# PHASE TRANSITIONS, LOGARITHMIC SOBOLEV INEQUALITIES, AND UNIFORM-IN-TIME PROPAGATION OF CHAOS FOR WEAKLY INTERACTING DIFFUSIONS

BY MATÍAS G. DELGADINO[1], RISHABH S. GVALANI[2], GRIGORIOS A. PAVLIOTIS[3] AND SCOTT A. SMITH[4]

[1]*Department of Mathematics, The University of Texas at Austin* matias.delgadino@math.utexas.edu

[2]*Max-Planck-Institut für Mathematik in den Naturwissenschaften* gvalani@mis.mpg.de

[3]*Department of Mathematics, Imperial College London* g.pavliotis@imperial.ac.uk

[4]*Academy of Mathematics and Systems Sciences, Chinese Academy of Sciences* ssmith@amss.ac.cn

**Abstract** In this article, we study the mean field limit of weakly interacting diffusions for confining and interaction potentials that are not necessarily convex. We explore the relationship between the large $N$ limit of the constant in the logarithmic Sobolev inequality (LSI) for the $N$-particle system and the presence or absence of phase transitions for the mean field limit. The non-degeneracy of the LSI constant is shown to have far reaching consequences, especially in the context of uniform-in-time propagation of chaos and the behaviour of equilibrium fluctuations. Our results extend previous results related to unbounded spin systems and recent results on propagation of chaos using novel coupling methods. As incidentals, we provide concise and, to our knowledge, new proofs of a generalised form of Talagrand's inequality and of quantitative propagation of chaos by employing techniques from the theory of gradient flows, specifically the Riemannian calculus on the space of probability measures.

**1. Introduction.** Interacting particle systems have attracted a lot of attention in recent years since they appear in diverse areas ranging from plasma physics and galactic dynamics to machine learning and optimization. For systems of identical (or exchangeable) particles in which the pair-wise interactions scale like the inverse of the number of particles, it is possible to pass to the mean field limit and obtain a coarse-grained description of the system via a nonlinear nonlocal PDE that governs the evolution of the one-particle density. In this paper, we consider systems of weakly interacting diffusions driven by pair-wise interactions, confinement and independent Brownian motions (see (2.1) ). In this case the mean field PDE is the so-called McKean–Vlasov equation.

A natural problem that one would like to address is how to obtain sharp quantitative estimates on the rate at which the empirical measure of the particle system converges to the mean field limit, as the number of particles $N$ goes to infinity. When considering arbitrarily long time scales, this problem is intimately connected to the rate of convergence to steady states as time $t$ goes to infinity. For the study of such quantitative results, a crucial role is played by the Poincaré (PI) and logarithmic Sobolev (LSI) inequalities. Our focus in this paper is to elucidate the connection between the validity of the LSI for the $N$-particle Gibbs measure uniformly in the number of particles $N$ and the properties of the mean field limit. We establish connections with uniform-in-time propagation of chaos, (non-)uniqueness of steady states of the mean field equation, exponential convergence to equilibrium, and the behaviour

*MSC2020 subject classifications:* Primary 60K35, 82B26; secondary 39B62.

*Keywords and phrases:* Interacting particle systems, log Sobolev inequalities, phase transitions, propagation of chaos.

of equilibrium fluctuations. We show the effect that the presence of multiple steady states for the mean field equation, which correspond to invariant measures of the associated nonlinear McKean SDE, has on the various quantitative and qualitative features of the underlying interacting particle system.

In the equilibrium statistical mechanics of lattice systems, the non-uniqueness of the infinite-volume grand canonical Gibbs measure is referred to as a phase transition [30]. The presence of such a phase transition can then be detected with the help of some order parameter, for example the thermodynamic pressure for the nearest-neighbour Ising model, or the average magnetisation of the ensemble for the mean-field Curie–Weiss model. Based on the behaviour of these quantities (or rather on the behaviour of the infinite-volume partition function), one can then characterise the phase transitions as either continuous (second-order) or discontinuous (first-order).

The situation in our setting is more complicated but closely mirrors the one for lattice systems. Indeed, the system we consider can be thought of as spin system with mean field interaction and the "spins" taking values in some uncountable state space $\Omega$. The difference between the system we consider and the Ising and Curie–Weiss model lies in the fact that, except in a small number of specific examples, it is extremely hard to specify an order-parameter or understand the exact behaviour of the infinite-volume partition function. An additional important difference is that for our system we will be more interested in the non-uniqueness of critical points of the free energy, which correspond to important changes in the features of the system, as opposed to non-uniqueness of its minimisers. In either case, the similarity with spin systems is instructive enough that it will serve the reader well to remember this analogy as we discuss the notion of phase transition we work with.

A detailed characterisation of phase transitions for McKean-Vlasov PDEs on the torus without a confining potential was given in [11]. In particular, the presence of phase transitions for this setting as it relates to non-uniqueness of minimisers of the free energy functional was discussed in detail. However, in this paper, we are more interested in characterising these phase transitions as they relate to non-uniqueness of critical points of the free energy.

For convex confining and interaction potentials (when the state space is Euclidean), the system does not undergo phase transitions. In fact, uniform-in-time propagation of chaos and uniqueness of the steady state for the mean field PDE have been established, for e.g. in [44]. Moreover, in [13] under a uniform convexity assumption of the potentials, the authors show exponentially fast relaxation to the unique steady state of the mean field system. Our focus in this paper is to deal with non-convex potentials which may exhibit phase transitions and thus could not be expected to always (for all temperatures) exhibit uniform-in-time propagation of chaos. In terms of the LSI, we show that the existence of a non-minimising steady state implies the (quantitative) degeneracy of the constant in the LSI of the $N$-particle Gibbs measure in the limit as $N \to +\infty$. On the other hand, we show that the non-degeneracy of the LSI, implies uniform-in-time propagation of chaos and Gaussianity of the fluctuations around the mean field limit at equilibrium. Furthermore, we conjecture that the limit of the LSI constant for the $N$-particle system can be uniquely characterised in terms of the dissipation inequality for the mean field system.

The relation between the non-degeneracy of the constant in the PI or the LSI, the absence of phase transitions, and the exponentially fast decay of correlations has been studied extensively for unbounded spin systems [63]. Conversely, in these works the equivalence between the slow decay of correlations and the fact that the constant in the LSI becomes degenerate at the phase transition has been established. Uniform estimates on the constant in the LSI beyond the convex case have been established recently [32], under the Lipschitzian spectral gap condition for the single particle. We remark that this assumption is reminiscent to the assumptions on the conditional measures for the two-scale LSI [49, 31, 41]. We utilize the

latter approach to show the non-degeneracy of the LSI for our $N$-particle Gibbs measure in the high temperature/weak interaction regime.

In the probability literature, the study of the LSI in the context of linear Fokker–Planck equations goes back to the classical $\Gamma_2$ functional introduced by Bakry and Emery [3]. More recently, contractivity for interacting particle systems has been studied in the context of entropic interpolation and Schrödinger bridges [51, 29, 2, 28]. These techniques yield proofs of both the Talagrand [16] and Sobolev [22] inequality, under lower curvature conditions on the underlying manifold. We also mention the novel coupling techniques introduced by Eberle in [24] that produce contractivity estimates in a tailored transportation cost distance. This approach was then later used to prove uniform-in-time propagation of chaos estimates under a smallness assumptions on the interaction potential [23].

Our approach in this paper exploits the fact that both the $N$-particle system (or rather its Fokker–Planck equation) and its mean field limit are gradient flows of a particular energy functional with respect to the 2-Wasserstein distance. We can use this structural feature of the system to study the limit of all the relevant quantities as $N \to \infty$. This approach was pioneered by Hauray and Mischler in [35], and later used by the authors in [10, 20] to study both propagation of chaos and periodic homogenization for the interacting particle system. The advantage of this approach is that we can often make minimal assumptions on the regularity of the confining and interaction potentials: we will essentially assume that they are both only semi-convex, which is natural for 2-Wasserstein gradient flows (see [1]). We refer to [54, 38, 9] for the reader interested in propagation of chaos results with more singular potentials.

*Organization of the paper.*. Section 2 sets up the problem we are interested in studying along with our notation and main assumptions. Section 3 contains the statements of all our main results. Section 4 connects our results to the phenomenon of phase transitions and discusses possible properties that could capture the radical change of behavior in our system in the presence multiple steady states of the mean field equation. Section 5 contains some technical results on the convergence of the relevant quantities and functionals as $N \to \infty$ which play an important role in the proofs of our main theorems. Sections 6 to 12 contain the proofs of Theorems 3.2 to 3.4, 3.6, 3.7, 3.11 and 3.14, respectively.

**2. Set up, assumptions, and notation.** We consider $\{X_t^i\}_{i=1,\dots,N} \subset \mathbb{R}^d$, the positions of $N$ indistinguishable interacting particles at time $t \geq 0$, satisfying the following system of SDEs:

$$\begin{cases} dX_t^i = -\nabla V(X_t^i)\,\mathrm{d}t - \dfrac{1}{N}\displaystyle\sum_{j=1}^{N} \nabla_1 W(X_t^i, X_t^j)\,\mathrm{d}t + \sqrt{2\beta^{-1}}dB_t^i \\ \mathrm{Law}(X_0^1, \dots, X_0^N) = \rho_{\mathrm{in}}^{\otimes N} \in \mathcal{P}_{2,\mathrm{sym}}((\mathbb{R}^d)^N), \end{cases} \tag{2.1}$$

where $V : \mathbb{R}^d \to \mathbb{R}$, $W : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$, $\beta^{-1} > 0$ is the inverse temperature, $B_t^i, i = 1, \dots, N$ are independent $d$-dimensional Brownian motions, and the initial position of the particles is i.i.d with law $\rho_{\mathrm{in}}$. The chaoticity assumption on the initial data is not necessary but it greatly simplifies the exposition. Similarly, the state space $\mathbb{R}^d$ can be replaced by the periodic domain $\mathbb{T}^d$ or any convex set $\Omega \subset \mathbb{R}^d$ with normal reflecting boundary conditions, see, for example, [56]. We denote the space of symmetric Borel probability measures on $\Omega^N$ with finite second moment by $\mathcal{P}_{2,\mathrm{sym}}(\Omega^N)$, i.e. probability measures which are invariant under the relabeling of variables (or probability measures that arise as laws of exchangeable random variables). Throughout the paper, we will always work with probability measures that have finite second moment; to avoid burdensome notation, we forego the subscript 2 from now on and simply write $\mathcal{P}_{\mathrm{sym}}(\Omega^N)$.

To ensure well-posedness of the evolutionary flows and coercivity, we make the following assumptions.

ASSUMPTION 2.1. The confining potential $V$ is lower semicontinuous, bounded below, $K_V$-convex for some $K_V \in \mathbb{R}$ and there exists $R_0 > 0$ and $\delta > 0$, such that $V(x) \geq |x|^\delta$ for $|x| > R_0$.

ASSUMPTION 2.2. The interaction potential $W$ is lower semicontinuous, $K_W$-convex for some $K_W \in \mathbb{R}$, bounded below, symmetric $W(x,y) = W(y,x)$, vanishes along the diagonal $W(x,x) = 0$, and there exists $C$ such that

$$|\nabla_1 W(x,y)| \leq C(1 + |W(x,y)| + V(x) + V(y)) \tag{2.2}$$

REMARK 2.3. The $K$-Convexity assumptions on the potentials is short hand for global lower bounds on their Hessians

$$D^2 V \geq K_V I^{d\times d} \qquad \text{and} \qquad D^2 W \geq K_W I^{2d\times 2d}$$

with $K_V$, $K_W \in \mathbb{R}$. Unlike a convexity assumption on the potentials, i.e. $K$-convexity with $K = 0$ (see [44] for results in the convex case), these assumptions are weak enough to include models that exhibit phase transitions with respect to changes in the system's temperature, $\beta^{-1}$, for example, the double well potential $V(x) = (1 - |x|^2)^2$ with quadratic interactions $W(x,y) = |x-y|^2$, also known as the Desai–Zwanzig model (for more details see [18].

REMARK 2.4. The more technical bound (2.2) replaces the more classical doubling condition [1, Section 10.4.42] which is used to characterise the minimal sub-differential of the interaction energy [1, Theorem 10.4.11].

2.1. *The Fokker–Planck equation..* It is well known that the curve $\rho^N : [0,\infty) \to \mathcal{P}_{\text{sym}}(\Omega^N)$ which describes the evolution of the law of the process $(X_t^1, ..., X_t^N) \in \Omega^N$ satisfies the following linear Fokker–Planck equation

$$\begin{cases} \partial_t \rho^N = \beta^{-1}\Delta\rho^N + \nabla\cdot(\rho^N \nabla H_N) & \text{in } (0,\infty)\times\Omega^N, \\ (\nabla\rho^N + \rho^N \nabla H_N)\cdot \vec{n}_{\Omega^N} = 0 & \text{on } (0,\infty)\times\partial\Omega^N \\ \rho^N(0) = \rho_{\text{in}}^{\otimes N}, \end{cases} \tag{2.3}$$

where the Hamiltonian $H_N$ is given by

$$H_N(x) = \sum_{i=1}^N V(x_i) + \frac{1}{2N}\sum_{j=1}^N\sum_{i=1}^N W(x_i, x_j),$$

$\rho_{\text{in}} \in \mathcal{P}(\Omega)$, and $\vec{n}_{\Omega^N}$ is the unit normal to $\Omega^N$.

2.2. *de Finetti/Hewitt–Savage..* To take the limit as $N \to \infty$, we will crucially use the exchangeability of the underlying particle system whose law is governed by (2.3). This implies that the joint law $\rho^N$ is symmetric for all times, that is

$$\rho^N(t) \in \mathcal{P}_{\text{sym}}(\Omega) \qquad \text{for all } t \geq 0.$$

The main idea is that we can characterise the limit $N \to \infty$ of $\mathcal{P}_{\mathrm{sym}}(\Omega^N)$ as $\mathcal{P}(\mathcal{P}(\Omega))$, which denotes the Borel probability measures with bounded second moment defined over the metric space $(\mathcal{P}(\Omega), d_2)$, where $d_2$ is the 2-Wasserstein distance.

Following de Finetti [19] and Hewitt–Savage [37], we know that any tight sequence $(\rho^N)_{N\in\mathbb{N}}$, with $\rho^N \in \mathcal{P}_{\mathrm{sym}}(\Omega^N)$, i.e. with tight (in $\mathcal{P}(\Omega)$) $l^{\text{th}}$ marginals for all $l \in \mathbb{N}$, has a limit $P_\infty \in \mathcal{P}(\mathcal{P}(\Omega))$ along a subsequence which we do not relabel such that

$$\rho^N \rightharpoonup P_\infty,$$

where weak convergence is given by duality with cylindrical test functions. That is, for any $l \in \mathbb{N}$ and $\varphi \in C_c(\Omega^l)$, we have

$$\lim_{N\to\infty} \int_{\Omega^l} \varphi(y)\ \mathrm{d}\rho_l^N(y) = \int_{\mathcal{P}(\Omega)} \left( \int_{\Omega^l} \varphi(y)\ \mathrm{d}\rho^{\otimes l}(y) \right) \mathrm{d}P_\infty(\rho), \tag{2.4}$$

where

$$\rho_l^N \in \mathcal{P}_{\mathrm{sym}}(\Omega^l) \qquad \text{is the } l\text{-th marginal of } \rho^N.$$

In essence, this means that in the limit $N \to \infty$ of symmetric probability measures can be characterised as convex combinations of chaotic measures. For more details, we refer the reader to [35, 10, 52]. In the sequel, we will use the notation $\rho^N \rightharpoonup P_\infty \in \mathcal{P}(\mathcal{P}(\Omega))$ to denote this notion of weak convergence for any sequence $(\rho^N)_{N\in\mathbb{N}}$ with $\rho^N \in \mathcal{P}_{\mathrm{sym}}(\Omega^N)$.
Moreover, a metric version of this result can be obtained by considering the appropriately scaled 2-Wasserstein distance, i.e.

$$\overline{d}_2 := \frac{1}{\sqrt{N}} d_2,$$

where $d_2$ is the classical 2-Wasserstein on $\mathcal{P}(\Omega^N)$. More specifically, Hauray–Mischler [35] showed that under the topology of (2.4)

$$\lim_{N\to\infty} \overline{d}_2 = \mathfrak{D}_2,$$

where $\mathfrak{D}_2$ is the 2-Wasserstein distance defined on the probabilities with second moment bounded over the metric space $(\mathcal{P}_2(\mathbb{R}^d), d_2)$. For more details, see Theorem 5.3. This metric on $\mathcal{P}(\mathcal{P}(\Omega))$ is closely related to the convergence discussed above as will be seen in Proposition 5.5.

2.3. *The mean field limit and nonlinear behaviour.* Within this formalism, the limit $N \to \infty$ of the equation (2.3) can be written as

$$\rho^N(t) \rightharpoonup \delta_{\rho(t)} \in \mathcal{P}(\mathcal{P}(\Omega)) \qquad \text{for all } t \geq 0,$$

where $\delta_{\rho(t)}$ denotes the delta measure concentrated on $\rho(t)$, the unique solution of the nonlinear McKean equation

$$\begin{cases} \partial_t \rho = \beta^{-1}\Delta\rho + \nabla\cdot(\rho\,(\nabla V + \nabla W \star \rho)) & \text{on } (0,\infty)\times\Omega \\ (\nabla\rho + \rho(\nabla V + \nabla W \star \rho)\cdot \vec{n}_\Omega = 0 & \text{on } (0,\infty)\times\partial\Omega \\ \rho(0) = \rho_{\mathrm{in}}, \end{cases} \tag{2.5}$$

with

$$W \star \rho(x) := \int_{\mathbb{R}^d} W(x,y)\ \mathrm{d}\rho(y)\,.$$

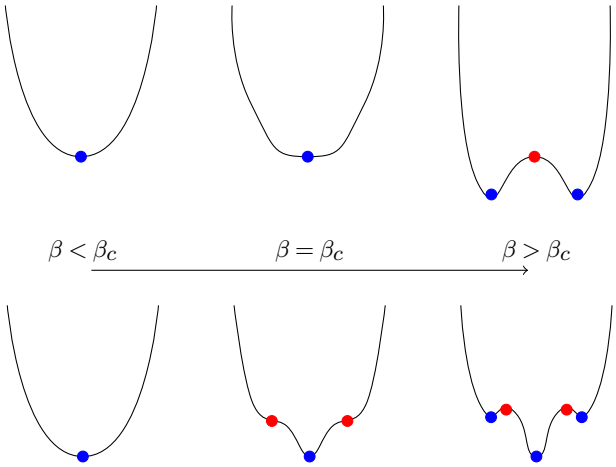


FIG 1. *A rough schematic showing two possible kinds of phase transition: The upper diagram shows a typical continuous phase transition. In this setting, the unique critical point (shown in blue) loses its local stability through a local (pitchfork) bifurcation which gives rise to new locally stable critical points. The lower diagram shows a typical discontinuous phase transition. In this setting, the unique critical point retains its local stability but new critical points arise in the free energy landscape through a saddle node bifurcation.*

One of the most salient differences between the particle dynamics (2.3) and the mean field dynamics (2.5) is that, whereas the Fokker-Planck equation governing the evolution of the $N$-particle system is linear, the mean field PDE (2.5) is nonlinear. As is well known, a consequence of this is that for non-convex confining/interaction potentials the mean field dynamics might have more than one stationary state, in contrast to the particle dynamics. Indeed, the unique steady state of the $N$-particle Fokker–Planck equation (2.3) is given by the Gibbs measure

$$M_N := \frac{e^{-\beta H_N(x)}}{Z_N}\,\mathrm{d}x \qquad \text{with} \qquad Z_N := \int_{\Omega^N} e^{-\beta H_N(y)}\,\mathrm{d}y.$$

On the other hand, the mean field limit can admit more than a single steady state, with the full characterisation being the set of solutions to the self-consistency equation

$$\beta^{-1}\log\rho_* + W\star\rho_* + V = C_* \qquad \text{on } \Omega \text{ for some } C_*\in\mathbb{R}, \tag{2.6}$$

This is discussed in detail in Proposition 4.1. See also [11, 18] and the references therein.

2.4. *Phase transitions.* The uniqueness/non-uniqueness of steady states of the mean field system (2.5) depends both the temperature of the system (or, equivalently, the strength of the interaction) and on the convexity properties of the confining and interaction potentials $V$ and $W$. At sufficiently high temperatures, the diffusion is strong enough that the expected escape time of particles from local minima of the potentials is bounded uniformly in the number of particles. Indeed, a perturbation argument shows that, at sufficiently high temperatures, the self-consistency equation (2.6) has a unique solution, see Proposition 4.2. When we cool the system and the potentials are non-convex, particles can get trapped for arbitrarily long time scales in local minima and condense [4]. In statistical physics terms, the system changes from a gaseous state to a liquid or solid state. In the mean field limit, this change of behavior can be characterised by the local or global instability of the minimisers of the mean field energy (see (2.10)), see for instance [11, 15] and Fig. 1 for a loose picture of the change in the mean field energy landscape as the temperature of the system is varied. For more details on the possible definitions of a phase transition, see Section 4.

2.5. *Qualitative long time behavior.* Due to the linearity of the $N$-particle system, we know that (2.3) admits a unique steady state (given by the Gibbs measure $M_N$). If the potentials prevent mass from escaping to infinity, by the La-Salle's principle for graident flows [12, Theorem 2.13], we know that independent of the initial condition

$$\lim_{t\to\infty} \rho^N(t) = M_N\,, \tag{2.7}$$

in the sense of weak convergence of probability measures. Furthermore, entropy methods and the LSI provide us with exponentially fast convergence to the steady state [45]. On the other hand, if the McKean–Vlasov equation admits multiple steady states, then the limit $t\to\infty$ for the mean field dynamics will depend on the choice of initial condition. In this case, the limiting dynamics (2.5) do not approximate the particle dynamics (2.1) for arbitrarily long times. In this paper we provide evidence that phase transitions constitute the natural obstruction to obtaining uniform-in-time propagation of chaos estimates. See Theorem 3.7 for sufficient conditions for the mean field approximation to be valid uniformly in time.

2.6. $\Gamma$*-convergence.* Taking advantage of the 2-Wasserstein gradient flow structure of (2.3) and (2.5) (see [1, 10, 53]), the main tool that we will use to obtain a quantitative understanding of the limit $N\to\infty$ is $\Gamma$-convergence with respect to the topology introduced by de Finetti/Hewitt-Savage-type convergence (2.4). To illustrate this technique, we use it to characterise the limit $N\to\infty$ of the Gibbs measure $M_N$. Following the pioneering work of Messer and Spohn [47], we notice that $M_N$ is the unique minimiser over probability measures of the energy per unit particle

$$E^N[\rho^N] = \frac{1}{N}\left(\beta^{-1}\int_{\Omega^N}\rho^N\log\rho^N\,\mathrm{d}x + \int_{\Omega^N} H_N\rho^N\,\mathrm{d}x\right). \tag{2.8}$$

Taking $N\to\infty$, we can characterise the thermodynamic limit of the energy $E^N$ as $E^N \to^\Gamma E^\infty : \mathcal{P}(\mathcal{P}(\Omega))\to\mathbb{R}\cup\{+\infty\}$, where

$$E^\infty[P] := \int_{\mathcal{P}(\Omega)} E^{MF}[\rho]\,\mathrm{d}P(\rho), \tag{2.9}$$

with the mean field energy $E^{MF}:\mathcal{P}(\Omega)\to\mathbb{R}\cup\{+\infty\}$ given by

$$E^{MF}[\rho] := \beta^{-1}\int_\Omega \rho\log(\rho)\,\mathrm{d}x + \frac{1}{2}\int_{\Omega^2} W(x,y)\,\mathrm{d}\rho(x)\,\mathrm{d}\rho(y) + \int_\Omega V(x)\,\mathrm{d}\rho(x)\,, \tag{2.10}$$

see [35] or Theorem 5.6 for a more modern proof. Using the fact that $M_N$ is the minimiser of (2.8), we know that any accumulation point of the sequence $M_N$ (in the sense of de Finetti/Hewitt–Savage) $P_\infty\in\mathcal{P}(\mathcal{P}(\Omega))$ needs to be a minimiser of (2.9). Hence, we may conclude that the minimal energy converges

$$\lim_{N\to\infty}\left(-\frac{1}{N}\log Z_N\right) = \lim_N E^N[M_N] = E^\infty[P_\infty] = \inf_{P\in\mathcal{P}(\mathcal{P}(\mathbb{R}^d))} E^\infty[P] = \inf_{\rho\in\mathcal{P}(\mathbb{R}^d)} E^{MF}[\rho],$$

where we have used that (2.9) is a potential energy, which also implies that $P_\infty$ needs to be supported on the minimisers of $E^{MF}$. This convergence and related results are discussed in further detail in Section 5.

Under our previous hypothesis, we have the following standard result.

THEOREM A. Under Assumptions 2.1 and 2.2, $E^{MF}$ is bounded below and has at least one minimiser.

The lower bound follows from Jensen's inequality, while the existence of a minimiser follows from the direct method of calculus of variations. Under the extra assumption that the $E^{MF}$ admits a unique minimiser $\rho_\beta \in \mathcal{P}(\mathbb{R}^d)$, we have

$$M_N \rightharpoonup \delta_{\rho_\beta} \in \mathcal{P}(\mathcal{P}(\Omega)).$$

The perspective of the proofs in this paper is that the evolution of the $N$-particle law (2.3) and the mean field limit (2.5) are respectively the gradient flows of $E^N$ and $E^{MF}$ with respect to the scaled 2-Wasserstein distance $\bar{d}_2$ and the 2-Wasserstein distance on $\mathcal{P}(\Omega)$. In fact, the $\lambda$-convexity assumption on the potentials and the doubling condition (2.2) can be used to obtain uniqueness of the gradient flow solutions. The next fundamental result is essentially a restatement of [1, Theorem 11.2.8], for the compact case see [53].

THEOREM B. If Assumptions 2.1 and 2.2 hold, then for any $\rho_{\text{in}} \in \mathcal{P}(\Omega)$ there exists unique distributional solutions $\rho^N \in C([0,\infty);\mathcal{P}_{\text{sym}}(\Omega^N)$ and $\rho \in C([0,\infty);\mathcal{P}(\Omega)$ to (2.3) and (2.5), respectively, which are the gradient flows of $E^N$ (2.8) and $E^{MF}$ (2.10) with respect to the scaled 2-Wasserstein distance $\bar{d}_2$ and the 2-Wasserstein distance $d_2$ on $\mathcal{P}(\Omega)$, respectively.

We remark here that in [10] an alternative proof of propagation of chaos is provided which employs the $\Gamma$-convergence result and the convergence of the gradient flow structures.

**3. Main results.** To quantify the convergence as $t \to \infty$ in (2.7), we can apply the standard relative entropy estimate [60]. More specifically, we consider the Lyapunov functional given by the scaled relative entropy of $\rho^N(t)$ with respect to the equilibrium measure $M_N$:

$$E^N[\rho^N(t)] - E^N[M_N] = \overline{\mathcal{E}}(\rho^N(t)|M_N) := \frac{1}{N}\int_{\Omega^N} \log\left(\frac{\rho^N(t)}{M_N}\right)\rho^N(t)\ \mathrm{d}x,$$

where we use the notation $\overline{\mathcal{E}} : \mathcal{P}(\Omega)\times\mathcal{P}(\Omega) \to [0,\infty]$ to denote the scaled relative entropy. Taking a time derivative and using the PDE (2.3) we obtain the scaled relative Fisher information $\overline{\mathcal{I}} : \mathcal{P}(\Omega)\times\mathcal{P}(\Omega) \to [0,\infty]$:

(3.1)
$$\frac{\mathrm{d}}{\mathrm{d}t}\overline{\mathcal{E}}(\rho^N(t)|M_N) = -\beta^{-1}\frac{1}{N}\int_{\Omega^N}\left|\nabla\log\left(\frac{\rho^N(t)}{M_N}\right)\right|^2\rho^N(t)\ \mathrm{d}x =: -\beta^{-1}\overline{\mathcal{I}}(\rho^N(t)|M_N).$$

The convergence (2.7) in relative entropy is exponential whenever we can show that the $N$-particle log Sobolev constant is bounded away from zero:

$$0 < \lambda_{\text{LS}}^N := \inf_{\rho^N\in\mathcal{P}(\Omega^N)\setminus\{M_N\}} \frac{\beta^{-1}\overline{\mathcal{I}}(\rho^N|M_N)}{\overline{\mathcal{E}}(\rho^N|M_N)}. \tag{3.2}$$

Following the classical work of Bakry–Emery [3, 40, 36], we can find mild conditions for the positivity of the log Sobolev constant whenever the domain $\Omega$ is $\mathbb{R}^d$.

THEOREM C. Under Assumptions 2.1 and 2.2, if there exists $R > 1$ and $\lambda > 0$ such that

$$D^2H^N(x) \geq \lambda I^{Nd\times Nd} \qquad \text{for every } |x| > R, \tag{3.3}$$

then we have that $\lambda_{\text{LS}}^N > 0$.

REMARK 3.1. The convexity condition (3.3) in the far field can arise from either the convexity of the interaction or the confining potential. We expect that the sharp condition for the Gibbs measure $M_N$ to satisfy $\lambda_{\text{LS}}^N > 0$ uniformly in $N$ is related to the behavior of the mean field limit dissipation inequality (3.5). We will discuss this in more detail in Conjecture 1.

For the mean field limit, we can perform a similar analysis with the relative mean field energy. More specifically, given $\rho(t)$ the solution to (2.5) we can differentiate to obtain the dissipation

(3.4)
$$\frac{\mathrm{d}}{\mathrm{d}t}E^{MF}[\rho(t)] - \inf E^{MF} = -\int_\Omega |\beta^{-1}\nabla \log \rho(t) + \nabla W \star \rho(t) + \nabla V|^2 \rho(t)\ \mathrm{d}x =: -D(\rho(t)).$$

Hence, we obtain exponential decay of the mean field energy to its minimum value, as long as the so-called infinite colume log Sobolev constant, given by

$$0 < \lambda_{\mathrm{LS}}^{\infty} := \inf_{\substack{\rho\in\mathcal{P}(\Omega)\\ \rho\notin\mathcal{K}}} \frac{D(\rho)}{E^{MF}[\rho(t)] - \inf E^{MF}}, \tag{3.5}$$

is positive, where

$$\mathcal{K} = \{\rho \in \mathcal{P}(\Omega)\ :\ E^{MF}[\rho] = \inf E^{MF}\}.$$

In both cases, when the log Sobolev constant is positive we can show that the relative energy behaves quadratically with respect to the 2-Wasserstein distance. This is essentially the content of Talagrand's inequality [58]. One of our contributions in this paper is a new proof of a generalised version of this inequality using gradient flow techniques.

THEOREM 3.2. *Under Assumptions 2.1 and 2.2,*

(3.6)
$$E^N[\rho^N] - E^N[M_N] \geq \frac{\lambda_{\mathrm{LS}}^N}{2}\overline{d}_2^2(\rho^N, M_N) \qquad \text{and} \qquad E^{MF}[\rho] - \inf E^{MF} \geq \frac{\lambda_{\mathrm{LS}}^\infty}{2} d_2^2(\rho, \mathcal{K}),$$

*where $\mathcal{K}$ is the set of minimisers of $E^{MF}$ and $d_2^2(\rho,\mathcal{K}) = \inf_{\mu\in\mathcal{K}} d_2^2(\rho,\mu)$.*

An optimal transport-based proof of inequality (3.6) for $E^N$ and $E^{MF}$ can be found in [61, Theorem 22.17]. We also refer to [16] for a proof using entropic interpolation. In Section 6, we provide a different more intuitive proof of (3.6) for general energies $E : \mathcal{P}(\Omega) \to \mathbb{R}\cup\{+\infty\}$. Our strategy is to use the associated gradient flow structure and what is sometimes referred to as Otto calculus [48], a formal Riemannian calculus on $(\mathcal{P}(\Omega), d_2)$. We should note that one of the main differences in Talagrand's inequality for the $N$-particle energy and for the mean field energy is that the set of minimisers $\mathcal{K}$ does not need to be a single point in the mean field case.

Having established the need for understanding the behavior of the log Sobolev constant, our first result relates the limit of the particle system log Sobolev constant (3.2) with the mean field or infinite volume log Sobolev constant(3.5):

THEOREM 3.3. *Under Assumptions 2.1 and 2.2, we have*

$$\limsup_{N\to\infty} \lambda_{\mathrm{LS}}^N \leq \lambda_{\mathrm{LS}}^\infty.$$

*Moreover, if the mean field energy $E^{MF}$ (2.10) admits a critical point that is not a minimiser, then $\lambda_{\mathrm{LS}}^\infty = 0$, and there exists $C > 0$ such that*

$$\lambda_{\mathrm{LS}}^N \leq \frac{C}{N}.$$

Our result complements similar results that have been obtained for unbounded spin systems [63] and the references therein. On the other hand, when $\lambda_{\mathrm{LS}}^\infty > 0$, we can show that the *regularized log Sobolev constant*

$$\lambda_{\mathrm{LS}}^{N,\varepsilon} := \inf_{\rho^N :\, \overline{\mathcal{E}}(\rho^N|M_N)>\varepsilon} \frac{\beta^{-1}\overline{\mathcal{I}}(\rho^N|M_N)}{\overline{\mathcal{E}}(\rho^N|M_N)}$$

does not degenerate. More specifically,

$$\lim_{N\to\infty} \lambda_{\mathrm{LS}}^{N,\varepsilon} \geq \lambda_{\mathrm{LS}}^{\infty} > 0.$$

This result implies that relaxation to neighborhoods of the stationary state of the particle dynamics (2.3) happens exponentially fast, uniformly in $N$.

THEOREM 3.4. *Under Assumptions 2.1 and 2.2, assume that* $\lambda_{\mathrm{LS}}^{\infty} > 0$, *and that* $\rho_{\mathrm{in}}$ *in* (2.1) *has finite energy and bounded higher order moments,*

$$E^{MF}[\rho_{\mathrm{in}}] < \infty \qquad \textit{and} \qquad \int_{\Omega} |x|^{2+\delta} \, \mathrm{d}\rho_{\mathrm{in}} < \infty, \qquad \textit{for some } \delta > 0. \tag{3.7}$$

*Then, for every* $\varepsilon > 0$, *there exists* $N_0 \in \mathbb{N}$, *such that for every* $N > N_0$ *we have*

$$\overline{\mathcal{E}}(\rho^N(t)|M_N) \leq \max\left\{\varepsilon, e^{-\frac{1}{2}\lambda_{\mathrm{LS}}^{\infty} t}\, \overline{\mathcal{E}}(\rho_{\mathrm{in}}^{\otimes N}|M_N)\right\}.$$

REMARK 3.5. For chaotic measures, we can take the limit $N \to \infty$ of the relative entropy obtaining

$$\lim_{N\to\infty} \overline{\mathcal{E}}(\rho_{\mathrm{in}}^{\otimes N}|M_N) = E^{MF}[\rho_{\mathrm{in}}] - \inf E^{MF}\,.$$

This is discussed in Theorem 5.6.

Unfortunately, we are not able to fully characterise the limit of $\lambda_{\mathrm{LS}}^{N}$ in terms of the mean field limit. Despite this, Theorem 3.3 and Theorem 3.4, provide us with evidence which is convincing enough to make the following conjecture.

CONJECTURE 1. Under Assumptions 2.1 and 2.2, we have the equality

$$\lim_{N\to\infty} \lambda_{\mathrm{LS}}^{N} = \lambda_{\mathrm{LS}}^{\infty}.$$

The results of our paper provide us with a strong indication that the absence of phase transitions (loosely defined to mean that the mean field limit has a unique stationary state), the non-degeneracy of the infinite volume log Sobolev constant, and the validity uniform-in-time propagation of chaos are all equivalent.

3.1. *Consequences of the non-degeneracy of the log Sobolev inequality.* Bearing Conjecture 1 in mind, we now explore the implications of the non-degeneracy of the LSI constant in the limit $N \to +\infty$. We begin by noticing that if the log Sobolev constant does not degenerate in $N$, then the invariant Gibbs measure $M_N$ of the $N$-particle system is well approximated by the unique minimiser of the mean field energy.

THEOREM 3.6. *Under Assumptions 2.1 and 2.2, assume that* $\limsup_{N\to\infty} \lambda_{\mathrm{LS}}^{N} > 0$. *Then, there exists a unique steady state* $\rho_\beta$ *to* (2.5). *Moreover, there exists* $C > 0$, *such that*

$$\overline{d}_2^2(\rho_\beta^{\otimes N}, M_N) \leq \frac{2}{\lambda_{\mathrm{LS}}^{N}} \overline{\mathcal{E}}(\rho_\beta^{\otimes N}|M_N) \leq \frac{2}{(\lambda_{\mathrm{LS}}^{N})^2} \overline{\mathcal{I}}(\rho_\beta^{\otimes N}|M_N) \leq \frac{C}{N}.$$

Interpolating the previous result with more standard propagation of chaos estimates that depend on the convexity constant of the potentials, we obtain the following uniform-in-time propagation of chaos result.

THEOREM 3.7. *Under Assumptions 2.1 and 2.2, let $\rho^N$ and $\rho$ denote the unique solutions to the particle (2.3) and mean field (2.5) dynamics given by Theorem B. Assume that $\rho_{\text{in}}$ has finite energy*

$$E^{MF}[\rho_{\text{in}}] < \infty\,,$$

*that the gradient of the square of the interaction potential is uniformly integrable*

$$\sup_{t\in[0,\infty]} \int_\Omega |\nabla_1 W|^2 \star \rho(t)\rho(t)\ \mathrm{d}x < \infty\,, \tag{3.8}$$

*and that $\liminf_{N\to\infty} \lambda^N_{\text{LS}} =: \lambda^\infty > 0$. Then,*

$$\overline{d}_2(\rho^N(t), \rho^{\otimes N}(t)) \leq \frac{C}{N^\theta} \qquad \text{for all } t > 0,$$

*where*

$$\theta = \begin{cases} 1/2 & \text{if } K_V + K_W(1-1/N) > 0 \\ \frac{1}{2}\frac{\lambda^\infty}{\lambda^\infty - 2(K_V + K_W(1-1/N))} & \text{if } K_V + K_W(1-1/N) < 0 \end{cases}$$

*with $K_V$ and $K_W$ the convexity constants of $V$ and $W$ in Assumptions 2.1 and 2.2. In the case where $K_V + K_W(1-1/N) = 0$, we can pick any $\theta < 1/2$.*

REMARK 3.8. The integrability assumption (3.8) is trivially true when $W$ is uniformly Lipschitz. Also, this assumption is satisfied when the potentials are attractive enough in the far field, i.e. outside a ball of radius $R$, that we can obtain uniform exponential bounds for the tail behavior of the mean field solution.

REMARK 3.9. We note that for convex potentials $K_V$, $K_W \geq 0$ the uniform propagation of chaos with $\theta = 1/2$ has already been shown in [44]. The main difference with this work is that our approach utilizes the convexity of the entropy along the 2-Wasserstein distance to obtain a contraction estimate. This approach can be easily extended to manifolds with Ricci curvature bounded from below [43], where we need to consider the sign of

$$K_{\text{Ric}} + K_V + K_W(1-1/N)$$

with $K_{\text{Ric}}$ the lower bound on the Ricci curvature of the underlying Riemannian manifold.

REMARK 3.10. We do not expect $\theta$ in the above theorem to be sharp for $K_V + K_W(1-1/N) < 0$. At sufficiently high temperatures and $K_V + K_W(1-1/N) < 0$, a comparable result for the 1-Wasserstein distance is shown by coupling methods in [25] with $\theta = 1/2$.

We can also use the non-degeneracy of the LSI constant to identify the fluctuations at equilibrium.

THEOREM 3.11. *Let $\Omega = \mathbb{T}^d$ and assume that $\liminf_{N\to\infty} \lambda^N_{\text{LS}} > 0$, then the fluctuations process*

$$\eta^N(t) = \sqrt{N}\left(\frac{1}{N}\sum_{i=1}^N \delta_{X_t^i} - \rho_\beta\right),$$

*where $(X_t^1, ..., X_t^N)$ are solutions to (2.1) with initial law given by the invariant Gibbs measure $M_N$, satisfies*

$$\sup_{N\in\mathbb{N}, t\in[0,T]} \mathbb{E}\left[\|\eta^N(t)\|^2_{H^{-s}(\mathbb{T}^d)}\right] < \infty,$$

*for any $T > 0$, $s > d/2 + 1$.*

*Moreover, assume that $V$, $W$ are smooth and that the linearised operator (2.5) around $\rho_\beta$*

$$\mathcal{L}_{\rho_\beta}\eta = \beta^{-1}\Delta\eta + \nabla\cdot(\rho_\beta\nabla W \star \eta) + \nabla\cdot(\eta\nabla W\star\rho_\beta) + \nabla\cdot(\nabla V\eta)\,,$$

*satisfies, for all $\phi \in C^\infty(\mathbb{T}^d)$, the following coercivity inequality*

$$\langle -\mathcal{L}_{\rho_\beta}\phi, \phi\rangle_{L^2(\mathbb{T}^d)} \geq c\|\nabla\phi\|^2_{L^2(\mathbb{T}^d)} \tag{3.9}$$

*for some $c > 0$.*

*Then, for any $m > d/2+3$, $\eta^N$ converges in law, as a $C([0,T];H^{-m}(\mathbb{T}^d))$-valued random variable, to the unique stationary solution $\eta^\infty$ of the following linear SPDE*

$$\partial_t\eta^\infty = \mathcal{L}_{\rho_\beta}\eta^\infty + \nabla\cdot(\sqrt{\rho_\beta}\xi) \tag{3.10}$$

*where $\xi$ is space-time white noise.*

REMARK 3.12. We mention the result of Fernandez and Meleard [26] (see also [59, 55]) which characterises the fluctuations of the particle dynamics with respect to the mean field limit for finite time horizons, under a stronger closeness assumption for the initial data. We also mention the recent preprint [62] which studies fluctuations for singular potentials.

To our knowledge, the first available results for fluctuations at equilibrium is due to Dawson [18] in which he shows that, for the specific case of Desai–Zwanzig model, at the phase transition temperature, equilibrium fluctuations are non-Gaussian. It is an interesting open problem if this behavior is universal for any system which undergoes a (continuous) phase transition. The different notions of phase transition will be discussed in Section 4.

REMARK 3.13. For The smoothness assumptions on $V$ and $W$ are used to have a well-defined semigroup associated to the linearised operator $\mathcal{L}_{\rho_\beta}$ which regularizes instantaneously arbitrary initial data in $H^{-m}(\mathbb{T}^d)$. Up to technical results, this can be quantified by requiring $V$ and $W \in W^{m+\epsilon,\infty}(\mathbb{T}^d)$ for any $\epsilon > 0$. For fluctuation results with singular potentials we refer the reader to the recent preprint [62].

The unique invariant measure of the SPDE (3.10) $\mathcal{G} \in \mathcal{P}(H^{-m}(\mathbb{T}^d))$ is a centred Gaussian measure with covariance operator $Q_\mathcal{G}$ given by

$$Q_\mathcal{G}(\varphi,\psi) := \lim_{t\to\infty}\int_0^t\int_{\mathbb{T}^d}\nabla e^{t\mathcal{L}^*_{\rho_\beta}}\varphi\cdot\nabla e^{t\mathcal{L}^*_{\rho_\beta}}\psi\,\rho_\beta\,\mathrm{d}x\,\mathrm{d}t\,,$$

for any mean-zero $\varphi$, $\psi \in C^\infty(\mathbb{T}^d)$ and where $\mathcal{L}^*_{\rho_\beta}$ denotes the flat $L^2$-adjoint.

In the specific case that $V \equiv 0$ and $W(x,y) = W(x-y)$, we can obtain a more explicit characterisation of $\mathcal{G}$. Indeed, since $\liminf_{N\to\infty}\lambda^N_{\mathrm{LS}} > 0$, we know from Proposition 4.2 (and the discussion following it) that $E^{MF}$ has a unique critical point which is given by $\rho_\infty(\mathrm{d}x) = \mathrm{d}x$. We can then write down an explicit representation of the action of the semigroup $e^{t\mathcal{L}^*_{\mathrm{d}x}}$ in Fourier space as follows

$$\hat{\varphi}_t(k) = e^{-4\pi^2|k|^2(\beta^{-1}+\hat{W}(k))t}\hat{\varphi}(k) \quad k\in\mathbb{Z}^d, k\neq 0\,,$$

where $\varphi_t = e^{t\mathcal{L}^*_{\mathrm{d}x}}\varphi$ for some mean-zero $\varphi \in C^\infty(\mathbb{T}^d)$. This leaves us with the formula

$$Q_\mathcal{G}(\varphi,\psi) = \sum_{k\in\mathbb{Z}^d,k\neq 0}\frac{\hat{\varphi}(k)\hat{\psi}(k)}{8\pi^2(\beta^{-1}+\hat{W}(k))}\,,$$

where we have used the fact that the coercivity inequality (3.9) is equivalent to the fact that $\beta^{-1} + \hat{W}(k) > 0$ for all $k \in \mathbb{Z}^d, k \neq 0$, which is also equivalent to the condition $\beta < \beta_\sharp$, see Proposition 4.3.

Another way of rewriting this is that $\mathcal{G}$ is the unique centred Gaussian measure with Cameron–Martin space given by the closure of all smooth mean-zero functions $\varphi$ under the norm

$$\|\varphi\|_{\mathcal{H}_\mathcal{G}}^2 = 8\pi^2\left(\beta^{-1}\int_{\mathbb{T}^d}\varphi^2\,\mathrm{d}x + \int_{\mathbb{T}^d}(W*\varphi)\varphi\,\mathrm{d}x\right).$$

Since, $\beta^{-1} + \hat{W}(k) > 0$ for all $k \in \mathbb{Z}^d, k \neq 0$, the above norm is equivalent to the standard $L^2(\mathbb{T}^d)$ norm. The above norm is also the same, up to a multiplicative constant, as the norm introduced in Property A.

One can further use the structure of the covariance operator $Q_\mathcal{G}$ to read off that $\mathcal{G}$ is supported on $H^{-\frac{d}{2}-}(\mathbb{T}^d)$ distributions. Thus, the limiting equilibrium fluctuations have the regularity of spatial white noise, which is not surprising considering the fact that their Cameron–Martin space is "*basically*" $L^2(\mathbb{T}^d)$.

3.2. *Non-degeneracy of the LSI constant in specific cases.* Putting aside for the time being the validity of Conjecture 1, we show that the LSI constant $\lambda_{\mathrm{LS}}^N$ does not degenerate in the high temperature regime when $\Omega$ is compact, or when the confinement $V$ satisfies an LSI inequality and the interaction strength is small enough.

THEOREM 3.14. *Assume that there exists a constant $C > 0$ such that*

$$\|W\|_{L^\infty(\Omega^2)},\ \|D_{xy}^2 W\|_{L^\infty(\Omega^2)} < C\,.$$

*We then have the following two scenarios:*

(a) *Compact case: Assume $\Omega$ is compact and its normalised Lebesgue measure $\mathrm{d}x$ satisfies a log Sobolev inequality. Then, there exists a $0 < \beta_{\mathrm{LS}} = \beta_{\mathrm{LS}}(C)$ such that for all $\beta < \beta_{\mathrm{LS}}$, we have*

$$\liminf_{N\to\infty} \lambda_{\mathrm{LS}}^N > 0\,.$$

(b) *weak interaction case: Assume $\Omega = \mathbb{R}^d$ and that the one-particle measure $Z_V^{-1}e^{-V}\,\mathrm{d}x$ satisfies a log Sobolev inequality with constant $\lambda_{\mathrm{LS}}^V > 0$. Then there exists an $\epsilon_{\mathrm{LS}} = \epsilon_{\mathrm{LS}}(C, \lambda_{\mathrm{LS}}^V, \beta) > 0$, such that for any $0 \leq \epsilon < \epsilon_{\mathrm{LS}}$ we have*

$$\liminf_{N\to\infty} \lambda_{\mathrm{LS}}^{\epsilon,N} > 0\,,$$

*where $\lambda_{\mathrm{LS}}^{\epsilon,N}$ is the log Sobolev constant of the Gibbs measure $\tilde{M}_N = Z_N^{-1}e^{-\beta H_N^\epsilon}\,\mathrm{d}x$ with*

$$H_N^\epsilon(x) = \sum_{i=1}^N V(x_i) + \frac{\epsilon}{2N}\sum_{j=1}^N\sum_{i=1}^N W(x_i, x_j)\,.$$

The above result relies crucially on the two-scale approach to log Sobolev inequalities introduced in [49], which is based on the analysis of the marginal and conditional measures of $M_N$. For the convenience of the reader, we describe the main result of [49] in Theorem 11.2. In addition to the above high temperature result, it is also possible to obtain a sharper result in certain specific scenarios. For instance, in the case in which $\Omega = \mathbb{T}$, $W(x, y) = -\cos(2\pi(x - y))$, and $V \equiv 0$. The corresponding system is referred to by many names: the noisy Kuramoto model, the mean field classical $XY$ model, the mean field $O(2)$ model or the Brownian mean field model [14]. It is known that this system exhibits a phase

transition (of type A, B, and C, see Definition 4.6). Due to the particularly simple nature of the model, it is possible to show that the $N$ particle log Sobolev constant $\lambda_{\mathrm{LS}}^N$ is asymptotically non-degenerate all the way up to the critical inverse temperature $\beta_c = 2$. We state without proof the following result due to Bauerschmidt and Bodineau [5], see also [6].

THEOREM D ([5, Theorem 1]). Consider the Gibbs measure $M_N$ of the mean field $O(2)$ model and denote by $\lambda_{\mathrm{LS}}^N$ its log Sobolev constant. Then, for all $0 < \beta < \beta_c = 2$, we have that

$$\liminf_{N\to\infty} \lambda_{\mathrm{LS}}^N > 0\,.$$

REMARK 3.15. An essentially similar argument as in [5, Theorem 1], can be used to show that the system with $\Omega = \mathbb{T}$, $W(x,y) = -\cos(2\pi k(x-y))$, and $V \equiv 0$ for some $k \in \mathbb{N}$ has a uniform LSI all the way up to the critical inverse temperature $\beta_c$ which coincides with $\beta_\sharp$ from Proposition 4.3.

**4. Phase transitions.** We start our discussion by stating and proving the following result which provides us with a particularly useful characterisation of steady states.

PROPOSITION 4.1. *Under Assumptions 2.1 and 2.2 the following statements are equivalent:*

1. *$\rho \in \mathcal{P}(\Omega)$ is a critical point of the mean field free energy $E^{MF}$, that is to say*
$$|\partial E^{MF}|(\rho) = D(\rho) = 0.$$
2. *$\rho \in \mathcal{P}(\Omega)$ is a steady state of the mean field equation (2.5), i.e. it is distributional weak solution of the PDE*
$$\beta^{-1}\Delta\rho + \nabla\cdot(\rho(\nabla W \star \rho + \nabla V)) = 0 \quad x \in \Omega\,.$$
3. *$\rho$ solves the self-consistency equation:*
$$\rho - \frac{1}{Z_\beta} e^{-\beta(W\star\rho+V)} = 0\,, \qquad Z_\beta = \int_\Omega e^{-\beta(W\star\rho+V)}\,\mathrm{d}x\,. \tag{4.1}$$

*Furthermore, for all $\beta > 0$, $E^{MF}$ has at least one global minimiser which is a critical point, and any critical point $\rho \in \mathcal{P}(\Omega)$ of $E^{MF}$ is Lipschitz, strictly positive, and has moments of all orders.*

PROOF. The proof of the equivalence of the three characterisations follows from similar arguments to [11, Proposition 2.4] and so we omit the proof. The fact that $E^{MF}$ has a critical point follows from Theorem A.

Now, using (4.1), we know that any critical point $\rho \in \mathcal{P}(\Omega)$ is of the form

$$\frac{1}{Z_\beta} e^{-\beta(W\star\rho+V)}\,,$$

which is Lipschitz and strictly positive. We now use Assumption 2.2 to assert that, since $W$ is bounded below,

$$\rho \le Z_\beta^{-1} e^{\beta C} e^{-\beta V}\,.$$

Thus, by Assumption 2.1, any critical point $\rho \in \mathcal{P}(\Omega)$ of $E^{MF}$ has moments of all orders. □

Furthermore, we have the following result regarding uniqueness of critical points.

PROPOSITION 4.2. *Under the assumptions of Theorem 3.14 there exists a unique critical point $\rho_\beta \in \mathcal{P}(\Omega)$ of the mean field free energy $E^{MF}$.*

PROOF. The proof of this result follows by combining the results of Theorems 3.6 and 3.14. We know from Theorem 3.14 that the logarithmic Sobolev inequality holds uniformly for $M_N$. We can then use Theorem 3.6 to argue that for $\beta$ sufficiently small, the mean field free energy $E^{MF}$ must have a unique critical point $\rho_\beta$. □

When we discuss non-uniqueness of critical points, we will often restrict ourselves to the case in which $V = 0, W(x, y) = W(x - y)$, and $\Omega = \mathbb{T}^d$. This case lends itself particularly well to analysis as the Lebesgue measure $\rho_\infty(\mathrm{d}x) = \mathrm{d}x$, what we shall hereafter refer to as the "*flat state*", is always a critical point of $E^{MF}$ for all $\beta > 0$. We shall see later that in this setting it is possible to provide relatively clean examples of the different types of phase transitions that we consider in this paper. For an example of what a typical phase transition looks like, see Fig. 1 for a schematic of the free energy landscape in the vicinity of a phase transition.

We now introduce and motivate a list of properties that may serve as a proxy for the absence of well-defined order parameter. We discuss how these properties relate to each other, specifically in the "*flat case*". We start by looking at the linearisation of the mean field dynamics (2.5):

PROPERTY A. Fix $\beta > 0$. We say that the system (2.5) satisfies *Property A* if $E^{MF}$ has a unique critical point $\rho_\beta$ and the linearised operator associated to right hand side of (2.5) has a spectral gap under the interaction weighted inner product. More specifically, we have that

$$\text{(A)} \qquad \inf_{\eta \in C_0^\infty(\Omega), \int_\Omega \eta \,\mathrm{d}x = 0} \frac{\langle -\mathcal{L}_{\rho_\beta}\eta, \eta\rangle_{W,\rho_\beta}}{\|\eta\|^2_{W,\rho_\beta}} > 0\,,$$

where

$$\mathcal{L}_{\rho_\beta}\eta = \beta^{-1}\Delta\eta + \nabla\cdot(\nabla V \eta) + \nabla\cdot(\rho_\beta \nabla W \star \eta) + \nabla\cdot(\eta \nabla W \star \rho_\beta)\,,$$

and

$$\langle \eta, \nu\rangle_{W,\rho_\beta} = \frac{\beta^{-1}}{2}\int_\Omega \eta\nu\rho_\beta^{-1}\,\mathrm{d}x + \frac{1}{2}\int_\Omega (W\star\eta)\nu\,\mathrm{d}x$$

with $\|\eta\|^2_{W,\rho_\beta} = \langle \eta, \eta\rangle_{W,\rho_\beta}$.

The above property captures essentially the local stability of the critical point $\rho_\beta$. Loss of the local stability of $\rho_\beta$ is a precursor to a phase transition. In the setting of a continuous phase transition (see the upper half of Fig. 1), one expects this property to fail exactly at the critical temperature $\beta = \beta_c$. The weighted inner product arises naturally through the linearisation of the mean field log Sobolev constant, or more specifically through the linearisation of the mean field energy $E^{MF}$, see (4.2). We note that the bilinear form $\langle\cdot,\cdot\rangle_{W,\rho_\beta}$ is positive semi-definite when $\rho_\beta$ is the unique critical point. For the classical $XY$ model (see the discussion before Theorem D), it is known that the system exhibits a continuous phase transition [11, Proposition 6.1]. In this case, the non-local inner product above it is related to the inner product that was introduced in [8] to understand the spectral gap ahead of the phase transition. In the absence of the interaction term, it reduces to the standard weighted inner product that symmetrises the Fokker-Planck operator, see [50]. In the periodic translation invariant case $\Omega = \mathbb{T}^d$, $W(x, y) = W(x - y)$, and $V \equiv 0$, when $\rho_\infty = \mathrm{d}x$ is the unique critical point of $E^{MF}$, the weighted inner product $\langle\cdot,\cdot\rangle_{W,\rho_\infty}$ is equivalent to the standard $L^2$ inner

product. Thus, in this situation, checking Property A is satisfied is equivalent to checking that $\mathcal{L}_{\rho_\infty}$ has a spectral gap in the standard $L^2$ inner product. This relationship will be made clearer in Proposition 4.3. Before discussing Property A in more detail, we introduce the next property which measures the local degeneracy of the self-consistency equation (4.1).

PROPERTY B. Fix $\beta > 0$. We denote by $T : L^2(\Omega) \to L^2(\Omega)$ the map associated to the self-consistency equation (4.1):

$$T(\rho) = \rho - \frac{1}{Z_\beta} e^{-\beta(W\star\rho+V)} = 0\,, \qquad Z_\beta = \int_\Omega e^{-\beta(W\star\rho+V)}\,\mathrm{d}x\,. \tag{B}$$

We say that the system (2.5) satisfies *Property B* if $E^{MF}$ has a unique critical point $\rho_\beta$ and the Fréchet derivative $D_\rho T$ of $T$ at $\rho_\beta$ is non-degenerate, that is to say it has a trivial kernel consisting only of constant functions.

Under additional conditions, see [11], the above property being violated implies the presence of a local bifurcation around the minimiser $\rho_\beta$. A simple condition that implies the presence of a local bifurcation is when the algebraic multiplicity of the 0 eigenvalue of $D_\rho T$ at $\rho_\beta$ is odd. Local bifurcations also arise, if the map $T$ can be rewritten as a so-called potential operator and its Frechét derivative $D_\rho T$ has a non-zero crossing number, see [39]. The above property exactly captures the second-order degeneracy of the mean field free energy around the unique critical point $\rho_\beta$. Indeed, formally expanding $E^{MF}[\rho]$ about $\rho_\beta$ we obtain

$$\begin{aligned} E^{MF}[\rho] =& E^{MF}[\rho_\beta] + \int_\Omega \beta^{-1}\frac{\eta^2}{2}\rho_\beta^{-1} + \frac{1}{2}W\star\eta\eta\,\mathrm{d}x + O(\eta^3) \\ =& E^{MF}[\rho_\beta] + \beta^{-1}\int_\Omega (D_\rho T[\rho_\beta]\eta)\eta\rho_\beta^{-1}\,\mathrm{d}x + O(\eta^3)\,, \end{aligned} \tag{4.2}$$

with $\eta = \rho - \rho_\beta$. Thus, if Property B is satisfied the mean field free energy is non-degenerate at second order near the critical point $\rho_\beta$.

The third and final property considers the validity of the Dissipation Inequality.

PROPERTY C. Fix $\beta > 0$. We say that the system (2.5) satisfies *Property C* if the infinite-volume log Sobolev constant is positive. More precisely, if we have that

$$\lambda_{\mathrm{LS}}^\infty := \inf_{\rho\in\mathcal{P}(\Omega),\rho\notin\mathcal{K}} \frac{D(\rho)}{E^{MF}[\rho] - \min_{\mathcal{P}(\Omega)} E^{MF}} > 0\,. \tag{C}$$

The third and final property captures global aspects of the free energy landscape. Indeed, one would expect it to be violated in both the situations described in Fig. 1. For the case of the discontinuous phase transition, the lower half of Fig. 1, Property C would be violated because of the presence of a non-minimising critical point of $E^{MF}$ which is represented by the red circles. Thus, the numerator of (C) vanishes while the denominator is strictly positive. As an explicit example of a system which exhibits such a phase transition, one can consider $\Omega = \mathbb{T}$, with the bi-chromatic interaction potential $W(x,y) = -\cos(2\pi(x-y)) - \cos(4\pi(x-y))$, and $V \equiv 0$ (cf. [11, Theorem 5.11]). On the other hand, in the case of a continuous phase transition, the upper half of Fig. 1, the fact that Property C should be violated is more subtle. To observe this, we linearise the right hand side of (C) about the unique minimiser $\rho_\beta$ at $\beta = \beta_c$. For the dissipation, we have that

$$D(\rho) = 2\int_\Omega \left|\beta^{-1}\nabla\frac{\eta}{\rho_\beta} + \nabla W\star\eta\right|^2 \mathrm{d}\rho_\beta + O(\eta^3)\,,$$

with $\eta = \rho - \rho_\beta$. Combining the above expression with (4.2), we obtain to leading order

$$\frac{D(\rho)}{E^{MF}[\rho] - E^{MF}[\rho_\beta]} \approx \frac{4\int_\Omega \left|\beta^{-1}\nabla\frac{\eta}{\rho_\beta} + \nabla W \star \eta\right|^2 \mathrm{d}\rho_\beta}{\int_\Omega \beta^{-1}\eta^2 + W \star \eta\eta\rho_\beta \,\mathrm{d}\rho_\beta^{-1}}\,. \tag{4.3}$$

Moreover, we notice that

$$(-\mathcal{L}_{\rho_\beta}\eta, \eta)_{W,\rho_\beta} = \frac{1}{2}\int_\Omega \left|\beta^{-1}\nabla\frac{\eta}{\rho_\beta} + \nabla W \star \eta\right|^2 \mathrm{d}\rho_\beta\,.$$

Using (4.3), it follows that to leading order

$$\frac{D(\rho)}{E^{MF}[\rho] - E^{MF}[\rho_\beta]} \approx 4\frac{\langle -\mathcal{L}_{\rho_\beta}\eta, \eta\rangle^2_{W,\rho_\beta}}{\|\eta\|^2_{W,\rho_\beta}}\,.$$

Hence, in the setting of a continuous phase transition the infinite volume log Sobolev constant captures the spectral gap of $\mathcal{L}_{\rho_\beta}$, and thus also captures the loss of local stability of $\rho_\beta$.

We now present the following result which characterises how the various properties we have discussed relate to each other in the periodic spatially homogeneous case.

PROPOSITION 4.3. *Assume $\Omega = \mathbb{T}^d$, $V \equiv 0$, $W(x,y) = W(x-y)$, and denote*

$$\beta_\sharp := \frac{1}{\min(0, -\min_{k\in\mathbb{Z}^d\setminus\{0\}} \hat{W}(k))} \in (0,\infty].$$

*Then, for $\beta < \beta_\sharp$ the linearised operator $\mathcal{L}_{\mathrm{d}x}$ satisfies the spectral gap property (10.3), and the kernel of the Fréchet derivative of the self-consistency equation (10.4) is non degenerate at the "*flat state*". For $\beta \geq \beta_\sharp$, Properties A and B are violated.*

*Furthermore, for all $\beta \neq \beta_\sharp$, Property C implies Property A and Property B.*

REMARK 4.4. We emphasize that Proposition 4.3 does not ensure that Properties A and B are satisfied for $\beta < \beta_\sharp$. In fact, for the bi-chromatic potential case $W(x,y) = -\cos(2\pi(x-y)) - \cos(4\pi(x-y))$, we know that the "*flat state*" is not the unique steady state for some $\beta < \beta_\sharp$ violating the uniqueness requirement of Properties A and B, see [11, Theorem 5.11].

PROOF. We will first argue that Properties A and B are equivalent. It follows from the fact that $\rho_\infty(\mathrm{d}x) \equiv \mathrm{d}x$ is always a critical point of $E^{MF}$ (in the flat case) that if $E^{MF}$ has a unique critical point, it must be $\rho_\infty$. For any mean-free $\eta \in C_0^\infty(\Omega)$, we have from the previous calculation, that

$$\frac{\langle -\mathcal{L}_{\rho_\infty}\eta, \eta\rangle_{W,\rho_\beta}}{\|\eta\|^2_{W,\rho_\beta}} = \frac{1}{2}\frac{\int_\Omega \left|\beta^{-1}\nabla\eta + \nabla W * \eta\right|^2 \mathrm{d}x}{\int_\Omega (\beta^{-1}\eta + W*\eta)\eta\,\mathrm{d}x}\,,$$

where we have used the fact that $\rho_\beta$ is the Lebesgue measure on $\mathbb{T}^d$. It is easy to check that the above expression is strictly positive if and only if $\beta < \frac{1}{\min(0, -\min_{k\in\mathbb{Z}^d\setminus\{0\}} \hat{W}(k))} = \beta_\sharp$. Similarly, computing the Fréchet derivative of $T$ at $\rho_\infty$, we obtain

$$D_\rho T[\rho_\infty]\eta = \eta + \beta W * \eta\,.$$

Again, the above linear operator has a trivial kernel if and only if $\beta < \beta_\sharp$. It follows that Properties A and B are equivalent.

We will now show that Property C implies Property B for $\beta \neq \beta_\sharp$. We know that Property B is satisfied if and only if $E^{MF}$ has unique critical point and $\beta < \beta_\sharp$. Assume it is violated. Then, either $E^{MF}$ has more than one critical point or $\beta \geq \beta_\sharp$ (or both).

Consider the case in which $\beta < \beta_\sharp$ but $E^{MF}$ has more than one critical point one of which is $\rho_\infty$. Furthermore, we know from [33, Lemma 5.6] that $\rho_\infty$ is a strict local minimum. If at least one of the other critical points is not a minimiser then clearly Property C is violated since the numerator can be chosen to be zero while the denominator remains positive. We will now argue that this is the case. Assume it is not, that is all other critical points are minimisers. Then, we can apply the mountain pass theorem in $\mathcal{P}(\Omega)$ (using the fact that $\rho_\infty$ is a strict local minimum) [33, Theorem 1.1] to construct a new non-minimising critical point thus obtaining a contradiction.

Consider now the case in which $\beta > \beta_\sharp$. In this situation, we note from [11, Proposition 5.3] that $\rho_\infty$ is a non-minimising critical point of $E^{MF}$. Thus, Property C is again violated. □

REMARK 4.5. We note here that one would expect the Properties A to C to be equivalent to each other. However, we unable to show that this holds true in general.

Given the above result, we are now finally in a position to present our definition of a phase transition in the "*flat case*".

DEFINITION 4.6 (Phase transition). Assume $V \equiv 0, W(x,y) = W(x-y)$, and $\Omega = \mathbb{T}^d$. Then, we say that the system (2.5) exhibits a phase transition of type A (resp. type B, type C) if there exists a $0 < \beta_c < \infty$ such that for all $0 < \beta < \beta_c$ Property A (resp. type B, type C) is satisfied, and for $\beta_c < \beta$ Property A (resp. type B, type C) is violated.

Except for the Brownian mean field model see Theorem D, we cannot ensure that the existence of a phase transition in the sense of Definition 4.6. Combining Theorem 3.14 with analysis in [15, 11], we can show the following result.

THEOREM 4.7. *Assume $V \equiv 0$, $W(x,y) = W(x-y)$, and $\Omega = \mathbb{T}^d$. If $W$ is $\mathbf{H}$-stable, that is to say $\hat{W}(k) \geq 0$ for all $k \in \mathbb{Z}^d \setminus \{0\}$, then the system (2.5) satisfies Properties A and B for all $\beta \in (0, \infty)$.*

*If there exists $k \in \mathbb{Z}^d \setminus \{0\}$ such that $\hat{W}(k) < 0$, then there exists $\beta_1 < \beta_\sharp$ such that Properties A to C are satisfied for all $\beta < \beta_1$, and Properties A to C are violated for all $\beta > \beta_\sharp$.*

PROOF. If $W$ is $\mathbf{H}$-stable, it follows from linear convexity $\rho_\infty = \mathrm{d}x$ is the unique critical point of $E^{MF}$ for all $0 < \beta < \infty$, hence Properties A and B cannot be violated, see [11, Proposition 5.8]. On the other hand, if $\beta_\sharp < \infty$, we know from Proposition 4.3 that Properties A to C are violated for all $\beta > \beta_\sharp$.

By Theorem 3.14 and Theorem 3.3, we know that there exists $\beta_{\mathrm{LS}}$ such that $0 < \beta < \beta_{\mathrm{LS}}$ then Property C is satisfied. Finally, by Proposition 4.3 we have that Properties A and B are also satisfied. □

**5. The limit $\boldsymbol{\mathcal{P}_{\mathrm{sym}}(\Omega^N) \to \mathcal{P}(\mathcal{P}(\Omega))}$.** In this section, we recall some useful results that allows us the characterise the various relevant objects in the limit as $N \to \infty$. We follow the approach taken by Hauray and Mischler in [35]. We start by defining the empirical measure.

DEFINITION 5.1. Given some $\rho^N \in \mathcal{P}_{\mathrm{sym}}(\Omega^N)$, the empirical measure is the $\mathcal{P}(\Omega)$-valued random variable which is given by

$$\mu^{(N)} := \frac{1}{N}\sum_{i=1}^{N} \delta_{x_i} \qquad \text{where } (x_1, ..., x_N) \text{ is distributed according } \rho^N.$$

We denote the law of the empirical measure $\mu^{(N)}$ on $\mathcal{P}(\Omega)$ by

$$\hat{\rho}^N \in \mathcal{P}(\mathcal{P}(\Omega))\,.$$

The topology we consider throughout most of the manuscript is the one induced by the scaled 2-Wasserstein distance.

DEFINITION 5.2. Given $\rho_1^N,\ \rho_2^N \in \mathcal{P}_{\mathrm{sym}}(\Omega^N)$, the scaled 2-Wassertein distance between them is given by

$$\overline{d}_2(\rho_1^N, \rho_2^N) = \inf_{\substack{X\sim\rho_1^N \\ Y\sim\rho_2^N}} \left(\frac{1}{N}\mathbb{E}[|X-Y|^2]\right)^{1/2},$$

where $X,\ Y$ are $\Omega^N$-valued random variables.

Similarly, given $P_1,\ P_2 \in \mathcal{P}(\mathcal{P}(\Omega))$, the 2-Wasserstein distance between them is given by

$$\mathfrak{D}_2(P_1, P_2) = \inf_{\substack{\mathcal{X}\sim P_1 \\ \mathcal{Y}\sim P_2}} \left(\mathbb{E}[d_2^2(\mathcal{X}, \mathcal{Y})]\right)^{1/2}$$

where $\mathcal{X},\ \mathcal{Y}$ are $\mathcal{P}(\Omega)$-valued random variables.

A fundamental result we need to understand the convergence $\mathcal{P}_{\mathrm{sym}}(\Omega^N) \to \mathcal{P}(\mathcal{P}(\Omega))$ is the fact that the mapping $\rho^N \mapsto \hat{\rho}^N$ is an isometry for the appropriately scaled 2-Wasserstein distance defined in Definition 5.2. Indeed, we have the following result.

THEOREM 5.3 ( [35, Proposition 2.14]). *For each $\mu^N,\ \nu^N \in \mathcal{P}_{\mathrm{sym}}(\Omega^N)$, we have*

$$\overline{d}_2(\mu^N, \nu^N) = \mathfrak{D}_2(\hat{\mu}^N, \hat{\nu}^N)\,.$$

PROOF. We present only a sketch of the proof of this result. The first inequality $\overline{d}_2(\mu^N, \nu^N) \geq \mathfrak{D}_2(\hat{\mu}^N, \hat{\nu}^N)$ follows from the mapping in Definition 5.1 $X \mapsto \mathcal{X}$, i.e. given $X = (X_1, ..., X_N)$ an $\Omega^N$-valued random variable with law $\mu^N$, we define

$$\mathcal{X} := \frac{1}{N}\sum_{i=1}^{N} \delta_{X_i}\,,$$

is a $\mathcal{P}(\Omega)$-valued random variable with law $\hat{\mu}^N$. The converse inequality follows from taking the inverse mapping and exploiting the symmetry of $\mu^N$ and $\nu^N$. □

Using this result, we can provide a metric notion of the de Finetti/Hewitt–Savage convergence.

DEFINITION 5.4. We say that the sequence $(\rho^N)_{N\in\mathbb{N}}$ such that $\rho^N \in \mathcal{P}_{\mathrm{sym}}(\Omega^N)$ converges in the 2-Wasserstein distance to $P_\infty \in \mathcal{P}(\mathcal{P}(\Omega))$ if

$$\lim_{N\to\infty} \mathfrak{D}_2(\hat{\rho}^N, P_\infty) = 0,$$

where $\hat{\rho}^N$ is the law of the empirical measure (as $\mathcal{P}(\Omega)$-valued random variable) associated to $\rho^N$.

We now have the following result which connects the above convergence to the one introduced in (2.4).

PROPOSITION 5.5. *For any sequence $(\rho^N)_{N\in\mathbb{N}}$ with $\rho^N \in \mathcal{P}_{\mathrm{sym}}(\Omega^N)$ and*

$$\sup_{N\in\mathbb{N}} \int_\Omega |x|^\gamma \, \mathrm{d}\rho_1^N(x) < \infty \tag{5.1}$$

*for some $\gamma > 2$, the metric convergence in Definition 5.4 is equivalent to the original marginal convergence of (2.4).*

PROOF. We first notice that by the results of Diaconis and Freedman [21], for a fixed $l \in \mathbb{N}$, the marginal $\rho_l^N$ coincides in the limit as $N \to \infty$ with the $l^{\text{th}}$ product of the empirical measure. More specifically, for any $\varphi \in C_c^\infty(\Omega^l)$ we have that

$$\left| \int_{\Omega^l} \varphi \, \mathrm{d}\rho_l^N - \int_{\mathcal{P}(\Omega)} \left( \int_{\Omega^l} \varphi \, \mathrm{d}\mu^{\otimes l} \right) \mathrm{d}\hat{\rho}^N(\mu) \right| \leq l^2 \frac{\|\varphi\|_{L^\infty(\Omega^l)}}{N}. \tag{5.2}$$

Furthermore, (5.1) implies that $\hat{\rho}^N$ is tight in $\mathcal{P}(\Omega)$. Indeed, we have

$$\int_{\mathcal{P}(\Omega)} d_2^{2\gamma}(\rho, \delta_0) \, \mathrm{d}\hat{\rho}^N(\rho) = \int_\Omega |x|^\gamma \, \mathrm{d}\rho_1^N.$$

Now if $\hat{\rho}^N$ converges in the sense of Definition 5.4 to $P_\infty$ we know that it must converge when tested against every $C_b(\mathcal{P}(\Omega))$. Since cylindrical test functions are a subset of $C_b(\Omega)$, it is clear from (5.2) that $\rho^N$ must also converge to $P_\infty$ in the sense of (2.4).

On the other hand if $\rho^N$ converges to $P_\infty$ in the sense of (2.4) we know that, under (5.1), the sequence $\hat{\rho}^N$ is relatively compact in $(\mathcal{P}(\mathcal{P}(\Omega)), \mathfrak{D}_2)$. By the Stone–Weierstrass theorem cylindrical functions are dense in $C_b(\mathcal{P}(\Omega))$ (see [52, Theorem 2.1 ]), which tells us that the limits in (5.2) and Definition 5.4 coincide. □

Next, we discuss the limit of the associated free energy and dissipation functionals.

THEOREM 5.6. *Under Assumptions 2.1 and 2.2, consider the sequence $(\rho^N)_{N\in\mathbb{N}}$ with $\rho^N \in \mathcal{P}_{\mathrm{sym}}(\Omega^N)$ such that $\rho^N \to P_\infty \in \mathcal{P}(\mathcal{P}(\Omega))$ in the sense of Definition 5.4. Then,*

$$\liminf_{N\to\infty} E^N[\rho^N] \geq E^\infty[P_\infty] \qquad \text{and} \qquad \liminf_{N\to\infty} \overline{\mathcal{I}}(\rho^N | M_N) \geq \int_{\mathcal{P}(\Omega)} D(\rho) \, dP_\infty(\rho)$$

*where $E^N$, $E^\infty$, $\overline{\mathcal{I}}$, and $D(\rho)$ are defined by (2.8), (2.9), (3.1), and (3.4), respectively. Moreover, for any $P_\infty \in \mathcal{P}(\mathcal{P}(\Omega))$ the sequence*

$$\rho_*^N = \int_{\mathcal{P}(\Omega)} \rho^{\otimes N} \, \mathrm{d}P_\infty(\rho) \in \mathcal{P}_{\mathrm{sym}}(\Omega^N)$$

*attains the limits the limits, i.e.*

$$\lim_{N\to\infty} E^N[\rho_*^N] = E^\infty[P_\infty] \qquad \text{and} \qquad \lim_{N\to\infty} \overline{\mathcal{I}}(\rho_*^N | M_N) = \int_{\mathcal{P}(\Omega)} D(\rho) \, dP_\infty(\rho) \tag{5.3}$$

PROOF. The convergence of the relative entropy functional follows directly from the classical arguments in [47]. A more modern proof can be found in [35]. The convergence of the relative Fisher information with respect to the Lebesgue measure is covered in [35]. Our

case is slightly more involved due to the minimal regularity assumptions on the potentials. Formally, expanding the square we obtain after integrating by parts

$$\begin{aligned}\overline{\mathcal{I}}(\rho^N, M_N) =& \frac{1}{N}\int_{\Omega^N} |\nabla \log \rho^N - \nabla H^N|^2 \,\mathrm{d}\rho^N \\ =& \int_{\Omega^N} (|\nabla \log \rho^N|^2 + 2\Delta H^N + |\nabla H^N|^2)\,\mathrm{d}\rho^N\,. \end{aligned} \tag{5.4}$$

The first term of the above expression is exactly the relative Fisher information with respect to the Lebesgue measure and so is already covered in [35]. Under stronger regularity assumptions on the potentials, the second and third term fall within the type of functionals already considered in the classical [47]. The main obstruction to conclude under Assumptions 2.1 and 2.2 is that $\Delta V$ and $\Delta W$ are merely signed measures bounded below, so we need to adapt the proofs of [35] to a lower regularity setting. We will circumvent the regularity problem, by appealing again and again to convexity. More specifically, by Assumptions 2.1 and 2.2, we know that

$$\overline{\mathcal{I}}^{1/2}(\rho^N|M_N) = \sup_{\nu^N\in\mathcal{P}_{\mathrm{sym}}(\Omega^N)} \left(\frac{E^N[\rho^N] - E^N[\nu^N]}{\overline{d}_2(\rho^N,\nu^N)} - \frac{K_V + K_W}{2}\overline{d}_2(\rho^N,\nu^N)\right)_+, \tag{5.5}$$

see [1, Theorem 2.4.9]. We pick $\nu^N$ to be the recovery sequence for a generic $Q_\infty \in \mathcal{P}(\mathcal{P}(\Omega))$, using the lower semicontinuous convergence of $E^N$, and the isometry Theorem 5.3, to obtain

$$\liminf_{N\to\infty} \overline{\mathcal{I}}^{1/2}(\rho^N|M_N) \geq \sup_{Q_\infty\in\mathcal{P}(\mathcal{P}(\Omega))} \left(\frac{E^\infty[P_\infty] - E^\infty[Q_\infty]}{\mathfrak{D}_2(P_\infty,Q_\infty)} - \frac{K_V + K_W}{2}\mathfrak{D}_2(P_\infty,Q_\infty)\right)_+.$$

To obtain equality, we consider $S_t : \mathcal{P}(\Omega)\to\mathcal{P}(\Omega)$ the solution operator associated to (2.5) which is well-defined by Theorem B and consider the curve $Q_\infty = S_t\#P_\infty$, which coincides with the unique gradient flow of $E^\infty$ with respect to $\mathfrak{D}_2$ (see [10, Lemma 19]). Taking $t\to 0^+$, we obtain

$$\begin{aligned}&\lim_{t\to 0^+}\left(\frac{E^\infty[P_\infty] - E^\infty[S_t\#P_\infty]}{\mathfrak{D}_2(P_\infty,S_t\#P_\infty)} - \frac{K_V+K_W}{2}\mathfrak{D}_2(P_\infty,S_t\#P_\infty)\right)_+ \\ &\quad = \left(\int_{\mathcal{P}(\Omega)} D(\rho)\ \mathrm{d}P_\infty(\rho)\right)^{1/2},\end{aligned}$$

obtaining the desired lower semicontinuity result.

$$\liminf_{N\to\infty}\mathcal{I}(\rho^N|M_N) \geq \int_{\mathcal{P}(\Omega)} D(\rho)\ \mathrm{d}P_\infty(\rho)$$

Now we focus on showing that given $P\in\mathcal{P}(\mathcal{P}(\Omega))$, the sequence

$$\rho_*^N = \int_{\mathcal{P}(\Omega)} \rho^{\otimes N}\ \mathrm{d}P(\rho) \tag{5.6}$$

attains the limit (5.3). For convenience, we consider the auxiliary functional $\mathcal{I}' : \mathcal{P}(\mathcal{P}(\Omega))\to\mathbb{R}\cup\{\infty\}$, which is given by

$$\mathcal{I}'[P] = \lim_{N\to\infty}\overline{\mathcal{I}}(\rho_*^N|M_N),$$

where $\rho_*^N$ is given by (5.6). Next, we show that $\mathcal{I}'$ is well defined and that it is weakly lower semicontinuous.

To show this, we regularize $\rho_*^N$ by using the $N$-particle dynamics (2.3). Using the convexity in Assumptions 2.1 and 2.2, we obtain that

$$e^{t(\lambda_V+\lambda_W)}\overline{\mathcal{I}}(S_t^N\rho_*^N|M_N) \leq \overline{\mathcal{I}}(\rho_*^N|M_N)\,,$$

where $S_t^N : \mathcal{P}(\Omega^N) \to \mathcal{P}(\Omega^N)$ is the solution operator associated to the Fokker–Planck equation (2.3). Using the weak lower semicontinuity of $\overline{\mathcal{I}}$ from the expression (5.5), we obtain

$$\lim_{t\to 0^+} \overline{\mathcal{I}}(S_t^N\rho_*^N|M_N) = \overline{\mathcal{I}}(\rho_*^N|M_N)\,. \tag{5.7}$$

Using the fact that $S_t^N$ is immediately regularizing, we can use the formal expression (5.4) freely for the measure $S_t^N\rho_*^N$. By using the sub-additivity of the Fisher information, we can notice that for any $3 < j < N$

$$\overline{\mathcal{I}}((S_t^N\rho_*^N)_j|M_j) \leq \overline{\mathcal{I}}(S_t^N\rho_*^N|M_N)\,, \tag{5.8}$$

where $(S_t^N\rho_*^N)_j$ is the $j^{\text{th}}$ marginal of $S_t^N\rho_*^N$. Taking $t \to 0^+$, using (5.7) and the weak lower semicontinuity from (5.5), we obtain for any $3 < j < N$

$$\overline{\mathcal{I}}(\rho_*^j|M_j) \leq \overline{\mathcal{I}}(\rho_*^N|M_N)\,,$$

which implies that we can characterise

$$\mathcal{I}'[P] = \sup_{N\geq 3} \overline{\mathcal{I}}(\rho_*^N|M_N).$$

By the weak lower semicontinuity of $\overline{\mathcal{I}}$, we obtain the weak lower semicontinuity of $\mathcal{I}'$.

Using (5.8) and the propagation of chaos result of [10], we obtain that

$$e^{t(\lambda_V+\lambda_W)}\overline{\mathcal{I}}(\rho_{t,*}^j|M_j) \leq e^{t(\lambda_V+\lambda_W)} \liminf_{N\to\infty} \overline{\mathcal{I}}((S_t^N\rho_*^N)_j|M_j) \leq \mathcal{I}'[P]\,, \tag{5.9}$$

where

$$\rho_{t,*}^j = \int_{\mathcal{P}(\Omega)} \rho^{\otimes j}\ \mathrm{d}(S_t\#P)(\rho) = \int_{\mathcal{P}(\Omega)} (S_t\rho)^{\otimes j}\ \mathrm{d}P(\rho)$$

with $S_t$ the solution operator to (2.5). Taking $j \to \infty$ in (5.9), we obtain the inequality

$$e^{t(\lambda_V+\lambda_W)} \int_{\mathcal{P}(\Omega)} D(S_t\#\rho)\ \mathrm{d}P(\rho) = e^{t(\lambda_V+\lambda_W)}\mathcal{I}'[S_t\#P] \leq \mathcal{I}'[P]\,,$$

where the equality in the left hand side follows from using the expression (5.4) and the result for the standard Fisher information in [35]. Taking $t \to 0^+$ and using the lower semicontinuity of $\mathcal{I}'$ we have showed above, we obtain the desired conclusion (5.3)

$$\mathcal{I}'[P] = \int_{\mathcal{P}(\Omega)} D(\rho)\ \mathrm{d}P(\rho).$$

□

Standard arguments also yield the convergence of the minimal energy.

COROLLARY 5.7. *Any accumulation point $P_\infty \in \mathcal{P}(\mathcal{P}(\Omega))$ of the sequence $(M_N)_{N\in\mathbb{N}}$ satisfies*

$$\mathrm{supp}(P_\infty) \subset \{\rho\,:\, E^{MF}[\rho] = \min_{\mathcal{P}(\Omega)} E^{MF}\}$$

*and*

$$E^N[M_N] = -\frac{1}{N}\log(Z_N) \overset{N\to\infty}{\to} \min_{\mathcal{P}(\Omega)} E^{MF}\,. \tag{5.10}$$

We remark that the convergence (5.10) coincides with the standard definition of the thermodynamic limit from statistical mechanics (see [36, Ch. 3]).

THEOREM 5.8 (HWI inequality). *Let $V$ and $W$ be $K$-convex. Given two arbitrary probability measures $\mu^N, \nu^N \in \mathcal{P}_{\mathrm{sym}}(\Omega^N)$, it holds*

$$\overline{\mathcal{E}}(\mu^N|M_N) \leq \overline{\mathcal{E}}(\nu^N|M_N) + \overline{d}_2(\mu^N,\nu^N)\sqrt{\overline{\mathcal{I}}(\mu^N|M_N)} - \frac{K}{2}\overline{d}_2^2(\mu^N,\nu^N)\,. \tag{5.11}$$

PROOF. The proof of this result can be found in [61, Theorem 30.22]. We just present the main idea of the proof for the reader's convenience. We consider $\gamma : [0,1] \to \mathcal{P}_{\mathrm{sym}}(\Omega^N)$ the 2-Wasserstein geodesic between $\mu^N$ and $\nu^N$. Under the hypothesis of $K$ convexity of the potentials, we obtain that $t \to \overline{\mathcal{E}}(\gamma(t)|M_N)$ is $K$-convex, which implies the desired inequality (5.11). □

COROLLARY 5.9. *Assume that Assumptions 2.1 and 2.2 hold true and that the sequence $(\rho^N)_{N\in\mathbb{N}}$ with $\rho^N \in \mathcal{P}_{\mathrm{sym}}(\Omega^N)$ such that $\rho^N$ converges to $P_\infty \in \mathcal{P}(\mathcal{P}(\Omega))$ in the sense of Definition 5.4 and*

$$\sup_{N\in\mathbb{N}} \overline{\mathcal{I}}(\rho^N|M_N) < \infty\,,$$

*then*

$$\lim_{N\to\infty} \overline{\mathcal{E}}(\rho^N|M_N) = E^\infty[P_\infty] - \min_{\mathcal{P}(\Omega)} E^{MF}.$$

PROOF. Given $P_\infty$, we consider the recovery sequence $(\rho_*^N)_{N\in\mathbb{N}}$ from Theorem 5.6. By the HWI inequality (5.11), we have

$$\overline{\mathcal{E}}(\rho^N|M_N) \leq \overline{\mathcal{E}}(\rho_*^N|M_N) + \overline{d}_2(\rho^N,\rho_*^N)\sqrt{\overline{\mathcal{I}}(\rho^N|M_N)} - \frac{K}{2}\overline{d}_2^2(\rho^N,\rho_*^N)\,.$$

Using that both sequences converge to $P_\infty$ and that $(\rho_*^N)_{N\in\mathbb{N}}$ is a recovery sequence, we obtain

$$\limsup_{N\to\infty}\left(E^N[\rho^N] - E^N[M_N]\right) = \limsup_{N\to\infty}\overline{\mathcal{E}}(\rho^N|M_N) \leq E^\infty[P_\infty] - \min_{\mathcal{P}(\Omega)} E^{MF}\,.$$

The reverse inequality follows from Theorem 5.6. □

REMARK 5.10. A version of this result without the potentials $V$ and $W$ can be found in [35].

## 6. Proof of Theorem 3.2.

PROOF OF THEOREM 3.2. We prove the result only for $E^{MF}$, as the proof for $E^N$ and even more general energies $E : \mathcal{P}(\Omega) \to [0,\infty]$ is analogous, see Remark 6.1. We consider $\rho : [0,\infty) \to \mathcal{P}(\Omega)$, the unique 2-Wasserstein gradient flow of $E^{MF}$ with initial condition $\rho_0 \in \mathcal{P}(\Omega)$, see Theorem B.

We notice that if $\lambda_{\mathrm{LS}}^\infty = 0$, then there is nothing left to prove. Hence, we can assume without loss of generality that $\lambda_{\mathrm{LS}}^\infty > 0$, which implies (cf. Property C and Proposition 4.1) that $E^{MF}$ does not admit non-minimising steady state. Using the version of LaSalle's invariance principle for gradient flows proved in [12, Theorem 4.11], we know that $\rho(t)$ accumulates on

the set of steady states of $E^{MF}$, as $t \to \infty$. Using the fact that all steady states are minimisers, we can find a sequence of times $t_n \to \infty$ such that

$$\rho(t_n) \rightharpoonup \rho_\infty \tag{6.1}$$

for some $\rho_\infty \in \mathcal{K}$. Differentiating $E^{MF}[\rho(t)]$ with respect to time and using that $\lambda_{\mathrm{LS}}^\infty > 0$ we obtain

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\left(E^{MF}[\rho(t)] - \inf_{\mathcal{P}(\Omega)} E^{MF}\right) &= -\int_\Omega \left|\beta^{-1}\nabla \log \rho(t) + \nabla V + \nabla W \star \rho(t)\right|^2 \,\mathrm{d}\rho(t)\\ &\leq -\sqrt{\lambda_{\mathrm{LS}}^\infty}\left(E^{MF}[\rho(t)] - \inf_{\mathcal{P}(\Omega)} E^{MF}\right)^{1/2}\\ &\quad\times\left(\int_\Omega \left|\beta^{-1}\nabla \log \rho(t) + \nabla V + \nabla W \star \rho(t)\right|^2 \,\mathrm{d}\rho(t)\right)^{1/2}.\end{aligned}$$

Integrating from 0 to $t_n$ for some $n \in \mathbb{N}$, we obtain

$$\begin{aligned}&\left(E^{MF}[\rho(t_n)] - \inf_{\mathcal{P}(\Omega)} E^{MF}\right)^{1/2} - \left(E^{MF}[\rho_0] - \inf_{\mathcal{P}(\Omega)} E^{MF}\right)^{1/2}\\ &\leq -\left(\frac{\lambda_{\mathrm{LS}}^\infty}{2}\right)^{1/2}\int_0^{t_n}\left(\frac{1}{2}\int_\Omega |\beta^{-1}\nabla \log \rho(s) + \nabla V + \nabla W \star \rho(s)|^2 \,\mathrm{d}\rho(s)\right)^{1/2}\mathrm{d}s\\ &\leq -\left(\frac{\lambda_{\mathrm{LS}}^\infty}{2}\right)^{1/2} d_2(\rho_0, \rho(t_n)),\end{aligned}$$

where the last inequality follows from the Benamou–Brenier [7] formulation of the 2-Wasserstein distance. Taking $t_n \to \infty$, applying (6.1), and rearranging, we end up with

$$\left(E^{MF}[\rho_0] - \inf_{\mathcal{P}(\Omega)} E^{MF}\right)^{1/2} \geq \left(\frac{\lambda_{\mathrm{LS}}^\infty}{2}\right)^{1/2} d_2(\rho_0, \rho_\infty) \geq \left(\frac{\lambda_{\mathrm{LS}}^\infty}{2}\right)^{1/2} d_2(\rho_0, \mathcal{K}).$$

The desired inequality (3.6) follows by squaring both sides. □

REMARK 6.1. This proof can easily be adapted to general $E : \mathcal{P}(\Omega) \to (0, \infty]$, that are regular enough to admit gradient flow solutions from arbitrary initial data, and that sub-level sets are weakly compact to ensure convergence of the gradient flows to steady states [12, Theorem 2.12].

## 7. Proof of Theorem 3.3.

PROOF OF THEOREM 3.3. For the proof of the first part of the theorem, we consider some $\rho \in \mathcal{P}(\Omega)$ such that $E^{MF}[\rho] > \inf_{\mathcal{P}(\Omega)} E^{MF} > -\infty$, see Theorem A. By Theorem 5.6 and Corollary 5.9, we can find a recovery sequence $P^N = \rho^{\otimes N}$, such that

$$\lim_{N\to\infty} \overline{\mathcal{E}}(P^N|M_N) = E^{MF}[\rho] - \inf_{\mathcal{P}(\Omega)} E^{MF} \qquad \text{and} \qquad \lim_{N\to\infty} \overline{\mathcal{I}}(P^N|M_N) = D(\rho).$$

Using the definition of the log Sobolev constant and passing to the limit as $N \to \infty$, we observe

$$\limsup_{N\to\infty} \lambda_{\mathrm{LS}}^N \leq \limsup_{N\to\infty} \frac{\overline{\mathcal{I}}(P^N|M_N)}{\overline{\mathcal{E}}(P^N|M_N)} = \frac{D(\rho)}{E^{MF}[\rho] - \inf E^{MF}}.$$

Taking the infimum over all $\rho \notin \mathcal{K}$, we obtain

$$\limsup_{N\to\infty} \lambda_{\mathrm{LS}}^N \leq \lambda_{\mathrm{LS}}^\infty = \inf_{\rho\notin\mathcal{K}} \frac{D(\rho)}{E^{MF}[\rho] - \inf_{\mathcal{P}(\Omega)} E^{MF}}.$$

This completes the proof of the first half of the theorem.

Next, we consider the case when there exists $\rho_* \in \mathcal{P}(\Omega)$ a non-minimising steady state, that is to say

$$D(\rho_*) = 0 \qquad \text{and} \qquad E^{MF}[\rho_*] > \inf_{\mathcal{P}(\Omega)} E^{MF}.$$

Consider now the sequence $\rho_*^{\otimes N} \in \mathcal{P}_{\mathrm{sym}}(\Omega^N)$. We have

$$\lambda_N^{\mathrm{LS}} \leq \frac{\overline{\mathcal{I}}(\rho_*^{\otimes N}|M_N)}{\overline{\mathcal{E}}(\rho_*^{\otimes N}|M_N)}$$

By Theorem 5.6 and Corollary 5.9, we can pass to the limit in the denominator to obtain

$$\lim_{N\to\infty} \overline{\mathcal{E}}(\rho_*^{\otimes N}|M_N) = E^{MF}[\rho_*] - \inf_{\mathcal{P}(\Omega)} E^{MF} > 0.$$

Thus, we have that for $N$ large enough

$$\lambda_N^{\mathrm{LS}} \leq \frac{2}{E^{MF}[\rho_*] - \inf_{\mathcal{P}(\Omega)} E^{MF}} \overline{\mathcal{I}}(\rho_*^{\otimes N}|M_N)\,.$$

The desired bound

$$\lambda_N^{\mathrm{LS}} \leq \frac{C}{N}$$

follows from the decay estimate for the Fisher information proved in Lemma 7.1. □

LEMMA 7.1. *Under Assumptions 2.1 and 2.2, assume that $\rho_*$ is a critical point of the mean field free energy $E^{MF}$, then we have the following bound*

$$\begin{aligned}\mathcal{I}(\rho_*^{\otimes N}|M_N) &= \left(1-\frac{1}{N}\right)\int_\Omega \left(|\nabla_1 W|^2 \star \rho_*(x_1) - |\nabla_1 W \star \rho_*(x_1)|^2\right)\rho_*(x_1)\,\mathrm{d}x_1 \\ &\quad + \frac{1}{N}\int |\nabla_1 W \star \rho_*(x_1)|^2 \rho_*(x_1)\,\mathrm{d}x_1 \\ &\leq C\,.\end{aligned}$$

PROOF OF LEMMA 7.1. We start by expanding $\mathcal{I}(\rho_*^{\otimes N}|M_N)$ as follows

$$\begin{aligned}\mathcal{I}(\rho_*^{\otimes N}|M_N) &= \int_{\Omega^N} \left|\nabla\left(\log\rho_*^{\otimes N} + \beta\sum_i V(x_i) + \frac{\beta}{2N}\sum_{i,j} W(x_i,x_j)\right)\right|^2 \rho_*^{\otimes N}\,\mathrm{d}x \\ &= \sum_i \int_{\Omega^N} \left|\frac{\nabla\rho_*(x_i)}{\rho_*(x_i)} + \beta\nabla V(x_i) + \frac{\beta}{N}\sum_j \nabla_1 W(x_i,x_j)\right|^2 \rho_*^{\otimes N}\,\mathrm{d}x \\ &= N\int_{\Omega^N} \left|-\beta\nabla_1 W\star\rho_*(x_1) + \frac{\beta}{N}\sum_j \nabla_1 W(x_1,x_j)\right|^2 \rho_*^{\otimes N}\,\mathrm{d}x\end{aligned}$$

$$=N\beta^2\int_{\Omega^N}\left(|\nabla_1 W\star\rho_*(x_1)|^2-\frac{2}{N}\nabla_1 W\star\rho_*(x_1)\cdot\sum_j\nabla_1 W(x_1,x_j)\right.$$
$$\left.+\frac{1}{N^2}\sum_{j,k}\nabla_1 W(x_1,x_j)\cdot\nabla_1 W(x_1,x_k)\rho_*^{\otimes N}\right)\,\mathrm{d}x,$$

where we have used symmetry of the particle system to write the integrand in terms of the variable $x_1$ and that $\rho_*$ is critical point of the mean field energy $E^{MF}$ which itself implies that

$$\beta^{-1}\log\rho_*+W\star\rho_*+V=C.$$

Next, proceeding term by term, we notice the following cancellation (for simplicity, we assume that $W(x,x)=0$, otherwise we can change $V$ by an additive constant such that this holds):

$$\int_{\Omega^N}|\nabla_1 W\star\rho_*(x_1)|^2\rho_*^{\otimes N}\,\mathrm{d}x=\int_\Omega|\nabla_1 W\star\rho_*(x_1)|^2\rho_*(x_1)\,\mathrm{d}x_1,$$
$$-\int_{\Omega^N}\frac{2}{N}\nabla_1 W\star\rho_*(x_1)\cdot\sum_j\nabla_1 W(x_1,x_j)\rho_*^{\otimes N}\,\mathrm{d}x=-\left(2-\frac{2}{N}\right)\int_\Omega|\nabla_1 W\star\rho_*(x_1)|^2\rho_*(x_1)\,\mathrm{d}x_1,$$

and

$$\int_{\Omega^N}\frac{1}{N^2}\sum_{j,k}\nabla_1 W(x_1,x_j)\cdot\nabla_1 W(x_1,x_k)\rho_*^{\otimes N}\,\mathrm{d}x=\frac{(N-1)(N-2)}{N^2}\int_\Omega|\nabla_1 W\star\rho_*(x_1)|^2\rho_*(x_1)\,\mathrm{d}x_1$$
$$+\frac{N-1}{N^2}\int|\nabla_1 W|^2\star\rho_*(x_1)\rho_*(x_1)\,\mathrm{d}x_1\,.$$

Putting the previous identities together, we obtain

$$\mathcal{I}(\rho_\beta^{\otimes N}|M_N)=\left(1-\frac{1}{N}\right)\int_\Omega\left(|\nabla_1 W|^2\star\rho_*(x_1)-|\nabla_1 W\star\rho_*(x_1)|^2\right)\rho_*(x_1)\,\mathrm{d}x_1$$
$$+\frac{1}{N}\int|\nabla_1 W\star\rho_*(x_1)|^2\rho_*(x_1)\,\mathrm{d}x_1\,.$$

Finally, the second inequality in the statement follows from applying Jensen's inequality to obtain

$$|\nabla_1 W|^2\star\rho_*(x)\geq|\nabla_1 W\star\rho_*(x)|^2\qquad\text{for every } x\in\Omega.$$

The final bound follows from the assumption that $W$ grows at most polynomially (cf. Assumption 2.2) and the fact that all steady states have finite moments of every order by Proposition 4.1. □

## 8. Proof of Theorem 3.4.

PROOF OF THEOREM 3.4. We prove the statement of the theorem by contradiction. To this end, we assume that there exists a sequence $N_i\to\infty$ and a sequence of times $t_i>0$

$$0<t_i<\frac{2}{\lambda_{\mathrm{LS}}^\infty}\log((E^{N_i}[\rho_{\mathrm{in}}^{\otimes N_i}]-E^{N_i}[M_{N_i}])/\varepsilon)$$

such that

$$\overline{\mathcal{E}}(\rho^{N_i}(t_i)|M_{N_i})>\max(\varepsilon,e^{-\frac{\lambda_{\mathrm{LS}}^\infty t_i}{2}}(E^{N_i}[\rho_{\mathrm{in}}^{\otimes N_i}]-E^{N_i}[M_{N_i}]).$$

Using the finite energy assumption on the initial condition (3.7), and Assumptions 2.1 and 2.2 which imply that $E^N$ is uniformly bounded below, we have that

$$\sup_i t_i \leq \sup_i \frac{2}{\lambda^\infty_{LSI}} \log((E^{N_i}[\rho_{\text{in}}^{\otimes N_i}] - E^{N_i}[M_{N_i}])/\varepsilon) =: T(\varepsilon) < \infty. \tag{8.1}$$

By the relative entropy dissipation estimate (3.1), we obtain

$$\overline{\mathcal{E}}(\rho^{N_i}(t_i)|M_{N_i}) > e^{-\frac{\lambda^\infty_{\text{LS}} t_i}{2}}(E^{N_i}[\rho_{\text{in}}^{\otimes N_i}] - E^{N_i}[M_{N_i}])$$

which implies that there exists at least one $s_i \in (0, t_i)$ such that

$$\frac{\beta^{-1}\overline{\mathcal{I}}(\rho^{N_i}(s_i)|M_{N_i})}{\overline{\mathcal{E}}(\rho^{N_i}(s_i)|M_{N_i})} < \frac{\lambda^\infty_{\text{LS}}}{2}.$$

The contradiction will arise if we can show that

$$\liminf_{i\to\infty} \frac{\beta^{-1}\overline{\mathcal{I}}(\rho^{N_i}(s_i)|M_{N_i})}{\overline{\mathcal{E}}(\rho^{N_i}(s_i)|M_{N_i})} \geq \lambda^\infty_{\text{LS}} = \inf_{\rho\notin\mathcal{K}} \frac{D(\rho)}{E^{MF}[\rho] - \inf_{\mathcal{P}(\Omega)} E^{MF}}.$$

First, we consider the case when

$$\liminf_{i\to\infty} \overline{\mathcal{I}}(\rho^{N_i}(s_i)|M_{N_i}) = \infty\,.$$

By using the monotonicity of the relative entropy (3.1) and the finite energy assumption in the statement of the theorem, we obtain

$$\overline{\mathcal{E}}(\rho^{N_i}(s_i)|M_{N_i}) \leq \overline{\mathcal{E}}(\rho_{\text{in}}^{\otimes N_i}|M_{N_i}) = E^{MF}[\rho_{\text{in}}] - E^N[M_N] < \infty.$$

Hence, we can conclude

$$\liminf_{i\to\infty} \frac{\beta^{-1}\overline{\mathcal{I}}(\rho^{N_i}(s_i)|M_{N_i})}{\overline{\mathcal{E}}(\rho^{N_i}(s_i)|M_{N_i})} = \infty > \lambda^\infty_{\text{LS}}.$$

Therefore, we can assume that up to a subsequence, which we do not relabel,

$$\sup_i \overline{\mathcal{I}}(\rho^{N_i}(s_i)|M_{N_i}) < \infty.$$

By the bounded higher moment hypothesis in (3.7), the uniform boundedness of $s_i$ (8.1), and the propagation of moments along the flow for $K$-convex potentials [10], we have that $\rho^{N_i}(s_i)$ has uniformly bounded moments of order $2+\delta$. By Proposition 5.5, we obtain that there exists $P_* \in \mathcal{P}(\mathcal{P}(\mathbb{R}^d))$, such that up to a (not relabeled) subsequence, $\rho^{N_i}(s_i) \to P_*$ in the sense of Definition 5.4. By the HWI inequality Theorem 5.8 and Corollary 5.9, we obtain the strong convergence in the relative entropy term, which yields

$$\varepsilon \leq \lim_{i\to\infty} \overline{\mathcal{E}}(\rho^{N_i}(s_i)|M_{N_i}) = \int_{\mathcal{P}(\mathbb{R}^d)} E^{MF}[\rho] - \inf E^{MF}\ \mathrm{d}P_*(\rho).$$

Combining this limit with the $\liminf$-inequality of Theorem 5.6 for the dissipation we obtain that

$$\liminf_{i\to\infty} \frac{\beta^{-1}\overline{\mathcal{I}}(\rho^{N_i}(s_i)|M_{N_i})}{\overline{\mathcal{E}}(\rho^{N_i}(s_i)|M_{N_i})} \geq \frac{\int_{\mathcal{P}(\mathbb{R}^d)} D(\rho)\ \mathrm{d}P_*(\rho)}{\int_{\mathcal{P}(\mathbb{R}^d)} E^{MF}[\rho] - \inf E^{MF}\ \mathrm{d}P_*(\rho)} \geq \lambda^\infty_{\text{LS}},$$

where the last inequality follows from the point-wise inequality

$$D(\rho) \geq \lambda^\infty_{\text{LS}}(E^{MF}[\rho] - \inf E^{MF})\,.$$

This is the desired contradiction and the result now follows. □

**9. Proof of Theorem 3.6.**

PROOF OF THEOREM 3.6. The uniqueness of the minimiser in the limit follows from applying the Talagrand inequality (3.6), which states that the energy grows quadratically around the Gibbs measure $M_N$. We take $\rho_1$ and $\rho_2$, two minimisers of $E^{MF}$, and show that they must coincide. By the triangle and Talangrand inequality, we have

$$\begin{aligned} d_2^2(\rho_1,\rho_2) = \overline{d}_2^2(\rho_1^{\otimes N},\rho_2^{\otimes N}) &\leq 2\overline{d}_2^2(\rho_1^{\otimes N},M_N) + 2\overline{d}_2^2(M_N,\rho_2^{\otimes N}) \\ &= \frac{4}{\lambda_{\mathrm{LS}}^N}(E^{MF}[\rho_1] + E^{MF}[\rho_2] - 2E^N[M_N]). \end{aligned}$$

The fact that $\rho_1$ is equal to $\rho_2$ follows from taking the limit in the previous inequality, using the hypothesis that $\limsup_{N\to\infty} \lambda_{\mathrm{LS}}^N > 0$, and Corollary 5.9 to obtain

$$\lim_{N\to\infty} E^N[M_N] = E^{MF}[\rho_1] = E^{MF}[\rho_2] = \inf_{\mathcal{P}(\Omega)} E^{MF}.$$

The quantitative convergence of $M_N$ to $\rho_\beta$, follows from the bound in Lemma 9.1. □

LEMMA 9.1. *Under Assumptions 2.1 and 2.2, assume that* $\liminf_{N\to\infty} \lambda_{\mathrm{LS}}^N =: \lambda^\infty > 0$. *Consider $M_N$ the $N$ particle Gibbs measure and $\rho_\beta$ the unique minimiser of the mean field energy. Then, for $N$ large enough,*

$$\overline{d}_2(M_N,\rho_\beta^{\otimes N}) \leq \frac{2}{\lambda^\infty}\frac{\displaystyle\int_\Omega |\nabla_1 W|^2 \star \rho_\beta\rho_\beta \,\mathrm{d}x}{\sqrt{N}} \leq \frac{C}{\sqrt{N}}.$$

PROOF OF LEMMA 9.1. Using the Talagrand and log Sobolev inequality, we have for $N$ large enough

$$\overline{d}_2^2(M_N,\rho_\beta^{\otimes N}) \leq \frac{2}{\lambda_{\mathrm{LS}}^N}\frac{\mathcal{E}(\rho_\beta^{\otimes N}|M_N)}{N} \leq \frac{2}{(\lambda_{\mathrm{LS}}^N)^2}\frac{\mathcal{I}(\rho_\beta^{\otimes N}|M_N)}{N} \leq \frac{4}{(\lambda^\infty)^2}\frac{\mathcal{I}(\rho_\beta^{\otimes N}|M_N)}{N}.$$

The result now follows from bounding the Fisher information using Lemma 7.1. □

**10. Proof of Theorem 3.7.** We start by revisiting the classical propagation of chaos results [57, 44] by using a convexity approach based on the 2-Wasserstein distance.

THEOREM 10.1. *Under Assumptions 2.1 and 2.2, if* $K_V + K_W(1-1/N) \neq 0$, *then*

(10.1)
$$\overline{d}_2(\rho^N(t),\rho^{\otimes N}(t)) \leq \frac{1-e^{-\frac{K_V+K_W(1-1/N)}{2}t}}{K_V+K_W(1-1/N)}\left(\frac{\sup_{s\in[0,t]}\left(\int_\Omega |\nabla_1 W|^2\star\rho(s)\rho(s)\,\mathrm{d}x\right)^{1/2}}{N^{1/2}}\right),$$

*else if* $K_V + K_W(1-1/N) = 0$, *then*

$$\overline{d}_2(\rho^N(t),\rho^{\otimes N}(t)) \leq \frac{t}{2}\left(\frac{\sup_{s\in[0,t]}\left(\int_\Omega |\nabla_1 W|^2\star\rho(s)\rho(s)\,\mathrm{d}x\right)^{1/2}}{N^{1/2}}\right), \tag{10.2}$$

PROOF OF THEOREM 10.1. Using [1, Theorem 8.4.7], we differentiate the 2-Wasserstein distance between $\rho^N$ and $\rho^{\otimes N}$ along their respective flows (2.3) and (2.5), to obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}\overline{d}_2^2(\rho^N,\rho^{\otimes N}) = -\frac{1}{N}\int_{\Omega^N\times\Omega^N}(x-y)\cdot\Big(\nabla_x\left(\beta^{-1}\log\rho^N(x) + H_N(x)\right)$$

$$-\nabla_y\left(\sum_{i=1}^N \beta^{-1}\log\rho(y_i)+V(y_i)+W\star\rho(y_i)\right)\Big)\ \mathrm{d}\Pi(x,y) \tag{10.3}$$

where $\Pi\in\mathcal{P}(\Omega^N\times\Omega^N)$ denotes the optimal transport plan between $\rho^N$ and $\rho^{\otimes N}$. The convexity along 2-Wasserstein geodesics of the entropy functional as discussed in [46] implies that (cf. [1, Section 10.1.1])

$$\int_{\Omega^N\times\Omega^N}(x-y)\cdot(\beta^{-1}\nabla_x\log\rho^N(x)-\beta^{-1}\nabla_y\log\rho^{\otimes N}(y))\ \mathrm{d}\Pi(x,y)\geq 0, \tag{10.4}$$

where we have used that $\nabla\log\rho^N$ and $\nabla\log\rho^{\otimes N}$ are in the sub-differential of the entropy at $\rho^N$ and $\rho^{\otimes N}$, respectively (cf. [1, Theorem 10.4.6]).

Applying inequality (10.4) to (10.3), we obtain

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\overline{d}_2^2(\rho^N,\rho^{\otimes N})&\leq -\frac{1}{N}\int_{\Omega^N\times\Omega^N}(x-y)\cdot\Bigg(\nabla_x H_N(x)-\nabla_y H_N(y)\\&\qquad+\nabla_y\left(\frac{1}{2N}\sum_{i,j=1}^N W(y_i,y_j)-\sum_{i=1}^N W\star\rho(y_i)\right)\Bigg)\ \mathrm{d}\Pi(x,y)\\&\leq -(K_V+K_W(1-1/N))\underbrace{\frac{1}{N}\int_{\Omega^N\times\Omega^N}|x-y|^2\ \mathrm{d}\Pi(x,y)}_{=\overline{d}_2^2(\rho^N,\rho^{\otimes N})}\end{aligned}$$

$$-\underbrace{\frac{1}{N}\int_{\Omega^N\times\Omega^N}(x-y)\cdot\nabla_y\left(\frac{1}{2N}\sum_{i=1}^N\sum_{j=1}^N W(y_i,y_j)-\sum_{i=1}^N W\star\rho(y_i)\right)\ \mathrm{d}\Pi(x,y)}_{\mathcal{R}}, \tag{10.5}$$

where the last inequality follows from the convexity hypothesis on the potentials (cf. Assumptions 2.1 and 2.2). To estimate the second term $\mathcal{R}$, we employ Cauchy-Schwarz inequality and use the fact that $\Pi$ is the optimal transference plan to obtain

$$\mathcal{R}\leq\overline{d}_2(\rho^N,\rho^{\otimes N})\left(\underbrace{\frac{1}{N}\int_{\Omega^N}\left|\nabla\left(\frac{1}{2N}\sum_{i=1}^N\sum_{j=1}^N W(y_i,y_j)-\sum_{i=1}^N W\star\rho(y_i)\right)\right|^2\ \mathrm{d}\rho^{\otimes N}}_{I}\right)^{1/2} \tag{10.6}$$

Expanding the square, using the symmetry of the underlying system and the fact that $W(y,y)=0$, we obtain

$$\begin{aligned}I&=\int_{\Omega^N}\left|\frac{1}{N}\sum_{j=2}^N\nabla_1 W(y_1,y_j)-\nabla_1 W\star\rho(y_1)\right|^2\ \mathrm{d}\rho^{\otimes N}\\&=\int_{\Omega^N}\underbrace{\frac{1}{N^2}\sum_{j=2}^N\sum_{k=2}^N\nabla_1 W(y_1,y_j)\nabla_1 W(y_1,y_k)}_{A}-\underbrace{\frac{2}{N}\sum_{j=2}^N\nabla_1 W(y_1,y_j)\nabla_1 W\star\rho(y_1)}_{B}\\&\qquad+(\nabla_1 W\star\rho(y_1))^2\ \mathrm{d}\rho^{\otimes N}.\end{aligned}$$

Going term by term, we have

$$A = \frac{(N-1)(N-2)}{N^2}\int_\Omega (\nabla_1 W \star \rho(y_1))^2 \, \mathrm{d}\rho(y_1) + \frac{(N-1)}{N^2}\int_\Omega |\nabla_1 W|^2 \star \rho(y_1) \, \mathrm{d}\rho(y_1),$$

$$B = -2\frac{N-1}{N}\int_\Omega (\nabla_1 W \star \rho(y_1))^2 \, \mathrm{d}\rho(y_1).$$

Using these identities, we are left with

$$I = \frac{1}{N}\left(1-\frac{1}{N}\right)\int_\Omega |\nabla_1 W|^2 \star \rho(y_1) - (\nabla_1 W \star \rho(y_1))^2 \, \mathrm{d}\rho(y_1) + \frac{1}{N^2}\int_\Omega (\nabla_1 W \star \rho(y_1))^2 \, \mathrm{d}\rho(y_1)$$
$$\leq \frac{1}{N}\int_\Omega |\nabla_1 W|^2 \star \rho(y_1) \, \mathrm{d}\rho(y_1),$$

where the last inequality follows from Jensen's inequality. Replacing the previous equation in (10.6), combined with (10.5), we obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}\overline{d}_2^2(\rho^N, \rho^{\otimes N}) \leq -(K_V + K_W(1-1/N))\overline{d}_2^2(\rho^N, \rho^{\otimes N}) + \overline{d}_2(\rho^N, \rho^{\otimes N})\frac{\left(\int_\Omega |\nabla_1 W|^2 \star \rho\rho \, \mathrm{d}x\right)^{1/2}}{N^{1/2}}.$$

The estimates (10.1) and (10.2) now follow from Grönwall's inequality. □

We are now ready to prove Theorem 3.7.

PROOF OF THEOREM 3.7. Using the uniform integrability assumption on the gradients of $W$ (3.8), the estimates of Theorem 10.1 simplify to

$$\overline{d}_2(\rho^N(t), \rho^{\otimes N}(t)) \leq \frac{C}{N^{1/2}}\begin{cases} \frac{1-e^{t(K_V+K_W(1-1/N))/2}}{K_V+K_W(1-1/N)} & \text{if } K_V + K_W(1-1/N) \geq 0 \\ \frac{t}{2} & \text{if } K_V + K_W(1-1/N) = 0 \\ \frac{e^{t|K_V+K_W(1-1/N)|/2}-1}{|K_V+K_W(1-1/N)|} & \text{if } K_V + K_W(1-1/N) \leq 0 \end{cases} \tag{10.7}$$

Next, we derive a competing estimate by employing the triangle inequality and the long time behavior of the flows. More specifically,

$$\begin{aligned}\overline{d}_2(\rho^N(t), \rho^{\otimes N}(t)) &\leq \overline{d}_2(\rho^N(t), M_N) + \overline{d}_2(M_N, \rho_\beta^{\otimes N}) + \overline{d}_2(\rho_\beta^{\otimes N}, \rho^{\otimes N}(t)) \\ &= \overline{d}_2(\rho^N(t), M_N) + \overline{d}_2(M_N, \rho_\beta^{\otimes N}) + d_2(\rho_\beta, \rho(t)).\end{aligned} \tag{10.8}$$

For the first term, we use the Talagrand inequality (3.6)

$$\overline{d}_2(\rho^N(t), M_N) \leq \left(\frac{2}{\lambda_{\mathrm{LS}}^N}\right)^{1/2} \overline{\mathcal{E}}^{1/2}(\rho^N(t)|M_N).$$

By the log Sobolev inequality we obtain exponential contraction of the relative entropy

$$\overline{d}_2(\rho^N(t), M_N) \leq e^{-\frac{\lambda_{\mathrm{LS}}^N}{2}t}\left(\frac{2}{\lambda_{\mathrm{LS}}^N}\right)^{1/2} \overline{\mathcal{E}}^{1/2}(\rho_{\mathrm{in}}^{\otimes N}(t)|M_N) \leq Ce^{-\frac{\lambda^\infty}{4}t}, \tag{10.9}$$

where in the last equality we have used the hypothesis $\rho_{\mathrm{in}}$ has finite energy and that the log Sobolev constant does not degenerate.
For the second term, we use Lemma 9.1 to obtain

$$\overline{d}_2(M_N, \rho_\beta^{\otimes N}) \leq \frac{C}{\sqrt{N}}. \tag{10.10}$$

For the third term, we use the limiting Talgrand inequality and the limiting log Sobolev inequality to obtain the exponential contraction estimate

$$\begin{aligned} d_2(\rho(t),\rho_\beta) &\le \left(\frac{2}{\lambda_\infty}\right)^{1/2} (E^{MF}[\rho(t)] - E^{MF}[\rho_\beta])^{1/2} \\ &\le \left(\frac{2}{\lambda_\infty}\right)^{1/2} e^{-\frac{\lambda_\infty}{2}t} (E^{MF}[\rho_{\text{in}}] - E^{MF}[\rho_\beta])^{1/2} \\ &\le C e^{-\frac{\lambda_\infty}{2}t}. \end{aligned} \tag{10.11}$$

Combining (10.8) with (10.9), (10.10) and (10.11) we obtain the estimate

$$\overline{d}_2(\rho^N(t), \rho^{\otimes N}(t)) \le C\left(e^{-\frac{\lambda_\infty}{4}t} + \frac{1}{\sqrt{N}}\right). \tag{10.12}$$

The result now follows from interpolating the estimates (10.7) and (10.12). In the case, $K_V + K_W(1-1/N) > 0$ the desired estimate follows directly from (10.7). For $K_- := K_V + K_W(1-1/N) < 0$, we consider the distinguished time scale $T_N := \log N^\gamma$, for some $\gamma > 0$ to be chosen in terms of $K_-$. Applying (10.12), we obtain

$$\overline{d}_2(\rho^N(t), \rho^{\otimes N}(t)) \le C\left(N^{-\gamma\frac{\lambda_\infty}{4}} + N^{-\frac{1}{2}}\right) \qquad \text{for } t > T_N.$$

For $t < T_N$, we apply (10.7)1 to obtain

$$\overline{d}_2(\rho^N(t), \rho^{\otimes N}(t)) \le C N^{-\frac{1+\gamma K_-}{2}}.$$

Choosing

$$\gamma = \left(\frac{\lambda_\infty}{2} - K_-\right)^{-1},$$

we obtain that for every $t \in (0,\infty)$

$$\overline{d}_2(\rho^N(t), \rho^{\otimes N}(t)) \le \frac{C}{N^\theta}$$

is satisfied with

$$\theta = \frac{1}{2}\frac{\lambda_\infty}{\lambda_\infty - 2K_-}.$$

□

**11. Proof of Theorem 3.14.** As mentioned earlier, our proof of Theorem 3.14 will rely on the two-scale approach to log Sobolev inequalities introduced in [49] and discussed further in [31], see also [41]. Before we introduce the main result of [49], we introduce some preliminary notions.

DEFINITION 11.1 (Conditional measures). Given a probability measure $\mu_N \in \mathcal{P}(\Omega^N)$ we define the conditional measure $\mu_{N,i}, i \in 1,\dots,N$ as the family of measures indexed by $x_j$, $j \neq i$ such that for all $\varphi \in C_b(\Omega^N)$

$$\int_{\Omega^N} \varphi \,\mathrm{d}\mu_N = \int_{\Omega^{N-1}} \int_\Omega \varphi \,\mathrm{d}\mu_{N,i} \,\mathrm{d}\mu_{N\setminus\{i\}},$$

where $\mu_{N\setminus\{i\}}$ is the marginal of $\mu_N$ obtained by integrating out $x_i \in \Omega$.

We can now the state the result of interest.

THEOREM 11.2 ([49, Theorem 1]). *Let $\Omega$ be a smooth, connected, and complete Riemannian manifold and assume that the measure $\mu_N$ has a Gibbs structure, that is to say*

$$\mu_N(\mathrm{d}x) = Z_N^{-1} e^{-\beta H_N}\,\mathrm{d}x\,,$$

*where $\mathrm{d}x$ is the Riemannian volume measure and $H_N : \Omega^N \to \mathbb{R}$ is some smooth Hamiltonian. Assume there exists some constants $\kappa_{ij}$, such that for all $i \neq j$*

$$\|D^2_{x_i x_j} H_N\| \leq \kappa_{ij} \qquad \text{for all } x \in \Omega^N,$$

*where $\|\cdot\|$ is the operator norm of $D^2_{x_i x_j} H_N$. Furthermore, assume that the conditional measures $\mu_{N,i}$ satisfy a log Sobolev inequality with uniform constant $\lambda_{\mathrm{LS}}^{N,i}$ for all $\hat{x} \in \Omega^{N-1}$. Consider the matrix $A \in \mathbb{R}^{N\times N}$ with entries $A_{ii} = \lambda_{\mathrm{LS}}^{N,i}$ and $A_{ij} = -\beta\kappa_{ij}$; if*

$$A \geq C I^{N\times N}\,,$$

*the measure $\mu_N$ satisfies a log Sobolev inequality with constant $C$.*

Relying on Theorem 11.2, we present now the proof of Theorem 3.14.

PROOF OF THEOREM 3.14. We note first that from the Definition 11.1, the conditional measure $M_{N,i}$ of $M_N$ can be expressed as

$$M_{N,i}(\mathrm{d}x_i) = M_N(\mathrm{d}x_i | x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_N) = \frac{M_N}{(M_N)_{N\setminus\{i\}}}\,,$$

where $(M_N)_{N\setminus\{i\}}$ is the marginal of $M_N$ obtained by integrating out $x_i$. We are thus left with

$$M_{N,i} = \frac{\exp\Big(-\beta(V(x_i) + \frac{1}{N}\sum_{j=1}^N W(x_i,x_j)) - \beta(\sum_{j\neq i, j=1}^N V(x_j) + \frac{1}{2N}\sum_{j,k=1, j,k\neq i} W(x_j,x_k))\Big)}{\int_\Omega \exp(-\beta H_N)\,\mathrm{d}x_i}$$

$$= Z_{N,i}^{-1} \exp\left(-\beta(V(x_i) + \frac{1}{N}\sum_{j=1}^N W(x_i,x_j))\right),$$

where

$$Z_{N,i} = \int_\Omega \exp\left(-\beta(V(x_i) + \frac{1}{N}\sum_{j=1}^N W(x_i,x_j))\right)\mathrm{d}x_i\,.$$

We now assert that the conditional measure $M_{N,i}$ satisfies a log Sobolev inequality. We first treat the case in which $\Omega$ is compact, that is to say Theorem 3.14 (a). By the Holley–Stroock perturbation Theorem [3, Proposition 5.1.6], we have that

$$\lambda_{\mathrm{LS}}^{N,i} \geq e^{-2\beta(\|W\|_{L^\infty(\Omega^2)} + \|V\|_{L^\infty(\Omega)})}\lambda_{\mathrm{LS}}^{\Omega}\,,$$

for all $x_j, j \neq i, j = 1, \ldots, N$ and where $\lambda_{\mathrm{LS}}^{\Omega}$ is the optimal log Sobolev constant of the Lebesgue measure on $\Omega$.

Note that because of the exchangeability of the underlying particle system, we have that $\lambda_{\mathrm{LS}}^{N,i} = \lambda_{\mathrm{LS}}^{N,j}$ for all $i, j = 1, \ldots, N$.

Note now that

$$D^2_{x_i x_j} H_N(x_1, \dots, x_N) = \frac{1}{N} D^2_{x_i x_j} W(x_i, x_j)\,.$$

Using the hypothesis that $W \in W^{2,\infty}(\Omega^2)$, we can bound

$$\|D^2_{x_i x_j} H_N\|_{L^\infty(\Omega^N)} \leq \frac{1}{N} \|D^2_{xy} W\|_{L^\infty(\Omega^2)} =: \kappa_{ij}\,,$$

for all $i, j = 1, \dots, N$. We will show that the matrix $A \in \mathbb{R}^{N\times N}$ from Theorem 11.2 is positive definite, by showing that it is diagonally dominant. In fact, for $\beta$ sufficiently small we have that

$$A_{ii} - \sum_{j\neq i} |A_{ij}| \geq e^{-2\beta(\|W\|_{L^\infty(\Omega^2)} + \|V\|_{L^\infty(\Omega)})} \lambda^{\Omega}_{\mathrm{LS}} - \beta \frac{N-1}{N} \|D^2_{xy} W\|_{L^\infty(\Omega^2)} > c > 0\,,$$

holds true for all $N$ with the constant $c$ independent of $N$. Applying Theorem 11.2, Theorem 3.14 (a) now follows.

For the proof of Theorem 3.14 (b) we can apply essentially the same perturbative argument as before but now around the measure $Z_V^{-1} e^{-V}\,\mathrm{d}x$. For $\epsilon$ sufficiently small, we obtain that the analogous bound

$$A_{ii} - \sum_{j\neq i} |A_{ij}| \geq e^{-2\epsilon\beta\|W\|_{L^\infty(\Omega^2)}} \lambda^{V}_{\mathrm{LS}} - \epsilon\beta \frac{N-1}{N} \|D^2_{xy} W\|_{L^\infty(\Omega^2)} > c > 0$$

holds true for all $N$ with the constant $c$ independent of $N$. □

**12. Proof of Theorem 3.11.** To simplify the computations in this section, we will take the following definition of the negative Sobolev norm $H_0^{-s}(\mathbb{T}^d)$ of mean-zero distributions is given by

$$\|h\|^2_{H^{-s}(\mathbb{T}^d)} := \sum_{j\in\mathbb{N}} |\langle h, \phi_j\rangle|^2, \tag{12.1}$$

where $(\phi_j)_{j\in\mathbb{N}}$ is a given a smooth orthonormal basis for the Sobolev space $H^s_0(\mathbb{T}^d)$.

We remark that Theorem 3.11 can also be proved when $\Omega = \mathbb{R}^d$, under appropriate assumptions on the confining and interaction potentials. In particular, we can construct an appropriate orthonormal basis using the eigenfunctions of the linearised McKean-Vlasov operator with a weighted inner-product. Conditions on the growth of the confining potential so that the Sobolev embedding theorems needed in the proof in the appropriate weighted spaces are given in [42].

LEMMA 12.1 (Law of large numbers). *Assume that* $\liminf_{N\to\infty} \lambda^N_{\mathrm{LS}} > 0$. *Let* $\mu^{(N)}$ *be the empirical measure associated to the* $N$*-particle Gibbs measure* $M_N \in \mathcal{P}_{\mathrm{sym}}((\mathbb{T}^d)^N)$. *Then, for any* $s > \frac{d+2}{2}$, *there exists* $C > 0$ *such that*

$$\mathbb{E}\Big[\|\mu^{(N)} - \rho_\beta\|^2_{H^{-s}(\mathbb{T}^d)}\Big] \leq \frac{C}{N}\,,$$

*where* $\rho_\beta \in \mathcal{P}(\mathbb{T}^d)$ *is the unique critical point of the mean field energy* $E^{MF}$.

REMARK 12.2. The solutions to the linear SPDE (12.6) are supported in $H^{-s}$ with $s > d/2$, see Lemma 12.4. Hence, the Law of large numbers does not hold for any $H^{-s}$ with $s < d/2$.

PROOF. We consider $\rho_\beta^{(N)}$ the empirical measure associated to $\rho_\beta^{\otimes N} \in \mathcal{P}_{\mathrm{sym}}((\mathbb{T}^d)^N)$, i.e. $\rho_\beta^{(N)}$ the probability-measure valued random variable defined as

$$\rho_\beta^{(N)} = \frac{1}{N}\sum_{i=1}^N \delta_{X_i} \qquad \text{such that } (X_1, ..., X_N) \text{ are distributed according to } \rho_\beta^{\otimes N}.$$

By the triangle inequality we have that

(12.2)
$$\mathbb{E}\Big[\|\mu^{(N)} - \rho_\beta\|^2_{H^{-s}(\mathbb{T}^d)}\Big] \le 2\Big(\mathbb{E}\Big[\|\mu^{(N)} - \rho_\beta^{(N)}\|^2_{H^{-s}(\mathbb{T}^d)}\Big] + \mathbb{E}\Big[\|\rho_\beta^{(N)} - \rho_\beta\|^2_{H^{-s}(\mathbb{T}^d)}\Big]\Big)$$

We start by controlling the first term $\mathbb{E}\Big[\|\mu^{(N)} - \rho_\beta^{(N)}\|^2_{H^{-s}(\mathbb{T}^d)}\Big]$. We consider the optimal coupling between $\mu^{(N)}$ and $\rho_\beta^{(N)}$ such that

$$\mathbb{E}\Big[d_2^2(\mu^{(N)}, \rho_\beta^{(N)})\Big] = \mathfrak{D}_2^2\left(\hat{\mu}^N, \hat{\rho}_\beta^N\right),$$

where $\hat{\mu}^N = \mathrm{Law}\left(\mu^{(N)}\right) \in \mathcal{P}(\mathcal{P}(\mathbb{T}^d))$, $\hat{\rho}_\beta^N = \mathrm{Law}\left(\rho_\beta^{(N)}\right) \in \mathcal{P}(\mathcal{P}(\mathbb{T}^d))$, and $\mathfrak{D}_2$ is the 2-Wassertein distance defined on the space of probability measures of the metric space $(\mathcal{P}(\mathbb{T}^d), d_2)$. Then by the isometry from Theorem 5.3, we obtain

$$\mathfrak{D}_2^2\left(\hat{\mu}^N, \hat{\rho}_\beta^N\right) = \overline{d}_2^2(M_N, \rho_\beta^{\otimes N}) \le \frac{C}{N},$$

where we have used Lemma 9.1 and the fact that $\liminf_{N\to\infty} \lambda_{\mathrm{LS}}^N > 0$ for the last inequality. Next, using the compactness of $\mathbb{T}^d$ and the fact that $s > d/2+1$, we have the pointwise bound

$$d_2^2(\mu^{(N)}, \rho_\beta^{(N)}) \ge c_1\|\mu^{(N)} - \rho_\beta^{(N)}\|^2_{W^{-1,\infty}(\mathbb{T}^d)} \ge c_2\|\mu^{(N)} - \rho_\beta^{(N)}\|^2_{H^{-s}(\mathbb{T}^d)},$$

for some constants $c_1, c_2 > 0$ independent of $\mu^{(N)}, \rho_\beta^{(N)}$. Combining the previous three expressions, we obtain the first desired bound

$$\mathbb{E}\Big[\|\mu^{(N)} - \rho_\beta^{(N)}\|^2_{H^{-s}(\mathbb{T}^d)}\Big] \le \frac{C}{N}. \tag{12.3}$$

Next, we bound $\mathbb{E}\Big[\|\rho_\beta^{(N)} - \rho_\beta\|^2_{H^{-s}(\mathbb{T}^d)}\Big]$. We take $\{\phi_j\}_{j=0}^\infty$ to be the orthonormal basis of $H_0^s(\mathbb{T}^d)$ from (12.1). Expanding the square, we obtain

$$\begin{aligned}
\mathbb{E}\Big[\|\rho_\beta^{(N)} - \rho_\beta\|^2_{H^{-s}(\mathbb{T}^d)}\Big] &= \mathbb{E}\left[\sum_{j=1}^\infty \Big|\langle \rho_\beta^{(N)} - \rho_\beta, \phi_j\rangle\Big|^2\right] \\
&= \sum_{j=1}^\infty \mathbb{E}\left[\left|\frac{1}{N}\sum_{i=1}^N \phi_j(X_i) - \int \phi_j \rho_\beta \, \mathrm{d}x\right|^2\right] \\
&= \frac{1}{N}\sum_{j=1}^\infty \int \phi_j^2 \rho_\beta \, \mathrm{d}x - \frac{1}{N}\left(\int \phi_j \rho_\beta \, \mathrm{d}x\right)^2 \\
&\le \frac{\|\rho_\beta\|_{L^\infty(\mathbb{T}^d)}}{N}\sum_{j=1}^\infty \|\phi_j\|^2_{L^2(\mathbb{T}^d)}.
\end{aligned}$$

Noticing that $s > d/2$ implies that the embedding $H_0^s(\mathbb{T}^d) \to L^2(\mathbb{T}^d)$ is Hilbert-Schmidt, we can use the boundedness of $\rho_\beta$ to obtain

$$\mathbb{E}\Big[\|\rho_\beta^{(N)} - \rho_\beta\|^2_{H^{-s}(\mathbb{T}^d)}\Big] \leq \frac{C}{N}\,. \tag{12.4}$$

Combining (12.2) with (12.3) and (12.4), we obtain the desired conclusion. □

REMARK 12.3. An alternative proof of inequality (12.4) can be found in [35, Theorem 5.1].

We now consider the implications of having uniform control of the log Sobolev constant the fluctuations of the stationary solutions to (2.1), i.e. solutions of (2.1) started at stationarity. To this end, we consider the empirical measure process $t \mapsto \mu^{(N)}(t)$ defined by

$$\mu^{(N)}(t) := \frac{1}{N}\sum_{i=1}^N \delta_{X_t^i},$$

where $X_t = (X_t^1, \dots, X_t^N)$ is the solution to (2.1) started from the unique invariant Gibbs measure $M_N$. Our goal is to analyze the corresponding fluctuation process $t \mapsto \eta^N(t)$ defined by

$$\eta^N(t) := \sqrt{N}(\mu^{(N)}(t) - \rho_\beta),$$

As a direct consequence of the estimates from the previous lemma and using the fact that $\liminf_{N\to\infty} \lambda_{LS}^N > 0$, we have the uniform bound

$$\sup_{N,t} \mathbb{E}\Big[\|\eta^N(t)\|^2_{H^{-s}(\mathbb{T}^d)}\Big] < \infty \qquad \text{for any } s > d/2 + 1. \tag{12.5}$$

In the sequel, we will use the above estimate together with the classical martingale method, c.f. [17, Chapter 8], to establish convergence in law as $N \to \infty$ of $\eta^N$ to the stationary solution $\eta$ of the following linear SPDE

$$\partial_t \eta = \mathcal{L}_{\rho_\beta}\eta + \nabla\cdot(\sqrt{\rho_\beta}\xi), \tag{12.6}$$

where $\xi$ is a mean-zero space-time white noise on $\mathbb{R}_+ \times \mathbb{T}^d$ and $\mathcal{L}_{\rho_\beta}$ is the linearisation of the McKean-Vlasov operator (2.5) around the unique invariant measure $\rho_\beta$ defined by

$$\mathcal{L}_{\rho_\beta}\psi = \beta^{-1}\Delta\psi + \nabla\cdot(\nabla W \star \psi\,\rho_\beta) + \nabla\cdot(\nabla W \star \rho_\beta\,\psi) + \nabla\cdot(\nabla V\,\psi).$$

The SPDE (12.6) can be solved using classical methods, the three typical notions of solution being the mild, weak, and martingale formulations. As is typical, the martingale formulation is the most convenient for identifying the limiting law of a tight subsequence of $\eta^N$, while the mild formulation provides a clearer picture of the uniqueness in law and hence the convergence in law of the full sequence.

Denote $U$ as the closed mean-zero subspace of $L^2(\mathbb{T}^d;\mathbb{T}^d)$ and $H := H^{-s}(\mathbb{T}^d)$ for $s > \frac{d}{2}$. Let $(\mathcal{X}, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ be a stochastic basis, i.e a complete filtered probability space with a right continuous filtration endowed with an $\mathcal{F}_t$-adapted $U$-valued cylindrical Wiener process $\mathbf{W}$ and let $\eta_0$ be an $H$-valued $\mathcal{F}_0$-measurable random variable independent of $\mathbf{W}$. The mild formulation of (12.6) with initial condition $\eta_0$ involves stochastic integration in Hilbert spaces, c.f. [17, Chapter 4], which we quickly review for our specific case below. Let $\mathbf{L}_0^2(U;H)$ denote the Hilbert–Schmidt operators from $U$ to $H$ equipped with the standard

Hilbert–Schmidt norm $\|\cdot\|_{\mathbf{L}_2^0(U;H)}$. Given $T>0$ and $\Phi\in L^2([0,T];\mathbf{L}_0^2(U;H))$, the stochastic integral $[0,T]\ni t\mapsto\int_0^t\Phi(s)d\mathbf{W}(s)$ is well defined as a continuous $\mathcal{F}_t$-martingale with trajectories in $C([0,T];H)$. We remind the reader that the relevance of $\mathbf{L}_0^2(U;H)$ in this context is that the Itô isometry takes the following form

$$\mathbb{E}\left[\left\|\int_0^t\Phi(s)d\mathbf{W}(s)\right\|_H^2\right]=\int_0^t\|\Phi(s)\|_{\mathbf{L}_0^2(U;H)}^2\,\mathrm{d}s.$$

The mild solution to $t\mapsto\eta(t)\in H$ to (12.6) with initial condition $\eta_0$ is then given by the following stochastic convolution:

$$\eta^\infty(t):=e^{t\mathcal{L}_{\rho_\beta}}\eta_0+\int_0^t e^{(t-s)\mathcal{L}_{\rho_\beta}}\nabla\cdot\big(\sqrt{\rho_\beta}d\mathbf{W}(s)\big). \tag{12.7}$$

LEMMA 12.4. *The mild solution* (12.7) *is well-defined as a stochastic process with trajectories in* $C([0,T];H)$.

PROOF. In light of our remarks in the preceding paragraph, it suffices to show that $s\mapsto\Phi(s)$ defined by $U\ni u\mapsto e^{(t-s)\mathcal{L}_{\rho_\beta}}\nabla\cdot(\sqrt{\rho_\beta}u)\in H$ belongs to $L^2([0,T];\mathbf{L}_0^2(U;H))$. Given an orthornormal basis $\{e_k\}_{k=1}^\infty$ of $U$, using the definition of the Hilbert–Schmidt norm and integrating by parts leads to

$$\begin{aligned}\int_0^T\|\Phi(s)\|_{\mathbf{L}_0^2(U;H)}^2\,\mathrm{d}s&=\sum_{k=1}^\infty\int_0^T\|e^{(t-s)\mathcal{L}_{\rho_\beta}}\nabla\cdot\big(\sqrt{\rho_\beta}e_k\big)\|_H^2\,\mathrm{d}s\\&=\sum_{j,k=1}^\infty\int_0^T\langle e_k,\sqrt{\rho_\beta}\nabla e^{(t-s)\mathcal{L}_{\rho_\beta}^*}\phi_j\rangle_U^2\,\mathrm{d}s\\&=\sum_{j=1}^\infty\int_0^T\|\sqrt{\rho_\beta}\nabla e^{(t-s)\mathcal{L}_{\rho_\beta}^*}\phi_j\|_U^2\,\mathrm{d}s,\end{aligned} \tag{12.8}$$

where in the last step we used Parseval's identity in $U$. Finally, we note that the coercivity hypothesis for $\mathcal{L}_{\rho_\beta}$ (see Lemma 12.8) implies that

$$\int_0^t\|\nabla e^{(t-s)\mathcal{L}_{\rho_\beta}^*}\phi_j\|_U^2\,\mathrm{d}s\le C\|\phi_j\|_{L^2(\mathbb{T}^d)}^2.$$

Combining this with (12.8) and using the fact that $\rho_\beta\in L^\infty(\mathbb{T}^d)$, yields

$$\int_0^T\|\Phi(s)\|_{\mathbf{L}_0^2(U;H)}^2\,\mathrm{d}s\le C\|\rho_\beta\|_{L^\infty(\mathbb{T}^d)}\sum_{j=1}^\infty\|\phi_j\|_{L^2(\mathbb{T}^d)}^2. \tag{12.9}$$

Since $s>\frac{d}{2}$, the embedding of $H_0^s(\mathbb{T}^d)$ into $L^2(\mathbb{T}^d)$ is Hilbert–Schmidt. Thus, the above series converges, completing the proof of the lemma. □

REMARK 12.5. We note that the representation (12.7) immediately implies that solutions to (12.6) are unique in law, i.e. (12.6) satisfies weak uniqueness. That is to say, given two different stochastic bases $(\mathcal{X},\mathcal{F},(\mathcal{F}_t)_{t\ge0},\mathbb{P},\mathbf{W})$ and $(\tilde{\mathcal{X}},\tilde{\mathcal{F}},(\tilde{\mathcal{F}}_t)_{t\ge0},\tilde{\mathbb{P}},\widetilde{\mathbf{W}})$ defining solutions $\eta$ and $\tilde\eta$ to (12.6) on their respective probability spaces through the formula (12.7), the laws of $\eta$ and $\tilde\eta$ agree on $C([0,T];H)$ for any $T>0$, as long as $\eta_0$ and $\tilde\eta_0$ are equal in law on $H$.

For our purposes, it is easier to work with the martingale formulation of (12.6), which is in turn motivated by the weak formulation of (12.6). Hence, we note in passing that the mild solution (12.7) has the property that for each $t \in [0,T]$ the following equality holds in $H^{-s}(\mathbb{T}^d)$, for $s > \frac{d+2}{2}$,

$$\eta^\infty(t) = \eta_0 + \int_0^t \mathcal{L}_{\rho_\beta} \eta^\infty(s)\ \mathrm{d}s + \int_0^t \nabla \cdot (\sqrt{\rho_\beta} d\mathbf{W}_s). \tag{12.10}$$

LEMMA 12.6. *Let $(\mathcal{X}, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ be a filtered probability space. Assume that $\eta$ is a continuous-time $\mathcal{F}_t$-adapted $H$-valued stochastic process and define $t \mapsto M(t)$ by*

$$M(t) := \eta(t) - \eta_0 - \int_0^t \mathcal{L}_{\rho_\beta} \eta(s)\ \mathrm{d}s. \tag{12.11}$$

*Assume that the following two conditions hold:*

- *For all $\varphi \in C^\infty(\mathbb{T}^d)$ it holds that*

  $$t \mapsto \langle M(t), \varphi \rangle \tag{12.12}$$

  *is an $\mathcal{F}_t$-martingale.*
- *For all $\varphi, \psi \in C^\infty(\mathbb{T}^d)$ it holds that*

  $$t \mapsto \langle M(t), \varphi \rangle \langle M(t), \psi \rangle - t \int_{\mathbb{T}^d} \nabla\phi \cdot \nabla\psi \rho_\beta\ \mathrm{d}x \tag{12.13}$$

  *is an $\mathcal{F}_t$ martingale.*

*Then, $\eta$ is equal in law on $C([0,T]; H)$ to the mild solution (12.7).*

PROOF. By [17, Theorem 8.2], it follows that on a suitable extension of the probability space, $\eta$ is a weak solution in the sense of (12.10). By [17, Theorem 5.4], the weak and mild solutions coincide on that probability space, so equality in law follows from Remark 12.5. □

Using the above lemma, we are now finally in a position to prove the convergence of the fluctuations.

THEOREM 12.7. *Assume that $\liminf_{N\to\infty} \lambda_{LS}^N > 0$ and that $V$ and $W$ are smooth. Then, for any $m > d/2 + 3$ the fluctuation process $\eta^N$ converges in law on $C([0,T]; H^{-m}(\mathbb{T}^d))$ to the unique stationary mild solution $\eta^\infty$ of the SPDE (12.6).*

PROOF. The proof has four steps. In Step 1, we apply Itô's formula to show that $\eta^N$ satisfies (12.14), an approximate version of the weak formulation (12.10). In Step 2, we combine Step 1 with the uniform bound (12.5) to show that the laws of $(\eta^N)_{N\in\mathbb{N}}$ on $C([0,T]; H^{-m}(\mathbb{T}^d))$ are uniformly tight for $m$ large enough. In Step 3, we pass to the limit in the martingale problem and verify the assumptions of Lemma (12.6) to identify the limit along any tight subsequence. In Step 4, we conclude the uniqueness of the limit and hence the proof of the theorem.

STEP 1. In this step, we show that for all $\varphi \in C^\infty(\mathbb{T}^d)$ it holds

$$\mathrm{d}\langle \eta^N, \varphi \rangle = \langle \eta^N, \mathcal{L}^*_{\rho_\beta} \varphi \rangle\, \mathrm{d}t + \langle R_N, \varphi \rangle\, \mathrm{d}t + \sqrt{2}(\beta N)^{-\frac{1}{2}} \sum_{i=1}^N \nabla\varphi(X_t^i) \cdot dB_t^i\,, \tag{12.14}$$

with $X_t^i$ the solution to (2.1) and where $R_N$ is defined as

$$R_N := N^{-\frac{1}{2}} \nabla \cdot (\eta^N \nabla_1 W \star \eta^N) .$$

Indeed, Itô's formula gives

$$\begin{aligned} \mathrm{d}\varphi(X_t^i) =& \big(\beta^{-1}\Delta - \nabla V \cdot \nabla\big)\varphi(X_t^i)\,\mathrm{d}t - \frac{1}{N}\sum_{j=1}^N \nabla_1 W(X_t^i, X_t^j)\cdot\nabla\varphi(X_t^i)\,\mathrm{d}t \\ &+ \sqrt{2}\beta^{-1}\nabla\varphi(X_t^i)\cdot dB_t^i . \end{aligned}$$

We now sum over $i = 1, \dots, N$, divide by $N$, and use the identity

$$\begin{aligned} \frac{1}{N^2}\sum_{i,j=1}^N \nabla_1 W(X_t^i, X_t^j)\cdot\nabla\varphi(X_t^i) =& \frac{1}{N}\sum_{i=1}^N \nabla_1 W * \mu^{(N)}(X_t^i)\cdot\nabla\varphi(X_t^i) \\ =& \langle \mu^{(N)}, \nabla_1 W \star \mu^{(N)} \cdot \nabla\varphi\rangle , \end{aligned}$$

to obtain

$$\mathrm{d}\langle\mu^{(N)},\varphi\rangle = \langle\mu^{(N)}, \beta^{-1}\Delta\varphi - (\nabla V + \nabla_1 W \star \mu^{(N)})\cdot\nabla\varphi\rangle\,\mathrm{d}t + \sqrt{2}\beta^{-\frac{1}{2}}\frac{1}{N}\sum_{i=1}^N \nabla\varphi(X_t^i)\cdot dB_t^i$$

Next, we insert the identity $\mu^{(N)} = \rho_\beta + N^{-\frac{1}{2}}\eta^N$, to deduce

$$\begin{aligned} \mathrm{d}\langle\eta^N,\varphi\rangle =& N^{1/2}\underbrace{\langle\rho_\beta, \beta^{-1}\Delta\varphi - (\nabla V + \nabla_1 W\star\rho_\beta)\cdot\nabla\varphi\rangle}_{=0}\,\mathrm{d}t \\ &+ \underbrace{\langle\eta^N, \beta^{-1}\Delta\varphi - (\nabla V + \nabla_1 W\star\rho_\beta)\cdot\nabla\varphi\rangle + \langle\rho_\beta, \nabla_1 W\star\eta^N\cdot\nabla\varphi\rangle}_{=\langle\eta^N, \mathcal{L}^*_{\rho_\beta}\varphi\rangle}\,\mathrm{d}t \\ &+ \underbrace{N^{-1/2}\langle\eta^N, \nabla_1 W\star\eta^N\cdot\nabla\varphi\rangle}_{\langle R^N,\varphi\rangle} + \frac{\sqrt{2}}{\beta N^{\frac{1}{2}}}\sum_{i=1}^N \nabla\varphi(X_t^i)\cdot dB_t^i \end{aligned}$$

The first identity follows from the fact that $\rho_\beta$ is a steady state and the second follows from integration by parts and Fubini's theorem (using the symmetry of $W$).

STEP 2. In this step, we will show that the laws of $(\eta^N)_{N\in\mathbb{N}}$ on $C([0,T]; H^{-m}(\mathbb{T}^d))$ for $m > m_0 := d/2+3$ are uniformly tight. To this end, we define a decomposition of $t \mapsto \eta^N(t)$ via the equality $\eta^N(t) = Y_N(t) + M_N(t)$, where

$$Y_N(t) := \eta^N(0) + \int_0^t \big(\mathcal{L}_{\rho_\beta}\eta^N(r) + R^N(r)\big)\,\mathrm{d}r.$$

We claim that for $m > m_0$ and all $p \geq 1$ there exist constants $C, C_p > 0$ such that

$$\mathbb{E}\Big[\|Y^N(t_1) - Y^N(t_2)\|^2_{H^{-m}(\mathbb{T}^d)}\Big] \leq C|t_1 - t_2|^2 \tag{12.15}$$

$$\mathbb{E}\Big[\|M^N(t_1) - M^N(t_2)\|^p_{H^{-m}(\mathbb{T}^d)}\Big] \leq C_p|t_1 - t_2|^{\frac{p}{2}}. \tag{12.16}$$

To obtain (12.15), first observe that at each fixed time we have

$$\begin{aligned} \|R_N\|_{H^{-m}(\mathbb{T}^d)} =& \|\nabla\cdot(\eta^N\nabla_1 W\star(\mu^{(N)} - \rho_\beta))\|_{H^{-m}(\mathbb{T}^d)} \\ \leq& \|\eta^N\nabla_1 W\star(\mu^{(N)} - \rho_\beta)\|_{H^{1-m}(\mathbb{T}^d)} \\ \leq& C_s\|\eta^N\|_{H^{1-m}(\mathbb{T}^d)}\|\nabla W\star(\mu^{(N)} - \rho_\beta)\|_{W^{m-1,\infty}(\mathbb{T}^d)} \\ \leq& C_s\|W\|_{W^{m,\infty}(\mathbb{T}^d)}\|\eta^N\|_{H^{1-m}(\mathbb{T}^d)}. \end{aligned} \tag{12.17}$$

Furthermore, note that the regularity of $V$, $W$, and $\rho_\beta$ implies the operator $\mathcal{L}_{\rho_\beta}$ is bounded from $H^{2-m}(\mathbb{T}^d)$ to $H^{-m}(\mathbb{T}^d)$. Therefore, integrating in time, taking the second moment, and applying the Cauchy–Schwartz inequality we find

$$\begin{aligned}\mathbb{E}\Big[\|Y^N(t_1)-Y^N(t_2)\|^2_{H^{-m}(\mathbb{T}^d)}\Big] \leq& \mathbb{E}\left(\int_{t_1}^{t_2} \|\mathcal{L}_{\rho_\beta}\eta^N(r)\|_{H^{-m}(\mathbb{T}^d)} + \mathbb{E}\|R_N\|_{H^{-m}(\mathbb{T}^d)}\,\mathrm{d}r\right)^2\\ \leq& C|t_2-t_1|\int_{t_1}^{t_2} \mathbb{E}\|\eta^N\|^2_{H^{2-s}(\mathbb{T}^d)} + \mathbb{E}\|\eta^N\|^2_{H^{1-m}(\mathbb{T}^d)}\,\mathrm{d}r\\ \leq& C|t_2-t_1|^2\mathbb{E}\|\eta^N(0)\|^2_{H^{2-m}(\mathbb{T}^d)},\end{aligned}$$

where we used the stationarity of $\eta^N$ in the last step. The inequality (12.15) now follows immediately from the bound (12.5), since our choice of $m$ implies $m-2>1+\frac{d}{2}$. We now turn our attention to the estimate (12.16). Note that for any smooth $\varphi$ it holds that

$$\begin{aligned}\mathbb{E}\Big[\big|\langle M^N(t)-M^N(s),\varphi\rangle\big|^p\Big] \leq& C\mathbb{E}\left(\frac{1}{N}\sum_{i=1}^N\int_s^t |\nabla\varphi(X^i_t)|^2\,\mathrm{d}t\right)^{\frac{p}{2}}\\ \leq& C|t-s|^{\frac{p}{2}}\|\nabla\varphi\|^2_{L^\infty(\mathbb{T}^d)}.\end{aligned}\tag{12.18}$$

We take $\varphi$ to be elements of a basis of $H^m_0(\mathbb{T}^d)$, the inequality (12.16) holds as long as

$$\sum_{j\in\mathbb{N}}\|\nabla\varphi_j\|^2_{L^\infty(\mathbb{T}^d)}<\infty.$$

The above estimate follows when we take $m>d/2+1$. Finally, we note that by an argument entirely analogous to the one showing (12.15) and (12.16), we can also show

$$\sup_{N\in\mathbb{N}}\mathbb{E}\left[\sup_{t\in[0,T]}\|\eta^N(t)\|^2_{H^{-m}(\mathbb{T}^d)}\right]<\infty,\tag{12.19}$$

by using that $M^N(0)=0$ and $Y^N(0)=\eta^N(0)$ together with (12.5). Combining (12.15), (12.16), and (12.19) we obtain the tightness of the laws of $(\eta^N)_{N\in\mathbb{N}}$ as a consequence of [27, Theorem 2.2] and Chebyshev's inequality. Specifically, we use the embedding of $W^{1-\varepsilon,2}([0,T];H^{-m+\varepsilon}(\mathbb{T}^d))+W^{\frac{1}{2},p}([0,T];H^{-m+\epsilon}(\mathbb{T}^d))$ into $C([0,T];H^{-m}(\mathbb{T}^d))$ for $p>2$ and $\varepsilon>0$ sufficiently small.

Step 3. In light of Step 2 and the Skorokhod representation theorem, passing to a subsequence (which we do not relabel) we can find a new probability space $(\tilde{\mathcal{X}},\tilde{\mathcal{F}},\tilde{\mathbb{P}})$, a new sequence $(\tilde\eta^N)_{N\in\mathbb{N}}$, and a limiting random variable $\eta$ such that $\tilde\eta^N$ is equal in law to $\eta^N$ for all $N\in\mathbb{N}$ and converges $\tilde{\mathbb{P}}$-a.s to $\eta$ in $C([0,T];H^{-m}(\mathbb{T}^d))$. In this step, we claim that (12.12) and (12.13) hold true.

To this end, for each $t>0$, we denote by $r_t$ the restriction operator from $C([0,T];H^{-m}(\mathbb{T}^d))$ to $C([0,t];H^{-m}(\mathbb{T}^d))$ and define a filtration $(\tilde{\mathcal{F}}_t)_{t\geq 0}$ by letting $\tilde{\mathcal{F}}_t=\sigma(r_t\eta)$ for $t>0$, i.e. the sigma algebra generated by $r_t\eta$. Recalling the definition (12.11) of $t\mapsto M(t)$, we will show that that for all times $s<t$, functions $\varphi,\psi\in C^\infty(\mathbb{T}^d)$, and bounded, continuous functions $\Gamma: C([0,s];H^{-m}(\mathbb{T}^d))\mapsto\mathbb{R}$ it holds that

$$\mathbb{E}[\Gamma(r_s\eta)\langle M(t)-M(s),\varphi\rangle]=0.\tag{12.20}$$

$$\begin{aligned}\mathbb{E}\big[\Gamma(r_s\eta)\big(&\langle M(t),\varphi\rangle\langle M(t),\psi\rangle-\langle M(s),\varphi\rangle\langle M(s),\psi\rangle\\ &-(t-s)\langle\rho_\beta\nabla\varphi,\nabla\psi\rangle\big)\big]=0.\end{aligned}\tag{12.21}$$

By the definition of conditional expectation and Egorov's theorem, (12.20) and (12.21) imply (12.12) and (12.13) hold true with respect to the filtration $(\tilde{\mathcal{F}}_t)_{t\geq 0}$. To prove (12.20) and (12.21), define $t \mapsto \tilde{M}^N(t)$ as in Step 2 but with $\tilde{\eta}_N$ in place of $\eta^N$.

Since $\tilde{\eta}^N$ and $\eta^N$ are equal in law, (12.14) implies that

$$\mathbb{E}\big[\Gamma(r_s\tilde{\eta}_N)\langle \tilde{M}^N(t) - \tilde{M}^N(s), \varphi\rangle\big] = 0. \tag{12.22}$$

$$\begin{aligned}\mathbb{E}\big[\Gamma(r_s\tilde{\eta}_N)\big(&\langle \tilde{M}_N(t),\varphi\rangle\langle \tilde{M}_N(t),\psi\rangle - \langle \tilde{M}_N(s),\varphi\rangle\langle \tilde{M}_N(s),\psi\rangle \\ &- (t-s)\langle \rho_\beta \nabla\varphi, \nabla\psi\rangle\big)\big] = 0.\end{aligned} \tag{12.23}$$

By a calculation similar to (12.17), it follows that for each $t \leq T$

$$\left\| \int_0^t \tilde{R}_N(s)ds \right\|_{H^{-(m+1)}(\mathbb{T}^d)} \leq CN^{-1/2}\|\tilde{\eta}_N\|^2_{C([0,T];H^{-m}(\mathbb{T}^d))},$$

which converges to zero $\tilde{P}$-a.s. As a consequence, we obtain $\langle \tilde{M}^N(t),\varphi\rangle$ converges $\tilde{P}$-a.s. to $\langle M(t),\varphi\rangle$ as a consequence of the a.s. convergence of $\tilde{\eta}^N$ to $\eta^\infty$. In addition, for all $t>0$, the sequence $\langle \tilde{M}^N(t),\varphi\rangle$ is uniformly bounded in $L^p(\tilde{\mathcal{X}})$ for all $p \geq 1$ as a consequence of the equality in law of $\tilde{\eta}^N$ and $\eta^N$ and the estimate (12.18). Using the Vitali convergence theorem, we may pass to the limit in (12.22) and (12.23) to obtain (12.20) and (12.21) as desired.

STEP 4. In light of Step 3 and Lemma 12.6, every subsequence of $\eta^N$ has a further subsequence which converges to a stationary mild solution to (12.6) on some probability space. Note that $\eta$ inherits stationarity from $\tilde{\eta}^N$ in the $N\to\infty$ limit. Hence, it suffices to show that all limit points induce the same law on $C([0,T];H^{-m}(\mathbb{T}^d))$. In light of Remark 12.5, the problem further reduces to showing that the initial distributions are the same. However, if $\eta$ satisfies (12.7) and is stationary, we can explicitly check that its law is a Gaussian on $H^{-m}(\mathbb{T}^d)$ for all $t\geq 0$. This follows from [34, Theorem 5.22 and Proposition 5.23], (12.9), and the fact that for all $f \in H^{-m}(\mathbb{T}^d)$, we have

$$\lim_{t\to\infty} \|e^{t\mathcal{L}_{\rho_\beta}} f\|_{H^{-m}} = 0\,.$$

The fact that this holds true follows from the coercivity assumption and the fact that the semigroup $e^{t\mathcal{L}_{\rho_\beta}}$ is smoothing, i.e. it maps $H^{-m}(\mathbb{T}^d)$ to $L^2(\mathbb{T}^d)$ for all $t>0$. □

We finish this section with the implication of the coercivity property, that is used in the proof to show the Hilbert-Schmidt property of the appropriate operators.

LEMMA 12.8. *Assume that $\mathcal{L}_{\rho_\beta}$ satisfies*

$$-\langle \mathcal{L}_{\rho_\beta}\phi,\phi\rangle \geq c\|\nabla\phi\|^2_{L^2(\mathbb{T}^d)}$$

*for some $c>0$, then*

$$\int_0^\infty \|\nabla e^{t\mathcal{L}^*_{\rho_\beta}}\phi\|^2_{L^2(\mathbb{T}^d)}\,\mathrm{d}t \leq \frac{\|\phi\|^2_{L^2}}{2c}.$$

PROOF. We use the short hand notation $\phi_t = e^{t\mathcal{L}^*_{\rho_\beta}}\phi$. We then have

$$\partial_t\phi_t - \mathcal{L}^*_{\rho_\beta}\phi_t = 0\,.$$

Multiplying by $\phi_t$ and integrating, we obtain the identity

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|\phi_t\|^2_{L^2(\mathbb{T}^d)} - \langle \mathcal{L}_{\rho_\beta}\phi_t,\phi_t\rangle = 0.$$

Applying the coercivity bound, we obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}\|\phi_t\|^2_{L^2(\mathbb{T}^d)} + 2c\|\nabla\phi_t\|^2_{L^2} \leq 0.$$

Integrating in time we obtain

$$2c\int_0^\infty \|\nabla\phi_t\|^2_{L^2(\mathbb{T}^d)}\ \mathrm{d}t \leq \|\phi_0\|^2_{L^2} = \|\phi\|^2_{L^2(\mathbb{T}^d)}\,,$$

which is the desired estimate. □

**Acknowledgements.** The authors would like to thank Martin Hairer and Luigia Ripani for useful discussions during the course of this work. G.A.P. was partially supported by the EPSRC through the grant number EP/P031587/1 and by JPMorgan Chase & Co under a J.P. Morgan A.I. Research Award 2019. M.G.D. was partially supported by CNPq-Brazil (#308800/2019-2) and Instituto Serrapilheira.

## REFERENCES

[1] AMBROSIO, L., GIGLI, N. and SAVARÉ, G. (2008). *Gradient flows in metric spaces and in the space of probability measures*, second ed. *Lectures in Mathematics ETH Zürich*. Birkhäuser Verlag, Basel. MR2401600

[2] BACKHOFF, J., CONFORTI, G., GENTIL, I. and LÉONARD, C. (2020). The mean field Schrödinger problem: ergodic behavior, entropy estimates and functional inequalities. *Probab. Theory Related Fields* **178** 475–530. MR4146543

[3] BAKRY, D., GENTIL, I. and LEDOUX, M. (2013). *Analysis and geometry of Markov diffusion operators* **348**. Springer Science & Business Media.

[4] BASHIRI, K. and MENZ, G. (2021). Metastability in a continuous mean-field model at low temperature and strong interaction. *Stochastic Process. Appl.* **134** 132–173. MR4200855

[5] BAUERSCHMIDT, R. and BODINEAU, T. (2019). A very simple proof of the LSI for high temperature spin systems. *J. Funct. Anal.* **276** 2582–2588. MR3926125

[6] BECKER, S. and MENEGAKI, A. (2020). Spectral gap in mean-field $\mathcal{O}(n)$-model. *Comm. Math. Phys.* **380** 1361–1400. MR4179730

[7] BENAMOU, J.-D. and BRENIER, Y. (2000). A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. *Numerische Mathematik* **84** 375–393.

[8] BERTINI, L., GIACOMIN, G. and PAKDAMAN, K. (2010). Dynamical aspects of mean field plane rotators and the Kuramoto model. *J. Stat. Phys.* **138** 270–290. MR2594897

[9] BRESCH, D., JABIN, P.-E. and WANG, Z. (2019). On mean-field limits and quantitative estimates with a large class of singular kernels: application to the Patlak-Keller-Segel model. *C. R. Math. Acad. Sci. Paris* **357** 708–720. MR4018082

[10] CARRILLO, J. A., DELGADINO, M. G. and PAVLIOTIS, G. A. (2020). A $\lambda$-convexity based proof for the propagation of chaos for weakly interacting stochastic particles. *Journal of Functional Analysis* **279** 108734.

[11] CARRILLO, J. A., GVALANI, R. S., PAVLIOTIS, G. A. and SCHLICHTING, A. (2020). Long-time behaviour and phase transitions for the McKean-Vlasov equation on the torus. *Arch. Ration. Mech. Anal.* **235** 635–690. MR4062483

[12] CARRILLO, J. A., GVALANI, R. S. and WU, J. (2020). An invariance principle for gradient flows in the space of probability measures. *arXiv e-prints* arXiv:2010.00424.

[13] CARRILLO, J. A., MCCANN, R. J. and VILLANI, C. (2003). Kinetic equilibration rates for granular media and related equations: entropy dissipation and mass transportation estimates. *Rev. Mat. Iberoamericana* **19** 971–1018. MR2053570

[14] CHAVANIS, P. H. (2014). The Brownian mean field model. *Eur. Phys. J. B* **87** Art. 120, 33. MR3210154

[15] CHAYES, L. and PANFEROV, V. (2010). The McKean–Vlasov equation in finite volume. *Journal of Statistical Physics* **138** 351–380.

[16] CONFORTI, G. and RIPANI, L. (2020). Around the entropic Talagrand inequality. *Bernoulli* **26** 1431–1452. MR4058373

[17] DA PRATO, G., IANNELLI, M. and TUBARO, L. (1982). An existence result for a linear abstract stochastic equation in Hilbert spaces. *Rendiconti del Seminario Matematico della Università di Padova* **67** 171–180. MR682709

[18] DAWSON, D. A. (1983). Critical dynamics and fluctuations for a mean-field model of cooperative behavior. *Journal of Statistical Physics* **31** 29–85.

[19] DE FINETTI, B. (1937). La prévision : ses lois logiques, ses sources subjectives. *Ann. Inst. H. Poincaré* **7** 1–68. MR1508036

[20] DELGADINO, M. G., GVALANI, R. S. and PAVLIOTIS, G. A. (2020). On the diffusive-mean field limit for weakly interacting diffusions exhibiting phase transitions. *arXiv preprint arXiv:2001.03920*.

[21] DIACONIS, P. and FREEDMAN, D. (1980). Finite exchangeable sequences. *The Annals of Probability* 745–764.

[22] DUPAIGNE, L., GENTIL, I. and ZUGMEYER, S. (2020). Sobolev's inequality under a curvature-dimension condition. *arXiv preprint arXiv:2011.07840*.

[23] DURMUS, A., EBERLE, A., GUILLIN, A. and ZIMMER, R. (2020). An elementary approach to uniform in time propagation of chaos. *Proc. Amer. Math. Soc.* **148** 5387–5398. MR4163850

[24] EBERLE, A. (2011). Reflection coupling and Wasserstein contractivity without convexity. *C. R. Math. Acad. Sci. Paris* **349** 1101–1104. MR2843007

[25] EBERLE, A. (2016). Reflection couplings and contraction rates for diffusions. *Probab. Theory Related Fields* **166** 851–886. MR3568041

[26] FERNANDEZ, B. and MÉLÉARD, S. (1997). A Hilbertian approach for fluctuations on the McKean-Vlasov model. *Stochastic Process. Appl.* **71** 33–53. MR1480638

[27] FLANDOLI, F. and GATAREK, D. (1995). Martingale and stationary solutions for stochastic Navier-Stokes equations. *Probability Theory and Related Fields* **102** 367–391.

[28] GENTIL, I., LÉONARD, C. and RIPANI, L. (2020). Dynamical aspects of the generalized Schrödinger problem via Otto calculus—a heuristic point of view. *Rev. Mat. Iberoam.* **36** 1071–1112. MR4130828

[29] GENTIL, I., LÉONARD, C., RIPANI, L. and TAMANINI, L. (2020). An entropic interpolation proof of the HWI inequality. *Stochastic Process. Appl.* **130** 907–923. MR4046525

[30] GEORGII, H. O. (2011). *Gibbs measures and phase transitions*, second ed. *De Gruyter Studies in Mathematics* **9**. Walter de Gruyter & Co., Berlin. MR2807681

[31] GRUNEWALD, N., OTTO, F., VILLANI, C. and WESTDICKENBERG, M. G. (2009). A two-scale approach to logarithmic Sobolev inequalities and the hydrodynamic limit. *Ann. Inst. Henri Poincaré Probab. Stat.* **45** 302–351. MR2521405

[32] GUILLIN, A., LIU, W., WU, L. and ZHANG, C. (2019). Uniform Poincaré and logarithmic Sobolev inequalities for mean field particles systems.

[33] GVALANI, R. S. and SCHLICHTING, A. (2020). Barriers of the McKean-Vlasov energy via a mountain pass theorem in the space of probability measures. *J. Funct. Anal.* **279** 108720, 34. MR4141490

[34] HAIRER, M. (2009). An Introduction to Stochastic PDEs. *arXiv e-prints* arXiv:0907.4178.

[35] HAURAY, M. and MISCHLER, S. (2014). On Kac's chaos and related problems. *J. Funct. Anal.* **266** 6055–6157. MR3188710

[36] HELFFER, B. (2002). *Semiclassical analysis, Witten Laplacians, and statistical mechanics*. *Series in Partial Differential Equations and Applications* **1**. World Scientific Publishing Co., Inc., River Edge, NJ. MR1936110

[37] HEWITT, E. and SAVAGE, L. J. (1955). Symmetric measures on Cartesian products. *Trans. Amer. Math. Soc.* **80** 470–501. MR76206

[38] JABIN, P.-E. and WANG, Z. (2018). Quantitative estimates of propagation of chaos for stochastic systems with $W^{-1,\infty}$ kernels. *Invent. Math.* **214** 523–591. MR3858403

[39] KIELHÖFER, H. (1988). A bifurcation theorem for potential operators. *J. Funct. Anal.* **77** 1–8. MR930387

[40] LEDOUX, M. (2001). Logarithmic Sobolev inequalities for unbounded spin systems revisited. In *Séminaire de Probabilités, XXXV*. *Lecture Notes in Math.* **1755** 167–194. Springer, Berlin. MR1837286

[41] LELIÈVRE, T. (2009). A general two-scale criteria for logarithmic Sobolev inequalities. *J. Funct. Anal.* **256** 2211–2221. MR2498763

[42] LORENZI, L. and BERTOLDI, M. (2006). *Analytical Methods for Markov Semigroups*. CRC Press, New York.

[43] LOTT, J. and VILLANI, C. (2009). Ricci curvature for metric-measure spaces via optimal transport. *Annals of Mathematics* 903–991.

[44] MALRIEU, F. et al. (2003). Convergence to equilibrium for granular media equations and their Euler schemes. *The Annals of Applied Probability* **13** 540–560.

[45] MARKOWICH, P. A. and VILLANI, C. (2000). On the trend to equilibrium for the Fokker-Planck equation: an interplay between physics and functional analysis. *Mat. Contemp.* **19** 1–29. MR1812873 (2002d:82058)

[46] MCCANN, R. J. (1997). A convexity principle for interacting gases. *Advances in mathematics* **128** 153–179.

[47] MESSER, J. and SPOHN, H. (1982). Statistical mechanics of the isothermal Lane-Emden equation. *J. Statist. Phys.* **29** 561–578. MR704588

[48] OTTO, F. (2001). The geometry of dissipative evolution equations: the porous medium equation. *Comm. Partial Differential Equations* **26** 101–174. MR1842429

[49] OTTO, F. and REZNIKOFF, M. G. (2007). A new criterion for the logarithmic Sobolev inequality and two applications. *J. Funct. Anal.* **243** 121–157. MR2291434

[50] PAVLIOTIS, G. A. (2014). *Stochastic processes and applications. Texts in Applied Mathematics* **60**. Springer, New York Diffusion processes, the Fokker-Planck and Langevin equations. MR3288096

[51] RIPANI, L. (2019). Convexity and regularity properties for entropic interpolations. *J. Funct. Anal.* **277** 368–391. MR3952157

[52] ROUGERIE, N. (2015). De finetti theorems, mean-field limits and bose-Einstein condensation. *arXiv e-prints* arXiv:1506.05263.

[53] SANTAMBROGIO, F. (2017). {Euclidean, metric, and Wasserstein} gradient flows: an overview. *Bulletin of Mathematical Sciences* **7** 87–154.

[54] SERFATY, S. (2020). Mean field limit for Coulomb-type flows. *Duke Math. J.* **169** 2887–2935. With an appendix by Mitia Duerinckx and Serfaty. MR4158670

[55] SHIGA, T. and TANAKA, H. (1985). Central limit theorem for a system of Markovian particles with mean field interactions. *Z. Wahrsch. Verw. Gebiete* **69** 439–459. MR787607

[56] SZNITMAN, A.-S. (1984). Nonlinear reflecting diffusion process, and the propagation of chaos and fluctuations associated. *Journal of functional analysis* **56** 311–336.

[57] SZNITMAN, A.-S. (1991). Topics in propagation of chaos. In *Ecole d'été de probabilités de Saint-Flour XIX—1989* 165–251. Springer.

[58] TALAGRAND, M. (1995). Concentration of measure and isoperimetric inequalities in product spaces. *Publications Mathématiques de l'Institut des Hautes Etudes Scientifiques* **81** 73–205.

[59] TANAKA, H. and HITSUDA, M. (1981). Central limit theorem for a simple diffusion model of interacting particles. *Hiroshima Math. J.* **11** 415–423. MR620546

[60] VARADHAN, S. R. S. (1991). Scaling limits for interacting diffusions. *Comm. Math. Phys.* **135** 313–353. MR1087387

[61] VILLANI, C. (2008). *Optimal transport: old and new* **338**. Springer Science & Business Media.

[62] WANG, Z., ZHAO, X. and ZHU, R. (2021). Gaussian fluctuations for interacting particle systems with singular kernels. *arXiv preprint arXiv:2105.13201*.

[63] YOSHIDA, N. (2003). Phase transition from the viewpoint of relaxation phenomena. *Rev. Math. Phys.* **15** 765–788. MR2018287