ESAIM: PROCEEDINGS AND SURVEYS, December 2023, Vol. 75, p. 2-23
Editors: Marie Doumic, Sébastien Gadat, Quentin Mérigot - Journées SMAI 2021, Université Toulouse 3 Paul Sabatier

RECENT PROGRESS ON LIMIT THEOREMS FOR LARGE STOCHASTIC PARTICLE SYSTEMS *

Max Fathi[1], Pierre Le Bris[2], Angeliki Menegaki[3], Pierre Monmarche[4], Julien Reygner[5] and Milica Tomasevic[6]

Abstract. This article presents a selection of recent results in the mathematical study of physical systems described by a large number of particles, with various types of interactions (mean-field, moderate, nearest-neighbor). Limit theorems are obtained concerning either the large-scale or the long-time behavior of these systems. These results rely on the use of a large range of mathematical tools, arising from both probability theory and the analysis of partial differential equations, and thereby illustrate fruitful interactions between these two disciplines.

Résumé. Cet article présente une sélection de résultats récents dans l'étude mathématique de systèmes physiques décrits par un grand nombre de particules, soumis à des interactions de diverses natures (champ moyen, interaction modérée, plus proches voisins). On y obtient des théorèmes limites concernant le comportement à grande échelle ou en temps long de ces systèmes. Ces résultats reposent sur l'emploi d'une large gamme d'outils mathématiques, provenant de la théorie des probabilités et de l'analyse des équations aux dérivées partielles, et illustrent ainsi les interactions fécondes entre ces deux disciplines.

1. Introduction

The derivation of macroscopic laws for particle systems defined by microscopic evolution rules is the primary purpose of statistical physics. Beyond natural applications to the kinetic theory of gases or the atomistic description of condensed matter, this approach has also been successfully employed in various other fields, such as neuroscience, social or ecological collective behavior modeling, machine learning, game theory, to name but a few. However, establishing rigorous limit theorems, for instance known as *hydrodynamic* or *thermodynamic* limits, depending on the scalings which are considered, is often a challenge to mathematicians. This may be due to various phenomena: the lack of satisfactory ergodic properties of the microscopic dynamics, the nonlinearity of macroscopic equations, a nontrivial interplay between relevant space and time scales... The present article aims

* *This work benefited from the support of the projects ANR EFI (ANR-17-CE40-0030), ANR QuAMProcs (ANR-19-CE40-0010), ANR SWIDIMS (ANR-20-CE40-0022) and ANR MESA (ANR-18-CE40-006) from the French National Research Agency.*

[1] Université Paris Cité and Sorbonne Université, CNRS, Laboratoire Jacques-Louis Lions & Laboratoire de Probabilités, Statistique et Modélisation, 8 place Aurélie Nemours, F-75013 Paris, France
[2] LJLL, Sorbonne Université, Paris, France
[3] IHES, Bures-sur-Yvette, France
[4] LJLL and LCT, Sorbonne Université, Paris, France
[5] CERMICS, Ecole des Ponts, Marne-la-Vallée, France
[6] CNRS, CMAP Ecole Polytechnique, France

Article published online by EDP Sciences and available at https://www.esaim-proc.org or https://doi.org/10.1051/proc/202375002

at presenting recent results which illustrate a selection of topics of current interest, both for the communities of partial differential equations and probability theory, in the study of large stochastic particle systems.

1.1. **Overview of the models considered**

A first class of systems which is considered concerns particles in *mean-field interaction*, which interact through their empirical measure. For such systems, the macroscopic behavior is expected to be induced by the *propagation of chaos* phenomenon [15, 16, 63]: when the size of the system grows to infinity, particles decorrelate and behave independently, following a nonlinear evolution equation which encodes the interaction at the macroscopic level. Mathematical proofs of such a behavior can be traced back to the works of Kac [45] and McKean [50, 51].

In many cases of interest, in particular for systems involving electrostatic interactions or arising from random matrix theory or chemotaxis models, particles interact through a pairwise potential which becomes singular when two particles approach each other. This prevents standard techniques from being employed, and even sometimes makes the particle system ill-defined. In such cases, a possible approach due to Oelschläger [55] consists in smoothing out the interaction kernel in the definition of the particle system, which leads to so-called *moderate interactions*. While seminal propagation of chaos results for such systems are purely qualitative, Section 2 presents a *quantitative* rate of convergence. This is made possible by the use of a recently developed technique, dubbed the *semigroup approach* [29], which is especially designed to approximate nonlinear partial differential equations by smoothed empirical measures.

Section 3 also presents a quantitative propagation of chaos result, for a system of interacting Langevin diffusions. There, the interaction kernel is smooth and the mean-field particle system is well-defined. However, the emphasis is laid on the *uniformity in time* of the propagation of chaos estimate. This result is obtained by a probabilistic construction, which combines in a nontrivial fashion synchronous and reflection couplings, following recent works by Eberle and coauthors [21, 24–26]. Remarkably, the argument allows to address situations where particles evolve in a non-convex confining potential. Furthermore, it may also be applied directly at the level of the macroscopic dynamics of the system, for which it yields quantitative *rates of convergence to equilibrium*. Overall, this provides a typical example of a situation where a purely deterministic result, the description of the long time behavior of a nonlinear partial differential equation (the Vlasov–Fokker–Planck equation), is obtained through a genuinely probabilistic construction. Such a probabilistic approach has long been employed for the study of nonlinear partial differential equations in various settings, see for instance [5, 6, 12, 69].

When propagation of chaos does not hold uniformly in time, *local equilibria* can arise for large but finite particle systems. In such cases, the particle system is typically observed to remain in a seemingly stable subset of the configurational space over a long time period, and then escapes this subset in an 'unpredictable' way — which means, in more probabilistic terms, that the escape time is close to exponentially distributed. This phenomenon is referred to as a *metastable* behavior [58]. Metastability induced by finite-size effects in mean-field particle systems was for instance studied, for a continuous version of the Curie–Weiss model, by Dawson and Gärtner [18]. Section 4 presents another instance of such a phenomenon, for a system of neurons described by mean-field interacting piecewise deterministic Markov processes. There again, the proof relies on a coupling argument, which thus illustrates the versatility of this technique.

Another powerful method to study the long time behavior of large particle systems is the use of *functional inequalities*, such as Poincaré or logarithmic Sobolev inequalities [1]. Section 5 specifically addresses the dependency of constants involved in such inequalities upon the number of particles. The framework is different from that of the three other sections. Indeed, the interaction is no longer of mean-field (or moderate) type, but particles are rather chained together and only interact with their nearest neighbors. Such models of *chains of oscillators*, when in contact with *thermostats*, are ubiquitous in the study of heat transfer, with the overall purpose to derive the Fourier law from microscopic dynamics [8, 19, 32, 46]. In this highly difficult research program, the quantification of the ergodicity of the microscopic dynamics in terms of the size of the system is an important step. The results from Section 5, which rely on a generalization of Bakry and Émery's Γ_2 *calculus* and on a spectral study of Schrödinger operators, provide explicit and, in certain cases, optimal estimates.

As a conclusion, the present article provides a selection of modern techniques employed to describe the large scale and long time behavior of stochastic particle systems with various types of interaction. All of them build on both analytic and probabilistic tools, and thus they exemplify how fruitful interactions between these two fields may be.

1.2. **Notation and common features**

While the four sections of this article present four different works, regarding four different particle systems, they share some common features, some of which have already been emphasized in the previous section: the interest in large scale and long time behaviors, as well as the interplay between these asymptotics; the emergence of nonlinearity at the macroscopic scale as the trace of interaction at the microscopic level; the use of probabilistic constructions to get information on solutions to deterministic evolution equations.

Throughout the article, we use standard notation of differential calculus to write our partial differential equations, and call on the usual theory of stochastic processes, in particular Itô diffusion processes and continuous-time Markov processes. In all sections, some assumptions and results turn out to be formulated in terms of L^1 and L^2 *Wasserstein distances*. We recall here their definition.

Definition 1.1 (L^p Wasserstein distance)**.** Let $p \in [1, +\infty)$. The L^p Wasserstein distance between two probability measures μ and ν on $\mathbb{R}^k$ is defined by

$$\mathcal{W}_p(\mu,\nu) = \left[\inf_{\pi\in C(\mu,\nu)} \int |x-y|^p \pi(dx,dy)\right]^{1/p},$$

where $C(\mu,\nu)$ is the set of probability measures on $\mathbb{R}^k \times \mathbb{R}^k$ with respective marginals μ and ν.

There are several reasons why Wasserstein distances come handy in the context of this article. Perhaps the most important one is the fact that, by its very definition, in order to bound $\mathcal{W}_p(\mu,\nu)^p$ from above it suffices to construct a *coupling* (X, Y) of μ and ν, namely a random pair whose law belongs to $C(\mu,\nu)$, and to estimate $\mathbb{E}[|X-Y|^p]$. Other useful properties of these distances include a nice dual formulation, particularly in the case $p = 1$, which is called the *Kantorovitch-Rubinstein duality*, as well as *transport inequalities* which allow to compare these distances with other divergences whose use is natural in the study of Markov processes, such as the relative entropy. We refer to [69] for an in-depth exposition of these properties.

2. Rate of convergence of moderately interacting particle systems towards non-linear Fokker-Planck Equations with singular kernels

In this section we are interested in the stochastic approximation of nonlinear Fokker-Planck Partial Differential Equations (PDEs) of the form

$$\begin{cases} \partial_t u(t,x) = \Delta u(t,x) - \nabla\cdot(u(t,x)\ K *_x u(t,x)), \quad t>0,\ x\in\mathbb{R}^d, \\ u(0,x) = u_0(x), \end{cases} \tag{1}$$

by means of moderately interacting particle systems. The main interest here is for the kernels K with a singularity at the origin. The results we will present concern the work [57] to which we refer for more details and proofs.

The motivation comes from the cases in which K is an attractive kernel whose singular nature may cause an explosion in finite time in (1) as it is the case in the parabolic-elliptic Keller-Segel model ($K(x) = -\frac{x}{|x|^d}, d \geq 2$). On the microscopic level, the kernels we are interested in lead to particle systems for which well posedness and/or propagation of chaos are not known to hold. A typical family of singular kernels that we consider

derives from Riesz potentials, defined in any dimension d as

$$V_s(x) := \begin{cases} |x|^{-s} & \text{if } s \in (0,d), \\ -\log|x| & \text{if } s = 0, \end{cases} \qquad x \in \mathbb{R}^d. \tag{2}$$

The associated kernel is then $K_s := \pm\nabla V_s$, the sign deciding whether the interaction is attractive or repulsive. Even when it is possible to define the particles, the propagation of chaos may not always hold. For example, in the tricky case of the $2d$ parabolic-elliptic Keller-Segel model, the mean-field particle system was shown to be well-defined [14, 33], but the convergence (on the level of measures on the space of trajectories) was known only to hold for small values of the critical parameter of the equation (see [33]). This was recently improved for subcritical and critical values in the paper [65]. In addition, Bresch-Jabin-Wang [9] showed quantitative convergence of the joint law of N particles towards the chaotic law $u^{\otimes N}$ (built from the solution u to the Keller-Segel model) in the subcritical case when $\mathbb{R}^2$ is replaced by the torus. For the d-dimensional parabolic-elliptic Keller-Segel model with $d \geq 3$ or for the attractive Riesz kernel with $s \in (d-2, d)$, so far the existence result for the associated particle system in mean field interaction has not been proved.

That is why, we will consider moderately interacting particle systems in the sense of Oëlschlager [55] which rely on a smoothing of the interaction kernel at the scale $N^{-\alpha}$, $\alpha \in [0,1]$ (here N is the total number of particles). The main difficulty is then to prove that such particle systems behave, as $N \to \infty$, in the same way as the classical mean-field non-smoothed systems would (if they would be well-defined). That is, we show for a class of singular kernels K that, as $N \to \infty$, our particles behave like independent copies of a non-linear stochastic process in the sense of McKean-Vlasov. As a consequence, we show that the empirical measure of our system converges to the mild solution of (1) on its maximal existence time. Moreover, we will quantify the latter convergence in terms of N and α. For the Riesz kernels, we are able to treat the kernels that derive from (2) for $s \in (0, d-1)$ in both attractive and repulsive case.

The present work is inspired by the new semigroup approach developed in [29], which allows one to approximate nonlinear PDEs by smoothed empirical measures in strong functional topologies. More precisely, the convergence of the mollified empirical measure of the moderately interacting particle system is obtained, and the approach was initially proposed for the FKPP equations. It has already found many applications: see [30] for a PDE-ODE system related to aggregation phenomena; [62] for non-local conservation laws; [31] for the $2d$ Navier-Stokes equation; and [56] for the parabolic-elliptic Keller-Segel systems. However, in these recent works this convergence was not quantified and the propagation of chaos was not considered.

2.1. **The particle system and main hypothesis**

Let $A > 0$ and consider $F_A : \mathbb{R}^d \to \mathbb{R}^d$ a sufficiently smooth version of the cut-off function

$$x \mapsto \mathbb{1}_{\{|x| \leq A\}} x + \mathbb{1}_{\{|x| > A\}} A\mathbf{1},$$

where the notation $\mathbf{1}$ refers to the vector $(1, \dots, 1)$ in $\mathbb{R}^d$. Let $V : \mathbb{R}^d \to \mathbb{R}_+$ be a smooth probability density function, and assume further that V is compactly supported. For any $x \in \mathbb{R}^d$, define

$$V^N(x) := N^{d\alpha} V(N^\alpha x), \qquad \text{for some } \alpha \in [0,1]. \tag{3}$$

Below, α will be restricted to some interval $(0, \alpha_0)$, see Assumption $(\mathbf{A}_\alpha)$.

Let $T > 0$. For each $N \in \mathbb{N}$, the particle system we consider is:

$$\begin{cases} dX_t^{i,N} = F_A\big(\frac{1}{N}\sum_{k=1}^N (K * V^N)(X_t^{i,N} - X_t^{k,N})\big)\, dt + \sqrt{2}\, dW_t^i, \quad t \leq T,\ 1 \leq i \leq N, \\ X_0^{i,N},\ 1 \leq i \leq N, \quad \text{are independent of } \{W^i,\ 1 \leq i \leq N\}, \end{cases} \tag{4}$$

where $\{(W_t^i)_{t\in[0,T]},\ i\in\mathbb{N}\}$ is a family of independent standard $\mathbb{R}^d$-valued Brownian motions defined on a filtered probability space $(\Omega,\mathcal{F},(\mathcal{F}_t)_{t\geq 0},\mathbb{P})$. Observe that the interaction kernel $K*V^N(x)$ is very close to $K(x)$ when $|x|$ is sufficiently large compared to $N^{-\alpha}$. The cut-off may seem unnecessary as the kernel is smoothed with V^N, but it will be well chosen so it is lifted at the limit as $N\to\infty$ and it facilitates the analysis for a fixed N.

Our main goal is to prove that under suitable conditions on K (see Hypothesis $(\mathbf{A}^K)$) and for a well chosen A, one has:

(a) $\{\mu_t^N=\frac{1}{N}\sum_{i=1}^N\delta_{X_t^{i,N}},\ t\in[0,T]\}$, the marginals of the empirical measure of (4), converge to the solution of the PDE (1) for values of the smoothing parameter α which can be up to $(\frac{1}{d})^-$ for some of the models we consider, which is the typical distance between the particles in problems coming from statistical physics with repulsive kernels.

We get a rate of convergence of order $N^{-\varrho}$, with ϱ which can be up to $(\frac{1}{d+2})^-$. Moreover, one can take $F_A(x)\equiv x$ in (4), at the price of a weaker form of convergence.

(b) The system (4) propagates chaos towards the following nonlinear equation (without the cut-off and the mollifier):

$$\begin{cases} dX_t = K*u_t(X_t)\,dt+\sqrt{2}dW_t, \quad t\leq T,\\ \mathrm{Law}(X_t)=u_t,\ \mathrm{Law}(X_0)=u_0. \end{cases} \tag{5}$$

Notice that it is not *a priori* clear that (5) is well-posed, due to the singularity of K. Hence, we also obtain a well-posedness result for (5). As an application, one can use the rate in (a) and a time discretization of (4) to propose a numerical approximation of the PDE (1). We leave this line of investigation for a future work.

Let us denote the empirical measure on $\mathcal{C}([0,T],\mathbb{R}^d)$ of N particles by $\mu^N=\frac{1}{N}\sum_{i=1}^N\delta_{X^{i,N}}$ and the mollified empirical measure by $u_t^N:=V^N*\mu_t^N$, $t\in[0,T]$. The general hypothesis $(\mathbf{A}^K)$ on the kernel K is the following:

$(\mathbf{A}_i^K)$ $K\in L^{\boldsymbol{p}}(\mathcal{B}_1)$, for some $\boldsymbol{p}\in[1,+\infty]$;

$(\mathbf{A}_{ii}^K)$ $K\in L^{\boldsymbol{q}}(\mathcal{B}_1^c)$, for some $\boldsymbol{q}\in[1,+\infty]$;

$(\mathbf{A}_{iii}^K)$ There exists $\boldsymbol{r}\geq\max(\boldsymbol{p}',\boldsymbol{q}')$, $\boldsymbol{\zeta}\in(0,1]$ and $C>0$ such that for any $f\in L^1\cap L^{\boldsymbol{r}}(\mathbb{R}^d)$, one has

$$\mathcal{N}_{\boldsymbol{\zeta}}(K*f)\leq C\|f\|_{L^1\cap L^{\boldsymbol{r}}(\mathbb{R}^d)},$$

where $\mathcal{N}_{\boldsymbol{\zeta}}$ is the Hölder seminorm of parameter $\boldsymbol{\zeta}$.

Assumption $(\mathbf{A}^K)$ is rather mild, and we provide in Section 2.3 a sufficient condition which is easier to check in concrete examples. One can check that the Riesz potentials (see (2)) up to $s\leq d-2$, whether repulsive or attractive, satisfy $(\mathbf{A}^K)$. The case of more singular kernels, e.g. Riesz kernels with $s\in(d-2,d-1)$ is treated separately, but in a very similar fashion adapting the hypothesis $(\mathbf{A}^K)$ and the corresponding proofs to accommodate for broader functional spaces (see Section 4 in [57]).

Throughout this section, it will be supposed that the parameter $\boldsymbol{r}$ is such that $\boldsymbol{r}\geq\max(\boldsymbol{p}',\boldsymbol{q}')$, where $(\boldsymbol{p},\boldsymbol{q})$ are given in $(\mathbf{A}^K)$. The restriction with respect to the key parameters is given by the following assumption:

$(\mathbf{A}_\alpha)$ The parameters α and $\boldsymbol{r}$ (which appear respectively in (3) and $(\mathbf{A}_{iii}^K)$) satisfy

$$0<\alpha<\frac{1}{d+2d(\frac{1}{2}-\frac{1}{\boldsymbol{r}})\vee 0}.$$

Notice that if the integrability of the kernel is such that $\boldsymbol{r}$ in $(\mathbf{A}_{iii}^K)$ could be chosen in the interval $[1,2]$, then α could be arbitrarily close to, but smaller than, $\frac{1}{d}$ (although this will not yield the best possible rate of convergence, see the discussion after Theorem 2.1). This scaling of order $N^{-(\frac{1}{d})^-}$ is close to the typical interparticle distance in physical models with repulsive interactions. This is what we get for instance in the case of $2d$ Navier-Stokes equation and $2d$ Coulomb potential. However, in the case of an attractive interaction, we may wonder what typical inter-particle distance represents and if it is necessarily of the order $N^{-\frac{1}{d}}$ as the particles may often collide or even agglomerate. Indeed, in the super-critical case of the parabolic-elliptic Keller-Segel equation, the non-smoothed particles collide with positive probability (see [33]) and there are, for example,

infinitely many binary collisions [34], whilst in the fully parabolic case they seem to agglomerate (numerically), see [66, Fig. 7.1, p. 120].

Finally, let us state the assumptions on the initial conditions of the system:

(**A**) Fix $\boldsymbol{r}$ from ($\mathbf{A}^K$). For any $m \geq 1$, $\sup_{1\leq i\leq N,\, N\in\mathbb{N}} \mathbb{E}|X_0^{i,N}|^m < \infty$ and $\sup_{N\in\mathbb{N}} \mathbb{E}\left[\left\|\mu_0^N * V^N\right\|_{L^{\boldsymbol{r}}(\mathbb{R}^d)}^m\right] < \infty$.

A sufficient condition for (**A**) to hold is that particles are initially i.i.d. with a law which is in $L^{\boldsymbol{r}}$ and $\alpha d < 1$.

Solutions to (1) are understood in the mild sense. That is, given K satisfying ($\mathbf{A}_i^K$)-($\mathbf{A}_{ii}^K$), $u_0 \in L^1 \cap L^{\boldsymbol{r}}(\mathbb{R}^d)$ with $\boldsymbol{r} \geq \max(\boldsymbol{p}', \boldsymbol{q}')$, and $T > 0$, a function u on $[0,T] \times \mathbb{R}^d$ such that $u \in \mathcal{C}([0,T]; L^1 \cap L^{\boldsymbol{r}}(\mathbb{R}^d))$ and

$$u_t = e^{t\Delta}u_0 - \int_0^t \nabla \cdot (e^{(t-s)\Delta}(u_s\, K * u_s))\, ds, \quad 0 \leq t \leq T, \tag{6}$$

is a mild solution to (1) on $[0,T]$. In [57] the existence and uniqueness of mild solutions for a general K satisfying ($\mathbf{A}_i^K$)-($\mathbf{A}_{ii}^K$) is shown locally in time. One may also check case by case, depending on the kernel, if this solution is global. Denote by $T_{\max}$ the maximal time of existence of a solution to (1) in the above mild sense.

Before we pass to the main result, we show how to choose the cut-off parameter A (for the function F_A) in the definition of the particle system. Indeed, notice that there exists $C_{K,d} > 0$ (which depends on a given K and d only) such that for any $f \in L^1 \cap L^{\boldsymbol{r}}(\mathbb{R}^d)$,

$$\|K * f\|_{L^\infty(\mathbb{R}^d)} \leq C_{K,d}\, \|f\|_{L^1\cap L^{\boldsymbol{r}}(\mathbb{R}^d)}.$$

Hence, for a local mild solution u on $[0,T]$, we will use the cut-off $A_T := C_{K,d}\, \sup_{t\leq T} \|u_t\|_{L^1\cap L^{\boldsymbol{r}}(\mathbb{R}^d)}$.

2.2. **Main results**

The first main result is the following claim:

Theorem 2.1. *Assume that the initial conditions $\{\mu_0^N\}_{N\in\mathbb{N}}$ satisfy (**A**) and that the kernel K satisfies ($\mathbf{A}^K$). Moreover, let ($\mathbf{A}_\alpha$) hold true. Let $T_{\max}$ be the maximal existence time for (1) and fix $T \in (0, T_{\max})$. In addition, let the dynamics of the particle system be given by (4) with A greater than suitably chosen A_T.*

Then, for any $\varepsilon > 0$ and any $m \geq 1$, there exists a constant $C > 0$ such that for all $N \in \mathbb{N}^*$,

$$\left\|\|u^N - u\|_{T,L^1\cap L^{\boldsymbol{r}}(\mathbb{R}^d)}\right\|_{L^m(\Omega)} \leq C\left\|\|u_0^N - u_0\|_{L^1\cap L^{\boldsymbol{r}}(\mathbb{R}^d)}\right\|_{L^m(\Omega)} + CN^{-\varrho+\varepsilon},$$

where

$$\varrho = \min\left(\alpha\boldsymbol{\zeta},\, \frac{1}{2}\left(1 - \alpha(d + d(1 - \frac{2}{\boldsymbol{r}}) \vee 0)\right)\right). \tag{7}$$

It is clear from the definition (7) of ϱ that there is a trade-off between choosing α close to $\frac{1}{d}$ (as imposed by ($\mathbf{A}_\alpha$), provided $\boldsymbol{r} \leq 2$) so as to have more physical particles; and choosing α smaller (such that $\alpha\boldsymbol{\zeta} = \frac{1}{2}(1 - \alpha d)$, for a given $\boldsymbol{\zeta}$ and assuming again that $\boldsymbol{r} \leq 2$) so as to maximize the rate of convergence. The latter case could be of importance in numerical applications.

As a corollary, we obtain the same rate for the genuine empirical measure of the system (4):

$$\left\|\sup_{t\in[0,T]} \mathcal{W}_1(\mu_t^N, u_t)\right\|_{L^m(\Omega)} \leq C\left\|\|u_0^N - u_0\|_{L^1\cap L^{\boldsymbol{r}}(\mathbb{R}^d)}\right\|_{L^m(\Omega)} + C\, N^{-\varrho+\varepsilon},$$

where $\mathcal{W}_1$ denotes the L^1 Wasserstein distance defined in Section 1.2.

We can also remove the cutoff F_A from the drift of the particle system (i.e. we choose $F_A(x) = x$ in (4), or equivalently $A = \infty$), but the convergence will then be in probability.

Now we look at the behavior of the trajectories of our particles when $N \to \infty$. We have the following result:

Theorem 2.2. *Let the hypotheses of Theorem 2.1 hold. In particular, recall that $u \in \mathcal{C}([0,T], L^1 \cap L^{\boldsymbol{r}}(\mathbb{R}^d))$. Assume further that the family of random variables $\{X_0^i,\ i \in \mathbb{N}\}$ is identically distributed and that $\langle u_0^N, \varphi\rangle \to \langle u_0, \varphi\rangle$ in probability, for any $\varphi \in \mathcal{C}_b(\mathbb{R}^d)$. Then, the empirical measure $\mu_\cdot^N$ (defined as a probability measure on $\mathcal{C}([0,T];\mathbb{R}^d)$) converges in probability towards $\mathbb{Q}$, which is the law of the unique weak solution of* (5).

We emphasize here that without the convergence of u^N in the convenient functional framework, it would not be possible to obtain the propagation of chaos in this singular setting. Hence, the result of Theorem 2.1 is very much related to the propagation of chaos and should be considered as the most important ingredient when proving Theorem 2.2.

2.3. **Important examples**

First, we comment here how to check Assumption $(\mathbf{A}^K)$. The first two points of Assumption $(\mathbf{A}^K)$ are simple technical conditions and may not require specific comments, except that it would be interesting to lift the first integrability condition in order to be able to consider more singular kernels. The third assumption is much more interesting. A sufficient easier-to-check condition that in practice replaces $(\mathbf{A}^K_{iii})$ is the following:

$(\widetilde{\widetilde{\mathbf{A}}}^K_{iii})$ There exists $\boldsymbol{r} \geq \max(\boldsymbol{p}', \boldsymbol{q}')$ and $z \in [\boldsymbol{p} \vee \boldsymbol{q}, +\infty] \cap (d, +\infty]$ such that the matrix-valued kernel ∇K defines a convolution operator which is bounded component-wise from $L^1 \cap L^{\boldsymbol{r}}(\mathbb{R}^d)$ to $L^z(\mathbb{R}^d)$.

Then, K satisfies $(\mathbf{A}^K_{iii})$ with the same parameter $\boldsymbol{r}$ and with $\boldsymbol{\zeta} = 1 - \frac{d}{z}$.

Let us now consider the application of Theorem 2.1 to several classes of models:

- For Coulomb-type kernels, which includes the Biot-Savart kernel in dimension 2, the Riesz kernel with $s = d-2$ and the Keller-Segel kernel, the convergence happens for any $\alpha < \frac{1}{2(d-1)}$ (note that in dimension 2, this accounts for choosing $\alpha = (\frac{1}{2})^-$); the best possible rate of convergence is $\varrho = \left(\frac{1}{2(d+1)}\right)^-$ for the choice $\alpha = \left(\frac{1}{2(d+1)}\right)^+$.
- The $2d$ Keller-Segel model ($K(x) = -\chi \frac{x}{2\pi|x|^2}$). We recall that the PDE has a global solution whenever the critical parameter χ satisfies $\chi < 8\pi$ (see [54]), and explodes in finite time otherwise (see [42, 59]). In Theorem 2.1, we get a rate for any value of χ which is almost $\frac{1}{2(d+1)}$. This result holds even if the PDE explodes in finite time ($\chi > 8\pi$). In that case, one works on $[0,T]$ for any $T < T_{\max}$.
- The Riesz kernels with $s > d-2$ do not satisfy Assumption $(\mathbf{A}^K_{iii})$. However, by imposing more regularity on the initial conditions and smaller values of α ($\alpha < \frac{1}{2d}$), we obtain a rate of convergence for singular Riesz kernels with $s \in (d-2, d-1)$, see [57, Thm. 4.2]. In this framework $(\mathbf{A}^K)$ is replaced by another set of hypothesis $(\widetilde{\mathbf{A}}^K)$ where the first one is $(\mathbf{A}^K_i)$ with $\boldsymbol{p} = 1$, the second one is identical to $(\mathbf{A}^K_{ii})$ and $(\mathbf{A}^K_{iii})$ is replaced by:

$(\widetilde{\mathbf{A}}^K_{iii})$ There exists $\boldsymbol{r} \in (d, \infty)$ and $\beta \in (\frac{d}{\boldsymbol{r}}, 1)$ and $\xi \in (0, 1]$ such that for any $f \in L^1 \cap H^\beta_{\boldsymbol{r}}(\mathbb{R}^d)$, one has for some $C > 0$

$$\mathcal{N}_{\boldsymbol{\zeta}}(K * f) \leq C \|f\|_{L^1 \cap H^\beta_{\boldsymbol{r}}(\mathbb{R}^d)}.$$

2.4. **Main Ideas**

We give here the main ideas for proving Theorem 2.1. To quantify the convergence of u^N towards u one derives the following mild formulation of the mollified empirical measure:

$$u_t^N(x) = e^{t\Delta} u_0^N(x) - \int_0^t \nabla \cdot e^{(t-s)\Delta} \langle \mu_s^N, V^N(x - \cdot) F(K * u_s^N(\cdot))\rangle \, ds - \frac{1}{N} \sum_{i=1}^N \int_0^t e^{(t-s)\Delta} \nabla V^N(x - X_s^{i,N}) \cdot dW_s^i. \quad (8)$$

A crucial property of the mollified empirical measure that one can prove using the above mild form is the following: for $q \geq 1$, one has

$$\sup_{N\in\mathbb{N}^*} \mathbb{E}\left[\sup_{t\in[0,T]} \left\|u_t^N\right\|_{L^r(\mathbb{R}^d)}^q\right] < \infty. \tag{9}$$

In view of the mild formulations for u and u^N one can then develop $\left\|\|u^N - u\|_{T,L^1\cap L^r(\mathbb{R}^d)}\right\|_{L^m(\Omega)}$ in several terms and using the assumptions on the kernel K and (9) the terms coming from the drift part are controlled in such a way that the Gronwall lemma can be applied at the end. This is where the $\alpha\xi$ term comes from in (7).

It remains to control in terms of N the moments of $\|M_t^N\|_{L^1(\mathbb{R}^d)}$ and $\|M_t^N\|_{L^r(\mathbb{R}^d)}$ where M_t^N denotes the last term in (8) (this control is also required in order to obtain (9)). This is the trickiest part of the proof. Note that $\left(\|M_s^N\|_{L^1(\mathbb{R}^d)}\right)_{s\geq 0}$ is not a martingale, but a stochastic convolution integral. To achieve the above controls, we use a generalization of the Burkholder-Davis-Gundy (BDG) inequality in UMD Banach spaces (see [67]). There is a classical trick to apply BDG-type inequalities to stochastic convolution integrals, however it only leads to a bound on $\|\|M_t^N\|_{L^1(\mathbb{R}^d)}\|_{L^m(\Omega)}$ for a fixed $t > 0$, instead of a bound on $\|\sup_{s\in[0,t]}\|M_s^N\|_{L^1(\mathbb{R}^d)}\|_{L^m(\Omega)}$. In order to keep the supremum in time inside the expectation, we will also use the lemma of Garsia, Rodemich and Rumsey [36]. Besides, there is an additional difficulty here which is that L^1 is not a UMD Banach space, hence the infinite-dimensional version of the BDG inequality cannot be applied directly. Once this is overcome, we obtain a control in N that leads to the second term in the minimum appearing in (7) and this is where our restriction on α comes from.

3. Convergence rates for the Vlasov-Fokker-Planck equation and uniform in time propagation of chaos in non convex cases

This section is based on [37], in which the authors proved the existence of a contraction rate for Vlasov-Fokker-Planck equation in Wasserstein distance, provided the interaction force is Lipschitz continuous and the confining force is both (locally) Lipschitz continuous and provides a sufficiently restoring effect, without necessarily being the gradient of a convex potential. Their strategy relies on coupling methods suggested by Eberle [24] adapted to the kinetic setting enabling also to obtain uniform in time propagation of chaos in a non convex setting.

3.1. **Framework and results**

Let U and W be two functions in $\mathcal{C}^1\left(\mathbb{R}^d\right)$ and let us consider the N-particle system in $\mathbb{R}^d$ in mean field interaction

$$\forall i \in [\![1, N]\!], \qquad \begin{cases} dX_t^i = V_t^i dt, \\ dV_t^i = \sqrt{2}dB_t^i - V_t^i dt - \nabla U\left(X_t^i\right) dt - \dfrac{1}{N}\displaystyle\sum_{j=1}^N \nabla W\left(X_t^i - X_t^j\right) dt, \end{cases} \tag{10}$$

where X_t^i and V_t^i are respectively the position and the velocity of the i-th particle, and $\left(B_t^i, 1 \leq i \leq N\right)$ are independent Brownian motions in dimension d.

In statistical physics, (10) is a Langevin equation that describes the motion of N particles subject to damping, random collisions, a *confining potential* U and interacting with one another through an *interaction potential* W, which can be polynomial (granular media), Newtonian (interacting stellar) or Coulombian (charged matter).

Our goal is to understand what happens as N, the number of particles, goes to infinity. Intuitively, *in a system of N exchangeable particles in mean-field interaction, as N increases, two particles become more and more statistically independent.* This phenomenon has been stated under the name *propagation of chaos*, an idea motivated by Kac [45], and greatly developed by Sznitman [63]. The notion of *chaos* refers to the independence of particles, and the notion of *propagation* to the fact that it will often be sufficient to prove such limit as N tends to infinity at time 0 for it to also hold at later time t.

To prove such result, let us first determine a good candidate for the limiting process. The dynamics of the first particle can be rewritten

$$\begin{cases} dX_t^1 = V_t^1 dt, \\ dV_t^1 = \sqrt{2} dB_t^1 - V_t^1 dt - \nabla U\left(X_t^1\right) dt - \nabla W * \mu_t^N\left(X_t^1\right) dt, \end{cases}$$

where we denote $\mu_t^N = \frac{1}{N}\sum_{i=1}^N \delta_{X_t^i}$ the empirical measure. The symbol $*$ refers to the operation of convolution. As N goes to infinity, if there is indeed independence, we expect μ_t^N to converge towards the law of X_t^1. We thus define the non linear stochastic differential equation of *McKean-Vlasov* type

$$\begin{cases} dX_t = V_t dt \\ dV_t = \sqrt{2} dB_t - V_t dt - \nabla U\left(X_t\right) dt - \nabla W * \mu_t\left(X_t\right) dt \\ \mu_t = \mathrm{Law}\left(X_t\right). \end{cases} \tag{11}$$

Here, $(X_t, V_t) \in \mathbb{R}^d \times \mathbb{R}^d$, $(B_t)_{t\geq 0}$ is a Brownian motion in dimension d, and μ_t is the law of the position X_t. The Liouville equation associated to the process (11) is refered to as the *Vlasov-Fokker-Planck* equation:

$$\partial_t \nu_t(x,v) = -\nabla_x \cdot \left(v \nu_t(x,v)\right) + \nabla_v \cdot \left(\left(v + \nabla U(x) + \nabla W * \mu_t(x)\right) \nu_t(x,v) + \nabla_v \nu_t(x,v)\right), \tag{12}$$

where $\nu_t(x,v)$ is a probability density in the space of positions $x \in \mathbb{R}^d$ and velocities $v \in \mathbb{R}^d$, and

$$\mu_t(x) = \int_{\mathbb{R}^d} \nu_t(x, dv)$$

is the space marginal of ν_t. In order to prove independence at the limit, we thus compare the law of the N particle system to the law of N independent copies of process (11).

Let us detail the assumptions on the potentials U and W.

Assumption 3.1. *The potential U is non-negative and there exist $\lambda > 0$ and $A \geq 0$ such that*

$$\forall x \in \mathbb{R}^d, \qquad \frac{1}{2}\nabla U(x) \cdot x \geq \lambda\left(U(x) + \frac{|x|^2}{4}\right) - A. \tag{13}$$

The condition (13) implies that the force $-\nabla U$ has a confining effect, bringing back particles toward some compact set. Since only the gradient of U is involved in the dynamics, the condition $U \geq 0$ is thus not restrictive as it can be enforced without loss of generality by adding a sufficient large constant to U. This condition is added in order to simplify some calculations.

Assumption 3.2. *There is a constant $L_U > 0$ such that*

$$\forall (x,y) \in \mathbb{R}^d \times \mathbb{R}^d, \qquad |\nabla U(x) - \nabla U(y)| \leq L_U |x - y|.$$

In particular, non convex potentials U may satisfy the two assumptions.

Example 3.3. Assume $d = 1$. The double-well potential given by

$$U(x) = \begin{cases} \left(x^2 - 1\right)^2 & \text{if } |x| \leq 1, \\ \left(|x| - 1\right)^2 & \text{otherwise,} \end{cases}$$

satisfies Assumptions 3.1 and 3.2.

Assumption 3.4. *The potential W is even, i.e. $W(x) = W(-x)$ for all $x \in \mathbb{R}^d$, in particular $\nabla W(0) = 0$. Moreover, there exists $L_W < \lambda/8$ (where λ is given in Assumption 3.1) such that*

$$\forall x, y \in \mathbb{R}^d \times \mathbb{R}^d, \qquad |\nabla W(x) - \nabla W(y)| \leq L_W |x - y|. \tag{14}$$

In particular $|\nabla W(x)| \leq L_W |x|$ for all $x \in \mathbb{R}^d$.

Here we consider an interaction force that is the gradient of a potential W, as we stick to the formalism of other related works (for instance [21]). Nevertheless, all the results and proofs still hold if ∇W is replaced by some $F : \mathbb{R}^d \to \mathbb{R}^d$ satisfying the same conditions. The confinement force may also be non gradient, however considering ∇U simplifies the construction of a Lyapunov function.

The condition $L_W \leq \lambda/8$ is related to the fact the interaction is considered as a perturbation of the non-interacting process studied in [25]. Therefore, ∇W has to be controlled by ∇U in some sense.

Before proving a result of propagation of chaos, we are first interested in the long-time convergence of the solution of (11) toward an equilibrium. The reason is that the coupling method used to prove the long-time convergence will apply, up to some technical modifications, to the proof of propagation of chaos, while being less cumbersome.

Our main results are stated in terms of an L^1 and L^2 Wasserstein distances on $\mathbb{R}^{2d}$ as introduced in Section 1.2. Please note below that the L^1 Wasserstein distance is slightly different, as we consider the distance between two elements in $(\mathbb{R}^{2d})^k$, denoted $((X^1, V^1), ..., (X^k, V^k))$ and $((X'^1, V'^1), ..., (X'^k, V'^k))$, to be $|X^1 - X'^1| + |V^1 - V'^1| + ... + |X^k - X'^k| + |V^k - V'^k|$.

Theorem 3.5. *Let $U \in \mathcal{C}^1(\mathbb{R}^d)$ satisfy Assumption 3.1 and Assumption 3.2. There is an explicit $c^W > 0$ such that, for all $W \in \mathcal{C}^1(\mathbb{R}^d)$ satisfying Assumption 3.4 with $L_W < c^W$, there is an explicit $\tau > 0$ such that for all probability measures ν_0^1 and ν_0^2 on $\mathbb{R}^{2d}$ with a finite second moment, there are explicit constants $C_1, C_2 > 0$ such that for all $t \geq 0$,*

$$\mathcal{W}_1(\nu_t^1, \nu_t^2) \leq e^{-\tau t} C_1, \qquad \mathcal{W}_2(\nu_t^1, \nu_t^2) \leq e^{-\tau t} C_2,$$

where ν_t^1 and ν_t^2 are solutions of (12) with respective initial distributions ν_0^1 and ν_0^2.

In particular, we have existence and uniqueness of, as well as convergence towards, a stationary solution.

The second of our main results is then a uniform in time convergence as $N \to +\infty$ of (10) toward (11).

Theorem 3.6. *Let $\tilde{\mathcal{C}}^0 > 0$ and $\tilde{a} > 0$. Let $U \in \mathcal{C}^1(\mathbb{R}^d)$ satisfy Assumptions 3.1 and 3.2. There is an explicit $c^W > 0$ such that, for all $W \in \mathcal{C}^1(\mathbb{R}^d)$ satisfying Assumption 3.4 with $L_W < c^W$, there exist explicit $B_1, B_2 > 0$, such that for all probability measures ν_0 on $\mathbb{R}^{2d}$ satisfying $\mathbb{E}_{\nu_0}(e^{\tilde{a}(|X|+|V|)}) \leq \tilde{\mathcal{C}}^0$,*

$$\mathcal{W}_1\left(\nu_t^{k,N}, \bar{\nu}_t^{\otimes k}\right) \leq \frac{kB_1}{\sqrt{N}}, \qquad \mathcal{W}_2^2\left(\nu_t^{k,N}, \bar{\nu}_t^{\otimes k}\right) \leq \frac{kB_2}{\sqrt{N}},$$

for all $k \in \mathbb{N}$, where $\nu_t^{k,N}$ is the marginal distribution at time t of the first k particles $\left((X_t^1, V_t^1), ..., (X_t^k, V_t^k)\right)$ of an N particle system (10) with initial distribution $(\nu_0)^{\otimes N}$, while $\bar{\nu}_t$ is a solution of (12) with initial distribution ν_0.

The speed of convergence in $\frac{1}{\sqrt{N}}$ for the L^1 Wasserstein distance in the theorem above is consistent with the usual speed given in [63] and following works. Notice however that the speed in L^2 Wasserstein distance is not as good as usual results. This is to be expected as we will consider a purely L^1 coupling method, from which the convergence in L^2 Wasserstein distance is merely a byproduct.

3.2. **Description of the method**

The idea behind the coupling method comes from [25]. To prove Theorem 3.5, we construct simultaneously two solutions of (11) that have a trend to get closer with time, since an upper bound on the Wasserstein distance

between two probability distributions is given by the construction of any pair of random variables distributed respectively according to those. Have (X_t, V_t) be a solution of (11) driven by some Brownian motion $(B_t)_{t\geq 0}$ and let (X'_t, V'_t) solve

$$\begin{cases} dX'_t = V'_t dt \\ dV'_t = \sqrt{2} dB'_t - V'_t dt - \nabla U\left(X'_t\right) dt - \nabla W * \mu_t\left(X'_t\right) dt \\ \mu'_t = \mathrm{Law}\left(X'_t\right), \end{cases}$$

with $(B'_t)_{t\geq 0}$ a d-dimensional Brownian motion. A coupling of (X, V) and (X', V') then follows from a coupling of the Brownian motions B and B'.

We identify three main behaviors. First, when one of the particles ventures at infinity, i.e when either $|X_t|, |V_t|, |X'_t|$, or $|V'_t|$ becomes too big, the friction and the confinement potential will tend to bring the particle back into a compact set of $\mathbb{R}^{2d}$. To translate this effect, we construct a Lyapunov function H.

We then observe that, by choosing $B = B'$, the Brownian noise cancels out in the infinitesimal evolution of the difference $(Z_t, W_t) = (X_t - X'_t, V_t - V'_t)$. This choice of coupling is named *synchronous* coupling. In that case, the difference (Z_t, W_t) goes to 0 only thanks to the deterministic drift, as in [7] in the kinetic setting for example (and earlier in the space homogeneous setting, see for instance [11,48,49] and references therein). Such a deterministic contraction *a priori* holds under very restrictive conditions, in particular U should be strongly convex. In our case, and this is the second identified behavior, the calculation of the evolution of Z_t and W_t shows that there is still some deterministic contraction when $Z_t + W_t = 0$. We can therefore use a synchronous coupling in the vicinity of this subspace.

The last behavior, when the particles are outside of the contracting space $\{(z, w) \in \mathbb{R}^{2d}, z + w = 0\}$ and in the compact set where the Lyapunov function is not sufficient, we use noise to get the processes closer together. In order to maximize the variance of this noise, we use a so-called *reflection* coupling, which consists in B and B' being *antithetic* (i.e $B'_t = -B_t$) in the direction of space given by the difference of the processes, and synchronous in the orthogonal direction. In other words, writing

$$e_t = \begin{cases} \frac{Z_t + W_t}{|Z_t + W_t|} & \text{if } Z_t + W_t \neq 0, \\ 0 & \text{otherwise,} \end{cases}$$

we consider $dB'_t = \left(I_d - 2e_t e_t^T\right) dB_t$, where I_d denotes the d-dimensional identity matrix and e_t^T is the transpose of e_t. Levy's characterization then ensures that it is indeed a Brownian motion.

We are thus led to the study of a suitable distance between the two processes of the form

$$\rho_t := f(r_t)(1 + \epsilon H(X_t, V_t) + \epsilon H(X'_t, V'_t)),$$

with $r_t = \alpha|Z_t| + |Z_t + W_t|$, where $\alpha, \epsilon > 0$ and the function f need to be carefully chosen in order for $\mathbb{E}\rho_t$ to decay exponentially fast. This leads to several constraints, and we have to prove that it is possible to meet all these conditions simultaneously.

Let us highlight the main points of the construction of the semimetrics. First, in order to deal with the kinetic process (11), the standard Euclidean norm $|x|^2 + |v|^2$ is not suitable and one should instead consider a linear change of variables, like $(x, v) \mapsto (x, x + \beta v)$ for some $\beta \in \mathbb{R}$. This is the case when using coupling methods as in [7,25] but also when using hypocoercive modified entropies involving mixed derivatives as in [3,13,41,64,68] or modified Lyapunov functionals in [70] for instance, the link being made in [53]. This motivates the definition of r above. Then, when using a reflection coupling, because of the symmetry of the noise, there is *a priori* no reason why the noise should decrease r rather than increase it. To deal with this issue we modify this distance r by some concave function f. The concavity is well adapted to Itô's formula, enabling the diffusion to provide a contraction effect (in a compact set), as a random decrease in r has more effect on $f(r)$ than a random increase of the same amount. This method has been applied to elliptic diffusions in [24], see also [26]. Finally, we multiply the distance by a Lyapunov function G, which has first been used for Wasserstein distances in [39]. That way, on average, $f(r)G$ tends to decay because, when r is small, $f(r)$ tends to decay and, when r is large, G tends to decay.

We now write this coupling. Let $\xi > 0$, and let $rc,\ sc : \mathbb{R}^{2d} \to [0,1]$ be two Lipschitz continuous functions such that :

$$\begin{aligned}
&rc^2 + sc^2 = 1,\\
&rc\,(z,w) = 0 \text{ if } |z+w| \leq \frac{\xi}{2} \text{ or } \alpha|z| + |z+w| \geq R_1 + \xi,\\
&rc\,(z,w) = 1 \text{ if } |z+w| \geq \xi \text{ and } \alpha|z| + |z+w| \leq R_1.
\end{aligned}$$

These two functions translate into mathematical terms the regions in which we use a reflection coupling (represented by $rc = 1$) and the ones where we use a synchronous coupling (represented by $sc = 1$). Finally, ξ is a parameter that will vanish to zero in the end. We therefore consider the following coupling:

$$\begin{cases}
dX_t = V_t dt\\
dV_t = -V_t dt - \nabla U\,(X_t)\,dt - \nabla W * \mu_t\,(X_t)\,dt + \sqrt{2} rc\,(Z_t, W_t)\,dB_t^{rc} + \sqrt{2} sc\,(Z_t, W_t)\,dB_t^{sc}\\
\mu_t = \mathrm{Law}\,(X_t)\\
d\tilde{X}_t = \tilde{V}_t dt\\
d\tilde{V}_t = -\tilde{V}_t dt - \nabla U(\tilde{X}_t)dt - \nabla W * \tilde{\mu}_t(\tilde{X}_t)dt + \sqrt{2} rc\,(Z_t, W_t)\left(I_d - 2e_t e_t^T\right) dB_t^{rc} + \sqrt{2} sc\,(Z_t, W_t)\,dB_t^{sc}\\
\tilde{\mu}_t = \mathrm{Law}(\tilde{X}_t),
\end{cases}$$

where B^{rc} and B^{sc} are independent Brownian motions, and

$$Z_t = X_t - \tilde{X}_t, \quad W_t = V_t - \tilde{V}_t, \quad Q_t = Z_t + W_t, \quad e_t = \begin{cases} \frac{Q_t}{|Q_t|} & \text{if } Q_t \neq 0,\\ 0 & \text{otherwise.} \end{cases}$$

Calculations of the time evolution of the coupling semimetric yield

$$\forall t \geq 0, \qquad e^{ct}\rho_t \leq \rho_0 + \int_0^t e^{cs} K_s ds + M_t,$$

where M_t is a continuous local martingale and, by regrouping the terms according to how we will use them

$$\begin{aligned}
K_t = &\left(cf\,(r_t) + \left(\alpha \frac{d|Z_t|}{dt} + (L_U + L_W)|Z_t|\right) f'\,(r_t)\right) G_t && (15)\\
&+ 4\left(f''\,(r_t)\,G_t + 24\epsilon \max\left(1, \frac{1}{2\alpha}\right) r_t f'\,(r_t)\right) rc\,(Z_t, W_t)^2 && (16)\\
&+ \epsilon\left(2B - \gamma H\,(X_t, V_t) - \gamma H(\tilde{X}_t, \tilde{V}_t)\right) f\,(r_t) && (17)\\
&+ L_W f'\,(r_t)\,\mathbb{E}\,(|Z_t|)\,G_t + \epsilon L_W\,(6 + 8\lambda)\left(\mathbb{E}\,(|X_t|)^2 + \mathbb{E}(|\tilde{X}_t|)^2\right) f\,(r_t)\,. && (18)
\end{aligned}$$

Briefly,

- lines (15) and (16) will be non positive thanks to the construction of a 'sufficiently concave' function f when using the reflection coupling,
- when only using the synchronous coupling, i.e when the deterministic drift is contracting, line (15) alone will be sufficiently small,
- line (17) translates the effect the Lyapunov function H has in bringing back processes that would have ventured at infinity,
- finally, line (18) contains the non linearity and will be tackled by taking L_W sufficiently small.

Then, Gronwall's lemma allows us to conclude on the long time behavior of the non linear process, i.e Theorem 3.5, as the quantity ρ_t controls the L^1 and L^2 Wasserstein distances.

Obviously, for conciseness purposes, the reader is spared from the full calculations. We however wish to draw their attention to what may constitute one of the most interesting aspect of a coupling method. Notice how, in (15)-(18), the various expected behaviors of the stochastic particle can be observed. This shows that there is a back-and-forth between the probabilistic understanding of the dynamics and the calculations. The choice of the semimetrics was for instance motivated by the expected behavior of the processes, and conversely the difficulties appearing in the calculations may highlight some physical phenomena concerning the particles, that then motivate modifications in the studied quantities.

In order to prove Theorem 3.6 on propagation of chaos, the same method is applied, considering a coupling of the N particle system (10) with N independent copies of the non linear SDE (11). Some technical difficulties (for instance the expectations appearing in (18), coming from the non linear aspect of the processes, are replaced by empirical means) coming from the interactions between particles must be dealt with by modifying the function ρ, thus considering a new semimetrics.

We thus assemble various elements scattered in previous works such as [21] (which proves uniform in time propagation of chaos in the space homogeneous case), [25] (which focuses on the long-time of convergence of (11) without interactions, i.e in a linear case) and [26] (which proves the long-time of convergence of (11) in the space homogeneous case) and doing so requires to perform several careful estimates to improve upon these works and obtain both Theorems 3.5 and 3.6. We refer to [37] for more details.

4. Metastability for a system of interacting neurons

This section is based on [47], to which we refer for details and proofs.

4.1. The model and questions

Let us consider a very elementary model for a network of N interacting neurons, introduced in [35]. A neuron is only described by its membrane potential $u_i \geq 0$, $i \in [\![1, N]\!]$, and it emits spikes at a rate $\lambda(u_i)$. When a neuron spikes, its potential is reset to 0 and the potential of all the other neurons is increased by a fixed value h/N for some $h > 0$. Between these jumps, the potential depletes at constant rate, namely $\partial_t u_i = -\alpha u_i$ for some $\alpha > 0$. As a summary, the network is described by the Markov process $U^N = (U_1^N, \dots, U_N^N)$ on $\mathbb{R}_+^N$ with generator

$$A\varphi(u) = \sum_{i=1}^{N} \lambda(u_i) \left[\varphi(u + \Delta_i(u)) - \varphi(u)\right] - \alpha u \cdot \nabla\varphi(u),$$

where

$$(\Delta_i(u))_j = \begin{cases} \frac{h}{N} & j \neq i, \\ -u_i & j = i. \end{cases}$$

We assume that the jump rate is bounded, increasing (which means the interaction has an excitatory effect), Lipschitz, and that $\lambda(0) = 0$.

An interesting feature about this model is the following. Denote by $(T_n)_{n\in\mathbb{N}}$ the random jump times of the process. Duarte and Ost have proven in [20] that almost surely there is only a finite number of spikes, i.e.

$$L_N := \sup\{T_n, n \in \mathbb{N}, T_n < \infty\} < \infty.$$

This is similar to results on models for population dynamics which exhibit an almost sure extinction. However, an important difference is that L_N is not a stopping time. Indeed, the event $\{L_N \leq t\}$ is measurable with respect to the future trajectory $\{U^N(s), s > t\}$, not the past. After time L_N, the process goes at rate α toward zero, and in particular the Dirac mass at zero is the only invariant measure of the process.

Besides, when N is large, the number of spikes in a given time interval is of order N, approximately deterministic due to the law of large numbers, each of them inducing an increase of order $1/N$ on the potentials of

the neurons. It is clear that this mean field interaction will produce, as $N \to \infty$, a deterministic drift. In other words as $N \to \infty$, the trajectory of a neuron is expected to converge to a process $\bar{U}$ solving

$$d\bar{U}(t) = -\alpha \bar{U}(t)dt + h\mathbb{E}(\lambda(\bar{U}(t)))dt - \bar{U}(t-)\int_{\mathbb{R}_+} \mathbb{1}_{\{z \leq \lambda(\bar{U}(t-))\}} \pi(dt, dz), \tag{19}$$

where $\pi(dt, dz)$ is a Poisson random measure on $\mathbb{R}_+ \times \mathbb{R}_+$ having intensity $dtdz$. In other words, $\bar{U}$ is reset to 0 at rate $\lambda(\bar{U}(t))$ and, otherwise, follows the deterministic flow $\partial_t u = -\alpha u + h\mathbb{E}\left(\lambda(\bar{U}(t))\right)$.

Now, assume that there exists a non-zero stationary distribution of the non-linear equation satisfied by the law of $\bar{U}$. If we consider a system of N interacting neurons with independent initial conditions distributed according to this positive stationary distribution then, on any finite time interval, for N large enough, the system will behave similarly to N independent copies of the non-linear process (19), and in particular the empirical measure of the system should stay approximately constant. This implies that L_N should go to infinity in law as $N \to +\infty$. Again, we recover a situation which is classical for population models, where a stochastic birth-and-death chain with almost sure extinction converges to an ODE with a positive attractive stable point. Then, the process remains for very long time close to the deterministic equilibrium, before a large random deviation from this limit behavior leads to extinction. Moreover, in this case, due to some renewal mechanism, the extinction time, normalized by its expectation, converges in law toward an exponential variable, which means the extinction is unpredictable. This behavior (fast convergence toward a local equilibrium, long time spent there, unpredictable random exit toward another equilibrium) is called a metastable behavior.

The topic of [47] is the study of the metastable behavior of the system of interacting neurons U^N. More precisely, this means addressing the following questions: can we prove that L_N goes to $+\infty$ as $N \to +\infty$, with a quantitative speed? Are there non-zero stationary distributions for the non-linear limit process, and in that case are they stable and is the Dirac mass at zero unstable? Does $L_N/\mathbb{E}(L_N)$ converge to an exponential variable?

Some difficulties with respect to the study of birth-and-death processes are the following: the limit process is a PDE instead of a one-dimensional ODE. The state space of the process U^N depends on N. The last spike time L_N is not a stopping time. Also, by comparison with other classical works (for instance in the other sections of this proceeding) the non-linear limit process is not a diffusion, it corresponds to a non-local PDE with no regularization properties.

4.2. **The results**

Similarly to the population dynamics cases, the metastable phenomenon is not expected for all values of the parameters but only when the excitatory part is stronger than the inhibitory part. For simplicity, assume that $\lambda(u) = \min(ku, \lambda_*)$ for some parameters $k, \lambda_* > 0$ (the framework of [47] is slightly more general, but this particular form is needed for some results). When a neuron's potential is close to 0, its jump rate is increased by kh/N when another neuron spikes. On the other hand, when a neuron spikes, at most, its jump rate goes from λ_* to zero. Hence, setting $a = \alpha/(kh)$ and $b = \lambda_*/(kh)$, the excitatory regime corresponds to small values of a and b. The results of [47], which partially answer the previous questions, are represented in Figure 1 and listed here:

- First, if $a > 1$ (blue part of Figure 1), the only stationary distribution for the non-linear system is the Dirac mass at zero, and it is asymptotically stable: all other solutions converge exponentially fast toward this equilibrium (with the Wasserstein distance).
- On the contrary, if $a < 1$ (yellow, orange and red regions of Figure 1), there exists at least one non-zero stationary solution for the limit equation, and the solution zero is unstable: for all non-zero initial condition, $\liminf \mathbb{E}(\lambda(\bar{U}(t))) > 0$ as $t \to +\infty$.
- If $a + b < 1$ (orange and red parts of Figure 1), there exists an explicit $\kappa > 0$ such that

$$\mathbb{P}\left(L_N \geq e^{\kappa N}\right) \underset{N \to +\infty}{\longrightarrow} 1\,,$$

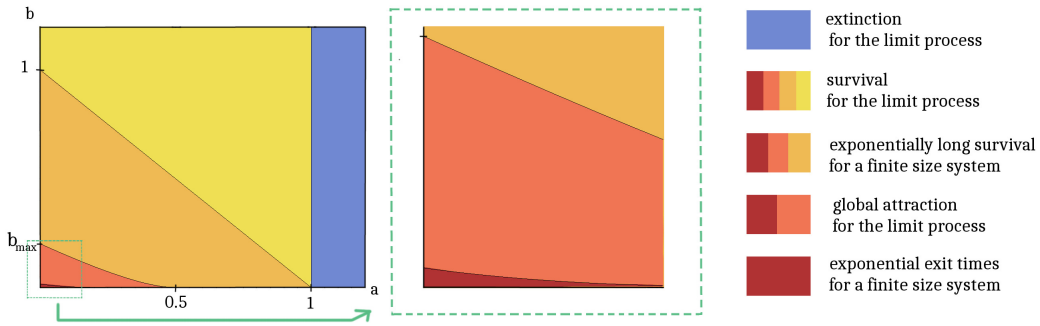

FIGURE 1. Summary of the results of [47], dependending on the parameters $a = \alpha/(kh)$ and $b = \lambda_*/(kh)$.

under mild conditions on the initial distribution (which hold for instance if initially the particles are independent and identically distributed according to a non-zero distribution).

- If a and b are small enough (with an explicit condition represented as the red parts in Figure 1), the non-zero stationary distribution for the limit process is unique and globally attractive: all solutions converge exponentially fast to it (with the Wasserstein distance).
- If a and b are small enough (with an explicit condition, stronger than the previous one, which is represented in deep red in Figure 1), considering the stopping time $\tau_N = \inf\{t \geq 0, \sum_{i=1}^N \lambda(U_i^N(t)) \leq N\gamma\}$ for some sufficiently small threshold $\gamma > 0$, then $\tau_N/\mathbb{E}(\tau_N)$ converges in distribution to a standard exponential variable.

As we see, the asymptotic exponentiality of $L_N/\mathbb{E}(L_N)$ is not established. However, τ_N is related to L_N, since the probability that $L_N \leq t$ given that $\tau_N > t$ is exponentially small with N and, conversely, if $L_N \leq t$, then the process will deterministically decay after time t and thus will fastly reach the time τ_N (see [47, Remark 3.3] for details on the comparison between τ_N and L_N).

We refer to [47] for the formal statements and the complete proofs of the results mentioned above. However, in the next section, let us discuss a key ingredient in these arguments.

4.3. **Synchronous couplings**

In our context, a synchronous coupling is a simultaneous definition of two different jump processes in such a way that, as much as possible, they jump at the same times. This can be done by using the same Poisson measure to define the jump times (see also [23] for a detailed construction). If λ_1 and λ_2 are the respective jump times of the processes $(X_t)_{t\geq 0}$ and $(Y_t)_{t\geq 0}$, then synchronous jumps for the pair (X_t, Y_t) occur at rate $\min(\lambda_1(X_t), \lambda_2(X_t))$ and asynchronous jumps (at which only the process with the highest jump rate jumps) occur at rate $|\lambda_1(X_t) - \lambda_2(Y_t)|$. This is somehow an analogous of the parallel coupling of two diffusion processes solution to some SDEs, where the same Brownian motion is used to drive both equations, as in some part of the space in the problem studied in Section 3. Notice that, by contrast to Section 3, in the present case, we will not use the noise to get two processes closer. If we wanted to, it would not be clear how to define an analogous of the reflection coupling of Section 3 for general jump processes, however we refer to [22] where this is done for a velocity jump process.

Most of the proofs of [47] are based on the synchronous coupling of two suitable processes.

First example: in order to obtain quantitative propagation of chaos estimates, namely the convergence in some sense of the interacting process U^N toward the limit process $\bar{U}$, an interacting system U^N is synchronously coupled with N independent copies of $\bar{U}$, with $\bar{U}_i(0) = U_i^N(0)$ (more precisely, for all $i \in [\![1, N]\!]$, the pair of particles $(U_i^N, \bar{U}_i)$ are synchronously coupled). We get that the distance between these two systems remains of

order $\sqrt{N}$ (and not N as would be the case of independent processes), which means the distance between U_i^N and $\bar{U}_i$ for a fixed $i \in [\![1, N]\!]$ is of order $1/\sqrt{N}$.

Second example: in order to obtain the long-time convergence of the limit system toward a unique stationary distribution, two non-linear processes $\bar{U}$ and $\hat{U}$ with different initial distributions are synchronously coupled. We get, first, that $\mathbb{E}|\lambda(\bar{U}(t)) - \lambda(\hat{U}(t))| \leq Ce^{-rt}$ for some $C, r > 0$, which in turns allows to prove that $\mathbb{E}|\bar{U}(t) - \hat{U}(t)| \leq Ce^{-rt}$, and thus a convergence in Wasserstein distance.

Last example: in order to bound L_N, one would like to understand the behavior of the average jump rate $\Lambda_N(t) = 1/N \sum_{i=1}^N \lambda(U_i^N(t))$. Unfortunately, this is not a Markov process, namely its evolution does not depend only on its value at time t but on the whole system $U^N(t)$. However, we can get some worst case bounds on the evolution of λ_N, namely $\partial_t \Lambda_N \geq -r\Lambda_N$ for some $r > 0$ between jumps and $\Lambda_N(T) \geq f(\Lambda_N(T-))$ for some explicit function f at a jump time T. Moreover, the jump times occur at rate $N\Lambda_N$. Thus, we can couple synchronously Λ_N with a Markov process $(Z_N(t))_{t\geq 0}$ on $\mathbb{R}_+$ which jumps at rate $NZ_N(t)$ to $f(Z_N(t-))$ and otherwise follows the deterministic flow $\partial_t Z_N = -rZ_N$. That way, we ensure that, almost surely, $Z_N(t) \leq \Lambda_N(t)$ and moreover all jump times of Z_N are jump times for Λ_N (and in particular L_N is larger than the last jump time of Z_N). Now, Z_N is a very simple one-dimensional Markov process which can be shown to converge to a deterministic ODE as $N \to +\infty$, to which standard Large Deviations results apply, yielding the desired bounds.

5. Rates of convergence to steady states for oscillator chains

The objective of this section is to present quantitative rates of convergence to a stationary state for a family of heat conducting systems consisting of a chain of interacting oscillators. This is based on the results of the articles [4, 52] to which we refer for the proofs.

The motivation for this study is the rigorous mathematical understanding of Fourier's law. Fourier's law is a physical macroscopic law that relates the local thermal flux $J(t, x)$ to small variations of temperature $\nabla T(t, x)$ through a proportionality constant $\kappa(T)$ known as *thermal conductivity*:

$$J(t, x) = -\kappa(T)\nabla T(t, x).$$

At the microscopic scale, matter is made out of particles assumed to evolve according to the classical laws of mechanics, and one of the goals of statistical physics is to model heat conductivity through a system of interacting atoms and to achieve a rigorous derivation of constitutive laws such as Fourier's law [8, 19, 32, 46]. Understanding macroscopic laws of matter when starting from a microscopic system of interacting atoms is a challenge addressed to mathematicians by Hilbert in his 6th problem [43].

A paradigmatic set up where Fourier's law is observed to hold with high precision is when one considers a fluid in a cylindrical slab of height h and uniform cross sectional area A, coupled at the two boundaries, the top and the bottom of the cylinder, to two heat reservoirs at different temperatures. This is known as the Benard experiment [8]. The two heat reservoirs keep the system out of equilibrium and produce a stationary heat flow. If there is a non-equilibrium steady state (NESS) that is described by a phase-space measure, one would like to prove that the following limit exists:

$$0 < \kappa(N) := \lim_{N\to\infty} \frac{\langle J^N(t, x)\rangle}{(A(\delta T/N))} < \infty$$

where N is the microscopic length of the cylinder,

$$\frac{\delta T}{N} = \frac{T_2 - T_1}{N}$$

is the effective temperature gradient, $\langle J^N(t,x)\rangle$ is the expectation of the heat flux with respect to the non equilibrium steady state and where we write $J^N(t,x)$ to stress the dependence of J on N. The above limit allows us to define the thermal conductivity and the very existence of the limit is a formulation of Fourier's law.

Our purpose is therefore to investigate how certain quantities, such as the relaxation rates to the NESS of such systems (the spectral gap of the associated dynamics), scale with the system size, since these are crucial to making sure that the thermal conductivity has a thermodynamic limit.

5.1. **The model and state of the art**

The model we focus on is a prototypical example of out-of-equilibrium systems and is a generalized version of the historical Fermi-Pasta-Ulam (FPU) chain. It consists of a chain of N interacting oscillators on the phase space $\mathbb{R}^{2dN}$, where the variables are q_i, p_i for $i = 1, \dots, N$: the displacements of the particles from their equilibrium positions and their momenta, respectively. Each particle has its own *pinning potential* and it interacts with its nearest neighbours through an *interaction potential*. We call H the Hamiltonian energy, whose precise expression as a function of the pairs (q_i, p_i) will be given below.

The dynamics of this model is such that the particles at the boundary are coupled to heat baths, modelled by *Langevin* (Ornstein–Uhlenbeck) processes at (possibly) different temperatures $\beta_i^{-1}, i \in \mathcal{F}$ and they are subject to friction. $\mathcal{F} \subset \{1, \dots, N\}$ here is the subset of the particles on which we impose friction and noise and we also denote by $\gamma_i > 0$ the friction strength at the i-th particle.

The time evolution is then for particles $i \in \{1, \dots, N\}$ described by a coupled system of SDEs:

$$\begin{aligned} dq_i(t) &= \nabla_{p_i} H dt, \\ dp_i(t) &= -\nabla_{q_i} H dt + \delta_{i\in\mathcal{F}} \left(-\gamma_i p_i \, dt + \sqrt{\frac{2\gamma_i}{\beta_i}} \, dW_i(t) \right). \end{aligned} \tag{20}$$

The generator of the dynamics when $\mathcal{F} = \{1, N\}$ is:

$$\mathcal{L} = \sum_{j=1}^{N} [\nabla_{p_j} H] \cdot \nabla_{q_j} - [\nabla_{q_j} H] \cdot \nabla_{p_j} - \gamma_1 p_1 \cdot \nabla_{p_1} - \gamma_N p_N \cdot \nabla_{p_N} + \gamma_1 T_L \Delta_{p_1} + \gamma_N T_R \Delta_{p_N}, \tag{21}$$

where T_L, T_R are the (possibly) different temperatures at the left and right boundary of the network of oscillators.

One of the main difficulties when studying out-of-equilibrium systems is the degeneracy of the systems. In particular, apart from the fact that the kinetic operator $\mathcal{L}$ is not self-adjoint, we notice that it is not elliptic, neither coercive. Therefore the validity of functional inequalities such as Poincare or Log-Sobolev inequalities is not straightforward.

5.1.1. *State of the art*

The non-equilibrium steady state for the purely harmonic chain, i.e. when both potentials are quadratic (harmonic), was made precise in [61]. Anharmonic chains were studied in various works [10, 17, 27, 28, 44, 60], where existence, uniqueness of a non-equilibrium steady state and exponential convergence towards it were proven in certain cases. More specifically the existence, uniqueness of a steady state and exponential convergence, hold under the assumptions that both the interaction and pinning potentials behave as polynomials near infinity and that the interaction is stronger than the pinning potential. The last assumption is important as there are some works which exhibit cases where the relaxation rate is not exponential, i.e. where there is lack of spectral gap [38, 40].

The proof of the above-mentioned results that provide convergence to the stationary measure relies on compactness arguments and thus they do not give information on how the speed of the convergence behaves as a function of N. Attempts have been made through hypocoercive techniques to get N-dependent estimates under certain conditions on the potentials: see the discussion in [68, Section 9.2] where this question was first raised. The techniques discussed in Villani's monograph however only yield non-optimal estimates.

5.2. **Main results**

5.2.1. *On the long time behavior*

Regarding the long time behavior of the system, we provide explicit rates of convergence to the non-equilibrium steady state (with optimal lower bound) in a 1-dimensional weakly anharmonic scenario, i.e. when both potentials are N-dependent perturbations of the harmonic ones.

The first statement concerns a contraction property in L^2 Wasserstein distance (see Section 1.2).

Theorem 5.1. *We consider a 1-dimensional chain of coupled oscillators with rigidly fixed edges so that the dynamics are described by the system* (20) *with*

$$H(p,q)=\sum_{i=1}^{N}\left(\frac{p_i^2}{2}+a\frac{q_i^2}{2}+U_{pin}^N(q_i)\right)+\sum_{i=1}^{N-1}\left(c\frac{(q_{i+1}-q_i)^2}{2}+U_{int}^N(q_{i+1}-q_i)\right)+c\frac{q_1^2}{2}+c\frac{q_N^2}{2}$$

for $a\geq 0, c>0$ and under the assumption that

$$\sup_{q_i}\|\nabla^2 U_{pin}^N(q_i)\|_2,\quad \sup_{r_i}\|\nabla^2 U_{int}^N(r_i)\|_2\leq C^N$$

where $r_i=q_{i+1}-q_i$ and $C^N\lesssim N^{-9/2}$. [1] *For a fixed number of particles N, there is a unique stationary state, in particular, for initial data f_0^1, f_0^2 we have:*

$$\mathcal{W}_2(P_t^* f_0^1, P_t^* f_0^2)\leq C_{a,c}N^{\frac{3}{2}}e^{-\frac{\lambda_0}{N^3}t}\ \mathcal{W}_2(f_0^1,f_0^2)$$

for $C_{a,c},\lambda_0$ dimensionless constants.

The proof relies on

- an application of a generalized version of the Γ_2-calculus of Bakry-Emery [2] for elliptic operators recently generalized by Baudoin for hypoelliptic operators [3] and
- a careful analysis of a high-dimensional matrix equation.

The generalised Γ_2-calculus allows us to prove the validity of a Log-Sobolev inequality for the invariant measure, with constant $C_N\lesssim N^3$. With this inequality in hand we also give a convergence to the stationary measure in relative entropy as in [68, Section 6]. We first recall the definitions of the following functionals:

For two probability measures μ and ν on $\mathbb{R}^{2N}$ with $\nu\ll\mu$, we define the Boltzmann H functional (relative entropy)

$$H_\mu(\nu)=\int_{\mathbb{R}^{2N}} h\log h\ d\mu,\ \nu=h\mu$$

and the relative Fisher information

$$I_\mu(\nu)=\int_{\mathbb{R}^{2N}}\frac{|\nabla h|^2}{h}d\mu,\ \nu=h\mu.$$

Theorem 5.2. *We consider a weakly anharmonic 1-dimensional chain of coupled oscillators with rigidly fixed edges whose dynamics are described by the system* (20) *under the same assumptions as in the Theorem 5.1 above. For a fixed number of particles N, assuming that (i) μ is the invariant measure for P_t and (ii) μ satisfies a Log-Sobolev inequality with constant $0<C_N\lesssim C_{T_L,\gamma}N^3$, for all $0<f\in L^1(\mu)$ with*

$$\mathcal{E}(f)<\infty,\ \textit{and}\ \int f d\mu=1,$$

[1] This is what we call a weakly anharmonic chain of oscillators.

we have a convergence to the non-equilibrium steady state in the following sense:

$$H_\mu(P_t f\mu) + I_\mu(P_t f\mu) \leq \lambda_{a,c} N^3 e^{-\lambda_0 N^{-3} t}(H_\mu(f\mu) + I_\mu(f\mu)) \tag{22}$$

for dimensionless constants $\lambda_{a,c}, \lambda_0$.

5.2.2. *On the spectral gap*

Furthermore, we study the spectral gap for purely harmonic chains and d-dimensional grids of oscillators, and prove the optimal lower and upper bounds. We also treat non-homogeneous scenarios where the coefficients of the pinning potentials are not identical. In particular we look at chains of oscillators with an impurity (so that the particle in the middle of the chain has pinning potential significantly weaker than the pinning potential of all the other particles) as well as at disordered chains of oscillators. As regards the d-dimensional grids, the spectral gap depends on which particles are exposed to friction. These are explained in the statement below.

Our setting is the following. From now on we use the notation $[N]^d := \{1, \dots, N\}^d$. We look at the system (20) with $\mathcal{F} \subset [N]^d$ and

$$H(q,p) = \frac{\langle p, m_{[N]^d}^{-1} p\rangle}{2} + V_{a,c}(q) \text{ where } V_{a,c}(q) = \sum_{i\in[N]^d} a_i |q_i|^2 + \sum_{i\sim j} c_{ij}|q_i - q_j|^2. \tag{23}$$

Theorem 5.3. *Let the positive masses* m_i *and interaction strengths* c_i *of all oscillators coincide,* N^d *be the number of oscillators, placed in a square grid with* N *oscillators on each side, and* d *the dimension of the network.*

- ***(Homogeneous chain):*** *Let the pinning strength* a_i *be the same for all oscillators, then*
  (1) *if two particles located at the corners* $(1,\dots,1), (N,\dots,N)$*, are exposed to the same non-zero friction and non-zero diffusion, the spectral gap of the generator decays at the optimal rate* N^{-3d}*:*

  $$\lambda_N = \mathcal{O}(N^{-3d}).$$

  In particular for the one-dimensional chain of oscillators $\lambda_N = \mathcal{O}(N^{-3})$.
  (2) *if the same non-zero friction and non-zero diffusion for particles located at the center of two opposite edges of the network*

  $$(1, \lceil N/2\rceil, \dots, \lceil N/2\rceil), (N, \lceil N/2\rceil, \dots, \lceil N/2\rceil),$$

  the spectral gap of the generator decays at the optimal rate $N^{-3-(d-1)}$*:* $\lambda_N = \mathcal{O}(N^{-3-(d-1)})$.
  (3) *if* $d = 2$ *and the particles exposed to the same non-zero friction are located at opposite edges of the network, the spectral gap satisfies* $\lambda_N \leq \mathcal{O}(N^{-5/2})$.
- ***(Chain with impurity):*** *Let* N *be even. We assume that all masses and interaction parameters are positive and coincide and the friction parameters* γ_i *of the boundary particles*

  $$\partial[N]^d := \{i \in [N]^d; \exists i_n : i_n \in \{1, N\}\} \quad of \quad [N]^d$$

  satisfy $\sup_{i\in\partial[N]^d} \gamma_i \in (0, c)$ *where* c *is independent of* N *and the friction is non-zero on at least one boundary edge. Then, if the pinning strength* $a_{c_d(N)}$ *at the center point* $c_d(N) = (N/2, .., N/2)$ *of the network is sufficiently small compared to the pinning strength of all other oscillators, the spectral gap* λ_N *of the generator decays at least exponentially fast in* N*, for all* $d \geq 1$.
  In dimension 1 *this rate is the optimal one.*
- ***(Disordered chain):*** *We assume that all masses and interaction parameters are positive and coincide and the friction parameters* γ_i *of the particles at the boundary*

  $$\partial[\pm N]^d := \{i \in [\pm N]^d; \|i\|_\infty = N\} \text{ of the network } [\pm N]^d := \{-N, ..., N\}^d$$

satisfy $\sup_{i\in\partial[\pm N]^d}\gamma_i \in (0,c)$ *where* c *is independent of* N *and the friction is non-zero on at least one boundary edge. Then, if the pinning strengths are iid random variables distributed according to some compactly supported density* $\rho \in C_c(0,\infty)$*, the spectral gap* λ_N *of the generator decays exponentially fast in* N*, for all* $d \geq 1$*, for all but finitely many* N*.*

The proof relies on an approach for studying non-symmetric spectral problems that reduces the problem to a spectral analysis of discrete Schrödinger operators. Using a *Wigner matrix* representation we reduce the study of this high dimensional spectral analysis to the study of resolvents involving only the heat bath sites.

The reason for the different behaviors of the spectral gap λ_N as a function of the number of particles N in these three different physical scenarios, is because the spectral gap of the generator (21) with the quadratic potentials (23) is *determined* -under a constraint on the spacing between its eigenvalues- by the decay rate of eigenstates of the discrete Schrödinger operator $-\Delta_{[N]^d} + V$. Here $\Delta_{[N]^d}$ is the Neumann Laplacian weighted with the interaction strengths and V is a potential fully defined through the pinning coefficients: $V = \sum_{i\in[N]^d} a_i\delta_i$, where $(\delta_i(u))(j) = \delta_{ij}u(i)$. Considering now that in the limit $N \to \infty$, the associated Schrödinger operator in the homogeneous setting has absolutely continuous spectrum with extended eigenstates, while in the strong impurity and the disordered setting it has -at least one- exponentially localized state, explains the polynomial decaying of the λ_N in the homogeneous case and the exponential decaying in the non-homogeneous settings.

References

[1] C. Ané, S. Blachère, D. Chafaï, P. Fougères, I. Gentil, F. Malrieu, C. Roberto, and G. Scheffer. *Sur les inégalités de Sobolev logarithmiques*, volume 10. Société Mathématique de France Paris, 2000.

[2] D. Bakry and M. Émery. Diffusions hypercontractives. In *Séminaire de probabilités, XIX, 1983/84*, volume 1123 of *Lecture Notes in Math.*, pages 177–206. Springer, Berlin, 1985.

[3] F. Baudoin. Bakry-Émery meet Villani. *J. Funct. Anal.*, 273(7):2275–2291, 2017.

[4] S. Becker and A. Menegaki. The optimal spectral gap for regular and disordered harmonic networks of oscillators. *J. Funct. Anal.*, 282(2):109286, 2022.

[5] S. Benachour, B. Roynette, D. Talay, and P. Vallois. Nonlinear self-stabilizing processes. I. Existence, invariant probability, propagation of chaos. *Stochastic Process. Appl.*, 75(2):173–201, 1998.

[6] S. Benachour, B. Roynette, and P. Vallois. Nonlinear self-stabilizing processes. II. Convergence to invariant probability. *Stochastic Process. Appl.*, 75(2):203–224, 1998.

[7] F. Bolley, A. Guillin, and F. Malrieu. Trend to equilibrium and particle approximation for a weakly selfconsistent Vlasov–Fokker–Planck equation. *ESAIM, Math. Model. Numer. Anal.*, 44(5):867–884, 2010.

[8] F. Bonetto, J. L. Lebowitz, and L. Rey-Bellet. Fourier's law: a challenge to theorists. In *Mathematical physics 2000*, pages 128–150. Imp. Coll. Press, London, 2000.

[9] D. Bresch, P-E. Jabin, and Z. Wang. On mean-field limits and quantitative estimates with a large class of singular kernels: Application to the Patlak–Keller–Segel model. *Comptes Rendus. Mathématique*, 357(9):708–720, 2019.

[10] P. Carmona. Existence and uniqueness of an invariant measure for a chain of oscillators in contact with two heat baths. *Stochastic Process. Appl.*, 117(8):1076–1092, 2007.

[11] J. A. Carrillo, R. J. McCann, and C. Villani. Contractions in the 2-Wasserstein length space and thermalization of granular media. *Arch. Ration. Mech. Anal.*, 179(2):217–263, 2006.

[12] P. Cattiaux, A. Guillin, and F. Malrieu. Probabilistic approach for granular media equations in the non-uniformly convex case. *Probab. Theory Related Fields*, 140(1-2):19–40, 2008.

[13] P. Cattiaux, A. Guillin, P. Monmarché, and C. Zhang. Entropic multipliers method for Langevin diffusion and weighted log Sobolev inequalities. *J. Funct. Anal.*, 277(11):108288, 24, 2019.

[14] P. Cattiaux and L. Pédèches. The 2-D stochastic Keller-Segel particle model: existence and uniqueness. *ALEA*, 13:447–463, 2016.

[15] L.-P. Chaintron and A. Diez. Propagation of chaos: A review of models, methods and applications. I. Models and methods. *Kinet. Relat. Models*, 15(6):895–1015, 2022.

[16] L.-P. Chaintron and A. Diez. Propagation of chaos: A review of models, methods and applications. II. Applications. *Kinet. Relat. Models*, 15(6):1017–1173, 2022.

[17] N. Cuneo, J.-P. Eckmann, M. Hairer, and L. Rey-Bellet. Non-equilibrium steady states for networks of oscillators. *Electron. J. Probab.*, 23:28 pp., 2018.

[18] D. A. Dawson and J. Gärtner. Large deviations, free energy functional and quasi-potential for a mean field model of interacting diffusions. *Mem. Amer. Math. Soc.*, 78(398):iv+94, 1989.

[19] A. Dhar. Heat transport in low-dimensional systems. *Advances in Physics*, 57, 08 2008.

[20] A. Duarte and G. Ost. A model for neuronal activity in the absence of external stimuli. *Markov Process. Related. Fields*, 22:37–52, 2016.
[21] A. Durmus, A. Eberle, A. Guillin, and R. Zimmer. An elementary approach to uniform in time propagation of chaos. *Proc. Amer. Math. Soc.*, 148(12):5387–5398, 2020.
[22] A. Durmus, A. Guillin, and P. Monmarché. Geometric ergodicity of the bouncy particle sampler. *Ann. Appl. Probab.*, 30(5):2069–2098, 10 2020.
[23] A. Durmus, A. Guillin, and P. Monmarché. Piecewise deterministic Markov processes and their invariant measures. *Ann. Inst. Henri Poincaré Probab. Stat.*, 57(3):1442 – 1475, 2021.
[24] A. Eberle. Reflection couplings and contraction rates for diffusions. *Probab. Theory Relat. Fields*, 166(3-4):851–886, 2016.
[25] A. Eberle, A. Guillin, and R. Zimmer. Couplings and quantitative contraction rates for Langevin dynamics. *Ann. Probab.*, 47(4):1982–2010, 2019.
[26] A. Eberle, A. Guillin, and R. Zimmer. Quantitative Harris-type theorems for diffusions and McKean–Vlasov processes. *Trans. Am. Math. Soc.*, 371(10):7135–7173, 2019.
[27] J.-P. Eckmann, C.-A. Pillet, and L. Rey-Bellet. Entropy production in nonlinear, thermally driven Hamiltonian systems. *J. Statist. Phys.*, 95(1-2):305–331, 1999.
[28] J.-P. Eckmann, C.-A. Pillet, and L. Rey-Bellet. Non-equilibrium statistical mechanics of anharmonic chains coupled to two heat baths at different temperatures. *Comm. Math. Phys.*, 201(3):657–697, 1999.
[29] F. Flandoli, M. Leimbach, and C. Olivera. Uniform convergence of proliferating particles to the FKPP equation. *J. Math. Anal. Appl.*, 473(1):27–52, 2019.
[30] F. Flandoli and M. Leocata. A particle system approach to aggregation phenomena. *J. Appl. Probab.*, 56(1):282–306, 2019.
[31] F. Flandoli, C. Olivera, and M. Simon. Uniform approximation of 2 dimensional Navier-Stokes equation by stochastic interacting particle systems. *SIAM J. Math. Anal.*, 52(6):5339–5362, 2020.
[32] P. Flandrin and C. Bernardin, editors. *Fourier and the Science of Today / Fourier et la Science d'aujourd'hui*, volume 20, Issue 5. Comptes Rendus Physique, 2019.
[33] N. Fournier and B. Jourdain. Stochastic particle approximation of the Keller–Segel equation and two-dimensional generalization of Bessel processes. *Ann. Appl. Probab.*, 27(5):2807–2861, 2017.
[34] N. Fournier and Y. Tardy. Collisions of the supercritical Keller-Segel particle system. Preprint arXiv, 2023.
[35] A. Galves and E. Löcherbach. Infinite systems of interacting chains with memory of variable length–a stochastic model for biological neural nets. *J. Stat. Phys. 151*, 5:896–921, 2013.
[36] A. M. Garsia, E. Rodemich, and H. Rumsey, Jr. A real variable lemma and the continuity of paths of some Gaussian processes. *Indiana Univ. Math. J.*, 20:565–578, 1970/71.
[37] A. Guillin, P. Le Bris, and P. Monmarché. Convergence rates for the Vlasov-Fokker-Planck equation and uniform in time propagation of chaos in non convex cases. *Electron. J. Probab.*, 27:Paper No. 124, 44, 2022.
[38] M. Hairer. How hot can a heat bath get? *Comm. Math. Phys.*, 292(1):131–177, 2009.
[39] M. Hairer and J. C. Mattingly. Spectral gaps in Wasserstein distances and the 2D stochastic Navier-Stokes equations. *Ann. Probab.*, 36(6):2050–2091, 2008.
[40] M. Hairer and J. C. Mattingly. Slow energy dissipation in anharmonic oscillator chains. *Comm. Pure Appl. Math.*, 62(8):999–1032, 2009.
[41] F. Hérau. Short and long time behavior of the Fokker-Planck equation in a confining potential and applications. *J. Funct. Anal.*, 244(1):95–118, 2007.
[42] M. A. Herrero and J. J. L. Velázquez. Singularity patterns in a chemotaxis model. *Math. Ann.*, 306(3):583–623, 1996.
[43] D. Hilbert. Mathematical problems. *Bull. Amer. Math. Soc.*, 8(10):437–479, 07 1902.
[44] V. Jakšić and C.-A. Pillet. Ergodic properties of classical dissipative systems I. *Acta Math.*, 181(2):245–282, 1998.
[45] M. Kac. Foundations of kinetic theory. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 3: Contributions to Astronomy and Physics*, pages 171–197, Berkeley, Calif., 1956. University of California Press.
[46] S. Lepri. *Thermal Transport in Low Dimensions: From Statistical Physics to Nanoscale Heat Transfer*, volume 921. 01 2016.
[47] E. Löcherbach and P. Monmarché. Metastability for systems of interacting neurons. *Ann. Inst. Henri Poincaré Probab. Stat.*, 58(1):343–378, 2022.
[48] F. Malrieu. Logarithmic Sobolev inequalities for some nonlinear PDE's. *Stochastic Process. Appl.*, 95(1):109–132, 2001.
[49] F. Malrieu. Convergence to equilibrium for granular media equations and their Euler schemes. *Ann. Appl. Probab.*, 13(2):540–560, 2003.
[50] H. P. McKean, Jr. A class of Markov processes associated with nonlinear parabolic equations. *Proc. Nat. Acad. Sci. U.S.A.*, 56:1907–1911, 1966.
[51] H. P. McKean, Jr. Propagation of chaos for a class of non-linear parabolic equations. In *Stochastic Differential Equations (Lecture Series in Differential Equations, Session 7, Catholic Univ., 1967)*, pages 41–57. Air Force Office Sci. Res., Arlington, Va., 1967.
[52] A. Menegaki. Quantitative Rates of Convergence to Non-equilibrium Steady State for a Weakly Anharmonic Chain of Oscillators. *J. Stat. Phys.*, 181(1):53–94, 2020.

[53] P. Monmarché. Generalized Γ calculus and application to interacting particles on a graph. *Potential Anal.*, 50(3):439–466, 2019.
[54] T. Nagai. Global existence and decay estimates of solutions to a parabolic-elliptic system of drift-diffusion type in $\mathbb{R}^2$. *Differential Integral Equations*, 24(1–2):29–68, 2011.
[55] K. Oelschläger. A law of large numbers for moderately interacting diffusion processes. *Z. Wahrsch. Verw. Gebiete*, 69(2):279–322, 1985.
[56] C. Olivera, A. Richard, and M. Tomašević. Quantitative approximation of the Keller-Segel and Burgers equations by moderately interacting particles. Preprint arXiv, 2020.
[57] C. Olivera, A. Richard, and M. Tomašević. Quantitative particle approximation of nonlinear Fokker-Planck equations with singular kernel. *Ann. Sc. Norm. Super. Pisa Cl. Sci. (5)*, page 44, 12 2022.
[58] E. Olivieri and M. E. Vares. *Large deviations and metastability*, volume 100 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 2005.
[59] B. Perthame. PDE models for chemotactic movements: parabolic, hyperbolic and kinetic. *Appl. Math.*, 49(6):539–564, 2004.
[60] L. Rey-Bellet and L. E. Thomas. Exponential convergence to non-equilibrium stationary states in classical statistical mechanics. *Comm. Math. Phys.*, 225(2):305–329, 2002.
[61] Z. Rieder, J. L. Lebowitz, and E. Lieb. Properties of a harmonic crystal in a stationary nonequilibrium state. *J. Math. Phys.*, 8(5):1073–1078, 1967.
[62] M. Simon and C. Olivera. Non-local conservation law from stochastic particle systems. *J. Dynam. Differential Equations*, 30(4):1661–1682, 2018.
[63] A.-S. Sznitman. Topics in propagation of chaos. In *École d'Été de Probabilités de Saint-Flour XIX—1989*, volume 1464 of *Lecture Notes in Math.*, pages 165–251. Springer, Berlin, 1991.
[64] D. Talay. Stochastic Hamiltonian systems: exponential convergence to the invariant measure, and discretization by the implicit Euler scheme. *Markov Process. Related Fields*, 8(2):163–198, 2002.
[65] Y. Tardy. Convergence of the empirical measure for the Keller-Segel model in both subcritical and critical cases. Preprint arXiv, 2022.
[66] M. Tomašević. *On a prbabilistic interpretation of the parabolic-parabolic Keller-Segel model*. PhD thesis, Université Côte d'Azur, 2018.
[67] J. M. A. M. van Neerven, M. C. Veraar, and L. Weis. Stochastic integration in UMD Banach spaces. *Ann. Probab.*, 35(4):1438–1478, 2007.
[68] C. Villani. Hypocoercivity. *Mem. Amer. Math. Soc.*, 202(950):iv+141, 2009.
[69] C. Villani. *Optimal transport*, volume 338 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2009. Old and new.
[70] L. Wu. Large and moderate deviations and exponential convergence for stochastic damping Hamiltonian systems. *Stochastic Process. Appl.*, 91(2):205–238, 2001.