) &= \int_0^t r(s, N\bar{X}^N(s), N\bar{\Phi}^N(s)) ds + \varepsilon_t^2,\end{aligned}$$

with $\varepsilon_t^1, \varepsilon_t^2$ satisfying $\sup_{0 \leq t \leq T} |\varepsilon_t^i| \rightarrow 0$ in probability as $N \rightarrow \infty$. Recalling from (33) that $v(\cdot, \cdot, \cdot), r(\cdot, \cdot, \cdot)$ are approximately equal to $V(\cdot, \cdot, \cdot), R(\cdot, \cdot, \cdot)$ after rescaling of the arguments, we obtain

$$\begin{aligned}\bar{X}^N(t) - \bar{X}^N(0) &= \int_0^t V(s, \bar{X}^N(s), \bar{\Phi}^N(s)) ds + o(1), \\ \bar{\Phi}^N(t) - \bar{\Phi}^N(0) &= \int_0^t R(s, \bar{X}^N(s), \bar{\Phi}^N(s)) ds + o(1),\end{aligned}$$

with the $o(1)$ terms going to 0 in probability (in the supremum norm over t) as $N \rightarrow \infty$.

Thus $(\bar{X}^N, \bar{\Phi}^N)$ approximately satisfies the same ODE as (X, Φ) and an application of Grönwall's inequality gives that for any $\varepsilon > 0$ with probability approaching 1 as $N \rightarrow \infty$ we have

$$\|\bar{P}^N - P\|_{sup} \leq C \max\{|\bar{X}^N(0) - X(0)| + \varepsilon, |\bar{\Phi}^N(0) - \Phi(0)| + \varepsilon\} e^{KT}$$

for some $C > 0$, where $K > 0$ depends only on the Lipschitz constants of V and R .

By definition of the processes \bar{P}^N and P the initial conditions $\bar{X}^N(0), X(0)$ and $\bar{\Phi}^N(0), \Phi(0)$ differ by at most $\frac{1}{N}$, which implies that $\mathbb{E} \|\bar{P}^N - P\|_{sup} \rightarrow 0$ as $N \rightarrow \infty$.

Thus the distribution $\bar{\nu}^N$ of the random particle process \bar{P}^N converges to ν in the $d_{\mathcal{W}}^{sup}$ metric as desired.

Now we can show that the random measures ν^N converge in distribution to the deterministic measure ν . By the characterization of tightness for random measures (see, e.g., [Kal21, Theorem 23.15]) the family ν^N will be tight, as a family of $\mathcal{M}(\tilde{\mathcal{D}})$ -valued random variables, if for any $\varepsilon > 0$ there exists a compact set $K \subseteq \tilde{\mathcal{D}}$ such that $\limsup_{N \rightarrow \infty} \mathbb{E} \left(\nu^N(K) \right) \geq 1 - \varepsilon$, or, more simply put, $\limsup_{N \rightarrow \infty} \tilde{\mathbb{P}}^N \left(P^N \in K \right) \geq 1 - \varepsilon$. Exactly the same calculation as for the processes (A^N, B^N) before shows the processes P^N satisfy the tightness criterion (44), which guarantess the existence of desired compact sets K and in turn tightness of ν^N .

Now to finish the proof we only need to show uniqueness of subsequential limits for the family ν^N . Since any such (possibly random) limit must have the associated

random particle process distributed according to ν , it is enough to show that the limit is deterministic.

Consider an outcome of ν^{η^N} , which is a measure from $\mathcal{M}(\tilde{\mathcal{D}})$, and sample independently two paths P_1^N, P_2^N from it. This corresponds to sampling η^N according to the biased interchange process, then choosing uniformly at random a pair of particles i, j (possibly with $i = j$, but this event has vanishing probability) and following their trajectories in η^N . By the already established convergence $\bar{\nu}^N \rightarrow \nu$ in $\mathcal{M}(\tilde{\mathcal{D}})$, each path P_1^N and P_2^N separately has distribution converging to ν . Moreover, due to stationarity of η^N the initial colors $\phi_i(\eta_0^N), \phi_j(\eta_0^N)$ of any two particles i, j are chosen uniformly at random, in particular they are independent for $i \neq j$. Thus the joint distribution of (P_1^N, P_2^N) converges to the distribution of two independent paths sampled from ν , as a path P sampled from ν is uniquely determined by its initial conditions. Since we already have tightness, applying Lemma 2.1 gives that any limit of a subsequence has to be deterministic, which finishes the proof. \square

6 One Block Estimate

In this section we prove the one block estimate of Lemma 5.4, needed for the proof of Theorem 5.1. Since another, simpler variant of this estimate will also be needed for the proof of the large deviation upper bound (Lemma 8.2), we prove the result in generality suited for both of these applications.

Let us fix a continuous function $w : [0, 1]^2 \rightarrow \mathbb{R}$ and let $I_w^N = \left\{ w \left(\frac{i}{N}, \frac{j}{N} \right) \right\}_{i,j=1}^N$. Let $I^N = \left\{ \frac{1}{N}, \dots, 1 \right\}$. Consider the interchange process on an extended configuration space E' in which each particle in addition to its label i has two colors (a_i, ϕ_i) , with $a_i \in I_w^N, \phi_i \in I^N$. The dynamics is given by the usual generator \mathcal{L} – adjacent particles are making swaps at rate $\frac{1}{2}N^\alpha$ and the colors a_i, ϕ_i of the particle i do not evolve in time. Since the one block estimate concerns only the distribution of colors, from now on we ignore the labels of the particles altogether. Similarly as before we use the notation $a_x = a_x(\eta), \phi_x = \phi_x(\eta)$ to denote the colors of the particle at site x in configuration η . The configuration at time s is denoted by η_s .

Consider a continuous function $g : [0, 1] \rightarrow [-1, 1]$ and for $\eta \in E'$ let $h_x(\eta) = a_x(\eta)b_{x-1}(\eta)$, where $b_x(\eta) = g(\phi_x(\eta))$ or $b_x(\eta) = a_x(\eta)$. As in the previous section let $\Lambda_{x,l} = \{x - l, x - l + 1, \dots, x + l\}$ denote the box of size l around x (with an appropriate truncation if the endpoints $x - l$ or $x + l$ exceed 1 or N , which we neglect in the notation from now on) and let $\hat{\mu}_{x,l}^\eta$ be the empirical distribution of colors in $\Lambda_{x,l}$ in configuration η , given for any $(\alpha, \varphi) \in I_w^N \times I^N$ by

$$\hat{\mu}_{x,l}^\eta(\alpha, \varphi) = \frac{1}{|\Lambda_{x,l}|} \#\{z \in \Lambda_{x,l} \mid (a_z(\eta), \phi_z(\eta)) = (\alpha, \varphi)\}.$$

Consider the associated i.i.d. distribution on configurations restricted to $\Lambda_{x,l}$, given for $(\alpha_y, \varphi_y)_{y=x-l}^{x+l} \in (I_w^N \times I^N)^{2l+1}$ by

$$\mu_{x,l}^\eta \left((\alpha_y, \varphi_y)_{y=x-l}^{x+l} \right) = \prod_{y=x-l}^{x+l} \widehat{\mu}_{x,l}^\eta (\alpha_y, \varphi_y).$$

Since h_x depends on η only through the colors at x and $x-1$, we will slightly abuse notation by writing $\mathbb{E}_{\mu_{x,l}^\eta}(h_x)$ for the expectation of h_x with respect to $\mu_{x,l}^\eta$.

Let $\psi : [0, 1] \rightarrow \mathbb{R}$ be a continuous function and let

$$U_{x,l}^N(\eta) = \psi(x) \left(h_x(\eta) - \mathbb{E}_{\mu_{x,l}^\eta}(h_x) \right).$$

We define

$$U_l^N(\eta) = \frac{1}{N} \sum_{x=1}^N |U_{x,l}^N(\eta)|.$$

Let μ denote the uniform distribution on E' . Note that the dynamics given by \mathcal{L} is reversible with respect to μ and the associated Dirichlet form is given by

$$D^N(f) = \frac{1}{4} N^\alpha \int \sum_{x=1}^{N-1} \left(\sqrt{f(\eta^{x,x+1})} - \sqrt{f(\eta)} \right)^2 d\mu(\eta)$$

for any $f : E' \rightarrow [0, \infty)$.

LEMMA 6.1. *With μ denoting the uniform distribution on E' , we have for any $C_0 > 0$*

$$\limsup_{l \rightarrow \infty} \limsup_{N \rightarrow \infty} \sup_{\substack{f \\ D^N(f) \leq C_0 N^\gamma}} \int U_l^N(\eta) f(\eta) d\mu(\eta) = 0,$$

where $\gamma = 3 - \alpha$ and the supremum is over all densities f with respect to μ such that $D^N(f) \leq C_0 N^\gamma$.

Proof. Let us decompose $a_x = a_x(\eta)$ and $b_x = b_x(\eta)$ into their positive and negative parts, $a_x = a_x^+ - a_x^-$, $b_x = b_x^+ - b_x^-$. Since

$$h_x = a_x b_{x-1} = a_x^+ b_{x-1}^+ - a_x^+ b_{x-1}^- - a_x^- b_{x-1}^+ + a_x^- b_{x-1}^-,$$

by the triangle inequality it is enough to prove the lemma with h_x replaced by one of the terms in the sum above, say, $a_x^+ b_{x-1}^+$. Let $K = \max\{1, \|w\|_\infty\}$ and let us write

$$a_x^+(\eta) = \int_0^K \mathbb{1}_{\{a_x(\eta) > \lambda\}} d\lambda,$$

$$b_x^+(\eta) = \int_0^K \mathbb{1}_{\{b_x(\eta) > \theta\}} d\theta.$$

We have

$$\begin{aligned} & \frac{1}{N} \sum_{x=1}^N \left| \psi(x) \left(a_x^+(\eta) b_{x-1}^+(\eta) - \mathbb{E}_{\mu_{x,l}^\eta} \left[a_x^+(\eta) b_{x-1}^+(\eta) \right] \right) \right| \\ & \leq \int_0^K \int_0^K \frac{1}{N} \sum_{x=1}^N \left| \psi(x) \left(\mathbb{1}_{\{a_x(\eta) > \lambda\}} \mathbb{1}_{\{b_{x-1}(\eta) > \theta\}} - \mathbb{E}_{\mu_{x,l}^\eta} \left[\mathbb{1}_{\{a_x(\eta) > \lambda\}} \mathbb{1}_{\{b_{x-1}(\eta) > \theta\}} \right] \right) \right| d\lambda d\theta, \end{aligned}$$

where the inequality comes from pulling the integrals over λ and θ outside the absolute value. Let us denote the expression under the integrals on the right hand side by $U_{l,\lambda,\theta}^N$. Since it is nonnegative and bounded, we can write

$$\sup_f \int \left(\int_0^K \int_0^K U_{l,\lambda,\theta}^N(\eta) d\lambda d\theta \right) f(\eta) d\mu(\eta) \leq \int_0^K \int_0^K \left(\sup_f \int U_{l,\lambda,\theta}^N(\eta) f(\eta) d\mu(\eta) \right) d\lambda d\theta,$$

where the supremum is over all densities f satisfying $D^N(f) \leq C_0 N^\gamma$. By the same token, when taking the limsup first over N and then over l , we can bound the resulting limit from above by one with the integral over λ and θ outside the limsup. Thus we see that it is enough to prove for fixed $\lambda, \theta \in [0, K]$

$$\limsup_{l \rightarrow \infty} \limsup_{N \rightarrow \infty} \sup_{\substack{f \\ D^N(f) \leq C_0 N^\gamma}} \int U_{l,\lambda,\theta}^N(\eta) f(\eta) d\mu(\eta) = 0.$$

Since

$$U_{l,\lambda,\theta}^N(\eta) = \frac{1}{N} \sum_{x=1}^N \left| \psi(x) \left(\mathbb{1}_{\{a_x(\eta) > \lambda\}} \mathbb{1}_{\{b_{x-1}(\eta) > \theta\}} - \mathbb{E}_{\mu_{x,l}^\eta} \left[\mathbb{1}_{\{a_x(\eta) > \lambda\}} \mathbb{1}_{\{b_{x-1}(\eta) > \theta\}} \right] \right) \right|,$$

we have reduced the problem to proving the one block estimate for the interchange process in which each particle has only four possible colors, corresponding to the possible values of the pair $(\mathbb{1}_{\{a_i(\eta) > \lambda\}}, \mathbb{1}_{\{b_i(\eta) > \theta\}})$. This in turn follows by essentially the same argument as for the simple exclusion process, which can be thought of as interchange process with just two colors (see e.g., [KL99, Lemma 5.3.1]). Since the argument is by now standard and used in several places in the literature (see e.g., [FT04] for the case of three possible colors), let us only explain that the bound on the Dirichlet form under the supremum is of the right order. The argument for the simple exclusion process goes through (see the remark following the proof of [KL99, Lemma 5.4.2]) if we assume that the Dirichlet form corresponding to the generator without time scaling is $o(N)$ and the process is speeded up by N^2 . In our case the generator \mathcal{L} has a scaling factor of N^α , so if $N^{-\alpha} D^N(f)$ is the Dirichlet form corresponding to the process without time scaling, then our bound on this Dirichlet form is $\leq C_0 N^{\gamma-\alpha} = C_0 N^{3-2\alpha}$. Since $\alpha \in (1, 2)$, this is $o(N)$, which agrees with the assumptions for the simple exclusion process. \square

LEMMA 6.2. *Let \mathbb{P}^N denote the law of the interchange process on E' with an arbitrary initial distribution. With the notation as above we have for any $t \geq 0$ and $\delta > 0$*

$$\limsup_{l \rightarrow \infty} \limsup_{N \rightarrow \infty} N^{-\gamma} \log \mathbb{P}^N \left(\int_0^t U_l^N(\eta_s) ds > \delta \right) = -\infty.$$

Proof. Let μ_0 be an arbitrary initial distribution. Let $\mathbb{P}_{0,\mu}^N$, resp. \mathbb{P}_{0,μ_0}^N , denote the distribution of the process started from μ , resp μ_0 , and let \mathbb{E}_μ , resp. \mathbb{E}_{μ_0} , denote the corresponding expectation.

By Chebyshev's inequality we have for any $c > 0$

$$\mathbb{P}_{0,\mu_0}^N \left(\int_0^t U_l^N(\eta_s) ds > \delta \right) \leq e^{-cN^\gamma} \mathbb{E}_{\mu_0} \exp \left\{ cN^\gamma \int_0^t U_l^N(\eta_s) ds \right\}. \quad (47)$$

We also have

$$\begin{aligned} \mathbb{E}_{\mu_0} \exp \left\{ cN^\gamma \int_0^t U_l^N(\eta_s) ds \right\} &= \mathbb{E}_\mu \left[\frac{d\mathbb{P}_{0,\mu_0}^N}{d\mathbb{P}_{0,\mu}^N}(t) \exp \left\{ cN^\gamma \int_0^t U_l^N(\eta_s) ds \right\} \right] \\ &\leq \left\| \frac{d\mathbb{P}_{0,\mu_0}^N}{d\mathbb{P}_{0,\mu}^N} \right\|_\infty \mathbb{E}_\mu \exp \left\{ cN^\gamma \int_0^t U_l^N(\eta_s) ds \right\}. \end{aligned}$$

Let $M = |I_w^N|$. Since $M \leq N^2$ and under μ each initial configuration has probability $(MN)^N = e^{\rho(N^\gamma)}$, the supremum norm of the Radon–Nikodym derivative above is $e^{\rho(N^\gamma)}$ as well, so to prove (47) it is in fact enough to show that for any $c > 0$

$$\limsup_{l \rightarrow \infty} \limsup_{N \rightarrow \infty} N^{-\gamma} \log \exp \left\{ cN^\gamma \int_0^t U_l^N(\eta_s) ds \right\} \leq 0 \quad (48)$$

and then take $c \rightarrow \infty$.

An application of Feynman–Kac formula to the semigroup generated by \mathcal{L} shows (see e.g., [KL99, Theorem 10.3.1 and Section A1.7]) that to obtain (48) it is sufficient to prove for any $c > 0$

$$\limsup_{l \rightarrow \infty} \limsup_{N \rightarrow \infty} \sup_f \left\{ \int cU_l^N(\eta) f(\eta) d\mu(\eta) - N^{-\gamma} D^N(f) \right\} \leq 0,$$

where the supremum is taken over all densities with respect to μ . Since U_l^N is bounded by a constant $C > 0$ depending only on ψ and g , the expression under the

supremum becomes negative if $D^N(f) > cCN^\gamma$. Thus it is enough to show that for any constant $C_0 > 0$ we have

$$\limsup_{l \rightarrow \infty} \limsup_{N \rightarrow \infty} \sup_f^{D^N(f) \leq C_0 N^\gamma} \int U_l^N(\eta) f(\eta) d\mu(\eta) \leq 0,$$

which exactly the statement of Lemma 6.1. □

This estimate will be enough for application in the proof of Lemma 8.2. As for the proof of Lemma 5.4, we will first show that the one block estimate holds for the unbiased process with color evolution, but with all rates equal to 1, i.e., the process with state space E' and the generator

$$\begin{aligned} (\mathcal{L}_0 f)(\eta) &= \frac{1}{2} N^\alpha \sum_{x=1}^{N-1} (f(\eta^{x,x+1}) - f(\eta)) \\ &\quad + \frac{1}{2} N^\alpha \sum_{x=1}^N [(f(\eta^{x,+}) - f(\eta)) + (f(\eta^{x,-}) - f(\eta))]. \end{aligned}$$

Here as usual $\eta^{x,\pm}$ denotes the configuration obtained from η by changing the color ϕ_x of the particle at site x to $\phi_x \pm 1$ (note that the colors a_i do not evolve in time here). We will then transfer the result to the biased process by estimating its Radon–Nikodym derivative.

LEMMA 6.3. *Let \mathbb{P}_0^N be the law of the unbiased process with rates 1 described above (with an arbitrary initial distribution). With the notation from Lemma 6.2, we have for any $t \geq 0$ and $\delta > 0$*

$$\limsup_{l \rightarrow \infty} \limsup_{N \rightarrow \infty} N^{-\gamma} \log \mathbb{P}_0^N \left(\int_0^t U_l^N(\eta_s) ds > \delta \right) = -\infty.$$

Proof. Let us write $\mathcal{L}_0 = \mathcal{L} + \mathcal{L}_c$, where \mathcal{L} is the first term in the definition of \mathcal{L}_0 and \mathcal{L}_c is the second term. The dynamics induced by \mathcal{L} and by \mathcal{L}_c is reversible with respect to μ , so the Dirichlet forms associated respectively to \mathcal{L}_c and \mathcal{L}_0 can be written as

$$\begin{aligned} D_c^N(f) &= \frac{1}{4} N^\alpha \int \sum_{x=1}^N \left[\left(\sqrt{f(\eta^{x,+})} - \sqrt{f(\eta)} \right)^2 + \left(\sqrt{f(\eta^{x,-})} - \sqrt{f(\eta)} \right)^2 \right] d\mu(\eta), \\ D_0^N(f) &= D^N(f) + D_c^N(f). \end{aligned}$$

By repeating the argument from the proof of Lemma 6.2 with the generator \mathcal{L}_0 instead of \mathcal{L} we obtain that it is enough to prove that for any $c > 0$

$$\limsup_{l \rightarrow \infty} \limsup_{N \rightarrow \infty} \sup_f \left\{ \int cU_l^N(\eta) f(\eta) d\mu(\eta) - N^{-\gamma} D_0^N(f) \right\} \leq 0,$$

where the supremum is taken over all densities with respect to μ .

Now observe that since $D_c^N(f) \geq 0$ for any nonnegative f , it is in fact enough to prove the statement above with $D_0^N(f)$ replaced by $D^N(f)$. Thus we have eliminated color evolution and the conclusion follows as in the proof of Lemma 6.2. \square

We can now prove the superexponential estimate for the biased process.

Proof of Lemma 5.4. Recall that $f_x(\eta) = L_x(\eta)v_{x-1}(\eta)$. Since we can uniformly approximate $v(0, x, \phi)$ by finite sums of terms which are product in x and ϕ , by using the triangle inequality we can without loss of generality assume that $v_x(\eta) = \psi(x)g(\phi_x)$ for some continuous functions $\psi : [0, 1] \rightarrow \mathbb{R}$, $g : [0, 1] \rightarrow [-1, 1]$. Applying Lemma 6.3 with $w(x, \phi) = v(0, x, \phi)$, $a_i = L_i$ and $h_x = L_x g(\phi_{x-1})$ provides us with the superexponential estimate for the process \mathbb{P}_0^N . To transfer the estimate to the biased process $\tilde{\mathbb{P}}^N$ we will need to estimate the Radon–Nikodym derivative of the two processes.

If \mathbb{P} is a Markov process with jump rates $\lambda(x)p(x, y)$ and $\tilde{\mathbb{P}}$ is another process on the same state space with rates $\tilde{\lambda}(x)\tilde{p}(x, y)$, the Radon–Nikodym derivative up to time t is given by (see, e.g., [KL99, Proposition A1.2.6])

$$\frac{d\tilde{\mathbb{P}}}{d\mathbb{P}}(t) = \exp \left\{ - \int_0^t \left(\tilde{\lambda}(X_s) - \lambda(X_s) \right) ds + \sum_{s \leq t} \log \frac{\tilde{\lambda}(X_{s-})\tilde{p}(X_{s-}, X_s)}{\lambda(X_{s-})p(X_{s-}, X_s)} \right\}, \quad (49)$$

where the sum is over jump times $s \leq t$.

Let us look at $\frac{d\tilde{\mathbb{P}}^N}{d\mathbb{P}_0^N}$. By the form (28) of the generator of $\tilde{\mathbb{P}}^N$ the sum of outgoing rates for any η is equal to

$$\frac{1}{2}N^\alpha \left(\sum_{x=1}^{N-1} [1 + \varepsilon(v_x(\eta) - v_{x+1}(\eta))] + \sum_{x=1}^N [1 + \varepsilon r_x(\eta)] + \sum_{x=1}^N [1 - \varepsilon r_x(\eta)] \right).$$

Since the sum of $\varepsilon(v_x - v_{x+1})$ telescopes, the rates v_x are 0 at the boundaries $x = 1, N$ and rates r_x for the color change cancel out, the intensities $\tilde{\lambda}$ and λ cancel out as well. The Radon–Nikodym derivative takes the form

$$\begin{aligned} \frac{d\tilde{\mathbb{P}}^N}{d\mathbb{P}_0^N}(t) = \exp \left\{ \sum_{s \leq t} \log \left(1 + \varepsilon \left[v(x_{j_s}, \phi_{x_{j_s}}(\eta_s)) - v(x_{j_s} + 1, \phi_{x_{j_s} + 1}(\eta_s)) \right] \right) \right. \\ \left. + \sum_{s_+ \leq t} \log \left(1 + \varepsilon \left[r(x_{j_{s_+}}, \phi_{x_{j_{s_+}}}(\eta_{s_+})) \right] \right) \right. \\ \left. + \sum_{s_- \leq t} \log \left(1 - \varepsilon \left[r(x_{j_{s_-}}, \phi_{x_{j_{s_-}}}(\eta_{s_-})) \right] \right) \right\}, \quad (50) \end{aligned}$$

where j_s is the label of the particle which makes a swap at time s and j_{s_\pm} is the label of the particle that changes its color by ± 1 at time s_\pm .

To simplify this formula we will use the fact that empirical currents across edges can be approximated by their averages, modulo a small martingale. More precisely, let us denote for simplicity

$$(\nabla_x v)(\eta) = v_x(\eta) - v_{x+1}(\eta).$$

We will sometimes use this notation with $x = N$, in which case we assume $(\nabla_x v)(\eta) = 0$. For brevity of notation whenever sums involving both r_x and $-r_x$ appear, we will write them as one term with a \pm sign, that is, with $\sum_x (1 \pm \varepsilon r_x)$ serving as a shorthand for $\sum_x (1 + \varepsilon r_x) + \sum_x (1 - \varepsilon r_x)$ and so on.

We introduce the following extension of the dynamics under \mathbb{P}_0^N – for any functions $h(x, \eta), h^\pm(x, \eta), x \in \{1, \dots, N\}$ consider the extended state space E' , consisting of pairs $(\eta, J), J \in \mathbb{R}$, and the generator \mathcal{L}' acting by

$$\begin{aligned} (\mathcal{L}' f)(\eta, J) = & \frac{1}{2} N^\alpha \left[\sum_{x=1}^{N-1} (f(\eta^{x,x+1}, J + h(x, \eta)) - f(\eta, J)) \right. \\ & + \sum_{x=1}^N (f(\eta^{x,+}, J + h^+(x, \eta)) - f(\eta, J)) \\ & \left. + \sum_{x=1}^N (f(\eta^{x,-}, J + h^-(x, \eta)) - f(\eta, J)) \right]. \end{aligned}$$

In other words, in the evolution of the extended configuration (η_t, J_t) each time the process makes a jump, J_t is increased by $h(x, \eta_t), h^+(x, \eta_t)$ or $h^-(x, \eta_t)$, depending on the type of the jump (swap or color change). Now if we take

$$\begin{aligned} h(x, \eta) &= \log [1 + \varepsilon(\nabla_x v)(\eta)], \\ h^\pm(x, \eta) &= \log [1 \pm \varepsilon r_x(\eta)], \end{aligned}$$

we see that J_t is simply equal to the sum over jumps appearing in the exponent in (50). Thus to bound the Radon–Nikodym derivative we only need to bound J_t .

This is done by use of an exponential martingale – for any $\lambda > 0$ the following process

$$Z_t = \exp \left\{ \lambda J_t - \int_0^t e^{-\lambda J_s} \mathcal{L}' e^{\lambda J_s} ds \right\}$$

is a local martingale with respect to \mathbb{P}_0^N . We will actually only need to consider $\lambda = 2$. Writing out the action of \mathcal{L}' on the function $g(\eta, J) = e^{2J}$ we obtain

$$Z_t = \exp \left\{ 2J_t - \frac{1}{2} N^\alpha \int_0^t \sum_{x=1}^N \left[\left(e^{2 \log(1 + \varepsilon(\nabla_x v)(\eta_s))} - 1 \right) + \left(e^{2 \log(1 \pm \varepsilon r_x(\eta_s))} - 1 \right) \right] ds \right\}.$$

Now we have

$$\begin{aligned} e^{2\log(1+\varepsilon(\nabla_x v)(\eta_s))} - 1 &= (1 + \varepsilon(\nabla_x v)(\eta_s))^2 - 1 = 2\varepsilon(\nabla_x v)(\eta_s) + \varepsilon^2 [(\nabla_x v)(\eta_s)]^2, \\ e^{2\log(1\pm\varepsilon r_x(\eta_s))} - 1 &= (1 \pm \varepsilon r_x(\eta_s))^2 - 1 = \pm 2\varepsilon r_x(\eta_s) + \varepsilon^2 r_x(\eta_s)^2. \end{aligned}$$

The sum of terms linear in ε vanishes – the rates r for ± 1 color change have opposite sign and the sum involving $\nabla_x v$ telescopes. Recalling that $\varepsilon = N^{1-\alpha}$ and $\gamma = 3 - \alpha$, so $N^{\alpha+1}\varepsilon^2 = N^\gamma$, we can then write

$$Z_t = \exp \left\{ 2J_t - \frac{1}{2}N^\gamma \int_0^t \frac{1}{N} \sum_{x=1}^N [(\nabla_x v)(\eta_s)^2 + r_x(\eta_s)^2] ds \right\}.$$

Since the rates v and r are bounded, we have $Z_t = e^{2J_t - N^\gamma X_t}$, where $|X_t| \leq C$ for some constant $C > 0$ depending only on v , r and T . In particular we get

$$\mathbb{E}e^{2J_t} = \mathbb{E} \left(e^{2J_t - N^\gamma X_t} e^{N^\gamma X_t} \right) = \mathbb{E} \left(Z_t e^{N^\gamma X_t} \right) \leq e^{CN^\gamma} \mathbb{E}Z_t.$$

Since Z_t is a local martingale bounded from below, it is a supermartingale, so we have $\mathbb{E}Z_t \leq \mathbb{E}Z_0 = 1$ and thus

$$\mathbb{E}e^{2J_t} \leq e^{CN^\gamma}. \quad (51)$$

Now we can transfer the superexponential bound of Lemma 6.3 from \mathbb{P}_0^N to $\tilde{\mathbb{P}}^N$. Let $\mathcal{O}_{N,l}$ be the event from the statement of the lemma and let us write simply $\frac{d\tilde{\mathbb{P}}^N}{d\mathbb{P}_0^N} = \frac{d\tilde{\mathbb{P}}^N}{d\mathbb{P}_0^N}(T)$. Denoting by $\tilde{\mathbb{E}}$ the expectation with respect to $\tilde{\mathbb{P}}^N$ we have

$$\tilde{\mathbb{P}}^N(\mathcal{O}_{N,l}) = \tilde{\mathbb{E}}(\mathbb{1}_{\mathcal{O}_{N,l}}) = \mathbb{E} \left(\frac{d\tilde{\mathbb{P}}^N}{d\mathbb{P}_0^N} \mathbb{1}_{\mathcal{O}_{N,l}} \right).$$

Applying the Cauchy-Schwarz inequality gives

$$\tilde{\mathbb{P}}^N(\mathcal{O}_{N,l}) \leq \left[\mathbb{E} \left(\frac{d\tilde{\mathbb{P}}^N}{d\mathbb{P}_0^N} \right)^2 \right]^{1/2} \cdot \mathbb{P}_0^N(\mathcal{O}_{N,l})^{1/2}.$$

Recalling that $\frac{d\tilde{\mathbb{P}}^N}{d\mathbb{P}_0^N} = e^{J_T}$ and applying the bound (51) we obtain

$$\tilde{\mathbb{P}}^N(\mathcal{O}_{N,l}) \leq e^{cN^\gamma} \mathbb{P}_0^N(\mathcal{O}_{N,l})^{1/2}$$

with $c = \frac{C}{2}$. Thus

$$\limsup_{N \rightarrow \infty} N^{-\gamma} \log \tilde{\mathbb{P}}^N(\mathcal{O}_{N,l}) \leq c + \frac{1}{2} \limsup_{N \rightarrow \infty} N^{-\gamma} \log \mathbb{P}_0^N(\mathcal{O}_{N,l})$$

and taking lim sup as $l \rightarrow \infty$ together with an application of Lemma 6.3 finishes the proof. \square

7 Large Deviation Lower Bound

In this section we prove the large deviation lower bound of Theorem A. Let us assume that the permuton process X satisfies Eq. 29. Since we already know how to construct a biased interchange process that will typically display the behavior of X , to bound the probability that the trajectory of a random particle in the interchange process is close in distribution to X we only need to compare the unbiased process with the biased one by means of calculating their Radon–Nikodym derivative.

Since these two processes have different configuration spaces, for convenience we introduce the *unbiased interchange process with colors*, which has the same configuration space as the biased process associated to 29 and the generator \mathcal{L}^u obtained by putting all velocities v to 0

$$\begin{aligned}
 (\mathcal{L}_t^u f)(\eta) &= \frac{1}{2} N^\alpha \sum_{y=1}^{N-1} (f(\eta^{y,y+1}) - f(\eta)) \\
 &\quad + \frac{1}{2} N^\alpha \sum_{x=1}^N [1 \pm \varepsilon r(t, x, \phi_x(\eta))] (f(\eta^{x,\pm}) - f(\eta)). \tag{52}
 \end{aligned}$$

Since here the colors do not influence the dynamics of swaps, the corresponding permutation process X^{η^N} will be the same as for the ordinary unbiased interchange process (and we will never be interested in the distribution of Φ^{η^N} for the unbiased process with colors).

Let us start by deriving the formula for the Radon–Nikodym derivative of the unbiased process with colors with respect to the biased one. Recall that $v_x(s, \eta_s) = v(s, x, \phi_x(\eta_s))$ denotes the velocity at time s of the particle at site x . Let \mathbb{P}_u^N denote the law of the unbiased process with colors. We will prove the following statement

LEMMA 7.1. *We have*

$$\frac{d\mathbb{P}_u^N}{d\tilde{\mathbb{P}}^N}(T) = \exp \left\{ -\frac{1}{2} N^\gamma \left[\int_0^T \frac{1}{N} \sum_{x=1}^N v_x(s, \eta_s)^2 ds + o(1) \right] \right\},$$

where the $o(1)$ term goes to 0 in probability as $N \rightarrow \infty$.

Proof. The calculation is similar as in the proof of Lemma 5.4, with the difference that we are using generator $\tilde{\mathcal{L}}$ instead of \mathcal{L}_0 . By the analog of formula (49) for time-inhomogeneous processes we have

$$\frac{d\mathbb{P}_u^N}{d\tilde{\mathbb{P}}^N}(t) = \exp \left\{ -\sum_{s \leq t} \log \left(1 + \varepsilon [v(s, x_{j_s}, \phi_{x_{j_s}}(\eta_s)) - v(s, x_{j_s} + 1, \phi_{x_{j_s} + 1}(\eta_s))] \right) \right\},$$

where the sum is over jump times $s \leq t$.

Denoting the sum in the exponent by J_t , we obtain by (34) (by considering as before the generator $\tilde{\mathcal{L}}$ acting on an extended configuration space) that

$$J_t = M_t + \frac{1}{2}N^\alpha \sum_{x=1}^{N-1} \int_0^t [1 + \varepsilon(\nabla_x v)(s, \eta_s)] \log [1 + \varepsilon(\nabla_x v)(s, \eta_s)] ds,$$

where M_t is a local martingale with respect to $\tilde{\mathbb{P}}^N$. Expanding all terms up to order ε^2 allows us to write

$$\frac{d\mathbb{P}_u^N}{d\tilde{\mathbb{P}}^N}(t) = \exp \left\{ -M_t - \frac{1}{2}N^\alpha \int_0^t \sum_{x=1}^{N-1} \left[\varepsilon(\nabla_x v)(s, \eta_s) + \frac{\varepsilon^2}{2} [(\nabla_x v)(s, \eta_s)]^2 \right] ds + O(N^{\alpha+1}\varepsilon^3) \right\}.$$

As before the term linear in ε vanishes. Recalling that $\varepsilon = N^{1-\alpha}$ and $\gamma = 3 - \alpha$ we have

$$\frac{d\mathbb{P}_u^N}{d\tilde{\mathbb{P}}^N}(t) = \exp \left\{ -M_t - \frac{1}{4}N^\gamma \int_0^t \frac{1}{N} \sum_{x=1}^{N-1} [(\nabla_x v)(s, \eta_s)]^2 ds + o(N^\gamma) \right\}.$$

Expanding $[(\nabla_x v)(s, \eta_s)]^2 = (v_x(s, \eta_s) - v_{x+1}(s, \eta_s))^2$ leads us to

$$\begin{aligned} \frac{d\mathbb{P}_u^N}{d\tilde{\mathbb{P}}^N}(t) = \exp \left\{ -M_t - \frac{1}{2}N^\gamma \int_0^t \frac{1}{N} \sum_{x=1}^N v_x(s, \eta_s)^2 ds \right. \\ \left. + \frac{1}{2}N^\gamma \int_0^t \frac{1}{N} \sum_{x=1}^N v_x(s, \eta_s)v_{x+1}(s, \eta_s) ds + o(N^\gamma) \right\}. \end{aligned} \quad (53)$$

The martingale term will be typically $o(N^\gamma)$. To see this, we use formula (35) – by performing a calculation similar to the one above we get that

$$N_t = M_t^2 - \frac{1}{2}N^\alpha \int_0^t \sum_{x=1}^N [1 + \varepsilon(\nabla_x v)(s, \eta_s)] [\log(1 + \varepsilon(\nabla_x v)(s, \eta_s))]^2 ds$$

is a local martingale with respect to $\tilde{\mathbb{P}}^N$. By expanding the log terms up to ε^2 we see that the second term above is bounded by $CN^{\alpha+1}\varepsilon^2 = CN^\gamma$ for some $C > 0$. In particular N_t is bounded from below, so it is a supermartingale. Thus $\mathbb{E}N_T \leq \mathbb{E}N_0 = 0$ and $\mathbb{E}M_T^2 \leq CN^\gamma$, so Chebyshev's inequality implies that $M_T = o(N^\gamma)$ with high probability.

The second sum in the exponent in (53) will be small by invariance of the uniform distribution of colors in the biased process. More precisely, at fixed time s for each x the correlation term $v_x(s, \eta_s)v_{x+1}(s, \eta_s)$ has mean 0, since η_s has stationary

distribution and by Proposition 4.1 in stationarity velocities at different sites are independent with mean 0. Moreover, for the same reason these terms are uncorrelated for different x , so by the weak law of large numbers we get that for any $s \leq T$ and $\delta > 0$

$$\tilde{\mathbb{P}}^N \left(\left| \frac{1}{N} \sum_{x=1}^N v_x(s, \eta_s) v_{x+1}(s, \eta_s) \right| > \delta \right) \rightarrow 0$$

as $N \rightarrow \infty$.

Since this holds for any fixed s and the random variables are bounded, we also have

$$\tilde{\mathbb{P}}^N \left(\left| \int_0^T \frac{1}{N} \sum_{x=1}^N v_x(s, \eta_s) v_{x+1}(s, \eta_s) ds \right| > \delta \right) \rightarrow 0$$

as $N \rightarrow \infty$, which proves that the correlation term is $o(N^\gamma)$ with high probability. Together with the bound on M_t this proves the desired formula for the Radon–Nikodym derivative. \square

We can now use Lemma 7.1 and the law of large numbers established in Theorem 5.1 to prove a large deviation lower bound for the interchange process. As the formula from the lemma suggests, the large deviation rate function will be related to the energy of the process to which the biased interchange process converges.

Recall from (10) that for any process $\pi \in \mathcal{P}$ its energy was defined by

$$I(\pi) = \mathbb{E}_{\gamma \sim \pi} \mathcal{E}(\gamma),$$

where $\mathcal{E}(\gamma)$ is the Dirichlet energy of the path γ defined by (7). We have the following large deviation lower bound

Theorem 7.2. *Let \mathbb{P}^N be the law of the unbiased interchange process η^N and let μ^{η^N} be the (random) distribution of the corresponding permutation process X^{η^N} . Let $P = (X, \Phi)$ be the colored trajectory process associated to the equation (29) and let μ denote the distribution of X . For any open set $\mathcal{O} \subseteq \mathcal{P}$ such that $\mu \in \mathcal{O}$ we have*

$$\liminf_{N \rightarrow \infty} N^{-\gamma} \log \mathbb{P}^N \left(\mu^{\eta^N} \in \mathcal{O} \right) \geq -I(\mu).$$

Proof. It will be enough to show the bound above for \mathcal{O} being any open ball $B(\mu, \varepsilon)$ in \mathcal{P} around μ . Let \mathbb{P}_u^N be the distribution of the unbiased process with colors, ν^{η^N} the distribution of the colored permutation process $P^{\eta^N} = (X^{\eta^N}, \Phi^{\eta^N})$ associated to η^N . Let ν denote the distribution of $P = (X, \Phi)$ and $\tilde{B}(\nu, \varepsilon)$ an open ball around ν in $\mathcal{M}(\tilde{\mathcal{D}})$. Since the projection $(X, \Phi) \mapsto X$ is continuous as a map from $\tilde{\mathcal{D}}$ to \mathcal{D} , the corresponding projection from $\mathcal{M}(\tilde{\mathcal{D}})$ to $\mathcal{M}(\mathcal{D})$ is also continuous. As μ^{η^N} has the same law under \mathbb{P}^N and \mathbb{P}_u^N (remember that in the latter process the colors do not influence the dynamics of swaps), we have that for any $\varepsilon > 0$ there exists $\varepsilon' > 0$ such

that $\mathbb{P}^N \left(\mu^{\eta^N} \in B(\mu, \varepsilon) \right) \geq \mathbb{P}_u^N \left(\nu^{\eta^N} \in \tilde{B}(\nu, \varepsilon') \right)$. Thus to prove the large deviation bound it is sufficient to prove the local lower bound

$$\liminf_{\varepsilon \rightarrow 0} \liminf_{N \rightarrow \infty} N^{-\gamma} \log \mathbb{P}_u^N \left(\nu^{\eta^N} \in \tilde{B}(\nu, \varepsilon) \right) \geq -I(\mu). \quad (54)$$

Recall that $\tilde{\mathbb{P}}^N$ denotes the distribution of the biased process associated to (29) and consider the Radon–Nikodym derivative $\frac{d\mathbb{P}_u^N}{d\tilde{\mathbb{P}}^N}(t)$. By Lemma 7.1 we have

$$\frac{d\mathbb{P}_u^N}{d\tilde{\mathbb{P}}^N}(T) = \exp \left\{ -\frac{1}{2} N^\gamma \left[\int_0^T \frac{1}{N} \sum_{x=1}^N v_x(\eta_s)^2 ds + Y_N \right] \right\}, \quad (55)$$

where Y_N goes to 0 in probability as $N \rightarrow \infty$.

Now by the law of large numbers from Theorem 5.1 and Remark 5.2 the distributions ν^{η^N} converge in probability in the $d_{\mathcal{W}}^{sup}$ metric to ν when η^N is sampled according to $\tilde{\mathbb{P}}^N$. Thus for any $\varepsilon > 0$ and an open ball $\tilde{B}_\varepsilon = \{ \zeta \in \mathcal{M}(\tilde{\mathcal{D}}) \mid d_{\mathcal{W}}^{sup}(\zeta, \nu) < \varepsilon \}$ around ν in the $d_{\mathcal{W}}^{sup}$ metric we have $\lim_{N \rightarrow \infty} \tilde{\mathbb{P}}^N \left(\nu^{\eta^N} \in \tilde{B}_\varepsilon \right) = 1$. Since convergence in $d_{\mathcal{W}}^{sup}$ implies convergence in $d_{\mathcal{W}}$, to prove (54) it is enough to analyze the probability $\mathbb{P}_u^N \left(\nu^{\eta^N} \in \tilde{B}_\varepsilon \right)$.

Fix arbitrary $\delta > 0$ and let $U_N = \{ |Y_N| \leq \delta \}$. Let $V_{N,\varepsilon} = U_N \cap \{ \nu^{\eta^N} \in \tilde{B}_\varepsilon \}$ and $\frac{d\mathbb{P}_u^N}{d\tilde{\mathbb{P}}^N} = \frac{d\mathbb{P}_u^N}{d\tilde{\mathbb{P}}^N}(T)$. With \mathbb{E} denoting the expectation with respect to \mathbb{P}_u^N and $\tilde{\mathbb{E}}$ with respect to $\tilde{\mathbb{P}}^N$ we have for any $\varepsilon > 0$ and sufficiently large N

$$\begin{aligned} \mathbb{P}_u^N \left(\nu^{\eta^N} \in \tilde{B}_\varepsilon \right) &= \mathbb{E} \left(\mathbb{1}_{\{ \nu^{\eta^N} \in \tilde{B}_\varepsilon \}} \right) \geq \mathbb{E} \left(\mathbb{1}_{V_{N,\varepsilon}} \right) \\ &= \tilde{\mathbb{E}} \left(\frac{d\mathbb{P}_u^N}{d\tilde{\mathbb{P}}^N} \mathbb{1}_{V_{N,\varepsilon}} \right) \geq \tilde{\mathbb{P}}^N(V_{N,\varepsilon}) \left(\inf_{V_{N,\varepsilon}} \frac{d\mathbb{P}_u^N}{d\tilde{\mathbb{P}}^N} \right). \end{aligned}$$

We have $\lim_{N \rightarrow \infty} \tilde{\mathbb{P}}^N(V_{N,\varepsilon}) = 1$ and on the event U_N we have

$$\frac{d\mathbb{P}_u^N}{d\tilde{\mathbb{P}}^N}(T) \geq \exp \left\{ -\frac{1}{2} N^\gamma \left[\int_0^T \frac{1}{N} \sum_{x=1}^N v_x(\eta_s)^2 ds + \delta \right] \right\}.$$

This implies

$$N^{-\gamma} \log \mathbb{P}_u^N \left(\nu^{\eta^N} \in \tilde{B}_\varepsilon \right) \geq - \inf_{\eta \in V_{N,\varepsilon}} I_N(\eta) - \delta, \quad (56)$$

where

$$I_N(\eta) = \frac{1}{2} \left(\frac{1}{N} \sum_{x=1}^N \int_0^T v_x(s, \eta_s)^2 ds \right).$$

Now it is not difficult to see that the infimum on the right hand side of (56) converges to $I(\mu)$ as $N \rightarrow \infty$ and then $\varepsilon \rightarrow 0$. When (X, Φ) is sampled from ν , X is the solution of (29) with a uniformly random initial condition, so the energy $I(\mu)$ is simply equal to

$$\mathbb{E} \int_0^T V(s, X(s), \Phi(s))^2 ds,$$

where the expectation is with respect to the choice of $(X(0), \Phi(0))$. Recall the notation $X_i(\eta_t^N) = \frac{1}{N}x_i(\eta_t^N)$, $\Phi_i(\eta_t^N) = \frac{1}{N}\phi_i(\eta_t^N)$. In light of (33) what we need to show is that

$$\inf_{\eta \in V_{N,\varepsilon}} \left(\frac{1}{N} \sum_{i=1}^N \int_0^T V(s, X_i(\eta_s^N), \Phi_i(\eta_s^N))^2 ds \right) \rightarrow \mathbb{E} \int_0^T V(s, X(s), \Phi(s))^2 ds \quad (57)$$

as $N \rightarrow \infty$ and then $\varepsilon \rightarrow 0$.

Consider the trajectory η^N and for any particle i let $(X_i(t), \Phi_i(t))$ denote the solution of (29) corresponding to the initial condition $(X_i(\eta_0^N), \Phi_i(\eta_0^N))$. Since the velocities V are bounded, we can write

$$\begin{aligned} & \left| \int_0^T [V(s, X_i(\eta_s^N), \Phi_i(\eta_s^N))^2 - V(s, X_i(s), \Phi_i(s))^2] ds \right| \\ & \leq C \int_0^T |V(s, X_i(\eta_s^N), \Phi_i(\eta_s^N)) - V(s, X_i(s), \Phi_i(s))| ds \\ & \leq KT \max \left\{ \sup_{t \leq T} |X_i(\eta_t^N) - X_i(t)|, \sup_{t \leq T} |\Phi_i(\eta_t^N) - \Phi_i(t)| \right\} \end{aligned} \quad (58)$$

for some $C, K > 0$ depending on the bound on V and the Lipschitz constant of V .

Now note that if $\nu^{\eta^N} \in \tilde{B}_\varepsilon$, then by considering the same coupling as in the proof of Theorem 5.1 we have

$$\limsup_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \left| \max \left\{ \sup_{t \leq T} |X_i(\eta_t^N) - X_i(t)|, \sup_{t \leq T} |\Phi_i(\eta_t^N) - \Phi_i(t)| \right\} \right| \leq \varepsilon' \quad (59)$$

for some $\varepsilon' > 0$ satisfying $\varepsilon' \rightarrow 0$ as $\varepsilon \rightarrow 0$. Since $\{\nu^{\eta^N} \in \tilde{B}_\varepsilon\} \subseteq V_{N,\varepsilon}$, combining this with (58) we obtain that the left hand side of (57) converges to

$$\frac{1}{N} \sum_{i=1}^N \int_0^T V(s, X_i(s), \Phi_i(s))^2 ds$$

as $N \rightarrow \infty$ and then $\varepsilon \rightarrow 0$.

Since $(X_i(t), \Phi_i(t))$ is a solution of (29) and V is the derivative of X , the integral is equal simply to the energy of the path $X_i(t)$. Since for each i the initial condition $\Phi_i(\eta_0^N)$ has uniform distribution on $\{\frac{1}{N}, \dots, 1\}$, independently for all i , it follows easily that this expression converges with high probability to the expected energy on the right hand side of (57). This implies $\inf_{\eta \in V_{N,\varepsilon}} I_N(\eta) \rightarrow I(\nu)$ as $N \rightarrow \infty$ and then $\varepsilon \rightarrow 0$. Since in (56) we can take δ to be arbitrarily small, this proves (54) and finishes the proof of the lower bound. \square

With this theorem the large deviation lower bound for generalized solutions to Euler equations, announced as Theorem A in the introduction, is now an easy corollary.

Theorem 7.3. *Let \mathbb{P}^N be the law of the interchange process η^N and let μ^{η^N} be the (random) distribution of the corresponding permutation process X^{η^N} . Let π be a permutation process which is a generalized solution to Euler equations (19). Provided π satisfies Assumptions (3.1), for any open set $\mathcal{O} \subseteq \mathcal{P}$ such that $\pi \in \mathcal{O}$ we have*

$$\liminf_{N \rightarrow \infty} N^{-\gamma} \log \mathbb{P}^N \left(\mu^{\eta^N} \in \mathcal{O} \right) \geq -I(\pi).$$

Proof. Let $\pi^{\beta,\delta}$ be the distribution of the process $X^{\beta,\delta}$ defined in Sect. 3. By the first part of Proposition 3.7 we have $d_{\mathcal{W}}^{sup}(\pi, \pi^{\beta,\delta}) \rightarrow 0$ as first δ and then $\beta \rightarrow 0$, in particular for small enough δ and β we have $\pi^{\beta,\delta} \in \mathcal{O}$. Then Theorem 7.2 implies that

$$\liminf_{N \rightarrow \infty} N^{-\gamma} \log \mathbb{P}^N \left(\mu^{\eta^N} \in \mathcal{O} \right) \geq -I(\pi^{\beta,\delta}).$$

Since by the second part of Proposition 3.7 we have $\lim_{\beta \rightarrow 0} \lim_{\delta \rightarrow 0} I(\pi^{\beta,\delta}) = I(\pi)$, the lower bound is proved. \square

8 Large Deviation Upper Bound

In this section we prove Theorem B, a large deviation upper bound for the distribution of the interchange process (we will drop the term “unbiased” from now on). As a first step we will bound the probability that after a (possibly short) time $t > 0$ we see a fixed permutation in the interchange process. This is summarized in the following

PROPOSITION 8.1. *Let \mathbb{P}^N be the law of the interchange process, with $\eta = \eta^N$ denoting the trajectory of the process. Let $\sigma^N \in \mathcal{S}_N$ be a sequence of permutations. For any $t > 0$ we have*

$$\limsup_{N \rightarrow \infty} N^{-\gamma} \log \mathbb{P}^N(\eta_0^{-1} \eta_t = \sigma_N) \leq -\frac{1}{t} \left(\liminf_{N \rightarrow \infty} I(\sigma^N) \right),$$

where $I(\sigma)$ is the energy of the permutation σ defined in (8).

In other words, the large deviation rate of seeing a permutation σ at time t in the interchange process is asymptotically bounded from above by $\frac{1}{t}$ times the energy of the permutation σ .

To prove Proposition 8.1 we will employ exponential martingales. The idea is as follows – if $M_S(\eta)$ is a function of the process (depending on some set of parameters S) which is a positive mean one martingale, then for any permutation $\sigma \in \mathcal{S}_N$ we can write

$$\begin{aligned} \mathbb{P}^N(\eta_0^{-1}\eta_t = \sigma) &= \mathbb{E}(\mathbb{1}_{\{\eta_0^{-1}\eta_t = \sigma\}}) = \mathbb{E}\left(M_S(\eta)M_S(\eta)^{-1}\mathbb{1}_{\{\eta_0^{-1}\eta_t = \sigma\}}\right) \\ &\leq \sup_{\{\chi_0^{-1}\chi_t = \sigma\}} M_S(\chi)^{-1}\mathbb{E}\left(M_S(\eta)\mathbb{1}_{\{\eta_0^{-1}\eta_t = \sigma\}}\right) \leq \sup_{\{\chi_0^{-1}\chi_t = \sigma\}} M_S(\chi)^{-1}, \end{aligned} \tag{60}$$

where the supremum is over all deterministic permutation-valued paths $\chi = (\chi_s, 0 \leq s \leq T)$ satisfying $\chi_0^{-1}\chi_t = \sigma$ and the last inequality comes from the fact that $M_S(\chi)$ is a positive mean one martingale. If M_S depends only on the increment $\chi_0^{-1}\chi_t$, we obtain a particularly simple expression

$$\mathbb{P}^N(\eta_0^{-1}\eta_t = \sigma) \leq M_S(\sigma)^{-1}.$$

We can then optimize over the set of parameters S to obtain a large deviation upper bound. The family of martingales we will use is similar to the one used in analyzing large deviations for a simple random walk.

Fix $t > 0$ and a sequence $S = (s_1, \dots, s_N)$, with $s_i \in \left\{ \frac{-1+\frac{1}{N}}{t}, \frac{-1+\frac{2}{N}}{t}, \dots, \frac{1-\frac{2}{N}}{t}, \frac{1-\frac{1}{N}}{t} \right\}$. We will think of s_i as “velocity” assigned to the particle i . Consider the function

$$F_S(\eta_t) = \varepsilon \sum_{i=1}^N s_i x_i(\eta_t),$$

where $x_i(\eta_t)$ is the position of the particle i in the configuration η_t . If \mathcal{L} is the generator of the interchange process, given by (12), then by the formula (36) for exponential martingales we obtain that

$$M_t^S = \exp \left\{ F_S(\eta_t) - F_S(\eta_0) - \int_0^t e^{-F_S(\eta_s)} \mathcal{L} e^{F_S(\eta_s)} ds \right\}$$

is a mean one positive martingale with respect to \mathbb{P}^N .

For simplicity we will use the same notation $s_x(\eta) = s_{\eta^{-1}(x)}$ as for velocities v_x of particles in the previous sections (with the convention that i denotes labels of particles and x denotes the positions), although bear in mind that now s_x are just parameters, not related in any way to the the biased interchange process considered in the preceding sections. We have

$$\mathcal{L} e^{F_S(\eta)} = \frac{1}{2} N^\alpha \sum_{x=1}^{N-1} \left(e^{F_S(\eta^{x,x+1})} - e^{F_S(\eta)} \right) = \frac{1}{2} N^\alpha \sum_{x=1}^{N-1} \left(e^{F_S(\eta) + \varepsilon [s_x(\eta) - s_{x+1}(\eta)]} - e^{F_S(\eta)} \right),$$

so

$$M_t^S = \exp \left\{ \varepsilon \sum_{i=1}^N s_i (x_i(\eta_t) - x_i(\eta_0)) - \frac{1}{2} N^\alpha \int_0^t \sum_{x=1}^{N-1} \left(e^{\varepsilon [s_x(\eta_s) - s_{x+1}(\eta_s)]} - 1 \right) ds \right\}.$$

Expanding up to order ε^2 we get

$$\begin{aligned} M_t^S = \exp \left\{ \varepsilon \sum_{i=1}^N s_i (x_i(\eta_t) - x_i(\eta_0)) - \frac{1}{2} N^\alpha \varepsilon \int_0^t \sum_{x=1}^{N-1} [s_x(\eta_s) - s_{x+1}(\eta_s)] ds \right. \\ \left. - \frac{1}{4} N^\alpha \varepsilon^2 \int_0^t \sum_{x=1}^{N-1} (s_x(\eta_s) - s_{x+1}(\eta_s))^2 ds + O(N^{\alpha+1} \varepsilon^3) \right\}, \end{aligned} \quad (61)$$

where the constants in the $O(\cdot)$ notation depend on t (which is fixed). Observe that the sum of $s_x - s_{x+1}$ telescopes, leaving only terms with s_1 and s_N , which are $O(N^\alpha \varepsilon) = o(N^\gamma)$. Rescaling by appropriate powers of N and expressing the exponents in terms of the large deviation exponent γ we get

$$\begin{aligned} M_t^S = \exp \left\{ N^\gamma \left[\frac{1}{N} \sum_{i=1}^N s_i \left(\frac{x_i(\eta_t) - x_i(\eta_0)}{N} \right) \right. \right. \\ \left. \left. - \frac{1}{4} \int_0^t \frac{1}{N} \sum_{x=1}^{N-1} (s_x(\eta_s) - s_{x+1}(\eta_s))^2 ds \right] + o(N^\gamma) \right\}. \end{aligned}$$

Expanding $(s_x - s_{x+1})^2$ we obtain (after adding and subtracting the boundary terms s_1^2, s_N^2 which are only $o(1)$ after rescaling) twice the sum of s_x^2 and the sum of mixed terms $s_x s_{x+1}$. Since $\sum_{x=1}^N s_x^2 = \sum_{i=1}^N s_i^2$ does not depend on time, we can write

$$\begin{aligned} M_t^S = \exp \left\{ N^\gamma \left[\frac{1}{N} \sum_{i=1}^N s_i \left(\frac{x_i(\eta_t) - x_i(\eta_0)}{N} \right) - \frac{t}{2} \left(\frac{1}{N} \sum_{i=1}^N s_i^2 \right) \right. \right. \\ \left. \left. + \frac{1}{2} \int_0^t \frac{1}{N} \sum_{x=1}^{N-1} s_x(\eta_s) s_{x+1}(\eta_s) ds + o(1) \right] \right\}. \end{aligned}$$

As in the proof of the law of large numbers we want to use the one block estimate to get rid of the sum involving correlations between s_x for adjacent x . This time the correlation term might not be small, since s_i are arbitrary, but typically it will be nonnegative, so we can neglect it for the sake of the upper bound. More precisely, we have the following

LEMMA 8.2. Let \mathbb{P}^N be the law of the interchange process. Fix $t > 0$ and let $s_i \in \left\{ \frac{-1}{t}, \frac{-1+\frac{1}{N}}{t}, \dots, \frac{1-\frac{1}{N}}{t}, \frac{1}{t} \right\}$. Then, with notation as above, we have for any $\delta > 0$

$$\limsup_{N \rightarrow \infty} N^{-\gamma} \log \mathbb{P}^N \left(\int_0^t \frac{1}{N} \sum_{x=1}^{N-1} s_x(\eta_s) s_{x+1}(\eta_s) ds \leq -\delta \right) = -\infty.$$

Proof. We employ Lemma 6.2 with $w(x, \phi) = \frac{2x-1}{t}$, $a_i = s_i$ and $b_x(\eta) = a_x(\eta)$, in particular $h_x(\eta) = s_x(\eta) s_{x-1}(\eta)$. As in the lemma consider $\mathbb{E}_{\mu_{x,l}^{\eta_s}}(s_x(\eta) s_{x+1}(\eta))$, where $\mu_{x,l}^{\eta_s}$ is the empirical distribution of a_i in a box $\Lambda_{x,l}$. Let us write

$$\begin{aligned} & \int_0^t \frac{1}{N} \sum_{x=1}^{N-1} s_x(\eta_s) s_{x+1}(\eta_s) ds \\ &= \int_0^t \frac{1}{N} \sum_{x=1}^{N-1} \left(s_x(\eta_s) s_{x+1}(\eta_s) - \mathbb{E}_{\mu_{x,l}^{\eta_s}} [s_x(\eta) s_{x+1}(\eta)] \right) ds \\ & \quad + \int_0^t \frac{1}{N} \sum_{x=1}^{N-1} \mathbb{E}_{\mu_{x,l}^{\eta_s}} [s_x(\eta) s_{x+1}(\eta)] ds. \end{aligned} \tag{62}$$

Since under $\mu_{x,l}^{\eta_s}$ the colors are i.i.d. random variables, we have $\mathbb{E}_{\mu_{x,l}^{\eta_s}} [s_x(\eta) s_{x+1}(\eta)] = \left(\mathbb{E}_{\mu_{x,l}^{\eta_s}} s_x(\eta) \right)^2 \geq 0$, so the second term on the right hand side of (62) is nonnegative for every l . Lemma 6.2 guarantees that for any $\delta > 0$

$$\begin{aligned} & \limsup_{l \rightarrow \infty} \limsup_{N \rightarrow \infty} N^{-\gamma} \log \mathbb{P}^N \left(\int_0^t \frac{1}{N} \sum_{x=1}^{N-1} (s_x(\eta_s) s_{x+1}(\eta_s) \right. \\ & \quad \left. - \mathbb{E}_{\mu_{x,l}^{\eta_s}} [s_x(\eta) s_{x+1}(\eta)] \right) ds \leq -\delta \Big) = -\infty. \end{aligned}$$

Since the left hand side of (62) does not depend on l , this finishes the proof. □

With this lemma the proof of Proposition 8.1 is rather straightforward.

Proof of Proposition 8.1. Lemma 8.2 implies that for any $a > 0$ there exist sets $\mathcal{O}_{N,a}$ such that on $\mathcal{O}_{N,a}$ we have

$$M_t^S(\eta) \geq \exp \left\{ N^\gamma \left[\frac{1}{N} \sum_{i=1}^N s_i \left(\frac{x_i(\eta_t) - x_i(\eta_0)}{N} \right) - \frac{t}{2} \left(\frac{1}{N} \sum_{i=1}^N s_i^2 \right) - a + o(1) \right] \right\}, \tag{63}$$

with the $o(1)$ term depending on t , and $\mathbb{P}^N(\mathcal{O}_{N,a}^c) \rightarrow 0$ as $N \rightarrow \infty$ superexponentially fast

$$\limsup_{N \rightarrow \infty} N^{-\gamma} \log \mathbb{P}^N(\mathcal{O}_{N,a}^c) = -\infty.$$

Now we can use the strategy outlined earlier with the positive mean one martingale $M_t^S(\eta)$. We write

$$\begin{aligned} \mathbb{P}^N(\eta_0^{-1}\eta_t = \sigma^N) &= \mathbb{E}\left(\mathbb{1}_{\{\eta_0^{-1}\eta_t = \sigma^N\}}\right) = \mathbb{E}\left(M_t^S(\eta)^{-1}M_t^S(\eta)\mathbb{1}_{\{\eta_0^{-1}\eta_t = \sigma^N\}}\right) \\ &= \mathbb{E}\left(M_t^S(\eta)^{-1}M_t^S(\eta)\mathbb{1}_{\{\eta_0^{-1}\eta_t = \sigma^N\}}\mathbb{1}_{\mathcal{O}_{N,a}}\right) + \mathbb{E}\left(\mathbb{1}_{\{\eta_0^{-1}\eta_t = \sigma^N\}}\mathbb{1}_{\mathcal{O}_{N,a}^c}\right). \end{aligned}$$

On $\mathcal{O}_{N,a}$ we can use the bound (63) obtained above. Note also that on the event $\{\eta_0^{-1}\eta_t = \sigma^N\}$ we have $x_i(\eta_t) - x_i(\eta_0) = \sigma^N(i) - i$, which together with (60) leads us to

$$\mathbb{P}^N(\eta_0^{-1}\eta_t = \sigma^N) \leq e^{-N^\gamma(I_S(\sigma^N) - a + o(1))} + \mathbb{P}^N(\mathcal{O}_{N,a}^c), \tag{64}$$

where

$$I_S(\sigma^N) = \frac{1}{N} \sum_{i=1}^N s_i \left(\frac{\sigma^N(i) - i}{N} \right) - \frac{t}{2} \left(\frac{1}{N} \sum_{i=1}^N s_i^2 \right).$$

To optimize over the choice of $S = (s_1, \dots, s_N)$, observe that $I_S(\sigma^N)$ is quadratic in s_i , so an easy calculation shows that the optimal choice is

$$s_i = \frac{\sigma^N(i) - i}{tN},$$

which is valid, since we assumed $s_i \in \left\{ \frac{-1 + \frac{1}{N}}{t}, \frac{-1 + \frac{2}{N}}{t}, \dots, \frac{1 - \frac{1}{N}}{t} \right\}$. This gives the maximal value of $I_S(\sigma^N)$ equal to

$$\frac{1}{2} \left(\frac{1}{N} \sum_{i=1}^N \frac{1}{t} \left(\frac{\sigma^N(i) - i}{N} \right)^2 \right),$$

which is exactly the energy $I(\sigma^N)$ rescaled by t . Inserting this into (64) gives us

$$\mathbb{P}^N(\eta_0^{-1}\eta_t = \sigma^N) \leq e^{-N^\gamma(I(\sigma^N) - a + o(1))} + \mathbb{P}^N(\mathcal{O}_{N,a}^c)$$

Since $\limsup_{n \rightarrow \infty} \frac{1}{n} \log(a_n + b_n) = \max\{\limsup_{n \rightarrow \infty} \frac{1}{n} \log a_n, \limsup_{n \rightarrow \infty} \frac{1}{n} \log b_n\}$, we obtain

$$\begin{aligned} &\limsup_{N \rightarrow \infty} N^{-\gamma} \log \mathbb{P}^N(\eta_0^{-1}\eta_t = \sigma^N) \\ &\leq \max \left\{ -\liminf_{N \rightarrow \infty} I(\sigma^N) + a, \limsup_{N \rightarrow \infty} N^{-\gamma} \log \mathbb{P}^N(\mathcal{O}_{N,a}^c) \right\}. \end{aligned}$$

The second lim sup is $-\infty$ and by taking $a \rightarrow 0$ we arrive at

$$\limsup_{N \rightarrow \infty} N^{-\gamma} \log \mathbb{P}^N(\eta_0^{-1}\eta_t = \sigma^N) \leq -\liminf_{N \rightarrow \infty} I(\sigma^N)$$

as desired. □

We can readily extend the bound from Proposition 8.1 to all finite-dimensional distributions of the interchange process as follows. Fix a finite set of times $0 \leq t_0 < t_1 < \dots < t_k \leq T$ and for clarity of notation let us write $\vec{\eta}_{t_0, \dots, t_k} = (\eta_{t_0}^{-1} \eta_{t_1}, \dots, \eta_{t_{k-1}}^{-1} \eta_{t_k})$ for the corresponding sequence of increments of η . Suppose we want to bound the probability $\mathbb{P}^N(\vec{\eta}_{t_0, \dots, t_k} = (\sigma_1^N, \dots, \sigma_k^N))$, where $(\sigma_1^N, \dots, \sigma_k^N)$ is a fixed sequence of permutations for each N , $\sigma_j^N \in \mathcal{S}_N$. Recall that the interchange process has independent increments, i.e., the permutations $(\eta_{t_j})^{-1} \eta_{t_{j+1}}$ for any family non-overlapping intervals $[t_j, t_{j+1})$ are independent. Therefore we can write

$$\mathbb{P}^N(\vec{\eta}_{t_0, \dots, t_k} = (\sigma_1^N, \dots, \sigma_k^N)) = \prod_{j=1}^k \mathbb{P}^N(\eta_{t_{j-1}}^{-1} \eta_{t_j} = \sigma_j^N).$$

As the interchange process is stationary, we have $\mathbb{P}^N(\eta_{t_{j-1}}^{-1} \eta_{t_j} = \sigma_j^N) = \mathbb{P}^N(\eta_0^{-1} \eta_{t_j - t_{j-1}} = \sigma_j^N)$. Thus by applying Proposition 8.1 we obtain

$$\limsup_{N \rightarrow \infty} N^{-\gamma} \log \mathbb{P}^N(\vec{\eta}_{t_0, \dots, t_k} = (\sigma_1^N, \dots, \sigma_k^N)) \leq - \liminf_{N \rightarrow \infty} \sum_{j=1}^k \frac{1}{t_j - t_{j-1}} I(\sigma_j^N). \tag{65}$$

Recall that μ^{η^N} denotes the distribution of the random permutation process associated to η^N (defined by (5)) and for a finite partition Π by $I^\Pi(\mu^{\eta^N})$ we denote the approximation of energy of μ^{η^N} associated to Π (defined by (11)). From Eq. (65) we obtain the following corollary which will be useful later

COROLLARY 8.3. *For any $C > 0$ and any finite partition $\Pi = \{0 = t_0 < t_1 < \dots < t_k = T\}$ we have*

$$\limsup_{N \rightarrow \infty} N^{-\gamma} \log \mathbb{P}^N(I^\Pi(\mu^{\eta^N}) \geq C) \leq -C.$$

Proof. Consider the set A_C^N of all sequences of permutations $(\sigma_1^N, \dots, \sigma_k^N)$, $\sigma_j^N \in \mathcal{S}_N$, such that $\sum_{j=1}^k \frac{1}{t_j - t_{j-1}} I(\sigma_j^N) \geq C$. By performing a union bound over all such sequences we get

$$\mathbb{P}^N(I^\Pi(\mu^{\eta^N}) \geq C) \leq N!^k \sup_{A_C^N} \mathbb{P}^N(\vec{\eta}_{t_0, \dots, t_k} = (\sigma_1^N, \dots, \sigma_k^N)),$$

Now it is enough to observe that for fixed k we have $\log(N!^k) = o(N^\gamma)$ and apply (65). □

Now we can proceed to prove a general large deviation upper bound, announced as Theorem B in the introduction,. Recall that $\mathcal{P} \subseteq \mathcal{M}(\mathcal{D})$ denotes the space of all permuton and approximate permuton processes.

Theorem 8.4. *Let \mathbb{P}^N be the law of the interchange process η^N and let μ^{η^N} be the (random) distribution of the corresponding random permutation process X^{η^N} . For any closed set $\mathcal{C} \subseteq \mathcal{P}$ we have*

$$\limsup_{N \rightarrow \infty} N^{-\gamma} \log \mathbb{P}^N \left(\mu^{\eta^N} \in \mathcal{C} \right) \leq - \inf_{\pi \in \mathcal{C}} I(\pi),$$

where $I(\pi)$ is the energy of the process π defined by (10).

Proof. It is standard (see, e.g., [Var16, Lemma 2.3]) that the large deviation upper bound for closed sets follows from a local upper bound for open balls and exponential tightness of the sequence μ^{η^N} . The exponential tightness part will be proved in Proposition 8.5 below, so here we focus on the first part, that is, we will prove that for any $\pi \in \mathcal{P}$ we have

$$\limsup_{\varepsilon \rightarrow 0} \limsup_{N \rightarrow \infty} N^{-\gamma} \log \mathbb{P}^N \left(\mu^{\eta^N} \in B(\pi, \varepsilon) \right) \leq -I(\pi), \tag{66}$$

where $B(\pi, \varepsilon)$ denotes the open ε -ball around π in the Wasserstein distance $d_{\mathcal{W}}$ on \mathcal{P} .

Fix a finite set of times $0 = t_0 < t_1 < \dots < t_k = T$. Since almost surely the interchange process does not make jumps at any of the prescribed times t_0, t_1, \dots, t_k , by continuity of projections for any $\varepsilon > 0$ there exists $\varepsilon' > 0$ such that

$$\mathbb{P}^N \left(d_{\mathcal{W}}(\mu^{\eta^N}, \pi) < \varepsilon' \right) \leq \mathbb{P}^N \left(d(\mu_{t_0, t_1}^{\eta^N}, \pi_{t_0, t_1}) < \varepsilon \right) \wedge \dots \wedge d(\mu_{t_{k-1}, t_k}^{\eta^N}, \pi_{t_{k-1}, t_k}) < \varepsilon \Big), \tag{67}$$

where d denotes the Wasserstein distance on $\mathcal{M}([0, 1]^2)$. Furthermore, note that the permutation process with distribution μ^{η^N} has independent increments, i.e., the permutations $\left(\eta_{t_j}^N\right)^{-1} \eta_{t_{j+1}}^N$ for any family non-overlapping intervals $[t_j, t_{j+1})$ are independent. Thus we can write

$$\begin{aligned} & \mathbb{P}^N \left(d(\mu_{t_0, t_1}^{\eta^N}, \pi_{t_0, t_1}) < \varepsilon \right) \wedge \dots \wedge d(\mu_{t_{k-1}, t_k}^{\eta^N}, \pi_{t_{k-1}, t_k}) < \varepsilon \Big) \\ &= \prod_{i=0}^{k-1} \mathbb{P}^N \left(d(\mu_{t_i, t_{i+1}}^{\eta^N}, \pi_{t_i, t_{i+1}}) < \varepsilon \right). \end{aligned} \tag{68}$$

In this way we have reduced the problem to bounding the probability that the random measure $\mu_{t_i, t_{i+1}}^{\eta^N}$ is close to a fixed permuton $\pi_{t_i, t_{i+1}}$.

Fix i and consider all permutations $\sigma \in \mathcal{S}_N$ such that the empirical measure μ_σ satisfies $d(\mu_\sigma, \pi_{t_i, t_{i+1}}) < \varepsilon$. As there are at most $N!$ such permutations, by performing a union bound over this set we obtain

$$\mathbb{P}^N \left(d(\mu_{t_i, t_{i+1}}^{\eta^N}, \pi_{t_i, t_{i+1}}) < \varepsilon \right) \leq N! \sup_{\substack{\sigma \in \mathcal{S}_N \\ d(\mu_\sigma, \pi_{t_i, t_{i+1}}) < \varepsilon}} \mathbb{P}^N \left(\mu_{t_i, t_{i+1}}^{\eta^N} = \mu_\sigma \right),$$

where on the right hand side we have the probability that the random measure $\mu_{t_i, t_{i+1}}^{\eta^N}$ is equal to μ_σ . This probability is simply equal to $\mathbb{P}^N \left((\eta_{t_i}^N)^{-1} \eta_{t_{i+1}}^N = \sigma \right)$ and by stationarity of the interchange process this is the same as $\mathbb{P}^N \left((\eta_0^N)^{-1} \eta_{t_{i+1}-t_i}^N = \sigma \right)$. By employing Proposition 8.1, with $\sigma^N \in \mathcal{S}_N$ being any permutation attaining the supremum above, and noticing that $\log N! = o(N^\gamma)$ we get

$$\limsup_{N \rightarrow \infty} N^{-\gamma} \log \mathbb{P}^N \left(d(\mu_{t_i, t_{i+1}}^{\eta^N}, \pi_{t_i, t_{i+1}}) < \varepsilon \right) \leq \limsup_{N \rightarrow \infty} \left(-\frac{1}{t_{i+1} - t_i} I(\sigma^N) \right).$$

Now observe that for any σ such that $d(\mu_\sigma, \pi_{t_i, t_{i+1}}) < \varepsilon$ the energy $I(\sigma) = I(\mu_\sigma)$ has to be close to $I(\pi_{t_i, t_{i+1}})$, the energy of the permutation $\pi_{t_i, t_{i+1}}$ (recall definition 9), since I is continuous in the weak topology on $\mathcal{M}([0, 1]^2)$. Thus upon taking $\varepsilon \rightarrow 0$ we obtain

$$\limsup_{\varepsilon \rightarrow 0} \limsup_{N \rightarrow \infty} N^{-\gamma} \log \mathbb{P}^N \left(d(\mu_{t_i, t_{i+1}}^{\eta^N}, \pi_{t_i, t_{i+1}}) < \varepsilon \right) \leq -\frac{1}{t_{i+1} - t_i} I(\pi_{t_i, t_{i+1}}).$$

Applying this estimate to the product in (68) and observing that in (67) without loss of generality we can assume $\varepsilon' \leq \varepsilon$, we arrive at the following bound

$$\limsup_{\varepsilon \rightarrow 0} \limsup_{N \rightarrow \infty} N^{-\gamma} \log \mathbb{P}^N \left(d_{\mathcal{W}}(\mu^{\eta^N}, \pi) < \varepsilon \right) \leq -\sum_{i=0}^{k-1} \frac{1}{t_{i+1} - t_i} I(\pi_{t_i, t_{i+1}}).$$

Since t_0, t_1, \dots, t_k were arbitrary, by optimizing over all finite partitions $\Pi = \{0 = t_0 < t_1 < \dots < t_k = T\}$ we obtain

$$\limsup_{\varepsilon \rightarrow 0} \limsup_{N \rightarrow \infty} N^{-\gamma} \log \mathbb{P}^N \left(d_{\mathcal{W}}(\mu^{\eta^N}, \pi) < \varepsilon \right) \leq -\sup_{\Pi} \sum_{i=0}^{k-1} \frac{1}{t_{i+1} - t_i} I(\pi_{t_i, t_{i+1}}).$$

Recalling the definitions (6), (10) and (11), to prove (66) it remains to show that we have $I(\pi) = \sup_{\Pi} I^{\Pi}(\pi)$, which is exactly the statement of Lemma 2.2. \square

PROPOSITION 8.5. *The family of measures μ^{η^N} is exponentially tight, that is, there exists a sequence of compact sets $K_m \subseteq \mathcal{P}$ such that*

$$\limsup_{N \rightarrow \infty} N^{-\gamma} \log \mathbb{P}^N (\mu^{\eta^N} \notin K_m) \leq -m.$$

Proof. The idea of the proof is to show that having many particles whose trajectories have poor modulus of continuity (which would spoil compactness) necessarily implies the process having high energy, which by Corollary 8.3 is unlikely.

Recall that $\mathcal{D} = \mathcal{D}([0, T], [0, 1])$ is the space of all càdlàg paths from $[0, T]$ to $[0, 1]$. It will be convenient to work with the following notion of càdlàg modulus of continuity – for a path $f \in \mathcal{D}$ we define

$$w''_{\delta}(f) = \sup_{\substack{t_1 \leq t \leq t_2 \\ t_2 - t_1 \leq \delta}} \{|f(t) - f(t_1)| \wedge |f(t_2) - f(t)|\}.$$

By a characterization of compactness in the Skorokhod space ([Bil13, Theorem 12.4]) a set $A \subseteq \mathcal{D}$ has compact closure if and only if the following conditions hold

$$\begin{cases} \sup_{f \in A} \sup_{t \in [0, T]} |f(t)| < \infty, \\ \lim_{\delta \rightarrow 0} \sup_{f \in A} w''_{\delta}(f) = 0, \\ \lim_{\delta \rightarrow 0} \sup_{f \in A} |f(\delta) - f(0)| = 0, \\ \lim_{\delta \rightarrow 0} \sup_{f \in A} |f(T-) - f(T - \delta)| = 0. \end{cases}$$

In our setting the first condition is trivially satisfied. To exploit the other conditions let us introduce for any $m, r \geq 1$ the following sets

$$\begin{aligned} K_{m,r}^w &= \bigcap_{k \geq 1} \left\{ f \in \mathcal{D} \mid w''_{\delta_k(m,r)}(f) \leq \varepsilon_k \right\}, \\ K_{m,r}^0 &= \bigcap_{k \geq 1} \left\{ f \in \mathcal{D} \mid |f(\delta_k(m,r)) - f(0)| \leq \varepsilon_k \right\}, \\ K_{m,r}^T &= \bigcap_{k \geq 1} \left\{ f \in \mathcal{D} \mid |f(T-) - f(T - \delta_k(m,r))| \leq \varepsilon_k \right\}, \end{aligned}$$

and

$$K_{m,r} = K_{m,r}^w \cap K_{m,r}^0 \cap K_{m,r}^T,$$

where $\varepsilon_k = 4^{-k}$ and $\delta_k(m, r)$ will be appropriately chosen later. We will assume that for fixed m, r we have $\lim_{k \rightarrow \infty} \delta_k(m, r) = 0$ and that for any $k \geq 1$ both $\frac{T}{\delta_k}$ and $\frac{\delta_k(m,r)}{\delta_{k+1}(m,r)}$ are integer (the latter assumption is for simplicity of notation only). Note that by the aforementioned compactness conditions each set $K_{m,r}$ has compact closure in \mathcal{D} .

Let

$$K_m = \bigcap_{r \geq 1} \left\{ \mu \in \mathcal{M}(\mathcal{D}) \mid \mu(K_{m,r}) \geq 1 - \frac{1}{r} \right\}.$$

We claim that K_m has compact closure in $\mathcal{M}(\mathcal{D})$. Indeed, by Prokhorov's theorem it is enough to prove that K_m is tight. If $\mu \in K_m$, then for any $r \geq 1$ we have $\mu(K_{m,r}^c) \leq \frac{1}{r}$, so the sets $K_{m,r}$ form the family of compact sets needed for tightness of K_m .

The sets K_m (possibly after taking their closures) will form the family of compact sets needed for exponential tightness. Thus our goal is to bound $\mathbb{P}^N(\mu^{\eta^N} \notin K_m)$. Let us write

$$\mathbb{P}^N(\mu^{\eta^N} \notin K_m) = \mathbb{P}^N \left(\exists_{r \geq 1} \mu^{\eta^N}(K_{m,r}^c) \geq \frac{1}{r} \right) \leq \sum_{r \geq 1} \mathbb{P}^N \left(\mu^{\eta^N}(K_{m,r}^c) \geq \frac{1}{r} \right).$$

It is enough to show that for any $m, r \geq 1$ and any $N \geq 1$ we have

$$\mathbb{P}^N \left(\mu^{\eta^N}(K_{m,r}^c) \geq \frac{1}{r} \right) \leq C e^{-mrN^\gamma}, \tag{69}$$

where $C > 0$ is some global constant.

For any given m and r , observe that $\mu^{\eta^N}(K_{m,r}^c) \geq \frac{1}{r}$ means that in η^N we have at least $\frac{N}{r}$ particles with paths $f \notin K_{m,r}$. Since $K_{m,r} = K_{m,r}^w \cap K_{m,r}^0 \cap K_{m,r}^T$, clearly it is enough to estimate separately the probabilities that at least $\frac{N}{3r}$ particles have paths respectively not in $K_{m,r}^w$, $K_{m,r}^0$ or $K_{m,r}^T$. The argument for $K_{m,r}^0$ and $K_{m,r}^T$ is much simpler, so we skip it and concentrate only on the case of $K_{m,r}^w$. For simplicity we will write $\alpha(r) = \frac{1}{3r}$.

For fixed m and r we will call a path f *bad* if $w''_{\delta_k(m,r)}(f) > \varepsilon_k(m, r)$ for some $k \geq 1$. We will call f *bad exactly at scale k* if $w''_{\delta_k(m,r)}(f) > \varepsilon_k(m, r)$, but $w''_{\delta_j(m,r)}(f) \leq \varepsilon_j(m, r)$ for all $j \geq k + 1$. Recalling the definition of the set $K_{m,r}^w$, the event whose probability we would like to bound is

$$A_N^{m,r} = \{ \text{there exist } \geq \alpha(r)N \text{ particles with bad paths} \}.$$

Consider now the events

$$B_N^{m,r,k} = \left\{ \text{there exist } \geq \frac{\alpha(r)}{2^k} N \text{ particles whose paths are bad exactly at scale } k \right\}.$$

Note that if f is a bad path with jumps of fixed size $\frac{1}{N}$, then there exists $k \geq 1$ such that f is bad exactly at scale k (since all paths we are considering are càdlàg). Thus we have $A_N^{m,r} \subseteq \bigcup_{k \geq 1} B_N^{m,r,k}$, so

$$\mathbb{P}^N \left(\mu^{\eta^N}((K_{m,r}^w)^c) \geq \alpha(r) \right) = \mathbb{P}^N(A_N^{m,r}) \leq \sum_{k \geq 1} \mathbb{P}^N(B_N^{m,r,k}).$$

Thus it is enough to show that for any $m, r, k \geq 1$ and any $N \geq 1$ we have

$$\mathbb{P}^N(B_N^{m,r,k}) \leq e^{-mrkN^\gamma}. \tag{70}$$

From now on we fix m, r, k and N . All paths we are considering are assumed to come from the interchange process η^N , in particular they have jumps of fixed size $\frac{1}{N}$. For the sake of brevity we will simply write $\delta_k = \delta_k(m, r)$.

Let us divide the interval $[0, T]$ into $J = \frac{T}{\delta_k}$ intervals of the form $[j\delta_k, (j+1)\delta_k]$, $j = 0, \dots, J-1$. Consider any path f which is bad exactly at scale k . The condition $w''_{\delta_k}(f) > \varepsilon_k$ implies that for some t, t_1, t_2 such that $t, t_2 \in [t_1, t_1 + \delta_k]$ we have $|f(t) - f(t_1)| > \varepsilon_k$ and $|f(t_2) - f(t)| > \varepsilon_k$. A simple application of the triangle inequality implies that there exists $j \in \{0, \dots, J-1\}$ and $t' \in [j\delta_k, (j+1)\delta_k]$ such that $|f(j\delta_k) - f(t')| > \frac{\varepsilon_k}{2}$.

Let us consider the interval $[s, s'] = [j\delta_k, (j + 1)\delta_k]$ obtained above. Subdivide it into $L = \frac{\delta_k}{\delta_{k+1}}$ intervals of the form $[s_\ell, s_{\ell+1}]$, $\ell = 0, \dots, L - 1$, where $s_\ell = s + \ell\delta_{k+1}$. For $\ell = 0, \dots, L - 1$ let $\Delta_\ell(f) = |f(s_\ell) - f(s_{\ell+1})|$.

The crucial observation is that $\sum_{\ell=0}^{L-1} \Delta_\ell(f) > \frac{\varepsilon_k}{4}$. To see this, consider t' such that $|f(s) - f(t')| > \frac{\varepsilon_k}{2}$, obtained above, and let $\tilde{\ell}$ be such that $t' \in (s_{\tilde{\ell}}, s_{\tilde{\ell}+1}]$. By the triangle inequality we have

$$\frac{\varepsilon_k}{2} < |f(s) - f(t')| \leq \sum_{\ell=0}^{\tilde{\ell}-2} \Delta_\ell(f) + |f(s_{\tilde{\ell}}) - f(t')|.$$

Since f is bad exactly at scale k , we have $w''_{\delta_{k+1}}(f) \leq \varepsilon_{k+1}$, which together with $|s_{\tilde{\ell}} - t'| \leq \delta_{k+1}$ implies $|f(s_{\tilde{\ell}}) - f(t')| \leq \varepsilon_{k+1} = 4^{-(k+1)} = \frac{\varepsilon_k}{4}$. Thus necessarily $\sum_{\ell=0}^{\tilde{\ell}-2} \Delta_\ell(f) > \frac{\varepsilon_k}{4}$. From this we obtain

$$\frac{\varepsilon_k^2}{16} < \left(\sum_{\ell=0}^{L-1} \Delta_\ell(f) \right)^2 \leq L \sum_{\ell=0}^{L-1} \Delta_\ell(f)^2, \tag{71}$$

where the right-hand side estimate follows from the Cauchy-Schwarz inequality.

Now let us suppose that the event $B_N^{m,r,k}$ holds. Then there exist at least $\frac{\alpha(r)}{2^k}N$ paths f_i for which the estimate (71) holds. Consider the partition $\Pi = \{0 = t_0 < t_1 < \dots < t_n = T\}$ where $n = \frac{T}{\delta_{k+1}}$, $t_j = j\delta_{k+1}$ for $j = 0, \dots, n$. Recalling that $f_i = \frac{1}{N}\eta^N(i)$, the definition of $\Delta_\ell(f)$ and the definition (11) of the energy $I^\Pi(\mu^{\eta^N})$ we obtain that on $B_N^{m,r,k}$ we have

$$\begin{aligned} I^\Pi(\mu^{\eta^N}) &= \frac{1}{N} \sum_{i=1}^N \left(\frac{1}{2} \sum_{j=1}^n \frac{|f_i(t_j) - f_i(t_{j-1})|^2}{t_j - t_{j-1}} \right) \\ &= \frac{1}{N} \sum_{i=1}^N \left(\frac{1}{2\delta_{k+1}} \sum_{j=1}^n |f_i(t_j) - f_i(t_{j-1})|^2 \right) > \frac{1}{N} \cdot \frac{\alpha(r)}{2^k} N \cdot \left(\frac{1}{2\delta_{k+1}} \frac{\varepsilon_k^2}{16L} \right) \\ &= \frac{\alpha(r)}{2^{k+5}} \frac{1}{\delta_{k+1}} \frac{\varepsilon_k^2}{\delta_{k+1}} = \frac{\varepsilon_k^2}{\delta_k} \frac{\alpha(r)}{2^{k+5}}. \end{aligned}$$

Writing again $\delta_k = \delta_k(m, r)$, we have thus obtained the bound

$$\mathbb{P}^N(B_N^{m,r,k}) \leq \mathbb{P}^N \left(I^\Pi(\mu^{\eta^N}) \geq \frac{\varepsilon_k^2}{\delta_k(m, r)} \frac{\alpha(r)}{2^{k+5}} \right).$$

Recalling $\varepsilon_k = 4^{-k}$, $\alpha(r) = \frac{1}{3r}$, we see that to prove (70) it is sufficient to take $\delta_k(m, r)$ small enough so that

$$\frac{4^{-2k}}{\delta_k(m, r)} \frac{1}{3r2^{k+5}} \geq 2mrk.$$

By applying Corollary 8.3 we obtain that

$$\mathbb{P}^N(B_N^{m,r,k}) \leq \mathbb{P}^N\left(I^\Pi(\mu^{\eta^N}) \geq 2mrk\right) \leq e^{-mrkN^\gamma}$$

for N large enough. By taking $\delta_k(m, r)$ even smaller if necessary we can make this estimate true for all values of $N \geq 1$, which proves (70) and finishes the proof of exponential tightness. \square

9 Asymptotics of Relaxed Sorting networks

In this section we prove the limiting behavior of random relaxed sorting networks, given by Theorem 1.1, and the asymptotic counting formula of Theorem 1.2. With the large deviation bounds obtained in the preceding sections both of the proofs are now rather straightforward.

Proof of Theorem 1.1. Let $\mathcal{R} \subseteq \mathcal{P}$ be the set of permutation processes X reaching exactly the reverse permutation at time 1, i.e., such that $(X_0, X_1) \sim (X, 1 - X)$, and likewise let \mathcal{R}_N be the set of permutation processes on N elements reaching exactly the reverse permutation $\text{rev}_N = (N \dots 2 1)$ at time 1. Let \mathcal{R}_δ denote the δ -neighborhood in the Wasserstein distance on \mathcal{P} of the set $\mathcal{R} \cup \bigcup_{N \geq 1} \mathcal{R}_N$.

Let η^N be the interchange process with $\alpha = 1 + \kappa \in (1, 2)$ and let μ^{η^N} be the distribution of the corresponding permutation process. By definition of a random relaxed sorting network, for any given $\delta > 0$ we have for sufficiently large N

$$\mathbb{P}^N(\pi_\delta^N \in B(\pi_{\mathcal{A}}, \varepsilon)) = \mathbb{P}^N(\mu^{\eta^N} \in B(\pi_{\mathcal{A}}, \varepsilon) | \mu^{\eta^N} \in \mathcal{R}_\delta).$$

Now, we have

$$\begin{aligned} &\mathbb{P}^N(\mu^{\eta^N} \notin B(\pi_{\mathcal{A}}, \varepsilon) | \mu^{\eta^N} \in \mathcal{R}_\delta) \\ &= \frac{1}{\mathbb{P}^N(\mu^{\eta^N} \in \mathcal{R}_\delta)} \mathbb{P}^N(\{\mu^{\eta^N} \notin B(\pi_{\mathcal{A}}, \varepsilon)\} \cap \{\mu^{\eta^N} \in \mathcal{R}_\delta\}). \end{aligned}$$

By the large deviation lower bound of Theorem 7.3 we have

$$\mathbb{P}^N(\mu^{\eta^N} \in \mathcal{R}_\delta) \geq \exp\{-N^\gamma(I(\pi_{\mathcal{A}}) + o(1))\},$$

where $\gamma = 3 - \alpha$.

Let $\mathcal{C}_{\varepsilon,\delta} = B(\pi_{\mathcal{A}}, \varepsilon)^c \cap \overline{\mathcal{R}_\delta}$. By the large deviation upper bound of Theorem 8.4 we have

$$\mathbb{P}^N(\mu^{\eta^N} \in \mathcal{C}_{\varepsilon,\delta}) \leq \exp\left\{-N^\gamma\left(\inf_{\mu \in \mathcal{C}_{\varepsilon,\delta}} I(\mu) + o(1)\right)\right\}.$$

Since $\pi_{\mathcal{A}}$ is the unique minimizer of energy on \mathcal{R} ([Bre89, RVV19]), given $\varepsilon > 0$ there exists $\beta = \beta(\varepsilon) > 0$ such that

$$\inf_{\mu \in B(\pi_{\mathcal{A}}, \varepsilon)^c \cap \mathcal{R}} I(\mu) \geq I(\pi_{\mathcal{A}}) + \beta. \tag{72}$$

Since $I(\cdot)$ is lower semi-continuous, by (72) we obtain (possibly after adjusting δ to replace $\overline{\mathcal{R}}_\delta$ with \mathcal{R}_δ) that for all sufficiently small δ we have

$$\inf_{\mu \in B(\pi_{\mathcal{A}}, \varepsilon)^c \cap \mathcal{R}_\delta} I(\mu) \geq I(\pi_{\mathcal{A}}) + \frac{\beta}{2}.$$

Altogether we obtain that for all sufficiently small δ (depending on ε only)

$$\mathbb{P}^N \left(\mu^{\eta^N} \notin B(\pi_{\mathcal{A}}, \varepsilon) \mid \mu^{\eta^N} \in \mathcal{R}_\delta \right) \leq e^{N\gamma(I(\pi_{\mathcal{A}}) + o(1))} e^{-N\gamma(I(\pi_{\mathcal{A}}) + \frac{\beta}{2} + o(1))} = e^{-N\gamma(\frac{\beta}{2} + o(1))}$$

and the right-hand side goes to 0 as $N \rightarrow \infty$ □

Proof of Theorem 1.2. Let \mathbb{P}^N denote the law of the interchange process with $\alpha = 1 + \kappa$. Let J be the number of all particle swaps in the process and let $M = \lfloor \frac{1}{2}N^\alpha(N - 1) \rfloor$.

Let \mathcal{S}_δ denote the δ -neighborhood of the reverse permuton in the Wasserstein distance on $\mathcal{M}([0, 1]^2)$. Observe that for sufficiently large N we have for any $k \leq M$

$$\mathbb{P}^N \left(\mu_{0,T}^{\eta^N} \in \mathcal{S}_\delta \mid J = k \right) \leq \mathbb{P}^N \left(\mu_{0,T}^{\eta^N} \in \mathcal{S}_\delta \mid J = M \right) (1 + o(1)). \tag{73}$$

This is because if the process has done k swaps up to time T_k and $\mu_{0,T_k}^{\eta^N} \in \mathcal{S}_\delta$, then with high probability $\mu_{0,T_M}^{\eta^N} \in \mathcal{S}_\delta$ as well, since \mathcal{S}_δ is an open set in $\mathcal{M}([0, 1]^2)$ and the additional number of steps done between T_k and T_M is $\leq \frac{1}{2}N^\alpha(N - 1) = o(N^3)$, so typically almost all particles have negligible displacement.

On the other hand, since in the interchange process each sequence of swaps of given length is equally likely, we have

$$\mathbb{P}^N \left(\mu_{0,T}^{\eta^N} \in \mathcal{S}_\delta \mid J = M \right) = \frac{|\mathcal{S}_{\kappa, \delta}^N|}{|\mathcal{P}_M^N|},$$

where \mathcal{P}_M^N is the set of all sequences of adjacent transpositions of length M . Summing (73) over $k \leq M$ we obtain

$$\mathbb{P}^N \left(\mu_{0,T}^{\eta^N} \in \mathcal{S}_\delta \cap \{J \leq M\} \right) \leq \mathbb{P}^N (J \leq M) \frac{|\mathcal{S}_{\kappa, \delta}^N|}{|\mathcal{P}_M^N|} (1 + o(1)).$$

Since under \mathbb{P}^N J has Poisson distribution with mean $\frac{1}{2}N^\alpha(N - 1)$, we have $\mathbb{P}^N(J \leq M) \rightarrow 1/2$ as $N \rightarrow \infty$.

To estimate the left-hand side, let $\tilde{\mathbb{P}}^N$ be the law of the biased interchange process corresponding to the sine curve process $\pi_{\mathcal{A}}$. Recall Lemma 7.1 and for fixed $\varepsilon > 0$

let A be the event that the $o(1)$ term in the formula for $\frac{d\mathbb{P}_\mu^N}{d\mathbb{P}^N}(T)$ is at most ε . Let us write

$$\mathbb{P}^N \left(\mu_{0,T}^{\eta^N} \in \mathcal{S}_\delta \cap \{J \leq M\} \cap A \right) \leq \mathbb{P}^N \left(\mu_{0,T}^{\eta^N} \in \mathcal{S}_\delta \cap \{J \leq M\} \right)$$

and

$$\begin{aligned} & \mathbb{P}^N \left(\mu_{0,T}^{\eta^N} \in \mathcal{S}_\delta \cap \{J \leq M\} \cap A \right) \\ &= \tilde{\mathbb{P}}^N \left(\mu_{0,T}^{\eta^N} \in \mathcal{S}_\delta \cap \{J \leq M\} \cap A \right) \frac{\mathbb{P}^N \left(\mu_{0,T}^{\eta^N} \in \mathcal{S}_\delta \cap \{J \leq M\} \cap A \right)}{\tilde{\mathbb{P}}^N \left(\mu_{0,T}^{\eta^N} \in \mathcal{S}_\delta \cap \{J \leq M\} \cap A \right)}. \end{aligned}$$

By Theorem 5.1 $\mu_{0,T}^{\eta^N} \in \mathcal{S}_\delta$ has high probability under $\tilde{\mathbb{P}}^N$ and, since the particle swap rates for the biased process sum up to $\frac{1}{2}N^\alpha(N-1)$ (recall (28)), we have similarly as for the unbiased process $\tilde{\mathbb{P}}^N(J \leq M) \rightarrow 1/2$ as $N \rightarrow \infty$. By Lemma 7.1 A is a high probability event under $\tilde{\mathbb{P}}^N$ as well.

To estimate the remaining probabilities, we employ the formula for the Radon–Nikodym derivative from Lemma 7.1. Since in the biased process with high probability the energy term in the derivative is close to $I(\pi_{\mathcal{A}}) = \frac{\pi^2}{6}$, we obtain

$$\frac{\mathbb{P}^N \left(\mu_{0,T}^{\eta^N} \in \mathcal{S}_\delta \cap \{J \leq M\} \cap A \right)}{\tilde{\mathbb{P}}^N \left(\mu_{0,T}^{\eta^N} \in \mathcal{S}_\delta \cap \{J \leq M\} \cap A \right)} \geq e^{-N^\gamma \left(\frac{\pi^2}{6} + \varepsilon \right) + o(N^\gamma)},$$

where $\gamma = 3 - \alpha$.

Altogether we obtain

$$|\mathcal{S}_{\kappa,\delta}^N| \geq |\mathcal{P}_M^N| e^{-N^\gamma \left(\frac{\pi^2}{6} + \varepsilon \right) + o(N^\gamma)}.$$

Since $|\mathcal{P}_M^N| = (N-1)^M = e^{\lfloor \frac{1}{2}N^\alpha(N-1) \rfloor \log(N-1)}$ and ε was arbitrary, we obtain the asymptotic lower bound on $|\mathcal{S}_{\kappa,\delta}^N|$ as claimed.

For the upper bound, let \mathcal{R}_δ be as in the previous theorem. By the large deviation upper bound of Theorem 8.4 we have

$$\mathbb{P}^N \left(\mu_{0,T}^{\eta^N} \in \mathcal{R}_\delta \right) \leq \exp \left\{ -N^\gamma \left(\inf_{\mu \in \mathcal{R}_\delta} I(\mu) + o(1) \right) \right\},$$

Since I is lower semi-continuous, given any $\varepsilon > 0$ we have for all sufficiently small $\delta > 0$

$$\inf_{\mu \in \mathcal{R}_\delta} I(\mu) \geq \inf_{\mu \in \mathcal{R}} I(\mu) - \varepsilon = I(\pi_{\mathcal{A}}) - \varepsilon,$$

where again we have used the energy minimization property of $\pi_{\mathcal{A}}$. Since $I(\pi_{\mathcal{A}}) = \frac{\pi^2}{6}$, this implies that for any $\varepsilon > 0$ and sufficiently small $\delta > 0$

$$\mathbb{P}^N \left(\mu^{\eta^N} \in \mathcal{R}_\delta \right) \leq e^{-N^\gamma (I(\pi_{\mathcal{A}}) - \varepsilon + o(1))}.$$

Now we estimate

$$\mathbb{P}^N \left(\mu^{\eta^N} \in \mathcal{R}_\delta \right) \geq \mathbb{P}^N \left(\mu^{\eta^N} \in \mathcal{R}_\delta \mid J = M \right) \mathbb{P}^N (J = M) = \frac{|\mathcal{S}_{\kappa, \delta}^N|}{|\mathcal{P}_M^N|} \mathbb{P}^N (J = M)$$

and use the same asymptotic estimate for $|\mathcal{P}_M^N|$ as in the lower bound. Since J is Poisson with mean $\frac{1}{2}N^\alpha(N-1)$ under \mathbb{P}^N , the second term on the right-hand side is $e^{O(\log N)}$. Altogether we obtain

$$|\mathcal{S}_{\kappa, \delta}^N| \leq \exp \left\{ \frac{1}{2}N^{1+\kappa}(N-1) \log(N-1) - N^{2-\kappa} (I(\pi_{\mathcal{A}}) - \varepsilon) + o(N^{2-\kappa}) \right\},$$

which proves the desired asymptotic upper bound on $|\mathcal{S}_{\kappa, \delta}^N|$. \square

Acknowledgements

We would like to thank Maxim Arnold, Duncan Dauvergne, Boris Khesin and Mustazee Rahman for interesting discussions. MK was supported by the National Science Centre, Poland, grant no. 2019/32/C/ST1/00525. BV was supported by the Canada Research Chair program and the NSERC Discovery Accelerator grant.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

- [AF09] L. AMBROSIO and A. FIGALLI. Geodesics in the space of measure-preserving maps and plans, *Arch. Ration. Mech. Anal.*, 194, 2 (2009), 421–462
- [AHRV07] O. ANGEL, A.E. HOLROYD, D. ROMIK, and B. VIRÁG. Random sorting networks, *Adv. Math.*, 215, 2 (2007), 839–868
- [Ald83] D. ALDOUS, *Random walks on finite groups and rapidly mixing Markov chains*, Séminaire de Probabilités XVII 1981/82 (Berlin, Heidelberg) (Jacques Azéma and Marc Yor, eds.). Springer Berlin Heidelberg, 1983, pp. 243–297
- [BFS09] M. BERNOT, A. FIGALLI, and F. SANTAMBROGIO. Generalized solutions for the Euler equations in one and two dimensions, *J. de Math. Pures et Appl.*, 91 (2009)

- [Bil13] P. BILLINGSLEY. *Convergence of probability measures*, Wiley Series in Probability and Statistics. Wiley (2013)
- [Bre89] Y. BRENIER. The least action principle and the related concept of generalized flows for incompressible perfect fluids, *J. Am. Math. Soc.*, 2, 2 (1989), 225–255
- [Bre99] Y. BRENIER. Minimal geodesics on groups of volume-preserving maps and generalized solutions of the Euler equations, *Commun. Pure Appl. Math.*, 52, 4 (1999), 411–452
- [Bre08] Y. BRENIER. Generalized solutions and hydrostatic approximation of the Euler equations, *Physica D*, 237 (2008), 1982–1988
- [CL55] A. CODDINGTON and N. LEVINSON, *Theory of ordinary differential equations*, International series in pure and applied mathematics. McGraw-Hill Companies, New York (1955)
- [Dau22] D. DAUVERGNE. The Archimedean limit of random sorting networks, *J. Am. Math. Soc.*, 35, 4 (2022), 1215–1267
- [FT04] J. FRITZ and B. TÓTH. Derivation of the Leroux system as the hydrodynamic limit of a two-component lattice gas, *Commun. Math. Phys.*, 249, 1 (2004), 1–27
- [HKM+13] C. HOPPEN, Y. KOHAYAKAWA, C.G. MOREIRA, B. RÁTH, and R.M. SAMPAIO. Limits of permutation sequences, *J. Combin. Theory Ser. B*, 103, 1 (2013), 93–113
- [Kal21] O. KALLENBERG, *Foundations of modern probability*, third ed., Probability Theory and Stochastic Modelling. Springer, Berlin (2021)
- [KL99] C. KIPNIS and C. LANDIM, *Scaling limits of interacting particle systems*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 320. Springer, Berlin (1999)
- [Lac16] H. LACOIN. Mixing time and cutoff for the adjacent transposition shuffle and the simple exclusion, *Ann. Probab.*, 44, 2 (2016), 1426–1487
- [Lov12] L. LOVÁSZ, *Large networks and graph limits*, American Mathematical Society Colloquium Publications, vol. 60. American Mathematical Society, Providence, RI (2012)
- [RV17] M. RAHMAN and B. VIRÁG, *Brownian motion as limit of the interchange process*. 1609.07745 (2017).
- [RVV19] M. RAHMAN, B. VIRÁG, and M. VIZER. Geometry of permutation limits, *Combinatorica*, 39, 4 (2019), 933–960
- [Shn87] A. SHNIRELMAN. On the geometry of the group of diffeomorphisms and the dynamics of an ideal incompressible fluid, *Math. Sbornik USSR*, 56 (1987), 79–105
- [Sta84] R.P. STANLEY. On the number of reduced decompositions of elements of Coxeter groups, *Eur. J. Combin.*, 5, 4 (1984), 359–372
- [Var16] S.R.S. VARADHAN, *Large deviations*, Courant Lecture Notes, Courant Institute of Mathematical Sciences (2016)

M. KOTOWSKI

Institute of Mathematics of the Polish Academy of Sciences,

ul. Śniadeckich 8, 00-656 Warsaw, Poland.

michal.kotowski@mimuw.edu.pl

B. VIRÁG

Department of Mathematics, University of Toronto, 40 St George St.,
Toronto, ON M5S 2E4, Canada.

balint@math.toronto.edu

Received: October 26, 2021

Accepted: October 7, 2022