Journal of Statistical Physics (2026) 193:51
https://doi.org/10.1007/s10955-026-03609-y

Hydrodynamic Limit and Large Deviations for Run-And-Tumble Particles with Mean-Field Switching Rates

Elena Pulvirenti[1] · Frank Redig[1] · Hidde van Wiechen[1]

Received: 25 September 2025 / Accepted: 2 April 2026

Abstract
In this paper, we study run-and-tumble particles moving on two copies of the discrete torus (referred to as layers), where the switching rate between layers depends on a mean-field interaction among the particles. We derive the hydrodynamic limit of this model, as well as the large deviations from the hydrodynamic limit. Our main tool is the introduction of a weakly perturbed version of the system, whose hydrodynamic equations precisely characterize the trajectories associated with large deviations.

1 Introduction

Run-and-tumble particles serve as simple models of active matter [1]. Systems of active particles constitute an important class of non-equilibrium systems where at the microscopic scale energy is dissipated in order to produce directed motion. They show a rich phenomenology such as clustering and long-range order [1–4].

In the mathematics literature on interacting particle systems, results on active particles are not very abundant (in contrast with the physics literature). To our knowledge, the first rigorous result on hydrodynamic limits for active particles is [5], where a system of locally interacting active particles was studied. In [5, 6], the authors prove the hydrodynamic limit (using the non-gradient method) and identify a motility-induced phase separation and a transition to collective motion from the equations obtained in the hydrodynamic limit. See also e.g. [7] for recent results in the physics literature on locally interacting active particles and collective effects therein.

In this paper we consider a simpler model where the interaction between the run-and-tumble particles is of mean-field type, i.e., via their empirical distribution. For this model we can then both prove the hydrodynamic limit and the large deviations from the hydrodynamic limit. In our model, the particles move on the one-dimensional torus, and have an internal state which takes values ± 1 and determines the direction of motion. The interaction between the particles arises implicitly via the flip rate at which particles flip their internal state which depends on the magnetization.

Communicated by Christian Maes.

✉ Hidde van Wiechen
hidde1996@live.nl

1 Delft University of Technology: Technische Universiteit Delft, Delft, Netherlands

Published online: 18 April 2026

First, we start by deriving the hydrodynamic limit, which is a coupled system of partial differential equations for the densities of particles with internal state ± 1, and an ordinary differential equation for the "magnetization". Second, we consider a weakly perturbed model where the influence of an external field which weakly depends on time and (microscopic) space is added. This field influences the rate at which particles flip their internal state and the rate at which they jump in the direction of their internal state. This is reminiscent of the weakly asymmetric exclusion process [8] which is an essential tool to study the large deviations for the trajectory of the density in the symmetric exclusion process (SEP). Contrary to the situation of the SEP, in our model, the perturbation does not act on the direction of the motion, which is always in the direction of the internal state.

We prove the hydrodynamic limit of this weakly perturbed model and use it to prove a large deviation principle for the trajectory of the densities in the original model. The technique of proof is based on a change of measure between the original and the weakly perturbed model and the associated exponential martingale (the Radon-Nikodym derivative of the perturbed model w.r.t. the original model). Due to the mean-field character of the interaction, no super-exponential replacement lemmas are needed, i.e., the quantities appearing in the Radon-Nikodym derivative between the perturbed and unperturbed model are a function of the empirical densities and the magnetization.

The rest of our paper is organized as follows. We start in Section 1.2 by an informal and non-technical description of the main results and techniques of proofs. In Section 2 we introduce the model, the weakly perturbed model and state the hydrodynamic limit of both. In Section 3 we prove large deviations for the trajectory of the densities. In Section 4 we provide the proof of the hydrodynamic limits stated in Section 2.

1.1 Motivation, Outline and Connection with Existing Literature

In this subsection we give an informal and high-level description of our results, the essence of the techniques which we use to prove the large deviation results, and the connection with existing literature. Regarding large deviation results for active particles, existing results in the physics literature are mostly based on macroscopic fluctuation theory, see e.g. [9, 10]. In the setting of models of active particles with local interaction such as in [5], to our knowledge, no mathematically rigorous large deviation results are available, and this seems (to us) a difficult problem. This is for us the main motivation to simplify the interaction between the particles and consider a mean field interaction, i.e., the rate at which the internal state is flipped is depending on the global magnetization, rather than on the local magnetization.

An important advantage of the mean-field interaction (as opposed to the local interaction) is the fact that in the proof of the large deviations one does not need the analog of the super-exponential replacement lemma from [8], which is difficult to obtain in the setting of locally interacting active particles.

To prove a large deviation result (which is done in Section 3), we have to compute the exponentially small probability of a trajectory deviating from the solution of the macroscopic PDE. It is a common strategy to introduce a modified dynamics, driven by a space and time dependent "potential" which makes the deviating trajectory "typical". This is reminiscent of the simple exponential change of measure in Cramer's theorem (cf. [11, 12] for an introduction to large deviation theory aimed at physicists). This "modified"dynamics is what we call the "weakly perturbed model" (reminiscent of the "weakly asymmetric exclusion process" in [8]) and it is introduced in Section 2.2 below. The large deviation rate function is then simply equal to the relative entropy of the modified dynamics (with the potential) w.r.t.

the original dynamics. This relative entropy can be computed explicitly because the Radon-Nikodym density of the modified dynamics w.r.t. original dynamics is an explicit exponential martingale (Girsanov's formula). A more technical aspect of the proof of the large deviations consists in proving that this perturbation of the dynamics via a space-time dependent potential exhaust the possible large deviation events, i.e., if there does not exist a corresponding field making the deviating trajectory typical, then the probability of the deviating trajectory is super-exponential, i.e., the corresponding rate function is infinite (this is important for the large deviation lower bound). Because we work with mean-field interactions, the exponential change of measure needed to make the deviating trajectory typical is automatically a function of the trajectory of the empirical distribution, which implies that no replacement lemmas have to be proved to close the macroscopic equation of the dynamics with a space-time dependent potential. On the other hand, to prove that the perturbation of the dynamics via a space-time dependent potential exhaust the possible large deviation events is more involved than in the setting of [8] where this is proved via Riesz representation theorem for Hilbert spaces, which in turn is due to the fact that in the setting of [8], the rate function has the (quadratic) structure of the square of a Hilbert space norm. This is no longer the case in our setting, and we have to use the Hahn-Banach theorem instead (see Lemma 3.8 below). Another ingredient of the proof of large deviations is to establish the hydrodynamic limit of the "modified dynamics" which we need for the large deviations as described above. First we have to prove well-posedness of the macroscopic PDE, which we do via a classical contraction argument, combined with the Banach fixed point theorem, to control the non-linear part of the macroscopic PDE (Section 4.1 below). Once this well-posedness is proved, the rest of the proof of the hydrodynamic limit of the modified dynamics follows a classical scheme (as explained e.g. in [13]) based on the Dynkin martingale ((4.17)), of which we show that in the hydrodynamic limit the predictable quadratic variation converges to zero. This then immediately implies the weak form of the macroscopic PDE in the hydrodynamic limit (cf. (4.14) for the rewriting of the PDE in the weak form).

2 Run-And-Tumble Particles with Mean-Field Switching Rates

In this section we describe the run-and-tumble particle model with mean-field switching rates. Later on we will also define a weakly perturbed version of this model which we will need for the large deviations. For both models, we will consider particles on the two-layered torus $V_N := \mathbb{T}_N \times S$, where $\mathbb{T}_N = \mathbb{Z}/N\mathbb{Z}$ is the discrete torus, and $S = \{-1, 1\}$ which we will call the internal state space. We then say that a particle $(x, \sigma) \in V_N$ has position $x \in \mathbb{T}_N$ and internal state $\sigma \in S$. The parameter N is a scaling parameter, and we will be interested in the limiting dynamics of macroscopic quantities (the empirical measure described below) when $N \to \infty$.

We will consider processes of particle configurations $\{\eta_t^N : t \geq 0\}$ on the state space $\Omega_N = \mathbb{N}^{V_N}$. We denote by $\eta_t^N(x, \sigma)$ the number of particles at site $(x, \sigma) \in V_N$ at time $t \geq 0$. The Markovian dynamics is then described as follows:

i) Active jump: with rate N a particle jumps from (x, σ) to $(x + \sigma, \sigma)$.
ii) Internal state flip: a particle jumps from (x, σ) to $(x, -\sigma)$, with a mean-field rate denoted by $c(\sigma, m_N(\eta_t^N))$

Here by "mean-field" we mean that the flip rate $c(\sigma, m_N(\eta_t^N))$ depends on the whole configuration η only via its magnetization $m_N(\eta_t^N)$. For $\eta \in \Omega_N$, this magnetization is defined

via

$$m_N(\eta) := \frac{1}{|\eta|} \sum_{x \in \mathbb{T}_N} \big(\eta(x,1) - \eta(x,-1)\big), \tag{2.1}$$

where $|\eta| := \sum_{(x,\sigma)\in V_N} \eta(x,\sigma)$ denotes the total number of particles in the configuration η. Here we also use the convention that if $|\eta| = 0$ then $m_N(\eta) = 0$. We assume $c(\sigma,\cdot)$ to be bounded from above and below (away from zero) and Lipschitz-continuous in the second variable, i.e., there exists an $L > 0$ such that for $\sigma \in \{-1,1\}$ and $m_1, m_2 \in [-1,1]$ we have that

$$|c(\sigma, m_1) - c(\sigma, m_2)| \le L|m_1 - m_2|. \tag{2.2}$$

The generator of the process is described above, working on functions $f : \Omega_N \to \mathbb{R}$, is defined as follows.

$$\begin{aligned}\mathscr{L}_N f(\eta) = N &\sum_{(x,\sigma)\in V_N} \eta(x,\sigma)\big[f(\eta^{(x,\sigma)\to(x+\sigma,\sigma)}) - f(\eta)\big] \\ &+ \sum_{(x,\sigma)\in V_N} c(\sigma, m_N(\eta))\eta(x,\sigma)\big[f(\eta^{(x,\sigma)\to(x,-\sigma)}) - f(\eta)\big],\end{aligned} \tag{2.3}$$

where $\eta^{(x,\sigma)\to(y,\sigma')}$ denotes the configuration η where a single particle has jumped from (x,σ) to (y,σ'), if possible.

Remark 2.1 If we choose the rates $c(\sigma, m) \equiv 1$, then the particles do not interact with each other, and we recover the run-and-tumble particle process, studied, for instance, in [1, 14–17]. An actual example of mean-field rates where particles do interact with one another is, for instance, given by the Curie-Weiss Glauber rates $c(\sigma, m) = e^{-\sigma\beta m}$ with $\beta > 0$.

2.1 Hydrodynamic Limit

For $N \in \mathbb{N}$, we define the empirical measure of a configuration $\eta \in \Omega_N$ by

$$\pi^N(\eta) := \frac{1}{N} \sum_{(x,\sigma)\in V_N} \eta(x,\sigma)\delta_{(\frac{x}{N},\sigma)}, \tag{2.4}$$

where δ denotes the Dirac measure. For a given η, $\pi^N(\eta)$ is a positive Radon measure on the macroscopic space $V := \mathbb{T} \times S$, where $\mathbb{T} = [0,1]$ is the torus. We denote the space of positive Radon measures on V by $\mathcal{M}_V$. Given $t \ge 0$, we further denote the empirical measure of the configuration η_t^N as

$$\pi_t^N := \pi^N(\eta_t^N) = \frac{1}{N} \sum_{(x,\sigma)\in V_N} \eta_t^N(x,\sigma)\delta_{(\frac{x}{N},\sigma)}, \tag{2.5}$$

For every $N \in \mathbb{N}$, this produces a process $\pi_\cdot^N := \{\pi_t^N : t \in [0,T]\}$ with trajectories in the Skorokhod space $D([0,T]; \mathcal{M}_V)$. The corresponding space of test functions is given by,

$$C^\infty(V) := \{\phi : V \to \mathbb{R} \mid \phi(\cdot,\sigma) \in C^\infty(\mathbb{T}) \text{ for all } \sigma \in S\}. \tag{2.6}$$

For $\phi \in C^\infty(V)$ we now denote the pairing

$$\left\langle \pi_t^N, \phi \right\rangle = \frac{1}{N} \sum_{(x,\sigma)\in V_N} \eta_t^N(x,\sigma)\phi(\tfrac{x}{N},\sigma). \tag{2.7}$$

For $N \in \mathbb{N}$ and smooth $\varrho(x,\sigma) : V \to \mathbb{R}_{\geq 0}$ we define the product Poisson measures

$$\mu_N^{\varrho} = \bigotimes_{(x,\sigma)\in V_N} \mathrm{Pois}(\varrho\left(\tfrac{x}{N},\sigma\right)), \tag{2.8}$$

which is the local equilibrium measure for the non-interacting ($c(\sigma,m) \equiv 1$) case (see [14]). We assume that the process $\{\eta_t^N : t \geq 0\}$ starts at $t = 0$ from the configuration $\eta^N = \eta_0^N$ distributed according to μ_N^{ϱ}, which fixes the initial density profile. More precisely, π_0^N converges as $N \to \infty$ to the positive measure with density $\varrho(x,\sigma) : V \to \mathbb{R}_{\geq 0}$, i.e., we have that for every $\phi \in C^\infty(V)$ and $\varepsilon > 0$

$$\lim_{N\to\infty} \mathbb{P}_N^{\varrho}\left(\left|\langle \pi_0^N, \phi\rangle - \sum_{\sigma\in S}\int \phi(x,\sigma)\varrho(x,\sigma)\mathrm{d}x\right| > \varepsilon\right) = 0. \tag{2.9}$$

Here $\mathbb{P}_N^{\varrho}$ denotes the path-space measure of the process with initial distribution μ_N^{ϱ}, where we omit the dependence on the terminal time T since we assume that this is fixed.

The question of the hydrodynamic limit is to find the limiting PDE for π_t^N as $N \to \infty$. We start with the following preliminary computation

$$\begin{aligned}\mathscr{L}_N\langle \pi^N(\eta),\phi\rangle &= N\sum_{(x,\sigma)\in V_N}\eta(x,\sigma)\left(\left\langle \pi^N\left(\eta^{(x,\sigma)\to(x+\sigma,\sigma)}\right),\phi\right\rangle - \langle \pi^N(\eta),\phi\rangle\right)\\ &\quad+\sum_{(x,\sigma)\in V_N}\eta(x,\sigma)c(\sigma,m_N(\eta))\left(\left\langle \pi^N\left(\eta^{(x,\sigma)\to(x,-\sigma)}\right),\phi\right\rangle - \langle \pi^N(\eta),\phi\rangle\right)\\ &= \frac{1}{N}\sum_{(x,\sigma)\in V_N}\eta(x,\sigma)\left[N\left(\phi(\tfrac{x+\sigma}{N},\sigma)-\phi(\tfrac{x}{N},\sigma)\right)\right.\\ &\quad\left.+c(\sigma,m_N(\eta))\left(\phi(x,-\sigma)-\phi(x,\sigma)\right)\right].\end{aligned} \tag{2.10}$$

From this computation, we observe that the evolution of the empirical measure depends on the evolution of the magnetization $m_N(\eta_t^N)$. Note however that the magnetization, defined in (2.1), can be expressed in terms of the empirical measure as follows,

$$m_N(\eta) = m(\pi^N(\eta)) := \frac{\langle \pi^N(\eta), \mathbb{1}_{\sigma=1} - \mathbb{1}_{\sigma=-1}\rangle}{\langle \pi^N(\eta), 1\rangle}. \tag{2.11}$$

This representation of the magnetization motivates the following definition of the magnetization corresponding to a density $\varrho(x,\sigma) : V \to \mathbb{R}_{\geq 0}$ as

$$m(\varrho) := \frac{\langle \varrho, \mathbb{1}_{\sigma=1} - \mathbb{1}_{\sigma=-1}\rangle_{L^2(V)}}{\langle \varrho, 1\rangle_{L^2(V)}}, \tag{2.12}$$

where $\langle\cdot,\cdot\rangle_{L^2(V)}$ denotes the inner product of L^2-functions on the space V, given by

$$\langle \psi,\phi\rangle_{L^2(V)} = \sum_{\sigma\in S}\int_{\mathbb{T}}\psi(x,\sigma)\phi(x,\sigma)\mathrm{d}x, \tag{2.13}$$

for any $\psi,\phi \in L^2(V)$.

Remark 2.2 Occasionally, with slight abuse of notation, we make the dependence on x and σ explicit inside the inner product $\langle\cdot,\cdot\rangle_{L^2(V)}$ or the dual pairing $\langle\cdot,\cdot\rangle$ to make clear how a function depends on these two variables. This means that sometimes we will write

$\langle \psi, \phi(x,\sigma)\rangle_{L^2(V)}$ while still intending that the function $\phi(x,\sigma)$ is integrated over both x and σ.

Theorem 2.1 *Under $\mathbb{P}_N^{\varrho}$, the trajectory $\pi_\cdot^N$ converges weakly in the Skorokhod topology to α, where $\alpha \in D([0,T];\mathcal{M}_V)$ is the deterministic trajectory of measures with at time $t \in [0,T]$ density $\varrho_t(x,\sigma)$ solving the following partial differential equation,*

$$\dot{\varrho}_t(x,\sigma) = -\sigma\partial_x\varrho_t(x,\sigma) + c(-\sigma, m(\varrho_t))\varrho_t(x,-\sigma) - c(\sigma, m(\varrho_t))\varrho_t(x,\sigma), \tag{2.14}$$

with initial condition $\varrho_0(x,\sigma) = \varrho(x,\sigma)$.

This theorem will be a consequence of a more general hydrodynamic limit of a weakly perturbed modification of the model, which we will introduce in Section 2.2. Apart from the hydrodynamic limit of the empirical measure, we can also find a limiting equation of the magnetization.

Corollary 2.2 *Under $\mathbb{P}_N^{\varrho}$, the trajectory of the magnetization $m(\pi_\cdot^N)$ converges weakly in the Skorokhod topology to $m(\varrho_\cdot)$, where ϱ_t solves (2.14). Moreover, $m_t := m(\varrho_t)$ solves the following ordinary differential equation,*

$$\dot{m}_t = c(-1, m_t)\cdot(1-m_t) - c(1,m_t)\cdot(1+m_t), \tag{2.15}$$

with initial condition $m_0 = m(\varrho)$.

Remark 2.3 When considering the Curie-Weiss Glauber rates $c(\sigma,m) = e^{-\sigma\beta m}$ with inverse temperature $\beta > 0$, the evolution of the magnetization is given by

$$\dot{m}_t = 2\sinh(\beta m_t) - 2m_t\cosh(\beta m_t). \tag{2.16}$$

As $t \to \infty$ the process m_t converges to a solution of the mean-field equation

$$m^* = \tanh(\beta m^*). \tag{2.17}$$

For $\beta \le 1$ the only solution is $m^* = 0$. However, for $\beta > 1$, there exist two nonzero solutions, and if $m_0 \neq 0$, the process m_t converges to either the positive or negative solution, depending on the initial value m_0. This implies flocking behavior of the particles, i.e., particles tend to move in the same direction, similar to the behavior of models with local interaction such as the Vicsek model [2, 4].

2.2 Weakly Perturbed Model

We furthermore introduce a weakly perturbed version of our model, which will be a key tool for the study of large deviations since it produces the deviating trajectories which have a finite rate function. The weak perturbation will be parametrized by a time-dependent potential $H : [0,T]\times V \to \mathbb{R}$, which we will assume to be differentiable in time and continuous in space, and which we will denote by $H \in C^{1,0}([0,T]\times V)$. The time-dependent generator of this model, acting on functions $f : \Omega_N \to \mathbb{R}$, is given as follows:

$$\begin{aligned}\mathscr{L}_{N,t}^H f(\eta) = N &\sum_{(x,\sigma)\in V_N} \eta(x,\sigma) e^{H_t(\frac{x+\sigma}{N},\sigma) - H_t(\frac{x}{N},\sigma)}\big[f(\eta^{(x,\sigma)\to(x+\sigma,\sigma)}) - f(\eta)\big] \\ &+ \sum_{(x,\sigma)\in V_N} \eta(x,\sigma)c(\sigma, m_N(\eta)) e^{H_t(\frac{x}{N},-\sigma) - H_t(\frac{x}{N},\sigma)}\big[f(\eta^{(x,\sigma)\to(x,-\sigma)}) - f(\eta)\big].\end{aligned} \tag{2.18}$$

Note that for $H = 0$ we recover the original model. We will denote by $\mathbb{P}_N^{\varrho,H}$ the path-space measure of this process, where η_0^N is distributed as μ_N^{ϱ} (cf. (2.8)). We further abbreviate $\widetilde{H}_t(x) := H_t(x,1) - H_t(x,-1)$. The hydrodynamic limit of this process is then given in the following theorem.

Theorem 2.3 *Under $\mathbb{P}_N^{\varrho,H}$, the trajectory $\pi_\cdot^N$ converges weakly in the Skorokhod topology to α^H, where $\alpha^H \in D([0,T];\mathcal{M}_V)$ is the deterministic trajectory of measures with at time $t \in [0,T]$ density $\varrho_t^H(x,\sigma)$ which solves the following partial differential equation,*

$$\begin{aligned}\dot{\varrho}_t^H(x,\sigma) = &-\sigma\partial_x\varrho_t^H(x,\sigma) + c(-\sigma, m(\varrho_t^H))e^{\sigma\widetilde{H}_t(x)}\varrho_t^H(x,-\sigma)\\ &-c(\sigma, m(\varrho_t^H))e^{-\sigma\widetilde{H}_t(x)}\varrho_t^H(x,\sigma),\end{aligned} \tag{2.19}$$

with initial condition $\varrho_0^H(x,\sigma) = \varrho(x,\sigma)$.

Note that Theorem 2.1 follows from this Theorem 2.3 by choosing $H \equiv 0$. The proof of Theorem 2.3 will be postponed to Section 4. Below we give the evolution of the magnetization under the perturbed dynamics.

Corollary 2.4 *Under $\mathbb{P}_N^{\varrho,H}$, the trajectory of the magnetization $m(\pi_\cdot^N)$ converges weakly in the Skorokhod topology to $m(\varrho_\cdot^H)$, where ϱ_t^H solves (2.19). Moreover, $m_t^H := m(\varrho_t^H)$ solves the following equation,*

$$\dot{m}_t^H = \frac{1}{\langle\varrho,1\rangle_{L^2(V)}}\left\langle\varrho_t^H, -2\sigma e^{-\sigma\widetilde{H}_t(x)}c(\sigma, m_t^H)\right\rangle_{L^2(V)}. \tag{2.20}$$

with initial condition $m_0 = m(\varrho)$.

Proof Note that if $\alpha_N \to \alpha$ in $D([0,T]:\mathcal{M}_V)$ then $m(\alpha_N) \to m(\alpha)$ in $D([0,T]:[-1,1])$, hence $m(\pi_\cdot^N) \xrightarrow{d} m_\cdot$ under $\mathbb{P}_N^{\varrho,H}$. Therefore, we can find an equation for the evolution of the magnetization from the evolution of ϱ_t^H. First note that $\langle\varrho_t^H,1\rangle_{L^2(V)} = \langle\varrho,1\rangle_{L^2(V)}$ for all $t \geq 0$, which is due to the conservation of particles. Therefore, using the definition (2.12), we find

$$\dot{m}_t = \frac{\langle\dot{\varrho}_t^H, \mathbb{1}_{\sigma=1} - \mathbb{1}_{\sigma=-1}\rangle_{L^2(V)}}{\langle\varrho,1\rangle_{L^2(V)}} \tag{2.21}$$

By the periodic boundary conditions, we have that

$$\langle -\sigma\partial_x\varrho_t^H, \mathbb{1}_{\sigma=1} - \mathbb{1}_{\sigma=-1}\rangle_{L^2(V)} = 0, \tag{2.22}$$

hence (2.20) follows by filling in (2.19) into (2.21).

□

Corollary 2.2 follows from Corollary 2.4 by choosing $H \equiv 0$ and using that for any density ϱ we have that

$$\frac{\int_{\mathbb{T}}\varrho(x,\sigma)\mathrm{d}x}{\langle\varrho,1\rangle_{L^2(V)}} = \tfrac{1}{2}(1+\sigma m(\varrho)). \tag{2.23}$$

Note that, unlike in the unperturbed model, under the perturbed dynamics the evolution of the magnetization is not a closed equation, but depends on the density ϱ_t^H. This is because under the these dynamics the magnetization process is no longer a Markov process on its own, and additional information on the positions of the particles is required.

3 Large Deviations

In this section we will prove a large deviation principle for the run-and-tumble particle process with mean-field switching rates. We start by defining the rate function $\mathcal{I}^{\varrho}$: $D([0,T];\mathcal{M}_V) \to [0,\infty]$, which is given in two parts

$$\mathcal{I}^{\varrho}(\widehat{\alpha}) = h_0^{\varrho}(\widehat{\alpha}_0) + \mathcal{I}_{tr}(\widehat{\alpha}) \tag{3.1}$$

Here $h_0^{\varrho}(\widehat{\alpha}_0)$ is the static part of the large deviation rate function, only depending on the measure at time $t=0$. It can be informally written as

$$h_0^{\varrho}(\widehat{\alpha}_0) = \mu_N^{\varrho}(\pi_0^N \approx \widehat{\alpha}_0), \tag{3.2}$$

i.e, it corresponds to the large deviation principle of the initial density profile π_0^N under the starting distribution μ_N^{ϱ}. Since μ_N^{ϱ} is given by a product Poisson measure, the corresponding rate function is known, and given by

$$h_0^{\varrho}(\widehat{\alpha}_0) = \sup_{\phi} h_0^{\varrho}(\widehat{\alpha}_0;\phi), \qquad h_0^{\varrho}(\widehat{\alpha}_0;\phi) = \langle \widehat{\alpha}_0, \phi\rangle - \langle \varrho, e^{\phi}-1\rangle_{L^2(V)}. \tag{3.3}$$

Here the supremum is taken over all $\phi \in C^{\infty}(V)$.

The term $\mathcal{I}_{tr}(\widehat{\alpha})$ in (3.1) is the dynamic part of the rate function, and depends on the whole trajectory $\widehat{\alpha}$. It is given by the following:

$$\mathcal{I}_{tr}(\widehat{\alpha}) = \sup_{G} \mathcal{I}_{tr}(\widehat{\alpha};G), \qquad \mathcal{I}_{tr}(\widehat{\alpha};G) = \ell(\widehat{\alpha};G) - \int_0^T \left\langle \widehat{\alpha}_t, c(\sigma, m(\widehat{\alpha}_t))\left(e^{-\sigma \widetilde{G}_t(x)}-1\right)\right\rangle \mathrm{d}t. \tag{3.4}$$

Here the supremum is taken over all $G \in C^{\infty}([0,T]\times V)$, and we recall the notation $\widetilde{G}_t(x) = G_t(x,1) - G_t(x,-1)$. Furthermore, $\ell(\widehat{\alpha};G)$ is a linear functional, defined as follows

$$\ell(\widehat{\alpha};G) := \langle \widehat{\alpha}_T, G_T\rangle - \langle \widehat{\alpha}_0, G_0\rangle - \int_0^T \langle \widehat{\alpha}_t, (\partial_t + \sigma\partial_x)G_t(x,\sigma)\rangle \mathrm{d}t, \tag{3.5}$$

and $m(\widehat{\alpha}_t)$ is the magnetization of the measure, defined as

$$m(\widehat{\alpha}_t) = \frac{\langle \widehat{\alpha}_t, \mathbb{1}_{\sigma=1} - \mathbb{1}_{\sigma=-1}\rangle}{\langle \widehat{\alpha}_t, 1\rangle}. \tag{3.6}$$

Remark 3.1 While the rate function might look implicit because there is a supremum appearing in (3.4), we obtain an explicit formula of the dynamic part of the rate function in terms of the deviating trajectory $\widehat{\alpha}$. Indeed in Lemma 3.8 we identify a function H (cf. (3.64)) for which $\mathcal{I}_{tr}(\widehat{\alpha}) = \mathcal{I}_{tr}(\widehat{\alpha};H)$.

In order to prove the large deviation principle with rate function $\mathcal{I}$, we need to establish the upper and lower bound.

i. **Upper bound**: For every closed set $\mathcal{C} \subset D([0,T];\mathcal{M}_V)$ we have

$$\overline{\lim_{N\to\infty}} \frac{1}{N} \log \mathbb{P}_N^{\varrho}(\pi_\cdot^N \in \mathcal{C}) \le -\inf_{\widehat{\alpha}\in\mathcal{C}} \mathcal{I}^{\varrho}(\widehat{\alpha}) \tag{3.7}$$

ii. **Lower bound**: For every open set $\mathcal{O} \subset D([0,T];\mathcal{M}_V)$ we have

$$\underline{\lim_{N\to\infty}} \frac{1}{N} \log \mathbb{P}_N^{\varrho}(\pi_\cdot^N \in \mathcal{O}) \ge -\inf_{\widehat{\alpha}\in\mathcal{O}} \mathcal{I}^{\varrho}(\widehat{\alpha}) \tag{3.8}$$

The essence of the proof (which is the content of the rest of the section) can be summarized in the follwoing three steps.

1. For the large deviation upper bound we use an exponential martingale which is the Radon-Nikodym derivative or the path-space measure of the weakly perturbed dynamics w.r.t. the original dynamics. This leads immediately to the upper bound for compact sets.
2. We prove exponential tightness, which then extends the upper bound to closed sets. For this we have to slightly modify a standard approach from [18], because we do not know the invariant measure of the perturbed model. We therefore use a uniform estimate of the path space measure of the modified interacting model w.r.t. to the path space measure of a simple non-interacting model of which we know the invariant measures (this is Lemma 3.5 below). Then we can proceed along the lines of the standard approach (Lemma 3.6).
3. For the proof of large deviation lower bound, the essential ingredient is Lemma 3.8 below where we show that if the rate function of a deviating trajectory is finite, then there exists a weakly perturbed model (i.e., a potential H) such that the trajectory is "typical" for this H. We first prove this for deviating trajectories of which the densities remain strictly positive. For general trajectories, we proceed by approximation (Lemma 3.9).

3.1 Radon-Nikodym Derivatives

The goal of this section is to obtain an explicit form for the Radon-Nikodym derivative of the path-space measure of the weakly perturbed process $\mathrm{d}\mathbb{P}_N^{\varrho,H}$ with respect to the path-space measure of the original process $\mathrm{d}\mathbb{P}_N^{\varrho}$. This Radon-Nikodym derivative is given by the so-called exponential martingale of the process, and is given in the following lemma.

Lemma 3.1 *For all $T > 0$, $N \in \mathbb{N}$ and $H \in C^\infty(V)$, we have that*

$$\begin{aligned}\frac{\mathrm{d}\mathbb{P}_N^{\varrho,H}}{\mathrm{d}\mathbb{P}_N^{\varrho}} &= \mathcal{Z}_{N,T}^H(\pi_\cdot^N)\\ &:= \exp\left(N\langle\pi_T^N, H_T\rangle - N\langle\pi_0^N, H_0\rangle - \int_0^T e^{-N\langle\pi_t^N,H_t\rangle}(\partial_t + \mathscr{L}_N)e^{N\langle\pi_t^N,H_t\rangle}\mathrm{d}t\right).\end{aligned} \tag{3.9}$$

Proof By Palmowski and Rolski [19], the exponential martingale $\mathcal{Z}_{N,T}^H(\pi_\cdot^N)$ is equal to the Radon-Nikodym derivative $\frac{\mathrm{d}\widetilde{\mathbb{P}}_N}{\mathrm{d}\mathbb{P}_N^{\varrho}}$, where $\widetilde{\mathbb{P}}_N$ is the path-space measure (up to time T) of the process corresponding to the time-dependent Markov generator on $\widehat{\varrho}_N$ given by

$$\widetilde{\mathscr{L}}_{N,t}f(\eta) = e^{-N\langle\pi^N(\eta),H_t\rangle}\left[\mathscr{L}_N\big(f(\eta)\cdot e^{N\langle\pi^N(\eta),H_t\rangle}\big) - f(\eta)\cdot\mathscr{L}_N e^{N\langle\pi^N(\eta),H_t\rangle}\right]. \tag{3.10}$$

where π^N now denotes the empirical measure as a function of $\eta \in \Omega_N$, as defined in (2.4). We will prove that $\widetilde{\mathbb{P}}_N = \mathbb{P}_N^{\varrho,H}$ by showing that $\widetilde{\mathscr{L}}_{N,t}$ is equal to the generator of the weakly

perturbed process $\mathscr{L}^H_{N,t}$ as defined in (2.18). We compute, using (2.3)

$$\begin{aligned} e^{-N\langle \pi^N(\eta),H_t\rangle}\mathscr{L}_N\big(f(\eta)\cdot e^{N\langle \pi^N(\eta),H_t\rangle}\big) &= N\sum_{(x,\sigma)\in V_N}\eta(x,\sigma)\big[f(\eta^{(x,\sigma)\to(x+\sigma,\sigma)})e^{H_t(\frac{x+\sigma}{N},\sigma)-H_t(\frac{x}{N},\sigma)}-f(\eta)\big] \\ &\quad+\sum_{(x,\sigma)\in V_N}c(\sigma,m(\eta))\eta(x,\sigma)\big[f(\eta^{(x,\sigma)\to(x,-\sigma)})e^{H_t(\frac{x}{N},-\sigma)-H_t(\frac{x}{N},\sigma)}-f(\eta)\big], \end{aligned} \tag{3.11}$$

where we used that

$$\langle \pi_N(\eta^{(x,\sigma)\to(y,\sigma')}),H_t\rangle-\langle \pi_N(\eta),H_t\rangle = H_t(\tfrac{y}{N},\sigma')-H_t(\tfrac{x}{N},\sigma). \tag{3.12}$$

Similarly we compute

$$\begin{aligned} e^{-N\langle \pi^N(\eta),H_t\rangle}\mathscr{L}_N e^{N\langle \pi^N(\eta),H_t\rangle}(\eta) &= N\sum_{(x,\sigma)\in V_N}\eta(x,\sigma)\big[e^{H_t(\frac{x+\sigma}{N},\sigma)-H_t(\frac{x}{N},\sigma)}-1\big] \\ &\quad+\sum_{(x,\sigma)\in V_N}c(\sigma,m(\eta))\eta(x,\sigma)\big[e^{H_t(\frac{x}{N},-\sigma)-H_t(\frac{x}{N},\sigma)}-1\big]. \end{aligned} \tag{3.13}$$

Substituting (3.11) and (3.13) into (3.10), we indeed find that $\widetilde{\mathscr{L}}_{N,t}f=\mathscr{L}^H_{N,t}f$ for functions $f:\Omega_N\to\mathbb{R}$, completing the proof.

□

Corollary 3.2

$$\frac{1}{N}\log\Big(\mathcal{Z}^H_{N,T}(\pi^N_\cdot)\Big)=\mathcal{I}_{tr}(\pi^N_\cdot;H)+\mathcal{O}(\tfrac{1}{N}). \tag{3.14}$$

Proof Making use of the following approximation

$$e^{H_t(\frac{x+\sigma}{N},\sigma)-H_t(\frac{x}{N},\sigma)}-1=\frac{\sigma}{N}\partial_x H_t(\tfrac{x}{N},\sigma)+\mathcal{O}(\tfrac{1}{N^2}), \tag{3.15}$$

we are able to write (3.13) as

$$e^{-N\langle \pi^N_t,H_t\rangle}\mathscr{L}_N e^{N\langle \pi^N_t,H_t\rangle}=N\langle \pi^N_t,\sigma\partial_x H_t\rangle+N\Big\langle \pi^N_t,c(\sigma,m(\pi^N_t))\Big(e^{-\sigma\widetilde{H}_t(x)}-1\Big)\Big\rangle+\mathcal{O}(1). \tag{3.16}$$

By now plugging this into (3.9), we find that

$$\begin{aligned} \mathcal{Z}^H_{N,T}(\pi^N_\cdot) &= \exp\Big(N\langle \pi^N_T,H_T\rangle-N\langle \pi^N_0,H_0\rangle-N\int_0^T\langle \pi^N_t,(\partial_t+\sigma\partial_x)H_t\rangle\mathrm{d}t+\mathcal{O}(1)\Big) \\ &\quad\times\exp\Big(-N\int_0^T\Big\langle \pi^N_t,c(\sigma,m(\pi^N_t))\Big(e^{-\sigma\widetilde{H}_t(x)}-1\Big)\Big\rangle\mathrm{d}t\Big) \\ &= \exp\Big(N\ell(\pi^N_\cdot;H)-N\int_0^T\Big\langle \pi^N_t,c(\sigma,m(\pi^N_t))\Big(e^{-\sigma\widetilde{H}_t(x)}-1\Big)\Big\rangle\mathrm{d}t+\mathcal{O}(1)\Big), \end{aligned} \tag{3.17}$$

finishing the proof.

□

3.2 Upper Bound

In this section we will prove the large deviation upper bound (3.7). A crucial ingredient is to prove that the path-space measures of $\pi_\cdot^N$ are exponentially tight, which then reduces the proof of (3.7) to compact sets.

Theorem 3.3 (Exponential Tightness) *For any $n \in \mathbb{N}$ there exists a compact set $\mathcal{K}_n \subset D([0,T],\mathcal{M}_V)$ such that*

$$\overline{\lim_{N\to\infty}} \frac{1}{N} \log \mathbb{P}_N^{\varrho}(\pi_\cdot^N \notin \mathcal{K}_n) = -n. \tag{3.18}$$

Before proving exponential tightness, we first give the proof of the upper bound for compact sets.

Theorem 3.4 (Upper bound for compact sets) *For every compact set $\mathcal{K} \subset D([0,T];\mathcal{M}_V)$ we have that*

$$\overline{\lim_{N\to\infty}} \frac{1}{N} \log \mathbb{P}_N^{\varrho}(\pi_\cdot^N \in \mathcal{K}) \le -\inf_{\widehat{\alpha}\in\mathcal{K}} \mathcal{I}^{\varrho}(\widehat{\alpha}). \tag{3.19}$$

Proof We start with the following computation

$$\begin{aligned}\frac{1}{N}\log \mathbb{P}_N^{\varrho}(\pi_\cdot^N \in \mathcal{K}) &= \frac{1}{N}\log \mathbb{E}_N^{\varrho}\left[\mathbb{1}_{\pi_\cdot^N\in\mathcal{K}} \cdot \frac{e^{Nh_0^{\varrho}(\pi_0^N;\phi)}}{e^{Nh_0^{\varrho}(\pi_0^N;\phi)}} \cdot \frac{\mathcal{Z}_{N,T}^G(\pi_\cdot^N)}{\mathcal{Z}_{N,T}^G(\pi_\cdot^N)}\right] \\ &\le -\inf_{\widehat{\alpha}\in\mathcal{K}} \frac{1}{N}\log\left[e^{Nh_0^{\varrho}(\widehat{\alpha}_0;\phi)} \cdot \mathcal{Z}_{N,T}^G(\widehat{\alpha})\right] \\ &\quad + \frac{1}{N}\log \mathbb{E}_N^{\varrho}\left[e^{Nh_0^{\varrho}(\pi_0^N;\phi)} \cdot \mathcal{Z}_{N,T}^G(\pi_\cdot^N)\right],\end{aligned} \tag{3.20}$$

where we recall the definition of h_0^{ϱ} in (3.3) and of $\mathcal{Z}_{N,T}^G$ in (3.9). Since $\mathcal{Z}_{N,T}^G(\pi_\cdot^N)$ is a martingale with $\mathcal{Z}_{N,0}^G(\pi_\cdot^N) = 1$, we actually find that

$$\mathbb{E}_N^{\varrho}\left[e^{Nh_0^{\varrho}(\pi_0^N;\phi)} \cdot \mathcal{Z}_{N,T}^G(\pi_\cdot^N)\right] = \mathbb{E}_{\mu_N^{\varrho}}\left[e^{Nh_0^{\varrho}(\pi_0^N;\phi)}\right] = \mathbb{E}_{\mu_N^{\varrho}}\left[e^{N\left(\langle\pi_0^N,\phi\rangle - \langle\varrho, e^{\phi}-1\rangle_{L^2(V)}\right)}\right] = 1, \tag{3.21}$$

where we used that μ_N^{ϱ} is a product Poisson distribution. Therefore the second term in (3.20) vanishes. For the first term, note that we took ϕ and G arbitrarily, so we have

$$\overline{\lim_{N\to\infty}} \frac{1}{N}\log \mathbb{P}_N^{\varrho}(\pi_\cdot^N \in \mathcal{K}) \le -\sup_{\phi,G} \overline{\lim_{N\to\infty}} \inf_{\widehat{\alpha}\in\mathcal{K}} \frac{1}{N}\log\left[e^{Nh_0^{\varrho}(\widehat{\alpha}_0;\phi)} \cdot \mathcal{Z}_{N,T}^G(\widehat{\alpha})\right] \tag{3.22}$$

using Corollary 3.2, we find that

$$\begin{aligned}-\sup_{\phi,G} \overline{\lim_{N\to\infty}} \inf_{\widehat{\alpha}\in\mathcal{K}} \frac{1}{N}\log\left[e^{Nh_0^{\varrho}(\widehat{\alpha}_0;\phi)} \cdot \mathcal{Z}_{N,T}^G(\widehat{\alpha})\right] &= -\sup_{\phi,G}\inf_{\widehat{\alpha}\in\mathcal{K}} h_0^{\varrho}(\widehat{\alpha}_0;\phi) + \mathcal{I}_{tr}(\widehat{\alpha};G) \\ &= -\inf_{\widehat{\alpha}\in\mathcal{K}} \mathcal{I}^{\varrho}(\widehat{\alpha}),\end{aligned} \tag{3.23}$$

where we were able to interchange the supremum over ϕ and G with the infimum over $\widehat{\alpha}$ using the argument of Lemma 11.3 in [20], using that $\mathcal{K}$ is compact.

□

In order to prove the exponential tightness, we want to use the method used in [18, pages 271-273]. However, since we do not know the invariant measure of this system, we first turn to a perturbed model of which we do know the invariant measures, namely the independent run-and-tumble particle system, as defined in [15] (without diffusive jumps). The generator of this process corresponds to setting $c(\sigma, m) \equiv 1$ in the generator $\mathscr{L}_N$ defined in (2.3), and we will denote it by $\mathscr{L}_N^{\mathrm{RTP}}$. For this process, we know that the Product Poisson measures with constant density ϱ_c are invariant. We denote the path-space measure of this process, started from μ_N^{ϱ}, by $\mathbb{P}_N^{\mathrm{RTP},\varrho}$.

Lemma 3.5 *There exists a constant $C > 0$ such that*

$$\mathbb{E}_{\mathbb{P}_N^{\mathrm{RTP},\varrho}}\left[\left(\frac{\mathrm{d}\mathbb{P}_N^{\varrho}}{\mathrm{d}\mathbb{P}_N^{\mathrm{RTP},\varrho_c}}\right)^2\right] \leq e^{CN}. \tag{3.24}$$

Proof Note that we can split the Radon-Nikodym derivative in the following way

$$\frac{\mathrm{d}\mathbb{P}_N^{\varrho}}{\mathrm{d}\mathbb{P}_N^{\mathrm{RTP},\varrho_c}} = \frac{\mathrm{d}\mu_N^{\varrho}}{\mathrm{d}\mu_N^{\varrho_c}} \cdot \frac{\mathrm{d}\mathbb{P}_N^{\varrho}}{\mathrm{d}\mathbb{P}_N^{\mathrm{RTP},\varrho}}. \tag{3.25}$$

The Radon-Nikodym derivative of the Poisson distributions can be bounded in the following way

$$\begin{aligned}\frac{\mathrm{d}\mu_N^{\varrho}}{\mathrm{d}\mu_N^{\varrho_c}}(\eta) &= \exp\left(\sum_{(x,\sigma)\in V_N} \eta(x,\sigma)\log\left(\frac{\varrho(\frac{x}{N},\sigma)}{\varrho_c}\right) - \sum_{(x,\sigma)\in V_N}(\varrho(\tfrac{x}{N},\sigma)-\varrho_c)\right)\\ &\leq \exp\left(\log\left(\frac{||\varrho||_\infty}{\varrho_c}\right)|\eta^N| + 2\varrho_c N\right).\end{aligned} \tag{3.26}$$

Since the jump rates of the processes corresponding to $\mathbb{P}_N^{\mathrm{RTP},\varrho}$ and $\mathbb{P}_N^{\varrho}$ only differ at the internal state jumps, by the Girsanov formula, we find that the Radon-Nikodym derivative is given by

$$\begin{aligned}\frac{\mathrm{d}\mathbb{P}_N^{\varrho}}{\mathrm{d}\mathbb{P}_N^{\mathrm{RTP},\varrho}} = \exp\Bigg(&\sum_{(x,\sigma)\in V_N}\int_0^T \log\big(c(\sigma, m_N(t))\big)\mathrm{d}J_t^{(x,\sigma)\to(x,-\sigma)}\\ &- \sum_{(x,\sigma)\in V_N}\int_0^T \eta_t^N(x,\sigma)\big(c(\sigma,m_N(t))-1\big)\mathrm{d}t\Bigg),\end{aligned} \tag{3.27}$$

where $J_t^{(x,\sigma)\to(x,-\sigma)}$ is the number of jumps made from (x,σ) to $(x,-\sigma)$ up to time t. Since $c(\sigma, m_N(t))$ is bounded from above and below, we can find constants $c_1, c_2 > 0$ such that

$$\left(\frac{\mathrm{d}\mathbb{P}_N^{\varrho}}{\mathrm{d}\mathbb{P}_N^{\mathrm{RTP},\varrho}}\right)^2 \leq \exp\left(c_1\sum_{(x,\sigma)\in V_N} J_T^{(x,\sigma)\to(x,-\sigma)} + c_2 T|\eta^N|\right), \tag{3.28}$$

where we recall that $|\eta^N|$ is the total number of particles in the configuration η^N. Note that $\sum_{(x,\sigma)\in V_N} J_T^{(x,\sigma)\to(x,-\sigma)}$ is the total number of internal state jumps up to time T, which under $\mathbb{P}_N^{\mathrm{RTP},\varrho}$ is a Poisson process with intensity $|\eta^N|$. Combining (3.26) and (3.28), we can

find a constant $c_3 > 0$ such that

$$\mathbb{E}_{\mathbb{P}_N^{\mathrm{RTP},\varrho}}\left[\left(\frac{\mathrm{d}\mathbb{P}_N^{\varrho}}{\mathrm{d}\mathbb{P}_N^{\mathrm{RTP},\varrho_c}}\right)^2\right] \leq e^{4\varrho_c N}\mathbb{E}_{\mathbb{P}_N^{\mathrm{RTP},\varrho}}\left[\exp\left(c_3 T|\eta^N|\right)\right]$$
$$\leq \exp\left(4\varrho_c N + 2||\varrho||_\infty\left(e^{c_3 T}-1\right)N\right), \tag{3.29}$$

hence (3.24) holds.

□

Now we can proceed with the proof of exponential tightness, using the approach from [18]. We will need the following result.

Lemma 3.6 *For every $\varepsilon > 0$ and $G \in C^\infty(V)$*

$$\lim_{\delta\to 0}\lim_{N\to\infty}\frac{1}{N}\log\mathbb{P}_N^{\varrho}\left(\sup_{|s-t|<\delta}\left|\left\langle\pi_t^N, G\right\rangle - \left\langle\pi_s^N, G\right\rangle\right| \geq \varepsilon\right) = -\infty. \tag{3.30}$$

Proof Let $\varepsilon > 0$ be given. First note that we have the following

$$\mathbb{P}_N^{\varrho}\left(\sup_{|s-t|<\delta}\left|\left\langle\pi_t^N, G\right\rangle - \left\langle\pi_s^N, G\right\rangle\right| \geq \varepsilon\right) \leq \mathbb{P}_N^{\varrho}\left(\sup_{|s-t|<\delta}\left\langle\pi_t^N, G\right\rangle - \left\langle\pi_s^N, G\right\rangle \geq \varepsilon\right)$$
$$+ \mathbb{P}_N^{\varrho}\left(\sup_{|s-t|<\delta}\left\langle\pi_t^N, -G\right\rangle - \left\langle\pi_s^N, -G\right\rangle \geq \varepsilon\right). \tag{3.31}$$

Since we are considering every $G \in C^\infty(V)$, we can neglect the absolute value in (3.30). Furthermore, by Hölder's inequality we have that for a general event A

$$\mathbb{P}_N^{\varrho}(A) = \mathbb{E}_{\mathbb{P}_N^{\mathrm{RTP},\varrho_c}}\left[\mathbb{1}_A\frac{\mathrm{d}\mathbb{P}_N^{\varrho}}{\mathrm{d}\mathbb{P}_N^{\mathrm{RTP},\varrho_c}}\right] \leq \left(\mathbb{P}_N^{\mathrm{RTP},\varrho_c}(A)\right)^{\frac{1}{2}}\left(\mathbb{E}_{\mathbb{P}_N^{\mathrm{RTP},\varrho}}\left[\left(\frac{\mathrm{d}\mathbb{P}_N^{\varrho}}{\mathrm{d}\mathbb{P}_N^{\mathrm{RTP},\varrho_c}}\right)^2\right]\right)^{\frac{1}{2}}. \tag{3.32}$$

Therefore, by Lemma 3.5, it is enough to prove the result for $\mathbb{P}_N^{\mathrm{RTP},\varrho_c}$.

Using the following inclusion.

$$\left\{\sup_{|s-t|<\delta}\langle\pi_t^N, G\rangle - \langle\pi_s^N, G\rangle \geq \varepsilon\right\} \subset \bigcup_{k=0}^{[T\delta^{-1}]}\left\{\sup_{0\leq t<\delta}\langle\pi_{k\delta+t}^N, G\rangle - \langle\pi_{k\delta}^N, G\rangle > \frac{\varepsilon}{4}\right\}, \tag{3.33}$$

we are able to find that

$$\overline{\lim_{N\to\infty}}\frac{1}{N}\log\mathbb{P}_N^{\mathrm{RTP},\varrho_c}\left(\sup_{|s-t|<\delta}\langle\pi_t^N, G\rangle - \langle\pi_s^N, G\rangle \geq \varepsilon\right)$$
$$\leq \overline{\lim_{N\to\infty}}\frac{1}{N}\log\max_{k=0}^{[T\delta^{-1}]}\mathbb{P}_N^{\mathrm{RTP},\varrho_c}\left(\sup_{0\leq t<\delta}\langle\pi_{k\delta+t}^N, G\rangle - \langle\pi_{k\delta}^N, G\rangle \geq \tfrac{1}{4}\varepsilon\right)$$
$$= \overline{\lim_{N\to\infty}}\frac{1}{N}\log\mathbb{P}_N^{\mathrm{RTP},\varrho_c}\left(\sup_{0\leq t<\delta}\langle\pi_t^N, G\rangle - \langle\pi_0^N, G\rangle \geq \tfrac{1}{4}\varepsilon\right), \tag{3.34}$$

where in the last step we used that $\mu_N^{\varrho_c}$ is invariant for the RTP system. By denoting the exponential martingale corresponding to $\mathscr{L}_N^{\mathrm{RTP}}$ as $\mathcal{Z}_{N,t}^{\mathrm{RTP},G}(\pi_\cdot^N)$ (recall the definition of the exponential martingale in (3.9)), and multiplying both sides by a constant $\lambda > 0$, we find that

$$\begin{aligned}
&\mathbb{P}_N^{\mathrm{RTP},\varrho_c}\left(\sup_{0\le t<\delta}\langle\pi_t^N,\lambda G\rangle-\langle\pi_0^N,\lambda G\rangle\ge\tfrac{1}{4}\lambda\varepsilon\right)\\
&\quad\le\mathbb{P}_N^{\mathrm{RTP},\varrho_c}\left(\sup_{0\le t<\delta}\left\{\frac{1}{N}\log\mathcal{Z}_{N,t}^{\mathrm{RTP},\lambda G}(\pi_\cdot^N)\right.\right.\\
&\quad\left.\left.+\frac{1}{N}\int_0^t e^{-N\langle\pi_s^N,\lambda G_s\rangle}(\partial_s+\mathscr{L}_N^{\mathrm{RTP}})e^{N\langle\pi_s^N,\lambda G_s\rangle}\mathrm{d}s\right\}\ge\tfrac{1}{4}\lambda\varepsilon\right)\\
&\quad\le\mathbb{P}_N^{\mathrm{RTP},\varrho_c}\left(\sup_{0\le t<\delta}\mathcal{Z}_{N,t}^{\mathrm{RTP},\lambda G}(\pi_\cdot^N)\ge e^{\frac{1}{8}N\lambda\varepsilon}\right)\\
&\qquad+\mathbb{P}_N^{\mathrm{RTP},\varrho_c}\left(\sup_{0\le t<\delta}\frac{1}{N}\int_0^t e^{-N\langle\pi_s^N,\lambda G_s\rangle}(\partial_s+\mathscr{L}_N^{\mathrm{RTP}})e^{N\langle\pi_s^N,\lambda G_s\rangle}\mathrm{d}s\ge\tfrac{1}{8}\lambda\varepsilon\right).
\end{aligned}\tag{3.35}$$

Recalling that $\mathcal{Z}_{N,t}^{\mathrm{RTP},\lambda G}(\pi_\cdot^N)$ is a non-negative martingale, by Doob's martingale inequality we can upper bound the first part by

$$\mathbb{P}_N^{\mathrm{RTP},\varrho_c}\left(\sup_{0\le t<\delta}\mathcal{Z}_{N,t}^{\mathrm{RTP},\lambda G}(\pi_\cdot^N)\ge e^{\frac{1}{8}N\lambda\varepsilon}\right)\le\mathbb{E}_N^{\mathrm{RTP},\varrho_c}\left[\mathcal{Z}_{N,\delta}^{\mathrm{RTP},\lambda G}\right]\cdot e^{-\frac{1}{8}N\lambda\varepsilon}=e^{-\frac{1}{8}N\lambda\varepsilon}.\tag{3.36}$$

For the second part, using (3.16), we are able to find the following upper bound for the integrand

$$\frac{1}{N}e^{-N\langle\pi_s^N,\lambda G_s\rangle}(\partial_s+\mathscr{L}_N^{\mathrm{RTP}})e^{N\langle\pi_s^N,\lambda G_s\rangle}\le\frac{1}{N}|\eta^N|\left(\lambda||\partial_x G||_\infty+M\left(e^{\lambda||\widetilde{G}||_\infty}-1\right)\right).\tag{3.37}$$

Therefore, using that $|\eta^N|$ is Poisson distributed with parameter $N\varrho_c$ under $\mathbb{P}_N^{\mathrm{RTP},\varrho_c}$, we find that by the Markov inequality,

$$\mathbb{P}_N^{\mathrm{RTP},\varrho_c}\left(\sup_{0\le t<\delta}\frac{1}{N}\int_0^t e^{-N\langle\pi_s^N,\lambda G_s\rangle}(\partial_s+\mathscr{L}_N^{\mathrm{RTP}})e^{N\langle\pi_s^N,\lambda G_s\rangle}\mathrm{d}s\ge\tfrac{1}{8}\lambda\varepsilon\right)=\mathcal{O}(\delta).\tag{3.38}$$

Combining (3.36) and (3.38), we find that

$$\lim_{\delta\to0}\lim_{N\to\infty}\frac{1}{N}\log\mathbb{P}_N^{\mathrm{RTP},\varrho_c}\left(\sup_{|s-t|<\delta}\left|\left\langle\pi_t^N,G\right\rangle-\left\langle\pi_s^N,G\right\rangle\right|\ge\varepsilon\right)\le-\frac{1}{8}\lambda\varepsilon,\tag{3.39}$$

and since we can take λ arbitrarily large, this concludes the proof.

□

We are now ready to give a proof of the exponential tightness.

Proof of Theorem 3.3 We start by defining the following set

$$\mathcal{E}_K=\left\{\widehat{\alpha}\in D([0,T];\mathcal{M}_V):\sup_{t\in[0,T]}\widehat{\alpha}_t(V)\le K\right\}.\tag{3.40}$$

For this set, by the Chernoff inequality, we find that

$$\mathbb{P}_N^{\varrho}(\pi_\cdot^N \notin \mathcal{E}_K) = \mathbb{P}_N^{\varrho}\left(\sup_{t\in[0,T]} \pi_t^N(V) > K\right) \leq e^{-NK}\mathbb{E}_N^{\varrho}\left[\exp\left(\sup_{t\in[0,T]} N\pi_t^N(V)\right)\right], \tag{3.41}$$

where the expectation can be upper bounded in the following way

$$\mathbb{E}_N^{\varrho}\left[\exp\left(\sup_{t\in[0,T]} N\pi_t^N(V)\right)\right] = \mathbb{E}_N^{\varrho}\left[\exp\left(|\eta^N|\right)\right] \leq e^{N||\varrho||_\infty(e-1)}. \tag{3.42}$$

Combining (3.41) and (3.42), we can find a sequence of numbers $(K_n)_{n\in\mathbb{N}}$ such that for every $n \in \mathbb{N}$

$$\overline{\lim_{N\to\infty}} \frac{1}{N} \log \mathbb{P}_N^{\varrho}(\pi_\cdot^N \notin \mathcal{E}_{K_n}) \leq -n \tag{3.43}$$

Next we consider a countable uniformly dense family $\{\phi_j\}_{j\in\mathbb{N}} \subset C^\infty(V)$ and define for each $\delta > 0$ and $\varepsilon > 0$ the following set

$$\mathcal{C}_{j,\delta,\varepsilon} = \left\{\widehat{\alpha} \in D([0,T];\mathcal{M}_V) : \sup_{|t-s|<\delta} \left|\langle\widehat{\alpha}_t, \phi_j\rangle - \langle\widehat{\alpha}_s, \phi_j\rangle\right| \leq \varepsilon\right\}. \tag{3.44}$$

Note that for any choice of the parameters the set $\mathcal{C}_{j,\delta,\varepsilon}$ is closed, and by Lemma 3.6 there exists a $\delta = \delta(j, m, n)$ such that

$$\mathbb{P}_N^{\varrho}(\pi_\cdot^N \notin \mathcal{C}_{j,\delta,1/m}) \leq \exp(-Nnmj) \tag{3.45}$$

for large enough N, and so

$$\mathbb{P}_N^{\varrho}\left(\pi_\cdot^N \notin \bigcap_{j\geq 1, m\geq 1} \mathcal{C}_{j,\delta(j,m,n),1/m}\right) \leq \sum_{j\geq 1, m\geq 1} \exp(-Nnmj) \leq C\exp(-Nn), \tag{3.46}$$

for some constant $C > 0$. By now considering the set

$$\mathcal{K}_n = \mathcal{E}_K \cap \bigcap_{j\geq 1, m\geq 1} \mathcal{C}_{j,\delta(j,m,n),1/m}, \tag{3.47}$$

it follows that (3.18) holds for this choice of $\mathcal{K}_n$. Since we furthermore know that it is closed we only need to show that it is relatively compact, which can be done by proving the following two things [18, Proposition 4.1.2]:

1. $\{\widehat{\alpha}_t : \widehat{\alpha} \in \mathcal{K}_n, t \in [0,T]\}$ is relatively compact in $\mathcal{M}_V$.
2. $\lim_{\delta\to 0} \sup_{\widehat{\alpha}\in\mathcal{K}_n} w_\delta(\widehat{\alpha}) = 0$, where

$$w_\delta(\widehat{\alpha}) := \sup_{|t-s|\leq\delta} \sum_{k=1}^{\infty} \frac{1}{2^k}\left(1 \wedge |\langle\widehat{\alpha}_t, \phi_j\rangle - \langle\widehat{\alpha}_s, \phi_j\rangle|\right) = 0. \tag{3.48}$$

Here item 1. is satisfied since $\mathcal{K}_n \subset \mathcal{E}_{K_n}$ and closed balls are compact in $\mathcal{M}_V$, and 2. follows from the definition of the sets $C_{j,\delta,\varepsilon}$.

□

3.3 Lower Bound

In this section we will prove the large deviation lower bound, as given in (3.8). The main idea is to show that if $\mathcal{I}^{\varrho}(\widehat{\alpha}) < \infty$, then there exists a function H such that

$$\mathcal{I}^{\varrho}(\widehat{\alpha}) = \lim_{N\to\infty} \frac{1}{N}\mathbb{E}_N^{\widehat{\alpha}_0}\left[\frac{\mathrm{d}\mathbb{P}_N^{\widehat{\alpha}_0,H}}{\mathrm{d}\mathbb{P}_N^{\varrho}}\right]. \tag{3.49}$$

To achieve this, we have to show two things. First we have to show that if $h_0^{\varrho}(\widehat{\alpha}_0) < \infty$, then $\widehat{\alpha}_0$ has a density $\widehat{\varrho}_0$, and h_0^{ϱ} can be written as the relative entropy of product Poisson distributions with the respective densities ϱ and $\widehat{\varrho}_0$. After that, we have to show that if $\mathcal{I}_{tr}(\widehat{\alpha}) < \infty$, then there exists a measurable $H : [0, T] \times V \to \mathbb{R}$ such that $\widehat{\alpha}$ satisfies the hydrodynamic equation of the weakly perturbed model, given in (2.19), and that H is then a function for which the supremum in the definition of $\mathcal{I}_{tr}$ in (3.4) is attained.

The first step follows from the following Lemma

Lemma 3.7 *If $h_0^{\varrho}(\widehat{\alpha}_0) < \infty$, then $\widehat{\alpha}_0$ has a density $\widehat{\varrho}_0 : V \to \mathbb{R}$, and*

$$h_0^{\varrho}(\widehat{\alpha}_0) = \lim_{N\to\infty} \frac{1}{N}\mathbb{E}_{\mu_N^{\widehat{\varrho}_0}}\left[\log \frac{\mathrm{d}\mu_N^{\widehat{\varrho}_0}}{\mathrm{d}\mu_N^{\varrho}}\right]. \tag{3.50}$$

Proof Assume $h_0^{\varrho}(\widehat{\alpha}_0) < \infty$ If $\widehat{\alpha}_0$ is not absolutely continuous, then there exists a Borel set $A \subset V$ such that $\widehat{\alpha}_0(A) > 0$ and $\lambda_V(A) = 0$, with λ_V the Lebesgue measure on V. By the definition of $h_0^{\varrho}(\widehat{\alpha}_0)$ in (3.3), we have that for $\phi \in C^{\infty}(V)$

$$\langle \widehat{\alpha}_0, \phi\rangle \le h_0^{\varrho}(\widehat{\alpha}_0) + \langle \varrho, e^{\phi} - 1\rangle_{L^2(V)}. \tag{3.51}$$

For $n \in \mathbb{N}$ now take a sequence $(\phi_k^{(n)})_{k\in\mathbb{N}} \subset C^{\infty}(V)$ such that $\phi_k^{(n)} \to n\mathbb{1}_A$ pointwise as $k \to \infty$. It then follows that

$$\langle \widehat{\alpha}_0, \phi_k^{(n)}\rangle \to n\widehat{\alpha}_0(A), \qquad \langle \varrho, e^{\phi_k^{(n)}} - 1\rangle_{L^2(V)} \to 0, \tag{3.52}$$

as $k \to \infty$. By taking n large enough this contradicts (3.51), hence we can conclude that $\widehat{\alpha}_0$ has a density $\widehat{\varrho}_0$.

The rest of the proof of (3.50) follows then from calculating the supremum.

$$\begin{aligned} h_0^{\varrho}(\widehat{\alpha}_0) &= \sup_{\phi}\left\{\langle \widehat{\varrho}_0, \phi\rangle_{L^2(V)} - \langle \varrho, e^{\phi} - 1\rangle_{L^2(V)}\right\} \\ &= \langle \widehat{\varrho}_0, \log(\tfrac{\widehat{\varrho}_0}{\varrho})\rangle_{L^2(V)} - \langle \widehat{\varrho}_0 - \varrho, 1\rangle_{L^2(V)} \\ &= \lim_{N\to\infty} \frac{1}{N}\mathbb{E}_{\mu_N^{\widehat{\varrho}_0}}\left[\log \frac{\mathrm{d}\mu_N^{\widehat{\varrho}_0}}{\mathrm{d}\mu_N^{\varrho}}\right]. \end{aligned} \tag{3.53}$$

where the supremum is attained for $\phi = \log(\frac{\widehat{\varrho}_0}{\varrho})$.

□

By a similar argument as in the previous lemma, we can show that if $\mathcal{I}_{tr}(\widehat{\alpha}) < \infty$ then there exists a density $\widehat{\varrho} : [0, T] \times V \to \mathbb{R}$ for the whole trajectory $\widehat{\alpha}$. In the following Lemma we prove an alternative formula for the dynamic part of the rate function in the case that $\widehat{\varrho}_t(v) > 0$ for all $t \in [0, T]$, $v \in V$.

Lemma 3.8 *If $\mathcal{I}_{tr}(\widehat{\alpha}) < \infty$ and $\widehat{\varrho} > 0$ then there exists a bounded measurable function H such that $\widehat{\alpha}$ satisfies the equation (2.19) in the weak sense. Furthermore, $\mathcal{I}_{tr}(\widehat{\alpha}) = \mathcal{I}_{tr}(\widehat{\alpha}; H)$ and*

$$\mathcal{I}_{tr}(\widehat{\alpha}) = \int_0^T \left\langle \widehat{\alpha}_t, \left(e^{-\sigma \widetilde{H}_t(x)}(-\sigma \widetilde{H}_t(x) - 1) + 1\right) c(\sigma, m(\widehat{\alpha}_t))\right\rangle \mathrm{d}t. \tag{3.54}$$

Proof By the definition of $\mathcal{I}_{tr}$ in (3.4), we have the following

$$\sup_{\substack{G \in C^\infty([0,T]\times V)\\ ||G||_\infty \le 1}} \left\{ \ell(\widehat{\alpha}; G) - \int_0^T \left\langle \widehat{\alpha}_t, c(\sigma, m(\widehat{\alpha}_t)) \left(e^{-\sigma \widetilde{G}_t(x)} - 1\right)\right\rangle \mathrm{d}t \right\} \le \mathcal{I}_{tr}(\widehat{\alpha}), \tag{3.55}$$

and so

$$\sup_{\substack{G \in C^\infty([0,T]\times V)\\ ||G||_\infty \le 1}} \ell(\widehat{\alpha}; G) \le \mathcal{I}_{tr}(\widehat{\alpha}) + \int_0^T \langle \widehat{\alpha}_t, c(\sigma, m(\widehat{\alpha}_t))\,(e-1)\rangle\, \mathrm{d}t < \infty. \tag{3.56}$$

Consequently, by the Hahn-Banach theorem, we can extend the linear functional $\ell(\widehat{\alpha}; \cdot)$ to a bounded linear functional in $C([0, T] \times V)$. Therefore, by the Riesz representation theorem, there exists a signed measure $\nu \in \mathcal{M}_{[0,T]\times V}$ such that

$$\ell(\widehat{\alpha}; G) = \langle \nu, G\rangle := \int_{[0,T]\times V} G d\nu. \tag{3.57}$$

Again, since we assume that $\mathcal{I}_{tr}(\widehat{\alpha}) < \infty$, this measure ν has a density $g : [0, T] \times V \to \mathbb{R}$. By then plugging in the definition of $\ell(\widehat{\alpha}; G)$ in (3.5), we find that $\widehat{\alpha}$ satisfies

$$\langle \widehat{\alpha}_T, G_T\rangle - \langle \widehat{\alpha}_0, G_0\rangle - \int_0^T \langle \widehat{\alpha}_t, (\partial_t + \sigma\partial_x) G_t\rangle \mathrm{d}t = \int_0^T \langle g_t, G_t\rangle_{L^2(V)} \mathrm{d}t, \tag{3.58}$$

i.e., it satisfies the following PDE in the weak sense

$$\dot{\widehat{\alpha}}_t(x, \sigma) = -\sigma\partial_x \widehat{\alpha}_t(x, \sigma) + g_t(x, \sigma). \tag{3.59}$$

We can now split up $g_t(x, \sigma) = \sigma f_t(x) + h_t(x)$, and we will show that $h \equiv 0$ almost everywhere. To see this, note that

$$\begin{aligned}
\mathcal{I}_{tr}(\widehat{\alpha}) &= \sup_G \left\{ \int_0^T \langle \sigma f_t, G_t\rangle_{L^2(V)} \mathrm{d}t + \int_0^T \langle h_t, G_t\rangle_{L^2(V)} \mathrm{d}t \right.\\
&\quad \left. - \int_0^T \left\langle \widehat{\alpha}_t, c(\sigma, m(\widehat{\alpha}_t)) \left(e^{-\sigma \widetilde{G}_t(x)} - 1\right)\right\rangle \mathrm{d}t \right\}\\
&= \sup_{\widetilde{G}, \overline{G}} \left\{ \int_0^T \langle f_t, \widetilde{G}_t\rangle_{L^2(\mathbb{T})} \mathrm{d}t + \int_0^T \langle h_t, \overline{G}_t\rangle_{L^2(\mathbb{T})} \mathrm{d}t \right.\\
&\quad \left. - \int_0^T \left\langle \widehat{\alpha}_t, c(\sigma, m(\widehat{\alpha}_t)) \left(e^{-\sigma \widetilde{G}_t(x)} - 1\right)\right\rangle \mathrm{d}t \right\},
\end{aligned} \tag{3.60}$$

where $\overline{G}_t(x) = G_t(x, 1) + G_t(x, -1)$. By considering functions where $G_t(x, 1) = G_t(x, -1)$, it follows that

$$\mathcal{I}_{tr}(\widehat{\alpha}) \ge \sup_{\overline{G}} \int_0^T \langle h_t, \overline{G}_t\rangle_{L^2(\mathbb{T})} \mathrm{d}t, \tag{3.61}$$

where if $h \not\equiv 0$ almost everywhere this supremum is infinite, which contradicts $\mathcal{I}^{\varrho}(\widehat{\alpha}) < \infty$. Therefore,

$$\dot{\widehat{\alpha}}_t(x,\sigma) = -\sigma\,\partial_x\widehat{\alpha}_t(x,\sigma) + \sigma\, f_t(x) \tag{3.62}$$

holds weakly, with $f : [0,T] \times \mathbb{T} \to \mathbb{R}$ some bounded measurable function.

Now we show that (3.62) implies that (2.19) is satisfied for some H fixed by f_t in (3.62). In order to find this H, we denote $\Psi_t(x,\sigma) := e^{-\sigma \widetilde{H}_t(x)}$ and we have to solve the following equation for Ψ_t,

$$\sigma\, f_t(x) = \frac{1}{\Psi_t(x,\sigma)} c(-\sigma, m(\widehat{\alpha}_t))\widehat{\varrho}_t(x,-\sigma) - \Psi_t(x,\sigma)c(\sigma, m(\widehat{\alpha}_t))\widehat{\varrho}_t(x,\sigma) \tag{3.63}$$

This is a quadratic equation in Ψ_t, with the following positive and bounded solution

$$\Psi_t(x,\sigma) = \frac{-\sigma\, f_t(x) + \sqrt{f_t(x)^2 + 4\widehat{\varrho}_t(x,\sigma)c(\sigma, m(\widehat{\alpha}_t))\widehat{\varrho}_t(x,-\sigma)c(-\sigma, m(\widehat{\alpha}_t))}}{2\widehat{\varrho}_t(x,\sigma)c(\sigma, m(\widehat{\alpha}_t))}. \tag{3.64}$$

It is a straightforward calculation to show that $\Psi(x,\sigma)\cdot\Psi(x,-\sigma) = 1$, and so (2.19) holds for $\widetilde{H}_t(x) = -\log(\Psi(x,1))$.

Now, assuming that $\widehat{\alpha}$ satisfies (2.19) we find that

$$\begin{aligned}
\mathcal{I}_{tr}(\widehat{\alpha}) &= \sup_G \left\{ \ell(\widehat{\alpha}; G) - \int_0^T \langle \widehat{\alpha}_t, \left(e^{-\sigma \widetilde{G}_t(x)} - 1\right) c(\sigma, m(\widehat{\alpha}_t))\rangle \mathrm{d}t \right\} \\
&= \sup_G \left\{ \int_0^T \left\langle \widehat{\alpha}_t, \left(-e^{-\sigma \widetilde{H}_t(x)}\sigma \widetilde{G}_t(x) - e^{-\sigma \widetilde{G}_t(x)} + 1\right) c(\sigma, m(\widehat{\alpha}_t))\right\rangle \mathrm{d}t \right\} \\
&= \int_0^T \left\langle \widehat{\alpha}_t, \sup_{p\in\mathbb{R}} \left(-e^{-\sigma \widetilde{H}_t(x)}\sigma\, p - e^{-\sigma\, p} + 1\right) c(\sigma, m(\widehat{\alpha}_t))\right\rangle \mathrm{d}t,
\end{aligned} \tag{3.65}$$

where we can interchange the supremum in the last integral by dominated convergence. This supremum is attained for $p = \widetilde{H}_t(x)$, indeed showing that $\mathcal{I}_{tr}(\widehat{\alpha}) = \mathcal{I}_{tr}(\widehat{\alpha}; H)$. After filling this in, it follows that (3.54) holds. □

We now have a clear formulation of the rate function when $\widehat{\alpha}$ has positive density, however if the density can be zero then the formulation in (3.64) is not well-defined. Furthermore, in order for the hydrodynamic limit of the weakly perturbed model to hold in Theorem 2.1, we need to assume that $H \in C^{1,0}([0,T]\times V)$. We therefore define the following space

$$\begin{aligned}
&D_0([0,T]; \mathcal{M}_V) \\
&\quad := \left\{ \widehat{\alpha} \in D([0,T]; \mathcal{M}_V) : \widehat{\varrho} > 0,\ \widehat{\alpha} \text{ satisfies (2.19) with } H \in C^{1,0}([0,T]\times V) \right\}
\end{aligned} \tag{3.66}$$

We will now show that the rate function of trajectories outside this set can be approximated by the rate function of trajectories within this set.

Lemma 3.9 *Let $\widehat{\alpha} \in D([0,T]; \mathcal{M}_V)$ such $\mathcal{I}_{tr}(\widehat{\alpha}) < \infty$, then there exists a sequence $(\widehat{\alpha}_k)_{k\in\mathbb{N}} \subset D_0([0,T]; \mathcal{M}_V)$ such that $\widehat{\alpha}_k \to \widehat{\alpha}$ weakly and*

$$\mathcal{I}_{tr}(\widehat{\alpha}) = \lim_{k\to\infty} \mathcal{I}_{tr}(\widehat{\alpha}_k). \tag{3.67}$$

Proof We first show that $\widehat{\alpha} \in D([0,T]; \mathcal{M}_V)$ can be approximated by trajectories with positive density. We define the following measure for any $\varepsilon > 0$

$$\widehat{\alpha}_\varepsilon = (1-\varepsilon)\widehat{\alpha} + \varepsilon 1 \tag{3.68}$$

where 1 on the right-hand side denotes the measure with constant density equal to 1. It follows that $\widehat{\alpha}_\varepsilon$ has positive density and that $\widehat{\alpha}_\varepsilon \to \widehat{\alpha}$ weakly as $\varepsilon \to 0$. Therefore, by convexity and lower semicontinuity of the rate function $\mathcal{I}^\varrho$, we then find that

$$\overline{\lim_{\varepsilon\to 0}}\, \mathcal{I}_{tr}(\widehat{\alpha}_\varepsilon) \le \mathcal{I}_{tr}(\widehat{\alpha}) \le \underline{\lim_{\varepsilon\to 0}}\, \mathcal{I}_{tr}(\widehat{\alpha}_\varepsilon). \tag{3.69}$$

Hence we have indeed found a good approximation.

Now assume that $\widehat{\alpha} \in D([0,T];\mathcal{M}_V)$ has density $\widehat{\varrho} > 0$ and that $\mathcal{I}_{tr}(\widehat{\alpha}) < \infty$. By Lemma 3.8, there exists a bounded measurable $H : [0,T]\times V \to \mathbb{R}$ such that (2.19) holds weakly. Now find a sequence $\widehat{\alpha}_k$ with densities $\widehat{\varrho}_k \in C^{2,1}([0,T]\times V)$ such that $\widehat{\varrho}_k \to \widehat{\varrho}$ pointwise as $k \to \infty$. It then follows that $\widehat{\alpha}_k \to \widehat{\alpha}$ weakly and, by (3.62) and (3.64), each $\widehat{\alpha}_k$ satisfies (2.19) for some function $H_k \in C^{1,0}([0,T]\times V)$ where $H_k \to H$ pointwise. By the formulation of $\mathcal{I}_{tr}(\widehat{\alpha})$ in (3.54), we can then indeed conclude that (3.67) holds. □

Theorem 3.10 (Lower bound) *Given a $\widehat{\alpha} \in D([0,T];\mathcal{M}_V)$, for every neighborhood $\mathcal{O} \subset D([0,T];\mathcal{M}_V)$ of $\widehat{\alpha}$ we have that*

$$\underline{\lim_{N\to\infty}}\, \frac{1}{N}\log \mathbb{P}_N^\varrho(\pi_\cdot^N \in \mathcal{O}) \ge -\mathcal{I}^\varrho(\widehat{\alpha}). \tag{3.70}$$

As a consequence, for every open set $\mathcal{O} \subset D([0,T];\mathcal{M}_V)$ we have that

$$\underline{\lim_{N\to\infty}}\, \frac{1}{N}\log \mathbb{P}_N^\varrho(\pi_\cdot^N \in \mathcal{O}) \ge -\inf_{\widehat{\alpha}\in\mathcal{O}} \mathcal{I}^\varrho(\widehat{\alpha}). \tag{3.71}$$

Proof If $\mathcal{I}^\varrho(\widehat{\alpha}) = \infty$, the result is immediate, therefore we can assume that $\mathcal{I}^\varrho(\widehat{\alpha}) < \infty$. By Lemma 3.9 it is then enough to prove it for $\widehat{\alpha} \in D_0([0,T];\mathcal{M}_V)$, and so by Theorem 2.3 there exists an $H \in C^{1,0}([0,T]\times V)$ such that

$$\mathbb{P}_N^{\widehat{\varrho}_0,H}(\pi_\cdot^N \in \cdot) \to \delta_{\widehat{\alpha}}. \tag{3.72}$$

where $\widehat{\varrho}_0$ is the density of $\widehat{\alpha}_0$. Therefore, we have that

$$\begin{aligned}
\underline{\lim_{N\to\infty}}\, \frac{1}{N}\log \mathbb{P}_N^\varrho(\pi_\cdot^N \in \mathcal{O}) &= -\underline{\lim_{N\to\infty}}\, \frac{1}{N}\log \mathbb{E}_N^{\widehat{\varrho}_0,H}\left[\mathbb{1}_{\pi_\cdot^N\in\mathcal{O}}\frac{\mathrm{d}\mu_N^{\widehat{\varrho}_0}}{\mathrm{d}\mu_N^{\varrho}}\cdot\frac{\mathrm{d}\mathbb{P}_N^{\widehat{\varrho}_0,H}}{\mathrm{d}\mathbb{P}_N^{\widehat{\varrho}_0}}\right]\\
&\ge -\underline{\lim_{N\to\infty}}\, \frac{1}{N}\mathbb{E}_N^{\widehat{\varrho}_0,H}\left[\log\frac{\mathrm{d}\mu_N^{\widehat{\varrho}_0}}{\mathrm{d}\mu_N^{\varrho}}\cdot\frac{\mathrm{d}\mathbb{P}_N^{\widehat{\varrho}_0,H}}{\mathrm{d}\mathbb{P}_N^{\widehat{\varrho}_0}}\right]\\
&= -\mathcal{I}^\varrho(\widehat{\alpha}),
\end{aligned} \tag{3.73}$$

where we used Lemma 3.7, Corollary 3.2 and that $\mathcal{I}_{tr}(\widehat{\alpha}) = \mathcal{I}_{tr}(\widehat{\alpha};H)$. □

4 Proof of the Hydrodynamic Limit

In this section we prove the hydrodynamic limit of the weakly perturbed process defined in Section 2.2. We first prove that the PDE given in (2.19) is well-posed, and afterwards we prove Theorem (2.3).

4.1 Well-Posedness of the PDE

The approach is to define a sequence of densities $\varrho^{(n)}$ through the following recursive relation

$$\partial_t \varrho_t^{(n+1)}(x,\sigma) = -\sigma \partial_x \varrho_t^{(n+1)}(x,\sigma) + c(-\sigma, m(\varrho_t^{(n)}))e^{\sigma \widetilde{H}_t(x)} \varrho_t^{(n+1)}(x,-\sigma) - c(\sigma, m(\varrho_t^{(n)}))e^{-\sigma \widetilde{H}_t(x)} \varrho_t^{(n+1)}(x,\sigma), \quad (4.1)$$

where every $\varrho^{(n)}$ starts from $\varrho_0^{(n)} = \varrho$. Setting $\mathcal{T}(\varrho^{(n)}) = \varrho^{(n+1)}$, by the Banach fixed-point theorem, it is enough to show that $\mathcal{T}$ is a contraction (up to some finite time $T > 0$) in the space $L^\infty([0,T]; L^1(V))$. First note that for every trajectory $\varrho^{(n)}$, the trajectory $\varrho^{(n+1)}$ solving (4.1) satisfies the conservation of particles, hence for any $t \geq 0$ we have that $||\varrho_t^{(n+1)}||_{L^1(V)} = ||\varrho||_{L^1(V)}$ with ϱ the initial profile. Therefore, we indeed have that $\mathcal{T} : L^\infty([0,T]; L^1(V)) \to L^\infty([0,T]; L^1(V))$. Now let ϱ_1, ϱ_2 be any two trajectories of densities, and denote the following

$$\begin{aligned} \psi_1 &= \mathcal{T}(\varrho_1), & \psi_2 &= \mathcal{T}(\varrho_2), \\ m_{1,t} &= m(\varrho_{1,t}), & m_{2,t} &= m(\varrho_{2,t}). \end{aligned} \quad (4.2)$$

Furthermore, denote $\delta = \psi_1 - \psi_2$ and $\varepsilon = m_1 - m_2$. For $[m] = \{m_t : t \in [0,T]\}$ a deterministic trajectory of magnetizations, we define the mapping

$$Q_{[m]}[\psi](x,\sigma,t) = c(-\sigma, m_t)e^{\sigma \widetilde{H}_t(x)}\psi_t(x,-\sigma) - c(\sigma, m_t)e^{-\sigma \widetilde{H}_t(x)}\psi_t(x,\sigma). \quad (4.3)$$

Note that this mapping is linear in ψ. We then have that

$$\partial_t \delta_t(x,\sigma) = -\sigma \partial_x \delta_t(x,\sigma) + Q_{[m_1]}[\psi_1](x,\sigma,t) - Q_{[m_2]}[\psi_2](x,\sigma,t). \quad (4.4)$$

We can rewrite the difference of the last two terms in a linear and non-linear part as follow

$$Q_{[m_1]}[\psi_1] - Q_{[m_2]}[\psi_2] = Q_{[m_1]}[\psi_1 - \psi_2] + \left(Q_{[m_1]}[\psi_2] - Q_{[m_2]}[\psi_2]\right). \quad (4.5)$$

For the linear part, we have that

$$\begin{aligned} ||Q_{[m_1]}[\delta](\cdot,\cdot,t)||_{L^1(V)} &= \left|\left|c(-\sigma, m_{1,t})e^{\sigma \widetilde{H}_t(x)}\delta_t(x,-\sigma) - c(\sigma, m_{1,t})e^{-\sigma \widetilde{H}_t(x)}\delta_t(x,\sigma)\right|\right|_{L^1(V)} \\ &\leq C_1 ||\delta_t||_{L^1(V)}, \end{aligned} \quad (4.6)$$

with C_1 some constant depending on the bounded functions c and H. For the non-linear part, we have that

$$\begin{aligned} &Q_{[m_1]}[\psi_2](x,\sigma,t) - Q_{[m_2]}[\psi_2](x,\sigma,t) \\ &\quad = [c(-\sigma, m_{1,t}) - c(-\sigma, m_{2,t})]e^{\sigma \widetilde{H}_t(x)}\psi_{2,t}(x,-\sigma) \\ &\quad - [c(\sigma, m_{1,t}) - c(\sigma, m_{2,t})]e^{-\sigma \widetilde{H}_t(x)}\psi_{2,t}(x,\sigma). \end{aligned} \quad (4.7)$$

By the Lipschitz-continuity of $c(\sigma, m)$ we now have that

$$||Q_{[m_1]}[\psi_2](\cdot,\cdot,t) - Q_{[m_2]}[\psi_2](\cdot,\cdot,t)||_{L^1(V)} \leq L|\varepsilon_t| \cdot C_2 ||\psi_{2,t}||_{L^1(V)}, \quad (4.8)$$

with L the Lipschitz-constant and C_2 some constant depending on H. Here $||\psi_{2,t}||_{L^1(V)} = ||\varrho||_{L^1(V)}$ by conservation of particles. Furthermore, by (2.12), we have that

$$|\varepsilon_t| = |m_{1,t} - m_{2,t}| = \frac{1}{||\varrho||_{L^1(V)}} \cdot ||\varrho_{1,t} - \varrho_{2,t}||_{L^1(V)}. \quad (4.9)$$

From (4.6) and (4.8), we find that

$$\begin{aligned}\partial_t||\delta_t||_{L^1(V)} &\leq ||Q_{[m_1]}[\delta](\cdot,\cdot,t)||_{L^1(V)} + ||Q_{[m_1]}[\psi_2](\cdot,\cdot,t) - Q_{[m_2]}[\psi_2](\cdot,\cdot,t)||_{L^1(V)}\\ &\leq C_1||\delta_t||_{L^1(V)} + LC_2||\varrho_{1,t} - \varrho_{2,t}||_{L^1(V)}\\ &\leq C_1||\delta_t||_{L^1(V)} + LC_2||\varrho_1 - \varrho_2||_{L^\infty([0,T];L^1(V))}. \end{aligned} \tag{4.10}$$

Note that the transport term vanishes since $\int_{\mathbb{T}} \partial_x|\delta_t(x,\sigma)|dx = 0$ by periodic boundary conditions. We can rewrite this in integral form as

$$||\delta_t||_{L^1(V)} \leq \int_0^t C_1||\delta_s||_{L^1(V)}\mathrm{d}s + LC_2t||\varrho_1 - \varrho_2||_{L^\infty([0,T];L^1(V))}. \tag{4.11}$$

By Gronwall's inequality, we find that

$$||\delta_t||_{L^1(V)} \leq LC_2te^{C_1t}||\varrho_1 - \varrho_2||_{L^\infty([0,T];L^1(V))}, \tag{4.12}$$

and so we can conclude that

$$||\mathcal{T}(\varrho_1) - \mathcal{T}(\varrho_2)||_{L^\infty([0,T];L^1(V))} \leq LC_2Te^{C_1T}||\varrho_1 - \varrho_2||_{L^\infty([0,T];L^1(V))}. \tag{4.13}$$

Taking $T > 0$ small enough, this is a contraction on $L^\infty([0,T];L^1(V))$, hence the sequence $\varrho^{(n)}$ converges locally to a unique solution of the PDE (2.19).

4.2 Proof of Theorem (2.3)

Recall that $\alpha^H \in D([0,T];\mathcal{M}_V)$ denotes the process such that for every $t \in [0,T]$ the measure α_t^H has density $\varrho_t^H(x,\sigma)$, which is the solution the the PDE given by (2.19). In this way, α^H is the unique measure-valued trajectory such that for every $G \in C^\infty([0,T]\times V)$ and $t \in [0,T]$ we have that

$$\mathscr{M}_t^{H,G}(\alpha^H) := \left\langle \alpha_t^H, G_t\right\rangle - \left\langle \alpha_0^H, G_0\right\rangle - \int_0^t \left\langle \alpha_s^H, \left(\partial_s + \left(A_{s,\alpha^H}^H\right)^*\right)G_s\right\rangle \mathrm{d}s = 0, \tag{4.14}$$

where $\left(A_{s,\alpha^H}^H\right)^*$ is the differential operator given by

$$\left(A_{s,\alpha^H}^H\right)^* G_s(x,\sigma) = \sigma\partial_x G_s(x,\sigma) + c(\sigma, m(\alpha_s^H))e^{-\sigma\widetilde{H}_s(x)}\big(G_s(x,-\sigma) - G_s(x,\sigma)\big), \tag{4.15}$$

which is the action of the L^2-adjoint of A_{s,α^H}^H on smooth functions, with A_{s,α^H}^H given by

$$\begin{aligned}A_{s,\alpha^H}^H G(x,\sigma) &= -\sigma\partial_x G(x,\sigma) + c(-\sigma, m(\alpha_s^H))e^{\sigma\widetilde{H}_s(x)}G(x,-\sigma)\\ &\quad - c(\sigma, m(\alpha_s^H))e^{-\sigma\widetilde{H}_s(x)}G(x,\sigma). \end{aligned} \tag{4.16}$$

For a given $G \in C^\infty([0,T]\times V)$ we define the Dynkin Martingale

$$\mathscr{M}_{N,t}^{H,G}(\pi_\cdot^N) := \langle \pi_t^N, G_t\rangle - \langle \pi_0^N, G_0\rangle - \int_0^t (\partial_s + \mathscr{L}_{N,s}^H)\langle \pi_s^N, G_s\rangle \mathrm{d}s. \tag{4.17}$$

Lemma 4.1 *For every $G \in C^\infty(V)$ we have*

$$\lim_{N\to\infty} \mathbb{E}_N^{\varrho,H}\left[\sup_{t\in[0,T]}\left|\mathscr{M}_{N,t}^{H,G}(\pi_\cdot^N) - \mathscr{M}_t^{H,G}(\pi_\cdot^N)\right|\right] = 0. \tag{4.18}$$

Proof Note that

$$\mathscr{M}^{H,G}_{N,t}(\pi^N_\cdot) - \mathscr{M}^{H,G}_t(\pi^N_\cdot) = \int_0^t \left(\mathscr{L}^H_{N,s}\langle \pi^N_s, G_s\rangle - \left\langle \pi^N_s, \left(A^H_{s,\pi^N_\cdot}\right)^* G_s\right\rangle\right) \mathrm{d}s. \tag{4.19}$$

So we need to calculate $\mathscr{L}^H_{N,s}\langle \pi^N_s, G_s\rangle$. In order to do that, we start with the preliminary calculation

$$\begin{aligned}
\mathscr{L}^H_{N,s}\langle \pi^N_s, G_s\rangle &= N \sum_{(x,\sigma)\in V_N} \eta^N_s(x,\sigma) e^{H_s(\frac{x+\sigma}{N},\sigma) - H_s(\frac{x}{N},\sigma)} \left[\left\langle \pi^N\left((\eta^N_s)^{(x,\sigma)\to(x+\sigma,\sigma)}\right), G_s\right\rangle - \langle \pi^N_s, G_s\rangle\right] \\
&\quad + \sum_{(x,\sigma)\in V_N} \eta^N_s(x,\sigma) c(\sigma, m(\pi^N_s)) e^{-\sigma \widetilde{H}_s(x)} \left[\left\langle \pi^N\left((\eta^N_t)^{(x,\sigma)\to(x,-\sigma)}\right), G_s\right\rangle - \langle \pi^N_s, G_s\rangle\right] \\
&= \sum_{(x,\sigma)\in V_N} \eta^N_s(x,\sigma) e^{H_s(\frac{x+\sigma}{N},\sigma) - H_s(\frac{x}{N},\sigma)} \left(G_s(\tfrac{x+\sigma}{N},\sigma) - G_s(\tfrac{x}{N},\sigma)\right) \\
&\quad + \frac{1}{N} \sum_{(x,\sigma)\in V_N} \eta^N_s(x,\sigma) c(\sigma, m(\pi^N_s)) e^{-\sigma \widetilde{H}_s(x)} \left(G_s(x,-\sigma) - G_s(x,\sigma)\right),
\end{aligned} \tag{4.20}$$

Using Taylor approximation, we can now write

$$\begin{aligned}
\mathscr{L}^H_{N,s}\langle \pi^N_s, G_s\rangle &= \frac{1}{N} \sum_{(x,\sigma)\in V_N} \eta^N_s(x,\sigma) \Big[\sigma \partial_x G_s(\tfrac{x}{N},\sigma) + c(\sigma, m(\pi^N_s)) e^{-\sigma \widetilde{H}_s(x)} \\
&\qquad \left(G_s(x,-\sigma) - G_s(x,\sigma)\right)\Big] \qquad (4.21) \\
&\qquad + R(N,H,G,s) \\
&= \left\langle \pi^N_s, \left(A^H_{s,\pi^N_\cdot}\right)^* G_s\right\rangle + R(N,H,G,s), \qquad (4.22)
\end{aligned}$$

where $R(N,H,G,s)$ is a remainder term satisfying

$$\begin{aligned}
|R(N,H,G,s)| &\le \frac{1}{N} \sum_{(x,\sigma)\in V_N} \eta(x,\sigma) ||\partial_x G||_\infty \sup_{(y,\sigma')\in V_N} \left|H_s(\tfrac{y+\sigma'}{N},\sigma') - H_s(\tfrac{y}{N},\sigma')\right| \\
&\quad + \frac{1}{N^2} \sum_{(x,\sigma)\in V_N} \eta(x,\sigma) ||\partial_{xx} G||_\infty e^{2||H||_\infty} \\
&= \frac{1}{N} |\eta^N| \left(||\partial_x G||_\infty \sup_{(y,\sigma')\in V_N} \left|H_s(\tfrac{y+\sigma'}{N},\sigma') - H_s(\tfrac{y}{N},\sigma')\right| + \frac{1}{N} ||\partial_{xx} G||_\infty e^{2||H||_\infty}\right).
\end{aligned} \tag{4.23}$$

We then conclude that

$$\begin{aligned}
&\mathbb{E}^{\varrho,H}_N \left[\sup_{t\in[0,T]} \left|\mathscr{M}^{H,G}_{N,t}(\pi^N_\cdot) - \mathscr{M}^{H,G}_t(\pi^N_\cdot)\right|\right] \\
&= \mathbb{E}^{\varrho,H}_N \left[\sup_{t\in[0,T]} \left|\int_0^t R(N,H,G,s)\mathrm{d}s\right|\right] \\
&\le \frac{T}{N} \left(||\partial_x G||_\infty \sup_{(y,\sigma')\in V_N} \left|H_s(\tfrac{y+\sigma'}{N},\sigma') - H_s(\tfrac{y}{N},\sigma')\right| + \frac{1}{N} ||\partial_{xx} G||_\infty e^{2||H||_\infty}\right) \mathbb{E}^{\varrho,H}_N \left[|\eta^N|\right] \to 0,
\end{aligned} \tag{4.24}$$

where we used that $\mathbb{E}^{\varrho,H}_N \left[|\eta^N|\right] \le ||\varrho||_\infty N$ and that H_s is continuous on $\mathbb{T}$.

□

Lemma 4.2 *For every $G \in C^\infty([0,T]\times V)$ we have that*

$$\lim_{N\to\infty} \mathbb{E}_N^{\varrho,H}\left[\sup_{t\in[0,T]}\left(\mathscr{M}_{N,t}^{H,G}(\pi_\cdot^N)\right)^2\right] = 0. \tag{4.25}$$

Proof By Doob's maximal inequality we find that

$$\mathbb{E}_N^{\varrho,H}\left[\sup_{t\in[0,T]}\left(\mathscr{M}_{N,t}^{H,G}(\pi_\cdot^N)\right)^2\right] \le 4\mathbb{E}_N^{\varrho,H}\left[\left(\mathscr{M}_{N,T}^{H,G}(\pi_\cdot^N)\right)^2\right] = 4\mathbb{E}_N^{\varrho,H}\left[\int_0^T \Gamma_{N,t}^{H,G}(\pi_t^N)\mathrm{d}t\right], \tag{4.26}$$

where in the last equality we use that the quadratic variation of the martingale $\mathscr{M}_{N,t}^{H,G}$ is given by the integral of the carré du champ operator $\Gamma_{H,t}^{N,G}$ defined by

$$\Gamma_{N,t}^{H,G}(\pi_t^N) = \mathscr{L}_{N,t}^H(\langle\pi_t^N, G_t\rangle)^2 - 2\langle\pi_t^N, G_t\rangle\cdot\mathscr{L}_{N,t}^H\langle\pi_t^N, G_t\rangle. \tag{4.27}$$

For a general jump process generator $Lf(\eta) = \sum_{\eta'} r(\eta,\eta')(f(\eta') - f(\eta))$ we have that

$$Lf^2(\eta) - 2f(\eta)\cdot Lf(\eta) = \sum_{\eta'} r(\eta,\eta')\big(f(\eta') - f(\eta)\big)^2, \tag{4.28}$$

and so we find that

$$\begin{aligned}\Gamma_{N,t}^{H,G}(\pi_t^N) = {}& \frac{1}{N}\sum_{(x,\sigma)\in V_N} \eta_t^N(x,\sigma)e^{H_t(\frac{x+\sigma}{N},\sigma)-H_t(\frac{x}{N},\sigma)}\big(G_t(\tfrac{x+\sigma}{N},\sigma) - G_t(\tfrac{x}{N},\sigma)\big)^2 \\ &+ \frac{1}{N^2}\sum_{(x,\sigma)\in V_N} \eta_t^N(x,\sigma)c(\sigma, m(\pi_t^N))e^{-\sigma\widetilde{H}_t(x)}\big(G_t(x,-\sigma) - G_t(x,\sigma)\big)^2.\end{aligned} \tag{4.29}$$

By the mean value theorem and since $c(\sigma, m)$ is bounded, we can find an upper bound given by

$$\Gamma_{N,t}^{H,G}(\pi_\cdot^N) \le \mathcal{O}(\tfrac{1}{N^2})\cdot|\eta^N|. \tag{4.30}$$

Using again that $\mathbb{E}_N^{\varrho,H}\left[|\eta^N|\right] \le ||\varrho||_\infty N$, we then find that

$$\mathbb{E}_N^{\varrho,H}\left[\sup_{t\in[0,T]}\left(\mathscr{M}_{N,t}^{H,G}(\pi_\cdot^N)\right)^2\right] \le 4T\mathcal{O}(\tfrac{1}{N^2})\cdot\mathbb{E}_N^{\varrho,H}\left[|\eta^N|\right] \to 0. \tag{4.31}$$

□

Proposition 4.3 *$\{\pi_\cdot^N : N\in\mathbb{N}\}$ is tight in $D([0,T];\mathcal{M}_V)$.*

Proof In order to prove tightness, using [18, Theorem 1.3], we have to show the following:

B.1 For all $t\in[0,T]$ and $\varepsilon > 0$ there exists a compact $K(t,\varepsilon)\subset\mathcal{M}_V$ such that

$$\sup_{N\in\mathbb{N}} \mathbb{P}_N^{\varrho,H}\big(\pi_t^N \notin K(t,\varepsilon)\big) \le \varepsilon. \tag{4.32}$$

B.2 For all $\varepsilon > 0$

$$\lim_{\delta\to 0}\limsup_{N\to\infty} \mathbb{P}_N^{\varrho,H}\big(\omega(\pi_\cdot^N,\delta) \ge \varepsilon\big) = 0, \tag{4.33}$$

where

$$\omega(\pi_\cdot^N,\delta) = \sup\{d(\pi_t^N,\pi_s^N)| : s,t\in[0,T], |t-s| < \delta\}, \tag{4.34}$$

with d the metric on $\mathcal{M}_V$ defined for $\alpha, \beta \in \mathcal{M}_V$ as

$$d(\alpha, \beta) = \sum_{j=1}^{\infty} 2^{-j}\Big(1 \wedge \big|\langle \alpha, \phi_j\rangle - \langle \beta, \phi_j\rangle\big|\Big). \tag{4.35}$$

We start with proving **B.1**. For every $C > 0$, we have that the set $K_C = \{\mu \in \mathcal{M}_V : \mu(V) \leq C\}$ is compact in $\mathcal{M}_V$. Furthermore,

$$\mathbb{P}_N^{\varrho,H}(\pi_t^N \notin K_C) = \mathbb{P}_N^{\varrho,H}(\pi_t^N(V) > C) \leq \frac{1}{C}\mathbb{E}_N^{\varrho,H}[\pi_t^N(V)], \tag{4.36}$$

where we used the Markov inequality in the last step. Here

$$\mathbb{E}_N^{\varrho,H}[\pi_t^N(V)] = \mathbb{E}_N^{\varrho,H}\left[\frac{1}{N}\sum_{(x,\sigma)\in V_N} \eta_t^N(x,\sigma)\delta_{(\frac{x}{N},\sigma)}(V)\right] = \frac{1}{N}\mathbb{E}_N^{\varrho,H}\big[|\eta^N|\big] \leq ||\varrho||_\infty. \tag{4.37}$$

Therefore

$$\mathbb{P}_N^{\varrho,H}(\pi_t^N \notin K_C) \leq \frac{1}{C}||\varrho||_\infty. \tag{4.38}$$

Since we took C arbitrarily, we can take $C > ||\varrho||_\infty \varepsilon^{-1}$, and **B.1** follows.

To prove **B.2**, take $\varepsilon' < \varepsilon$ and note that by the Markov inequality we have that

$$\begin{aligned}\lim_{\delta\to 0}\limsup_{N\to\infty} \mathbb{P}_N^{\varrho,H}(\omega(\pi_\cdot^N, \delta) > \varepsilon) &\leq \lim_{\delta\to 0}\limsup_{N\to\infty} \frac{1}{\varepsilon}\mathbb{E}_N^{\varrho,H}\Big[\omega(\pi_\cdot^N, \delta)\Big] \\ &\leq \lim_{\delta\to 0}\limsup_{N\to\infty} \frac{1}{\varepsilon'}\mathbb{E}_N^{\varrho,H}\Big[\omega(\pi_\cdot^N, \delta)\Big].\end{aligned} \tag{4.39}$$

Now for $\omega(\pi_\cdot^N, \delta)$ we have that

$$\begin{aligned}\omega(\pi_\cdot^N, \delta) &= \sup_{\substack{s,t\in[0,T]\\ |t-s|<\delta}} \sum_{j=1}^{\infty} 2^{-j}\Big(1 \wedge \big|\langle \pi_t^N, \phi_j\rangle - \pi_s^N(\phi_j)\big|\Big) \\ &\leq 2^{-m} + \sup_{\substack{s,t\in[0,T]\\ |t-s|<\delta}} \sum_{j=1}^{m} \big|\langle \pi_t^N, \phi_j\rangle - \pi_s^N(\phi_j)\big|,\end{aligned} \tag{4.40}$$

where we took $m \in \mathbb{N}$ arbitrarily. Using the martingale $\mathscr{M}_{N,t}^{H,\phi_j}(\pi_\cdot^N)$, we find that

$$\begin{aligned}\mathbb{E}_N^{\varrho,H}\Big[\big|\langle \pi_t^N, \phi_j\rangle - \pi_s^N(\phi_j)\big|\Big] &= \mathbb{E}_N^{\varrho,H}\left[\left|\mathscr{M}_{N,t}^{H,\phi_j}(\pi_\cdot^N) - \mathscr{M}_{N,s}^{H,\phi_j}(\pi_\cdot^N) - \int_s^t \mathscr{L}_{N,s}^H\langle \pi_r^N, \phi_j\rangle \mathrm{d}r\right|\right] \\ &\leq 2\mathbb{E}_N^{\varrho,H}\left[\sup_{t\in[0,T]}\left|\mathscr{M}_{N,t}^{H,\phi_j}(\pi_\cdot^N)\right|\right] + \mathbb{E}_N^{\varrho,H}\left[\left|\int_s^t \mathscr{L}_{N,s}^H\langle \pi_r^N, \phi_j\rangle \mathrm{d}r\right|\right].\end{aligned} \tag{4.41}$$

By Lemma 4.2 the first expectation vanishes as $N \to \infty$. By (4.21) we can upper bound the second expectation by

$$\begin{aligned}\mathbb{E}_N^{\varrho,H}\left[\left|\int_s^t \mathscr{L}_{N,s}^H\langle \pi_r^N, \phi_j\rangle \mathrm{d}r\right|\right] &\leq \mathbb{E}_N^{\varrho,H}\left[\left|\int_s^t \left\langle \pi_r^N, \left(A_{s,\pi_\cdot^N}^H\right)^* \phi_j\right\rangle \mathrm{d}r\right|\right] \\ &\quad + \mathbb{E}_N^{\varrho,H}\left[\left|\int_s^t R(N, \phi_j, r)\mathrm{d}r\right|\right].\end{aligned} \tag{4.42}$$

From (4.24) we see that the second expectation also vanishes as $N \to \infty$ (uniformly in s and t). For the first expectation note that

$$\begin{aligned}\mathbb{E}_N^{\varrho,H}\left[\left|\int_s^t \left\langle \pi_r^N, \left(A^H_{s,\pi_\cdot^N}\right)^* \phi_j\right\rangle \mathrm{d}r\right|\right] &\le \mathbb{E}_N^{\varrho,H}\left[\left|\int_s^t \frac{1}{N}\sum_{(x,\sigma)\in V_N} \eta_r^N(x,\sigma)\left|\left|\left(A^H_{s,\pi_\cdot^N}\right)^* \phi_j\right|\right|_\infty \mathrm{d}r\right|\right] \\ &\le \frac{1}{N}\left|\left|\left(A^H_{s,\pi_\cdot^N}\right)^* \phi_j\right|\right|_\infty |t-s|\cdot \mathbb{E}_N^{\varrho,H}\left[|\eta^N|\right] \\ &\le \left|\left|\left(A^H_{s,\pi_\cdot^N}\right)^* \phi_j\right|\right|_\infty ||\varrho||_\infty \delta. \end{aligned} \tag{4.43}$$

Combining all of the above, we find that

$$\begin{aligned}&\lim_{\delta\to 0}\limsup_{N\to\infty} \mathbb{P}_N^{\varrho,H}(\omega(\pi_\cdot^N,\delta) > \varepsilon) \\ &\le \frac{1}{\varepsilon'}2^{-m} + \lim_{\delta\to 0}\limsup_{N\to\infty}\frac{1}{\varepsilon'}\mathbb{E}_N^{\varrho,H}\left[\sup_{\substack{s,t\in[0,T]\\ |t-s|<\delta}}\sum_{j=1}^m \left|\langle \pi_t^N,\phi_j\rangle - \pi_s^N(\phi_j)\right|\right] \\ &\le \frac{1}{\varepsilon'}2^{-m} + \lim_{\delta\to 0}\frac{1}{\varepsilon'}\left|\left|\left(A^H_{s,\pi_\cdot^N}\right)^* \phi_j\right|\right|_\infty ||\varrho||_\infty \delta \\ &= \frac{1}{\varepsilon'}2^{-m}. \end{aligned} \tag{4.44}$$

Since we took m arbitrarily, we can choose it such that $2^{-m} \le (\varepsilon')^2$, i.e.,

$$\lim_{\delta\to 0}\limsup_{N\to\infty} \mathbb{P}_N^{\varrho,H}(\omega(\pi_\cdot^N,\delta) > \varepsilon) < \varepsilon', \tag{4.45}$$

and since we took ε' arbitrarily small, we indeed find that **B.2** holds. □

We are now ready to give the proof of the hydrodynamic limit of the weakly perturbed model.

Proof of Theorem 2.3 By Prokhorov's theorem, the tightness of the sequence $\{\pi_\cdot^N : N \in \mathbb{N}\}$ implies that the sequence is sequentially compact. If we then prove that every convergent subsequence converges to δ_α, then the theorem holds.

Take such a convergent subsequence $\mathbb{P}_{N_k}^{\varrho,H}(\pi_\cdot^N \in \cdot) \to P^*$, with P^* a probability measure on $D([0,T];\mathcal{M}_V)$. For a given $\varepsilon > 0$ and $G \in C^\infty(V)$, define the set

$$\Xi_\varepsilon^{H,G} = \left\{\beta \in D([0,T];\mathcal{M}_V) : \sup_{t\in[0,T]}\left|\mathscr{M}_t^{H,G}(\beta)\right| \le \epsilon\right\}, \tag{4.46}$$

which is closed in the Skorokhod topology. By Portmanteau's theorem, we now have that

$$\begin{aligned}P^*(\Xi_\varepsilon^{H,G}) &\ge \lim_{k\to\infty}\mathbb{P}_{N_k}^{\varrho,H}(\pi_\cdot^{N_k} \in \Xi_\varepsilon^{H,G}) \\ &= \lim_{k\to\infty}\mathbb{P}_{N_k}^{\varrho,H}\left(\sup_{t\in[0,T]}\left|\mathscr{M}_t^{H,G}(\pi_\cdot^{N_k})\right| \le \varepsilon\right) \\ &= \lim_{k\to\infty}\mathbb{P}_{N_k}^{\varrho,H}\left(\sup_{t\in[0,T]}\left|\mathscr{M}_{N_k,t}^{H,G}(\pi_\cdot^{N_k})\right| \le \varepsilon\right), \end{aligned} \tag{4.47}$$

where we used Lemma 4.1 for the last step. Now by using Chebyshev's inequality together with Lemma 4.2, we find that

$$\mathbb{P}_{N_k}^{\varrho,H}\left(\sup_{t\in[0,T]}\left|\mathscr{M}_{N_k,t}^{H,G}(\pi_\cdot^{N_k})\right|>\varepsilon\right)\leq\frac{1}{\varepsilon^2}\mathbb{E}_{N_k}^{\varrho,H}\left[\sup_{t\in[0,T]}\left|\mathscr{M}_{N_k,t}^{H,G}(\pi_\cdot^{N_k})\right|^2\right]\to 0, \tag{4.48}$$

and so indeed

$$P^*(\Xi_\varepsilon^{H,G})\geq\lim_{k\to\infty}\mathbb{P}_{N_k}^{\varrho,H}\left(\sup_{t\in[0,T]}\left|\mathscr{M}_{N_k,t}^{H,G}(\pi_\cdot^{N_k})\right|\leq\varepsilon\right)=1. \tag{4.49}$$

Since this is true for all $\varepsilon>0$ and $G\in C^\infty(V)$, it follows that $P^*=\delta_{\alpha^H}$. □

Data Availability Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

Declarations

Competing interests The authors have no relevant financial or non-financial interests to disclose.

References

1. Solon, A.P., Cates, M.E., Tailleur, J.: Active brownian particles and run-and-tumble particles: A comparative study. The European Physical Journal Special Topics **224**(7), 1231–1262 (July2015)
2. Czirók, A., Ben-Jacob, E., Cohen, I., Vicsek, T.: Formation of complex bacterial colonies via self-generated vortices. Phys. Rev. E **54**(2), 1791 (1996)
3. Schaller, V., Weber, C., Semmrich, C., Frey, E., Bausch, A.R.: Polar patterns of driven filaments. Nature **467**(7311), 73–77 (2010)
4. Degond, P., Liu, J.-G., Motsch, S., Panferov, V.: Hydrodynamic models of self-organized dynamics: derivation and existence theory. Methods and Applications of Analysis, 089–114, (2013)
5. Erignoux, C.: On the hydrodynamics of active matter models on a lattice. arXiv:2108.04003, (2021)
6. Kourbane-Houssene, M., Erignoux, C., Bodineau, T., Tailleur, J.: Exact hydrodynamic description of active lattice gases. Phys. Rev. Lett. **120**(26), 268003 (2018)
7. Noguchi, H.: Spatiotemporal patterns in active four-state potts models. Sci. Rep. **15**(1), 674 (2025)
8. Kipnis, C., Olla, S., Varadhan, S.R.S.: Hydrodynamics and large deviation for simple exclusion processes. Commun. Pure Appl. Math. **42**(2), 115–137 (March 1989)
9. Agranov, T., Ro, S., Kafri, Y., Lecomte, V.: Macroscopic fluctuation theory and current fluctuations in active lattice gases. SciPost Physics **14**(3), 045 (2023)
10. Keta, Y.-E., Fodor, É., van Wijland, F., Cates, M.E., Jack, R.L.: Collective motion in large deviations of active particles. Phys. Rev. E **103**(2), 022603 (2021)
11. Burenev, I., Cloete, D., Kharbanda, V., Touchette, H.: An introduction to large deviations with applications in physics. SciPost Physics Lecture Notes, 104, (2025)
12. Touchette, H.: The large deviation approach to statistical mechanics. Phys. Rep. **478**(1–3), 1–69 (2009)
13. Seppäläinen, T.: Translation invariant exclusion processes (book in progress). University of Wisconsin, Department of Mathematics, (2008)
14. Redig, F., van Wiechen, H.: Ergodic theory of multi-layer interacting particle systems. J. Stat. Phys. **190**(4), 88 (2023)

15. Redig, F., van Wiechen, H.: Stationary fluctuations of run-and-tumble particles. Markov Process. Related Fields **30**(2), 297–331 (2024)
16. van Ginkel, B., van Gisbergen, B., Redig, F.: Run-and-Tumble Motion: The Role of Reversibility. J. Stat. Phys. **183**(3), 44 (June2021)
17. Demaerel, T., Maes, C.: Active processes in one dimension. Phys. Rev. E **97**(3), 032604 (2018)
18. Kipnis, C., Landim, C.: Scaling Limits of Interacting Particle Systems. Grundlehren der mathematischen Wissenschaften. Springer, Berlin Heidelberg (1998)
19. Palmowski, Z., Rolski, T.: A technique for exponential change of measure for Markov processes. Bernoulli **8**(6), 767–785 (2002)
20. Varadhan, S.R.S.: Large deviations and applications. Society for Industrial and Applied Mathematics (SIAM), (1984)