



Heriot-Watt University  
Research Gateway

## Fast non mean field networks: uniform in time averaging

### Citation for published version:

Barré, J, Dobson, P, Ottobre, M & Zatorska, E 2021, 'Fast non mean field networks: uniform in time averaging', *SIAM Journal on Mathematical Analysis*, vol. 53, no. 1, pp. 937–972.  
<https://doi.org/10.1137/20M1328646>

### Digital Object Identifier (DOI):

[10.1137/20M1328646](https://doi.org/10.1137/20M1328646)

### Link:

[Link to publication record in Heriot-Watt Research Portal](#)

### Document Version:

Peer reviewed version

### Published In:

SIAM Journal on Mathematical Analysis

### Publisher Rights Statement:

© 2021, Society for Industrial and Applied Mathematics

Read More: <https://epubs.siam.org/doi/abs/10.1137/20M1328646?af=R&>

### General rights

Copyright for the publications made accessible via Heriot-Watt Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

### Take down policy

Heriot-Watt University has made every reasonable effort to ensure that the content in Heriot-Watt Research Portal complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [open.access@hw.ac.uk](mailto:open.access@hw.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.

# FAST NON MEAN-FIELD NETWORKS: UNIFORM IN TIME AVERAGING

J. BARRÉ<sup>(1)</sup>, P. DOBSON<sup>(2)</sup>, M. OTTOBRE<sup>(3)</sup> AND E. ZATORSKA<sup>(4)</sup>

**ABSTRACT.** We study a population of  $N$  particles, which evolve according to a diffusion process and interact through a dynamical network. In turn, the evolution of the network is coupled to the particles' positions. In contrast with the mean-field regime, in which each particle interacts with every other particle, i.e. with  $O(N)$  particles, we consider the a priori more difficult case of a *sparse network*; that is, each particle interacts, on average, with  $O(1)$  particles. We also assume that the network's dynamics is much faster than the particles' dynamics, with the time-scale of the network described by a parameter  $\varepsilon > 0$ . We combine the averaging ( $\varepsilon \rightarrow 0$ ) and the many particles ( $N \rightarrow \infty$ ) limits and prove that the evolution of the particles' empirical density is described (after taking both limits) by a non-linear Fokker-Planck equation; we moreover give conditions under which such limits can be taken *uniformly in time*, hence providing a criterion under which the limiting non-linear Fokker-Planck equation is a good approximation of the original system uniformly in time. The heart of our proof consists of controlling precisely the dependence in  $N$  of the averaging estimates.

**KEYWORDS.** Interacting particle systems, Markov Semigroups, Averaging methods, sparse interaction, non-linear Fokker-Planck equation.

**AMS CLASSIFICATION (MSC 2020).** 60K35, 47D07, 60J60, 35Q84, 35Q82, 82C31.

## 1. INTRODUCTION

Systems of diffusing particles with mean-field interactions, i.e. “all to all”-type interactions, have been extensively studied in various contexts. In particular it is well-known that, under mild conditions, the empirical density of  $N$  such particles can be described in the large  $N$  limit by a non-linear Fokker-Planck equation; classical references for this are [23, 32]. The fundamental reason behind this behavior is a law of large numbers: each particle interacts with many others, hence feels an almost deterministic force. A natural question, which is relevant for modelling purposes, regards interactions mediated by a network where two particles, or more generally two agents, interact only if they are linked together. When each particle is linked, on average, to  $O(N)$  particles –i.e., borrowing nomenclature from [13], when the *average degree* of the interaction network diverges – each particle again typically interacts with many others, hence one may still expect a kind of law of large numbers to apply, leading to a non-linear Fokker-Planck equation in the macroscopic limit. With some caveats and subtleties, this is indeed the case, as shown in [12, 20]; [8] and [25] go further in this direction, and tackle the large deviation regime. On the other hand, when the mean degree of the interaction network is bounded (in  $N$ ), i.e. when each particle interacts on average with  $O(1)$  particles so that the network is *sparse*, one may expect a very different, and more complicated, phenomenology; this case is studied e.g. in [26].

In many situations, it is actually important to take one further step, and consider a dynamical network whose evolution is coupled with the particles' dynamics (for reviews see for instance [17, 29]). This is clear in epidemiology, where the network of contacts between people is influenced by the spreading of the disease [18, 21]; in models of opinion formation, where people with similar opinions may tend to interact more together [16]; but also in the modelling of various biological phenomena [3, 11, 28, 33, 34] (note that [11, 28] tackle the extra difficulty of considering individual

units which are fibers, hence with an orientational degree of freedom, and not merely point particles).

Details of the model that we will consider can be found in Section 2.1 below; for the time-being we summarize the main features of the dynamics that we will be studying: in this work, which is inspired by the biological applications of [11, 28, 33, 34] and by the framework of [2, 4], we consider a system of  $N$  particles, each of them moving in  $\mathbb{R}^n$ , and interacting through a sparse graph/network (we will use the words network and graph interchangeably), i.e. particle  $i$  and  $j$  interact only when they are linked by the edge of an evolving network. In turn, the evolution of the network is coupled to the particles' dynamics through the following mechanism: the network's edges are created and deleted at random times described by Poisson processes; in particular, the rate of the Poisson process governing the link formation mechanism between particle  $i$  and particle  $j$  depends on the distance between such particles and it is designed in a way that a link can only be formed if the particles are close enough (i.e., within a fixed distance  $R > 0$  of each other). Moreover, our choice of the link formation and destruction rates guarantees that the graph remains sparse.

These coupled network/particles systems are difficult to analyze mathematically in full generality, and it is then important to understand some limit regimes which may be amenable to analysis. In this article, we study the limit of a time-scale separation between network and particles' dynamics and focus on the regime in which the network's dynamics evolves much faster than the particles' dynamics; the time-scale of the network is governed by a parameter  $\varepsilon > 0$ , which enters the evolution in such a way that  $\varepsilon \ll 1$  corresponds to accelerating the network's dynamics. In this regime one can employ averaging techniques to “average out” the effect of the network and consider an *effective* interaction between particles. In particular, we take both the averaging ( $\varepsilon \rightarrow 0$ ) and many particles ( $N \rightarrow \infty$ ) limits and prove that the *effective macroscopic description* of the dynamics (i.e. the one obtained after taking both the  $\varepsilon$  and  $N$  limits) is given by an appropriate non-linear Fokker-Planck equation. Let us emphasize from now that we take such limits *together*; this is not only needed if one is after a rigorous treatment but, on a technical level, this is what enables us to treat the a priori more difficult case of a sparse network (see comments in the next paragraph below, at the end of Section 3 and at the beginning of Section 5 on this point). Furthermore, since one is often interested in the asymptotic ( $t \rightarrow \infty$ ) behavior of such models, we identify sufficient conditions so that the effective macroscopic description is a valid approximation of the initial system *uniformly in time*. We will comment on the relevance of such uniform in time estimates in Note 2.2; for the time being we point out in passing that, as a bi-product, our approach produces the first (to the best of our knowledge) uniform in time averaging result.

A comprehensive literature review of the subject is out of reach, so we mention only the works which are most relevant to our context. The papers [2, 4, 11, 34] consider first the limit of a large population ( $N \rightarrow \infty$  with our notation), without the hypothesis of a fast network dynamics, obtaining, in the limit, kinetic-like equations describing the density of the law of the coupled evolution of the particles and the network. To do so, these articles make formal “propagation-of-chaos”-like ansatzes, which may be reasonable approximations, but are unlikely to be fully valid in the context we consider here of sparse networks. The works [2, 4, 11] then obtain an effective macroscopic description by taking formally the fast network (i.e.  $\varepsilon \rightarrow 0$ ) limit in these kinetic equations. They obtain the analog of the nonlinear Fokker-Planck equation that we obtain in this paper, equation (14). In [3] it was suggested that one should be able to rigorously derive the effective macroscopic description (14), without the formal intermediate step of the kinetic equations (i.e. without any formal ansatz). This is precisely what we achieve in this paper.

Let us also observe that this interplay between many particle limit and a fast randomly evolving interaction is relevant in the context of two-layers neural networks as well, as new exciting developments are linking the theory of neural networks with the framework of interacting

particle systems [22, 30, 31]: the units of the neural network are described by parameters to be optimized using the available data; the dynamics induced by the learning procedure (a stochastic gradient descent) translates as an interacting particles dynamics for these parameters, in which the interaction is random and changes as new learning data are taken into account. Further recent relevant literature on mean-field limits for dynamically evolving graphs can be found in [5, 6].

Our main result is Theorem 2.1, which proves the convergence of the empirical density of particles to the solution of an appropriate non-linear Fokker-Planck equation (namely equation (14) below), in the combined limit  $N \rightarrow \infty$  (large population) and  $\varepsilon \rightarrow 0$  (large time scale separation). Part b of the theorem provides sufficient conditions under which this convergence is uniform in time. Let us briefly comment on the main technical difficulties which are tackled in this paper. Firstly, the proof combines averaging techniques and “mean-field-like” ones; by themselves, these techniques are known, so the issue here is to combine them and control precisely the dependence in  $N$  of the averaging estimates, in order to be able to take the many particles limit (the key result in this respect is given by Theorem 4.1). We believe that the  $N$  and  $\varepsilon$  dependence of our estimates is optimal, and we heuristically explain at the end of Note 2.2 why we think this is the case. Secondly, the idea to obtain uniform in time estimates is borrowed and re-adapted from [9, 10] and it hinges on conquering exponentially fast decay (in time) for the space-derivatives of an appropriate Markov Semigroup; further to [9, 10], we gained explicit dimension-dependence (i.e. dependence on  $N$ ) of the constants appearing in such estimates, which was crucial to our proof. More comments on this and on related literature in Note 2.2. Finally, to tackle the many particles limit in the sparse-network regime, we show that the dynamics “preserves sparsity”: if the network is sparse at time zero, it will remain sparse at subsequent times (this is addressed in Lemma 4.2). To conclude, we point out that, while the set up of the problem is stochastic, the whole method of proof relies solely on analytic techniques.

The paper is organized as follows. In Section 2 we describe the model that we will be analysing, present an intuitive derivation of the effective macroscopic description, and state our main result; Section 3, Section 4 and Section 5 are devoted to the proofs: Section 3 outlines the general strategy to prove Theorem 2.1, Section 4 details the averaging estimates with the crucial proof of Theorem 4.1, and Section 5 presents the many-particles estimates.

## 2. DESCRIPTION OF THE MODEL AND STATEMENT OF MAIN RESULTS

**2.1. Description of the Model.** We consider a system of  $N$  interacting particles, each of them moving in  $\mathbb{R}^n$ . The position of the  $i$ -th particle is denoted by  $X_t^{i,N,\varepsilon}$  (the reason for the superscript  $\varepsilon$  will be clarified below) and, collectively, we will use the notation  $\mathbb{R}^{Nn} \ni X_t^{N,\varepsilon} := (X_t^{1,N,\varepsilon}, \dots, X_t^{N,N,\varepsilon})$  for the whole process; such particles are connected through an evolving graph which we describe by its adjacency matrix  $A^{N,\varepsilon}(t) = \{A_{ij}^{N,\varepsilon}(t)\}_{1 \leq i,j \leq N}$ . For each  $t \geq 0$ ,  $A^{N,\varepsilon}(t)$  is a symmetric  $N \times N$  matrix with entries 0 or 1: for any pair  $(i, j)$ , there is an edge between particles  $i$  and  $j$  at time  $t$  if and only if  $A_{ij}^{N,\varepsilon}(t) = 1$ . The particles are not linked to themselves, and so we take  $A_{ii}^{N,\varepsilon}(t) = 0$  for all  $i = 1, \dots, N$ . In other words, for every  $t \geq 0$ ,  $A^{N,\varepsilon}(t) \in \mathcal{A}^N$ , where

$$\mathcal{A}^N := \{\text{arrays } \mathbf{a} = \{a_{ij}\}_{1 \leq i,j \leq N} : a_{ij} = a_{ji} \in \{0, 1\} \text{ and } a_{ii} = 0 \text{ for every } i = 1, \dots, N\}.$$

The coupled equations of motion for the particles and the evolution of the graph are as follows:

$$dX_t^{i,N,\varepsilon} = -\nabla V(X_t^{i,N,\varepsilon})dt + \sum_{j \neq i} A_{ij}^{N,\varepsilon}(t)K(X_t^{j,N,\varepsilon} - X_t^{i,N,\varepsilon})dt + \sqrt{2D}dB_t^i \quad (1a)$$

$$dA_{ij}^{N,\varepsilon}(t) = -A_{ij}^{N,\varepsilon}(t_-)dN_{ij}^{d,\varepsilon}(t) + [1 - A_{ij}^{N,\varepsilon}(t_-)]dN_{ij}^{f,\varepsilon}(t). \quad (1b)$$

We describe in turn all the contributions to the above dynamics and make assumptions on the initial data:

- all the particles are subject to the same *external potential*  $V : \mathbb{R}^n \rightarrow \mathbb{R}$  (as customary  $\nabla$  denotes  $n$ -dimensional gradient w.r.t. the argument of  $V$ ) and to randomness, as described by the processes  $B_t^i$ 's,  $i = 1, \dots, N$ , which are  $n$ -dimensional independent standard Brownian motions;  $D > 0$  is the diffusion constant;
- if  $A_{ij}^{N,\varepsilon}(t) = 1$  then at time  $t$  particle  $i$  and particle  $j$  interact through the *interaction force*  $K : \mathbb{R}^n \rightarrow \mathbb{R}^n$  and they keep interacting as long as  $A_{ij}^{N,\varepsilon}(t)$  remains equal to one;
- in (1b),  $N_{ij}^{d,\varepsilon}, N_{ij}^{f,\varepsilon}$  are independent Poisson processes with intensities  $\nu_d \varepsilon^{-1}$  and  $\tilde{\nu}_f(x^i, x^j) \varepsilon^{-1}$ , respectively, where:  $\varepsilon$  is a fixed positive parameter;  $\nu_d > 0$  is a constant (independent of  $N, \varepsilon, i$  and  $j$ ), which we refer to as the *destruction rate*;  $\tilde{\nu}_f : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$  is a function, namely

$$\tilde{\nu}_f(x^i, x^j) := \nu_f^N \varphi_R(|x^i - x^j|) \quad (2)$$

with  $\nu_f^N$  a positive constant, the *formation rate*, which depends on  $N$  in a way that will be specified later (the  $N$  dependence is chosen in a way to ensure that the interaction network is sparse, see (11)) but is independent of everything else;  $|\cdot|$  denotes euclidean norm in  $\mathbb{R}^n$  and  $\varphi_R : \mathbb{R} \rightarrow [0, 1]$  is a smooth function with compact support in  $[0, R]$  for some fixed  $R > 0$ ;

- the initial data  $X_0^{N,\varepsilon}, A^{N,\varepsilon}(0)$  are i.i.d. random variables; we will make precise assumptions on the initial data in Section 2.3, see Hypothesis **[A.1]**;
- finally,  $t_-$  denotes left limit at  $t$ ; that is, for any càdlàg function  $f : \mathbb{R} \rightarrow \mathbb{R}$ , we set  $f(t_-) := \lim_{s \rightarrow t_-} f(s)$ .

Precise assumptions on the potential  $V$  and force kernel  $K$  and on the initial data will be given in Section 2.3. The evolution for the link  $A_{ij}^{N,\varepsilon}$  in (1b) is a piecewise constant process which switches between the values 0 and 1 at random times according to the Poisson clocks  $N_{ij}^{d,\varepsilon}$  and  $N_{ij}^{f,\varepsilon}$ . In particular, equation (1b) is coupled to the particles' evolution (1a) only through the process  $N_{ij}^{f,\varepsilon}(t)$ ; this is a bit hidden in the notation that we have used in (1b), so to clarify this fact let us point out that if  $\varphi_R$  was the indicator function of the interval  $[0, R]$  (denoted by  $\mathbf{1}_{[0,R]}(\cdot)$ ), (1b) could be replaced by the following equivalent dynamics

$$dA_{ij}^{N,\varepsilon}(t) = -A_{ij}^{N,\varepsilon}(t_-)dN_{ij}^{d,\varepsilon}(t) + [1 - A_{ij}^{N,\varepsilon}(t_-)]\mathbf{1}_{[0,R]}(|X_t^{i,N,\varepsilon} - X_t^{j,N,\varepsilon}|)d\check{N}_{ij}^{f,\varepsilon}(t), \quad (3)$$

where now  $\check{N}_{ij}^{f,\varepsilon}(t)$  is a simple Poisson process with rate  $\nu_f^N$ .

In order to gain further intuition about the dynamics (1a)-(1b) one may think for example of  $K$  as being an elastic force (i.e. a short-range repulsion and long-range attraction kind of interaction) so that, if  $A_{ij}^{N,\varepsilon}(t) = 1$  then at time  $t$  particle  $i$  and  $j$  are linked via a spring (as well as being subject to the external potential  $V$  and to randomness) and when the Poisson clock  $N_{ij}^{d,\varepsilon}$  goes off then the link between the two particles is destroyed, i.e.  $A_{ij}^{N,\varepsilon}$  switches to zero. However, if particle  $i$  and  $j$  are not linked at time  $t$ , i.e. if  $A_{ij}^{N,\varepsilon}(t) = 0$ , then two things can happen: if  $|X_t^{i,N,\varepsilon} - X_t^{j,N,\varepsilon}| \leq R$ , then the Poisson clock  $N_{ij}^{f,\varepsilon}$  may have a positive rate

(depending on whether  $\varphi_R(|X_t^{i,N,\varepsilon} - X_t^{j,N,\varepsilon}|)$  is bigger or equal to zero) and may go off, in which case a link is formed and  $A_{ij}^{N,\varepsilon}$  switches to one; if on the other hand  $|X_t^{i,N,\varepsilon} - X_t^{j,N,\varepsilon}| > R$ , then the Poisson clock  $N_{ij}^{f,\varepsilon}$  has zero rate (as in this case  $\varphi_R(|X_t^{i,N,\varepsilon} - X_t^{j,N,\varepsilon}|) = 0$ ) and does not go off, so that  $A_{ij}^{N,\varepsilon}$  remains equal to zero at least as long as  $|X_t^{i,N,\varepsilon} - X_t^{j,N,\varepsilon}| > R$ . In other words the function  $\varphi_R$  ensures that an edge/link can only be created between particles  $i$  and  $j$  if such particles are within distance  $R$  of each other.

The existence of both weak and strong solutions to system (1a)-(1b) is standard under our assumptions on the potentials  $V$  and force kernel  $K$  (see Note 2.3). For completeness, we quickly recall the strategy to build such solutions in Appendix A.

**2.2. Main Result.** We are concerned with understanding the behaviour of system (1a)-(1b) in the limit  $\varepsilon \rightarrow 0$  and  $N \rightarrow \infty$ . Let us emphasize that these limits are taken *together* – although, for expository purposes, in what follows we present the heuristic derivation of our main results by formally taking the  $\varepsilon$  limit first (averaging limit) and then the  $N \rightarrow \infty$  (large particle) limit. Exact proofs are contained in Section 3, Section 4 and Section 5.

When  $0 < \varepsilon \ll 1$ , the links' dynamics is much faster than the particles' dynamics; this kind of situation is dealt with within the framework of stochastic averaging, see e.g. [27]. The idea of stochastic averaging can be informally described as follows: because the process  $A^{N,\varepsilon}(t)$  (1b) evolves much faster than the process  $X_t^{N,\varepsilon}$  (1a), the former process will have effectively reached equilibrium while the latter has remained substantially unchanged. In particular, considering the slow process as “frozen”, say  $X_t^{N,\varepsilon} = \mathbf{x} \in \mathbb{R}^{Nn}$ , the fast process will converge, as  $\varepsilon \rightarrow 0$ , to a stationary distribution, denoted by  $\mu_{\mathbf{x}}$  (this measure clearly depends on the “frozen” values of the slow process). Note that, given the value of the slow process  $X_t^{N,\varepsilon}$ , the components  $A_{ij}^{N,\varepsilon}(t)$  of the fast process evolve independently of each other; the evolution of each component is a simple two-state Markov process with generator given by the matrix

$$L^{(ij)} = \frac{1}{\varepsilon} \begin{pmatrix} -\nu_f^N \varphi_R(|x^i - x^j|) & \nu_f^N \varphi_R(|x^i - x^j|) \\ \nu_d & -\nu_d \end{pmatrix}.$$

The invariant measure  $\mu_{\mathbf{x}}$  of  $A^{N,\varepsilon}(t)$  given  $X_t^{N,\varepsilon}$ , is a probability measure on the space  $\mathcal{A}^N$ . Let us emphasize that this is the invariant measure for (1b), given  $X_t^{N,\varepsilon} = \mathbf{x}$ . Because of conditional independence, one just needs to find the invariant measure  $\mu_{\mathbf{x}}^{(ij)}$  for the dynamics of  $A_{ij}^{N,\varepsilon}$  (given  $X_t^{N,\varepsilon}$ ) and then the overall invariant measure on  $\mathcal{A}^N$  is just a product measure. The invariant measure for the process  $A_{ij}^{N,\varepsilon}(t)$  is the only normalised element of the kernel of the transpose of the matrix  $L^{(ij)}$ , see [24]. A simple calculation yields

$$\mu_{\mathbf{x}}^{(ij)}(1) = \frac{\nu_f^N \varphi_R(|x^i - x^j|)}{\nu_f^N \varphi_R(|x^i - x^j|) + \nu_d}, \quad \mu_{\mathbf{x}}^{(ij)}(0) = 1 - \frac{\nu_f^N \varphi_R(|x^i - x^j|)}{\nu_f^N \varphi_R(|x^i - x^j|) + \nu_d}.$$

Hence, the sought after invariant measure  $\mu_{\mathbf{x}}$  on  $\mathcal{A}^N$  is given by

$$\mu_{\mathbf{x}}(\mathbf{a}) = \prod_{i \neq j} \left( \frac{\nu_f^N \varphi_R(|x^i - x^j|)}{\nu_f^N \varphi_R(|x^i - x^j|) + \nu_d} \delta_{a_{ij}-1} + \left[ 1 - \frac{\nu_f^N \varphi_R(|x^i - x^j|)}{\nu_f^N \varphi_R(|x^i - x^j|) + \nu_d} \right] \delta_{a_{ij}} \right), \quad (4)$$

for  $\mathbf{a} = \{a_{ij}\} \in \mathcal{A}^N$  (as customary,  $\delta_a$  is the Kronecker delta, i.e  $\delta_a = 1$  if  $a = 0$  and it is equal to zero otherwise).

According to the classical averaging paradigm, as  $\varepsilon \rightarrow 0$  the fast-slow system (1a)-(1b) converges to the so-called *averaged dynamics* (also often referred to as *effective dynamics*); the evolution equation for the averaged dynamics  $\{\bar{X}_t^{i,N}\}_{1 \leq i \leq N}$  is formally obtained by taking the

average of drift and diffusion coefficients of (1a) with respect to the measure  $\mu_{\mathbf{x}}$ . Because, given  $X_t^{N,\varepsilon}$ , we have

$$\mathbb{E}_{\mu_{X_t^{N,\varepsilon}}} \left[ \sum_{j \neq i} A_{ij}^{N,\varepsilon}(t) K(X_t^{j,N,\varepsilon} - X_t^{i,N,\varepsilon}) \right] = \sum_{j \neq i} \mathbb{E}_{\mu_{X_t^{N,\varepsilon}}} \left[ A_{ij}^{N,\varepsilon} \right] K(X_t^{j,N,\varepsilon} - X_t^{i,N,\varepsilon}) \quad (5)$$

$$= \sum_{j \neq i} \frac{\nu_f^N \varphi_R(|X_t^{i,N,\varepsilon} - X_t^{j,N,\varepsilon}|)}{\nu_f^N \varphi_R(|X_t^{i,N,\varepsilon} - X_t^{j,N,\varepsilon}|) + \nu_d} K(X_t^{j,N,\varepsilon} - X_t^{i,N,\varepsilon}), \quad (6)$$

the resulting equation for  $\bar{X}_t^{i,N}$  is

$$d\bar{X}_t^{i,N} = -\nabla V(\bar{X}_t^{i,N})dt + \sum_{j \neq i} \frac{\nu_f^N \varphi_R(|\bar{X}_t^{i,N} - \bar{X}_t^{j,N}|)}{\nu_f^N \varphi_R(|\bar{X}_t^{i,N} - \bar{X}_t^{j,N}|) + \nu_d} K(\bar{X}_t^{j,N} - \bar{X}_t^{i,N})dt + \sqrt{2D} dB_t^i \quad (7)$$

$$= -\nabla V(\bar{X}_t^{i,N})dt + \sum_{j \neq i} \bar{K}^N(\bar{X}_t^{j,N} - \bar{X}_t^{i,N})dt + \sqrt{2D} dB_t^i, \quad (8)$$

with initial conditions  $\bar{X}_0^{i,N} = X_0^{i,N,\varepsilon}$  for  $i \in \{1, \dots, N\}$  and having set

$$\bar{K}^N(x^j - x^i) := \frac{\nu_f^N \varphi_R(|x^i - x^j|)}{\nu_f^N \varphi_R(|x^i - x^j|) + \nu_d} K(x^j - x^i). \quad (9)$$

For brevity we will later on use a more compact notation for the whole drift of the averaged equation, namely

$$\bar{\mathcal{F}}^i(\mathbf{x}) := -\nabla V(x^i) + \sum_{j \neq i} \bar{K}^N(x^j - x^i). \quad (10)$$

We now assume the following scaling of  $\nu_f^N$  with  $N$ :

$$\nu_f^N = \nu_f / N, \quad (11)$$

where  $\nu_f > 0$  is a constant (independent of  $N, \varepsilon$  and  $i, j$ ). This choice of scaling for  $\nu_f$  ensures that the average number of links per particle is at most of order  $N \frac{\nu_f^N}{\nu_d + \nu_f^N} = O(1)$ . That is, heuristically, each particle interacts with  $O(1)$  particles (on average), so that the interaction graph is sparse. Equation (7) has now the canonical form of interacting diffusions considered e.g. in [32]. As  $N \rightarrow \infty$  the particles become independent of each other and, in the limit, each of them evolves according to the following  $n$ -dimensional non-linear diffusion (here the non-linearity is *in the sense of McKean*)

$$dX_t = -\nabla V(X_t)dt + (\bar{K} * \rho_t)(X_t)dt + \sqrt{2D}dB_t. \quad (12)$$

In the above  $*$  denotes convolution (over  $\mathbb{R}^n$ ), the function  $\bar{K} : \mathbb{R}^n \rightarrow \mathbb{R}^n$  is defined as

$$\bar{K}(x) = \frac{\nu_f}{\nu_d} \varphi_R(|x|) K(x), \quad x \in \mathbb{R}^n, \quad (13)$$

and  $\rho_t : \mathbb{R}^n \rightarrow \mathbb{R}$  is the (density of the) law of the process  $X_t$  itself; such a function solves the following non-linear Fokker-Planck equation:

$$\begin{aligned} \partial_t \rho_t(x) &= \nabla \cdot (\rho_t(x) \nabla V(x)) - \nabla \cdot (\rho_t(x) (\bar{K} * \rho_t)(x)) + D \Delta \rho_t(x), \\ \rho_t(x)|_{t=0} &= \rho_0(x), \end{aligned} \quad (14)$$

Note that the expression of the kernel  $\bar{K}$  can be formally obtained by taking sums over  $j$  in (9), using (11) and letting  $N \rightarrow \infty$ . Equations such as (14) find applications in many areas of natural sciences, and have a rich phenomenology (see [15] for a textbook reference). It is worth

mentioning that if  $\bar{K}$  is the gradient of some function, then (14) has a gradient flow structure, a very helpful fact to elucidate its qualitative behavior.

Our main result can then be stated as follows – the hypothesis under which such a result holds are detailed in Section 2.3.

**Theorem 2.1.** *With the notation introduced so far, let the potential  $V$  and force kernel  $K$  satisfy Hypothesis [H.1]. Let  $\hat{\mu}_{X_t^{N,\varepsilon}}$  be the empirical measure of the particle system  $X_t^{i,N,\varepsilon}$  evolving according to (1a)-(1b), namely*

$$\hat{\mu}_{X_t^{N,\varepsilon}}(x) := \frac{1}{N} \sum_{i=1}^N \delta_{X_t^{i,N,\varepsilon}}(x),$$

and let  $\rho_t$  be the solution to (14). If the initial data for the particle system (1a)-(1b) and for the PDE (14) satisfy Hypothesis [H.3] and Assumption [A.1] then,

**a):** for every  $u \in C_b^2(\mathbb{R}^n)$  (i.e. twice differentiable, bounded and with bounded first and second derivatives), there exists a constant  $C > 0$  (independent of  $\varepsilon, T$  and  $N$  but dependent on  $u$ ) such that

$$\sup_{t \in [0, T]} \mathbb{E} \left( \int_{\mathbb{R}^n} u(y) d\hat{\mu}_{X_t^{N,\varepsilon}}(y) - \int_{\mathbb{R}^n} u(y) \rho_t(y) dy \right)^2 \leq \frac{C}{N} e^{CT} + C(1+T)N\varepsilon, \quad (15)$$

for every  $T > 0$ ,  $\varepsilon > 0$  and  $N \in \mathbb{N}_+$ . Hence, as  $\varepsilon \rightarrow 0$  and  $N \rightarrow \infty$  in such a way that  $N\varepsilon \rightarrow 0$ , the particle system (1a)-(1b) converges to the solution of (14).

**b):** If, in addition, the external potential  $V$  satisfies Hypothesis [H.2] with  $\kappa_3 \geq \max(\kappa_{\text{av}}, \kappa_{\text{mf}})$  (where  $\kappa_{\text{av}}$  and  $\kappa_{\text{mf}}$  are defined in Lemma 4.4, and Lemma 5.1, respectively, and  $\kappa_3$  appears in Hypothesis [H.2]) then for every  $u \in C_b^2(\mathbb{R}^n)$  (there exists a constant  $C > 0$  (independent of  $\varepsilon$ , time and  $N$  but dependent on  $u$ ) such that

$$\sup_{t \geq 0} \mathbb{E} \left( \int_{\mathbb{R}^n} u(y) d\hat{\mu}_{X_t^{N,\varepsilon}}(y) - \int_{\mathbb{R}^n} u(y) \rho_t(y) dy \right)^2 \leq \frac{C}{N} + CN\varepsilon, \quad (16)$$

for every  $\varepsilon > 0$  and  $N \in \mathbb{N}_+$ . Hence, as  $\varepsilon \rightarrow 0$  and  $N \rightarrow \infty$  in such a way that  $N\varepsilon \rightarrow 0$ , the particle system (1a)-(1b) converges to the solution of (14) uniformly in time.

**Note 2.2.** Some comments on the above statement.

- The first part of the theorem proves that the macroscopic density  $\rho_t$ , solution of (14), is a good approximation of the empirical density of the particles as soon as  $N$  is large and  $N\varepsilon$  is small, but only on a finite time horizon. Strengthening this result to the *uniform in time* estimate (16) requires the strong confining properties stated in Hypothesis [H.2]. To prove Theorem 2.1, we need to combine averaging ( $\varepsilon \rightarrow 0$ ) and many-particles ( $N \rightarrow \infty$ ) estimates. In both (15) and (16), the first error term on the RHS comes from the many-particles estimate, and the second from the averaging procedure. In the uniform in time estimate (16), the requirements  $\kappa_3 \geq \kappa_{\text{av}}$  and  $\kappa_3 \geq \kappa_{\text{mf}}$  come respectively from the averaging and many-particles steps. One of the main difficulties in the proof is to track precisely the  $N$ -dependence of the averaging estimates. As already emphasized in the introduction, the  $\varepsilon$  and  $N$  limits are taken together. We further explain why at the end of Section 3.
- To further elaborate on the significance of the uniform in time result, let us point out that a long-standing criticism of averaging techniques is the following: while one can typically only prove that the averaged dynamics is a good approximation of the original slow-fast system for *finite-time* windows, the averaged dynamics is often used in practice to make predictions about the *long-time* behaviour of the slow-fast system. The fact that the averaged dynamics is a good approximation of the slow-fast system also for long times



is typically only conjectured (in specific cases) on the basis of numerical evidence. Part b of Theorem 2.1 allows to avoid this pitfall.

- As already mentioned in the introduction, from a technical point of view the idea used to obtain uniform in time estimates is borrowed from [9] and it requires proving exponentially fast decay (in time) for the space-derivatives of the Markov Semigroup associated to (7), i.e. estimates of the type (41). Such estimates, which are related to those studied in [7, 10, 14], are different from the ones usually appearing in the literature in two respects: first, similarly to [7, 10], these are not smoothing-type estimates; second, and further to [7, 10, 14], the dimension-dependence (i.e. dependence in  $N$ ) of the constants appearing in them is explicit. To further explain the former point, note that in contrast to (41), smoothing estimates for a given Markov Semigroup <sup>1</sup>  $\mathcal{P}_t$  are, generally, of the form

$$|\mathfrak{D}\mathcal{P}_t f(x)| \leq \frac{1}{t^\gamma} \|f\|_\infty, \quad f \in C_b,$$

where  $\mathfrak{D}$  is some appropriate differential operator and  $\gamma > 0$  depends on  $\mathfrak{D}$  (for example, if the semigroup is elliptic one can take  $\mathfrak{D}$  to be the usual gradient and  $\gamma = 1/2$ , if the process is hypoelliptic  $\gamma$  will depend on the number of commutators needed to obtain the direction  $\mathfrak{D}$ ), see [1] and references therein for a comprehensive review. Note that in the above  $f \in C_b$  while in (41) we take  $f \in C_b^2$ ; smoothing estimates can be seen as quantifying the “explosion” of heat-type semigroups as  $t \rightarrow 0$  so, in a way, they are more meaningful for  $t$  small. Here we want a specific (i.e. exponential) quantitative estimate for  $t$  large. Note that by the semigroup property the short and long-time estimates could be “glued” together, and this is routinely done in the literature; we do not do it here as the smoothing effect is not the main concern of the paper, what we are interested in is the long-time regime. Moreover, this would imply having to trace the  $N$  dependence in the short-time estimate as well, which is not central to the scope of this paper.

- As a bi-product of the above observation, because in (15) and (16) we take  $f \in C_b^2$  rather than  $f \in C_b$ , from our result we cannot deduce weak convergence in  $C([0, T]; \mathbb{P}r)$ , where  $\mathbb{P}r$  is the space of probability measures; hence, in the statement of the theorem, by *convergence* we mean convergence in the sense of (15) (and (16)).
- The theorem is meaningful in the regime  $N$  large,  $\varepsilon$  small and  $N\varepsilon$  small. This may be expected from (11): the rate of the Poisson process creating the links is of order  $(N\varepsilon)^{-1}$ ; good averaging estimates then require  $N\varepsilon \ll 1$ .

**2.3. Setting, Notation and Assumptions.** Because our process (1a)-(1b) evolves in  $\mathbb{R}^{Nn} \times \mathcal{A}^N$ , we will deal with vectors  $\mathbf{x} \in \mathbb{R}^{Nn}$  – the boldfont letter is reserved for elements of  $\mathbb{R}^{Nn}$ . In particular,  $\mathbf{x} = (x^1, \dots, x^N)$ , where  $x^j \in \mathbb{R}^n$  for every  $j \in 1, \dots, N$ . Elements  $\mathbf{a} \in \mathcal{A}^N$  will be regarded as arrays with components  $\mathbf{a} = (a_{12}, a_{13}, \dots, a_{NN})$ . As customary,  $C_b^k(\mathbb{R}^d)$  will denote the set of real-valued functions which are  $k$ -times continuously differentiable, bounded and with derivatives up to order  $k$  all bounded. Regarding initial data, we make the following standing assumption.

**[A.1]** We assume that the initial datum of (14), i.e.  $\rho_0(x)$ , is a smooth probability density on  $\mathbb{R}^n$  with finite moments of order two. The initial data for (1a)-(1b) are as follows:  $\{X_0^{i,N,\varepsilon}\}_{1 \leq i \leq N}$ , are i.i.d. random variables,  $X_0^{i,N,\varepsilon} \sim \rho_0$  for every  $1 \leq i \leq N$ ; the initial distribution of the links is i.i.d. as well,  $A_{i,j}^{N,\varepsilon}(0) \sim \pi_a$  for every  $1 \leq i < j \leq N$ , where  $\pi_a$  is any probability distribution on  $\{0, 1\}$ . The initial distribution of the particles,  $\rho_0$ , is independent of the initial distribution of the links,  $\pi_a$ , and it is also independent of every other source of randomness.

---

<sup>1</sup>The precise definition of Markov Semigroup is given in Section 2.3 below.

The dynamics (1a)-(1b) contains four sources of randomness: the Brownian motions  $B(t) = (B^1(t), \dots, B^N(t))$ , the Poisson processes driving formation and destruction of links, the initial distribution of the particles  $\rho_0$ , and the initial distribution of the links  $\pi_a$ . In the proofs below, we will first consider a fixed initial datum  $X_0^{N,\varepsilon} = \mathbf{x}$ ,  $A^{N,\varepsilon}(0) = \mathbf{a}$  and then, at the latest possible moment, we take expectation with respect to the noise in the initial data, i.e. with respect to  $\mathbf{x} = (x^1, \dots, x^N)$  and  $\mathbf{a} = (a_{11}, a_{12}, \dots, a_{NN})$  with  $\mathbf{x} \sim \Pi_{i=1}^N \rho_0$  and  $\mathbf{a} \sim \Pi_{i,j=1}^N \pi_a$ . Moreover,

- $\mathbb{E}_\pi$  will denote expectation with respect to both  $\mathbf{x} \sim \Pi_{i=1}^N \rho_0$  and  $\mathbf{a} \sim \Pi_{i,j=1}^N \pi_a$ ;
- $\mathbb{E}_B$  will denote expectation with respect to all the sources of noise except those in the initial data (of both the particles and the links);
- $\mathbb{E}$  is expectation with respect to all the sources of noise.

With this notation in mind, let  $\bar{\mathcal{P}}_t^N : C_b^2(\mathbb{R}^{Nn}) \rightarrow C_b^2(\mathbb{R}^{Nn})$ <sup>2</sup> be the semigroup associated with the SDE (7), namely

$$(\bar{\mathcal{P}}_t^N f)(\mathbf{x}) := \mathbb{E}_B[f(\bar{X}_t^N) | \bar{X}_0^N = \mathbf{x}];$$

where there is no risk of confusion, we will use interchangeably the notation  $\bar{\mathcal{P}}_t^N f(\mathbf{x}) = (\bar{\mathcal{P}}_t^N f)(\mathbf{x})$ . It is well-known that such a semigroup is a classical solution to the PDE

$$\begin{cases} \partial_t(\bar{\mathcal{P}}_t^N f)(\mathbf{x}) = \bar{\mathcal{L}}^N \bar{\mathcal{P}}_t^N f(\mathbf{x}) \\ (\bar{\mathcal{P}}_0^N f)(\mathbf{x}) = f(\mathbf{x}) \end{cases} \quad (17)$$

where  $\bar{\mathcal{L}}^N$  is the generator of the diffusion  $\bar{X}_t^N$ , i.e. the second order differential operator formally acting on smooth functions as

$$(\bar{\mathcal{L}}^N f)(\mathbf{x}) := \sum_{i=1}^N [\bar{\mathcal{F}}^i(\mathbf{x}) \cdot \nabla_i f(\mathbf{x})] + D \operatorname{Tr}(\operatorname{Hess} f)(\mathbf{x}), \quad (18)$$

where  $\operatorname{Tr}(\operatorname{Hess} f)$  is simply the Trace of the  $Nn \times Nn$  Hessian matrix of the function  $f$  and  $\nabla_i$  is the  $n$ -dimensional gradient with respect to the component  $x^i$  of  $\mathbf{x}$ . Analogously, the process  $(X_t^{N,\varepsilon}, A^{N,\varepsilon}(t))$  solution of (1a)-(1b) generates a semigroup  $\mathcal{P}_t^{N,\varepsilon} : C_b^2(\mathbb{R}^{nN} \times \mathcal{A}^N) \rightarrow C_b^2(\mathbb{R}^{nN} \times \mathcal{A}^N)$ , namely

$$(\mathcal{P}_t^{N,\varepsilon} h)(\mathbf{x}, \mathbf{a}) := \mathbb{E}_B \left[ h(X_t^{N,\varepsilon}, A^{N,\varepsilon}(t)) | (X_0^{N,\varepsilon}, A^{N,\varepsilon}(0)) = (\mathbf{x}, \mathbf{a}) \right].$$

Because we want to compare the dynamics  $\bar{X}_t^N$  with the dynamics  $X_t^{N,\varepsilon}$  (but  $X_t^{N,\varepsilon}$  alone does not generate a semigroup), for technical convenience we will restrict our attention to the case in which the functions  $h$  on which  $\mathcal{P}_t^{N,\varepsilon}$  acts are constant in the variable  $\mathbf{a}$ , i.e. to the case in which  $h$  does not depend on  $\mathbf{a}$  and for such functions, coherently with the above, we use the letter  $f$ . Note that while  $f$  acts only on the variable  $\mathbf{x}$ , the function  $(\mathcal{P}_t^{N,\varepsilon} f)(\mathbf{x}, \mathbf{a})$  depends on both variables. In other words, we will restrict to considering initial value problems where the initial profile is a function independent of  $\mathbf{a}$ :

$$\begin{cases} \partial_t(\mathcal{P}_t^{N,\varepsilon} f)(\mathbf{x}, \mathbf{a}) = \mathcal{L}^{N,\varepsilon}(\mathcal{P}_t^{N,\varepsilon} f)(\mathbf{x}, \mathbf{a}) \\ (\mathcal{P}_0^{N,\varepsilon} f)(\mathbf{x}, \mathbf{a}) = f(\mathbf{x}). \end{cases} \quad (19)$$

The generator  $\mathcal{L}^{N,\varepsilon}$  is the operator (formally) defined on sufficiently smooth functions  $h : \mathbb{R}^{nN} \times \mathcal{A}^N \rightarrow \mathbb{R}$  as

$$\mathcal{L}^{N,\varepsilon} h := \mathcal{L}_S h + \frac{1}{\varepsilon} \mathcal{L}_F h, \quad (20)$$

<sup>2</sup>The semigroup is usually defined on the set of continuous and bounded functions. As mentioned in Note 2.2, here we are not interested in smoothing results so, to avoid further technical complications, we just work with  $C_b^2$  functions.

where  $\mathcal{L}_S$  and  $\mathcal{L}_F$  are the parts of the generator corresponding to the slow and to the fast dynamics, respectively (both  $\mathcal{L}_S$  and  $\mathcal{L}_F$  should be denoted by  $\mathcal{L}_S^{N,\varepsilon}$  and  $\mathcal{L}_F^{N,\varepsilon}$ , we don't do this to avoid cumbersome formulas later on); namely,

$$\begin{aligned} (\mathcal{L}_S h)(\mathbf{x}, \mathbf{a}) &:= \sum_{i=1}^N \left[ \left( -\nabla V(x^i) + \sum_{j \neq i} a_{ij} K(x^j - x^i) \right) \cdot \nabla_i h(\mathbf{x}, \mathbf{a}) + D \operatorname{Tr}(\operatorname{Hess}_{\mathbf{x}} h(\mathbf{x}, \mathbf{a})) \right] \\ &:= \sum_{i=1}^N [\mathcal{F}^i(\mathbf{x}, \mathbf{a}) \cdot \nabla_i h(\mathbf{x}, \mathbf{a}) + D \operatorname{Tr}(\operatorname{Hess}_{\mathbf{x}} h(\mathbf{x}, \mathbf{a}))] \end{aligned} \quad (21)$$

having set

$$\mathcal{F}^i(\mathbf{x}, \mathbf{a}) := -\nabla V(x^i) + \sum_{j \neq i} a_{ij} K(x^j - x^i),$$

and

$$(\mathcal{L}_F h)(\mathbf{x}, \mathbf{a}) = h_d(\mathbf{x}, \mathbf{a}) + h_f(\mathbf{x}, \mathbf{a}), \quad (22)$$

where

$$h_d(\mathbf{x}, \mathbf{a}) := \nu_d \sum_{(ij): i < j} h_d^{(ij)}(\mathbf{x}, \mathbf{a}), \quad (23)$$

$$h_f(\mathbf{x}, \mathbf{a}) := \nu_f^N \sum_{(ij): i < j} \varphi_R(|x^i - x^j|) h_f^{(ij)}(\mathbf{x}, \mathbf{a}), \quad (24)$$

$$h_d^{(ij)}(\mathbf{x}, \mathbf{a}) := \left[ h(\mathbf{x}, a_{12}, \dots, \underbrace{0}_{ij}, \dots, a_{NN}) - h(\mathbf{x}, \mathbf{a}) \right], \quad (25)$$

$$h_f^{(ij)}(\mathbf{x}, \mathbf{a}) := \left[ h(\mathbf{x}, a_{12}, \dots, \underbrace{1}_{ij}, \dots, a_{NN}) - h(\mathbf{x}, \mathbf{a}) \right]. \quad (26)$$

Note that the operator  $\mathcal{L}_S$  acts on the variable  $\mathbf{x}$  only, while the operator  $\mathcal{L}_F$  acts on the variable  $\mathbf{a}$  only. In each addend of line (25) (line (26), respectively), the component  $a_{ij}$  of  $\mathbf{a}$  has been replaced by 0 (1, respectively) and we have used interchangeably the notation  $h(\mathbf{x}, \mathbf{a})$  and  $h(\mathbf{x}, a_{12}, \dots, a_{NN})$ . It remains understood that  $a_{ij} = a_{ji}$  so in (25) and (26), we effectively set to zero (or to one) the value of two components.

Let us now move on to stating the following assumptions for the kernel  $K$ , the confinement potential  $V$  and the initial data.

**[H.1]** The interaction force  $K : \mathbb{R}^n \rightarrow \mathbb{R}^n$  and the external potential  $V : \mathbb{R}^n \rightarrow \mathbb{R}$  are such that

**[H1.1]**  $K \in C^2(\mathbb{R}^n)$  and its first and second derivatives are bounded.

**[H1.2]**  $V \in C^3(\mathbb{R}^n)$  and its second and third derivatives are bounded.

**[H.2]**  $V$  is strongly convex:

$$\operatorname{Hess}V(x) \geq \kappa_3 I_n,$$

where  $\kappa_3 > 0$  will be chosen to be sufficiently large (see Theorem 2.1),  $\operatorname{Hess}V$  is the Hessian matrix of  $V$ , and the inequality is in the sense of quadratic forms.

**[H.3]** The initial data  $X_0^{N,\varepsilon} = \mathbf{x} \in \mathbb{R}^{nN}$ ,  $A^N(0) = \mathbf{a} \in \mathcal{A}^N$  of system (1a)-(1b) are as follows:

- If the interaction force  $K$  satisfies **[H.1]** and is unbounded then we assume that there exists a function  $g \in C^2(\mathbb{R}^n)$ ,  $g(x) \geq \max\{1, |K(x)|, |x|\}$ , and with first and second bounded derivatives such that, after setting

$$\psi^i(\mathbf{x}, \mathbf{a}) := \sum_{j:j \neq i} a_{ij} g(x^j - x^i) \quad (27)$$

we assume

$$\mathbb{E}_\pi \mathcal{I}_1^N(\mathbf{x}, \mathbf{a}) := \mathbb{E}_\pi \frac{1}{N} \sum_{i=1}^N \psi^i(\mathbf{x}, \mathbf{a}) < C, \quad (28)$$

and

$$\mathbb{E}_\pi \mathcal{I}_2^N(\mathbf{x}, \mathbf{a}) := \frac{1}{N} \mathbb{E}_\pi \sum_{i=1}^N [\psi^i(\mathbf{x}, \mathbf{a})]^2 < C, \quad (29)$$

where  $C > 0$  is a constant independent of  $N$  (and the notation  $\mathbb{E}_\pi$  has been introduced after Assumption **[A.1]**).

- if  $K$  satisfies **[H.1]** and is bounded then we assume that (28) and (29) are satisfied with  $g \equiv 1$ .

**Note 2.3.** Hypothesis **[H.1]** is (more than) sufficient to ensure that system (1a) -(1b) has a strong solution. Hypothesis **[H.2]** will be invoked to obtain uniform in time results. Hypothesis **[H.3]** and **[A.1]** are assumptions on the initial conditions. Note that if  $g \equiv 1$  then (27) boils down to assuming that at time  $t = 0$  each particle has a number of links which is on average of order one, i.e.

$$\frac{1}{N} \mathbb{E}_\pi \sum_{i=1}^N \sum_{j:j \neq i} A_{ij}(0) \leq C,$$

where  $C > 0$  is a constant independent of  $N$ . In general one should think of  $g(x)$  as being  $g(x) = \max(1, |x|)$ . In this case (28) implies both that the average number of links per particle is of order 1, and that there are not too many links between “far away” particles. If (28) is needed to control the growth in  $N$  of the expected value of  $\psi^i$ , assumption (29) is needed to control the growth in  $N$  of the second moment of  $\psi^i$ .

When  $K$  is unbounded, a crucial step in the proof of Theorem 4.1 consists of controlling the probability of two far away particles being connected by a link. Lemma 4.2 and Lemma 4.3 address this issue. See also Note 4.5 and the comments after the statement of Lemma 4.2 on the significance of assumption **[H.3]**.

### 3. PROOF OF THEOREM 2.1

Let us consider the LHS of (15), namely

$$\begin{aligned} E &:= \mathbb{E} \left( \int_{\mathbb{R}^n} u(y) d\hat{\mu}_{X_t^{N,\varepsilon}}(y) - \int_{\mathbb{R}^n} u(y) \rho_t(y) dy \right)^2 \\ &= \frac{1}{N^2} \mathbb{E} \sum_{i,j} u(X_t^{i,N,\varepsilon}) u(X_t^{j,N,\varepsilon}) + \left( \int_{\mathbb{R}^n} u(y) \rho_t(y) dy \right)^2 \\ &\quad - \frac{2}{N} \left( \int_{\mathbb{R}^n} u(y) \rho_t(y) dy \right) \mathbb{E} \sum_i u(X_t^{i,N,\varepsilon}). \end{aligned} \quad (30)$$

We now split the above expression introducing terms which contain the dynamics of the particles which evolve according to the averaged dynamics (7):

$$E = E^{\text{av}} + E^{\text{part}}$$

with

$$\begin{aligned} E^{\text{av}} &:= \frac{1}{N^2} \mathbb{E} \sum_{i,j} \left( u(X_t^{i,N,\varepsilon}) u(X_t^{j,N,\varepsilon}) - u(\bar{X}_t^{i,N}) u(\bar{X}_t^{j,N}) \right) \\ &\quad - \frac{2}{N} \left( \int u(y) \rho_t(y) dy \right) \mathbb{E} \left( \sum_i^N u(X_t^{i,N,\varepsilon}) - u(\bar{X}_t^{i,N}) \right) \end{aligned} \quad (31)$$

and

$$\begin{aligned} E^{\text{part}} &:= \frac{1}{N^2} \mathbb{E} \sum_{i,j}^N u(\bar{X}_t^{i,N}) u(\bar{X}_t^{j,N}) - \frac{2}{N} \mathbb{E} \left( \int u(y) \rho_t(y) dy \right) \mathbb{E} \sum_i^N u(\bar{X}_t^{i,N}) \\ &\quad + \left( \int u(y) \rho_t(y) dy \right)^2 \\ &= \mathbb{E} \left( \frac{1}{N} \sum_i^N u(\bar{X}_t^{i,N}) - \int u(y) \rho_t(y) dy \right)^2, \end{aligned} \quad (32)$$

where  $\bar{X}_t^{i,N}$  satisfies (7). To obtain (15), we need to show

$$\sup_{t \in [0, T]} |E^{\text{av}}| \leq C(1+T)N\varepsilon. \quad (33)$$

and

$$\sup_{t \in [0, T]} |E^{\text{part}}| \leq \frac{C}{N} e^{CT}. \quad (34)$$

To obtain (16) we need to prove, under the stated additional assumptions:

$$\sup_{t \geq 0} |E^{\text{av}}| \leq CN\varepsilon \quad (35)$$

and

$$\sup_{t \geq 0} |E^{\text{part}}| \leq \frac{C}{N}. \quad (36)$$

The averaging estimates (33) and (35) are proven in Section 4; the estimates (34) and (36) are studied in Section 5. Once such estimates are shown, the proof is concluded.

Note that, once the above estimates have been proved, the  $\varepsilon \rightarrow 0$  and  $N \rightarrow \infty$  limits are taken *together*. This raises a natural question about the commutativity of these two limits. At least formally, the limit  $\varepsilon \rightarrow 0$  first and  $N \rightarrow \infty$  could be considered and would produce the non-linear Fokker-Planck equation (14). It is much less clear how, even formally, one would let  $N \rightarrow \infty$  first in this sparse graph regime (and indeed, as mentioned in the introduction, the works [2, 4, 11, 34], which contain a formal derivation by taking the  $N$  limit first, do need to make propagation-of-chaos-type ansatzes to carry out the procedure). Note indeed that, once the  $\varepsilon$  limit is taken, i.e. once the dynamics (7) is obtained, our choice of scaling of  $\nu_f^N$  (11) makes (7) treatable through mean-field arguments, as anticipated in [3]. Either way, unless the two limits are taken together, the analysis can never be revealing of the correct critical regime  $\varepsilon N \ll 1$  to be considered.

## 4. THE AVERAGING LIMIT: PROOF OF (33) AND (35)

Using the exchangeability of the particles we can rewrite (31) as

$$\begin{aligned}
E^{\text{av}} &= \frac{1}{N} \mathbb{E} \left[ u^2(X_t^{1,N,\varepsilon}) - u^2(\bar{X}_t^{1,N,\varepsilon}) \right] \\
&\quad + \frac{N-1}{N} \left( \mathbb{E} \left[ u(X_t^{1,N,\varepsilon}) u(X_t^{2,N,\varepsilon}) \right] - \mathbb{E} \left[ u(\bar{X}_t^{1,N,\varepsilon}) u(\bar{X}_t^{2,N,\varepsilon}) \right] \right) \\
&\quad - 2 \left( \int u(y) \rho_t(y) dy \right) \mathbb{E} \left[ u(X_t^{1,N,\varepsilon}) - u(\bar{X}_t^{1,N,\varepsilon}) \right] \\
&\leq \frac{C}{N} + \frac{N-1}{N} \mathbb{E}_\pi \left| (\mathcal{P}_t^{N,\varepsilon} u^{(2)})(\mathbf{x}, \mathbf{a}) - (\bar{\mathcal{P}}_t^N u^{(2)})(\mathbf{x}, \mathbf{a}) \right| \\
&\quad + C \mathbb{E}_\pi \left| (\mathcal{P}_t^{N,\varepsilon} u^{(1)})(\mathbf{x}, \mathbf{a}) - (\bar{\mathcal{P}}_t^N u^{(1)})(\mathbf{x}, \mathbf{a}) \right|,
\end{aligned} \tag{37}$$

where  $u^{(k)} : \mathbb{R}^{Nn} \rightarrow \mathbb{R}$  is defined as  $u^{(k)}(\mathbf{x}) := u(x^1) \dots u(x^k)$ ; in the last inequality we have used the boundedness of  $u$  and the fact that, by independence of the sources of noise,  $\mathbb{E} = \mathbb{E}_\pi \mathbb{E}_B$ .

The result now follows once we can control the difference between the semigroups  $\mathcal{P}_t^{N,\varepsilon}$  and  $\bar{\mathcal{P}}_t^N$ . This is the purpose of the following theorem, which is the main result of this section. Before stating and proving such a theorem let us emphasize again that throughout this section we will first consider a fixed initial datum  $X_0^{N,\varepsilon} = \mathbf{x}$ ,  $A^{N,\varepsilon}(0) = \mathbf{a}$  and then at the latest possible moment we will take expectation  $\mathbb{E}_\pi$  with respect to the noise in the initial data.

**Theorem 4.1.** *With the notation introduced so far, suppose Assumption [A.1], Hypothesis [H.1] and Hypothesis [H.3] are satisfied. Then there exists some  $\varepsilon_0$  such that for any  $t \in [0, \infty)$ , and any  $N, \varepsilon$  with  $N\varepsilon < \varepsilon_0$  we have*

$$\mathbb{E}_\pi \left| \mathcal{P}_t^{N,\varepsilon} f(\mathbf{x}, \mathbf{a}) - \bar{\mathcal{P}}_t^N f(\mathbf{x}) \right| \leq \varepsilon N C (1+t) \mathbb{E}_\pi \left( C_1 + C_3 \mathcal{I}_1^N(\mathbf{x}, \mathbf{a}) e^{-\frac{C_2 t}{\varepsilon}} + \mathcal{I}_2^N(\mathbf{x}, \mathbf{a}) e^{-\frac{C_4 t}{\varepsilon}} \right)$$

for every  $f \in C_b^2(\mathbb{R}^{nN})$  (where the constant  $C$  does not depend on  $t$ ,  $N$  or  $\varepsilon$  but it will depend on a suitable norm of  $f$ , see (38)). Moreover, if Hypothesis [H.2] holds for some  $\kappa_3 \geq \kappa_{\text{av}}$  (the latter being defined in Lemma 4.4), then the above estimate is uniform in time, i.e.

$$\mathbb{E}_\pi \left| \mathcal{P}_t^{N,\varepsilon} f(\mathbf{x}, \mathbf{a}) - \bar{\mathcal{P}}_t^N f(\mathbf{x}) \right| \leq \varepsilon N C \mathbb{E}_\pi \left( C_1 + C_3 \mathcal{I}_1^N(\mathbf{x}, \mathbf{a}) e^{-C_2 \frac{t}{\varepsilon}} + \mathcal{I}_2^N(\mathbf{x}, \mathbf{a}) e^{-C_4 \frac{t}{\varepsilon}} \right).$$

Before proving Theorem 4.1 we state three lemmata, which are key to proving Theorem 4.1. We will prove such lemmata after the proof of Theorem 4.1. The first two lemmata, Lemma 4.2 and Lemma 4.3, let us express more precisely the idea that the average number of links per particle remains of order 1, if such a property is true at time zero. Lemma 4.4 contains bounds on the derivatives of the semigroup  $\bar{\mathcal{P}}_t^N$ , which are key to obtaining uniform in time results.

**Lemma 4.2.** *With the notation introduced so far, suppose Hypotheses [H.1] and [H.3] hold. If  $N$  and  $\varepsilon$  are such that the product  $N\varepsilon$  is sufficiently small, then for any  $t \in [0, \infty)$  there exist positive constants  $C_1, C_2$  independent of  $t$  such that*

$$(\mathcal{P}_t^{N,\varepsilon} \mathcal{I}_1^N)(\mathbf{x}, \mathbf{a}) = \mathbb{E}_B \left[ \mathcal{I}_1^N(X_t^{N,\varepsilon}, A^{N,\varepsilon}(t)) \mid (X_0^{N,\varepsilon}, A^{N,\varepsilon}(0)) = (\mathbf{x}, \mathbf{a}) \right] \leq C_1 + \mathcal{I}_1^N(\mathbf{x}, \mathbf{a}) e^{-\frac{C_2 t}{\varepsilon}}.$$

According to Lemma 4.2, the quantity  $\mathbb{E}_B[\mathcal{I}_1^N(X_t^{N,\varepsilon}, A^{N,\varepsilon}(t))]$  remains of order one (in  $N$ ) uniformly in time if it is of order one at the initial time. However Lemma 4.2 says also something more: because we are working in the regime  $N\varepsilon \ll 1$ , so e.g.  $\varepsilon = 1/N^\alpha$  for some  $\alpha > 1$ , if  $\mathcal{I}_1^N$  is large in  $N$  at time zero, then such initially large value quickly decays in expectation, due to the factor  $e^{-C_2 t/\varepsilon}$ . Therefore it is reasonable to assume that the average number of links in the initial configuration is  $O(1)$  in  $N$ . Lemma 4.3 below strengthens this result by giving some control on the second moment as well.

**Lemma 4.3.** *With the same assumptions and notation as in Lemma 4.2, there exists  $\varepsilon_0 > 0$  such that if  $N\varepsilon < \varepsilon_0$  then, for any  $t \in [0, \infty)$  and for some positive constants  $C_1, C_2, C_3, C_4 > 0$  independent of  $t$ , we have*

$$\mathcal{P}_t^{N,\varepsilon} \mathcal{I}_2^N(\mathbf{x}, \mathbf{a}) \leq C_1 + C_3 \mathcal{I}_1^N(\mathbf{x}, \mathbf{a}) e^{-C_2 \frac{t}{\varepsilon}} + \mathcal{I}_2^N(\mathbf{x}, \mathbf{a}) e^{-C_4 \frac{t}{\varepsilon}}.$$

Here  $\mathcal{I}_1^N$  is defined as in Lemma 4.2.

For  $f : \mathbb{R}^{Nn} \rightarrow \mathbb{R}$ ,  $f \in C_b^2(\mathbb{R}^{Nn})$ , we define the norm

$$\|f\|_{C_b^2(\mathbb{R}^{Nn})}^2 := \|f\|_\infty^2 + \sum_{i=1}^N \|\nabla_i f\|_\infty^2 + \sum_{i,j=1}^N \|\nabla_{ij}^2 f\|_\infty^2, \quad (38)$$

and seminorm

$$[[f]]_{C_b^2(\mathbb{R}^{Nn})}^2 := \sup_{\mathbf{x} \in \mathbb{R}^{Nn}} \left( \sum_{i=1}^N |\nabla_i f(\mathbf{x})|^2 + \sum_{i,j=1}^N |\nabla_{ij}^2 f(\mathbf{x})|^2 \right) \quad (39)$$

where  $\nabla_i f$  is the gradient with respect to the variable  $x^i \in \mathbb{R}^n$ ,  $x^i = (x_1^i, \dots, x_n^i)$ , so that  $\|\nabla_i f\|_\infty := \max_{k \in \{1, \dots, n\}} \|(\nabla_i f)_k\|_\infty$ , where  $(\nabla_i f)_k$  is the  $k$ -th component of the gradient  $\nabla_i$  and  $\|\cdot\|_\infty$  is the supremum norm in  $\mathbb{R}^n$ . Analogously,  $\nabla_{ij}^2 f$  is the  $n \times n$  matrix containing all the second derivatives with respect to the variables  $x^i, x^j \in \mathbb{R}^n$ . Let us emphasize that the dependence on  $n$  of this quantity is irrelevant to our purposes, as  $n$  is fixed; what matters here is the dependence on  $N$  so, to simplify notation, the proofs of all the lemmata will be done in the case  $n = 1$ . With this notation in place we can state the next lemma.

**Lemma 4.4.** *Suppose Hypothesis [H.1] holds. Recall the function  $\bar{K}^N : \mathbb{R}^n \rightarrow \mathbb{R}^n$  is defined in (9) and note that, as a consequence of our assumptions on  $K$  and  $\varphi_R$ , there exist positive constants  $\kappa_1, \kappa_2$  such that  $\|\nabla \bar{K}^N(z)\|_\infty \leq \kappa_1/N$  and  $\|\nabla^2 \bar{K}^N(z)\|_\infty \leq \kappa_2/N$ .<sup>3</sup> Let  $\kappa_3 \in \mathbb{R}, \kappa_4 > 0$  be constants such that  $\text{Hess}V(z) \geq \kappa_3 \text{Id}_n$  and  $\|\nabla^3 V(z)\|_\infty \leq \kappa_4$  for all  $z \in \mathbb{R}^n$ .<sup>4</sup> Then*

**i):** *There is a constant  $C > 0$  which depends only on  $\kappa_1, \kappa_2, \kappa_3$  and  $\kappa_4$  such that for all  $N \in \mathbb{N}$ ,  $f \in C_b^2(\mathbb{R}^{Nn})$  and  $t \geq 0$  we have*

$$[[\bar{\mathcal{P}}_t^N f]]_{C_b^2(\mathbb{R}^{Nn})} \leq C \|f\|_{C_b^2(\mathbb{R}^{Nn})}. \quad (40)$$

**ii):** *Moreover, if [H.2] holds with  $\kappa_3$  such that, for some  $\delta > 0$ ,*

$$\kappa_3 \geq \kappa_{\text{av}} := \frac{\delta}{2} + 2\kappa_1 + \frac{\kappa_4 + 4\kappa_2}{\delta + 4\kappa_2 + \kappa_4 + 4\kappa_1},$$

*then,*

$$[[\bar{\mathcal{P}}_t^N f]]_{C_b^2(\mathbb{R}^{Nn})} \leq C e^{-\delta t} \|f\|_{C_b^2(\mathbb{R}^{Nn})}^2 \quad (41)$$

*for some  $C > 0$  independent of  $N$  and  $t$ .*

**4.1. Heuristics: formal expansion of the semigroup  $\mathcal{P}_t^{N,\varepsilon}$ .** To motivate the structure of the proof of Theorem 4.1, we first present a heuristic computation. One can consider formally expanding the semigroup  $\mathcal{P}_t^{N,\varepsilon}$  in powers of  $\varepsilon$  (this procedure is similar in spirit to the procedure presented in [27], the only difference is that here we expand the semigroup, while in [27] the emphasis is on the expansion of the generator):

$$(\mathcal{P}_t^{N,\varepsilon} f)(\mathbf{x}, \mathbf{a}) = f_t^0(\mathbf{x}, \mathbf{a}) + \varepsilon f_t^1(\mathbf{x}, \mathbf{a}) + \dots \quad (42)$$

<sup>3</sup>Here  $\nabla$  and  $\nabla^2$  are gradient and Hessian with respect to  $z \in \mathbb{R}^n$ .

<sup>4</sup> $\nabla^3$  denotes all the third derivatives with respect to  $z \in \mathbb{R}^n$ .

Using that  $\mathcal{P}_t^{N,\varepsilon}$  solves the PDE (19), i.e.

$$\partial_t \mathcal{P}_t^{N,\varepsilon} f - \mathcal{L}^{N,\varepsilon} \mathcal{P}_t^{N,\varepsilon} f = 0,$$

and recalling that the generator  $\mathcal{L}^{N,\varepsilon}$  of  $\mathcal{P}_t^{N,\varepsilon}$  is given by (20), following [27], we can insert (42) in the above and, comparing terms containing the same power of  $\varepsilon$  we deduce:

$$O\left(\frac{1}{\varepsilon}\right) : \quad \mathcal{L}_F f_t^0 = 0 \quad (43)$$

$$O(1) : \quad \partial_t f_t^0 - \mathcal{L}_S f_t^0 = \mathcal{L}_F f_t^1 \quad (44)$$

As we have explained, if we fix  $X_t^{N,\varepsilon}$  then the dynamics of  $A^{N,\varepsilon}(t)$  is ergodic (i.e. it admits a unique invariant measure). This implies that the only solutions  $u(\mathbf{x}, \mathbf{a})$  to  $\mathcal{L}_F u = 0$  are constant in  $\mathbf{a}$ <sup>5</sup> so we have  $f_t^0(\mathbf{x}, \mathbf{a}) = f_t^0(\mathbf{x})$  from (43). Now integrating (44) with respect to  $\mu_{\mathbf{x}}$ , the unique invariant measure on  $\mathcal{A}^N$  with the  $\mathbf{x}$ -coefficient frozen, we obtain

$$\partial_t f_t^0(\mathbf{x}) - \bar{\mathcal{L}}^N f_t^0(\mathbf{x}) = 0, \quad (45)$$

where we recall that, from (10) and (18),  $\bar{\mathcal{L}}^N$  is given by

$$\begin{aligned} \bar{\mathcal{L}}^N f(\mathbf{x}) = & \sum_{i=1}^N \left[ \left( -\nabla_i V(x^i) + \sum_{j \neq i} \frac{\nu_f^N \varphi_R(|x^i - x^j|)}{\nu_f^N \varphi_R(|x^i - x^j|) + \nu_d} K(x^j - x^i) \right) \nabla_i f(\mathbf{x}) \right] \\ & + D\text{Tr}(\text{Hess}f)(\mathbf{x}). \end{aligned}$$

From (45) one must then have  $f_t^0(\mathbf{x}) = (\bar{\mathcal{P}}_t^N f)(\mathbf{x})$ , hence, from (42),

$$(\mathcal{P}_t^{N,\varepsilon} f)(\mathbf{x}, \mathbf{a}) = (\bar{\mathcal{P}}_t^N f)(\mathbf{x}) + \varepsilon f_t^1(\mathbf{x}, \mathbf{a}) + \dots \quad (46)$$

One can now solve<sup>6</sup> (44) to find  $f_t^1$ :

$$f_t^1(\mathbf{x}, \mathbf{a}) = \sum_{i=1}^N \sum_{j \neq i} \frac{1}{\nu_f^N \varphi_R(|x^i - x^j|) + \nu_d} K(x^j - x^i) a_{ij} \partial_i \bar{\mathcal{P}}_t^N f(\mathbf{x}).$$

Note that the solution to (44) is *not* unique<sup>7</sup>, however the above is one possible solution, and the one with which we will work. To simplify notation, define

$$\tilde{K}(z) = \frac{1}{\nu_f^N \varphi_R(z) + \nu_d} K(z). \quad (47)$$

(We should denote  $\tilde{K}$  by  $\tilde{K}^N$ , we don't do so to avoid cumbersome formulas later on). Now we can write

$$f_t^1(\mathbf{x}, \mathbf{a}) = \sum_{i=1}^N \sum_{j \neq i} \tilde{K}(x^j - x^i) a_{ij} \partial_i \bar{\mathcal{P}}_t^N f(\mathbf{x}). \quad (48)$$

---

<sup>5</sup>If  $\mathcal{P}_t^{A,x}$  denotes the semigroup associated to  $A$  with the  $x$ -coefficient frozen. Then let  $\mu^x$  denote the unique invariant measure of the frozen system, note that  $\mathcal{P}_t^{A,x} f$  converges to  $\int f d\mu^x$ . Now

$$\mathcal{P}_t^{A,x} f(A) = f(A) + \int_0^t \mathcal{P}_s^{A,x} \mathcal{L}_F f(A) ds.$$

Therefore, if  $f$  is a solution to  $\mathcal{L}_F f = 0$  then  $\mathcal{P}_t^{A,x} f(A) = f(A)$  and letting  $t$  tend to  $\infty$  we must have that  $f = \mu^x(f)$ , i.e.  $f$  is a constant function in the variable  $A$ .

<sup>6</sup>This solution was found by first considering the case when  $N = 2$  in which (44) reduces to two linear equations, then proposing an ansatz based on this solution and verifying that it does indeed solve (44).

<sup>7</sup>This solution is not unique since  $\mathcal{L}_F$  has a non-trivial kernel and in particular any function which does not depend on  $\mathbf{a}$  belongs to the kernel; this could be resolved by suitable boundary conditions but for our purposes we do not require the solution to be unique and will just work with the solution found in (48).



We shall later use this explicit expression for  $f_t^1$ . Before moving on to the actual proof of Theorem 4.1, let us make a remark.

**Note 4.5.** A few comments on the definition of  $f_t^1$ .

- It is clear from (46) that, in order to control the difference  $\mathcal{P}_t^{N,\varepsilon} - \bar{\mathcal{P}}_t^N$ , we need to control  $f_t^1$ ; comparing the above expression for  $f_t^1$  with the one for  $\mathcal{I}_1^N$ , one can see that the need to control  $f_t^1$  is where Lemma 4.2 and Lemma 4.3 stem from and, in turn, the reason for imposing Hypothesis **[H.3]** on the initial data. Note that if  $K$  is bounded (with bounded derivatives) then  $\tilde{K}$  is also bounded (with bounded derivatives). This relates to why **[H.3]** includes the two different cases of  $K$  being bounded/unbounded.
- If  $f_t^1$  is as in (48), then, as opposed to what it may look like by observing the heuristic expansion (42),  $\varepsilon f_t^1$  is not the whole term of order  $\varepsilon$  when we look at the difference between  $\mathcal{P}_t^{N,\varepsilon}$  and  $\bar{\mathcal{P}}_t^{N,\varepsilon}$  (cf (42) and (46)). This is in no way contradictory, because of the non-uniqueness of  $f_t^1$ . The whole term of order  $\varepsilon$  will appear during the rigorous proof, see (50) and (54) below.

We now have all the tools to prove Theorem 4.1.

#### 4.2. Proof of Theorem 4.1.

*Proof of Theorem 4.1.* To simplify notation we set the diffusion coefficient  $D = 1$  and  $n = 1$ , so that  $f \in C_b^2(\mathbb{R}^N)$  and  $K, V : \mathbb{R} \rightarrow \mathbb{R}$ . We will therefore write  $\partial_i$  instead of  $\nabla_i$  and  $V'$  denotes derivative of  $V$  with respect to its argument. Let  $f_t^1$  be the function defined in (48) and observe that, by a direct calculation, such a function is a solution to

$$\bar{\mathcal{L}}^N \bar{\mathcal{P}}_t^N f - \mathcal{L}_S \bar{\mathcal{P}}_t^N f = \mathcal{L}_F f_t^1. \quad (49)$$

Motivated by (42) and (46), we set

$$r_t^{N,\varepsilon}(\mathbf{x}, \mathbf{a}) := (\mathcal{P}_t^{N,\varepsilon} f)(\mathbf{x}, \mathbf{a}) - (\bar{\mathcal{P}}_t^N f)(\mathbf{x}) - \varepsilon f_t^1(\mathbf{x}, \mathbf{a}). \quad (50)$$

Differentiating  $r_t^{N,\varepsilon}$  with respect to time gives

$$\partial_t r_t^{N,\varepsilon}(\mathbf{x}, \mathbf{a}) = \partial_t (\mathcal{P}_t^{N,\varepsilon} f)(\mathbf{x}, \mathbf{a}) - \partial_t (\bar{\mathcal{P}}_t^N f)(\mathbf{x}) - \varepsilon \partial_t f_t^1(\mathbf{x}, \mathbf{a}).$$

Using (19) we have

$$\begin{aligned} \partial_t r_t^{N,\varepsilon}(\mathbf{x}, \mathbf{a}) &= \mathcal{L}^{N,\varepsilon} \mathcal{P}_t^{N,\varepsilon} f(\mathbf{x}, \mathbf{a}) - \partial_t \bar{\mathcal{P}}_t^N f(\mathbf{x}) - \varepsilon \partial_t f_t^1(\mathbf{x}, \mathbf{a}) \\ &\stackrel{(50)}{=} \mathcal{L}^{N,\varepsilon} r_t^{N,\varepsilon}(\mathbf{x}, \mathbf{a}) + \mathcal{L}^{N,\varepsilon} \bar{\mathcal{P}}_t^N f(\mathbf{x}) + \varepsilon \mathcal{L}^{N,\varepsilon} f_t^1(\mathbf{x}, \mathbf{a}) - \partial_t \bar{\mathcal{P}}_t^N f(\mathbf{x}) - \varepsilon \partial_t f_t^1(\mathbf{x}, \mathbf{a}). \end{aligned}$$

Since  $\bar{\mathcal{P}}_t^N f(\mathbf{x})$  does not depend on  $\mathbf{a}$  we have  $\mathcal{L}^{N,\varepsilon} \bar{\mathcal{P}}_t^N f(\mathbf{x}) = \mathcal{L}_S \bar{\mathcal{P}}_t^N f(\mathbf{x})$  and using (49) we obtain

$$\begin{aligned} \partial_t r_t^{N,\varepsilon}(\mathbf{x}, \mathbf{a}) &= \mathcal{L}^{N,\varepsilon} r_t^{N,\varepsilon}(\mathbf{x}, \mathbf{a}) - \mathcal{L}_F f_t^1(\mathbf{x}, \mathbf{a}) + \varepsilon \mathcal{L}^{N,\varepsilon} f_t^1(\mathbf{x}, \mathbf{a}) - \varepsilon \partial_t f_t^1(\mathbf{x}, \mathbf{a}) \\ &= \mathcal{L}^{N,\varepsilon} r_t^{N,\varepsilon}(\mathbf{x}, \mathbf{a}) + \varepsilon (\mathcal{L}_S f_t^1(\mathbf{x}, \mathbf{a}) - \partial_t f_t^1(\mathbf{x}, \mathbf{a})). \end{aligned}$$

Therefore, using the variation of constants formula, we have

$$r_t^{N,\varepsilon}(\mathbf{x}, \mathbf{a}) = \mathcal{P}_t^{N,\varepsilon} r_0^{N,\varepsilon}(\mathbf{x}, \mathbf{a}) + \varepsilon \int_0^t \mathcal{P}_{t-s}^{N,\varepsilon} (\mathcal{L}_S f_s^1 - \partial_s f_s^1)(\mathbf{x}, \mathbf{a}) ds.$$

The remainder of the proof is divided into 3 steps:

- Step 1: Show that

$$|\mathcal{P}_t^{N,\varepsilon} r_0^{N,\varepsilon}| \leq CN\varepsilon \left( C_1 + \mathcal{I}_1^N(\mathbf{x}, \mathbf{a}) e^{-C_2 \frac{t}{\varepsilon}} \right) \|f\|_{C_b^2(\mathbb{R}^N)} \quad (51)$$

- Step 2: Show that

$$|\mathcal{P}_{t-s}^{N,\varepsilon}(\mathcal{L}_S - \partial_s)f_s^1| \leq CN \left( C_1 + C_3 \mathcal{I}_1^N(\mathbf{x}, \mathbf{a}) e^{-C_2 \frac{t}{\varepsilon}} + \mathcal{I}_2^N(\mathbf{x}, \mathbf{a}) e^{-C_4 \frac{t}{\varepsilon}} \right) \|f\|_{C_b^2(\mathbb{R}^N)}.$$

Moreover, if ii) of Lemma 4.4 is satisfied,

$$|\mathcal{P}_{t-s}^{N,\varepsilon}(\mathcal{L}_S - \partial_s)f_s^1| \leq CN e^{-\delta s} \left( C_1 + C_3 \mathcal{I}_1^N(\mathbf{x}, \mathbf{a}) e^{-\frac{C_2 t}{\varepsilon}} + \mathcal{I}_2^N(\mathbf{x}, \mathbf{a}) e^{-\frac{C_4 t}{\varepsilon}} \right) \|f\|_{C_b^2(\mathbb{R}^N)}.$$

for some  $\delta > 0$ .

- Step 3: Show that

$$|f_t^1(\mathbf{x}, \mathbf{a})| \leq CN \mathcal{I}_1^N(\mathbf{x}, \mathbf{a}) \|f\|_{C_b^2(\mathbb{R}^N)}. \quad (52)$$

The above estimates are valid for any  $t \in [0, \infty)$  and for constants  $C, C_1, \dots, C_4$  that are independent of  $t$ . This completes the proof since Step 1 and Step 2 imply

$$|r_t^{N,\varepsilon}(\mathbf{x}, \mathbf{a})| \leq C\varepsilon N(1+t) \left( C_1 + (1+C_3)\mathcal{I}_1^N(\mathbf{x}, \mathbf{a}) e^{-C_2 \frac{t}{\varepsilon}} + \mathcal{I}_2^N(\mathbf{x}, \mathbf{a}) e^{-C_4 \frac{t}{\varepsilon}} \right) \|f\|_{C_b^2(\mathbb{R}^N)}. \quad (53)$$

Then, by (53), (50) and Step 3, we have

$$\begin{aligned} & \left| \mathcal{P}_t^{N,\varepsilon} f(\mathbf{x}, \mathbf{a}) - \bar{\mathcal{P}}_t^N f(\mathbf{x}) \right| \\ &= |\varepsilon f_t^1(\mathbf{x}, \mathbf{a}) + r_t^{N,\varepsilon}(\mathbf{x}, \mathbf{a})| \\ &\leq CN\varepsilon(1+t) \left( C_1 + (1+C_3)\mathcal{I}_1^N(\mathbf{x}, \mathbf{a}) e^{-C_2 \frac{t}{\varepsilon}} + \mathcal{I}_2^N(\mathbf{x}, \mathbf{a}) e^{-C_4 \frac{t}{\varepsilon}} \right) \|f\|_{C_b^2(\mathbb{R}^N)}. \end{aligned}$$

Moreover, if ii) of Lemma 4.4 is satisfied, from Step 1 and Step 2 we also have

$$|r_t^{N,\varepsilon}(\mathbf{x}, \mathbf{a})| \leq \frac{C\varepsilon N}{\delta} \left( C_1 + (1+C_3)\mathcal{I}_1^N(\mathbf{x}, \mathbf{a}) e^{-C_2 \frac{t}{\varepsilon}} + \mathcal{I}_2^N(\mathbf{x}, \mathbf{a}) e^{-C_4 \frac{t}{\varepsilon}} \right) \|f\|_{C_b^2(\mathbb{R}^N)}. \quad (54)$$

Hence,

$$\left| \mathcal{P}_t^{N,\varepsilon} f(\mathbf{x}, \mathbf{a}) - \bar{\mathcal{P}}_t^N f(\mathbf{x}) \right| \leq CN\varepsilon \left( C_1 + (1+C_3)\mathcal{I}_1^N(\mathbf{x}, \mathbf{a}) e^{-\frac{C_2 t}{\varepsilon}} + \mathcal{I}_2^N(\mathbf{x}, \mathbf{a}) e^{-\frac{C_4 t}{\varepsilon}} \right) \|f\|_{C_b^2(\mathbb{R}^N)}.$$

To prove Step 1, note that from (19) we have  $r_0^{N,\varepsilon}(\mathbf{x}, \mathbf{a}) = f(\mathbf{x}) - f(\mathbf{x}) - \varepsilon f_0^1(\mathbf{x}, \mathbf{a}) = -\varepsilon f_0^1(\mathbf{x}, \mathbf{a})$ . From (48) and the definition of  $\psi^i$  (see (27)) we get

$$|\mathcal{P}_t^{N,\varepsilon} r_0^{N,\varepsilon}(\mathbf{x}, \mathbf{a})| = \varepsilon |\mathcal{P}_t^{N,\varepsilon} f_0^1(\mathbf{x}, \mathbf{a})| \leq C\varepsilon \sum_{i=1}^N (\mathcal{P}_t^{N,\varepsilon} \psi^i)(\mathbf{x}, \mathbf{a}) \|\partial_i f\|_\infty \quad (55)$$

$$\stackrel{(38)}{\leq} CN\varepsilon \|f\|_{C_b^2(\mathbb{R}^N)} (\mathcal{P}_t^{N,\varepsilon} \mathcal{I}_1^N)(\mathbf{x}, \mathbf{a}). \quad (56)$$

Here  $C$  is a positive constant that varies line by line but is independent of  $N, \varepsilon$  and  $\mathbf{a}$ . By Lemma 4.2 and the above one obtains (51), hence we have proven Step 1.

To prove Step 2, we apply the operator  $(\mathcal{L}_S - \partial_s)$  to the definition of  $f_s^1$  (see (48)).

$$\begin{aligned} (\mathcal{L}_S - \partial_s)f_s^1 &= \sum_{i=1}^N \sum_{j \neq i} (\mathcal{L}_S - \partial_s) \left( a_{ij} \tilde{K}(x^j - x^i) \partial_i (\bar{\mathcal{P}}_s^N f)(\mathbf{x}) \right) \\ &= \sum_{i=1}^N \sum_{j \neq i} \mathcal{L}_S(\tilde{K}(x^j - x^i)) a_{ij} \partial_i \bar{\mathcal{P}}_s^N f(\mathbf{x}) \end{aligned} \quad (57)$$

$$\begin{aligned} &+ 2 \sum_{i=1}^N \sum_{j \neq i} \sum_{k=1}^N \partial_k(\tilde{K}(x^j - x^i)) a_{ij} \partial_k \partial_i \bar{\mathcal{P}}_s^N f(\mathbf{x}) \\ &+ \sum_{i=1}^N \sum_{j \neq i} \tilde{K}(x^j - x^i) a_{ij} (\mathcal{L}_S - \partial_s) (\partial_i \bar{\mathcal{P}}_s^N f(\mathbf{x})) \end{aligned} \quad (58)$$

We study the three addends on the RHS of the above; to this end, we start by considering  $\mathcal{L}_S(\tilde{K}(x^j - x^i))$  (contained in the first addend):

$$\begin{aligned} \mathcal{L}_S(\tilde{K}(x^j - x^i)) &= \sum_{k=1}^N \sum_{\ell \neq k} K(x^\ell - x^k) a_{k\ell} \partial_k(\tilde{K}(x^j - x^i)) + \sum_{k=1}^N \partial_{kk}^2(\tilde{K}(x^j - x^i)) \\ &\quad - \sum_{k=1}^N V'(x^k) \partial_k(\tilde{K}(x^j - x^i)) \\ &= - \sum_{\ell \neq i} K(x^\ell - x^i) a_{i\ell} \tilde{K}'(x^j - x^i) + \tilde{K}''(x^j - x^i) + \tilde{K}'''(x^j - x^i) \\ &\quad + \sum_{\ell \neq j} K(x^\ell - x^j) a_{j\ell} \tilde{K}'(x^j - x^i) + V'(x^i) \tilde{K}'(x^j - x^i) - V'(x^j) \tilde{K}'(x^j - x^i) \\ &\leq C(1 + \sum_{\ell \neq i} a_{i\ell} |K(x^\ell - x^i)| + \sum_{\ell \neq j} a_{j\ell} |K(x^j - x^\ell)|) + C|x^i - x^j|. \end{aligned}$$

To obtain the last line we have used that  $V'$  is Lipschitz and that  $K'$  and  $K''$  are bounded (hence the boundedness of  $\tilde{K}'$  and  $\tilde{K}''$ ). Recalling that  $g$  is as defined in **[H.3]**, we then have

$$\mathcal{L}_S(\tilde{K}(x^j - x^i)) \leq C(1 + \psi^i(\mathbf{x}, \mathbf{a}) + \psi^j(\mathbf{x}, \mathbf{a})) + Cg(x^i - x^j). \quad (59)$$

Next, we consider the term  $(\mathcal{L}_S - \partial_s) (\partial_i \bar{\mathcal{P}}_s^N f(\mathbf{x}))$  (contained in the third addend of (57)). Note first that, from (9) and (47), we have

$$\bar{K}^N(z) = \nu_f^N \varphi_R(z) \tilde{K}(z). \quad (60)$$

Hence,

$$\begin{aligned}
& (\mathcal{L}_S - \partial_s) (\partial_i \bar{\mathcal{P}}_s^N f(\mathbf{x})) \\
&= (\mathcal{L}_S \partial_i - \partial_i \bar{\mathcal{L}}^N) (\bar{\mathcal{P}}_s^N f(\mathbf{x})) \\
&= \sum_{k=1}^N \sum_{\ell \neq k} K(x^\ell - x^k) a_{k\ell} \partial_k \partial_i \bar{\mathcal{P}}_s^N f(\mathbf{x}) - \sum_{k=1}^N V'(x^k) \partial_k \partial_i \bar{\mathcal{P}}_s^N f(\mathbf{x}) \\
&\quad - \sum_{k=1}^N \sum_{\ell \neq k} \tilde{K}(x^\ell - x^k) \nu_f^N \varphi_R(x^\ell - x^k) \partial_i \partial_k \bar{\mathcal{P}}_s^N f(\mathbf{x}) + \sum_{k=1}^N V'(x^k) \partial_i \partial_k \bar{\mathcal{P}}_s^N f(\mathbf{x}) \\
&\quad - \sum_{k=1}^N \sum_{\ell \neq k} \nu_f^N \partial_i (\tilde{K}(x^\ell - x^k) \varphi_R(x^\ell - x^k)) \partial_k \bar{\mathcal{P}}_s^N f(\mathbf{x}) + V''(x^i) \partial_i \bar{\mathcal{P}}_s^N f(\mathbf{x}),
\end{aligned}$$

so that

$$\begin{aligned}
& (\mathcal{L}_S - \partial_s) (\partial_i \bar{\mathcal{P}}_s^N f(\mathbf{x})) \\
&\leq C \sum_{k=1}^N \sum_{\ell \neq k} a_{k\ell} |K(x^\ell - x^k)| |\partial_k \partial_i \bar{\mathcal{P}}_s^N f(\mathbf{x})| + C \sum_{k=1}^N (N-1) \nu_f^N |\partial_i \partial_k \bar{\mathcal{P}}_s^N f(\mathbf{x})| \\
&\quad + \sum_{\ell \neq i} \nu_f^N (\varphi_R \tilde{K})'(x^\ell - x^i) \partial_i \bar{\mathcal{P}}_s^N f(\mathbf{x}) - \sum_{k \neq i} \nu_f^N (\varphi_R \tilde{K})'(x^i - x^k) \partial_k \bar{\mathcal{P}}_s^N f(\mathbf{x}) + C |\partial_i \bar{\mathcal{P}}_s^N f(\mathbf{x})| \\
&\leq C \sum_{k=1}^N \sum_{\ell \neq k} a_{k\ell} |K(x^\ell - x^k)| |\partial_k \partial_i \bar{\mathcal{P}}_s^N f(\mathbf{x})| + C \sum_{k=1}^N (N-1) \nu_f^N |\partial_i \partial_k \bar{\mathcal{P}}_s^N f(\mathbf{x})| \\
&\quad + C(N-1) \nu_f^N |\partial_i \bar{\mathcal{P}}_s^N f(\mathbf{x})| + C \nu_f^N \sum_{k \neq i} |\partial_k \bar{\mathcal{P}}_s^N f(\mathbf{x})| + C |\partial_i \bar{\mathcal{P}}_s^N f(\mathbf{x})| \\
&\leq C \sum_{k=1}^N \psi^k(\mathbf{x}, \mathbf{a}) |\partial_k \partial_i \bar{\mathcal{P}}_s^N f(\mathbf{x})| + C \sum_{k=1}^N |\partial_i \partial_k \bar{\mathcal{P}}_s^N f(\mathbf{x})| + C |\partial_i \bar{\mathcal{P}}_s^N f(\mathbf{x})| + C \nu_f^N \sum_{k \neq i} |\partial_k \bar{\mathcal{P}}_s^N f(\mathbf{x})|.
\end{aligned} \tag{61}$$

From (57) we then have

$$\begin{aligned}
(\mathcal{L}_S - \partial_s) f_s^1 &\leq \sum_{i=1}^N \sum_{j \neq i}^N |\mathcal{L}_S(\tilde{K}(x^j - x^i))| a_{ij} |\partial_i \bar{\mathcal{P}}_s^N f(\mathbf{x})| \\
&\quad + 2 \sum_{i=1}^N \sum_{j \neq i}^N a_{ij} |\tilde{K}'(x^j - x^i)| |\partial_j \partial_i \bar{\mathcal{P}}_s^N f(\mathbf{x})| + 2 \sum_{i=1}^N \sum_{j \neq i}^N a_{ij} |\tilde{K}'(x^j - x^i)| |\partial_i \partial_i \bar{\mathcal{P}}_s^N f(\mathbf{x})| \\
&\quad + \sum_{i=1}^N \psi^i(\mathbf{x}, \mathbf{a}) |(\mathcal{L}_S - \partial_s)(\partial_i \bar{\mathcal{P}}_s^N f(\mathbf{x}))| \\
&\stackrel{(59), (61)}{\leq} \sum_{i=1}^N \sum_{j \neq i}^N C (1 + \psi^i(\mathbf{x}, \mathbf{a}) + \psi^j(\mathbf{x}, \mathbf{a}) + g(x^i - x^j)) a_{ij} |\partial_i \bar{\mathcal{P}}_s^N f(\mathbf{x})| \\
&\quad + 2C \sum_{i=1}^N \sum_{j \neq i}^N a_{ij} |\partial_j \partial_i \bar{\mathcal{P}}_s^N f(\mathbf{x})| + 2C \sum_{i=1}^N \psi^i(\mathbf{x}, \mathbf{a}) |\partial_i \partial_i \bar{\mathcal{P}}_s^N f(\mathbf{x})| \\
&\quad + C \sum_{i=1}^N \psi^i(\mathbf{x}, \mathbf{a}) \sum_{k=1}^N \psi^k(\mathbf{x}, \mathbf{a}) |\partial_k \partial_i \bar{\mathcal{P}}_s^N f(\mathbf{x})| + C \sum_{i=1}^N \psi^i(\mathbf{x}, \mathbf{a}) \sum_{k=1}^N |\partial_i \partial_k \bar{\mathcal{P}}_s^N f(\mathbf{x})| \\
&\quad + C \sum_{i=1}^N \psi^i(\mathbf{x}, \mathbf{a}) |\partial_i \bar{\mathcal{P}}_s^N f(\mathbf{x})| + C \nu_f^N \sum_{i=1}^N \psi^i(\mathbf{x}, \mathbf{a}) \sum_{k \neq i}^N |\partial_k \bar{\mathcal{P}}_s^N f(\mathbf{x})|.
\end{aligned}$$

Now we want to eliminate in the right hand side above all derivatives of  $\bar{\mathcal{P}}_s^N f$  in favor of  $[[\bar{\mathcal{P}}_s^N f]]_{C_b^2(\mathbb{R}^N)}$ , and all terms involving  $a_{ij}$  and  $\psi^i$  in favor of  $\mathcal{I}_1^N, \mathcal{I}_2^N$ . Using  $g \geq 1$  and the definitions of  $\psi^i, \mathcal{I}_1^N, \mathcal{I}_2^N$  and  $[[\cdot]]_{C_b^2(\mathbb{R}^N)}$  (ref. (27,28, 29) and (39)) we reach the following expression:

$$\begin{aligned}
(\mathcal{L}_S - \partial_s) f_s^1 &\leq CN \mathcal{I}_1^N [[\bar{\mathcal{P}}_s^N f]]_{C_b^2(\mathbb{R}^N)} + CN \mathcal{I}_2^N [[\bar{\mathcal{P}}_s^N f]]_{C_b^2(\mathbb{R}^N)} + C \sum_i \sum_j a_{ij} \psi^j |\partial_i \bar{\mathcal{P}}_s^N f(\mathbf{x})| \\
&\quad + C \sum_i \sum_k \psi^i |\partial_k \partial_i \bar{\mathcal{P}}_s^N f(\mathbf{x})| + C \sum_i \sum_k \psi^i \psi^k |\partial_k \partial_i \bar{\mathcal{P}}_s^N f(\mathbf{x})|.
\end{aligned}$$

The last three terms above are dealt with using Cauchy-Schwarz's inequality, and  $y^{1/2} \leq (1+y)/2$  for  $y \geq 0$ :

$$\begin{aligned} \sum_i \sum_j a_{ij} \psi^j |\partial_i \bar{\mathcal{P}}_s^N f(\mathbf{x})| &\leq \sum_j \psi^j \left( \sum_i a_{ij}^2 \right)^{1/2} [[\bar{\mathcal{P}}_s^N f]]_{C_b^2(\mathbb{R}^N)} \\ &\leq [[\bar{\mathcal{P}}_s^N f]]_{C_b^2(\mathbb{R}^N)} \frac{1}{2} \sum_j \psi^j (1 + \psi^j) \leq \frac{1}{2} N (\mathcal{I}_1^N + \mathcal{I}_2^N) [[\bar{\mathcal{P}}_s^N f]]_{C_b^2(\mathbb{R}^N)} ; \\ \sum_i \sum_k \psi^i |\partial_i \partial_k \bar{\mathcal{P}}_s^N f(\mathbf{x})| &\leq [[\bar{\mathcal{P}}_s^N f]]_{C_b^2(\mathbb{R}^N)} \left( \sum_{i,k} (\psi^i)^2 \right)^{1/2} \\ &\leq N \sqrt{\mathcal{I}_2^N} [[\bar{\mathcal{P}}_s^N f]]_{C_b^2(\mathbb{R}^N)} \leq \frac{1}{2} N (1 + \mathcal{I}_2^N) [[\bar{\mathcal{P}}_s^N f]]_{C_b^2(\mathbb{R}^N)} ; \\ \sum_i \sum_k \psi^i \psi^k |\partial_i \partial_k \bar{\mathcal{P}}_s^N f(\mathbf{x})| &\leq [[\bar{\mathcal{P}}_s^N f]]_{C_b^2(\mathbb{R}^N)} \left( \sum_{i,k} (\psi^i)^2 (\psi^k)^2 \right)^{1/2} \\ &\leq N \mathcal{I}_2^N [[\bar{\mathcal{P}}_s^N f]]_{C_b^2(\mathbb{R}^N)} . \end{aligned}$$

We now put everything together, and apply the semigroup  $\mathcal{P}_{t-s}^{N,\varepsilon}$ , obtaining

$$\mathcal{P}_{t-s}^{N,\varepsilon} (\mathcal{L}_S - \partial_s) f_s^1 \leq CN [[\bar{\mathcal{P}}_s^N f]]_{C_b^2(\mathbb{R}^N)} \left( 1 + \mathcal{P}_{t-s}^{N,\varepsilon} \mathcal{I}_1^N(\mathbf{x}, \mathbf{a}) + \mathcal{P}_{t-s}^{N,\varepsilon} \mathcal{I}_2^N(\mathbf{x}, \mathbf{a}) \right),$$

where  $C$  has been redefined. By Lemma 4.2 and Lemma 4.3 we then get

$$\begin{aligned} \mathcal{P}_{t-s}^{N,\varepsilon} (\mathcal{L}_S - \partial_s) f_s^1 &\leq CN [[\bar{\mathcal{P}}_s^N f]]_{C_b^2(\mathbb{R}^N)} \left( 1 + 2C_1 + (1 + C_3) \mathcal{I}_1^N(\mathbf{x}, \mathbf{a}) e^{-C_2 \frac{t}{\varepsilon}} + \mathcal{I}_2^N(\mathbf{x}, \mathbf{a}) e^{-C_4 \frac{t}{\varepsilon}} \right) \\ &\leq 3CN [[\bar{\mathcal{P}}_s^N f]]_{C_b^2(\mathbb{R}^N)} \left( C_1 + C_3 \mathcal{I}_1^N(\mathbf{x}, \mathbf{a}) e^{-C_2 \frac{t}{\varepsilon}} + \mathcal{I}_2^N(\mathbf{x}, \mathbf{a}) e^{-C_4 \frac{t}{\varepsilon}} \right). \end{aligned}$$

(The inequality holds provided  $C_1 \geq 1, C_3 \geq 2$ , which may be assumed without loss of generality.) Then Step 2 follows from applying Lemma 4.4. Finally we prove Step 3. By the definition of  $\mathcal{I}_1^N$  we have

$$|f_t^1(\mathbf{x}, \mathbf{a})| \leq CN \mathcal{I}_1^N(\mathbf{x}, \mathbf{a}) [[\bar{\mathcal{P}}_s^N f]]_{C_b^2(\mathbb{R}^N)}. \quad (62)$$

The inequality (52) then follows from Lemma 4.4 and the proof is complete.  $\square$

**4.3. Proof of Lemma 4.2, 4.3 and 4.4.** We start by proving the bounds on the semigroup  $\bar{\mathcal{P}}_t^N$  contained in Lemma 4.4 and then prove Lemma 4.2 and 4.3. Throughout this section, without loss of generality, we set the diffusion coefficient  $D = 1$  and  $n = 1$ , so that  $f \in C_b^2(\mathbb{R}^N)$  and  $K, V : \mathbb{R} \rightarrow \mathbb{R}$ . We will therefore write  $\partial_i$  instead of  $\nabla_i$  and  $V'$  denotes derivative of  $V$  with respect to its argument (similarly for  $K'$ ).

*Proof of Lemma 4.4.* Define the function

$$\Gamma(f) := \alpha |f(\mathbf{x})|^2 + \beta \sum_{i=1}^N |\partial_i f(\mathbf{x})|^2 + \sum_{i,j=1}^N |\partial_{ij}^2 f(\mathbf{x})|^2 \quad (63)$$

where  $\alpha, \beta$  are positive constants to be chosen later. Note that if we can prove

$$\Gamma(\bar{\mathcal{P}}_t^N f) \leq C \left( \sum_{i=1}^N \|\partial_i f\|_\infty^2 + \sum_{i,j=1}^N \|\partial_{ij}^2 f\|_\infty^2 \right) \quad (64)$$

for some constant  $C$  dependent only on  $\kappa_1, \kappa_2, \kappa_3$ , and  $\kappa_4$ , then (40) holds. On the other hand, to prove (41) it is sufficient to take  $\alpha = 0$  in the definition of  $\Gamma$  and, with this choice of  $\alpha$ , show that for some  $\delta > 0$  one has

$$(\partial_s - \bar{\mathcal{L}}^N)\Gamma(f_s) \leq -\delta\Gamma(f_s), \quad (65)$$

where we are using the shorthand notation  $f_t(\mathbf{x}) := (\bar{\mathcal{P}}_t^N f)\mathbf{x}$ . Indeed, by standard semigroup properties,  $\partial_s(\bar{\mathcal{P}}_{t-s}^N\Gamma(f_s)) = \bar{\mathcal{P}}_{t-s}^N((\partial_s - \bar{\mathcal{L}}^N)\Gamma(f_s))$ , and therefore (65) implies

$$\partial_s(\bar{\mathcal{P}}_{t-s}^N\Gamma(f_s)) \leq -\delta\bar{\mathcal{P}}_{t-s}^N\Gamma(f_s).$$

If we apply Gronwall's inequality (in the  $s$  variable) to the above and calculate the resulting expression in  $s = t$ , we obtain

$$\Gamma(\bar{\mathcal{P}}_t^N f) \leq e^{-\delta t}\bar{\mathcal{P}}_t^N(\Gamma(f));$$

by Markovianity,  $\bar{\mathcal{P}}_t^N$  is a contraction in the  $L^\infty$  norm, thus (41) holds.

Similarly, in order to prove (64), one just needs to show that (65) holds with  $\delta = 0$ , however this time with  $\alpha > 0$  to be chosen, as we will show. To this end, we expand the expression  $(\partial_s - \bar{\mathcal{L}}^N)\Gamma(f_s)$  as follows:

$$\begin{aligned} (\partial_s - \bar{\mathcal{L}}^N)\Gamma(f_s) &= -2\alpha \sum_{i=1}^N |\partial_i f_s|^2 + 2\beta \sum_{i=1}^N ([\partial_i, \bar{\mathcal{L}}^N]f_s)(\partial_i f_s) \\ &\quad - 2\beta \sum_{i,j=1}^N |\partial_{ij}^2 f_s|^2 + 2 \sum_{i,j=1}^N ([\partial_{ij}^2, \bar{\mathcal{L}}^N]f_s)(\partial_{ij}^2 f_s) - 2 \sum_{i,j,k=1}^N |\partial_{ijk}^3 f_s|^2, \end{aligned}$$

where in the above  $[\cdot, \cdot]$  denotes the commutator<sup>8</sup> between two differential operators. The above equality can be obtained via direct calculations or by using shortcuts similar to those typically employed in Bakry-Emery-type approaches, see for example [10, equation (26)] or also [19]. Using the definition of the function  $\bar{\mathcal{F}}^i$ , see (10), and (18), we can write

$$\bar{\mathcal{L}}^N f(\mathbf{x}) = \sum_{i=1}^N \bar{\mathcal{F}}^i(\mathbf{x})\partial_i f(\mathbf{x}) + \sum_{i=1}^N \partial_{ii}^2 f(\mathbf{x}).$$

Now

$$[\partial_i, \bar{\mathcal{L}}^N] = \sum_{j=1}^N [\partial_i, \bar{\mathcal{F}}^j(\mathbf{x})\partial_j] = \sum_{j=1}^N (\partial_i \bar{\mathcal{F}}^j(\mathbf{x}))\partial_j,$$

and similarly

$$\begin{aligned} [\partial_{ij}^2, \bar{\mathcal{L}}^N] &= \sum_{k=1}^N [\partial_{ij}^2, \bar{\mathcal{F}}^k(x)\partial_k] \\ &= \sum_{k=1}^N (\partial_{ij}^2 \bar{\mathcal{F}}^k(\mathbf{x}))\partial_k + (\partial_i \bar{\mathcal{F}}^k(\mathbf{x}))\partial_{jk}^2 + (\partial_j \bar{\mathcal{F}}^k(\mathbf{x}))\partial_{ik}^2. \end{aligned}$$

---

<sup>8</sup>We recall that if  $X$  and  $Y$  are two differential operators, the commutator  $[X, Y]$  is the differential operator defined as  $[X, Y]f := (XY)(f) - (YX)(f)$ .

For  $i, j, k$  all distinct we have

$$\begin{aligned}\partial_i \bar{\mathcal{F}}^i(\mathbf{x}) &= \sum_{\ell \neq i} \partial_i \bar{K}^N(x^\ell - x^i) - \partial_{ii}^2 V(x^i) \\ &= - \sum_{\ell \neq i} (\bar{K}^N)'(x^\ell - x^i) - V''(x^i) \leq \kappa_1 - \kappa_3, \\ \partial_i \bar{\mathcal{F}}^j(\mathbf{x}) &= (\bar{K}^N)'(x^i - x^j) \leq \kappa_1/N,\end{aligned}$$

and

$$\begin{aligned}\partial_{ij} \bar{\mathcal{F}}^k(\mathbf{x}) &= 0 \\ |\partial_{ii}^2 \bar{\mathcal{F}}^k(\mathbf{x})| &= |(\bar{K}^N)''(x^i - x^k)| \leq \kappa_2/N \\ |\partial_{ik}^2 \bar{\mathcal{F}}^k(\mathbf{x})| &= |-(\bar{K}^N)''(x^i - x^k)| \leq \kappa_2/N \\ |\partial_{ii}^2 \bar{\mathcal{F}}^i(\mathbf{x})| &= \left| \sum_{\ell \neq i} (\bar{K}^N)''(x^\ell - x^i) - V'''(x^i) \right| \leq \kappa_2 + \kappa_4.\end{aligned}$$

Therefore we have

$$\begin{aligned}(\partial_s - \bar{\mathcal{L}}^N)\Gamma(f_s) &= -2\alpha \sum_{i=1}^N |\partial_i f_s|^2 + 2\beta \sum_{i=1}^N \sum_{j=1}^N (\partial_i \bar{\mathcal{F}}^j(\mathbf{x})) (\partial_j f_s) (\partial_i f_s) \\ &\quad - 2\beta \sum_{i,j=1}^N |\partial_{ij}^2 f_s|^2 + 2 \sum_{i,j,k=1}^N (\partial_{ij}^2 \bar{\mathcal{F}}^k(\mathbf{x})) (\partial_k f_s) (\partial_{ij}^2 f_s) \\ &\quad + 2 \sum_{i,j,k=1}^N (\partial_i \bar{\mathcal{F}}^k(\mathbf{x})) (\partial_{jk}^2 f_s) (\partial_{ij}^2 f_s) - 2 \sum_{i,j,k=1}^N |\partial_{ijk}^3 f_s|^2 \\ &\leq -2\alpha \sum_{i=1}^N |\partial_i f_s|^2 + 2\beta \sum_{i=1}^N (\kappa_1 - \kappa_3) (\partial_i f_s)^2 \\ &\quad + \frac{2\beta\kappa_1}{N} \sum_{i=1}^N \sum_{j \neq i} |\partial_j f_s| |\partial_i f_s| - 2\beta \sum_{i,j=1}^N |\partial_{ij}^2 f_s|^2 \\ &\quad + 2 \sum_i^N (\kappa_2 + \kappa_4) |\partial_i f_s| |\partial_{ii}^2 f_s| + \frac{4\kappa_2}{N} \sum_i \sum_{j \neq i} |\partial_i f_s| |\partial_{ij}^2 f_s| \\ &\quad + \frac{2\kappa_2}{N} \sum_i \sum_{k \neq i} |\partial_k f_s| |\partial_{ii}^2 f_s| \\ &\quad + \frac{2\kappa_1}{N} \sum_{i,j=1}^N \sum_{k \neq i} |\partial_{jk}^2 f_s| |\partial_{ij}^2 f_s| + 2(\kappa_1 - \kappa_3) \sum_{i,j}^N (\partial_{ij}^2 f_s)^2.\end{aligned}$$



Now by using Young's inequality we have

$$\begin{aligned}
(\partial_s - \bar{\mathcal{L}}^N)\Gamma(f_s) &\leq -2\alpha \sum_{i=1}^N |\partial_i f_s|^2 + 2\beta \sum_{i=1}^N (\kappa_1 - \kappa_3)(\partial_i f_s)^2 + \beta\kappa_1 \sum_{i=1}^N |\partial_i f_s|^2 \\
&+ \beta\kappa_1 \sum_{j=1}^N |\partial_j f_s|^2 - 2\beta \sum_{i,j=1}^N |\partial_{ij}^2 f_s|^2 + (\kappa_2 + \kappa_4) \sum_{i=1}^N |\partial_i f_s|^2 + (\kappa_2 + \kappa_4) \sum_{i=1}^N |\partial_{ii}^2 f_s|^2 \\
&+ \frac{2\kappa_2}{N} \sum_i \sum_{j \neq i} |\partial_{ij}^2 f_s|^2 + 2\kappa_2 \sum_i |\partial_i f_s|^2 + \kappa_2 \sum_{k=1}^N |\partial_k f_s|^2 + \kappa_2 \sum_i |\partial_{ii}^2 f_s|^2 \\
&+ \kappa_1 \sum_{j=1}^N \sum_{k \neq i} |\partial_{jk}^2 f_s|^2 + \kappa_1 \sum_{i,j=1}^N |\partial_{ij}^2 f_s|^2 + 2(\kappa_1 - \kappa_3) \sum_{i,j} (\partial_{ij}^2 f_s)^2.
\end{aligned}$$

Finally, collecting all terms gives

$$\begin{aligned}
(\partial_s - \bar{\mathcal{L}}^N)\Gamma(f_s) &\leq (-2\alpha - 2\beta\kappa_3 + 4\beta\kappa_1 + \kappa_4 + 4\kappa_2) \sum_{i=1}^N |\partial_i f_s|^2 \\
&+ \left( -2\beta + 2\kappa_2 + \kappa_4 + \frac{2\kappa_2}{N} + 4\kappa_1 - 2\kappa_3 \right) \sum_{i,j} (\partial_{ij} f_s)^2.
\end{aligned}$$

If we choose

$$2\beta \geq \max\{0, 2\kappa_2 + \kappa_4 + \frac{2\kappa_2}{N} + 4\kappa_1 - 2\kappa_3\}$$

and

$$2\alpha = 4\beta\kappa_1 + \kappa_4 + 4\kappa_2$$

then we have  $(\partial_s - \bar{\mathcal{L}})\Gamma(f_s) \leq 0$  as required to prove (64). It should be emphasized that  $\alpha$  and  $\beta$  can be taken independent of  $N$ . Moreover if in the above calculations we set  $\alpha = 0$  and choose  $\delta > 0$  and  $\beta > 0$  such that

$$2\beta = \delta + 4\kappa_2 + \kappa_4 + 4\kappa_1,$$

and

$$-\kappa_3 \leq -\frac{\delta}{2} - 2\kappa_1 - \frac{\kappa_4}{2\beta} - \frac{2\kappa_2}{\beta},$$

we have

$$(\partial_s - \bar{\mathcal{L}}^N)\Gamma(f_s) \leq -\delta \left( \beta \sum_{i=1}^N |\partial_i f_s|^2 + \sum_{i,j} (\partial_{ij} f_s)^2 \right) = -\delta\Gamma(f_s),$$

where  $\Gamma(f_s)$  is given by (63) with  $\alpha = 0$ . This completes the proof of (41).  $\square$

*Proof of Lemma 4.2.* As in previous proofs, without loss of generality we set the diffusion coefficient  $D = 1$  and  $n = 1$ . From (22) and (27),

$$\begin{aligned}
\mathcal{L}_F \psi^i(\mathbf{x}, \mathbf{a}) &= \sum_{j:j \neq i} (-\nu_d a_{ij} + \nu_f(x^i, x^j)(1 - a_{ij})) g(x^j - x^i) \\
&\leq -\nu_d \psi^i(\mathbf{x}, \mathbf{a}) + \sum_{j:j \neq i} \nu_f(x^i, x^j) g(x^j - x^i),
\end{aligned}$$

having used the fact that  $1 - a_{ij} \leq 1$ . Note that if  $|x^j - x^i| > R$  then  $\tilde{\nu}_f(x^i, x^j) = \nu_f^N \varphi_R(|x^j - x^i|) = 0$ , hence

$$\tilde{\nu}_f(x^i, x^j)g(x^j - x^i) \leq \nu_f^N \sup_{|z| \leq R} g(z) \leq \frac{C}{N}, \quad (66)$$

where the constant on the r.h.s. depends on  $R$ , so that

$$\sum_{j:j \neq i} \tilde{\nu}_f(x^i, x^j)g(x^j - x^i) \leq N\nu_f^N \sup_{|z| \leq R} g(z) \leq C$$

for some constant  $C$  which depends on  $R$  but is independent of  $N$ . In what follows  $C$  may vary from line to line. We then have

$$\mathcal{L}_F \psi^i(\mathbf{x}, \mathbf{a}) \leq -\nu_d \psi^i(\mathbf{x}, \mathbf{a}) + C.$$

Taking the sum over  $i$  and dividing by  $N$  on both sides we obtain

$$\mathcal{L}_F \mathcal{I}_1^N(\mathbf{x}, \mathbf{a}) \leq -\nu_d \mathcal{I}_1^N(\mathbf{x}, \mathbf{a}) + C. \quad (67)$$

For  $j \neq i$ , using (21) we get

$$\begin{aligned} \mathcal{L}_S(g(x^j - x^i)) &= \sum_{k=1}^N \sum_{r \neq k} K(x^r - x^k) a_{rk} \partial_k(g(x^j - x^i)) + \sum_{k=1}^N \partial_{kk}^2(g(x^j - x^i)) \\ &\quad - \sum_{k=1}^N \partial_k V(x^k) \partial_k(g(x^j - x^i)) \\ &= \sum_{r \neq j} K(x^r - x^j) a_{rj} g'(x^j - x^i) - \sum_{r \neq i} K(x^r - x^i) a_{ri} g'(x^j - x^i) \\ &\quad + 2g''(x^j - x^i) + V'(x^i)g'(x^j - x^i) - V'(x^j)g'(x^j - x^i) \\ &\leq C \sum_{r \neq j} g(x^r - x^j) a_{rj} + C \sum_{r \neq i} g(x^r - x^i) a_{ri} + 2C + C|x^i - x^j|. \end{aligned}$$

To obtain the last inequality we have used that  $g'$  is bounded,  $V'$  is Lipschitz and  $|K(z)| \leq g(z)$ . Now using  $|z| \leq g(z)$  and the definition of  $\psi^i$  (27), we have

$$\mathcal{L}_S(g(x^j - x^i)) \leq C\psi^j(\mathbf{x}, \mathbf{a}) + C\psi^i(\mathbf{x}, \mathbf{a}) + 2C + Cg(x^j - x^i), \quad (68)$$

and so

$$\begin{aligned} \mathcal{L}_S \psi^i(\mathbf{x}, \mathbf{a}) &= \sum_{j:j \neq i} a_{ij} \mathcal{L}_S(g(x^j - x^i)) \\ &\leq \sum_{j:j \neq i} a_{ij} C\psi^j(\mathbf{x}, \mathbf{a}) + CN\psi^i(\mathbf{x}, \mathbf{a}) + 2C \sum_{j:j \neq i} a_{ij} + \sum_{j:j \neq i} a_{ij} Cg(x^i - x^j) \\ &\leq C \sum_{j:j \neq i} \psi^j(\mathbf{x}, \mathbf{a}) + CN\psi^i(\mathbf{x}, \mathbf{a}) + 2C\psi^i(\mathbf{x}, \mathbf{a}) + C\psi^i(\mathbf{x}, \mathbf{a}). \end{aligned}$$

To obtain the last inequality we have used  $1 \leq g(x^i - x^j)$  and once again the definition of  $\psi^i$ . Taking the sum over  $i$  and dividing by  $N$  we finally obtain

$$\mathcal{L}_S \mathcal{I}_1^N(\mathbf{x}, \mathbf{a}) \leq 2CN\mathcal{I}_1^N(\mathbf{x}, \mathbf{a}) + 2C\mathcal{I}_1^N(\mathbf{x}, \mathbf{a}) + C\mathcal{I}_1^N(\mathbf{x}, \mathbf{a}) \leq CN\mathcal{I}_1^N(\mathbf{x}, \mathbf{a}). \quad (69)$$

We can now estimate  $\mathcal{P}_t^{N,\varepsilon}\mathcal{I}_1^N(\mathbf{x}, \mathbf{a})$ ; indeed, using the fact that the semigroup and its generator commute, together with (67) and (69),

$$\begin{aligned}\partial_t \mathcal{P}_t^{N,\varepsilon}\mathcal{I}_1^N(\mathbf{x}, \mathbf{a}) &= \mathcal{P}_t^{N,\varepsilon} \left( \frac{1}{\varepsilon} \mathcal{L}_F \mathcal{I}_1^N + \mathcal{L}_S \mathcal{I}_1^N \right) (\mathbf{x}, \mathbf{a}) \\ &\leq -\frac{\nu_d}{\varepsilon} \mathcal{P}_t^{N,\varepsilon}\mathcal{I}_1^N(\mathbf{x}, \mathbf{a}) + \frac{C}{\varepsilon} + CN \mathcal{P}_t^{N,\varepsilon}\mathcal{I}_1^N(\mathbf{x}, \mathbf{a}).\end{aligned}$$

Now, by Gronwall's inequality, we obtain

$$\mathcal{P}_t^{N,\varepsilon}\mathcal{I}_1^N(\mathbf{x}, \mathbf{a}) \leq \frac{C}{\nu_d - CN\varepsilon} + \mathcal{I}_1^N(\mathbf{x}, \mathbf{a}) e^{(CN\varepsilon - \nu_d)\frac{t}{\varepsilon}}.$$

Therefore, for  $N\varepsilon$  sufficiently small we have

$$\mathcal{P}_t^{N,\varepsilon}\mathcal{I}_1^N(\mathbf{x}, \mathbf{a}) \leq C_1 + \mathcal{I}_1^N(\mathbf{x}, \mathbf{a}) e^{-C_2\frac{t}{\varepsilon}} \quad (70)$$

where  $C_1, C_2$  are positive constants independent of  $N$  and  $\varepsilon$  (as long as  $N\varepsilon \ll 1$ ).  $\square$

*Proof of Lemma 4.3.* The proof is similar in spirit to the proof of Lemma 4.2. Note that

$$\begin{aligned}\psi^i(\mathbf{x}, \mathbf{a})^2 &= \sum_{j:j \neq i} \sum_{k:k \neq i} a_{ij} a_{ik} g(x^j - x^i) g(x^k - x^i) \\ &= \sum_{j:j \neq i} \sum_{k:k \neq i, j} a_{ij} a_{ik} g(x^j - x^i) g(x^k - x^i) + \sum_{j:j \neq i} a_{ij} g(x^j - x^i)^2.\end{aligned}$$

When we apply  $\mathcal{L}_F$  to  $\psi^i(\mathbf{x}, \mathbf{a})^2$  we get

$$\begin{aligned}\mathcal{L}_F(\psi^i(\mathbf{x}, \mathbf{a})^2) &= -2\nu_d \sum_{j:j \neq i} \sum_{k:k \neq i, j} a_{ij} a_{ik} g(x^j - x^i) g(x^k - x^i) - \nu_d \sum_{j:j \neq i} a_{ij} g(x^j - x^i)^2 \\ &\quad + \sum_{j:j \neq i} \sum_{k:k \neq i, j} \tilde{\nu}_f(x^i, x^j) (1 - a_{ij}) a_{ik} g(x^j - x^i) g(x^k - x^i) \\ &\quad + \sum_{j:j \neq i} \sum_{k:k \neq i, j} \tilde{\nu}_f(x^i, x^k) a_{ij} (1 - a_{ik}) g(x^j - x^i) g(x^k - x^i) \\ &\quad + \sum_{j:j \neq i} \tilde{\nu}_f(x^i, x^j) (1 - a_{ij}) g^2(x^j - x^i) \\ &\leq -\nu_d \psi^i(\mathbf{x}, \mathbf{a})^2 + \sum_{j:j \neq i} \sum_{k:k \neq i, j} \nu_f(x^i, x^j) a_{ik} g(x^j - x^i) g(x^k - x^i) \\ &\quad + \sum_{j:j \neq i} \sum_{k:k \neq i, j} \tilde{\nu}_f(x^i, x^k) a_{ij} g(x^j - x^i) g(x^k - x^i) + \sum_{j:j \neq i} \tilde{\nu}_f(x^i, x^j) g^2(x^j - x^i).\end{aligned}$$

By using the same reasoning as in (66) (with  $g^2$  instead of  $g$ ),

$$\begin{aligned}\mathcal{L}_F(\psi^i(\mathbf{x}, \mathbf{a})^2) &\leq -\nu_d \psi^i(\mathbf{x}, \mathbf{a})^2 + \sum_{k:k \neq i, j} a_{ik} g(x^k - x^i) + \sum_{j:j \neq i} a_{ij} g(x^j - x^i) + C \\ &\leq -\nu_d \psi^i(\mathbf{x}, \mathbf{a})^2 + 2\psi^i(\mathbf{x}, \mathbf{a}) + C,\end{aligned}$$

which gives

$$\mathcal{L}_F \mathcal{I}_2^N(\mathbf{x}, \mathbf{a}) \leq -\nu_d \mathcal{I}_2^N(\mathbf{x}, \mathbf{a}) + \frac{2}{N} \sum_{i=1}^N \psi^i(\mathbf{x}, \mathbf{a}) + C \quad (71)$$

Now we shall calculate  $\mathcal{L}_S \mathcal{I}_2^N$ . Using (68) we have

$$\begin{aligned}
\mathcal{L}_S(g(x^j - x^i)g(x^k - x^i)) &= \mathcal{L}_S[g(x^j - x^i)]g(x^k - x^i) + g(x^j - x^i)\mathcal{L}_S[g(x^k - x^i)] \\
&\quad + \sum_{\ell=1}^N \partial_\ell(g(x^j - x^i))\partial_\ell(g(x^k - x^i)) \\
&\leq C\psi^j(\mathbf{x}, \mathbf{a})g(x^k - x^i) + C\psi^i(\mathbf{x}, \mathbf{a})g(x^k - x^i) \\
&\quad + 2Cg(x^k - x^i) + 2Cg(x^j - x^i)g(x^k - x^i) \\
&\quad + C\psi^k(\mathbf{x}, \mathbf{a})g(x^j - x^i) + C\psi^i(\mathbf{x}, \mathbf{a})g(x^j - x^i) \\
&\quad + 2Cg(x^j - x^i) + g'(x^j - x^i)g'(x^k - x^i).
\end{aligned}$$

Now we can insert this estimate into the definition of  $\mathcal{I}_2^N$  and, using the fact that  $\mathcal{L}_S$  only acts in the variable  $\mathbf{x}$ ,

$$\begin{aligned}
\mathcal{L}_S(\psi^i(\mathbf{x}, \mathbf{a})^2) &\leq C \sum_{j:j \neq i} \sum_{k:k \neq i} a_{ij}a_{ik}\psi^j(\mathbf{x}, \mathbf{a})g(x^k - x^i) + C \sum_{j:j \neq i} \sum_{k:k \neq i} a_{ij}a_{ik}\psi^i(\mathbf{x}, \mathbf{a})g(x^k - x^i) \\
&\quad + 2C \sum_{j:j \neq i} \sum_{k:k \neq i} a_{ij}a_{ik}g(x^k - x^i) + 2C \sum_{j:j \neq i} \sum_{k:k \neq i} a_{ij}a_{ik}g(x^j - x^i)g(x^k - x^i) \\
&\quad + C \sum_{j:j \neq i} \sum_{k:k \neq i} a_{ij}a_{ik}\psi^k(\mathbf{x}, \mathbf{a})g(x^j - x^i) + C \sum_{j:j \neq i} \sum_{k:k \neq i} a_{ij}a_{ik}\psi^i(\mathbf{x}, \mathbf{a})g(x^j - x^i) \\
&\quad + 2C \sum_{j:j \neq i} \sum_{k:k \neq i} a_{ij}a_{ik}g(x^j - x^i) + \sum_{j:j \neq i} \sum_{k:k \neq i} a_{ij}a_{ik}g'(x^j - x^i)g'(x^k - x^i).
\end{aligned}$$

Using that  $g'$  is bounded and  $a_{ij} \leq 1$ , we obtain

$$\begin{aligned}
\mathcal{L}_S(\psi^i(\mathbf{x}, \mathbf{a})^2) &\leq C \sum_{j:j \neq i} \psi^j(\mathbf{x}, \mathbf{a})\psi^i(\mathbf{x}, \mathbf{a}) + CN\psi^i(\mathbf{x}, \mathbf{a})^2 \\
&\quad + 2C \sum_{j:j \neq i} a_{ij}\psi^i(\mathbf{x}, \mathbf{a}) + 2C\psi^i(\mathbf{x}, \mathbf{a})^2 \\
&\quad + C \sum_{k:k \neq i} \psi^k(\mathbf{x}, \mathbf{a})\psi^i(\mathbf{x}, \mathbf{a}) + CN\psi^i(\mathbf{x}, \mathbf{a})^2 \\
&\quad + 2C \sum_{j:j \neq i} \sum_{k:k \neq i} a_{ij}a_{ik}g(x^j - x^i) + C \sum_{j:j \neq i} \sum_{k:k \neq i} a_{ij}a_{ik}.
\end{aligned}$$

Since  $1 \leq g$  we can write each of these expressions in terms of  $\psi^i, \psi^j$ .

$$\mathcal{L}_S(\psi^i(\mathbf{x}, \mathbf{a})^2) \leq 2C \sum_{j:j \neq i} \psi^j(\mathbf{x}, \mathbf{a})\psi^i(\mathbf{x}, \mathbf{a}) + 2CN\psi^i(\mathbf{x}, \mathbf{a})^2 + 7C\psi^i(\mathbf{x}, \mathbf{a})^2.$$

By Young's inequality

$$\mathcal{L}_S(\psi^i(\mathbf{x}, \mathbf{a})^2) \leq C \sum_{j:j \neq i} \psi^j(\mathbf{x}, \mathbf{a})^2 + CN\psi^i(\mathbf{x}, \mathbf{a})^2 + 2CN\psi^i(\mathbf{x}, \mathbf{a})^2 + 7C\psi^i(\mathbf{x}, \mathbf{a})^2.$$

Summing over  $i$ ,

$$\begin{aligned}
\mathcal{L}_S \mathcal{I}_2^N(\mathbf{x}, \mathbf{a}) &\leq CN\mathcal{I}_2^N(\mathbf{x}, \mathbf{a}) + CN\mathcal{I}_2^N(\mathbf{x}, \mathbf{a}) + 2CN\mathcal{I}_2^N(\mathbf{x}, \mathbf{a}) + 7C\mathcal{I}_2^N(\mathbf{x}, \mathbf{a}) \\
&\leq CN\mathcal{I}_2^N(\mathbf{x}, \mathbf{a}).
\end{aligned} \tag{72}$$

Finally, combining (71) and (72), we have

$$\begin{aligned} \partial_t \mathcal{P}_t^{N,\varepsilon} \mathcal{I}_2^N(\mathbf{x}, \mathbf{a}) &= \mathcal{P}_t^{N,\varepsilon} \left( \frac{1}{\varepsilon} \mathcal{L}_F \mathcal{I}_2^N(\mathbf{x}, \mathbf{a}) + \mathcal{L}_S \mathcal{I}_2^N(\mathbf{x}, \mathbf{a}) \right) \\ &\leq \mathcal{P}_t^{N,\varepsilon} \left( -\frac{\nu_d}{\varepsilon} \mathcal{I}_2^N(\mathbf{x}, \mathbf{a}) + \frac{2}{\varepsilon N} \sum_{i=1}^N \psi^i(\mathbf{x}, \mathbf{a}) + \frac{C}{\varepsilon} + CN \mathcal{I}_2^N(\mathbf{x}, \mathbf{a}) \right). \end{aligned}$$

By (70),

$$\partial_t \mathcal{P}_t^{N,\varepsilon} \mathcal{I}_2^N(\mathbf{x}, \mathbf{a}) \leq \left( -\frac{\nu_d}{\varepsilon} + CN \right) \mathcal{P}_t^{N,\varepsilon} \mathcal{I}_2^N(\mathbf{x}, \mathbf{a}) + \frac{C}{\varepsilon} + \frac{2}{\varepsilon} C_1 + \frac{2}{\varepsilon} \mathcal{I}_1^N(\mathbf{x}, \mathbf{a}) e^{-C_2 \frac{t}{\varepsilon}}.$$

If there exists  $\varepsilon_0 > 0$  such that  $N\varepsilon < \varepsilon_0$  then we can integrate this inequality to find constants  $C_1, C_2, C_3, C_4 > 0$  such that

$$\mathcal{P}_t^{N,\varepsilon} \mathcal{I}_2^N(\mathbf{x}, \mathbf{a}) \leq C_1 + C_3 \mathcal{I}_1^N(\mathbf{x}, \mathbf{a}) e^{-C_2 \frac{t}{\varepsilon}} + \mathcal{I}_2^N(\mathbf{x}, \mathbf{a}) e^{-C_4 \frac{t}{\varepsilon}}.$$

□

## 5. THE LARGE $N$ LIMIT: PROOF OF (34) AND (36)

The purpose of this section is to estimate the term  $E^{\text{part}}$  given by (32). In the original dynamics (1a)-(1b), any particle typically interacts at a given time with a bounded (in  $N$ ) number of other particles; hence mean-field techniques cannot be employed. However, the averaged process (7) is of mean-field type, so that classical techniques can be used. We introduce an auxiliary system of independent particles  $Y_t^i$ , such that  $Y_0^{i,N} = X_0^{i,N,\varepsilon} = \bar{X}_0^{i,N}$  and

$$dY_t^{i,N} = -\nabla V(Y_t^{i,N}) dt + (\bar{K} * \rho_t)(Y_t^{i,N}) dt + \sqrt{2D} dB_t^i. \quad (73)$$

The Brownian motions  $B_t^i$  are the same as those for the  $\bar{X}^i$  in equation (7). From now on for simplicity we drop the superscript  $N$  and we simply use the notation  $Y_t^i, \bar{X}_t^i$ , respectively, instead of  $Y_t^{i,N}, \bar{X}_t^{i,N}$ , respectively. The  $Y^i$  are independent, and the law  $\mu_t(y) dy$  of each  $Y^i$  solves the equation (recall that the  $X_0^i$  are iid with law  $\rho_0$ ):

$$\begin{aligned} \partial_t \mu_t(y) &= \nabla \cdot (\mu_t \nabla V(y)) - \nabla \cdot ((\bar{K} * \rho_t) \mu_t) + D \Delta \mu_t, \\ \mu_t(y)|_{t=0} &= \rho_0(y). \end{aligned} \quad (74)$$

Note that, under Assumption **[A.1]** and Hypothesis **[H.1]**, equation (14) has a unique solution (say in  $L^2$ )<sup>9</sup> So, comparing (14) and the above equation, we conclude that the law of any  $Y_i$  at

---

<sup>9</sup>While hard to find in the literature, this is standard under our assumptions on  $\rho_0, V, K$ ; indeed much more general results are available for this equation, e.g. uniqueness of measure-valued solutions, but this is not needed here.

time  $t$  is precisely  $\rho_t(y)dy$ . We therefore get

$$\begin{aligned}
E^{\text{part}} &= \mathbb{E} \left( \frac{1}{N} \sum_i u(\bar{X}_t^i) - \int u(y)\rho_t(y)dy \right)^2 \\
&\leq 2\mathbb{E} \left( \frac{1}{N} \sum_i u(\bar{X}_t^i) - \frac{1}{N} \sum_i u(Y_t^i) \right)^2 \\
&\quad + 2\mathbb{E} \left( \frac{1}{N} \sum_i u(Y_t^i) - \int u(y)\rho_t(y)dy \right)^2 \\
&\leq 2L_u^2 \mathbb{E} [(\bar{X}_t^i - Y_t^i)^2] \\
&\quad + 2\mathbb{E} \left( \frac{1}{N} \sum_i u(Y_t^i) - \int u(y)\rho_t(y)dy \right)^2, \tag{75}
\end{aligned}$$

where  $L_u$  stands for the Lipschitz constant of  $u$  and we have used the particles' exchangeability. Our goal now is to estimate both terms in (75). The main tool is the following classical lemma.

**Lemma 5.1.** *Assume Hypotheses [H.1] and Assumption [A.1]. Then there exists a constant  $C$  such that for any time  $T > 0$*

$$\sup_{t \in [0, T]} \mathbb{E} [(\bar{X}_t^i - Y_t^i)^2] \leq \frac{C}{N} e^{4LT},$$

where  $L$  is related to the Lipschitz constants of  $\bar{K}$  and  $\nabla V$ .

Assume in addition that Hypothesis [H.2] holds with  $\kappa_3 > \kappa_{\text{mf}} := 2L_{\bar{K}}$ , where  $L_{\bar{K}}$  is the Lipschitz constant of  $\bar{K}$ . Then there exists a constant  $C$  such that

$$\sup_{t \geq 0} \mathbb{E} [(\bar{X}_t^i - Y_t^i)^2] \leq \frac{C}{N}.$$

*Proof of Lemma 5.1.* We first compare equations (7) with (73), to obtain

$$\begin{aligned}
d(\bar{X}_t^i - Y_t^i) &= -(\nabla V(\bar{X}_t^i) - \nabla V(Y_t^i)) dt \\
&\quad + \left\{ \left[ \sum_{j \neq i} \bar{K}^N (\bar{X}_t^j - \bar{X}_t^i) \right] - (\bar{K} * \rho_t)(Y_t^i) \right\} dt, \tag{76}
\end{aligned}$$

which is a purely deterministic equation. Therefore, multiplying both sides by  $(\bar{X}_t^i - Y_t^i)$  and integrating with respect to time, we get:

$$\begin{aligned}
\frac{1}{2} (\bar{X}_t^i - Y_t^i)^2 (t) &= - \int_0^t [\nabla V(\bar{X}_s^i) - \nabla V(Y_s^i)] [\bar{X}_s^i - Y_s^i] ds \\
&+ \int_0^t \left[ \sum_{j \neq i} \left( \bar{K}^N - \frac{1}{N} \bar{K} \right) (\bar{X}_s^j - \bar{X}_s^i) \right] [\bar{X}_s^i - Y_s^i] ds \\
&+ \int_0^t \frac{1}{N} \left[ \sum_{j \neq i} \bar{K} (\bar{X}_s^j - \bar{X}_s^i) - \bar{K} (Y_s^j - Y_s^i) \right] [\bar{X}_s^i - Y_s^i] ds \\
&+ \int_0^t \left\{ \frac{1}{N} \left[ \sum_{i \neq j} \bar{K} (Y_s^j - Y_s^i) \right] - (\bar{K} * \rho_s)(Y_s^i) \right\} [\bar{X}_s^i - Y_s^i] ds \\
&=: \int_0^t (I_1 + \dots + I_4) ds.
\end{aligned} \tag{77}$$

To estimate  $I_1$  we either use the boundedness of the second derivative of  $V$  to get

$$\mathbb{E}[I_1] = -\mathbb{E} [\nabla V(\bar{X}_s^i) - \nabla V(Y_s^i)] [\bar{X}_s^i - Y_s^i] \leq C_V \mathbb{E} \left[ (\bar{X}_s^i - Y_s^i)^2 \right],$$

or the convexity of the external potential  $V$  from Hypothesis **[H.2]** to get:

$$\mathbb{E}[I_1] = -\mathbb{E} [\nabla V(\bar{X}_s^i) - \nabla V(Y_s^i)] [\bar{X}_s^i - Y_s^i] \leq -\kappa_3 \mathbb{E} \left[ (\bar{X}_s^i - Y_s^i)^2 \right],$$

Using expressions for  $\bar{K}^N$  and  $\bar{K}$ , (9) and (13), respectively, and the boundedness of  $\bar{K}$  and  $\varphi_R$ , we arrive at the inequality

$$|N\bar{K}^N(x) - \bar{K}(x)| \leq \frac{C}{N}.$$

Therefore, using the particles' exchangeability we obtain

$$\mathbb{E}[I_2] \leq \frac{C}{N} \mathbb{E} |\bar{X}_s^i - Y_s^i| \leq \frac{C}{N} \left( \mathbb{E} \left[ (\bar{X}_s^i - Y_s^i)^2 \right] \right)^{1/2}. \tag{78}$$

For the term  $I_3$  we use the Lipschitz continuity of  $\bar{K}$

$$\begin{aligned}
|I_3| &\leq L_{\bar{K}} \frac{1}{N} \sum_{j \neq i} [|\bar{X}_s^j - \bar{X}_s^i - (Y_s^j - Y_s^i)| |\bar{X}_s^i - Y_s^i|] \\
&\leq L_{\bar{K}} \frac{1}{N} \sum_{j \neq i} [|\bar{X}_s^j - Y_s^j| |\bar{X}_s^i - Y_s^i| + |\bar{X}_s^i - Y_s^i|^2]
\end{aligned}$$

and so, taking the expectation and using exchangeability again, we have

$$\mathbb{E}[I_3] \leq 2L_{\bar{K}} \mathbb{E} \left[ (\bar{X}_s^i - Y_s^i)^2 \right]. \tag{79}$$

Finally, for  $I_4$ , by direct expansion we obtain

$$\begin{aligned}
|\mathbb{E}[I_4]| &= \left| \mathbb{E} \left\{ \frac{1}{N} \left[ \sum_{i \neq j} \bar{K}(Y_s^j - Y_s^i) \right] - (\bar{K} * \rho_s)(Y_s^i) \right\} [\bar{X}_s^i - Y_s^i] \right| \\
&\leq \left( \mathbb{E} \left\{ \left[ \frac{1}{N} \sum_{j \neq i} \bar{K}(Y_s^j - Y_s^i) \right] - (\bar{K} * \rho_s)(Y_s^i) \right\}^2 \right)^{1/2} \left( \mathbb{E} [(\bar{X}_s^i - Y_s^i)^2] \right)^{1/2} \\
&\leq \frac{C}{\sqrt{N}} \left( \mathbb{E} [(\bar{X}_s^i - Y_s^i)^2] \right)^{1/2} + \frac{C'}{N},
\end{aligned} \tag{80}$$

where the pre-factor  $C/\sqrt{N}$  and the last addend  $C'/N$  in the last inequality follow from a direct computation; indeed:

$$\begin{aligned}
&\mathbb{E} \left\{ \left[ \frac{1}{N} \sum_{j \neq i} \bar{K}(Y_s^j - Y_s^i) \right] - (\bar{K} * \rho_s)(Y_s^i) \right\}^2 \\
&= \frac{1}{N^2} \sum_{k, j \neq i} \mathbb{E} \left[ (\bar{K}(Y_s^j - Y_s^i) - (\bar{K} * \rho_s)(Y_s^i)) (\bar{K}(Y_s^k - Y_s^i) - (\bar{K} * \rho_s)(Y_s^i)) \right] \\
&+ \frac{1}{N^2} \sum_{j \neq i} \mathbb{E} \left[ (\bar{K}(Y_s^j - Y_s^i) - (\bar{K} * \rho_s)(Y_s^i))^2 \right] + \left( \frac{1}{N} - \frac{1}{N-1} \right) \sum_{j \neq i} (\bar{K} * \rho_s)(Y_s^i);
\end{aligned} \tag{81}$$

the last term above is smaller than  $\|\bar{K}\|_\infty/N = C'/N$ . Notice now that the random variables  $\bar{K}(Y_s^j - Y_s^i) - (\bar{K} * \rho_s)(Y_s^i)$ ,  $\bar{K}(Y_s^k - Y_s^i) - (\bar{K} * \rho_s)(Y_s^i)$  are independent conditionally on  $Y_s^i$  for  $k \neq j$ , and have zero conditional expectation (because the law of any  $Y_s^l$  is equal to the same  $\rho_s(y)dy$ ); thus only the  $N-1$  terms in the first term of the second line of (81) are non zero, and they can be estimated by a constant proportional to  $\|\bar{K}\|_\infty^2$ . Therefore, taking expectation in (77), and gathering all estimates from above we obtain either

$$\begin{aligned}
\frac{1}{2} \mathbb{E} [(\bar{X}_t^i - Y_t^i)^2] (t) &\leq \frac{C}{N} \int_0^t \left( \mathbb{E} [(\bar{X}_s^i - Y_s^i)^2] \right)^{1/2} ds + (2L_{\bar{K}} + C_V) \int_0^t \mathbb{E} [(\bar{X}_s^i - Y_s^i)^2] ds \\
&+ \int_0^t \left[ \frac{C}{\sqrt{N}} \left( \mathbb{E} [(\bar{X}_s^i - Y_s^i)^2] \right)^{1/2} + \frac{C'}{N} \right] ds,
\end{aligned} \tag{82}$$

or, if  $V$  satisfies Hypothesis **[H.2]**

$$\begin{aligned}
\frac{1}{2} \mathbb{E} [(\bar{X}_t^i - Y_t^i)^2] (t) &\leq (-\kappa_3 + 2L_{\bar{K}}) \int_0^t \mathbb{E} [(\bar{X}_s^i - Y_s^i)^2] ds + \frac{C}{N} \int_0^t \left( \mathbb{E} [(\bar{X}_s^i - Y_s^i)^2] \right)^{1/2} ds \\
&+ \int_0^t \left[ \frac{C}{\sqrt{N}} \left( \mathbb{E} [(\bar{X}_s^i - Y_s^i)^2] \right)^{1/2} + \frac{C'}{N} \right] ds.
\end{aligned} \tag{83}$$

Recall that  $\kappa_3 > 0$ , so for  $\kappa_3 > 2L_{\bar{K}}$  we have  $-\lambda := -\kappa_3 + 2L_{\bar{K}} < 0$ , and, changing the constant  $C$ , we can rewrite the above inequality as

$$\alpha'(t) \leq -2\lambda\alpha(t) + \frac{C}{\sqrt{N}}\alpha^{1/2}(t) + \frac{C}{N}, \tag{84}$$

where we denoted  $\alpha(t) = \mathbb{E} [(\bar{X}_t^i - Y_t^i)^2] (t)$ . We choose  $\delta > 0$  such that  $\tilde{\lambda} = \lambda - \delta > 0$ . From (84), we have

$$\alpha'(t) \leq -2(\lambda - \delta)\alpha(t) + \frac{1}{N} \left( \frac{C^2}{8\delta} + C \right) = -2\tilde{\lambda}\alpha(t) + \frac{C}{N}, \tag{85}$$



where we have redefined  $C$  in the last equality. Then Gronwall's lemma ensures that  $\alpha(t) \leq C/(2\tilde{\lambda}N)$  (recall  $\alpha(0) = 0$ ). Hence (again considering  $C$  a generic constant)

$$\sup_{t \geq 0} \mathbb{E} \left[ (\bar{X}_t^i - Y_t^i)^2 \right] \leq \frac{C}{N}. \quad (86)$$

In the case when  $V$  satisfies only Hypothesis **[H.1]**, we go back to (82). A similar reasoning as above proves a local-in-time estimate:

$$\mathbb{E} \left[ (\bar{X}_t^i - Y_t^i)^2 \right] \leq \frac{C}{N} e^{Lt}, \quad (87)$$

for some  $L > 2(2L_{\bar{K}} + C_V)$ .  $\square$

Now we come back to (75). The first term on the r.h.s. is controlled by Lemma 5.1. Observe also that the second term on the r.h.s. can be controlled similarly to (81), and is therefore small of order  $\frac{1}{N}$ , and thus (75) implies

$$\sup_{t \in [0, T]} |E^{\text{part}}| \leq O\left(\frac{1}{N}\right) + \frac{C}{N} e^{LT},$$

for any  $T > 0$ , if  $V$  satisfies **[H.1]**, or

$$\sup_{t \geq 0} |E^{\text{part}}| \leq \frac{C}{N}$$

for  $V$  satisfying in addition Hypothesis **[H.2]**.

#### ACKNOWLEDGEMENTS

Our attention was drawn to this question by many discussions and works with Pierre Degond, whose contribution is gratefully acknowledged. In particular, the possibility of combining averaging and mean-field techniques was proposed in [3], a common work of two of the authors with P. Degond and D. Peurichard. We acknowledge the generous support of the International Centre for Mathematical Sciences, Research-in-Groups programme. P. Dobson was supported by the Maxwell Institute Graduate School in Analysis and its Applications (MIGSAA), a Centre for Doctoral Training funded by the UK Engineering and Physical Sciences Research Council (grant EP/L016508/01), the Scottish Funding Council, Heriot-Watt University and the University of Edinburgh. Last but not least, the authors are very grateful to Iain Souttar, who spotted a mistake in an earlier version.

#### APPENDIX A. SKETCH OF WELL-POSEDNESS OF THE SLOW-FAST SYSTEM (1a)-(1b)

The well-posedness for (1a)-(1b) is completely standard (under our assumptions on the potential  $V$  and on the interaction force  $K$ ), however we could not find in the literature a result that can be cited directly and that contains as a subcase systems of this form, so we just sketch the proof.

First notice that for a general SDE driven by a Poisson process, i.e. an SDE of the form

$$dY_t = b(Y_{t-})dt + \sigma(Y_{t-})dN_t, \quad Y_0 = y$$

to be well-posed, it is sufficient for  $\sigma$  to be say an everywhere defined function, and  $b$  to be such that the solution of the ODE  $dY_t = b(Y_t)dt$  exists globally (and, in particular, starting from any initial datum). Indeed, the solution  $Y_t$  can be iteratively constructed as follows: i) consider an outcome of Poisson events  $T_1, T_2, \dots$ . In between events, i.e. on each interval  $[T_j, T_{j+1})$ , the dynamics follows the ODE  $dY_t = b(Y_t)dt$ , so in particular we know the value of  $Y_{T_{j+1}-}$ . The value after the jump is given by  $Y_{T_{j+1}} = Y_{T_{j+1}-} + \sigma(Y_{T_{j+1}-})$ . After the jump the solution will resume evolving according to the ODE with initial datum  $Y_{T_{j+1}}$ .

The strong solution (i.e. pathwise solution) to (1a)-(1b) can be built using the same logic; to simplify notation, suppose first  $N = 2$  so that there are only two particles and one link,  $A(t) \in \{0, 1\}$ . In between firing events of either  $N^{d,\varepsilon}$  or  $N^{f,\varepsilon}$  the dynamics (1a) evolves according to either the simple Langevin equation

$$\begin{cases} dX^{1,2,\varepsilon}(t) = -\nabla V(X^{1,2,\varepsilon}(t)) + \sqrt{2\overline{D}}dB_t^1 \\ dX^{2,2,\varepsilon}(t) = -\nabla V(X^{2,2,\varepsilon}(t)) + \sqrt{2\overline{D}}dB_t^2 \end{cases}$$

or

$$\begin{cases} dX^{1,2,\varepsilon}(t) = -\nabla V(X^{1,2,\varepsilon}(t)) + K(X^{2,2,\varepsilon}(t) - X^{1,2,\varepsilon}(t)) + \sqrt{2\overline{D}}dB_t^1 \\ dX^{2,2,\varepsilon}(t) = -\nabla V(X^{2,2,\varepsilon}(t)) + K(X^{1,2,\varepsilon}(t) - X^{2,2,\varepsilon}(t)) + \sqrt{2\overline{D}}dB_t^2 \end{cases}$$

Under our assumptions both  $\nabla V$  and  $K$  are globally Lipschitz, so global well-posedness for both of the above SDEs is completely standard, once the initial datum has finite second moment. Similarly for any  $N > 2$ . Note also that, for every  $N$  fixed, there is almost surely a strictly positive amount of time between any two firing events of  $N^{d,\varepsilon}$  and  $N^{f,\varepsilon}$  (non-explosivity): with the notation used in the introduction, one can consider first the sum of  $N^{d,\varepsilon}$  and  $\check{N}^{f,\varepsilon}$ , which is still a Poisson process, with rate  $\nu_d + \nu_f$ . Hence, by elementary properties of simple Poisson processes, there is almost surely a strictly positive amount of time between any two firing events of  $N^{d,\varepsilon}$  and  $\check{N}^{f,\varepsilon}$ . But, by construction, the time between firing events of  $N^{d,\varepsilon}$  and  $N^{f,\varepsilon}$  is longer than the time between firing events of  $N^{d,\varepsilon}$  and  $\check{N}^{f,\varepsilon}$ , as the realization of the sum of  $N^{d,\varepsilon}$  and  $N^{f,\varepsilon}$  is a subsequence of the realization of the sum of  $N^{d,\varepsilon}$  and  $\check{N}^{f,\varepsilon}$ . Observe moreover that the paths of (1a) will be a.s. continuous, as the coefficients  $A_{ij}^{N,\varepsilon} \in \{0, 1\}$  only have the effect of switching on or off the interaction kernel  $K$ .

## REFERENCES

- [1] D. BAKRY, I. GENTIL, AND M. LEDOUX, *Analysis and geometry of Markov diffusion operators*, vol. 348, Springer Science & Business Media, 2013.
- [2] J. BARRÉ, J. A. CARRILLO, P. DEGOND, D. PEURICHARD, AND E. ZATORSKA, *Particle interactions mediated by dynamical networks: assessment of macroscopic descriptions*, Journal of nonlinear science, 28 (2018), pp. 235–268.
- [3] J. BARRÉ, P. DEGOND, D. PEURICHARD, AND E. ZATORSKA, *Modelling pattern formation through differential repulsion*, arXiv preprint arXiv:1906.00704, (2019).
- [4] J. BARRÉ, P. DEGOND, AND E. ZATORSKA, *Kinetic theory of particle interactions mediated by dynamical networks*, Multiscale Modeling & Simulation, 15 (2017), pp. 1294–1323.
- [5] E. BAYRAKTAR AND R. WU, *Mean field interaction on random graphs with dynamically changing multi-color edges*, arXiv preprint arXiv:1912.01785, (2019).
- [6] S. BHAMIDI, A. BUDHIRAJA, AND R. WU, *Weakly interacting particle systems on inhomogeneous random graphs*, Stochastic Processes and their Applications, 129 (2019), pp. 2174–2206.
- [7] T. CASS, D. CRISAN, P. DOBSON, AND M. OTTOBRE, *Long-time behaviour of degenerate diffusions: Ufg-type sdes and time-inhomogeneous hypoelliptic processes*, arXiv preprint arXiv:1805.01350, (2018).
- [8] F. COPPINI, H. DIETERT, AND G. GIACOMIN, *A law of large numbers and large deviations for interacting diffusions on erdős–rényi graphs*, Stochastics and Dynamics, (2019), p. 2050010.
- [9] D. CRISAN, P. DOBSON, AND M. OTTOBRE, *Uniform in time estimates for the weak error of the euler method for sdes and a pathwise approach to derivative estimates for diffusion semigroups*, arXiv preprint arXiv:1905.03524, (2018).
- [10] D. CRISAN AND M. OTTOBRE, *Pointwise gradient bounds for degenerate semigroups (of ufg type)*, Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 472 (2016), p. 20160442.
- [11] P. DEGOND, F. DELEBECQUE, AND D. PEURICHARD, *Continuum model for linked fibers with alignment interactions*, Mathematical Models and Methods in Applied Sciences, 26 (2016), pp. 269–318.
- [12] S. DELATTRE, G. GIACOMIN, AND E. LUÇON, *A note on dynamical models on random graphs and fokker–planck equations*, Journal of Statistical Physics, 165 (2016), pp. 785–798.
- [13] S. N. DOROGOVTSSEV AND J. F. MENDES, *Evolution of networks: From biological nets to the Internet and WWW*, OUP Oxford, 2013.
- [14] F. DRAGONI, V. KONTIS, AND B. ZEGARLINSKI, *Ergodicity of markov semigroups with hörmander type generators in infinite dimensions*, arXiv preprint arXiv:1012.0257, (2010).

- [15] T. D. FRANK, *Nonlinear Fokker-Planck equations: fundamentals and applications*, Springer Science & Business Media, 2005.
- [16] S. GIL AND D. H. ZANETTE, *Coevolution of agents and networks: Opinion spreading and community disconnection*, *Physics Letters A*, 356 (2006), pp. 89–94.
- [17] T. GROSS AND B. BLASIUS, *Adaptive coevolutionary networks: a review*, *Journal of the Royal Society Interface*, 5 (2008), pp. 259–271.
- [18] T. GROSS, C. J. D. D’LIMA, AND B. BLASIUS, *Epidemic dynamics on an adaptive network*, *Physical review letters*, 96 (2006), p. 208701.
- [19] V. KONTIS, M. OTTOBRE, AND B. ZEGARLINSKI, *Markov semigroups with hypocoercive-type generator in infinite dimensions: Ergodicity and smoothing*, *Journal of Functional Analysis*, 270 (2016), pp. 3173–3223.
- [20] E. LUÇON, *Quenched asymptotics for interacting diffusions on inhomogeneous random graphs*, arXiv preprint arXiv:1811.09229, (2018).
- [21] V. MARCEAU, P.-A. NOËL, L. HÉBERT-DUFRESNE, A. ALLARD, AND L. J. DUBÉ, *Adaptive networks: Coevolution of disease and topology*, *Physical Review E*, 82 (2010), p. 036116.
- [22] S. MEI, A. MONTANARI, AND P.-M. NGUYEN, *A mean field view of the landscape of two-layer neural networks*, *Proceedings of the National Academy of Sciences*, 115 (2018), pp. E7665–E7671.
- [23] S. MÉLÉARD, *Asymptotic behaviour of some interacting particle systems; mckean-vasov and boltzmann models*, in *Probabilistic models for nonlinear partial differential equations*, Springer, 1996, pp. 42–95.
- [24] J. R. NORRIS, *Markov chains*, no. 2, Cambridge university press, 1998.
- [25] R. I. OLIVEIRA AND G. H. REIS, *Interacting diffusions on random graphs with diverging average degrees: Hydrodynamics and large deviations*, *Journal of Statistical Physics*, 176 (2019), pp. 1057–1087.
- [26] R. I. OLIVEIRA, G. H. REIS, AND L. M. STOLERMAN, *Interacting diffusions on sparse graphs: hydrodynamics from local weak limits*, arXiv preprint arXiv:1812.11924, (2018).
- [27] G. PAVLIOTIS AND A. STUART, *Multiscale methods: averaging and homogenization*, Springer Science & Business Media, 2008.
- [28] D. PEURICHARD, F. DELEBECQUE, A. LORSIGNOL, C. BARREAU, J. ROUQUETTE, X. DESCOMBES, L. CASTEILLA, AND P. DEGOND, *Simple mechanical cues could explain adipose tissue morphology*, *Journal of theoretical biology*, 429 (2017), pp. 61–81.
- [29] M. A. PORTER, *Nonlinearity+ networks: A 2020 vision*, arXiv preprint arXiv:1911.03805, (2019).
- [30] G. M. ROTSKOFF AND E. VANDEN-EIJNDEN, *Neural networks as interacting particle systems: Asymptotic convexity of the loss landscape and universal scaling of the approximation error*, arXiv preprint arXiv:1805.00915, (2018).
- [31] J. SIRIGNANO AND K. SPILIOPOULOS, *Mean field analysis of neural networks: A law of large numbers*, *SIAM Journal on Applied Mathematics*, 80 (2020), pp. 725–752.
- [32] A.-S. SZNITMAN, *Topics in propagation of chaos*, in *Ecole d’été de probabilités de Saint-Flour XIX—1989*, Springer, 1991, pp. 165–251.
- [33] J. TAYLOR-KING, P. BUENZLI, S. J. CHAPMAN, C. C. LYNCH, AND D. BASANTA, *Modelling osteocyte network formation: Healthy and cancerous environments*, bioRxiv, (2019), p. 729046.
- [34] J. P. TAYLOR-KING, D. BASANTA, S. J. CHAPMAN, AND M. A. PORTER, *Mean-field approach to evolving spatial networks, with an application to osteocyte network formation*, *Physical Review E*, 96 (2017), p. 012301.

(1) INSTITUT DENIS POISSON, UNIVERSITÉ D’ORLÉANS, CNRS, UNIVERSITÉ DE TOURS, FRANCE AND INSTITUT UNIVERSITAIRE DE FRANCE, JULIEN.BARRÉ@UNIV-ORLEANS.FR

(2) DELFT INSTITUTE OF APPLIED MATHEMATICS, DELFT UNIVERSITY OF TECHNOLOGY, 2628 XE DELFT, THE NETHERLANDS, P.DOBSON@TUDELFT.NL

(3) HERIOT-WATT UNIVERSITY, MATHEMATICS DEPARTMENT, EDINBURGH EH14 4AS, UK. M.OTTOBRE AT HW.AC.UK

(4) UNIVERSITY COLLEGE LONDON, LONDON WC1E 6BT, UK. E.ZATORSKA@UCL.AC.UK