Calc. Var. (2017) 56:100
DOI 10.1007/s00526-017-1186-9

Calculus of Variations

Variational approach to coarse-graining of generalized gradient flows

Manh Hong Duong[1] · **Agnes Lamacz**[2] · **Mark A. Peletier**[3] · **Upanshu Sharma**[4]

Received: 12 July 2015 / Accepted: 4 May 2017 / Published online: 28 June 2017

Abstract In this paper we present a variational technique that handles coarse-graining and passing to a limit in a unified manner. The technique is based on a duality structure, which is present in many gradient flows and other variational evolutions, and which often arises from a large-deviations principle. It has three main features: (a) a natural interaction between the duality structure and the coarse-graining, (b) application to systems with non-dissipative effects, and (c) application to coarse-graining of approximate solutions which solve the equation only to some error. As examples, we use this technique to solve three limit problems, the overdamped limit of the Vlasov–Fokker–Planck equation and the small-noise limit of randomly perturbed Hamiltonian systems with one and with many degrees of freedom.

Mathematics Subject Classification 35K67 · 35B25 · 49S99 · 49J45 · 35K10 · 35K20 · 60F10 · 70F40 · 70G75 · 37L05 · 35Q99 · 60J60

Contents

Communicated by L. Ambrosio.

✉ Upanshu Sharma
upanshu.sharma@enpc.fr

1 Mathematics Institute, University of Warwick, Coventry, United Kingdom

2 Fakultät für Mathematik, Dortmund, Germany

3 Department of Mathematics and Computer Sciences and Institute for Complex Molecular Systems, Technische Universiteit Eindhoven, Eindhoven, The Netherlands

4 CERMICS, Ecole des Ponts ParisTech, Champs sur Marne, France

1 Introduction

Coarse-graining is the procedure of approximating a system by a simpler or lower-dimensional one, often in some limiting regime. It arises naturally in various fields such as thermodynamics, quantum mechanics, and molecular dynamics, just to name a few. Typically coarse-graining requires a separation of temporal and/or spatial scales, i.e. the presence of fast and slow variables. As the ratio of 'fast' to 'slow' increases, some form of averaging or homogenization should allow one to remove the fast scales, and obtain a limiting system that focuses on the slow ones.

Coarse-graining limits are by nature *singular limits*, since information is lost in the coarse-graining procedure; therefore rigorous proofs of such limits are always non-trivial. Although the literature abounds with cases that have been treated successfully, and some fields can even be called well-developed—singular limits in ODEs and homogenization theory, to name just two—many more cases seem out of reach, such as coarse-graining in materials [25], climate prediction [66], and complex systems [33,59].

All proofs of singular limits hinge on using certain *special structure* of the equations; well-known examples are compensated compactness [55,72], the theories of viscosity solutions [19] and entropy solutions [46,69], and the methods of periodic unfolding [16,17] and

two-scale convergence [5]. *Variational-evolution structure*, such as in the case of gradient flows and variational rate-independent systems, also facilitates limits [28,51,53,54,67,70,71].

In this paper we introduce and study such a structure, which arises from the theory of *large deviations* for stochastic processes. In recent years we have discovered that many gradient flows, and also many 'generalized' gradient systems, can be matched one-to-one to the large-deviation characterization of some stochastic process [2,3,24,26,27,52]. The large-deviation rate functional, in this connection, can be seen to *define* the generalized gradient system. This connection has many philosophical and practical implications, which are discussed in the references above.

We show how in such systems, described by a rate functional, 'passing to a limit' is facilitated by the duality structure that a rate function inherits from the large-deviation context, in a way that meshes particularly well with coarse-graining.

1.1 Variational approach—an outline

The systems that we consider in this paper are evolution equations in a space of measures. Typical examples are the forward Kolmogorov equations associated with stochastic processes, but also various nonlinear equations, as in one of the examples below.

Consider the family of evolution equations

$$\begin{aligned} \partial_t \rho^\varepsilon &= \mathcal{N}^\varepsilon \rho^\varepsilon, \\ \rho^\varepsilon|_{t=0} &= \rho_0^\varepsilon, \end{aligned} \tag{1}$$

where $\mathcal{N}^\varepsilon$ is a linear or nonlinear operator. The unknown ρ^ε is a time-dependent Borel measure on a state space $\mathcal{X}$, i.e. $\rho^\varepsilon : [0,T] \to \mathcal{M}(\mathcal{X})$. In the systems of this paper, (1) has a variational formulation characterized by a functional I^ε such that

$$I^\varepsilon \geq 0 \qquad \text{and} \qquad \rho^\varepsilon \text{ solves (1)} \iff I^\varepsilon(\rho^\varepsilon) = 0. \tag{2}$$

This variational formulation is closely related to the Brezis–Ekeland–Nayroles variational principle [10,41,57,71] and the integrated energy-dissipation identity for gradient flows [4]; see Sect. 5.

Our interest in this paper is the limit $\varepsilon \to 0$, and we wish to study the behaviour of the system in this limit. If we postpone the aspect of coarse-graining for the moment, this corresponds to studying the limit of ρ^ε as $\varepsilon \to 0$. Since ρ^ε is characterized by I^ε, establishing the limiting behaviour consists of answering two questions:

1. *Compactness* Do solutions of $I^\varepsilon(\rho^\varepsilon) = 0$ have useful compactness properties, allowing one to extract a subsequence that converges in a suitable topology, say ς?
2. *Liminf inequality* Is there a limit functional $I \geq 0$ such that

$$\rho^\varepsilon \xrightarrow{\varsigma} \rho \implies \liminf_{\varepsilon \to 0} I^\varepsilon(\rho^\varepsilon) \geq I(\rho)? \tag{3}$$

And if so, does one have

$$I(\rho) = 0 \iff \rho \text{ solves } \partial_t \rho = \mathcal{N}\rho,$$

for some operator $\mathcal{N}$?

A special aspect of the method of the present paper is that it also applies to *approximate* solutions. By this we mean that we are interested in sequences of time-dependent Borel measures ρ^ε such that $\sup_{\varepsilon>0} I^\varepsilon(\rho^\varepsilon) \leq C$ for some $C \geq 0$. The exact solutions are special

cases when $C = 0$. The main message of our approach is that all the results then follow from this uniform bound and assumptions on well-prepared initial data.

The compactness question will be answered by the first crucial property of the functionals I^ε, which is that they provide an *a priori* bound of the type

$$S^\varepsilon(\rho_t^\varepsilon) + \int_0^t R^\varepsilon(\rho_s^\varepsilon)\, ds \le S^\varepsilon(\rho_0^\varepsilon) + I^\varepsilon(\rho^\varepsilon), \tag{4}$$

where ρ_t^ε denotes the time slice at time t and S^ε and R^ε are functionals. In the examples of this paper S^ε is a free energy and R^ε a relative Fisher Information, but the structure is more general. This inequality is reminiscent of the energy-dissipation inequality in the gradient-flow setting. The uniform bound, by assumption, of the right-hand side of (4) implies that each term in the left-hand side of (4), i.e., the free energy at any time $t > 0$ and the integral of the Fisher information, is also bounded. This will be used to apply the Arzelà–Ascoli theorem to obtain certain compactness and 'local-equilibrium' properties. All this discussion will be made clear in each example in this paper.

The second crucial property of the functionals I^ε is that they satisfy a duality relation of the type

$$I^\varepsilon(\rho) = \sup_f \mathcal{J}^\varepsilon(\rho, f), \tag{5}$$

where the supremum is taken over a class of smooth functions f. It is well known how such duality structures give rise to good convergence properties such as (3), but the focus in this paper is on how this duality structure combines well with coarse-graining.

In this paper we define *coarse-graining* to be a shift to a reduced, lower dimensional description via a coarse-graining map $\xi : \mathcal{X} \to \mathcal{Y}$ which identifies relevant information and is typically highly non-injective. Note that ξ may depend on ε. A typical example of such a coarse-graining map is a 'reaction coordinate' in molecular dynamics. The coarse-grained equivalent of $\rho^\varepsilon : [0, T] \to \mathcal{M}(\mathcal{X})$ is the push-forward $\hat{\rho}^\varepsilon := \xi_\# \rho^\varepsilon : [0, T] \to \mathcal{M}(\mathcal{Y})$. If ρ^ε is the law of a stochastic process X^ε, then $\xi_\# \rho^\varepsilon$ is the law of the process $\xi(X^\varepsilon)$.

There might be several reasons to be interested in $\xi_\# \rho^\varepsilon$ rather than ρ^ε itself. The push-forward $\xi_\# \rho^\varepsilon$ obeys a dynamics with fewer degrees of freedom, since ξ is non-injective; this might allow for more efficient computation. Our first example (see Sect. 1.3), the overdamped limit in the Vlasov–Fokker–Planck equation, is an example of this. As a second reason, by removing certain degrees of freedom, some specific behaviour of ρ^ε might become clearer; this is the case with our second and third examples (Sect. 1.3), where the effect of ξ is to remove a rapid oscillation, leaving behind a slower diffusive movement. Whatever the reason, in this paper we assume that some ξ is given, and that we wish to study the limit of $\xi_\# \rho^\varepsilon$ as $\varepsilon \to 0$.

The core of the arguments of this paper, that leads to the characterization of the equation satisfied by the limit of $\xi_\# \rho^\varepsilon$, is captured by the following formal calculation:

$$\begin{aligned}
I^\varepsilon(\rho^\varepsilon) &= \sup_f \ \mathcal{J}^\varepsilon(\rho^\varepsilon, f) \\
&\overset{f=g\circ\xi}{\ge} \sup_g \ \mathcal{J}^\varepsilon(\rho^\varepsilon, g \circ \xi) \\
&\qquad\qquad \Big\downarrow \varepsilon \to 0 \\
&\qquad \sup_g \ \mathcal{J}(\rho, g \circ \xi) \\
&\overset{(*)}{=}: \sup_g \ \hat{\mathcal{J}}(\hat{\rho}, g) \quad \overset{(**)}{=}: \quad \hat{I}(\hat{\rho})
\end{aligned}$$

Let us go through the lines one by one. The first line is the duality characterization (5) of I^ε. The inequality in the second line is due to the reduction to a subset of special functions f, namely those of the form $f = g \circ \xi$. This is in fact an implementation of coarse-graining: in the supremum we decide to limit ourselves to observables of the form $g \circ \xi$ which only have access to the information provided by ξ. After this reduction we pass to the limit and show that $\mathcal{J}^\varepsilon(\rho^\varepsilon, g \circ \xi)$ converges to some $\mathcal{J}(\rho, g \circ \xi)$—at least for appropriately chosen coarse-graining maps.

In the final step $(*)$ one requires that the loss-of-information in passing from ρ to $\hat{\rho}$ is consistent with the loss-of-resolution in considering only functions $f = g \circ \xi$. This step requires a proof of *local equilibrium*, which describes how the behaviour of ρ that is *not* represented explicitly by the push-forward $\hat{\rho}$, can nonetheless be deduced from $\hat{\rho}$. This local-equilibrium property is at the core of various coarse-graining methods and is typically determined case by case.

We finally define $\hat{I}$ by duality in terms of $\hat{J}$ as in $(**)$. In a *successful* application of this method, the resulting functional $\hat{I}$ at the end has 'good' properties *despite* the loss-of-accuracy introduced by the restriction to functions of the form $g \circ \xi$, and this fact acts as a test of success. Such good properties should include, for instance, the property that $\hat{I} = 0$ has a unique solution in an appropriate sense.

Now let us explain the origin of the functionals I^ε.

1.2 Origin of the functional I^ε: large deviations of a stochastic particle system

The abstract methodology that we described above arises naturally in the context of *large deviations*, and we now describe this in the context of the three examples that we discuss in the next section. All three originate from (slight modifications of) one stochastic process, that models a collection of interacting particles with inertia in the physical space $\mathbb{R}^d$:

$$dQ_i^n(t) = \frac{P_i^n(t)}{m} dt, \tag{6a}$$

$$dP_i^n(t) = -\nabla V\left(Q_i^n(t)\right) dt - \frac{1}{n}\sum_{j=1}^{n} \nabla\psi\left(Q_j^n(t) - Q_i^n(t)\right) dt - \frac{\gamma}{m} P_i^n(t) dt + \sqrt{2\gamma\theta}\, dW_i(t). \tag{6b}$$

Here $Q_i^n \in \mathbb{R}^d$ and $P_i^n \in \mathbb{R}^d$ are the position and momentum of particles $i = 1, \ldots, n$ with mass m. Equation (6a) is the usual relation between $\dot{Q}_i^n$ and P_i^n, and (6b) is a force balance which describes the forces acting on the particle. For this system, corresponding to the first example below, these forces are (a) a force arising from a fixed potential V, (b) an interaction force deriving from a potential ψ, (c) a friction force, and (d) a stochastic force characterized by independent d-dimensional Wiener measures W_i. Throughout this paper we collect Q_i^n and P_i^n into a single variable $X_i^n = (Q_i^n, P_i^n)$.

The parameter γ characterizes the intensity of collisions of the particle with the solvent; it is present in both the friction term and the noise term, since they both arise from these collisions (and in accordance with the Einstein relation). The parameter $\theta = kT_a$, where k is the Boltzmann constant and T_a is the absolute temperature, measures the mean kinetic energy of the solvent molecules, and therefore characterizes the magnitude of collision noise. Typical applications of this system are for instance as a simplified model for chemical reactions, or as a model for particles interacting through Coulomb, gravitational, or volume-exclusion forces. However, our focus in this paper is on methodology, not on technicality, so we will assume that ψ is sufficiently smooth later on.

We now consider the many-particle limit $n \to \infty$ in (6). It is a well-known fact that the empirical measure

$$\rho_n(t) = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i^n(t)} \tag{7}$$

converges almost surely to the unique solution of the *Vlasov–Fokker–Planck (VFP) equation* [60]

$$\partial_t \rho = (\mathscr{L}_\rho)^* \rho, \qquad (\mathscr{L}_\mu)^* \rho := -\operatorname{div}_q\left(\rho \frac{p}{m}\right) + \operatorname{div}_p \rho\left(\nabla_q V + \nabla_q \psi * \mu + \gamma \frac{p}{m}\right) + \gamma\theta\, \Delta_p \rho, \tag{8}$$

$$= -\operatorname{div} \rho J \nabla (H + \psi * \mu) + \gamma \operatorname{div}_p \rho \frac{p}{m} + \gamma\theta \Delta_p \rho, \tag{9}$$

with an initial datum that derives from the initial distribution of X_i^n. The spatial domain here is $\mathbb{R}^{2d}$ with coordinates $(q, p) \in \mathbb{R}^d \times \mathbb{R}^d$, and subscripts such as in ∇_q and Δ_p indicate that differential operators act only on corresponding variables. The convolution is defined by $(\psi * \rho)(q) = \int_{\mathbb{R}^{2d}} \psi(q - q')\rho(q', p')dq'dp'$. In the second line above we use a slightly shorter way of writing $\mathscr{L}_\mu^*$, by introducing the Hamiltonian $H(q, p) = p^2/2m + V(q)$ and the canonical symplectic matrix $J = \left(\begin{smallmatrix} 0 & I \\ -I & 0 \end{smallmatrix}\right)$. This way of writing also highlights that the system is a combination of conservative effects, described by J, H, and ψ, and dissipative effects, which are parametrized by γ. The primal form $\mathscr{L}_\mu$ of the operator $(\mathscr{L}_\mu)^*$ is

$$\mathscr{L}_\mu f = J\nabla(H + \psi * \mu) \cdot \nabla f - \gamma \frac{p}{m} \cdot \nabla_p f + \gamma\theta \Delta_p f.$$

The almost-sure convergence of ρ_n to the solution ρ of the (deterministic) VFP equation is the starting point for a *large-deviation* result. In particular it has been shown that the sequence (ρ_n) has a *large-deviation property* [9,22,26] which characterizes the probability of finding the empirical measure far from the limit ρ, written informally as

$$\text{Prob}(\rho_n \approx \rho) \sim \exp\Big(-\frac{n}{2} I(\rho)\Big),$$

in terms of a *rate functional* $I : C([0, T]; \mathcal{P}(\mathbb{R}^{2d})) \to \mathbb{R}$. If we assume that the initial data X_i^n are chosen to be deterministic, and such that the initial empirical measure $\rho_n(0)$ converges narrowly to some ρ_0, then I has the form [26]

$$I(\rho) := \sup_{f \in C_b^{1,2}(\mathbb{R}\times\mathbb{R}^{2d})} \int_{\mathbb{R}^{2d}} f_T \, d\rho_T - \int_{\mathbb{R}^{2d}} f_0 \, d\rho_0 - \int_0^T \int_{\mathbb{R}^{2d}} \left(\partial_t f + \mathscr{L}_{\rho_t} f\right) d\rho_t dt$$

$$- \frac{1}{2}\int_0^T \int_{\mathbb{R}^{2d}} \Lambda(f, f)\, d\rho_t dt, \tag{10}$$

provided $\rho_t|_{t=0} = \rho_0$, where Λ is the carré-du-champ operator (e.g. [11, Sect. 1.4.2])

$$\Lambda(f, g) := \frac{1}{2}\big(\mathscr{L}_\mu(fg) - f\mathscr{L}_\mu g - g\mathscr{L}_\mu f\big) = \gamma\theta\, \nabla_p f \nabla_p g.$$

If the initial measure $\rho_t|_{t=0}$ is not equal to the limit ρ_0 of the stochastic initial empirical measures, then $I(\rho) = \infty$.

Note that the functional I in (10) is non-negative, since $f \equiv 0$ is admissible. If $I(\rho) = 0$, then by replacing f by λf and letting λ tend to zero we find that ρ is the weak solution of (8)

(which is unique, given initial data ρ_0 [35]). Therefore I is of the form that we discussed in Sect. 1.1: $I \geq 0$, and $I(\rho) = 0$ iff ρ solves (8), which is a realization of (1).

1.3 Concrete problems

We now apply the coarse-graining method of Sect. 1.1 to three limits: the *overdamped* limit $\gamma \to \infty$, and two *small-noise* limits $\theta \to 0$. In each of these three limits, the VFP Eq. (8) is the starting point, and we prove convergence to a limiting system using appropriate coarse-graining maps. Note that the convergence is therefore from one deterministic equation to another one; but the method makes use of the large-deviation structure that the VFP equation has inherited from its stochastic origin.

1.3.1 Overdamped limit of the Vlasov–Fokker–Planck equation

The first limit that we consider is the limit of large friction, $\gamma \to \infty$, in the Vlasov–Fokker–Planck equation (8), setting $\theta = 1$ for convenience. To motivate what follows, we divide (8) throughout by γ and formally let $\gamma \to \infty$ to find

$$\operatorname{div}_p \rho\left(\frac{p}{m}\right) + \Delta_p \rho = 0,$$

which suggests that in the limit $\gamma \to \infty$, ρ should be Maxwellian in p, i.e.

$$\rho_t(dq, dp) = Z^{-1} \exp\left(-\frac{p^2}{2m}\right) dp\, \sigma_t(dq), \tag{11}$$

where $Z = (2m\pi)^{d/2}$ is the normalization constant for the Maxwellian distribution. The main result in Sect. 2 shows that after an appropriate time rescaling, in the limit $\gamma \to \infty$, the remaining unknown $\sigma \in C([0, T]; \mathcal{P}(\mathbb{R}^d))$ solves the Vlasov–Fokker–Planck equation

$$\partial_t \sigma = \operatorname{div}(\sigma \nabla V(q)) + \operatorname{div}(\sigma(\nabla \psi * \sigma)) + \Delta\sigma. \tag{12}$$

In his seminal work [45], Kramers formally discussed these results for the 'Kramers equation', which corresponds to (8) with $\psi = 0$, and this limit has become known as the *Smoluchowski–Kramers approximation*. Nelson made these ideas rigorous [58] by studying the corresponding stochastic differential equations (SDEs); he showed that under suitable rescaling the solution to the Langevin equation converges almost surely to the solution of (12) with $\psi = 0$. Since then various generalizations and related results have been proved [18,34,43,56], mostly using stochastic and asymptotic techniques.

In this article we recover some of the results mentioned above for the VFP equation using the variational technique described in Sect. 1.1. Our proof is made up of the following three steps. Theorem 2.4 provides the necessary compactness properties to pass to the limit, Lemma 2.5 gives the characterization (11) of the limit, and in Theorem 2.6 we prove the convergence of the solution of the VFP equation to the solution of (12).

1.3.2 Small-noise limit of a randomly perturbed Hamiltonian system with one degree of freedom

In our second example we consider the following equation

$$\partial_t \rho = -\operatorname{div}_q\left(\rho \frac{p}{m}\right) + \operatorname{div}_p(\rho \nabla_q V) + \varepsilon \Delta_p \rho \qquad \text{on } [0, T] \times \mathbb{R}^2, \tag{13}$$

where $(q, p) \in \mathbb{R}^2$, $t \in [0, T]$ and div_q, div_p, Δ_p are one-dimensional derivatives. This equation can also be written as

$$\partial_t \rho = -\mathrm{div}(\rho J \nabla H) + \varepsilon \Delta_p \rho, \qquad \text{on } [0, T] \times \mathbb{R}^2. \tag{14}$$

This corresponds to the VFP Eq. (8) with $\psi = 0$, without friction and with small noise $\varepsilon = \gamma\theta$.

In addition to the interpretation as the many-particle limit of (6), Eq. (14) also is the forward Kolmogorov equation of a randomly perturbed Hamiltonian system in $\mathbb{R}^2$ with Hamiltonian H:

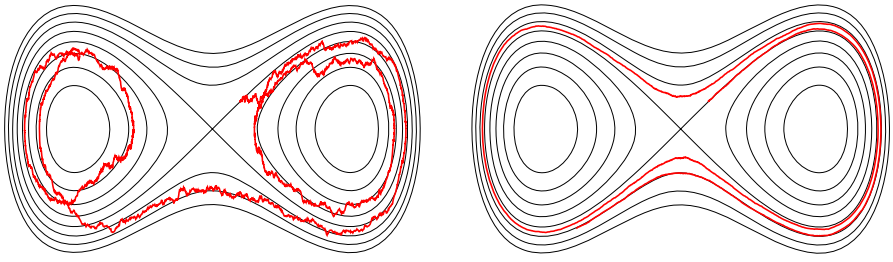
$$X = \begin{pmatrix} Q \\ P \end{pmatrix}, \qquad dX_t = J\nabla H(X_t) + \sqrt{2\varepsilon} \begin{pmatrix} 0 \\ 1 \end{pmatrix} dW_t, \tag{15}$$

where W_t is now a 1-dimensional Wiener process. When the amplitude ε of the noise is small, the dynamics (14) splits into fast and slow components. The fast component approximately follows an unperturbed trajectory of the Hamiltonian system, which is a level set of H. The slow component is visible as a slow modification of the value of H, corresponding to a motion transverse to the level sets of H. Figure 1 illustrates this.

Following [37] and others, in order to focus on the slow, Hamiltonian-changing motion, we rescale time such that the Hamiltonian, level-set-following motion is fast, of rate $O(1/\varepsilon)$, and the level-set-changing motion is of rate $O(1)$. In other words, the process (15) 'whizzes round' level sets of H at rate $O(1/\varepsilon)$, while shifting from one level set to another at rate $O(1)$.

This behaviour suggests choosing a coarse-graining map $\xi : \mathbb{R}^2 \to \Gamma$, which maps a whole *level set* to a single point in a new space Γ; because of the structure of level sets of H, the set Γ has a structure that is called a *graph*, a union of one-dimensional intervals locally parametrized by the value of the Hamiltonian. Figure 2 illustrates this, and in Sect. 3 we discuss it in full detail.

After projecting onto the graph Γ, the process turns out to behave like a diffusion process on Γ. This property was first made rigorous in [37] for a system with one degree of freedom, as here, and non-degenerate noise, using probabilistic techniques. In [38] the authors consider the case of degenerate noise by using probabilistic and analytic techniques based on hypoelliptic operators. More recently this problem has been handled using PDE techniques [44] (the elliptic case) and Dirichlet forms [15]. In Sect. 3 we give a new proof, using the structure outlined in Sect. 1.1.



(a) $\varepsilon = 0.005$ **(b)** $\varepsilon = 0.00005$

Fig. 1 Simulation of (15) for varying ε. Shown are the level curves of the Hamiltonian H and for each case a single trajectory

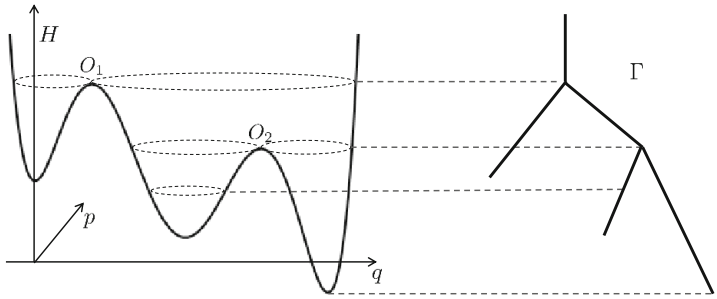


Fig. 2 Left: Hamiltonian $\mathbb{R}^2 \ni (q, p) \mapsto H(q, p)$, Right: Graph Γ

1.3.3 Small-noise limit of a randomly perturbed Hamiltonian system with d degrees of freedom

The convergence of solutions of (14) as $\varepsilon \to 0$ to a diffusion process on a graph requires that the non-perturbed system has a unique invariant measure on each connected component of a level set. While this is true for a Hamiltonian system with one degree of freedom, in the higher-dimensional case one might have additional first integrals of motion. In such a system the slow component will not be a one-dimensional process but a more complicated object—see [40]. However, by introducing an additional stochastic perturbation that destroys all first integrals except the Hamiltonian, one can regain the necessary ergodicity, such that the slow dynamics again lives on a graph.

In Sect. 4 we discuss this case. Equation (14) gains an additional noise term, and reads

$$\partial_t \rho = -\operatorname{div}(\rho J \nabla H) + \kappa \operatorname{div}(a \nabla \rho) + \varepsilon \Delta_p \rho, \tag{16}$$

where $a : \mathbb{R}^{2d} \to \mathbb{R}^{2d\times 2d}$ with $a\nabla H = 0$, $\dim(\text{Kernel}(a)) = 1$, and $\kappa, \varepsilon > 0$ with $\kappa \gg \varepsilon$. The spatial domain is $\mathbb{R}^{2d}$, $d > 1$ with coordinates $(q, p) \in \mathbb{R}^d \times \mathbb{R}^d$ and the unknown is a trajectory in the space of probability measures $\rho : [0, T] \to \mathcal{P}(\mathbb{R}^{2d})$. As before the aim is to derive the dynamics as $\varepsilon \to 0$. This problem was studied in [39] and the results closely mirror the previous case. The main difference lies in the proof of the local equilibrium statement, which we discuss in Sect. 4.

1.4 Comparison with other work

The novelty of the present paper lies in the following.

1. *In comparison with existing literature on the three concrete examples treated in this paper* The results of the three examples are known in the literature (see for instance [37–39,58]), but they are proved by different techniques and in a different setting. The variational approach of this paper, which has a clear microscopic interpretation from the large-deviation principle, to these problems is new. We provide alternative proofs, recovering known results, in a unified framework. In addition, we obtain all the results on compactness, local-equilibrium properties and liminf inequalities solely from the variational structures. The approach also is applicable to approximate solutions, which obey the original fine-grained dynamics only to some error. This allows us to work with larger class of measures and to relax many regularity conditions required by the exact solutions. Furthermore, our abstract setting has potential applications to many other systems.

2. *In comparison with recently developed variational-evolutionary methods* Many recently developed variational techniques for 'passing to a limit' such as the Sandier-Safety method based on the Ψ–Ψ^* structure [6,51,70] only apply to gradient flows, i.e. dissipative systems. The approach of this paper also applies to certain variational-evolutionary systems that include non-dissipative effects, such as GENERIC systems [26,62]; our examples illustrate this. Since our approach only uses the duality structure of the rate functionals, which holds true for more general systems, this method also works for other limits in non-gradient-flow systems such as the Langevin limit of the Nosé–Hoover–Langevin thermostat [31,61,68].
3. *Quantification of the coarse-graining error* The use of the rate functional as a central ingredient in 'passing to a limit' and coarse-graining also allows us to obtain quantitative estimates of the coarse-graining error. One intermediate result of our analysis is a functional inequality similar to the energy-dissipation inequality in the gradient-flow setting (see (4)). This inequality provides an upper bound on the free energy and the integral of the Fisher information by the rate functional and initial free energy. To keep the paper to a reasonable length, we address this issue in details separately in a companion article [23].

We provide further comments in Sect. 5.

1.5 Outline of the article

The rest of the paper is devoted to the study of three concrete problems: the overdamped limit of the VFP equation in Sect. 2, diffusion on a graph with one degree of freedom in Sect. 3, and diffusion on a graph with many degrees of freedom in Sect. 4. In each section, the main steps in the abstract framework are performed in detail. Section 5 provides further discussion. Finally, detailed proofs of some theorems are given in Appendices A and B.

1.6 Summary of notation

$\pm_{kj}$	± 1, depending on which end vertex O_j lies of edge I_k	Sect. 3.1
$\mathcal{F}$	Free energy	(22), (46)
γ (Sect. 2)	Large-friction parameter	
Γ, γ (Sect. 3)	The graph Γ and its elements γ	Sect. 3.1
$\mathcal{H}(\cdot\|\cdot)$	Relative entropy	(21)
$H(q, p)$	$H(q, p) = p^2/2m + V(q)$, the Hamiltonian	
$\mathscr{H}^n$	n-dimensional Hausdorff measure	
$\mathcal{I}(\cdot\|\cdot)$	relative Fisher Information	(24)
Int	The interior of a set	
I^ε	Large-deviation rate functional for the diffusion-on-graph problem	(48)
I^γ	Large-deviation rate functional for the VFP equation	(19)
J	$J = \left(\begin{smallmatrix} 0 & I \\ -I & 0 \end{smallmatrix}\right)$, the canonical symplectic matrix	
$\mathcal{L}$	Lebesgue measure	
$\mathscr{L}_\mu$, $(\mathscr{L}_\mu)^*$	Primal and dual generators	Sect. 1.2
$\mathcal{M}(\mathcal{X})$	Space of finite, non-negative Borel measures on $\mathcal{X}$	
$\mathcal{P}(\mathcal{X})$	Space of probability measures on $\mathcal{X}$	
$\hat{\rho}$	Push-forward under ξ of ρ	(45)
$T(\gamma)$	Period of the periodic orbit at $\gamma \in \Gamma$	(51)
$V(q)$	Potential on position ('on-site')	
x	$x = (q, p)$ joint variable	
ξ^γ, ξ	Coarse-graining maps	(30), (44)

Throughout we use measure notation and terminology. For a given topological space $\mathcal{X}$, the space $\mathcal{M}(\mathcal{X})$ is the space of non-negative, finite Borel measures on $\mathcal{X}$; $\mathcal{P}(\mathcal{X})$ is the space of probability measures on $\mathcal{X}$. For a measure $\rho \in \mathcal{M}([0,T]\times\mathbb{R}^{2d})$, for instance, we often write $\rho_t \in \mathcal{M}(\mathbb{R}^{2d})$ for the time slice at time t; we also often use both the notation $\rho(x)dx$ and $\rho(dx)$ when ρ is Lebesgue-absolutely-continuous. We equip $\mathcal{M}(\mathcal{X})$ and $\mathcal{P}(\mathcal{X})$ with the *narrow* topology, in which convergence is characterized by duality with continuous and bounded functions on $\mathcal{X}$.

2 Overdamped limit of the VFP equation

2.1 Setup of the system

In this section we prove the large-friction limit $\gamma \to \infty$ of the VFP Eq. (8). Setting $\theta = 1$ for convenience, and speeding time up by a factor γ, the VFP equation reads

$$\partial_t \rho = \mathscr{L}_\rho^* \rho, \qquad \mathscr{L}_\nu^* \rho := -\gamma \operatorname{div} \rho J \nabla (H + \psi * \nu) + \gamma^2 \left[\operatorname{div}_p \left(\rho \frac{p}{m} \right) + \Delta_p \rho \right], \quad (17)$$

where, as before, $J = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}$ and $H(q,p) = p^2/2m + V(q)$. The spatial domain is $\mathbb{R}^{2d}$ with coordinates $(q,p) \in \mathbb{R}^d \times \mathbb{R}^d$ with $d \geq 1$, and $\rho \in C([0,T]; \mathcal{P}(\mathbb{R}^{2d}))$. For later reference we also mention the primal form of the operator $\mathscr{L}_\nu^*$:

$$\mathscr{L}_\nu f = \gamma J \nabla (H + \psi * \nu) \cdot \nabla f - \gamma^2 \frac{p}{m} \cdot \nabla_p f + \gamma^2 \Delta_p f. \quad (18)$$

We assume

(V1) The potential $V \in C^2(\mathbb{R}^d)$ has globally bounded second derivative. Furthermore $V \geq 0$, $|\nabla V|^2 \leq C(1+V)$ for some $C > 0$, and $\mathrm{e}^{-V} \in L^1(\mathbb{R}^d)$.

(V2) The interaction potential $\psi \in C^2(\mathbb{R}^d) \cap W^{1,1}(\mathbb{R}^d)$ is symmetric, has globally bounded first and second derivatives, and the mapping $\nu \mapsto \int \nu * \psi \, d\nu$ is convex (or equivalently non-negative).

As we described in Sect. 1.1, the study of the limit $\gamma \to \infty$ contains the following steps:

1. Prove compactness;
2. Prove a local-equilibrium property;
3. Prove a liminf inequality.

According to the framework detailed by (1), (2), each of these results is based on the large-deviation structure, which for Eq. (17) is associated to the functional $I^\gamma : C([0,T]; \mathcal{P}(\mathbb{R}^{2d})) \to \mathbb{R}$ with

$$I^\gamma(\rho) = \sup_{f \in C_b^{1,2}(\mathbb{R}\times\mathbb{R}^{2d})} \left[\int_{\mathbb{R}^{2d}} f_T \, d\rho_T - \int_{\mathbb{R}^{2d}} f_0 \, d\rho_0 - \int_0^T \int_{\mathbb{R}^{2d}} \left(\partial_t f_t + \mathscr{L}_{\rho_t} f_t \right) d\rho_t dt \right.$$
$$\left. - \frac{\gamma^2}{2} \int_0^T \int_{\mathbb{R}^{2d}} \left| \nabla_p f_t \right|^2 d\rho_t dt \right], \quad (19)$$

where $\mathscr{L}_\nu$ is given in (18). Alternatively the rate functional can be written as [26, Theorem 2.5]

$$I^\gamma(\rho) = \begin{cases} \frac{1}{2}\int_0^T \int_{\mathbb{R}^{2d}} |h_t|^2 \, d\rho_t dt & \text{if } \partial_t \rho_t = \mathscr{L}^*_{\rho_t}\rho_t - \gamma \operatorname{div}_p(\rho_t h_t), \text{ for } h \in L^2(0,T; L^2_\nabla(\rho)), \text{ and } \rho|_{t=0} = \rho_0 \\ +\infty & \text{otherwise,} \end{cases} \tag{20}$$

where $\mathscr{L}^*_\nu$ is given in (17). For fixed t, the space $L^2_\nabla(\rho_t)$ is the closure of the set $\{\nabla_p \varphi : \varphi \in C_c^\infty(\mathbb{R}^{2d})\}$ in $L^2(\rho_t)$, the ρ_t-weighted L^2-space. Similarly, $L^2(0,T; L^2_\nabla(\rho))$ is defined as the closure of $\{\nabla_p \varphi : \varphi \in C_c^\infty((0,T)\times\mathbb{R}^{2d})\}$ in the L^2-space associated to the space–time density ρ. This second form of the rate functional shows clearly how $I^\gamma(\rho) = 0$ is equivalent to the property that ρ solves the VFP Eq. (17). It also shows that if $I^\gamma(\rho) > 0$, then ρ is an approximate solution in the sense that it satisfies the VFP equation up to some error $-\gamma \operatorname{div}_p(\rho_t h_t)$ whose norm is controlled by the rate functional.

2.2 A priori bounds

We give ourselves a sequence, indexed by γ, of solutions ρ^γ to the VFP Eq. (17) with initial datum $\rho^\gamma_t|_{t=0} = \rho_0$. We will deduce the compactness of the sequence ρ^γ from *a priori* estimates, that are themselves derived from the rate function I^γ.

For probability measures ν, ζ on $\mathbb{R}^{2d}$ we first introduce:

- Relative entropy:

$$\mathcal{H}(\nu|\zeta) = \begin{cases} \int_{\mathbb{R}^{2d}} [f \log f] \, d\zeta & \text{if } \nu = f\zeta, \\ \infty & \text{otherwise.} \end{cases} \tag{21}$$

- The free energy for this system:

$$\mathcal{F}(\nu) := \mathcal{H}(\nu|Z_H^{-1}\mathrm{e}^{-H}dx) + \frac{1}{2}\int_{\mathbb{R}^{2d}} \psi * \nu \, d\nu = \int_{\mathbb{R}^{2d}} \Big[\log g + H + \frac{1}{2}\psi * g\Big] g dx + \log Z_H, \tag{22}$$

where $Z_H = \int \mathrm{e}^{-H}$ and the second expression makes sense whenever $\nu = g dx$.

The convexity of the term involving ψ (condition (V2)) implies that the free energy $\mathcal{F}$ is strictly convex and has a unique minimizer $\mu \in \mathcal{P}(\mathbb{R}^{2d})$. This minimizer is a stationary point of the evolution (17), and has the implicit characterization

$$\mu \in \mathcal{P}(\mathbb{R}^{2d}) : \ \mu(dqdp) = Z^{-1} \exp\Big(-\big[H(q,p) + (\psi * \mu)(q)\big]\Big) dqdp, \tag{23}$$

where Z is the normalization constant for μ. Note that $\nabla_p \mu = -\mu \nabla_p H = -p\mu/m$.

We also define the *relative Fisher Information* with respect to μ (in the p-variable only):

$$\mathcal{I}(\nu|\mu) = \sup_{\varphi \in C_c^\infty(\mathbb{R}^{2d})} 2\int_{\mathbb{R}^{2d}} \Big[\Delta_p \varphi - \frac{p}{m}\nabla_p \varphi - \frac{1}{2}|\nabla_p \varphi|^2\Big] d\nu. \tag{24}$$

Note that the right hand side of (24) depends on μ via $\nabla_p(\log \mu) = -\nabla_p H(q,p) = -p/m$. In the more common case in which the derivatives Δ_p and ∇_p are replaced by the full derivatives Δ and ∇, the relative Fisher Information has an equivalent formulation in terms of the Lebesgue density of ν. In our case such equivalence only holds when ν is absolutely continuous with respect to the Lebesgue measure in both q and p:

Lemma 2.1 (Equivalence of relative-Fisher-Information expressions for a.c. measures) *If* $\nu \in \mathcal{P}(\mathbb{R}^{2d})$, $\nu(dx) = f(x)dx$ *with* $f \in L^1(\mathbb{R}^{2d})$, *then*

$$\mathcal{I}(\nu|\mu) = \begin{cases} \int_{\mathbb{R}^{2d}} \left| \frac{\nabla_p f}{f} \mathbb{1}_{\{f>0\}} + \frac{p}{m} \right|^2 f \, dqdp, & \text{if } \nabla_p f \in L^1_{\text{loc}}(dqdp), \\ \infty & \text{otherwise}, \end{cases} \tag{25}$$

where $\mathbb{1}_{\{f>0\}}$ *denotes the indicator function of the set* $\{x \in \mathbb{R}^{2d} \mid f(x) > 0\}$ *and* $\nabla_p f$ *is the distributional gradient of* f *in the* p*-variable only.*

For a measure of the form $\zeta(dq)f(p)dp$, with $\zeta \not\ll dq$, the functional $\mathcal{I}$ in (24) may be finite while the integral in (25) is not defined. Because of the central role of duality in this paper, definition (24) is a natural one, as we shall see below. The proof of Lemma 2.1 is given in Appendix A.

In the introduction we mentioned that we expect ρ^γ to become Maxwellian in the limit $\gamma \to \infty$. This will be driven by a vanishing relative Fisher Information, as we shall see below. For absolutely continuous measures, the characterization (25) already provides the property

$$\mathcal{I}(f\,dx|\mu) = 0 \quad \Longrightarrow \quad f(q,p) = \tilde{f}(q) \exp\Big(-\frac{p^2}{2m}\Big).$$

This property holds more generally:

Lemma 2.2 (Zero relative Fisher Information implies Maxwellian) *If* $\nu \in \mathcal{P}(\mathbb{R}^{2d})$ *with* $\mathcal{I}(\nu|\mu) = 0$, *then there exists* $\sigma \in \mathcal{P}(\mathbb{R}^d)$ *such that*

$$\nu(dqdp) = Z^{-1} \exp\left(-\frac{p^2}{2m}\right) \sigma(dq)dp,$$

where $Z = \int_{\mathbb{R}^d} \mathrm{e}^{-p^2/2m} dp$ *is the normalization constant for the Maxwellian distribution.*

Proof From

$$\mathcal{I}(\nu|\mu) = \sup_{\varphi \in C_c^\infty(\mathbb{R}^{2d})} 2 \int_{\mathbb{R}^{2d}} \left(\Delta_p \varphi - \frac{p}{m} \cdot \nabla_p \varphi - \frac{1}{2} |\nabla_p \varphi|^2 \right) d\nu = 0 \tag{26}$$

we conclude upon disintegrating ν as $\nu(dqdp) = \sigma(dq)\nu_q(dp)$,

$$\text{for } \sigma - \text{a.e. } q: \quad \sup_{\phi \in C_c^\infty(\mathbb{R}^d)} \int_{\mathbb{R}^d} \left(\Delta_p \phi - \frac{p}{m} \cdot \nabla_p \phi - \frac{1}{2} |\nabla_p \phi|^2 \right) \nu_q(dp) = 0.$$

By replacing ϕ by $\lambda\phi$, $\lambda > 0$, and taking $\lambda \to 0$ we find

$$\forall \phi \in C_c^\infty(\mathbb{R}^d): \int_{\mathbb{R}^d} \left(\Delta_p \phi - \frac{p}{m} \cdot \nabla_p \phi \right) \nu_q(dp) = 0,$$

which is the weak form of an elliptic equation on $\mathbb{R}^d$ with unique solution (see e.g. [13, Theorem 4.1.11])

$$\nu_q(dp) = \frac{1}{Z} \exp\left(-\frac{p^2}{2m}\right) dp.$$

This proves the lemma. □

In the following theorem we give the central *a priori* estimate, in which free energy and relative Fisher Information are bounded from above by the rate functional and the relative entropy at initial time.

Theorem 2.3 (A priori bounds) *Fix $\gamma > 0$ and let $\rho \in C([0,T];\mathcal{P}(\mathbb{R}^{2d}))$ with $\rho_t|_{t=0} =: \rho_0$ satisfy*

$$I^\gamma(\rho) < \infty, \ \mathcal{F}(\rho_0) < \infty. \tag{27}$$

Then for any $t \in [0,T]$ we have

$$\mathcal{F}(\rho_t) + \frac{\gamma^2}{2}\int_0^t \mathcal{I}(\rho_s|\mu)\,ds \le I^\gamma(\rho) + \mathcal{F}(\rho_0). \tag{28}$$

From (28) we obtain the separate inequality

$$\frac{1}{2}\int_{\mathbb{R}^{2d}} H\,d\rho_t \le \mathcal{F}(\rho_0) + I^\gamma(\rho) + \log\frac{\int_{\mathbb{R}^{2d}} \mathrm{e}^{-H/2}}{\int_{\mathbb{R}^{2d}} \mathrm{e}^{-H}}. \tag{29}$$

This estimate will lead to a priori bounds in two ways. First, the bound (29) gives tightness estimates, and therefore compactness in space and time (Theorem 2.4); secondly, by (28), the relative Fisher Information is bounded by C/γ^2 and therefore vanishes in the limit $\gamma \to \infty$. This fact is used to prove that the limiting measure is Maxwellian (Lemma 2.5).

Proof We give a heuristic motivation here; Appendix B contains a full proof. Given a trajectory ρ as in the theorem, note that by (20) ρ satisfies

$$\begin{aligned}\partial_t\rho_t &= -\gamma\,\mathrm{div}\,\rho_t J\nabla(H+\psi*\rho_t) + \gamma^2\Big(\mathrm{div}_p\,\rho_t\frac{p}{m} + \Delta_p\rho_t\Big) - \gamma\,\mathrm{div}_p\,\rho_t h_t,\\ &\qquad \text{with } h \in L^2(0,T;L^2_\nabla(\rho)).\end{aligned}$$

We then formally calculate

$$\begin{aligned}\frac{d}{dt}\mathcal{F}(\rho_t) &= \int_{\mathbb{R}^{2d}}\big[\log\rho_t + 1 + H + \psi*\rho_t\big]\Big(-\gamma\,\mathrm{div}\,\rho_t J\nabla(H+\psi*\rho_t)\\ &\quad + \gamma^2\big(\mathrm{div}_p\,\rho_t\frac{p}{m} + \Delta_p\rho_t\big) - \gamma\,\mathrm{div}_p\,\rho_t h_t\Big)\\ &= -\gamma^2\int_{\mathbb{R}^{2d}}\frac{1}{\rho_t}\Big|\nabla_p\rho_t + \rho_t\frac{p}{m}\Big|^2 + \gamma\int_{\mathbb{R}^{2d}} h_t\Big(\nabla_p\rho_t + \rho_t\frac{p}{m}\Big)\\ &\le -\frac{\gamma^2}{2}\int_{\mathbb{R}^{2d}}\frac{1}{\rho_t}\Big|\nabla_p\rho_t + \rho_t\frac{p}{m}\Big|^2 + \frac{1}{2}\int_{\mathbb{R}^{2d}}\rho_t h_t^2,\end{aligned}$$

where the first $O(\gamma)$ term cancels because of the anti-symmetry of J. After integration in time this latter expression yields (28).

For exact solutions of the VFP equation, i.e. when $I^\gamma(\rho) = 0$, this argument can be made rigorous following e.g. [8]. However, the fairly low regularity of the right-hand side in (20) prevents these techniques from working. 'Mild' solutions, defined using the variation-of-constants formula and the Green function for the hypoelliptic operator, are not well-defined either, for the same reason: the term $\iint \nabla_p G\cdot h\,d\rho$ that appears in such an expression is generally not integrable. In the appendix we give a different proof, using the method of dual equations.

Equation (29) follows by substituting

$$\mathcal{F}(\rho_t) = \mathcal{H}\Big(\rho_t\Big|Z_{H/2}^{-1}\mathrm{e}^{-H/2}dx\Big) + \frac{1}{2}\int_{\mathbb{R}^{2d}} H\,d\rho_t + \frac{1}{2}\int_{\mathbb{R}^{2d}}\psi*\rho_t\,d\rho_t + \log\frac{\int_{\mathbb{R}^{2d}}\mathrm{e}^{-H}}{\int_{\mathbb{R}^{2d}}\mathrm{e}^{-H/2}},$$

in (28), where $Z_{H/2} := \int_{\mathbb{R}^{2d}} \mathrm{e}^{-H/2}$. □

2.3 Coarse-graining and compactness

As we described in the introduction, in the overdamped limit $\gamma \to \infty$ we expect that ρ will resemble a Maxwellian distribution $Z^{-1} \exp\bigl(-p^2/2m\bigr)\sigma_t(dq)$, and that the q-dependent part σ will solve Eq. (12). We will prove this statement using the method described in Sect. 1.1.

It would be natural to define 'coarse-graining' in this context as the projection $\xi(q, p) := q$, since that should eliminate the fast dynamics of p and focus on the slower dynamics of q. However, this choice fails: it completely decouples the dynamics of q from that of p, thereby preventing the noise in p from transferring to q. Following the lead of Kramers [45], therefore, we define a slightly different coarse-graining map

$$\xi^\gamma : \mathbb{R}^{2d} \to \mathbb{R}^d, \qquad \xi^\gamma(q, p) := q + \frac{p}{\gamma}. \tag{30}$$

In the limit $\gamma \to \infty$, $\xi^\gamma \to \xi$ locally uniformly, recovering the projection onto the q-coordinate.

The theorem below gives the compactness properties of the solutions ρ^γ of the rescaled VFP equation that allow us to pass to the limit. There are two levels of compactness, a weaker one in the original space $\mathbb{R}^{2d}$, and a stronger one in the coarse-grained space $\mathbb{R}^d = \xi^\gamma(\mathbb{R}^{2d})$. This is similar to other multilevel compactness results as in e.g. [42].

Theorem 2.4 (Compactness) *Let a sequence $\rho^\gamma \in C([0, T]; \mathcal{P}(\mathbb{R}^{2d}))$ satisfy for a suitable constant $C > 0$ and every γ the estimate*

$$I^\gamma(\rho^\gamma) + \mathcal{F}(\rho_t^\gamma|_{t=0}) \le C. \tag{31}$$

Then there exist a subsequence (not relabeled) such that

1. *$\rho^\gamma \to \rho$ in $\mathcal{M}([0, T] \times \mathbb{R}^{2d})$ with respect to the narrow topology.*
2. *$\xi^\gamma_\# \rho^\gamma \to \xi_\# \rho$ in $C([0, T]; \mathcal{P}(\mathbb{R}^d))$ with respect to the uniform topology in time and narrow topology on $\mathcal{P}(\mathbb{R}^d)$.*

For a.e. $t \in [0, T]$ the limit ρ_t satisfies

$$\mathcal{I}(\rho_t|\mu) = 0 \tag{32}$$

Proof To prove part 1, note that the positivity of the convolution integral involving ψ and the free-energy-dissipation inequality (28) imply that $\mathcal{H}(\rho_t^\gamma | Z_H^{-1} \mathrm{e}^{-H} dx)$ is bounded uniformly in t and γ. By an argument as in [7, Prop. 4.2] this implies that the set of space–time measures $\{\rho^\gamma : \gamma > 1\}$ is tight, from which compactness in $\mathcal{M}([0, T] \times \mathbb{R}^{2d})$ follows.

To prove (32) we remark that

$$0 \le \sup_{\varphi \in C_c^\infty(\mathbb{R} \times \mathbb{R}^{2d})} 2 \int_0^T \int_{\mathbb{R}^{2d}} \Bigl[\Delta_p \varphi - \frac{p}{m} \nabla_p \varphi - \frac{1}{2} |\nabla_p \varphi|^2\Bigr] d\rho_t^\gamma \, dt \le \int_0^T \mathcal{I}(\rho_t^\gamma|\mu)\, dt$$
$$\le \frac{C}{\gamma^2} \overset{\gamma \to \infty}{\longrightarrow} 0,$$

and by passing to the limit on the left-hand side we find

$$\sup_{\varphi \in C_c^\infty(\mathbb{R} \times \mathbb{R}^{2d})} 2 \int_0^T \int_{\mathbb{R}^{2d}} \Bigl[\Delta_p \varphi - \frac{p}{m} \nabla_p \varphi - \frac{1}{2} |\nabla_p \varphi|^2\Bigr] d\rho_t \, dt = 0.$$

By disintegrating ρ in time as $\rho(dtdqdp) = \rho_t(dqdp)dt$, we find that $\mathcal{I}(\rho_t|\mu) = 0$ for (Lebesgue-) almost all t.

We prove part 2 with the Arzelà–Ascoli theorem. For any $t \in [0, T]$ the sequence $\xi^\gamma_\# \rho^\gamma_t$ is tight, which follows from the tightness of ρ^γ_t proved above and the local uniform convergence $\xi^\gamma \to \xi$ (see e.g. [4, Lemma 5.2.1]).

To prove equicontinuity we will show that

$$\sup_{\gamma>1} \sup_{t\in[0,T-h]} \sup_{\substack{\varphi\in C^2_c(\mathbb{R}^d)\\ \|\varphi\|_{C^2(\mathbb{R}^d)}\le 1}} \int_{\mathbb{R}^d} \varphi(\xi^\gamma_\# \rho^\gamma_{t+h} - \xi^\gamma_\# \rho^\gamma_t) \xrightarrow{h\to 0} 0. \tag{33}$$

In fact, (33) is a direct consequence of the following stronger statement

$$\int_{\mathbb{R}^d} \varphi(\xi^\gamma_\# \rho^\gamma_{t+h} - \xi^\gamma_\# \rho^\gamma_t) \le C\|\nabla\varphi\|_\infty \sqrt{h} \tag{34}$$

with C independent of t, γ and φ. Note that (34) in particular implies a uniform 1/2-Hölder estimate with respect to the L^1-Wasserstein distance.

Let us now give the proof of (34). Indeed, the boundedness of the rate functional, definition (20), and tightness of ρ^γ imply that there exists some $h^\gamma \in L^2(0, T; L^2_\nabla(\rho^\gamma_t))$ with

$$\partial_t \rho^\gamma_t = (\mathcal{L}_{\rho^\gamma_t})^* \rho^\gamma_t - \gamma \operatorname{div}_p(\rho^\gamma_t h^\gamma_t). \tag{35}$$

in duality with $C^2_b(\mathbb{R}^{2d})$, pointwise almost everywhere in $t \in [0, T]$. Therefore for any $f \in C^2_b(\mathbb{R}^{2d})$ we have in the sense of distributions on $[0, T]$,

$$\begin{aligned}\frac{d}{dt}\int_{\mathbb{R}^{2d}} f\rho^\gamma_t = \int_{\mathbb{R}^{2d}} \Big(&\gamma \frac{p}{m}\cdot\nabla_q f - \gamma\nabla_q V\cdot\nabla_p f - \gamma\nabla_p f\cdot(\nabla_q\psi * \rho^\gamma)\\ &- \gamma^2\frac{p}{m}\cdot\nabla_p f + \gamma^2\Delta_p f + \gamma\nabla_p f\cdot h^\gamma_t\Big) d\rho^\gamma_t.\end{aligned}$$

To prove (34), make the choice $f = \varphi \circ \xi^\gamma$ for $\varphi \in C^2_c(\mathbb{R}^d)$ and integrate over $[t, t + h]$. Note that due to the specific form of $\xi^\gamma = q + p/\gamma$ the terms $\gamma\frac{p}{m}\cdot\nabla_q f$ and $\gamma^2\frac{p}{m}\cdot\nabla_p f$ cancel and therefore

$$\begin{aligned}\int_{\mathbb{R}^d} \varphi(\xi^\gamma_\# \rho^\gamma_{t+h} - \xi^\gamma_\# \rho^\gamma_t) = \int_t^{t+h}\int_{\mathbb{R}^{2d}} \Big(&-\nabla V(q)\cdot\nabla\varphi\Big(q+\frac{p}{\gamma}\Big) - (\nabla_q\psi * \rho^\gamma_s)(q)\cdot\nabla\varphi\Big(q+\frac{p}{\gamma}\Big)\\ &+ \Delta\varphi\Big(q+\frac{p}{\gamma}\Big) + \nabla\varphi\Big(q+\frac{p}{\gamma}\Big)\cdot h^\gamma_s(q,p)\Big) d\rho^\gamma_s\, ds.\end{aligned}$$

We estimate the first term on the right hand side by using Hölder's inequality and growth condition (V1),

$$\begin{aligned}&\left|\int_t^{t+h}\int_{\mathbb{R}^{2d}} \nabla V(q)\cdot\nabla\varphi\Big(q+\frac{p}{\gamma}\Big) d\rho^\gamma_s\, ds\right|\\ &\le \|\nabla\varphi\|_\infty\sqrt{h}\left(\int_t^{t+h}\int_{\mathbb{R}^{2d}} |\nabla V(q)|^2 d\rho^\gamma_s\, ds\right)^{1/2}\\ &\le \|\nabla\varphi\|_\infty\sqrt{h}\left(\int_t^{t+h}\int_{\mathbb{R}^{2d}} C(1+V(q))\rho^\gamma_s\, ds\right)^{1/2}\\ &\le \tilde{C}\|\nabla\varphi\|_\infty\sqrt{h},\end{aligned}$$

where the last inequality follows from the free-energy-dissipation inequality (28). For the second term we use $|\nabla_q \psi * \rho_s^\gamma| \leq \|\nabla_q \psi\|_\infty$ and the last term is estimated by Hölder's inequality,

$$\left| \int_t^{t+h} \int_{\mathbb{R}^{2d}} \nabla\varphi \left(q + \frac{p}{\gamma} \right) h_s^\gamma(q,p) d\rho_s^\gamma ds \right| \leq \|\nabla\varphi\|_\infty \sqrt{h} \left(\int_t^{t+h} \int_{\mathbb{R}^{2d}} |h_s^\gamma|^2 d\rho_s^\gamma ds \right)^{\frac{1}{2}}$$
$$\leq \|\nabla\varphi\|_\infty \sqrt{h} \left(2I^\gamma(\rho^\gamma) \right)^{\frac{1}{2}} \leq C \|\nabla\varphi\|_\infty \sqrt{h}.$$

To sum up we have

$$\left| \int_{\mathbb{R}^d} \varphi(\xi_\#^\gamma \rho_{t+h}^\gamma - \xi_\#^\gamma \rho_t^\gamma) \right| \leq C \|\nabla\varphi\|_\infty \sqrt{h} \xrightarrow{h \to 0} 0,$$

where C is independent of t, γ and φ.

Thus by the Arzelà–Ascoli theorem there exists a $\nu \in C([0,T]; \mathcal{P}(\mathbb{R}^d))$ such that $\xi_\#^\gamma \rho^\gamma \to \nu$ with respect to uniform topology in time and narrow topology on $\mathcal{P}(\mathbb{R}^d)$. Since $\rho^\gamma \to \rho$ in $\mathcal{M}([0,T] \times \mathbb{R}^{2d})$ and $\xi^\gamma \to \xi$ locally uniformly, we have $\xi_\#^\gamma \rho^\gamma \to \xi_\# \rho$ in $\mathcal{M}([0,T] \times \mathbb{R}^d)$ (again using [4, Lemma 5.2.1]), implying that $\nu = \xi_\# \rho$. This concludes the proof of Theorem 2.4.

2.4 Local equilibrium

A central step in any coarse-graining method is the treatment of the information that is 'lost' upon coarse-graining. The lemma below uses the a priori estimate (28) to reconstruct this information, which for this system means showing that ρ^γ becomes Maxwellian in p as $\gamma \to \infty$.

Lemma 2.5 (Local equilibrium) *Under the assumptions of Theorem 2.4, let $\rho^\gamma \to \rho$ in $\mathcal{M}([0,T] \times \mathbb{R}^{2d})$ with respect to the narrow topology and $\xi_\#^\gamma \rho^\gamma \to \xi_\# \rho$ in $C([0,T]; \mathcal{P}(\mathbb{R}^d))$ with respect to the uniform topology in time and narrow topology on $\mathcal{P}(\mathbb{R}^d)$. Then there exists $\sigma \in C([0,T]; \mathcal{P}(\mathbb{R}^d))$, $\sigma(dtdq) = \sigma_t(dq)dt$, such that for almost all $t \in [0,T]$,*

$$\rho_t(dqdp) = Z^{-1} \exp\left(-\frac{p^2}{2m} \right) \sigma_t(dq)dp, \tag{36}$$

where $Z = \int_{\mathbb{R}^d} \mathrm{e}^{-p^2/2m} dp$ is the normalization constant for the Maxwellian distribution. Furthermore $\xi_\#^\gamma \rho^\gamma \to \sigma$ uniformly in time and narrowly on $\mathcal{P}(\mathbb{R}^d)$.

Proof Since $\rho^\gamma \to \rho$ narrowly in $\mathcal{M}([0,T] \times \mathbb{R}^{2d})$, the limit ρ also has the disintegration structure $\rho(dtdpdq) = \rho_t(dpdq)dt$, with $\rho_t \in \mathcal{P}(\mathbb{R}^{2d})$. From the *a priori* estimate (28) and the duality definition of $\mathcal{I}$ we have $\mathcal{I}(\rho_t|\mu) = 0$ for almost all t, and the characterization (36) then follows from Lemma 2.2. The uniform in time convergence of $\xi_\#^\gamma \rho^\gamma$ implies $\xi_\#^\gamma \rho^\gamma \to \xi_\# \rho = \sigma$ uniformly in time and narrowly on $\mathcal{P}(\mathbb{R}^d)$ and the regularity $\sigma \in C([0,T]; \mathcal{P}(\mathbb{R}^d))$.

2.5 Liminf inequality

The final step in the variational technique is proving an appropriate liminf inequality which also provides the structure of the limiting coarse-grained evolution. The following theorem makes this step rigorous.

Define the (limiting) functional $I : C([0,T];\mathcal{P}(\mathbb{R}^d)) \to \mathbb{R}$ by

$$\begin{aligned} I(\sigma) := \sup_{g \in C_b^{1,2}(\mathbb{R}\times\mathbb{R}^d)} & \int_{\mathbb{R}^d} g_T d\sigma_T - \int_{\mathbb{R}^d} g_0 d\sigma_0 - \int_0^T \int_{\mathbb{R}^d} \Big(\partial_t g \\ & - \nabla V \cdot \nabla g - (\nabla \psi * \sigma) \cdot \nabla g + \Delta g\Big) d\sigma_t dt \\ & - \frac{1}{2}\int_0^T \int_{\mathbb{R}^d} |\nabla g|^2 \, d\sigma_t dt. \end{aligned} \tag{37}$$

Note that $I \geq 0$ (since $g = 0$ is admissible); we have the equivalence

$$I(\sigma) = 0 \iff \partial_t \sigma = \operatorname{div} \sigma \nabla V(q) + \operatorname{div} \sigma (\nabla \psi * \sigma) + \Delta \sigma \quad \text{in } [0,T] \times \mathbb{R}^d.$$

Theorem 2.6 (Liminf inequality) *Under the same conditions as in Theorem 2.4 we assume that $\rho^\gamma \to \rho$ narrowly in $\mathcal{M}([0,T] \times \mathbb{R}^{2d})$ and $\xi^\gamma_\# \rho^\gamma \to \xi_\# \rho \equiv \sigma$ in $C([0,T];\mathcal{P}(\mathbb{R}^d))$. Then*

$$\liminf_{\gamma \to \infty} I^\gamma(\rho^\gamma) \geq I(\sigma).$$

Proof Write the large deviation rate functional $I^\gamma : C([0,T];\mathcal{P}(\mathbb{R}^{2d})) \to \mathbb{R}$ in (19) as

$$I^\gamma(\rho) = \sup_{f \in C_b^{1,2}(\mathbb{R}\times\mathbb{R}^{2d})} \mathcal{J}^\gamma(\rho, f), \tag{38}$$

where

$$\begin{aligned} \mathcal{J}^\gamma(\rho, f) = & \int_{\mathbb{R}^{2d}} f_T d\rho_T - \int_{\mathbb{R}^{2d}} f_0 d\rho_0 - \int_0^T \int_{\mathbb{R}^{2d}} \Big(\partial_t f + \gamma \frac{p}{m} \cdot \nabla_q f - \gamma \nabla_q V \cdot \nabla_p f \\ & - \gamma \nabla_p f \cdot (\nabla_q \psi * \rho_t) \\ & - \gamma^2 \frac{p}{m} \cdot \nabla_p f + \gamma^2 \Delta_p f \Big) d\rho_t dt - \frac{\gamma^2}{2} \int_0^T \int_{\mathbb{R}^{2d}} \left|\nabla_p f\right|^2 d\rho_t dt. \end{aligned}$$

Define $\mathcal{A} := \{f = g \circ \xi^\gamma \text{ with } g \in C_b^{1,2}(\mathbb{R} \times \mathbb{R}^d)\}$. Then we have

$$I^\gamma(\rho^\gamma) \geq \sup_{f \in \mathcal{A}} \mathcal{J}^\gamma(\rho^\gamma, f),$$

and

$$\begin{aligned} \mathcal{J}^\gamma(\rho^\gamma, g \circ \xi^\gamma) = & \int_{\mathbb{R}^{2d}} g_T \circ \xi^\gamma d\rho_T^\gamma - \int_{\mathbb{R}^{2d}} g_0 \circ \xi^\gamma d\rho_0^\gamma \\ & - \int_0^T \int_{\mathbb{R}^{2d}} \left[\partial_t (g \circ \xi^\gamma) - \nabla_q V(q) \cdot \nabla g \left(q + \frac{p}{\gamma}\right) \right. \\ & \left. + \Delta g \left(q + \frac{p}{\gamma}\right) - \nabla g \left(q + \frac{p}{\gamma}\right) \cdot (\nabla_q \psi * \rho_t^\gamma)(q) \right] d\rho_t^\gamma dt \\ & - \frac{1}{2} \int_0^T \int_{\mathbb{R}^{2d}} \left|\nabla (g \circ \xi^\gamma)\right|^2 d\rho_t^\gamma dt. \end{aligned} \tag{39}$$

Note how the specific dependence of $\xi^\gamma(q,p) = q + p/\gamma$ on γ has caused the coefficients γ and γ^2 in the expression above to vanish. Adding and subtracting $\nabla V(q+p/\gamma) \cdot \nabla g(q+p/\gamma)$

in (39) and defining $\hat{\rho}^\gamma := \xi^\gamma_\# \rho^\gamma$, $\mathcal{J}^\gamma$ can be rewritten as

$$\begin{aligned}\mathcal{J}^\gamma(\rho, g\circ\xi^\gamma) = &\int_{\mathbb{R}^d} g_T d\hat{\rho}^\gamma_T - \int_{\mathbb{R}^d} g_0 d\hat{\rho}^\gamma_0 - \int_0^T \int_{\mathbb{R}^d} (\partial_t g - \nabla V\cdot\nabla g + \Delta g)(\zeta)\hat{\rho}^\gamma_t(d\zeta)dt\\ &- \frac{1}{2}\int_0^T \int_{\mathbb{R}^d} |\nabla g|^2 d\hat{\rho}^\gamma_t dt\\ &- \int_0^T \int_{\mathbb{R}^{2d}} \left(\nabla V\left(q+\frac{p}{\gamma}\right) - \nabla V(q)\right)\cdot\nabla g\left(q+\frac{p}{\gamma}\right) d\rho^\gamma_t dt\\ &+ \int_0^T \int_{\mathbb{R}^{2d}} \nabla g\left(q+\frac{p}{\gamma}\right)\cdot(\nabla_q\psi * \rho^\gamma_t)(q) d\rho^\gamma_t dt.\end{aligned} \tag{40}$$

We now show that (40) converges to the right-hand side of (37), term by term. Since $\xi^\gamma_\#\rho^\gamma \to \xi_\#\rho = \sigma$ narrowly in $\mathcal{M}([0,T]\times\mathbb{R}^{2d})$ and $g \in C^{1,2}_b(\mathbb{R}\times\mathbb{R}^d)$ we have

$$\begin{aligned}&\int_0^T \int_{\mathbb{R}^d} \left(\partial_t g - \nabla V\cdot\nabla g + \Delta g + \frac{1}{2}|\nabla g|^2\right) d\hat{\rho}^\gamma_t dt \xrightarrow{\gamma\to\infty} \int_0^T \int_{\mathbb{R}^d}\\ &\left(\partial_t g - \nabla V\cdot\nabla g + \Delta g + \frac{1}{2}|\nabla g|^2\right) d\sigma_t dt.\end{aligned}$$

Taylor expansion of ∇V around q and estimate (29) give

$$\begin{aligned}&\left|\int_0^T \int_{\mathbb{R}^{2d}} \left(\nabla V\left(q+\frac{p}{\gamma}\right) - \nabla V(q)\right)\cdot\nabla g\left(q+\frac{p}{\gamma}\right) d\rho^\gamma_t dt\right| \le\\ &\le \|D^2 V\|_\infty \|\nabla g\|_\infty \sqrt{T}\left(\int_0^T \int_{\mathbb{R}^{2d}} \frac{p^2}{\gamma^2} d\rho^\gamma_t dt\right)^{1/2} \le \frac{C}{\gamma} \xrightarrow{\gamma\to\infty} 0.\end{aligned}$$

Adding and subtracting $\nabla g(q)\cdot(\nabla_q\psi * \rho^\gamma_t)(q)$ in (40) we find

$$\begin{aligned}&\int_0^T \int_{\mathbb{R}^{2d}} \nabla g\left(q+\frac{p}{\gamma}\right)\cdot(\nabla_q\psi * \rho^\gamma_t)(q) d\rho^\gamma_t dt = \int_0^T \int_{\mathbb{R}^{2d}} \nabla g(q)\cdot(\nabla_q\psi * \rho^\gamma_t)(q) d\rho^\gamma_t dt\\ &+ \int_0^T \int_{\mathbb{R}^{2d}} \left[\nabla g\left(q+\frac{p}{\gamma}\right) - \nabla g(q)\right]\cdot(\nabla_q\psi * \rho^\gamma_t)(q) d\rho^\gamma_t dt.\end{aligned}$$

Since $\rho^\gamma \to \rho$ we have $\rho^\gamma\otimes\rho^\gamma \to \rho\otimes\rho$ and therefore passing to the limit in the first term and using the local-equilibrium characterization of Lemma 2.5, we obtain

$$\int_0^T \int_{\mathbb{R}^{2d}} \nabla g(q)\cdot(\nabla_q\psi * \rho^\gamma)(q)\, d\rho^\gamma_t dt \xrightarrow{\gamma\to 0} \int_0^T \int_{\mathbb{R}^d} \nabla g\cdot(\nabla\psi * \sigma)\, d\sigma_t dt.$$

For the second term we calculate

$$\begin{aligned}&\left|\int_0^T \int_{\mathbb{R}^{2d}} \left[\nabla g\left(q+\frac{p}{\gamma}\right) - \nabla g(q)\right]\cdot(\nabla_q\psi * \rho^\gamma)(q) d\rho^\gamma_t dt\right| \le\\ &\le \|D^2 g\|_\infty \|\nabla_q\psi\|_\infty \sqrt{T}\left(\int_0^T \int_{\mathbb{R}^{2d}} \frac{p^2}{\gamma^2} d\rho^\gamma_t dt\right)^{1/2} \le \frac{C}{\gamma} \xrightarrow{\gamma\to\infty} 0.\end{aligned}$$

Therefore

$$\int_0^T \int_{\mathbb{R}^{2d}} \nabla g\left(q+\frac{p}{\gamma}\right)\cdot(\nabla_q\psi * \rho^\gamma)(q) d\rho^\gamma_t dt \xrightarrow{\gamma\to\infty} \int_0^T \int_{\mathbb{R}^d} \nabla g\cdot(\nabla\psi * \sigma)\, d\sigma_t dt.$$

□

2.6 Discussion

The ingredients of the convergence proof above are, as mentioned before, (a) a compactness result, (b) a local-equilibrium result, and (c) a liminf inequality. All three follow from the large-deviation structure, through the rate functional I^γ. We now comment on these.

Compactness Compactness in the sense of measures is, both for ρ^γ and for $\xi^\gamma_\# \rho^\gamma$, a simple consequence of the confinement provided by the growth of H. In Theorem 2.4 we provide a stronger statement for $\xi^\gamma_\# \rho^\gamma$, by showing continuity in time, in order for the limiting functional $I(\sigma)$ in (37) to be well defined. This continuity depends on the boundedness of I^γ.

Local equilibrium The local-equilibrium statement depends crucially on the structure of I^γ, and more specifically on the large coefficient γ^2 multiplying the derivatives in p. This coefficient also ends up as a prefactor of the relative Fisher Information in the *a priori* estimate (28), and through this estimate it drives the local-equilibrium result.

Liminf inequality As remarked in the introduction, the duality structure of I^γ is the key to the liminf inequality, as it allows for relatively weak convergence of ρ^γ and $\xi^\gamma_\# \rho^\gamma$. The role of the local equilibrium is to allow us to replace the p-dependence in some of the integrals by the Maxwellian dependence, and therefore to reduce all terms to dependence on the macroscopic information $\xi^\gamma_\# \rho^\gamma$ only.

As we have shown, the choice of the coarse-graining map has the advantage that it has caused the (large) coefficients γ and γ^2 in the expression of the rate functionals to vanish. In other words, it cancels out the inertial effects and transforms a Laplacian in p variable to a Laplacian in the coarse-grained variable while rescaling it to be of order 1. The choice $\xi(q, p) = q$, on the other hand, would lose too much information by completely discarding the diffusion.

3 Diffusion on a graph in one dimension

In this section we derive the small-noise limit of a randomly perturbed Hamiltonian system, which corresponds to passing to the limit $\varepsilon \to 0$ in (14). In terms of a rescaled time, in order to focus on the time scale of the noise, Eq. (14) becomes

$$\partial_t \rho^\varepsilon = -\frac{1}{\varepsilon} \operatorname{div}(\rho^\varepsilon J \nabla H) + \Delta_p \rho^\varepsilon. \tag{41}$$

Here $\rho^\varepsilon \in C([0, T], \mathcal{P}(\mathbb{R}^2))$, $J = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ is again the canonical symplectic matrix, Δ_p is the Laplacian in the p-direction, and the equation holds in the sense of distributions. The Hamiltonian $H \in C^2(\mathbb{R}^{2d}; \mathbb{R})$ is again defined by $H(q, p) = p^2/2m + V(q)$ for some potential $V : \mathbb{R}^d \to \mathbb{R}$. We make the following assumptions (that we formulate on H for convenience):

(A1) $H \geq 0$, and H is coercive, i.e. $H(x) \xrightarrow{|x| \to \infty} \infty$;
(A2) $|\nabla H|, |\Delta H|, |\nabla_p H|^2 \leq C(1 + H)$;
(A3) H has a finite number of non-degenerate (i.e. non-singular Hessian) saddle points $O_1, \ldots, O_n$ with $H(O_i) \neq H(O_j)$ for every $i, j \in \{1, \ldots, n\}$, $i \neq j$.

As explained in the introduction, and in contrast to the VFP equation of the previous section, Eq. (41) has two equally valid interpretations: as a PDE in its own right, or as the Fokker-Planck (forward Kolmogorov) equation of the stochastic process

$$X^\varepsilon = \begin{pmatrix} Q^\varepsilon \\ P^\varepsilon \end{pmatrix}, \qquad dX_t^\varepsilon = \frac{1}{\varepsilon} J \nabla H(X_t^\varepsilon) dt + \sqrt{2} \begin{pmatrix} 0 \\ 1 \end{pmatrix} dW_t. \tag{42}$$

For the sequel we will think of ρ^ε as the law of the process X_t^ε; although this is not strictly necessary, it helps in illustrating the ideas.

3.1 Construction of the graph Γ

As mentioned in the introduction, the dynamics of (41) has two time scales when $0 < \varepsilon \ll 1$, a fast and a slow one. The fast time scale, of scale ε, is described by the (deterministic) equation

$$\dot{x} = \frac{1}{\varepsilon} J \nabla H(x) \qquad \text{in } \mathbb{R}^2, \tag{43}$$

whereas the slow time scale, of order 1, is generated by the noise term.

The solutions of (43) follow level sets of H. There exist three types of such solutions: stationary ones, periodic orbits, and homoclinic orbits. Stationary solutions of (43) correspond to stationary points of H (where $\nabla H = 0$); periodic orbits to connected components of level sets along which $\nabla H \neq 0$; and homoclinic orbits to components of level sets of H that are terminated on each end by a stationary point. Since we have assumed in (A3) that there is at most one stationary point in each level sets, heteroclinic orbits do not exist, and the orbits necessarily connect a stationary point with itself.

Looking ahead towards coarse-graining, we define Γ to be the set of all connected components of level sets of H, and we identify Γ with a union of one-dimensional line segments, as shown in Fig. 3. Each periodic orbit corresponds to an interior point of one of the edges of Γ; the vertices of Γ correspond to connected components of level sets containing a stationary point of H. Each saddle point O corresponds to a vertex connected by three edges.

For practical purposes we also introduce a coordinate system on Γ. We represent the edges by closed intervals $I_k \subset \mathbb{R}$, and number them with numbers $k = 1, 2, \ldots, n$; the pair (h, k) is then a coordinate for a point $\gamma \in \Gamma$, if k is the index of the edge containing γ, and h the value of H on the level set represented by γ. For a vertex $O \in \Gamma$, we write $O \sim I_k$ if O

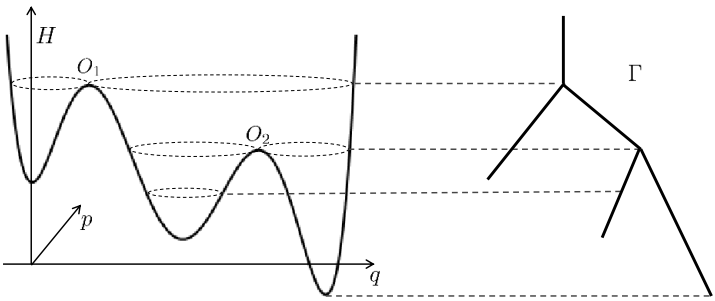


Fig. 3 Left: Hamiltonian $\mathbb{R}^2 \ni (q, p) \mapsto H(q, p)$, Right: Graph Γ

is at one end of edge I_k; we use the shorthand notation $\pm_{kj}$ to mean 1 if O_j is at the upper end of I_k, and -1 in the other case. Note that if $O \sim I_{k_1}$, $O \sim I_{k_2}$ and $O \sim I_{k_3}$ and h_0 is the value of H at the point corresponding to O, then the coordinates (h_0, k_1), (h_0, k_2) and (h_0, k_3) correspond to the same point O. With a slight abuse of notation, we also define the function $k : \mathbb{R}^2 \to \{1, \ldots, n\}$ as the index of the edge $I_k \subset \Gamma$ corresponding to the component containing (q, p).

The rigorous construction of the graph Γ and the topology on it has been done several times [15,36,37]; for our purposes it suffices to note that (a) inside each edge, the usual topology and geometry of $\mathbb{R}^1$ apply, and (b) across the whole graph there is a natural concept of distance, and therefore of continuity. It will be practical to think of functions $f : \Gamma \to \mathbb{R}$ as defined on the disjoint union $\sqcup_k I_k$. A function $f : \Gamma \to \mathbb{R}$ is then called well-defined if it is a single-valued function on Γ (i.e., it takes the same value on those vertices that are multiply represented). A well-defined function $f : \Gamma \to \mathbb{R}$ is *continuous* if $f|_{I_k} \in C(I_k)$ for every k.

We also define a concept of *differentiability* of a function $f : \Gamma \to \mathbb{R}$. A *subgraph* of Γ is defined as any union of edges such that each interior vertex connects exactly two edges, one from above and one from below—i.e., a subtree without bifurcations. A continuous function on Γ is called differentiable on Γ if it is differentiable on each of its subgraphs.

Finally, in order to integrate over Γ, we write $d\gamma$ for the measure on Γ which is defined on each I_k as the local Lebesgue measure dh. Whenever we write $\int_\Gamma$, this should be interpreted as $\sum_k \int_{I_k}$.

3.2 Adding noise: diffusion on the graph

In the noisy evolution (42), for small but finite $\varepsilon > 0$, the evolution follows fast trajectories that nearly coincide with the level sets of H; the noise breaks the conservation of H, and causes a slower drift of X_t across the levels of H. In order to remove the fast deterministic dynamics, we now define the coarse-graining map as

$$\xi : \mathbb{R}^2 \to \Gamma, \quad \xi(q, p) := (H(q, p), k(q, p)), \tag{44}$$

where the mapping $k : \mathbb{R}^2 \to \{1, \ldots, n\}$ indexes the edges of the graph, as above.

We now consider the process $\xi(X_t^\varepsilon)$, which contains no fast dynamics. For each finite $\varepsilon > 0$, $\xi(X_t^\varepsilon)$ is not a Markov process; but as $\varepsilon \to 0$, the fast movement should result in a form of averaging, such that the influence of the missing information vanishes; then the limit process is a diffusion on the graph Γ.

The results of this section are stated and proved in terms of the corresponding objects ρ^ε and $\hat{\rho}^\varepsilon$, where $\hat{\rho}^\varepsilon$ is the push-forward

$$\hat{\rho}^\varepsilon := \xi_\# \rho^\varepsilon, \tag{45}$$

as explained in Sect. 1.1, and similar to Sect. 2. The corresponding statement about ρ^ε and $\hat{\rho}^\varepsilon$ is that $\hat{\rho}^\varepsilon$ should converge to some $\hat{\rho}$, which in the limit satisfies a (convection-) diffusion equation on Γ. Theorems 3.2 and 3.6 make this statement precise.

3.3 Compactness

As in the case of the VFP equation, Eq. (41) has a free energy, which in this case is simply the Boltzmann entropy

$$\mathcal{F}(\rho) = \int_{\mathbb{R}^2} \rho \log \rho \, \mathcal{L}^2, \tag{46}$$

where $\mathcal{L}^2$ denotes the two dimensional Lebesgue measure in $\mathbb{R}^2$.

The corresponding 'relative' Fisher Information is the same as the Fisher Information in the p-variable,

$$\mathcal{I}(\rho|\mathcal{L}^2) = \sup_{\varphi \in C_c^\infty(\mathbb{R}^2)} 2 \int_{\mathbb{R}^2} \left[\Delta_p \varphi - \frac{1}{2} |\nabla_p \varphi|^2 \right] d\rho, \tag{47}$$

and satisfies for $\rho = f\mathcal{L}^2$,

$$\mathcal{I}(f\mathcal{L}^2|\mathcal{L}^2) = \int_{\mathbb{R}^2} |\nabla_p \log f|^2 \, f \, dq dp,$$

whenever this is finite.

The large deviation functional $I^\varepsilon : C([0,T]; \mathcal{P}(\mathbb{R}^2)) \to \mathbb{R}$ is given by

$$I^\varepsilon(\rho) = \sup_{f \in C_c^{1,2}(\mathbb{R}\times\mathbb{R}^2)} \left[\int_{\mathbb{R}^2} f_T d\rho_T - \int_{\mathbb{R}^2} f_0 d\rho_0 - \int_0^T \int_{\mathbb{R}^2} \left(\partial_t f + \frac{1}{\varepsilon} J\nabla H \cdot \nabla f + \Delta_p f \right) d\rho_t dt \right.$$
$$\left. - \frac{1}{2} \int_0^T \int_{\mathbb{R}^2} |\nabla_p f|^2 \, d\rho_t dt \right]. \tag{48}$$

For fixed $\varepsilon > 0$, ρ^ε solves (41) iff $I^\varepsilon(\rho^\varepsilon) = 0$.

The following theorem states the relevant *a priori* estimates in this setting.

Theorem 3.1 (A priori estimates) *Let $\varepsilon > 0$ and let $\rho \in C([0,T]; \mathcal{P}(\mathbb{R}^2))$ with $\rho_t|_{t=0} =: \rho_0$ satisfy*

$$I^\varepsilon(\rho) + \mathcal{F}(\rho_0) + \int_{\mathbb{R}^2} H \, d\rho_0 \leq C.$$

Then for any $t \in [0,T]$ we have

$$\int_{\mathbb{R}^2} H\rho_t \, dt < C', \tag{49}$$

where $C' > 0$ depends on C but is independent of ε. Furthermore, for any $t \in [0,T]$ we have

$$\mathcal{F}(\rho_t) + \frac{1}{2} \int_0^t \mathcal{I}(\rho_s|\mathcal{L}^2) \, ds \leq I^\varepsilon(\rho) + \mathcal{F}(\rho_0). \tag{50}$$

See Appendix D for a proof of Theorem 3.1.

Note that the estimate (50) implies that $\mathcal{F}(\rho_t) = \mathcal{H}(\rho_t|\mathcal{L}^2)$ is finite for all t, and therefore ρ_t is Lebesgue absolutely continuous. We will often therefore write $\rho_t(x)$ for the Lebesgue density of ρ_t. In addition, the integral of the relative Fisher Information is also bounded: $0 \leq \int_0^t \mathcal{I}(\rho_s|\mathcal{L}^2) \, ds \leq C$.

The next result summarizes the compactness properties for any sequence ρ^ε with $\sup_\varepsilon I^\varepsilon(\rho^\varepsilon) < \infty$.

Theorem 3.2 (Compactness) *Let a sequence $\rho^\varepsilon \in C([0,T]; \mathcal{P}(\mathbb{R}^2))$ with $\rho^\varepsilon|_{t=0} =: \rho_0^\varepsilon$ satisfy for a constant $C > 0$ and all $\varepsilon > 0$ the estimate*

$$I^\varepsilon(\rho^\varepsilon) + \mathcal{F}(\rho_0^\varepsilon) + \int_{\mathbb{R}^2} H d\rho_0^\varepsilon \leq C.$$

Then there exist subsequences (not relabelled) such that

1. $\rho^\varepsilon \to \rho$ *in* $\mathcal{M}([0,T]\times\mathbb{R}^2)$ *in the narrow topology;*
2. $\hat\rho^\varepsilon \to \hat\rho = \xi_\#\rho$ *in* $C([0,T];\mathcal{P}(\Gamma))$ *with respect to the uniform topology in time and narrow topology on* $\mathcal{P}(\Gamma)$.

Finally, we have the estimate

$$\mathcal{F}(\rho_t) + \frac{1}{2}\int_0^t \mathcal{I}(\rho_s|\mathcal{L}^2)\,ds \le C \qquad \text{for all } t\in[0,T].$$

The sequence ρ^ε is tight in $\mathcal{M}([0,T]\times\mathbb{R}^2)$ by estimate (49), which implies Part 1. The proof of part 2 is similar to Part 2 in Theorem 2.4, and the final estimate is a direct consequence of (50).

3.4 Local equilibrium

Theorem 3.2 states that ρ^ε converges narrowly on $[0,T]\times\mathbb{R}^2$ to some ρ. In fact we need a stronger statement, in which the behaviour of ρ on each connected component of H is fully determined by the limit $\hat\rho$.

Lemma 3.3 below makes this statement precise. Before proceeding we define $T:\Gamma\to\mathbb{R}$ as

$$T(\gamma) := \int_{\xi^{-1}(\gamma)} \frac{\mathscr{H}^1(dx)}{|\nabla H(x)|}, \tag{51}$$

where $\mathscr{H}^1$ is the one-dimensional Hausdorff measure. T has a natural interpretation as the period of the periodic orbit of the deterministic Eq. (43) corresponding to γ. When γ is an interior vertex, such that the orbit is homoclinic, not periodic, $T(\gamma)=+\infty$. T also has a second natural interpretation: the measure $T(\gamma)d\gamma = T(h,k)dh$ on Γ is the push-forward under ξ of the Lebesgue measure on $\mathbb{R}^2$, and the measure $T(\gamma)d\gamma$ therefore appears in various places.

Lemma 3.3 (Local Equilibrium) *Under the assumptions of Theorem 3.2, let* $\rho^\varepsilon\to\rho$ *in* $\mathcal{M}([0,T]\times\mathbb{R}^2)$ *with respect to the narrow topology. Let* $\hat\rho$ *be the push-forward* $\xi_\#\rho$ *of the limit* ρ, *as above.*

Then for a.e. t, *the limit* ρ_t *is absolutely continuous with respect to the Lebesgue measure,* $\hat\rho_t$ *is absolutely continuous with respect to the measure* $T(\gamma)d\gamma$, *where* $T(\gamma)$ *is defined in* (51). *Writing*

$$\rho_t(dx) = \rho_t(x)dx \qquad \text{and} \qquad \hat\rho_t(d\gamma) = \alpha_t(\gamma)T(\gamma)d\gamma,$$

we have

$$\rho_t(x) = \alpha_t(\xi(x)) \qquad \text{for almost all } x\in\mathbb{R}^2 \text{ and } t\in[0,T]. \tag{52}$$

Proof From the boundedness of $I^\varepsilon(\rho^\varepsilon)$ and the narrow convergence $\rho^\varepsilon\to\rho$ we find, passing to the limit in the rate functional (48), for any $f\in C_c^{1,2}(\mathbb{R}\times\mathbb{R}^2)$

$$\int_0^T \int_{\mathbb{R}^2} J\nabla H \cdot \nabla f \, d\rho_t dt = 0. \tag{53}$$

Now choose any $\varphi \in C_c^2([0,T]\times\mathbb{R}^2)$ and any $\zeta \in C_b^2(\Gamma)$ such that ζ is constant in a neighbourhood of each vertex; then the function $f(t,x) = \zeta(\xi(x))\varphi(t,x)$ is well-defined and in $C_c^2([0,T]\times\mathbb{R}^2)$. We substitute this special function in (53); since $J\nabla H\nabla(\zeta\circ\xi) = 0$, we have $J\nabla H\nabla f = (\zeta\circ\xi)J\nabla H\nabla\varphi$. Applying the disintegration theorem to ρ, writing $\rho_t(dx) = \hat{\rho}_t(d\gamma)\tilde{\rho}_t(dx|\gamma)$ with $\operatorname{supp}\tilde{\rho}_t(\cdot|\gamma) \subset \xi^{-1}(\gamma)$, we obtain

$$\begin{aligned} 0 &= \int_0^T \int_\Gamma \zeta(\gamma)\hat{\rho}_t(d\gamma) \int_{\xi^{-1}(\gamma)} \nabla\varphi \cdot \frac{J\nabla H}{|\nabla H|} |\nabla H|\tilde{\rho}(\cdot|\gamma) d\mathscr{H}^1 \\ &= \int_0^T \int_\Gamma \zeta(\gamma)\hat{\rho}_t(d\gamma) \int_{\xi^{-1}(\gamma)} \partial_\tau\varphi |\nabla H|\tilde{\rho}(\cdot|\gamma) d\mathscr{H}^1 dt, \end{aligned}$$

where ∂_τ is the tangential derivative. By varying ζ and φ we conclude that for $\hat{\rho}$-almost every (γ,t), $|\nabla H|\tilde{\rho}_t(\cdot|\gamma) = C_{\gamma,t}$ for some γ,t-dependent constant $C_{\gamma,t} > 0$, and since $\tilde{\rho}$ is normalized, we find that

$$\text{for } \hat{\rho}\text{-a.e. } (\gamma,t):\ \tilde{\rho}_t(dx|\gamma) = \frac{1}{T(\gamma)|\nabla H(x)|}\mathscr{H}^1\lfloor_{\xi^{-1}(\gamma)}(dx). \tag{54}$$

This also implies that $\tilde{\rho}_t(\cdot|\gamma)$ is in fact t-independent.

For measurable f we now compare the two relations

$$\begin{aligned} \int_{\mathbb{R}^2} f\, d\rho_t &= \int_{\mathbb{R}^2} f(y)\rho_t(y)\, dy &&= \int_\Gamma d\gamma \int_{\xi^{-1}(\gamma)} \frac{f(y)}{|\nabla H(y)|}\rho_t(y)\mathscr{H}^1(dy) \\ \int_{\mathbb{R}^2} f\, d\rho_t &= \int_\Gamma \hat{\rho}_t(d\gamma) \int_{\xi^{-1}(\gamma)} f(y)\tilde{\rho}(dy|\gamma) &&= \int_\Gamma \frac{\hat{\rho}_t(d\gamma)}{T(\gamma)} \int_{\xi^{-1}(\gamma)} \frac{f(y)}{|\nabla H(y)|}\mathscr{H}^1(dy) \end{aligned}$$

where we have used the co-area formula in the first line and (54) in the second one. Since f was arbitrary, (52) follows for almost all t. □

3.5 Continuity of ρ and $\hat{\rho}$

As a consequence of the local-equilibrium property (52) and the boundedness of the Fisher Information, we will show in the following that ρ and its push-forward $\hat{\rho}$ satisfy an important continuity property. We first motivate this property heuristically.

The local-equilibrium result Lemma 3.3 states that the limit measure ρ depends on x only through $\xi(x)$. Take any measure $\rho \in \mathcal{P}(\mathbb{R}^2)$ of that form, i.e. $\rho(dx) = f(\xi(x))dx$, with finite free energy and finite relative Fisher Information. Setting $\tilde{f} = f\circ\xi$, by Lemma 2.1, $\nabla_p\tilde{f}$ is well-defined and locally integrable.

Consider a section Ω_ε of the (q,p)-plane as shown in Fig. 4, bounded by $q=a$ and $q=b$ and level sets $H=h$ and $H=h+\varepsilon$. The top and bottom boundaries γ and γ_ε correspond to elements of Γ that we also call γ and γ_ε; they might be part of the same edge k of the graph, or they might belong to different edges. As $\varepsilon\to 0$, γ_ε converges to γ.

By simple integration we find that

$$\int_{\Omega_\varepsilon} \nabla_p\tilde{f} = \int_{\gamma_\varepsilon\cup\gamma} \tilde{f} n_p\, dr = (f(\gamma_\varepsilon) - f(\gamma))(b-a),$$

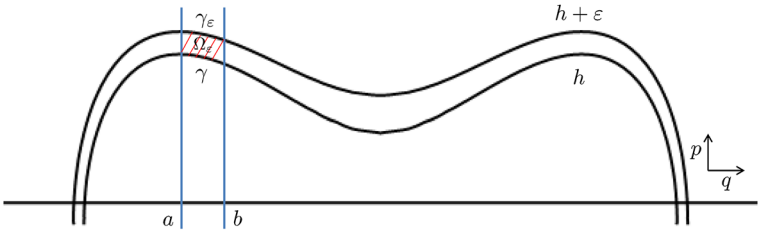


Fig. 4 Section Ω in which $H^{-1}(h)$ is transverse to p

where dr is the scalar line element and n_p the p-component of the normal n. Applying Hölder's inequality we find

$$|b-a|\,|f(\gamma_\varepsilon)-f(\gamma)| = \left|\int_{\Omega_\varepsilon} \nabla_p \rho\right| \le \left(\int_{\Omega_\varepsilon} \frac{1}{\rho}|\nabla_p \rho|^2\right)^{\frac{1}{2}} \left(\int_{\Omega_\varepsilon} \rho\right)^{\frac{1}{2}} \xrightarrow{\varepsilon\to 0} 0.$$

This argument shows that f is continuous from the right at the point $\gamma \in \Gamma$.

The following lemma generalizes this argument to the case at hand, in which ρ also depends on time. Note that Int Γ is the interior of the graph Γ, which is Γ without the lower exterior vertices.

Lemma 3.4 (Continuity of ρ) *Let $\rho \in \mathcal{P}([0,T]\times\mathbb{R}^2)$, $\rho(dtdx) = f(t,\xi(x))dtdx$ for a Borel measurable $f : [0,T]\times\Gamma \to \mathbb{R}$, and assume that*

$$\int_0^T \mathcal{I}(\rho_t|\mathcal{L}^2)\,dt + \sup_{t\in[0,T]} \mathcal{F}(\rho_t) < \infty.$$

Then for almost all $t \in [0,T]$, $\gamma \mapsto f(t,\gamma)$ is continuous on Int Γ.

Proof The argument is essentially the same as the one above. For almost all t, ρ_t is Lebesgue-absolutely-continuous and $\mathcal{I}(\rho_t|\mathcal{L})$ is finite, and the argument above can be applied to the neighbourhood of any point x with $\nabla H(x) \neq 0$, and to both right and left limits. The only elements of Γ that have no representative $x \in \mathbb{R}^2$ with $\nabla H(x) \neq 0$ are the lower ends of the graph, corresponding to the bottoms of the wells of H. At all other points of Γ we obtain continuity.

Corollary 3.5 (Continuity of $\hat\rho$) *Let ρ be the limit given by Theorem 3.2, and $\hat\rho := \xi_\# \rho$ its push-forward. For almost all t, $\hat\rho_t \ll T(\gamma)d\gamma$, and $d\hat\rho_t/T(\gamma)d\gamma$ is continuous on* Int Γ.

This corollary follows by combining Lemma 3.4 with Lemma 3.3.

3.6 Liminf inequality

We now derive the final ingredient of the proof, the liminf inequality. Define

$$\hat I(\hat\rho) := \begin{cases} \sup\limits_{g\in C_c^{1,2}(\mathbb{R}\times\Gamma)} \hat{\mathcal{J}}(\hat\rho, g) & \text{if } \hat\rho_t \ll T(\gamma)d\gamma,\ \hat\rho_t(d\gamma) = f_t(\gamma)T(\gamma)d\gamma \text{ with } f \text{ continuous on Int } \Gamma, \\ & \text{for almost all } t\in[0,T], \\ +\infty & \text{otherwise,} \end{cases} \tag{55}$$

where

$$\hat{\mathcal{J}}(\hat{\rho}, g) := \int_\Gamma g_T d\hat{\rho}_T - \int_\Gamma g_0 d\hat{\rho}_0 - \int_0^T \int_\Gamma \big(\partial_t g_t(\gamma) + A(\gamma) g_t''(\gamma) + B(\gamma) g_t'(\gamma)\big)\hat{\rho}_t(d\gamma)dt$$
$$-\frac{1}{2}\int_0^T \int_\Gamma A(\gamma)(g_t'(\gamma))^2 \hat{\rho}_t(d\gamma)dt, \tag{56}$$

and we use g' and g'' to indicate derivatives with respect to h. For $\gamma \in \Gamma$, the coefficients are defined by

$$A(\gamma) := \frac{1}{T(\gamma)} \int_{\xi^{-1}(\gamma)} \frac{(\nabla_p H)^2}{|\nabla H|} d\mathscr{H}^1, \quad B(\gamma) := \frac{1}{T(\gamma)} \int_{\xi^{-1}(\gamma)} \frac{\Delta_p H}{|\nabla H|} d\mathscr{H}^1,$$
$$T(\gamma) := \int_{\xi^{-1}(\gamma)} \frac{1}{|\nabla H|} d\mathscr{H}^1. \tag{57}$$

Note that for our particular choice of $H(q, p) = p^2/2m + V(q)$, we have $B(\gamma) = 1/m$.

The class of test functions in (55) is $C_c^{1,2}(\mathbb{R} \times \Gamma)$; recall that differentiability of a function $f : \Gamma \to \mathbb{R}$ is defined by restriction to one-dimensional subgraphs, and $C_c^{1,2}(\mathbb{R} \times \Gamma)$ therefore consists of functions $g : \Gamma \to \mathbb{R}$ that are twice continuously differentiable in h in this sense. The subscript c indicates that we restrict to functions that vanish for sufficiently large h (i.e. somewhere along the top edge of Γ).

Note that again $\hat{I} \geq 0$; formally, $\hat{I}(\hat{\rho}) = 0$ iff $\hat{\rho}$ satisfies the diffusion equation

$$\partial_t \hat{\rho} = (A\hat{\rho})'' - (B\hat{\rho})',$$

and we will investigate this equation in more detail in the next section.

Theorem 3.6 (Liminf inequality) *Under the same assumptions as in Theorem 3.2, let* $\rho^\varepsilon \to \rho$ *in* $\mathcal{M}([0, T]; \mathbb{R}^2)$ *and* $\hat{\rho}^\varepsilon := \xi_\# \rho^\varepsilon \to \xi_\# \rho =: \hat{\rho}$ *in* $C([0, T]; \mathcal{P}(\Gamma))$. *Then*

$$\liminf_{\varepsilon \to 0} I^\varepsilon(\rho^\varepsilon) \geq \hat{I}(\hat{\rho}).$$

Proof Recall the rate functional from (48)

$$I^\varepsilon(\rho^\varepsilon) = \sup_{f \in C_c^{1,2}(\mathbb{R} \times \mathbb{R}^2)} \mathcal{J}^\varepsilon(\rho^\varepsilon, f), \qquad \text{where} \tag{58}$$
$$\mathcal{J}^\varepsilon(\rho^\varepsilon, f) := \int_{\mathbb{R}^2} f_T d\rho_T^\varepsilon - \int_{\mathbb{R}^2} f_0 d\rho_0^\varepsilon - \int_0^T \int_{\mathbb{R}^2} \Big(\partial_t f + \frac{1}{\varepsilon} J\nabla H \cdot \nabla f + \Delta_p f\Big)$$
$$d\rho_t^\varepsilon dt - \frac{1}{2}\int_0^T \int_{\mathbb{R}^2} |\nabla_p f|^2 d\rho_t^\varepsilon dt.$$

Define $\hat{\mathcal{A}} := \Big\{ f = g \circ \xi : \ g \in C_c^{1,2}(\mathbb{R} \times \Gamma) \Big\}$. Then we have

$$I^\varepsilon(\rho^\varepsilon) \geq \sup_{f \in \hat{\mathcal{A}}} \mathcal{J}^\varepsilon(\rho^\varepsilon, f).$$

Since $J\nabla H \nabla(g \circ \xi) = 0$, upon substituing $f = g \circ \xi$ into $\mathcal{J}^\varepsilon$ the $O(1/\varepsilon)$ term vanishes. Using the notation g' for the partial derivative with respect to h, $\partial_t g$ for the time derivative, and suppressing the dependence of g on time, we find

$$\mathcal{J}^\varepsilon(\rho^\varepsilon, g\circ\xi) := \int_\Gamma g_T d\hat{\rho}_T^\varepsilon - \int_\Gamma g_0 d\hat{\rho}_0^\varepsilon - \int_0^T \int_{\mathbb{R}^2} \Big(\partial_t g(\xi(x)) + g''(\xi(x))(\nabla_p H(x))^2 + g'(\xi(x))\Delta_p H(x) \Big) \rho_t^\varepsilon(dx)dt - \frac{1}{2}\int_0^T \int_{\mathbb{R}^2} |g'(\xi(x))\nabla_p H(x)|^2 \rho_t^\varepsilon(dx)dt. \tag{59}$$

The limit of (59) is determined term by term. Taking the fourth term as an example, using the co-area formula and the local-equilibrium result of Lemma 3.3, the fourth term on the right-hand side of (59) gives

$$\int_0^T \int_{\mathbb{R}^2} g''(\xi(x))(\nabla_p H(x))^2 \rho_t^\varepsilon(dx)dt \xrightarrow{\varepsilon\to 0} \int_0^T \int_{\mathbb{R}^2} g''(\xi(x))(\nabla_p H(x))^2 \rho_t(dx)dt$$
$$= \int_0^T dt \int_\Gamma \frac{g''(\gamma)\hat{\rho}_t(d\gamma)}{T(\gamma)} \Big(\int_{\xi^{-1}(\gamma)} \frac{(\nabla_p H(y))^2}{|\nabla H(y)|} \mathscr{H}^1(dy) \Big) = \int_0^T \int_\Gamma A(\gamma) g''(\gamma) \hat{\rho}_t(d\gamma)dt,$$

where $A : \Gamma \to \mathbb{R}$ is defined in (57). Proceeding similarly with the other terms we find

$$\liminf_{\varepsilon\to 0} I^\varepsilon(\rho^\varepsilon) \geq \sup_{g\in C_c^{1,2}(\mathbb{R}\times\Gamma)} \hat{\mathcal{J}}(\hat{\rho}, g). \tag{60}$$

This concludes the proof of Theorem 3.6. □

3.7 Study of the limit problem

We now investigate the limiting functional $\hat{I}$ from (55) a little further. The two main results of this section are that $\hat{\mathcal{J}}$ can be written as

$$\hat{\mathcal{J}}(\hat{\rho}, g) = \int_\Gamma g_T d\hat{\rho}_T - \int_\Gamma g_0 d\hat{\rho}_0 - \int_0^T \int_\Gamma \Big[\partial_t g_t \, d\hat{\rho}_t + \Big((TA\, g_t')' + \frac{1}{2} TA\, g_t'^2 \Big) \frac{d\hat{\rho}_t}{T} \Big] dt, \tag{61}$$

and that $\hat{I}$ satisfies

$$\hat{I}(\hat{\rho}) \geq \sup_{g\in\mathcal{A}} \hat{\mathcal{J}}(\hat{\rho}, g) \qquad \text{for all } \hat{\rho} \in C([0,T]; \mathcal{P}(\Gamma)), \tag{62}$$

where $\mathcal{A}$ is the larger class

$$\mathcal{A} := \Big\{ g : C^{1,0}(\mathbb{R}\times\Gamma) : g\big|_{I_k} \in C_b^{1,2}(\mathbb{R}\times I_k), \quad \forall \text{ interior vertex } O_j \ \forall t : \sum_{k: I_k \sim O_j} \pm_{kj} \, g_t'(O_j, k)\, TA(O_j, k) = 0 \Big\}. \tag{63}$$

The admissible set $\mathcal{A}$ relaxes the conditions on g at interior vertices: instead of requiring g to have identical derivatives coming from each edge, only a single scalar combination of the derivatives has to vanish. (In fact it can be shown that equality holds in (62), but that requires a further study of the limiting equation that takes us too far here.)

Both results use some special properties of T, A, and B, which are given by the following lemma. In this lemma and below we use TA and TB for the functions obtained by multiplying T with A and B; these combinations play a special role, and we treat them as separate functions.

Lemma 3.7 (Properties of TA and TB) *The functions TA and TB have the following properties.*

1. *$TA \in C^1(I_k)$ for each k, and $(TA)' = TB$;*
2. *TA is bounded on compact subsets of Γ;*
3. *At each interior vertex O_j, for each k such that $I_k \sim O_j$, $TA(O_j,k) := \lim_{\substack{h\in I_k \\ h\to O_j}} TA(h,k)$ exists, and*

$$\sum_{k:I_k\sim O_j} \pm_{kj}\, TA(O_j,k) = 0. \tag{64}$$

From this lemma the expression (61) follows by simple manipulation.

With these two results, we can obtain a differential-equation characterization of those $\hat\rho$ with $\hat I(\hat\rho) = 0$. Assume that a $\hat\rho$ with $\hat I(\hat\rho) = 0$ is given. By rescaling we find that for all $g \in \mathcal{A}$,

$$\int_\Gamma g_T\, d\hat\rho_T - \int_\Gamma g_0 d\hat\rho_0 = \int_0^T \int_\Gamma \Big[\partial_t g_t\, d\hat\rho + (TA\, g_t')' \frac{d\hat\rho_t}{T}\Big] dt. \tag{65}$$

As already remarked we find a parabolic equation inside each edge of Γ,

$$\partial_t \hat\rho_t = \Big(TA\,\big(\frac{\hat\rho_t}{T}\big)'\Big)' = (A\hat\rho_t)'' - (B\hat\rho_t)'. \tag{66}$$

We next determine the boundary and connection conditions at the vertices.

Consider a single interior vertex O_j, and choose a function $g \in \mathcal{A}$ such that $\operatorname{supp} g$ contains no other vertices. Writing $\hat\rho_t(d\gamma) = f_t(\gamma)T(\gamma)d\gamma$ we find first that f_t is continuous at O_j, by the definition (55) of $\hat I$. Then, assuming that $\hat\rho$ is smooth enough for the following expressions to make sense[1], we perform two partial integrations in γ and one in time on (65) and substitute (66) to find

$$\begin{aligned} 0 \;&=\; \int_0^T f_t(O_j) \sum_{k:I_k\sim O_j} \pm_{kj}\, TA(O_j,k) g_t'(O_j,k)\, dt \\ &\quad - \int_0^T g_t(O_j) \sum_{k:I_k\sim O_j} \pm_{kj}\, TA(O_j,k) f_t'(O_j,k)\, dt. \end{aligned}$$

The first term vanishes since $g \in \mathcal{A}$, while the second term leads to the connection condition

$$\text{at each interior vertex } O_j: \quad \sum_{k:I_k\sim O_j} \pm_{kj}\, TA(O_j,k) f_t'(O_j,k) = 0.$$

The lower exterior vertices and the top vertex are *inaccessible*, in the language of [30,50], and therefore require no boundary condition. Summarizing, we find that if $\hat I(\hat\rho) = 0$, then $\hat\rho =: fTd\gamma$ satisfies a weak version of Eq. (66) with connection conditions

$$\text{at each interior vertex } O_j: \quad f \text{ is continuous and } \sum_{k:I_k\sim O_j} \pm_{kj}\, TA(O_j,k) f_t'(O_j,k) = 0.$$

This combination of equation and boundary conditions can be proved to characterize a well-defined semigroup using e.g. the Hille–Yosida theorem and the characterization of one-dimensional diffusion processes by Feller (e.g. [30]).

We now prove the inequality (62).

[1] This can actually be proved using the properties of A and B near the vertices and applying standard parabolic regularity theory on each of the edges.

Lemma 3.8 (Comparison of $\hat{I}$ and $\tilde{I}$) *We have*

$$\hat{I}(\hat{\rho}) \geq \tilde{I}(\hat{\rho}) := \sup_{g \in \mathcal{A}} \hat{\mathcal{J}}(\hat{\rho}, g).$$

Proof Take $\hat{\rho}$ such that $\hat{I}(\hat{\rho}) < \infty$, implying that $\hat{\rho}_t(d\gamma) = f_t(\gamma)T(\gamma)d\gamma$ with f_t continuous on Int Γ for almost all t. Choose $g \in \mathcal{A}$; we will show that $\hat{I}(\hat{\rho}) \geq \hat{\mathcal{J}}(\hat{\rho}, g)$, thus proving the lemma. For simplicity we only treat the case of a single interior vertex, called O; the case of multiple vertices is a simple generalization. For convenience we also assume that O corresponds to $h = 0$.

Define

$$g_{\delta,t}(h,k) = g_t(h,k)\zeta_\delta(h) + (1 - \zeta_\delta(h))g_t(0), \tag{67}$$

where ζ_δ is a sequence of smooth functions such that

- ζ_δ is identically zero in a δ-neighbourhood of O, and identically 1 away from a 2δ-neighbourhood of O;
- ζ_δ satisfies the growth conditions $|\zeta_\delta'| \leq 2/\delta$ and $|\zeta_\delta''| \leq 4/\delta^2$.

We calculate $\hat{\mathcal{J}}(\hat{\rho}, g_\delta)$. The limit of the first three terms is straightforward: by dominated convergence we obtain

$$\int_\Gamma g_{\delta,T}\, d\hat{\rho}_T - \int_\Gamma g_{\delta,0} d\hat{\rho}_0 - \int_0^T \int_\Gamma \partial_t g_{\delta,t}\, d\hat{\rho}_t \xrightarrow{\delta \to 0} \int_\Gamma g_T d\hat{\rho}_T - \int_\Gamma g_0 d\hat{\rho}_0 - \int_0^T \int_\Gamma \partial_t g_t\, d\hat{\rho}_t.$$

Next consider the term

$$\int_0^T \int_\Gamma A(\gamma) g_\delta''(\gamma) \hat{\rho}_t(d\gamma)dt = \int_0^T \int_\Gamma \Big[g''(h,k)\zeta_\delta(h) + 2\zeta_\delta'(h)g'(h,k) + \zeta_\delta''(h)\big[hg'(0,k) + O(h^2)\big] \Big] A(\gamma)\hat{\rho}_t(d\gamma)dt. \tag{68}$$

Since the function $(\gamma, t) \mapsto A(\gamma)g_t''(\gamma) \in L^\infty(\hat{\rho}_t)$ the first term in (68) again converges by dominated convergence:

$$\int_0^T \int_\Gamma g_t''(h,k)\zeta_\delta(h)A(h,k)\hat{\rho}_t(d\gamma)dt \xrightarrow{\delta \to 0} \int_0^T \int_\Gamma g_t''(h,k)A(h,k)\hat{\rho}_t(d\gamma)dt.$$

Abbreviate $f_t(\gamma)TA(\gamma)$ as $a(\gamma)$; note that a is continuous and bounded in a neighbourhood of O. Write the second term on the right-hand side in (68) as (supressing the time integral for the moment)

$$\begin{aligned} 2\int_\Gamma \zeta_\delta'(h)g'(h,k)a(h,k)dh &= 2\int_\Gamma \zeta_\delta'(h)g'(h,k)\big(a(h,k) - a(0,k)\big)d\gamma \\ &\quad + 2\sum_k a(0,k)\int_{I_k} \zeta_\delta'(h)\big(g'(h,k) - g'(0,k)\big)dh \\ &\quad + 2\sum_k a(0,k)g'(0,k)\int_{\Gamma_k} \zeta_\delta'(h)dh \\ &\xrightarrow{\delta \to 0} 0 + 0 - 2\sum_{k: I_k \sim O} \pm_{kO}\, g'(0,k)\, a(0,k) \end{aligned}$$

$$= 2 \sum_{k:I_k \sim O} \pm_{kO}\, g'(0,k)\, f(0,k)\, TA(0,k).$$

The limit above holds since $-\zeta_\delta'(\cdot,k)$ converges weakly to a signed Dirac, $\pm_{kO}\delta_0$, as $\delta \to 0$. Proceeding similarly with the remaining terms we have

$$\begin{aligned}\hat{I}(\hat{\rho}) \geq \hat{\mathcal{J}}(\hat{\rho}, g_\delta) \xrightarrow{\delta\to 0} & \int_\Gamma g_T d\hat{\rho}_T - \int_\Gamma g_0 d\hat{\rho}_0 - \int_0^T \int_\Gamma \big(\partial_t g_t + A(\gamma) g_t''(\gamma) \\ & + B(\gamma) g_t'(\gamma)\big)\hat{\rho}_t(d\gamma)dt \\ & - \frac{1}{2}\int_0^T \int_\Gamma A(\gamma) g_t'(\gamma)^2 \hat{\rho}_t(d\gamma)dt - \int_0^T f_t(0,k)\Big[\sum_{k:I_k\sim O} \pm_{kO} TA(0,k) g_t'(0,k)\Big]dt.\end{aligned}$$

Note that the final term vanishes by the requirement that $g \in \mathcal{A}$, and therefore the right-hand side above equals $\hat{\mathcal{J}}(\hat{\rho}, g)$. This concludes the proof of the lemma. □

We still owe the reader the proof of Lemma 3.7.

Proof of Lemma 3.7 We first prove part 1. For simplicity, assume first that H has a single well, and therefore Γ has only one edge, $k = 1$. Since

$$\operatorname{div}\begin{pmatrix} 0 \\ \nabla_p H \end{pmatrix} = \Delta_p H,$$

and remarking that the exterior normal n to the set $H \leq h$ equals $(0, \nabla_p H/|\nabla H|)^T$, we calculate that

$$\int_{\{H\leq h\}} \Delta_p H = \int_{\{H=h\}} \frac{(\nabla_p H)^2}{|\nabla H|}\, d\mathscr{H}^1 = TA(h). \tag{69}$$

By the smoothness of H, the derivative of the left-hand integral is well-defined for all h such that $\nabla H \neq 0$ at that level. At such h we then have

$$TB(h) = \int_{\{H=h\}} \frac{\Delta_p H}{|\nabla H|}\, d\mathscr{H}^1 = \partial_h \int_{\{H\leq h\}} \Delta_p H = \partial_h TA(h).$$

For the multi-well case, this argument can simply be applied to each branch of Γ.

For part 2, since H is coercive, $\{H \leq h\}$ is bounded for each h; since H is smooth, therefore $\Delta_p H$ is bounded on bounded sets. From (69) it follows that TA also is bounded on bounded sets of Γ.

Finally, for part 3, note first that TB is bounded near each interior vertex. This follows by an explicit calculation and our assumption that each interior vertex corresponds to exactly one, non-degenerate, saddle point. Since $(TA)' = TB$, TA has a well-defined and finite limit at each interior saddle. The summation property (64) follows from comparing (69) for values of h just above and below the critical value. For instance, in the case of a single saddle at value $h = 0$, with two lower edges $k = 1, 2$ and upper edge $k = 0$, we have

$$\begin{aligned}\lim_{h\uparrow 0} TA(h,1) + TA(h,2) &= \lim_{h\uparrow 0} \int_{\xi^{-1}((-\infty,h]\times\{1\})} \Delta_p H + \int_{\xi^{-1}((-\infty,h]\times\{2\})} \Delta_p H \\ &= \lim_{h\uparrow 0} \int_{\{H\leq h\}} \Delta_p H = \lim_{h\downarrow 0} \int_{\{H\leq h\}} \Delta_p H = \lim_{h\downarrow 0} TA(h,0).\end{aligned}$$

This concludes the proof of Lemma 3.7. □

3.8 Conclusion and discussion

The combination of Theorems 3.2 and 3.6 give us that along subsequences $\hat{\rho}^\varepsilon := \xi_\# \rho^\varepsilon$ converges in an appropriate manner to some $\hat{\rho}$, and that

$$\hat{I}(\hat{\rho}) \leq \liminf_{\varepsilon \to 0} I^\varepsilon(\rho^\varepsilon).$$

In addition, any $\hat{\rho}$ satisfying $I(\hat{\rho}) = 0$ is a weak solution of the PDE

$$\partial_t \hat{\rho} = (A\hat{\rho})'' - (B\hat{\rho})'$$

on the graph Γ. This is the central coarse-graining statement of this section. We also obtain the boundary conditions, similarly as in the conventional weak-formulation method, by expanding the admissible set of test functions.

In switching from the VFP Eqs. (9)–(41) we removed two terms, representing the friction with the environment and the interaction between particles. Mathematically, it is straightforward to treat the case with friction, which leads to an additional drift term in the limit equation in the direction of decreasing h. We left this out simply for the convenience of shorter expressions.

As for the interaction, represented by the interaction potential ψ, again there is no mathematical necessity for setting $\psi = 0$ in this section; the analysis continues rather similarly. However, the limiting equation will now be non-local, since the particles at some $\gamma \in \Gamma$, which can be thought of as 'living' on a full connected level set of H, will feel a force exerted by particles at a different $\gamma' \in \Gamma$, i.e. at a different level set component. This makes the interpretation of the limiting equation somewhat convoluted.

The results of the current and the next sections were proved by Freidlin and co-authors in a series of papers [36–40], using probabilistic techniques. Recently, Barret and Von Renesse [15] provided an alternative proof using Dirichlet forms and their convergence. The latter approach is closer to ours in the sense that it is mainly PDE-based method and of variational type. However, in [15] the authors consider a perturbation of the Hamiltonian by a friction term and a non-degenerate noise, i.e. the noise is present in both space and momentum variables; this non-degeneracy appears to be essential in their method. Moreover, their approach invokes a reference measure which is required to satisfy certain non-trivial conditions. In contrast, the approach of this paper is applicable to degenerate noise and does not require such a reference measure. In addition, certain non-linear evolutions can be treated, such as the example of the VFP equation.

4 Diffusion on a graph, $d > 1$

We now switch to our final example. As described in the introduction, the higher-dimensional analogue of the diffusion-on-graph system has an additional twist: in order to obtain unique stationary measures on level sets of ξ, we need to add an additional noise in the SDE, or equivalently, an additional diffusion term in the PDE. This leads to the equation

$$\partial_t \rho = -\frac{1}{\varepsilon} \operatorname{div}(\rho J \nabla H) + \frac{\kappa}{\varepsilon} \operatorname{div}(a \nabla \rho) + \Delta_p \rho, \tag{70}$$

where $a : \mathbb{R}^{2d} \to \mathbb{R}^{2d \times 2d}$ with $a\nabla H = 0$, $\dim(\operatorname{Ker}(a)) = 1$ and $\kappa, \varepsilon > 0$ with $\kappa \gg \varepsilon$. The spatial domain is $\mathbb{R}^{2d}$, $d > 1$, with coordinates $(q, p) \in \mathbb{R}^d \times \mathbb{R}^d$. Here the unknown is

trajectory in the space of probability measures $\rho : [0, T] \to \mathcal{P}(\mathbb{R}^{2d})$; the Hamiltonian is the same as in the previous section, $H : \mathbb{R}^{2d} \to \mathbb{R}$ given by $H(q, p) = p^2/2m + V(q)$.

The results for the limit $\varepsilon \to 0$ in (70) closely mirror the one-degree-of-freedom diffusion-on-graph problem of the previous section; the only real difference lies in the proof of local equilibrium (Lemma 3.3). For a rigorous proof of this lemma in this case, based on probabilistic techniques, we refer to [39, Lemma 3.2]; here we only outline a possible analytic proof.

Along the lines of Theorem 3.1, and using boundedness of the rate functional $I^\varepsilon(\rho^\varepsilon)$, one can show that

$$\frac{1}{2}\int_0^T \int_{\mathbb{R}^2} \frac{|\nabla_p \rho^\varepsilon|^2}{\rho^\varepsilon} + \frac{\kappa}{\varepsilon}\int_0^T \int_{\mathbb{R}^2} \frac{a\nabla\rho^\varepsilon \cdot \nabla\rho^\varepsilon}{\rho^\varepsilon} \leq C.$$

Multiplying this inequality by ε/κ and using the weak convergence $\rho^\varepsilon \rightharpoonup \rho$ along with the lower-semicontinuity of the Fisher information [32, Theorem D.45] we find

$$\int_0^T \int_{\mathbb{R}^2} \frac{a\nabla\rho \cdot \nabla\rho}{\rho} = 0,$$

or in variational form, for almost all $t \in [0, T]$,

$$\begin{aligned} 0 &= \sup_{\varphi \in C_c^\infty(\mathbb{R}^{2d})} \int_{\mathbb{R}^{2d}} \operatorname{div}(a\nabla\varphi)\rho_t - \frac{1}{2}\int_{\mathbb{R}^{2d}} a\nabla\varphi \cdot \nabla\varphi\rho_t \\ &\iff 0 = \int_{\mathbb{R}^{2d}} \operatorname{div}(a\nabla\varphi)\rho_t, \qquad \forall \varphi \in C_c^\infty(\mathbb{R}^{2d}). \end{aligned}$$

Applying the co-area formula we find

$$\int_{\xi^{-1}(\gamma)} \frac{\rho(x)}{|\nabla H(x)|} \operatorname{div}(a(x)\nabla\varphi(x))\, \mathscr{H}^{2d-1}(dx) = 0, \tag{71}$$

where $\mathscr{H}^{2d-1}$ is the $(2d-1)$ dimensional Haursdoff measure. Let $\mathcal{M}_\gamma$ be the $(2d-1)$ dimensional manifold $\xi^{-1}(\gamma)$ with volume element $|\nabla H|^{-1}\mathscr{H}^{2d-1}$. Then (71) becomes

$$\int_{\mathcal{M}_\gamma} \rho(x)\, \operatorname{div}_{\mathcal{M}}(a(x)\nabla_{\mathcal{M}}\varphi(x))\, \operatorname{vol}_{\mathcal{M}}(dx) = 0,$$

where $\operatorname{div}_{\mathcal{M}}$ and $\nabla_{\mathcal{M}}$ are the corresponding differential operators on $\mathcal{M}_\gamma$, and $\operatorname{vol}_{\mathcal{M}}$ is the induced volume measure. Since $a\nabla H = 0$, $\dim(\operatorname{Ker}(a)) = 1$, a is non-degenerate on the tangent space of $\mathcal{M}_\gamma$. Therefore, given $\psi \in C^\infty(\mathcal{M}_\gamma)$ with $\int_{\mathcal{M}_\gamma} \psi\, d\operatorname{vol}_{\mathcal{M}} = 0$, we can solve the corresponding Laplace–Beltrami–Poisson equation for φ,

$$\operatorname{div}_{\mathcal{M}}(a\nabla_{\mathcal{M}}\varphi) = \psi,$$

and therefore

$$\int_{\mathcal{M}_\gamma} \rho\, \psi\, \operatorname{dvol}_{\mathcal{M}} = 0, \ \forall \psi \in C^\infty(\mathcal{M}_\gamma) \text{ with } \int_{\mathcal{M}_\gamma} \psi\, d\operatorname{vol}_{\mathcal{M}} = 0.$$

Since $\mathcal{M}_\gamma$ is connected by definition, it follows that ρ constant on $\mathcal{M}_\gamma$; this is the statement of Lemma 3.3.

5 Conclusion and discussion

In this paper we have presented a structure in which coarse-graining and 'passing to a limit' combine in a natural way, and which extends also naturally to a class of approximate solutions. The central object is the rate function I, which is minimal and vanishes at solutions; in the dual formulation of this rate function, coarse-graining has a natural interpretation, and the inequalities of the dual formulation and of the coarse-graining combine in a convenient way.

We now comment on a number of issues related with this method.

Why does this method work? One can wonder why the different pieces of the arguments of this paper fit together. Why do the relative entropy and the relative Fisher information appear? To some extent this can be recognized in the similarity between the duality definition of the rate function I and the duality characterization of relative entropy and relative Fisher Information. The details of Appendix B show this most clearly, but the similarity between the duality definition of the relative Fisher information and the duality structure of I can readily be recognized: in (19) combined with (18) we collect the $O(\gamma^2)$ terms

$$\int_0^T \int_{\mathbb{R}^{2d}} \left[\Delta_p f_t - \frac{p}{m} \nabla_p f_t - \frac{1}{2} \left| \nabla_p f_t \right|^2 \right] d\rho_t dt,$$

and these match one-to-one to the definition (24). This shows how the structure of the relative Fisher Information is to some extent 'built-in' in this system.

Relation with other variational formulations Our variational formulation (2) to 'passing to a limit' is closely related to other variational formulations in the literature, notably the Ψ–Ψ^* formulation and the method in [7,64]. In the Ψ–Ψ^* formulation, a gradient flow of the energy $\mathcal{E}_\varepsilon : \mathcal{Z} \to \mathbb{R}$ with respect to the dissipation Ψ_ε^* is defined to be a curve $\rho^\varepsilon \in C([0,T], \mathcal{Z})$ such that

$$\mathcal{A}^\varepsilon(\rho) := \mathcal{E}_\varepsilon(\rho_T) - \mathcal{E}_\varepsilon(\rho_0) + \int_0^T \left[\Psi_\varepsilon(\dot{\rho}_t, \rho_t) + \Psi_\varepsilon^*(-\mathsf{D}\mathcal{E}_\varepsilon(\rho_t), \rho_t) \right] dt = 0. \tag{72}$$

'Passing to a limit' in a Ψ–Ψ^* structure is then accomplished by studying (Gamma-) limits of the functionals $\mathcal{A}^\varepsilon$. The method introduced in [7,64] is slightly different. Therein 'passing to a limit' in the evolution equation is executed by studying (Gamma-)limits of the functionals that appear in the approximating discrete minimizing-movement schemes.

The similarities between these two approaches and ours is that all the methods hinge on duality structure of the relevant functionals, allow one to obtain both compactness and limiting results, and can work with approximate solutions, see e.g. [6] and the papers above for details. In addition, all methods assume some sort of well-prepared initial data, such as bounded initial free energy and boundedness of the functionals. Our assumptions on the boundedness of the rate functionals arise naturally in the context of large-deviation principle since this assumption describes events of a certain degree of 'improbability'.

The main difference is that the method of this paper makes no use of the gradient-flow structure, and therefore also applies to non-gradient-flow systems as in this paper. The first example, of the overdamped limit of the VFP equation, also is interesting in the sense that it derives a dissipative system from a non-dissipative one. Since the GENERIC framework unifies both dissipative and non-dissipative systems, we expect that the method of this paper could be used to derive evolutionary convergence for GENERIC systems (see the next point). Finally, we emphasize that using the duality of the rate functional is mathematically convenient because we do not need to treat the three terms in the right-hand side of (72) separately. Note that although the entropy and energy functionals as well as the dissipation mechanism

are not explictly present in this formulation, we are still able to derive an energy-dissipation inequality in (4).

Relation with GENERIC As mentioned in the introduction, the Vlasov–Fokker–Planck system (8) combines both conservative and dissipative effects. In fact it can be cast into the GENERIC form by introducing an excess-energy variable e, depending only on time, that captures the fluctuation of energy due to dissipative effects (but does not change the evolution of the system). The building blocks of the GENERIC for the augmented system for (ρ, e) can be easily deduced from the conservative and dissipative effects of the original Vlasov–Fokker–Planck equation. Moreover, this GENERIC structure can be derived from the large-deviation rate functional of the empirical process (7). We refer to [26] for more information. This suggests that our method could be applied to other GENERIC systems.

Gradient flows and large-deviation principles As mentioned in the introduction, this approach using the duality formulation of the rate functionals is motivated by our recent results on the connection between generalised gradient flows and large-deviation principles [2,3,24,26,27,52]. We want to discuss here how the two overlap but are not the same. In [52], the authors show that if $\mathcal{N}^\varepsilon$ is the adjoint operator of a generator of a Markov process that satisfies a *detailed balance condition*, then the evolution (1) is the same as the generalised gradient flow induced from a large-deviation rate functional, which is of the form $\int_0^T \mathscr{L}^\varepsilon(\rho_t, \dot\rho_t)\,dt$, of the underlying empirical process. The generalised gradient flow is described via the Ψ–Ψ^* structure as in (72) with $\mathscr{L}^\varepsilon(z, \dot z) = \Psi_\varepsilon(z, \dot z) + \Psi_\varepsilon^*(z, -\mathsf{D}\mathcal{E}_\varepsilon(z)) + \langle \mathsf{D}\mathcal{E}_\varepsilon(z), \dot z\rangle$. Moreover, $\mathcal{E}_\varepsilon$ and Ψ_ε can be determined from $\mathscr{L}^\varepsilon$ [52, Theorem 3.3]. However, it is not clear if such characterisation holds true for systems that do not satisfy detailed balance. In addition, there exist (generalised) gradient flows for which we currently do not know of any corresponding microscopic particle systems, such as the Allen–Cahn and Cahn–Hilliard equations.

Quantification of coarse-graining error The use of the rate functional in a central role allows us not only to derive the limiting coarse-grained system but also to obtain quantitative estimates of the coarse-graining error. Existing quantitative methods such as [42,49] only work for gradient flows systems since they use crucially the gradient flow structures. The essential estimate that they need is the energy-dissipation inequality, which is similar to (4). Since we are able to obtain this inequality from the duality formulation of the rate functionals, our method would offer an alternative technique for obtaining quantitative estimate of the coarse-graining error for both dissipative and non-dissipative systems. We address this issue in detail in a companion article [23].

Other stochastic processes The key ingredient of the method is the duality structure of the rate functional (5) and (10). This duality formulation holds true for many other stochastic processes; indeed, the 'Feng–Kurtz' algorithm (see chapter 1 of [32]) suggests that the large-deviation rate functional for a very wide class of Markov processes can be written as

$$I(\rho) = \sup_f \left\{ \langle f_T, \rho_T\rangle - \langle f_0, \rho_0\rangle - \int_0^T \langle \dot f_t, \rho_t\rangle\,dt - \int_0^T \mathcal{H}(\rho_t, f_t)\,dt \right\},$$

where $\mathcal{H}$ is an appropriate limit of 'non-linear' generators. The formula (10) is a special case. As a result, we expect that the method can be extended to this same wide class of Markov processes.

Acknowledgements The authors would like to thank the anonymous referee for valuable suggestions and comments. US thanks Giovanni Bonaschi, Xiulei Cao, Joep Evers and Patrick van Meurs for insightful discussions regarding Theorems 2.4 and 3.3. MAP and US kindly acknowledge support from the Nederlandse

Organisatie voor Wetenschappelijk Onderzoek (NWO) VICI Grant 639.033.008. MHD was supported by the ERC Starting Grant 335120. Part of this work has appeared in Oberwolfach proceedings [63].

A Proof of Lemma 2.1

Define $\tilde{\mathcal{I}}(f)$ to be the right-hand side in (25),

$$\tilde{\mathcal{I}}(f) := \begin{cases} \displaystyle\int_{\mathbb{R}^{2d}} \left| \frac{\nabla_p f}{f} \mathbb{1}_{\{f>0\}} + \frac{p}{m} \right|^2 f \, dqdp, & \text{if } \ \nabla_p f \in L^1_{\mathrm{loc}}(dqdp), \\ \infty & \text{otherwise.} \end{cases}$$

for $f \in L^1(\mathbb{R}^{2d})$. We need to show that $\tilde{\mathcal{I}}(f) = \mathcal{I}(f\,dqdp|\mu)$.

First assume that $\tilde{\mathcal{I}}$ is finite. Then $\frac{\nabla_p f}{f}\mathbb{1}_{\{f>0\}} + \frac{p}{m} \in L^2(f\,dqdp)$, which implies the following stronger statement.

Lemma A.1 *One has*

$$\frac{\nabla_p f}{f}\mathbb{1}_{\{f>0\}} + \frac{p}{m} \in L^2_\nabla(f\,dqdp),$$

where the space $L^2_\nabla(f\,dqdp)$ is defined as the closure of $\left\{\nabla_p \varphi : \varphi \in C_c^\infty(\mathbb{R}^{2d})\right\}$ with respect to the norm $\|\cdot\|^2_{f\,dqdp} := \int_{\mathbb{R}^{2d}} |\cdot|^2 \, f\,dqdp$.

Assuming Lemma A.1 for the moment we rewrite $\tilde{\mathcal{I}}(f)$ as

$$\begin{aligned} \tilde{\mathcal{I}}(f) &= \int_{\mathbb{R}^{2d}} \left| \tfrac{\nabla_p f}{f}\mathbb{1}_{\{f>0\}} + \tfrac{p}{m} \right|^2 f\,dqdp = \left\| -\nabla_p \cdot \left(f \left(\tfrac{\nabla_p f}{f}\mathbb{1}_{\{f>0\}} + \tfrac{p}{m} \right) \right) \right\|^2_{-1,(f\,dqdp)} \\ &= \| -\nabla_p \cdot \left(\mathbb{1}_{\{f>0\}} \nabla_p f + f \tfrac{p}{m} \right) \|^2_{-1,(f\,dqdp)} \\ &= \| -\nabla_p \cdot \left(\nabla_p f + f \tfrac{p}{m} \right) \|^2_{-1,(f\,dqdp)}, \end{aligned}$$

where $\|\cdot\|_{-1,f\,dqdp}$ is the dual norm (in duality with $L^2_\nabla(f\,dqdp)$) from [26] and $\mathbb{1}_{\{f>0\}}\nabla_p f = \nabla_p f$ holds due to Stampacchia's Lemma [47, Theorem A.1]. Following the variational characterization of $\|\cdot\|_{-1,(f\,dqdp)}$ from [26, (11)] we finally obtain

$$\begin{aligned} \tilde{\mathcal{I}}(f) &= \sup_{\varphi \in C_c^\infty(\mathbb{R}^{2d})} 2 \int_{\mathbb{R}^{2d}} \left(\nabla_p \varphi \cdot \frac{p}{m} - \mathbb{1}_{\{f>0\}} \Delta_p \varphi - \tfrac{1}{2} |\nabla_p \varphi|^2 \right) f\,dqdp \\ &= \sup_{\varphi \in C_c^\infty(\mathbb{R}^{2d})} 2 \int_{\mathbb{R}^{2d}} \left(\nabla_p \varphi \cdot \frac{p}{m} - \Delta_p \varphi - \tfrac{1}{2} |\nabla_p \varphi|^2 \right) f\,dqdp, \end{aligned}$$

which is the claimed result. The same reference also provides that $\tilde{\mathcal{I}} = \infty$ iff $\mathcal{I}(f\,dqdp|\mu) = \infty$.

Proof of Lemma A.1 We assume that $\frac{\nabla_p f}{f}\mathbb{1}_{\{f>0\}} + \frac{p}{m} \in L^2(f\,dqdp)$ and show that the two individual terms $\frac{\nabla_p f}{f}\mathbb{1}_{\{f>0\}}$ and $\frac{p}{m}$ are in $L^2_\nabla(f\,dqdp)$. Choose a smooth cut-off function $\eta_R = \eta(x/R)$ with $\eta : \mathbb{R}^{2d} \to \mathbb{R}$, $\eta = 1$ on $B_1(0)$ and $\eta = 0$ in $\mathbb{R}^{2d} \setminus B_2(0)$. Then

$$-\int_{\mathbb{R}^{2d}} \eta_R \frac{p}{m}\cdot\frac{\nabla_p f}{f}\mathbb{1}_{\{f>0\}} f = -\int_{\mathbb{R}^{2d}} \eta_R \frac{p}{m}\cdot\nabla_p f\,\mathbb{1}_{\{f>0\}} = -\int_{\mathbb{R}^{2d}} \eta_R \frac{p}{m}\cdot\nabla_p(\mathbb{1}_{\{f>0\}} f)$$
$$= +\frac{1}{m}\int_{\mathbb{R}^{2d}}\Big[\eta_R\, d + p\cdot\nabla_p\eta_R\Big]\mathbb{1}_{\{f>0\}} f \le \frac{d}{m}$$
$$+\int_{\mathbb{R}^{2d}} p\cdot\nabla_p\eta_R\, f =: b(R).$$

As $R\to\infty$, the bound $b(R)$ converges to d/m.

Therefore we have

$$\int_{\mathbb{R}^{2d}} \eta_R\left[\left|\tfrac{\nabla_p f}{f}\mathbb{1}_{\{f>0\}}\right|^2 + \left|\tfrac{p}{m}\right|^2\right] f = \int_{\mathbb{R}^{2d}} \eta_R\left|\tfrac{\nabla_p f}{f}\mathbb{1}_{\{f>0\}} + \tfrac{p}{m}\right|^2 f - 2\int_{\mathbb{R}^{2d}} \eta_R\,\nabla_p f\cdot\tfrac{p}{m}\mathbb{1}_{\{f>0\}}$$
$$\le 2b(R) + \int_{\mathbb{R}^{2d}} \eta_R\left|\tfrac{\nabla_p f}{f}\mathbb{1}_{\{f>0\}} + \tfrac{p}{m}\right|^2 f.$$

By passing to the limit $R\to\infty$ we obtain

$$\lim_{R\to\infty}\int_{\mathbb{R}^{2d}} \eta_R\left[\left|\tfrac{\nabla_p f}{f}\mathbb{1}_{\{f>0\}}\right|^2 + \left|\tfrac{p}{m}\right|^2\right] f \le \int_{\mathbb{R}^{2d}}\left|\tfrac{\nabla_p f}{f}\mathbb{1}_{\{f>0\}} + \tfrac{p}{m}\right|^2 f + \frac{2d}{m} < \infty$$

and thus $\frac{\nabla_p f}{f}\mathbb{1}_{\{f>0\}}, \frac{p}{m} \in L^2(f\,dq dp)$. To conclude the proof of Lemma A.1 it remains to show that $\frac{\nabla_p f}{f}\mathbb{1}_{\{f>0\}}, \frac{p}{m}$ can be approximated by gradients of C_c^∞-functions. To this end we consider, for $\varepsilon>0$, the smooth cut-off function $\eta_\varepsilon := \eta(x\varepsilon)$ with η as above and define

$$\varphi_\varepsilon := \left[\log\left(\frac{1}{\varepsilon}\wedge(f\vee\varepsilon)\right) - \log\varepsilon\right]\eta_\varepsilon.$$

Then φ_ε has compact support in $\mathbb{R}^{2d}$. Note that φ_ε is not necessarily smooth, but by convolution with a mollifier we can also achieve smoothness. For the gradient one obtains

$$\nabla_p\varphi_\varepsilon = \begin{cases} \mathbb{1}_{B_{\frac{1}{\varepsilon}}(0)}\frac{\nabla_p f}{f} + \mathbb{1}_{B_{\frac{2}{\varepsilon}}(0)\setminus B_{\frac{1}{\varepsilon}}(0)}\left(\eta_\varepsilon\frac{\nabla_p f}{f} + \nabla_p\eta_\varepsilon(\log f - \log\varepsilon)\right) & \text{for } \{\varepsilon\le f\le\frac{1}{\varepsilon}\} \\ \mathbb{1}_{B_{\frac{2}{\varepsilon}}(0)\setminus B_{\frac{1}{\varepsilon}}(0)}\nabla_p\eta_\varepsilon\left(\log\frac{1}{\varepsilon} - \log\varepsilon\right) & \text{for } \{f>\frac{1}{\varepsilon}\} \\ 0 & \text{for } \{f<\varepsilon\} \end{cases}$$

Our aim is to show that $\left\|\frac{\nabla_p f}{f}\mathbb{1}_{\{f>0\}} - \nabla_p\varphi_\varepsilon\right\|_{f\,dq dp} \to 0$ as $\varepsilon\to 0$. Indeed,

$$\int_{\mathbb{R}^{2d}}\left|\tfrac{\nabla_p f}{f}\mathbb{1}_{\{f>0\}} - \nabla_p\varphi_\varepsilon\right|^2 f$$
$$= \int_{\{f<\varepsilon\}}\left|\tfrac{\nabla_p f}{f}\mathbb{1}_{\{f>0\}}\right|^2 f + \int_{\{f>\frac{1}{\varepsilon}\}}\left|\tfrac{\nabla_p f}{f}\mathbb{1}_{\{f>0\}} - \nabla_p\eta_\varepsilon\left(\log\frac{1}{\varepsilon} - \log\varepsilon\right)\mathbb{1}_{B_{\frac{2}{\varepsilon}}(0)\setminus B_{\frac{1}{\varepsilon}}(0)}\right|^2 f$$
$$+ \int_{\{\varepsilon\le f\le\frac{1}{\varepsilon}\}}\left|(1-\eta_\varepsilon)\tfrac{\nabla_p f}{f}\mathbb{1}_{\{f>0\}} - \nabla_p\eta_\varepsilon(\log f - \log\varepsilon)\right|^2\mathbb{1}_{B_{\frac{2}{\varepsilon}}(0)\setminus B_{\frac{1}{\varepsilon}}(0)} f$$
$$+ \int_{\{\varepsilon\le f\le\frac{1}{\varepsilon}\}}\left|\tfrac{\nabla_p f}{f}\mathbb{1}_{\{f>0\}}\right|^2\mathbb{1}_{\mathbb{R}^{2d}\setminus B_{\frac{2}{\varepsilon}}(0)} f$$
$$=: \mathrm{I}_\varepsilon + \mathrm{II}_\varepsilon + \mathrm{III}_\varepsilon + \mathrm{IV}_\varepsilon.$$

Since $\frac{\nabla_p f}{f}\mathbb{1}_{\{f>0\}} \in L^2(f\,dqdp)$ we directly conclude that I_ε and IV_ε vanish in the limit as $\varepsilon\to 0$. Concerning II_ε and III_ε we note that, for $\{\varepsilon\le f\le\frac{1}{\varepsilon}\}$, one has

$$\left|\nabla_p\eta_\varepsilon(\log f - \log\varepsilon)\right|^2 \le |\nabla_p\eta_\varepsilon|^2\,|\log 1/\varepsilon - \log\varepsilon|^2 = |\nabla_p\eta_\varepsilon|^2\left|2\log\frac{1}{\varepsilon}\right|^2 \le C\varepsilon,$$

where we exploited $|\nabla_p \eta_\varepsilon|^2 \le C\varepsilon^2$ and $\left(\log \frac{1}{\varepsilon}\right)^2 \le C\frac{1}{\varepsilon}$ for some ε-independent constant C. This shows that also II_ε and III_ε vanish in the limit as $\varepsilon \to 0$. To sum up, we conclude that $\frac{\nabla_p f}{f}\mathbb{1}_{\{f>0\}} \in L^2_\nabla(f\,dq\,dp)$. The calculation for $\frac{p}{m} = \nabla_p\left(\frac{|p|^2}{2m}\right)$ is similar. □

B Proof of Theorem 2.3

In this appendix, we prove Theorem 2.3 using the method of the duality equation; see e.g. [1, 12,29,65] or [13, Ch. 9] for examples. Throughout this appendix γ is fixed.

We recall the functional $I^\gamma : C([0,T];\mathcal{P}(\mathbb{R}^{2d})) \to \mathbb{R}$ defined in (19)

$$I^\gamma(\rho) = \sup_{f\in C_b^{1,2}(\mathbb{R}\times\mathbb{R}^{2d})}\Bigg[\int_{\mathbb{R}^{2d}} f_T\,d\rho_T - \int_{\mathbb{R}^{2d}} f_0\,d\rho_0 - \int_0^T\int_{\mathbb{R}^{2d}}\left(\partial_t f_t + \mathscr{L}_{\rho_t} f_t\right)d\rho_t dt$$
$$-\frac{\gamma^2}{2}\int_0^T\int_{\mathbb{R}^{2d}}\left|\nabla_p f_t\right|^2 d\rho_t dt\Bigg], \tag{73}$$

where $\mathscr{L}_\nu$ is given by

$$\mathscr{L}_\nu f = \gamma J\nabla(H+\psi*\nu)\cdot\nabla f - \gamma^2\frac{p}{m}\cdot\nabla_p f + \gamma^2\Delta_p f. \tag{74}$$

In addition to the duality definition of the Fisher Information (24) we will use the Donsker–Varadhan duality characterization of the relative entropy (21) for two probability measures (see e.g. [20, Lemma 1.4.3])

$$\mathcal{H}(\nu|\mu) = \sup_{\phi\in C_c^\infty(\mathbb{R}^{2d})}\int_{\mathbb{R}^{2d}}\phi d\nu - \log\int_{\mathbb{R}^{2d}} \mathrm{e}^{\phi}d\mu,$$

which implies the corresponding characterization of the free energy (22)

$$\mathcal{F}(\nu) = \sup_{\phi\in C_c^\infty(\mathbb{R}^{2d})}\int_{\mathbb{R}^{2d}}\left[\phi + \frac{1}{2}\psi*\nu\right]d\nu - \log\int_{\mathbb{R}^{2d}}\mathrm{e}^{\phi-H}dx + \log Z_H. \tag{75}$$

We first present some intermediate results which we will use to prove Theorem 2.3.

Lemma B.1 *Let $\rho \in C([0,T];\mathcal{P}(\mathbb{R}^{2d}))$.*

1. *The maps $t \mapsto \psi*\rho_t$ and $t \mapsto \nabla\psi*\rho_t$ are continuous from $[0,T]$ to $C_b(\mathbb{R}^d)$;*
2. *If $I^\gamma(\rho), \mathcal{H}(\rho_0|Z_H^{-1}\mathrm{e}^{-H}) < \infty$, then $\int H\rho_t < \infty$ for all $t \in [0,T]$.*

Proof The first part follows from the bound $\psi \in W^{1,1}(\mathbb{R}^d)\cap C_b^2(\mathbb{R}^d)$. Fix $\varepsilon > 0, t \in [0,T]$, and take a sequence $t_n \to t$. For each n, choose $x_n \in \mathbb{R}^{2d}$ such that $|\psi*(\rho_t-\rho_{t_n})|(x_n) \ge \|\psi*(\rho_t-\rho_{t_n})\|_\infty - \varepsilon/2$. Since $\rho_{t_n} \to \rho_t$ narrowly, $\{\rho_{t_n}\}_n$ is tight, implying that x_n can be chosen bounded; therefore there exists a subsequence (not relabelled) such that $x_n \to x$ as $n \to \infty$. Then

$$\begin{aligned}|(\psi*\rho_t)(x_n) - (\psi*\rho_{t_n})(x_n)| \le\ & |(\psi*\rho_t)(x_n) - (\psi*\rho_t)(x)| \\ &+ |(\psi*\rho_t)(x) - (\psi*\rho_{t_n})(x)| \\ &+ |(\psi*\rho_{t_n})(x) - (\psi*\rho_{t_n})(x_n)|.\end{aligned}$$

The last term on the right-hand side satisfies

$$|(\psi * \rho_{t_n})(x) - (\psi * \rho_{t_n})(x_n)| \le \int_{\mathbb{R}^{2d}} |\psi(x-y) - \psi(x_n - y)| \rho_{t_n}(y,z)\, dy\, dz \to 0$$

since $\psi(x_n - \cdot) \to \psi(x - \cdot)$ uniformly, and a similar argument applies to the first term. The middle term converges to zero by the narrow convergence of ρ_{t_n} to ρ_t. This proves that the function $t \mapsto \psi * \rho_t$ is continuous; a similar argument applies to $t \mapsto \nabla\psi * \rho_t$.

For the second part, we take in (73) the function $f(q,p,t) = \zeta(H(q,p))$, where $\zeta \in C^\infty([0,\infty))$ is a smooth, bounded, increasing truncation of the function $f(s) = s$, satisfying $0 \le \zeta' \le 1$ and $\zeta'' \le 0$. Then we find

$$\begin{aligned}\int_{\mathbb{R}^{2d}} \zeta(H)\rho_\tau - \int_{\mathbb{R}^{2d}} \zeta(H)\rho_0 - I^\gamma(\rho) &\le \int_0^\tau \int_{\mathbb{R}^{2d}} \Big(-\gamma\zeta' \frac{p}{m} \cdot \nabla_q \psi * \rho_t + \gamma^2 \Big(\zeta'' \\ &\quad + \tfrac{1}{2}\zeta'^2 - \zeta'\Big)\frac{p^2}{m^2} + \gamma^2 \zeta' \frac{d}{m}\Big)\, d\rho_t dt \\ &\le \int_0^\tau \int_{\mathbb{R}^{2d}} \left(\frac{1}{2}\zeta' |\nabla_q \psi * \rho_t|^2 + \gamma^2 \zeta' \frac{d}{m}\right) d\rho_t dt \le \frac{\tau}{2}\|\nabla_q \psi\|_\infty^2 + \gamma^2 \frac{d}{m}\,\tau.\end{aligned}$$

The result follows upon letting ζ converge to the identity.

Note that this inequality gives a bound on $\int H\rho_t$ for fixed γ, but this bound breaks down when $\gamma \to \infty$. The bound (29), which is directly derived from (28), gives a γ-independent estimate. □

In the next few results we study certain properties of an auxiliary PDE and its connection to the rate functional.

Theorem B.2 *Given $\phi \in C_c^\infty(\mathbb{R}^{2d})$ and $\varphi \in C_c^\infty([0,T] \times \mathbb{R}^d)$, there exists a function $f \in L^1_{\mathrm{loc}}([0,T] \times \mathbb{R}^{2d})$ which satisfies the following equation a.e. in $L^1_{\mathrm{loc}}([0,T] \times \mathbb{R}^{2d})$ (i.e. for each compact set $K \subset [0,T] \times \mathbb{R}^{2d}$, the equation is satisfied with all weak derivatives and all terms in $L^1(K)$):*

$$\partial_t f + \mathscr{L}_\rho f + \frac{\gamma^2}{2}|\nabla_p f|^2 + \gamma J\nabla H \cdot \nabla\psi * \rho_t = -\gamma^2\Big(\Delta_p \varphi - \nabla_p H \cdot \nabla_p \varphi - \frac{1}{2}|\nabla_p \varphi|^2\Big), \tag{76a}$$

$$f|_{t=T} = \phi \tag{76b}$$

where $\mathscr{L}_\rho$ is defined in (74). The final-time condition (76b) is satisfied in the sense of traces in $L^1_{\mathrm{loc}}(\mathbb{R}^{2d})$ (which are well-defined since $\partial_t f \in L^1_{\mathrm{loc}}([0,T] \times \mathbb{R}^{2d})$). The solution satisfies $|f| \le C(1+H)^{1/2}$ for each $t \in [0,T]$ and almost everywhere in $\mathbb{R}^{2d}$, for some constant $C > 0$. Finally,

$$t \mapsto \int_{\mathbb{R}^{2d}} \mathrm{e}^{f_t - H} dx \quad \textit{is non-decreasing.} \tag{77}$$

Proof The Hopf–Cole tranformation $f = 2\log g$ and the time reversal $t \mapsto T - t$ transform equation (76a) into

$$\partial_t g - \mathscr{L}_\rho g = -\frac{g}{2}\Big(-\gamma J\nabla H\cdot\nabla\psi * \rho_t - \gamma^2\Big[\Delta_p\varphi - \nabla_p H\cdot\nabla_p\varphi - \frac{1}{2}|\nabla_p\varphi|^2\Big]\Big), \quad (78)$$

with initial datum (now at time zero) $g_0 = \mathrm{e}^{\phi/2}$. The analysis of Eq. (78) is non-standard and therefore we study this equation separately in Appendix C. The existence and uniqueness of a solution, with this initial value, follow from Corollary C.7. The solution g satisfies (78) a.e. in $L^1_{\mathrm{loc}}([0,T]\times\mathbb{R}^{2d})$ by Proposition C.13. Furthermore, by Proposition C.10 there exist constants $\alpha_1, \alpha_2, \beta_1, \beta_2, \omega_1, \omega_2$ such that

$$\alpha_1 \exp\left(-\beta_1 t\sqrt{\omega_1 + H}\right) \le g \le \alpha_2 \exp\left(\beta_2 t\sqrt{\omega_2 + H}\right).$$

Finally, by Proposition C.11 we have

$$t \mapsto \int_{\mathbb{R}^{2d}} g_t^2 \mathrm{e}^{-H} dx \quad \text{is non-increasing.}$$

Transforming back to f we find the result. □

To prove the second main result on the auxiliary Eq. (76a), which is Proposition B.4 below, we will need the following lemma. For the rest of this appendix we write $*_t$ for convolution in time and $*_x$ for convolution in space ($x = (q, p)$). (The convolution $\psi *_x \rho$ is the same as the notation $\psi * \rho$ used in the rest of this paper.)

Lemma B.3 *Let f satisfy*

$$\partial_t f + \mathscr{L}_{\rho_t} f + \frac{\gamma^2}{2}|\nabla_p f|^2 = \Phi, \quad (79)$$

a.e. in $L^1_{\mathrm{loc}}(\mathbb{R}\times\mathbb{R}^{2d})$ with $\Phi \in L^1_{\mathrm{loc}}(\mathbb{R}\times\mathbb{R}^{2d})$. Define $f_\delta := \nu_\delta *_x f$ and $f_\varepsilon := \eta_\varepsilon *_t f$, where $\eta_\varepsilon = \eta_\varepsilon(t)$ is a regularizing sequence in the t-variable and $\nu_\delta = \nu_\delta(q,p)$ is a regularizing sequence in the (q,p)-variables. Then we have

$$\partial_t f_\delta + \mathscr{L}_{\rho_t} f_\delta + \frac{\gamma^2}{2}|\nabla_p f_\delta|^2 \le \nu_\delta *_x \Phi + \gamma\delta\|d^2 H\|_{L^\infty}(\nu_\delta *_x |\nabla f| + \gamma\nu_\delta *_x |\nabla_p f|) + \gamma\Big(J\nabla\psi *_x \rho_t\cdot\nabla f_\delta - \nu_\delta *_x (J\nabla\psi *_x \rho_t\cdot\nabla f)\Big), \quad (80)$$

$$\partial_t f_\varepsilon + \mathscr{L}_{\rho_t} f_\varepsilon + \frac{\gamma^2}{2}|\nabla_p f_\varepsilon|^2 \le \eta_\varepsilon *_t \Phi + \gamma\Big(J\nabla\psi *_x \rho_t\nabla f_\varepsilon - \eta_\varepsilon *_t (J\nabla\psi *_x \rho_t\cdot\nabla f)\Big). \quad (81)$$

Proof of Lemma B.3 Using (79) and the definition of $\mathscr{L}_\rho$ we have

$$\begin{aligned} 0 &= \int_{\mathbb{R}} \eta_\varepsilon(t-\tau)\Big(\partial_t f + \gamma^2\Delta_p f - \gamma^2\nabla_p H\cdot\nabla_p f + \gamma J\nabla(H + \psi *_x \rho_t)\cdot\nabla f \\ &\quad + \frac{\gamma^2}{2}|\nabla_p f|^2 - \Phi\Big)(\tau,x)d\tau \\ &= \Big(\partial_t f_\varepsilon + \gamma^2\Delta_p f_\varepsilon - \gamma^2\nabla_p H\cdot\nabla_p f_\varepsilon + \gamma J\nabla(H + \psi *_x \rho_t)\cdot\nabla f_\varepsilon\Big)(t,x) \\ &\quad + \frac{\gamma^2}{2}\eta_\varepsilon *_t |\nabla_p f|^2 - \eta_\varepsilon *_t \Phi(t,x) \\ &\quad + \gamma\int_{\mathbb{R}} \eta_\varepsilon(t-\tau)\Big(J\nabla\psi *_x \rho_\tau - J\nabla\psi *_x \rho_t\Big)\nabla f(\tau,x)d\tau. \end{aligned} \quad (82)$$

By Jensen's inequality we have $\eta_\varepsilon *_t |\nabla_p f|^2 \ge |\nabla_p f_\varepsilon|^2$. Substituting this inequality into the relation above completes the proof of (81). The proof of (80) follows similarly. □

The next result connects the solution of the auxiliary Eq. (76a) to the rate functional (73).

Proposition B.4 *Let f be the solution of* (76a), (76b) *in the sense of Theorem* B.2. *Then for* $\tau \in [0, T]$ *we have*

$$\begin{aligned}&\int_{\mathbb{R}^{2d}} \rho_\tau \Big(f_\tau + \frac{1}{2}\psi *_x \rho_\tau \Big) \\ &\quad - \int_0^\tau \int_{\mathbb{R}^{2d}} \Big\{ \partial_t f + \mathscr{L}_\rho f + \frac{\gamma^2}{2} |\nabla_p f|^2 + \gamma J \nabla H \cdot \nabla \psi *_x \rho_t \Big\} d\rho_t dt \\ &\quad \leq I(\rho) + \mathcal{F}(\rho_0) + \log \int_{\mathbb{R}^{2d}} \mathrm{e}^{f_0 - H} dx - \log Z_H. \end{aligned} \tag{83}$$

Proof We first show that for every $\tau \in [0, T]$,

$$\begin{aligned} I(\rho) \geq \sup_{\tilde f \in \mathcal{A}} &\int_{\mathbb{R}^{2d}} \rho \Big(\tilde f + \frac{1}{2}\psi *_x \rho \Big) \Big|_0^\tau - \int_0^\tau \int_{\mathbb{R}^{2d}} \Big\{ \partial_t \tilde f + \mathscr{L}_\rho \tilde f + \frac{\gamma^2}{2} |\nabla_p \tilde f|^2 \\ &+ \gamma J \nabla H \cdot \nabla \psi *_x \rho_t \Big\} d\rho_t dt, \end{aligned} \tag{84}$$

where

$$\mathcal{A} = \Big\{ \tilde f \in C^{1,2}([0, T] \times \mathbb{R}^{2d}) : |\partial_t \tilde f|, |\nabla \tilde f|^2, |\Delta \tilde f| \leq C(1 + H) \Big\}.$$

Formally, this follows from substituting in the rate functional (73) $f(t, x) = \big[\psi *_x \rho + \tilde f\big](t, x)\chi_{[0,\tau]}(t)$ with $\tilde f \in \mathcal{A}$, and where $\chi_{[0,\tau]}$ is the characteristic function of the interval $[0, \tau]$. The rigorous proof follows by choosing in the rate functional (73) the function

$$f_n = \delta_n *_t (\xi \delta_n *_t \psi *_x \rho) + \tilde f \xi,$$

for some $\tilde f \in \mathcal{A}$ and $\xi \in C_c^\infty((0, \tau))$. Here $\delta_n(t) := n\delta(nt)$ is an approximation of a Dirac. Upon rearranging the time convolutions, letting $n \to \infty$, using Lemma B.1, and letting ξ converge to the function $\chi_{[0,\tau]}$, we recover (84).

From (84) we now derive (83). From here onwards we denote the expression in the supremum on the right hand side of (84) by $\mathcal{J}(\rho, \tilde f)$ and use the notation

$$\Psi := -\gamma^2 \Big(\Delta_p \varphi - \nabla_p H \cdot \nabla_p \varphi - \frac{1}{2} |\nabla_p \varphi|^2 \Big). \tag{85}$$

Our aim is to substitute the solution f of (76a), (76b) into (84). To do this, we first extend f outside $[0, T] \times \mathbb{R}^{2d}$ by constants and define

$$f_\delta := \nu_\delta *_x f, \qquad f_{\delta,\varepsilon} := \eta_\varepsilon *_t f_\delta,$$

where $\eta_\varepsilon(t)$, $\nu_\delta(q, p)$ are again regularizing sequences in time and space. The rest of the proof is divided into the following steps:

1. We first show that $\mathcal{J}(\rho, f_{\delta,\varepsilon})$ is well defined.
2. We then successively take the limits $\varepsilon \to 0$ and $\delta \to 0$ in $\mathcal{J}(\rho, f_{\delta,\varepsilon})$.
3. We finally show that the limit satisfies (83).

Step 1 Let us first show that $\mathcal{J}(\rho, f_{\delta,\varepsilon})$ is well defined. From Theorem B.2 we know that f satisfies $|f| \leq C(1 + H)^{1/2}$, and therefore we find

$$|\partial_t f_{\delta,\varepsilon}|, \ |\Delta_p f_{\delta,\varepsilon}|, \ |J\nabla\psi *_x \rho_t \cdot \nabla f_{\delta,\varepsilon}|, \ |J\nabla H \cdot \nabla f_{\delta,\varepsilon}|, \ |\nabla H \cdot \nabla f_{\delta,\varepsilon}| \leq C(1 + H), \tag{86}$$

where the constant C depends on δ and ε. The last two objects are bounded since $|\nabla H|^2 \leq C(1+H)$; similar estimates hold for f_δ. These bounds combined with Lemma B.1 imply that the integrals in $\mathcal{J}(\rho, f_{\delta,\varepsilon})$ are well defined and using (84) it follows that

$$\mathcal{J}(\rho, f_{\delta,\varepsilon}) \leq I(\rho).$$

Step 2 Now we consider the convergence of $\mathcal{J}(\rho, f_{\delta,\varepsilon})$ as $\varepsilon \to 0$. Since all the derivatives of f in (76a) are in $L^1_{\text{loc}}((-2, T+2)\times\mathbb{R}^{2d})$ (recall that we have extended f by constant functions of (q,p) outside $[0,T]$) the same is true for the corresponding derivatives of $f_\delta := \nu_\delta *_x f$, and therefore using standard results, the following convergence results hold in $L^1_{\text{loc}}(\mathbb{R}\times\mathbb{R}^{2d})$ as $\varepsilon \to 0$,

$$f_{\delta,\varepsilon} \to f_\delta,\ \partial_t f_{\delta,\varepsilon} \to \partial_t f_\delta,\ \nabla f_{\delta,\varepsilon} \to \nabla f_\delta,\ \Delta_p f_{\delta,\varepsilon} \to \Delta_p f_\delta. \tag{87}$$

Let us first consider the single-integral terms in $\mathcal{J}(\rho, f_{\delta,\varepsilon})$. Since $f_\delta \in W^{1,1}(0,T;L^1(B_R))$ for any $R>0$, we have

$$f_{\delta,\varepsilon} \xrightarrow{\varepsilon\to 0} f_\delta \quad \text{in } W^{1,1}(0,T;L^1(B_R)),$$

which together with the trace theorem implies that

$$f_{\delta,\varepsilon}\Big|_{t=0,\tau} \xrightarrow{\varepsilon\to 0} f_\delta\Big|_{t=0,\tau} \quad \text{in } L^1(B_R) \text{ and a.e. along a subsequence.} \tag{88}$$

Since the traces of f_δ and $f_{\delta,\varepsilon}$ at $t=0,\tau$ are continuous in (q,p), this convergence holds everywhere in B_R. Combining this convergence statement with the estimate (86) and Lemma B.1 and using the dominated convergence theorem we find

$$\int_{\mathbb{R}^{2d}} \rho_t f_{\delta,\varepsilon,t}\Big|_{t=0,\tau} \xrightarrow{\varepsilon\to 0} \int_{\mathbb{R}^{2d}} \rho_t f_{\delta,t}\Big|_{t=0,\tau}.$$

Now consider the double integral in $\mathcal{J}(\rho, f_{\delta,\varepsilon})$. Using the estimate (81) with the choice

$$\Phi = \partial_t f_\delta + \mathscr{L}_{\rho_t} f_\delta + \frac{1}{2}|\nabla_p f_\delta|^2,$$

we have

$$\begin{aligned}
&\limsup_{\varepsilon\to 0} \int_0^\tau \int_{\mathbb{R}^{2d}} \Big(\partial_t f_{\delta,\varepsilon} + \mathscr{L}_{\rho_t} f_{\delta,\varepsilon} + \frac{\gamma^2}{2}|\nabla_p f_{\delta,\varepsilon}|^2 + \gamma J\nabla H\cdot\nabla\psi *_x \rho_t\Big)\, d\rho_t dt \\
&\leq \limsup_{\varepsilon\to 0} \int_0^\tau \int_{\mathbb{R}^{2d}} \Big(\eta_\varepsilon *_t \Big[\partial_t f_\delta + \mathscr{L}_{\rho_t} f_\delta + \frac{\gamma^2}{2}|\nabla_p f_\delta|^2\Big] + \gamma J\nabla H\cdot\nabla\psi *_x \rho_t\Big)\, d\rho_t dt \\
&\quad + \int_0^\tau \int_{\mathbb{R}^{2d}} \Big(\gamma J\nabla\psi *_x \rho_t\cdot\nabla f_{\delta,\varepsilon} - \eta_\varepsilon *_t (\gamma J\nabla\psi *_x \rho_t\cdot\nabla f_\delta)\Big)\, d\rho_t dt
\end{aligned}$$

Since $t \mapsto \nabla\psi *_x \rho_t$ is continuous (see Lemma B.1), it follows that for all $x\in\mathbb{R}^{2d}$

$$t \mapsto \int_{\mathbb{R}} \eta_\varepsilon(t-s)\Big[\gamma J\nabla\psi *_x \rho_t - \gamma J\nabla\psi *_x \rho_s\Big]\nabla f_\delta(s,x)\, ds \xrightarrow{\varepsilon\to 0} 0 \text{ in } L^1(0,\tau).$$

Using this convergence along with (87) we find

$$\begin{aligned}
&\limsup_{\varepsilon\to 0} \int_0^\tau \int_{\mathbb{R}^{2d}} \Big(\partial_t f_{\delta,\varepsilon} + \mathscr{L}_{\rho_t} f_{\delta,\varepsilon} + \frac{\gamma^2}{2}|\nabla_p f_{\delta,\varepsilon}|^2 + \gamma J\nabla H\cdot\nabla\psi *_x \rho_t\Big) d\rho_t dt \\
&\qquad \leq \int_0^\tau \int_{\mathbb{R}^{2d}} \Big(\partial_t f_\delta + \mathscr{L}_{\rho_t} f_\delta + \frac{\gamma^2}{2}|\nabla_p f_\delta|^2 + \gamma J\nabla H\cdot\nabla\psi *_x \rho_t\Big) d\rho_t dt.
\end{aligned} \tag{89}$$

Combining these terms and using $I(\rho) \geq \liminf_{\varepsilon\to 0} \mathcal{J}(\rho, f_{\delta,\varepsilon})$ we have

$$\int_{\mathbb{R}^{2d}} \rho\Big(f_\delta + \frac{1}{2}\psi *_x \rho\Big)\Big|_0^\tau - \int_0^\tau \int_{\mathbb{R}^{2d}} \Big(\partial_t f_\delta + \mathscr{L}_{\rho_t} f_\delta + \frac{1}{2}|\nabla_p f_\delta|^2 + \gamma J\nabla H\cdot\nabla\psi *_x \rho_t\Big) d\rho_t dt \leq I(\rho) \tag{90}$$

Now we study the $\delta \to 0$ limit of (90). Using a similar analysis as before, the following convergence results hold in $L^1_{\mathrm{loc}}(\mathbb{R}\times\mathbb{R}^{2d})$ as $\delta\to 0$,

$$f_\delta \to f,\ \partial_t f_\delta \to \partial_t f,\ \nabla f_\delta \to \nabla f,\ \Delta_p f_\delta \to \Delta_p f.$$

Since $f_T = \phi \in C_c^\infty(\mathbb{R}^{2d})$ (see Theorem B.2) and therefore $f_{\delta,T} \to f_T$ everywhere, we have

$$\int_{\mathbb{R}^{2d}} \rho_\tau f_{\delta,\tau} \xrightarrow{\delta\to 0} \int_{\mathbb{R}^{2d}} \rho_\tau f_\tau. \tag{91}$$

To pass to the limit in the right hand side of inequality (89), we use the estimate (80) with the choice $\Phi = \Psi - \gamma J\nabla H\cdot\nabla\psi *_x \rho_t$ (see (85) for the definition of Ψ), which leads to

$$\begin{aligned}
&\limsup_{\delta\to 0} \int_0^\tau \int_{\mathbb{R}^{2d}} \Big(\partial_t f_\delta + \mathscr{L}_{\rho_t} f_\delta + \frac{\gamma^2}{2}|\nabla_p f_\delta|^2 + \gamma J\nabla H\cdot\nabla\psi *_x \rho_t\Big) d\rho_t dt \\
&\leq \limsup_{\delta\to 0} \int_0^\tau \int_{\mathbb{R}^{2d}} \Big(\nu_\delta *_x \Psi - \nu_\delta *_x (\gamma J\nabla H\cdot\nabla\psi *_x \rho_t) + \gamma J\nabla H\cdot\nabla\psi *_x \rho_t\Big) d\rho_t dt \\
&\quad + \int_0^\tau \int_{\mathbb{R}^{2d}} \Big(\gamma\delta\|d^2 H\|_{L^\infty}(\nu_\delta *_x |\nabla f| + \gamma\nu_\delta *_x |\nabla_p f|) + \gamma\Big[J\nabla\psi *_x \rho_t\cdot\nabla f_\delta \\
&\quad - \nu_\delta *_x (J\nabla\psi *_x \rho_t\cdot\nabla f)\Big]\Big) d\rho_t dt \\
&= \int_0^\tau \int_{\mathbb{R}^{2d}} \Psi d\rho_t dt.
\end{aligned}$$

The only term left is the single-integral term at $t = 0$. Instead of passing to the limit, here we estimate as follows

$$\int_{\mathbb{R}^{2d}} \rho_0\Big(f_{\delta,0} + \frac{1}{2}\psi *_x \rho_0\Big) \leq \mathcal{F}(\rho_0) + \log\int_{R^{2d}} \mathrm{e}^{f_{\delta,0}-H} - \log Z_H. \tag{92}$$

Let us first prove (92). Recall from the proof of Theorem B.2 that

$$f_0 = 2\log g_0 \leq 2\log\alpha_2 + 2\beta_2 T\sqrt{\omega_2 + H},$$

where $\alpha_2, \beta_2, \omega_2$ are constants, and therefore

$$f_{\delta,0} = \nu_\delta *_x f_0 \leq 2\log\alpha_2 + \beta_2 T\Big(\delta^2\|D^2\sqrt{\omega_2+H}\|_{L_\infty} + 2\sqrt{\omega_2+H}\Big). \tag{93}$$

To arrive at the estimate above we have used

$$\begin{aligned}
\nu_\delta *_x f(x) &= \int f(x-y)\nu_\delta(y)dy \leq \int \Big(|f(x)| + |\nabla f(x)|y + \frac{1}{2}|y|^2\|d^2 f\|_{L^\infty}\Big)\nu_\delta(y)dy \\
&\leq |f(x)| + \frac{1}{2}\delta^2\|d^2 f\|_{L^\infty},
\end{aligned}$$

for any $f \in C_b^2(\mathbb{R}^{2d})$ and ν_δ satisfying $\int \nu_\delta = 1$ and $\int x\nu_\delta(x)dx = 0$.

Furthermore, using the growth conditions on $H = p^2/2m + V(q)$ (see (V1)) we find for the second derivative

$$d^2\sqrt{\omega_2 + H} = -\frac{\nabla H \otimes \nabla H}{4(\omega + H)^{3/2}} + \frac{d^2 H}{2\sqrt{\omega_2 + H}} \quad \Longrightarrow \quad \left\| d^2\sqrt{\omega_2 + H} \right\|_{L^\infty} < \infty,$$

and therefore (93) implies that $|f_{\delta,0}| \leq C(1+H)^{1/2}$. The estimate (92) then follows by using a truncated version of $f_{\delta,0}$ in the variational definition (75) of the free energy.

Substituting (92) into (90) we have

$$\begin{aligned} &\int_{\mathbb{R}^{2d}} \rho\Big(f_\delta + \frac{1}{2}\psi *_x \rho_\tau\Big)\Big|_{t=\tau} - \int_0^\tau \int_{\mathbb{R}^{2d}} \Big\{\partial_t f_\delta + \mathscr{L}_\rho f_\delta + \frac{\gamma^2}{2}|\nabla_p f_\delta|^2 + \gamma J\nabla H \cdot \nabla\psi *_x \rho_t\Big\} d\rho_t dt \\ &\leq I(\rho) + \mathcal{F}(\rho_0) + \log \int_{R^{2d}} \mathrm{e}^{f_{\delta,0} - H} - \log Z_H. \end{aligned} \tag{94}$$

Using the bound $|f_{\delta,0}| \leq C(1+H)^{1/2}$ and the dominated convergence theorem we find

$$\log \int_{R^{2d}} \mathrm{e}^{f_{\delta,0}-H} \xrightarrow{\delta\to 0} \log \int_{R^{2d}} \mathrm{e}^{f_0 - H},$$

and therefore passing to the limit $\delta \to 0$ in (94) gives

$$\begin{aligned} \int_{\mathbb{R}^{2d}} \rho_\tau\Big(f_\tau + \frac{1}{2}\psi *_x \rho_\tau\Big) - \int_0^\tau \int_{\mathbb{R}^{2d}} \Big\{\partial_t f + \mathscr{L}_\rho f + \frac{\gamma^2}{2}|\nabla_p f|^2 + \gamma J\nabla H \cdot \nabla\psi *_x \rho_t\Big\} d\rho_t dt \\ \leq I(\rho) + \mathcal{F}(\rho_0) + \log \int_{\mathbb{R}^{2d}} \mathrm{e}^{f_0 - H} dx - \log Z_H. \end{aligned}$$

□

We are now ready to prove Theorem 2.3.

Proof of Theorem 2.3 Combining (83) with Eq. (76a) we have

$$\begin{aligned} \int_{\mathbb{R}^{2d}} \rho_\tau\Big(f_\tau + \frac{1}{2}\psi *_x \rho_\tau\Big) \leq I(\rho) + \mathcal{F}(\rho_0) + \log \int_{\mathbb{R}^{2d}} \mathrm{e}^{f_0 - H} dx - \log Z_H \\ -\gamma^2 \int_0^\tau \int_{\mathbb{R}^{2d}} \Big(\Delta_p\varphi - \nabla_p H \cdot \nabla_p \varphi - \frac{1}{2}|\nabla_p\varphi|^2\Big) d\rho_t dt. \end{aligned}$$

Substituting this relation into the formula (75) for the free energy, and using $f|_{t=\tau} = \varphi$, we find

$$\begin{aligned} \mathcal{F}(\rho_\tau) &= \sup_{\phi \in C_c^\infty(\mathbb{R}^{2d})} \int_{\mathbb{R}^{2d}} \Big(\phi + \frac{1}{2}\psi *_x \rho_\tau\Big)\rho_\tau - \log \int_{\mathbb{R}^{2d}} \mathrm{e}^{\phi - H} dx + \log Z_H \\ &\leq \sup_{\phi \in C_c^\infty(\mathbb{R}^{2d})} I(\rho) + \mathcal{F}(\rho_0|\mu) + \log \int_{\mathbb{R}^{2d}} \mathrm{e}^{f_0 - H} dx - \log \int_{\mathbb{R}^{2d}} \mathrm{e}^{\phi - H} dx \\ &\quad - \gamma^2 \int_0^\tau \int_{\mathbb{R}^{2d}} \Big(\Delta_p\varphi - \nabla_p H \cdot \nabla_p\varphi - \frac{1}{2}|\nabla_p\varphi|^2\Big) d\rho_t dt. \end{aligned}$$

Rearranging and using (77) this becomes

$$\mathcal{F}(\rho_\tau) + \gamma^2 \int_0^\tau \int_{\mathbb{R}^{2d}} \Big(\Delta_p\varphi - \nabla_p H \cdot \nabla_p\varphi - \frac{1}{2}|\nabla_p\varphi|^2\Big) d\rho_t dt \leq I(\rho) + \mathcal{F}(\rho_0|\mu). \tag{95}$$

Taking the supremum over $\varphi \in C_c^\infty(\mathbb{R} \times \mathbb{R}^{2d})$ and using a standard argument, based on C^2-seperability of C_c^∞, we can move the supremum inside of the time integral and the definition of the relative Fisher Information (24) then gives

$$\mathcal{F}(\rho_\tau) + \frac{\gamma^2}{2} \int_0^\tau \mathcal{I}(\rho_t|\mu)\, dt \leq \mathcal{F}(\rho_0) + I(\rho).$$

This completes the proof. □

C Properties of the auxiliary PDE

In this appendix we will study the following equation in $[0, T] \times \mathbb{R}^{2d}$:

$$\begin{aligned} &\partial_t g - J\nabla H \cdot \nabla g - J\nabla(\psi * \rho_t) \cdot \nabla g + \nabla_p H \cdot \nabla_p g - \Delta_p g - \frac{g}{2}\left(J\nabla H \cdot \nabla\psi * \rho_t - \Psi\right) = U, \\ &g|_{t=0} = g^0. \end{aligned} \tag{96}$$

In addition to providing well-posednes results (see Sect. C.1), in this section we also prove certain important properties of this equations such as a comparison principle and bounds at infinity (see Sect. C.2).

Equation (78) is a special case of (96) with the choice

$$U = 0, \quad \Psi = -\Big(\Delta_p\varphi - \nabla_p\varphi \cdot \nabla_p H - \frac{1}{2}|\nabla_p\varphi|^2\Big).$$

Here and in the rest of this appendix we set $\gamma = 1$, since the value of γ plays no role in the discussion.

The results of this appendix are a generalization of [21, Appendix A]. In that reference Degond treats the case of Eq. (96) without on-site and interaction potentials and without the friction term $\nabla_p H \cdot \nabla_p g$. We generalize the equation, while closely following his line of argument, and proving what are essentially similar results.

The main difference in our treatment is the introduction of a weighted functional setting for the Eq. (96), in which the L^2-spaces, Sobolev spaces, and the weak formulation of the equation are all given a weight function e^{-H}. The choice of this weight function is closely connected to the fact that e^{-H} is a stationary measure both for the convective part of the equation $J\nabla H \cdot \nabla g$ and for the Ornstein–Uhlenbeck dissipative part $\nabla_p H \cdot \nabla_p g - \Delta_p g$. This weighted setting has the advantage of effectively eliminating all the unbounded coefficients in the equation.

C.1 Well-posedness

Following Degond [21] we introduce a change of variable

$$g \mapsto \mathrm{e}^{\lambda t} g, \quad \text{with } \lambda \geq \frac{1}{2}\|\Psi\|_{L^\infty} + 1, \tag{97}$$

which transforms (96) into

$$\begin{aligned} &\partial_t g - J\nabla H \cdot \nabla g - J\nabla(\psi * \rho_t) \cdot \nabla g + \nabla_p H \cdot \nabla_p g - \Delta_p g \\ &\quad - \frac{g}{2}\left(J\nabla H \cdot \nabla\psi * \rho_t\right) + \Big(\lambda + \frac{1}{2}\Psi\Big) g = \mathrm{e}^{-\lambda t} U, \\ &g|_{t=0} = g^0. \end{aligned} \tag{98}$$

In what follows we will study the well-posedness of (98), and at the end of the section we will extrapolate the results to (96).

Let us formally derive the weak formulation for (98). Multiplying with a test function $\phi \in C_c^\infty([0,T)\times\mathbb{R}^{2d})$ and a weight e^{-H}, and using integration by parts, for the left-hand side of (98) we get

$$\begin{aligned}
&\int_0^T \int_{\mathbb{R}^{2d}} \phi\Big\{\partial_t g - J\nabla H\cdot\nabla g - J\nabla(\psi * \rho_t)\cdot\nabla g + \nabla_p H\cdot\nabla_p g - \Delta_p g \\
&\quad - \frac{g}{2}\,(J\nabla H\cdot\nabla\psi * \rho_t) + \Big(\lambda + \frac{1}{2}\Psi\Big)g\Big\}\mathrm{e}^{-H} \\
&= \int_0^T \int_{\mathbb{R}^{2d}} \Big\{ g\Big(-\partial_t\phi + J\nabla H\cdot\nabla\phi + \frac{1}{2}J\nabla\psi * \rho_t\cdot\nabla\phi + \Big(\lambda+\frac{1}{2}\Psi\Big)\phi\Big) \\
&\quad - \frac{1}{2}\phi\, J\nabla\psi * \rho_t\cdot\nabla g + \nabla_p g\cdot\nabla_p\phi\Big\}\,\mathrm{e}^{-H} \\
&\quad - \int_{\mathbb{R}^{2d}} g\phi\big|_{t=0}\,\mathrm{e}^{-H}.
\end{aligned}$$

The weight e^{-H} causes cancellation of certain terms after integration by parts, as for instance for the two convolution terms,

$$\begin{aligned}
&\int_0^T \int_{\mathbb{R}^{2d}} \phi\Big(-J\nabla\psi * \rho_t\cdot\nabla g - \frac{1}{2}gJ\nabla H\cdot\nabla\psi * \rho_t\Big)\mathrm{e}^{-H} \\
&= \int_0^T \int_{\mathbb{R}^{2d}} \phi\Big(-\frac{1}{2}J\nabla\psi * \rho_t\cdot\nabla g - \frac{1}{2}J\nabla\psi * \rho_t\cdot\nabla g - \frac{1}{2}gJ\nabla H\cdot\nabla\psi * \rho_t\Big)\mathrm{e}^{-H} \\
&= \int_0^T \int_{\mathbb{R}^{2d}} \Big(-\frac{1}{2}\phi J\nabla\psi * \rho_t\cdot\nabla g + \frac{1}{2}gJ\nabla\psi * \rho_t\cdot\nabla\phi + \frac{1}{2}\phi g J\nabla H\cdot\nabla\psi * \rho_t \\
&\quad - \frac{1}{2}\phi g J\nabla H\cdot\nabla\psi * \rho_t\Big)\mathrm{e}^{-H} \\
&= \int_0^T \int_{\mathbb{R}^{2d}} \Big(-\frac{1}{2}\phi J\nabla\psi * \rho_t\cdot\nabla g + \frac{1}{2}gJ\nabla\psi * \rho_t\cdot\nabla\phi\Big)\mathrm{e}^{-H}.
\end{aligned}$$

These calculations suggest that we seek weak solutions in the space

$$X := \Big\{ g \in L^2(0,T;L^2(\mathbb{R}^{2d};\mathrm{e}^{-H})) : \nabla_p g \in L^2(0,T;L^2(\mathbb{R}^{2d};\mathrm{e}^{-H}))\Big\}, \tag{99}$$

endowed with the norm

$$\|g\|_X^2 := \|g\|^2_{L^2(L^2(\mathrm{e}^{-H}))} + \|\nabla_p g\|^2_{L^2(L^2(\mathrm{e}^{-H}))}.$$

The subscript in the norm is shorthand notation for $L^2(0,T;L^2(\mathbb{R}^{2d};\mathrm{e}^{-H}))$. Note that $C_c^\infty((0,T)\times\mathbb{R}^{2d})$ is dense in X.

We will use $\|\cdot\|_{L^2}$ to indicate the L^2 norm without any weight, and $\langle\cdot,\cdot\rangle_{X',X}$ for the dual bracket between X' (the dual of X) and X.

For all $g \in X$ we can consider the combination $\partial_t g - J\nabla H\cdot\nabla g$ as a linear form on $C_c^\infty((0,T)\times\mathbb{R}^{2d})$ by interpreting the derivatives in the sense of distributions:

$$\langle \partial_t g - J\nabla H\cdot\nabla g, \phi\rangle := -\int_0^T \int_{\mathbb{R}^{2d}} g(\partial_t\phi - J\nabla H\cdot\nabla\phi)\mathrm{e}^{-H} \qquad \text{for } \phi \in C_c^\infty((0,T)\times\mathbb{R}^{2d}).$$

Note that the weight function e^{-H} yields no extra terms upon partial integration If this linear form is bounded in the X'-norm, i.e. if the norm

$$\|\partial_t g - J\nabla H\cdot\nabla g\|_{X'} := \sup\left\{\int_0^T\int_{\mathbb{R}^{2d}} g(\partial_t\phi - J\nabla H\cdot\nabla\phi)\mathrm{e}^{-H} : \quad \phi \in C_c^\infty((0,T)\times\mathbb{R}^{2d}),\ \|\phi\|_X \le 1\right\}$$

is finite, then $\partial_t g - J\nabla H\cdot\nabla g \in X'$. We define Y to be the space of such functions g:

$$Y := \left\{g\in X : \partial_t g - J\nabla H\cdot\nabla g \in X'\right\},\quad \text{with norm } \|g\|_Y^2 := \|g\|_X^2 + \|\partial_t g - J\nabla H\cdot\nabla g\|_{X'}^2. \tag{100}$$

We now define the variational equation (which is a weak form of (98)) to be

$$E_\lambda(g,\phi) = L_\lambda(\phi),\quad \forall\phi\in C_c^\infty([0,T)\times\mathbb{R}^{2d}), \tag{101}$$

where $E_\lambda : X\times C_c^\infty([0,T)\times\mathbb{R}^{2d})\to\mathbb{R}$ and $L_\lambda : C_c^\infty([0,T)\times\mathbb{R}^{2d})\to\mathbb{R}$ are given by

$$E_\lambda(g,\phi) := \int_0^T\int_{\mathbb{R}^{2d}}\Big\{g\Big(-\partial_t\phi + J\nabla H\cdot\nabla\phi + \frac{1}{2}J\nabla\psi * \rho_t\cdot\nabla\phi + \Big(\lambda+\frac{1}{2}\Psi\Big)\phi\Big) - \frac{1}{2}\phi\, J\nabla\psi * \rho_t\cdot\nabla g + \nabla_p g\cdot\nabla_p\phi\Big\}\,\mathrm{e}^{-H}, \tag{102}$$

$$L_\lambda(\phi) := \langle \mathrm{e}^{-\lambda t}U,\phi\rangle_{X',X} + \int_{\mathbb{R}^{2d}} g^0\phi\big|_{t=0}\,\mathrm{e}^{-H}. \tag{103}$$

We use the subscript λ to indicate that that the variational Eq. (101) corresponds to the transformed Eq. (98).

We now state our main result.

Theorem C.1 (Well-posedness) *Assume that*

$$\Psi\in C_c^2(\mathbb{R}^{2d}),\quad U\in X',\quad \text{and}\quad g^0\in L^2(\mathbb{R}^{2d};\mathrm{e}^{-H}).$$

Then there exists a unique solution g in Y to the variational Eq. (101). Furthermore the solution g satisfies the initial condition in the sense of traces in $L^2(\mathbb{R}^{2d};\mathrm{e}^{-H})$.

To prove Theorem C.1, we require certain properties of Y. In the first lemma below, we prove an auxiliary result concerning the commutator of a mollification with a multiplication. In the second lemma we prove that $C_c^\infty([0,T]\times\mathbb{R}^{2d})$ is dense in Y. In order to give meaning to the initial conditions (as required in Theorem C.1) we need to prove a trace theorem. We prove this trace theorem and a Green formula (which gives meaning to 'integration by parts') in the third lemma. At the end of this section we prove Theorem C.1.

Lemma C.2 *Define $\nu_\delta(x) := \delta^{-n}\nu(\frac{x}{\delta})$ for some $\nu\in C_c^\infty(\mathbb{R}^n)$, and consider $f\in W^{1,q}(\mathbb{R}^n;R^n)$, $h\in W^{1,r}(\mathbb{R}^n)$ where $1\le q,r\le\infty$ and $1\le p<\infty$ satisfies $\frac{1}{p}=\frac{1}{q}+\frac{1}{r}$. Then for any $\delta>0$ we have*

$$\|\nu_\delta * (f\cdot\nabla h) - f\cdot\nu_\delta * \nabla h\|_{L^p} \le \left(\|\nabla f\|_{L^q}^p\left(\int_{\mathbb{R}^n}|z|\,|\nabla\nu(z)|dz\right)^p + \|\nu\|_{L^1}^p\|\operatorname{div} f\|_{L^q}^p\right)^{1/p}\|h\|_{L^r} \tag{104}$$

Proof The argument of the norm on the left hand side of (104) is

$$(\nu_\delta * (f\cdot\nabla h) - f\cdot\nu_\delta * \nabla h)(x) = \int_{\mathbb{R}^n}\nu_\delta(x-y)\,[f(x)-f(y)]\,\nabla h(y)dy$$

$$= \int_{\mathbb{R}^n} \Big(\nabla v_\delta(x-y)\,[f(x)-f(y)] + v_\delta(x-y)\,\mathrm{div}\, f(y) \Big) h(y)dy =: \mathrm{I} + \mathrm{II}.$$

Using Young's and Hölder's inequalities on the second term gives

$$\|\mathrm{II}\|_{L^p} = \|v_\delta * (h\,\mathrm{div}\, f)\|_{L^p} \le \|v_\delta\|_{L^1} \|h\,\mathrm{div}\, f\|_{L^p} \le \|v_\delta\|_{L^1} \|h\|_{L^r} \|\mathrm{div}\, f\|_{L^q}.$$

For the first term we calculate, writing $\kappa_\delta(z) := |z||\nabla v_\delta(z)|$ and $k := \|\kappa_\delta\|_{L^1}$, that

$$\begin{aligned}
\left| \int_{\mathbb{R}^n} \frac{1}{k} \nabla v_\delta(x-y)\big[f(x)-f(y)\big] h(y)dy \right|^p &\le \left(\frac{1}{k} \int_{\mathbb{R}^n} \kappa_\delta(x-y) \frac{|f(x)-f(y)|}{|x-y|} |h(y)|dy \right)^p \\
&\le \frac{1}{k} \int_{\mathbb{R}^n} \kappa_\delta(x-y) \frac{|f(x)-f(y)|^p}{|x-y|^p} |h(y)|^p dy \\
&\le \frac{\alpha^{q/p}}{k} \frac{p}{q} \int_{\mathbb{R}^n} \kappa_\delta(x-y) \frac{|f(x)-f(y)|^q}{|x-y|^q} dy + \frac{1}{k\alpha^{r/p}} \frac{p}{r} \int_{\mathbb{R}^n} \kappa_\delta(x-y) |h(y)|^r dy,
\end{aligned}$$

and therefore

$$\begin{aligned}
\|\mathrm{I}\|_{L^p}^p &= k^p \int_{\mathbb{R}^n} \left| \frac{1}{k} \int_{\mathbb{R}^n} \nabla v_\delta(x-y)\,[f(x)-f(y)]\,h(y)dy \right|^p dx \\
&\le \alpha^{q/p} k^{p-1} \frac{p}{q} \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} \kappa_\delta(x-y) \frac{|f(x)-f(y)|^q}{|x-y|^q} dy dx \\
&\quad + \frac{1}{\alpha^{r/p}} k^{p-1} \frac{p}{r} \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} \kappa_\delta(x-y) |h(y)|^r dy dx \\
&\le \alpha^{q/p} k^{p-1} \frac{p}{q}\, k\, \|\nabla f\|_q^q + \frac{1}{\alpha^{r/p}} k^{p-1} \frac{p}{r}\, k\, \|h\|_r^r.
\end{aligned}$$

By optimizing over α we find

$$\|\mathrm{I}\|_{L^p}^p \le k^p \|\nabla f\|_{L^q}^p \|h\|_{L^r}^p = \|\kappa_\delta\|_{L^1}^p \|\nabla f\|_{L^q}^p \|h\|_{L^r}^p.$$

Combining these estimates and using

$$\int_{\mathbb{R}^n} |z||\nabla v_\delta(z)| dz = \delta^{-n} \int_{\mathbb{R}^n} \frac{|z|}{\delta} |\nabla v| \left(\frac{z}{\delta} \right) dz = \int_{\mathbb{R}^n} |\tilde{z}||\nabla v(\tilde{z})| d\tilde{z} \tag{105}$$

we obtain the claimed result. □

Lemma C.3 *Let Y be the space defined in (100). Then $C_c^\infty([0,T] \times \mathbb{R}^{2d})$ is dense in Y.*

Proof We prove this lemma in two steps. In the first step we approximate functions in Y by spatially compactly supported functions. In the second step we approximate functions in Y with spatially compact support by smooth functions.

In both steps we construct an approximating sequence that converges strongly in X and weakly in X'; it then follows from Mazur's lemma that a convex combination of this sequence converges strongly in both X and X', and therefore in Y.

Step 1 For an arbitrary $g \in Y$, define $g_R(t,x) := g(t,x)\chi_R(\sqrt{H(x)})$, where $\chi_R \in C_c^\infty(\mathbb{R};\mathbb{R})$ is given by

$$\chi_R(x) = \begin{cases} 1, & |x| \le R \\ 0, & |x| > 2R \end{cases}, \quad \text{with} \quad \|\nabla \chi_R\|_{L^\infty} \le \frac{C}{R}. \tag{106}$$

Note that g_R is compactly supported in $\mathbb{R}^{2d}$. Using the dominated convergence theorem we find

$$\|g_R - g\|_X^2 = \int_0^T \int_{\mathbb{R}^{2d}} \Big[(1-\chi_R)^2(g^2 + |\nabla_p g|^2) + g^2|\nabla_p \chi_R|^2\Big] \mathrm{e}^{-H} \xrightarrow{R\to\infty} 0.$$

Here we have used $|\nabla H|^2 \le C(1+H)$ and the estimate

$$|\nabla_p \chi_R|^2 = (\chi_R'(\sqrt{H}))^2 \frac{1}{4H} |\nabla_p H|^2 \le C.$$

To conclude the first part of this proof we need to show that

$$\langle \partial_t g_R - J\nabla H \cdot \nabla g_R, \phi\rangle_{X',X} \xrightarrow{R\to\infty} \langle \partial_t g - J\nabla H \cdot \nabla g, \phi\rangle_{X',X}, \quad \forall \phi \in X. \tag{107}$$

Let $\phi \in C_c^\infty((0,T)\times\mathbb{R}^{2d})$. Then

$$\begin{aligned}
\big|\langle \partial_t g_R - J\nabla H \cdot \nabla g_R, \phi\rangle_{X',X}\big| &= \left|\int_0^T \int_{\mathbb{R}^{2d}} g_R(\partial_t \phi - J\nabla H\cdot\nabla\phi)\mathrm{e}^{-H}\right| \\
&\le \left|\int_0^T \int_{\mathbb{R}^{2d}} g\Big[\partial_t(\phi(\chi_R\circ\sqrt{H})) - J\nabla H\cdot\nabla((\chi_R\circ\sqrt{H})\phi)\Big]\mathrm{e}^{-H}\right| \\
&\quad + \left|\int_0^T \int_{\mathbb{R}^{2d}} g\phi J\nabla H\cdot\nabla(\chi_R\circ\sqrt{H})\mathrm{e}^{-H}\right| \\
&\le \|\partial_t g - J\nabla H\cdot\nabla g\|_{X'}\|\phi\|_X,
\end{aligned}$$

where we have used $J\nabla H\cdot\nabla(\chi_R\circ\sqrt{H}) = 0$ to arrive at the final inequality. As a result

$$\|\partial_t g_R - J\nabla H\cdot\nabla g_R\|_{X'} \le \|\partial_t g - J\nabla H\cdot\nabla g\|_{X'}, \tag{108}$$

and using the dominated convergence theorem we find

$$\langle \partial_t g_R - J\nabla H\cdot\nabla g_R, \phi\rangle_{X',X} \xrightarrow{R\to\infty} \langle \partial_t g - J\nabla H\cdot\nabla g, \phi\rangle_{X',X}, \quad \forall \phi\in C_c^\infty((0,T)\times\mathbb{R}^{2d}). \tag{109}$$

Estimate (108) together with the convergence statement (109) implies that (107) holds. As mentioned above, Mazur's lemma then gives the existence of a sequence that converges strongly in Y.

Step 2 In this step we approximate spatially compactly supported functions $g \in Y$ by smooth functions. Using a partition of unity (in time), it is sufficient to consider

$$\mathcal{A} := \{g \in Y : g \text{ has compact support in } [0,T)\times\mathbb{R}^{2d}\}.$$

We will show that these functions can be approximated by functions in $C_c^\infty([0,T)\times\mathbb{R}^{2d})$.

For any $g\in\mathcal{A}$, we define its translation to the left in time over $\tau > 0$ as $g_\tau(t,x) := g(t+\tau,x)$. Furthermore define $g_{\tau,\delta} = \nu_\delta * g_\tau$, where ν_δ is a symmetric regularising sequence in $\mathbb{R}\times\mathbb{R}^{2d}$. Note that $g_{\tau,\delta} \in C_c^\infty([0,T)\times\mathbb{R}^{2d})$ when δ is small enough. Using standard results it follows that $g_{\tau,\delta}\to g$ as $\tau,\delta\to 0$ in X. We will now show that

$$|\langle \partial_t g_{\tau,\delta} - J\nabla H\cdot\nabla g_{\tau,\delta}, \phi\rangle_{X',X}| \le C\|g\|_X\|\phi\|_X + \|\partial_t f - J\nabla H\cdot\nabla g\|_{X'}\|\phi\|_X, \tag{110}$$

where C is independent of τ and δ and of the test function ϕ. For any $\phi\in C_c^\infty((0,T)\times\mathbb{R}^{2d})$,

$$\langle \partial_t g_{\tau,\delta} - J\nabla H\cdot\nabla g_{\tau,\delta}, \phi\rangle_{X',X} = -\int_0^T \int_{\mathbb{R}^{2d}} (\nu_\delta * g_\tau)(\partial_t\phi - J\nabla H\cdot\nabla\phi)\mathrm{e}^{-H}$$

$$= -\int_0^T \int_{\mathbb{R}^{2d}} g_\tau \Big[\nu_\delta * (\partial_t \phi \, \mathrm{e}^{-H}) + \nu_\delta * (J \nabla \mathrm{e}^{-H} \cdot \nabla \phi) \Big] \tag{111}$$

$$\begin{aligned} &= -\int_0^T \int_{\mathbb{R}^{2d}} g_\tau \big[\partial_t (\nu_\delta * \phi) - J \nabla H \cdot (\nu_\delta * \nabla \phi) \big] \mathrm{e}^{-H} \\ &\quad - \int_0^T \int_{\mathbb{R}^{2d}} g_\tau \Big[\nu_\delta * (\partial_t \phi \, \mathrm{e}^{-H}) - (\nu_\delta * \partial_t \phi) \mathrm{e}^{-H} \Big] \\ &\quad - \int_0^T \int_{\mathbb{R}^{2d}} g_\tau \Big[\nu_\delta * (J \nabla \mathrm{e}^{-H} \cdot \nabla \phi) - J \nabla \mathrm{e}^{-H} \cdot (\nu_\delta * \nabla \phi) \Big]. \end{aligned} \tag{112}$$

We now estimate each term in the right hand side of (112). For the first term, extending the time integral to $\mathbb{R}$ and using a change of variables we find

$$\begin{aligned} &\left| \int_{\mathbb{R}} \int_{\mathbb{R}^{2d}} g(t+\tau, x) \Big(\partial_t (\nu_\delta * \phi) - J \nabla H \cdot (\nu_\delta * \nabla \phi) \Big)(t, x) \mathrm{e}^{-H(x)} dx dt \right| \\ &= \left| \int_{\mathbb{R}} \int_{\mathbb{R}^{2d}} g(s, x) \Big(\partial_t (\nu_\delta * \phi) - J \nabla H \cdot (\nu_\delta * \nabla \phi) \Big)(s-\tau, x) \mathrm{e}^{-H(x)} dx ds \right| \\ &= \left| \int_{\mathbb{R}} \int_{\mathbb{R}^{2d}} g(s, x) \Big(\partial_t (\eta \nu_\delta * \phi) - J \nabla H \cdot (\nu_\delta * \nabla \phi) \eta \Big)(s-\tau, x) \mathrm{e}^{-H(x)} dx ds \right| \\ &\le \| \partial_t g - J \nabla H \cdot \nabla g \|_{X'} \| \phi_\delta(\cdot - \tau, \cdot) \eta \|_X. \end{aligned}$$

Here $\eta \in C_c([0, T))$ is any smooth function satisfying $0 \le \eta \le 1$ and $\eta(t) = 1$ for $t \in \operatorname{supp}_t g$, and the final inequality follows by the definition of Y and $\phi_\delta(\cdot - \tau, \cdot)\eta \in C_c^\infty((0, T) \times \mathbb{R}^{2d})$. Using $\eta \le 1$ and a change of variable we obtain

$$\| \phi_\delta(\cdot - \tau, \cdot) \eta \|_X^2 \le \int_0^T \int_{\mathbb{R}^{2d}} \Big[|\phi_\delta(t-\tau, x)|^2 + |\nabla_p \phi_\delta(t-\tau, x)|^2 \Big] \mathrm{e}^{-H(x)} dx dt \le \|\phi\|_X^2,$$

and therefore for the for first term on the right hand side of (112) we have

$$\begin{aligned} &\left| \int_0^T \int_{\mathbb{R}^{2d}} g(t-\tau, x) \Big(\partial_t (\nu_\delta * \phi) - J \nabla H \cdot (\nu_\delta * \nabla \phi) \Big)(x, t) \mathrm{e}^{-H(x)} dx dt \right| \\ &\quad \le \| \partial_t g - J \nabla H \cdot \nabla g \|_{X'} \| \phi \|_X. \end{aligned}$$

For the final term in the right hand side of (112), using $\operatorname{div}(J \nabla \mathrm{e}^{-H}) = 0$ and applying Lemma C.2 with $f = J \nabla \mathrm{e}^{-H}$, $h = \phi$ and $r = p = 2$, $q = \infty$, we find

$$\begin{aligned} &\left| \int_0^T \int_{\mathbb{R}^{2d}} g_\tau \Big(\nu_\delta * (J \nabla \mathrm{e}^{-H} \cdot \nabla \phi) - J \nabla \mathrm{e}^{-H} \cdot (\nu_\delta * \nabla \phi) \Big) \right| \\ &\qquad \le \| g_\tau \|_{L^2(S)} \left\| \nu_\delta * (J \nabla \mathrm{e}^{-H} \cdot \nabla \phi) - J \nabla \mathrm{e}^{-H} \cdot (\nu_\delta * \nabla \phi) \right\|_{L^2(S)} \\ &\qquad \le \| g \|_{L^2(S)} \left\| D^2 \mathrm{e}^{-H} \right\|_{L^\infty(\mathbb{R}^{2d})} \| \phi \|_{L^2(S)} \Big(\int_S |z| |\nabla \nu(z)| dz dt \Big) \le \frac{C}{\alpha} \| g \|_X \| \phi \|_X. \end{aligned}$$

Here $S := \operatorname{supp} g$, $D^2 \mathrm{e}^{-H}$ is the Hessian of e^{-H} and $\alpha := \inf_{x \in S} \mathrm{e}^{-H(x)} > 0$. Repeating a similar calculation for the second term on the right hand side of (112), we find

$$\int_0^T \int_{\mathbb{R}^{2d}} g_\tau \Big[\nu_\delta * (\partial_t \phi \, \mathrm{e}^{-H}) - (\nu_\delta * \partial_t \phi) \mathrm{e}^{-H} \Big] \le C \| g \|_X \| \phi \|_X.$$

Combining all the terms we find (110). As a result, $\| \partial_t g_{\tau, \delta} - J \nabla H \cdot \nabla g_{\tau, \delta} \|_{X'}$ is bounded independently of τ and δ. Using the dominated convergence theorem we also have for all

$\phi \in C_c^\infty((0,T)\times\mathbb{R}^{2d})$

$$\forall\tau>0:\langle\partial_t g_{\tau,\delta}-J\nabla H\cdot\nabla g_{\tau,\delta},\phi\rangle_{X',X}\xrightarrow{\delta\to0}\langle\partial_t g_\tau-J\nabla H\cdot\nabla g_\tau,\phi\rangle_{X',X},\quad\text{and}$$
$$\langle\partial_t g_\tau-J\nabla H\cdot\nabla g_\tau,\phi\rangle_{X',X}\xrightarrow{\tau\to0}\langle\partial_t g-J\nabla H\cdot\nabla g,\phi\rangle_{X',X}$$

Taking two sequences $\tau_n\to 0$ and $\delta_n\to 0$ such that the translation and convolution operations above are allowed, we use the boundedness of $\partial_t g_{\tau,\delta}-J\nabla H\cdot\nabla g_{\tau,\delta}$ in the separable space X' to extract a subsequence that converges in the weak-star topology; we then use the density of $C_c^\infty((0,T)\times\mathbb{R}^{2d})$ in X and the convergence of $g_{\tau,\delta}$ to identify the limit. Again using Mazur's lemma it follows that there exists a strongly converging sequence. This concludes the proof of the lemma. □

Lemma C.4 *Let $g\in Y$. Then g admits (continuous) time trace values in $L^2(\mathrm{e}^{-H})$. Furthermore, for any $g,\tilde g\in Y$ we have*

$$\langle\partial_t g-J\nabla H\cdot\nabla g,\tilde g\rangle_{X',X}+\langle\partial_t\tilde g-J\nabla H\cdot\nabla\tilde g,g\rangle_{X',X}=\int_{\mathbb{R}^{2d}} g\tilde g\,\mathrm{e}^{-H}\Big|_{t=0}^{t=T}. \tag{113}$$

Proof We will prove that the mapping

$$C_c^\infty([0,T]\times\mathbb{R}^{2d})\ni g\mapsto(g(0),g(T))\in L^2(\mathrm{e}^{-H})\times L^2(\mathrm{e}^{-H}),$$

can be continuously extended to Y. This implies that any $f\in Y$ admits trace values in $L^2(\mathrm{e}^{-H})$ since $C_c^\infty([0,T]\times\mathbb{R}^{2d})$ is dense in Y by Lemma C.3. The proof of (113) follows by applying integration by parts to smooth functions and then passing to the limit in Y.

Consider $\eta\in C^\infty([0,T])$ with $0\le\eta\le 1$, $\eta(t)=1$ for $t\in[0,T/3]$, and $\eta(t)=0$ for $t\in[2T/3,T]$. We have for any $g\in C_c^\infty([0,T]\times\mathbb{R}^{2d})$

$$\begin{aligned}
\|g|_{t=0}\|^2_{L^2(\mathrm{e}^{-H})}&=\int_{\mathbb{R}^{2d}} g^2|_{t=0}\mathrm{e}^{-H}=\int_{\mathbb{R}^{2d}} g^2\eta^2|_{t=0}\mathrm{e}^{-H}=-2\int_0^T\int_{\mathbb{R}^{2d}} g\eta\,\partial_t(g\eta)\mathrm{e}^{-H}\\
&=-2\int_0^T\int_{\mathbb{R}^{2d}} g\eta(\partial_t(g\eta)-J\nabla H\cdot\nabla(g\eta))\mathrm{e}^{-H}-2\int_0^T\int_{\mathbb{R}^{2d}} g\eta\,J\nabla H\cdot\nabla(g\eta)\mathrm{e}^{-H}\\
&=2\langle(\partial_t-J\nabla H\cdot\nabla)(g\eta),g\eta\rangle_{X',X}+\int_0^T\int_{\mathbb{R}^{2d}} J\nabla\mathrm{e}^{-H}\cdot\nabla(g^2\eta^2)\\
&=2\langle(\partial_t-J\nabla H\cdot\nabla)(g\eta),g\eta\rangle_{X',X}\le 2\|(\partial_t-J\nabla H\cdot\nabla)(g\eta)\|_{X'}\|g\eta\|_X,
\end{aligned}\tag{114}$$

where the final equality follows by the anti-symmetry of J. Note that $\|g\eta\|_X\le\|g\|_X$. Furthermore

$$\begin{aligned}
\|(\partial_t-J\nabla H\cdot\nabla)(g\eta)\|_{X'}&=\sup_{\substack{\phi\in C_c^\infty((0,T)\times\mathbb{R}^{2d})\\ \|\phi\|_X=1}}\int_0^T\int_{\mathbb{R}^{2d}} g\eta(\partial_t\phi-J\nabla H\cdot\nabla\phi)\mathrm{e}^{-H}\\
&=\sup_\phi\int_0^T\int_{\mathbb{R}^{2d}} g(\partial_t(\phi\eta)-J\nabla H\nabla\phi\eta)\mathrm{e}^{-H}-\int_0^T\int_{\mathbb{R}^{2d}} g\phi\,\partial_t\eta\mathrm{e}^{-H}\\
&\le\|\partial_t g-J\nabla H\cdot\nabla g\|_{X'}+\|\partial_t\eta\|_\infty\|g\|_X\le C\|g\|_Y.
\end{aligned}$$

Substituting back into (114) we find

$$\|g|_{t=0}\|^2_{L^2(\mathrm{e}^{-H})}\le C\|g\|_Y,$$

which completes the proof for the initial time. The proof for the final time proceeds similarly. □

Now we are ready to prove Theorem C.1. We will make use of a result of Lions [48], which we state here for convenience.

Theorem C.5 *Let F be a Hilbert space, equipped with a norm $\|\cdot\|_F$ and an inner product $(\cdot,\cdot)$. Let Θ be a subspace of F, provided with a prehilbertian norm $\|\cdot\|_\Theta$, such that the injection $\Theta \hookrightarrow F$ is continuous. Consider a bilinear form E:*

$$E : F \times \Theta \ni (g,\phi) \mapsto E(g,\phi) \in \mathbb{R}$$

such that $E(\cdot,\phi)$ is continuous on F for any fixed $\phi \in \Theta$, and such that

$$|E(\phi,\phi)| \geq \alpha \|\phi\|_\Theta^2, \quad \forall \phi \in \Theta, \text{ with } \alpha > 0. \tag{115}$$

Then, given a continuous linear form L on Θ, there exists a solution g in F of the problem

$$E(g,\phi) = L(\phi), \quad \forall \phi \in \Theta.$$

Proof of Theorem C.1 We will use Theorem C.5 to show the existence of a solution to the variational equation (101). We choose $F = X$ and $\Theta = C_c^\infty([0,T)\times\mathbb{R}^{2d})$ with

$$\|\phi\|_\Theta^2 = \|\phi\|_X^2 + \frac{1}{2}\|\phi|_{t=0}\|^2_{L^2(\mathrm{e}^{-H})}, \quad \forall \phi \in \Theta.$$

By definition $\Theta \hookrightarrow X$.

The bilinear form E_λ defined in (101) satisfies property (115), since

$$\begin{aligned} E_\lambda(\phi,\phi) &= \int_0^T \int_{\mathbb{R}^{2d}} \Big\{ -\frac{1}{2}\partial_t\phi^2 + \frac{1}{4} J\nabla\psi * \rho_t \cdot \nabla\phi^2 + \Big(\lambda + \frac{1}{2}\Psi\Big)\phi^2 \\ &\quad - \frac{1}{4} J\nabla\psi * \rho_t \cdot \nabla\phi^2 + |\nabla_p\phi|^2 \Big\} \mathrm{e}^{-H} \\ &\geq \frac{1}{2}\|\phi|_{t=0}\|^2_{L^2(\mathrm{e}^{-H})} + \min\Big\{1, \lambda - \frac{1}{2}\|\Psi\|_{L^\infty}\Big\}\|\phi\|_X^2 \geq \|\phi\|_\Theta^2, \end{aligned}$$

where we have used (97).

Since all the conditions of Theorem C.5 are satisfied, the variational Eq. (101) admits a solution g in X. We have

$$\begin{aligned} &\int_0^T \int_{\mathbb{R}^{2d}} g(\partial_t\phi - J\nabla H\cdot\nabla\phi)\mathrm{e}^{-H} \\ &= \int_0^T \int_{\mathbb{R}^{2d}} g\left\{\frac{1}{2} J\nabla\psi * \rho_t \cdot \nabla\phi + \Big(\lambda + \frac{1}{2}\Psi\Big)\phi - \frac{1}{2}\phi J\nabla\psi * \rho_t \cdot \nabla g + \nabla_p g \cdot \nabla_p\phi\right\}\mathrm{e}^{-H} \\ &\quad + L_\lambda(\phi) \leq C\|g\|_X\|\phi\|_X, \end{aligned}$$

where we have used $J\nabla\psi * \rho_t \cdot \nabla\phi = -\nabla_q\psi * \rho_t \cdot \nabla_p\phi$. Note that $C > 0$ is independent of ϕ, and therefore the solution g belongs to Y.

Next we show that g^0 appearing in the definition of L_λ in (103) is the initial value for the solution g of (101). Choose $\phi(t,x) = \hat\phi(x)\bar\phi_\varepsilon(t)$, where $\hat\phi \in C_c^\infty(\mathbb{R}^{2d})$ and the sequence $\bar\phi_\varepsilon$ satisfies $\bar\phi_\varepsilon(0) = 1$, $\bar\phi_\varepsilon(t) \to 0$ for any $t \in (0,T)$ and $\bar\phi_\varepsilon' \to -\delta_0$ (Dirac delta at $t = 0$). Substituting ϕ in (101) we find

$$-\int_0^T \int_{\mathbb{R}^{2d}} g\hat\phi(x)\bar\phi_\varepsilon'(t)\mathrm{e}^{-H} = \int_{\mathbb{R}^{2d}} g^0\hat\phi(x)\mathrm{e}^{-H} + o(1) \tag{116}$$

as $\varepsilon \to 0$. By Lemma C.4, g admits trace values in $L^2(\mathbb{R}^{2d}; \mathrm{e}^{-H})$, and therefore passing $\varepsilon \to 0$ in (116) we find

$$\int_{\mathbb{R}^{2d}} \left[g(0,x) - g^0(x)\right] \hat{\phi}(x) \mathrm{e}^{-H} dx = 0, \quad \forall \hat{\phi} \in C_c^\infty(\mathbb{R}^{2d}).$$

Finally we prove the uniqueness in Y of the solution of (101). Consider two solutions $g_1, g_2 \in Y$ and let $g = g_1 - g_2$. Since the initial data and the right-hand side U in (101) vanish, we have $E_\lambda(g, \phi) = 0$ for all $\phi \in C_c^\infty([0,T) \times \mathbb{R}^{2d})$. Taking a sequence $\phi_n \in C_c^\infty([0,T) \times \mathbb{R}^{2d})$ that converges in X to g, we find

$$\begin{aligned}
0 &= \lim_{n\to\infty} E_\lambda(g, \phi_n) \\
&= \lim_{n\to\infty} \langle \partial_t g - J\nabla H \cdot \nabla g, \phi_n \rangle_{X',X} + \\
&\quad + \int_0^T \int_{\mathbb{R}^{2d}} \left\{ g \left(\frac{1}{2} J\nabla\psi * \rho_t \cdot \nabla\phi_n + \left(\lambda + \frac{1}{2}\Psi\right)\phi_n \right) - \frac{1}{2}\phi_n J\nabla\psi * \rho_t \cdot \nabla g + \nabla_p g \cdot \nabla_p \phi_n \right\} \mathrm{e}^{-H} \\
&= \langle \partial_t g - J\nabla H \cdot \nabla g, g \rangle_{X',X} + \int_0^T \int_{\mathbb{R}^{2d}} \left\{ \left(\lambda + \frac{1}{2}\Psi\right) g^2 + |\nabla_p g|^2 \right\} \mathrm{e}^{-H} \\
&\overset{(113)}{\geq} \frac{1}{2} \int_{\mathbb{R}^{2d}} g^2|_{t=T} \, \mathrm{e}^{-H} + \|g\|_X^2 \geq 0.
\end{aligned}$$

This proves uniqueness. □

Remark C.6 Using the same technique as in the uniqueness proof above we can prove the following result. If $g \in Y$ satisfies $E_\lambda(g, \phi) = L_\lambda(\phi)$ for all $\phi \in C_c^\infty([0,T) \times \mathbb{R}^{2d})$, then for all $\phi \in C([0,T] \times \mathbb{R}^{2d})$ we have

$$E_\lambda(g,\phi) = L_\lambda(\phi) - \int_{\mathbb{R}^{2d}} g\phi\big|_{t=T} \, \mathrm{e}^{-H} = \langle \mathrm{e}^{-\lambda t} U, \phi \rangle_{X',X} - \int_{\mathbb{R}^{2d}} g\phi\big|_{t=0}^{t=T} \, \mathrm{e}^{-H}.$$

Theorem C.1 proves the well-posedness of the variational Eq. (101) which is a weak form for the time-rescaled Eq. (98). Transforming back, we also conclude the well-posedness of the variational equation corresponding to the original Eq. (96). We state this in the following corollary.

Corollary C.7 *Assume that*

$$\Psi \in C_c^2(\mathbb{R} \times \mathbb{R}^{2d}), \quad U \in X', \quad \text{and} \quad g^0 \in L^2(\mathbb{R}^{2d}; \mathrm{e}^{-H}).$$

Then there exists a unique solution g to the variational equation

$$E(g,\phi) = L(\phi), \quad \forall \phi \in C_c^\infty([0,T) \times \mathbb{R}^{2d}), \tag{117}$$

in the class of functions Y. Here $E : X \times C_c^\infty([0,T) \times \mathbb{R}^{2d}) \to \mathbb{R}$ and $F : C_c^\infty([0,T) \times \mathbb{R}^{2d}) \to \mathbb{R}$ are given by

$$\begin{aligned}
E(g,\phi) &:= \int_0^T \int_{\mathbb{R}^{2d}} \left\{ g\left(-\partial_t\phi + J\nabla H \cdot \nabla\phi + \frac{1}{2} J\nabla\psi * \rho_t \cdot \nabla\phi + \frac{1}{2}\Psi\phi \right) \right. \\
&\quad \left. - \frac{1}{2}\phi \, J\nabla\psi * \rho_t \cdot \nabla g + \nabla_p g \cdot \nabla_p \phi \right\} \mathrm{e}^{-H}, \qquad (118) \\
L(\phi) &:= \langle U, \phi \rangle_{X',X} + \int_{\mathbb{R}^{2d}} g^0\phi\big|_{t=0} \, \mathrm{e}^{-H}. \qquad (119)
\end{aligned}$$

C.2 Bounds and regularity properties

Having discussed the well-posedness of Eq. (96), in this section we derive some properties of its solution. These properties play an important role in the proof of Theorem B.2.

C.2.1 Comparison principle and growth at infinity

We first provide an auxiliary lemma which we require to prove the comparison principle.

Lemma C.8 *For $g \in Y$, define $g^- \in X$ by $g^- := \max\{-g, 0\}$. Then*

$$\langle \partial_t g - J\nabla H \cdot \nabla g, g^- \rangle_{X',X} = -\frac{1}{2} \int_{\mathbb{R}^{2d}} (g^-)^2 \Big|_{t=0}^{t=T} \mathrm{e}^{-H}. \tag{120}$$

Proof Since $C_c^\infty([0,T] \times \mathbb{R}^{2d})$ is dense in Y by Lemma C.3, it is sufficient to prove (120) for $g \in C_c^\infty([0,T] \times \mathbb{R}^{2d})$. For $g \in C_c^\infty([0,T] \times \mathbb{R}^{2d})$, $g^- \in X \cap \mathrm{Lip}(\mathbb{R}^{2d})$ and there exists a sequence $\phi_n \in C_c^\infty([0,T] \times \mathbb{R}^{2d})$ such that $\phi_n \to g^-$ in X. We have

$$\begin{aligned}
\langle \partial_t g - J\nabla H \cdot \nabla g, g^- \rangle_{X',X} &= \lim_{n\to\infty} \langle \partial_t g - J\nabla H \cdot \nabla g, \phi_n \rangle_{X',X} \\
&= \lim_{n\to\infty} \int_0^T \int_{\mathbb{R}^{2d}} \phi_n (\partial_t g - J\nabla H \cdot \nabla g) \mathrm{e}^{-H} \\
&= \int_0^T \int_{\mathbb{R}^{2d}} g^- (\partial_t g - J\nabla H \cdot \nabla g) \mathrm{e}^{-H} \\
&= -\int_0^T \int_{\mathbb{R}^{2d}} g^- (\partial_t g^- - J\nabla H \cdot \nabla g^-) \mathrm{e}^{-H} \\
&\overset{(113)}{=} -\frac{1}{2} \int_{\mathbb{R}^{2d}} (g^-)^2 \Big|_{t=0}^{t=T} \mathrm{e}^{-H}.
\end{aligned}$$

□

We now prove the comparison principle.

Proposition C.9 (Comparison principle) *Let g be the solution given by Corollary C.7. Then*

1. *$g^0 \geq 0$ and $U \geq 0 \implies g \geq 0$.*
2. *$g^0 \in L^\infty(\mathbb{R}^{2d})$ and $U \in L^1(0,T; L^\infty(\mathbb{R}^{2d})) \implies g \in L^\infty([0,T] \times \mathbb{R}^{2d})$ with*

$$\|g(t)\|_{L^\infty} \leq \|g^0\|_{L^\infty} + \int_0^t \|U(s)\|_{L^\infty} ds.$$

Proof Let g be the solution of the transformed variational Eq. (101) provided by Theorem C.1, which reads explicitly

$$\begin{aligned}
0 = {} & \langle \partial_t g - J\nabla H \cdot \nabla g, \phi \rangle_{X',X} - \langle \mathrm{e}^{-\lambda t} U, \phi \rangle_{X',X} - \int_{\mathbb{R}^{2d}} g^0 \phi_0 \, \mathrm{e}^{-H} \\
& + \int_0^T \int_{\mathbb{R}^{2d}} \Big\{ g \Big(\frac{1}{2} J\nabla\psi * \rho_t \cdot \nabla\phi + \Big(\lambda + \frac{1}{2}\Psi \Big) \phi \Big) \\
& - \frac{1}{2} \phi \, J\nabla\psi * \rho_t \cdot \nabla g + \nabla_p g \cdot \nabla_p \phi \Big\} \mathrm{e}^{-H}.
\end{aligned}$$

Consider a sequence $\phi_n \to g^-$ in X as $n \to \infty$, with $\phi_n \geq 0$. Then by the assumptions on U and g^0 we have

$$\langle \mathrm{e}^{-\lambda t} U, \phi_n \rangle_{X',X} + \int_{\mathbb{R}^{2d}} g^0 \phi_n|_{t=0}\, \mathrm{e}^{-H} \geq 0,$$

and therefore

$$\begin{aligned}
0 \leq & \lim_{n\to\infty} \langle \partial_t g - J\nabla H \cdot \nabla g, \phi_n \rangle_{X',X} \\
& + \int_0^T \int_{\mathbb{R}^{2d}} \Big\{ g\Big(\frac{1}{2} J\nabla\psi * \rho_t \cdot \nabla\phi_n + \Big(\lambda + \frac{1}{2}\Psi\Big)\phi_n\Big) - \frac{1}{2}\phi_n\, J\nabla\psi * \rho_t \cdot \nabla g + \nabla_p g \cdot \nabla_p \phi_n \Big\} \mathrm{e}^{-H} \\
= & \langle \partial_t g - J\nabla H \cdot \nabla g, g^- \rangle_{X',X} \\
& + \int_0^T \int_{\mathbb{R}^{2d}} \Big\{ g\Big(\frac{1}{2} J\nabla\psi * \rho_t \cdot \nabla g^- + \Big(\lambda + \frac{1}{2}\Psi\Big) g^-\Big) - \frac{1}{2} g^-\, J\nabla\psi * \rho_t \cdot \nabla g + \nabla_p g \cdot \nabla_p g^- \Big\} \mathrm{e}^{-H} \\
= & -\frac{1}{2} \int_{\mathbb{R}^{2d}} (g^-)^2 \Big|_{t=0}^{t=T} \mathrm{e}^{-H} - \int_0^T \int_{\mathbb{R}^{2d}} \Big\{ \Big(\lambda + \frac{1}{2}\Psi\Big)|g^-|^2 + |\nabla_p g^-|^2 \Big\} \mathrm{e}^{-H},
\end{aligned}$$

where the last equality follows by Lemma C.8. Since $g^-|_{t=0} = 0$ and $\lambda \geq \frac{1}{2}\|\Psi\|_\infty + 1$ by assumption (97), this implies that $g^- = 0$.

This completes the proof of the first part of Proposition C.9. The second part is a simple consequence of the first part, by applying the first part to the function $\tilde{g} \in Y$, $\tilde{g}(t) := \|g^0\|_\infty + \int_0^t \|U(s)\|_{L^\infty}\, ds - g(t)$, which satisfies an equation of the same form. □

In the next result we use the comparison principle to prove explicit bound on the solution of Eq. (96) when $U = 0$.

Proposition C.10 (Growth) *Assume that* $\inf H = 0$ *and* $0 < \alpha_1 \leq g^0 \leq \alpha_2 < \infty$. *The solution for the variational problem* (117) *with* $U = 0$ *satisfies*

$$\alpha_1 \exp\Big(-\beta_1 t\sqrt{\omega_1 + H}\Big) \leq g \leq \alpha_2 \exp\Big(\beta_2 t\sqrt{\omega_2 + H}\Big)$$

for some fixed constants $\beta_1, \beta_2, \omega_1, \omega_2 > 0$.

Proof We first prove the second inequality in Proposition C.10. For some constants $\beta_2 > 0, \omega_2 > 1$ to be specified later, we define $g_2 := \alpha_2 \exp(\beta_2 t\sqrt{\omega_2 + H}) \in Y$, such that $g_2|_{t=0} = \alpha_2$. We will show that $g_2 - g$ satisfies the assumptions of Proposition C.9.

Substituting $g_2 - g$ in (118) and using the smoothness of g_2 we find

$$E(g_2 - g, \phi) = \langle U_2, \phi \rangle_{X',X} + \int_{\mathbb{R}^{2d}} \big(g_2|_{t=0} - g^0\big)\, \phi\, \mathrm{e}^{-H}$$

with

$$\begin{aligned}
U_2 = & \partial_t g_2 - J\nabla H \cdot \nabla g_2 - J\nabla(\psi * \rho_t) \cdot \nabla g_2 + \nabla_p H \cdot \nabla_p g_2 - \Delta_p g_2 \\
& - \frac{g_2}{2} \left(J\nabla H \cdot \nabla\psi * \rho_t - \Psi\right).
\end{aligned}$$

By construction $g_2|_{t=0} - g^0 \geq 0$. We now show that $U_2 \geq 0$. We calculate

$$\begin{aligned}
& \partial_t g_2 - J\nabla H \cdot \nabla g_2 - J\nabla(\psi * \rho_t) \cdot \nabla g_2 + \nabla_p H \cdot \nabla_p g_2 \\
& - \Delta_p g_2 - \frac{g_2}{2} \left(J\nabla H \cdot \nabla\psi * \rho_t - \Psi\right) \\
& \geq g_2\Big(\frac{1}{2}\beta_2\sqrt{\omega_2 + H} - \frac{1}{2}\left(J\nabla H \cdot \nabla\psi * \rho_t - \Psi\right) + \frac{1}{2}\beta_2\sqrt{\omega_2 + H} - c\beta_2 t - \tilde{c}\beta_2^2 t^2\Big),
\end{aligned}$$

where the constants $c, \tilde{c}$ are independent of β_2 and ω_2, using the uniform bounds on ΔH and the bound $|\nabla H|^2 \leq C(1+H)$. Because of this growth condition on ∇H, we can choose β_2, ω_2 large enough such that

$$\frac{1}{2}\beta_2\sqrt{\omega_2 + H} \geq \frac{1}{2}\left(J\nabla H \cdot \nabla\psi * \rho_t - \Psi\right).$$

Then we choose ω_2 even larger such that for any $t \in [0, T]$

$$\frac{1}{2}\beta_2\sqrt{\omega_2 + H} \geq \frac{1}{2}\beta_2\sqrt{\omega_2} \geq c\beta_2 t + \tilde{c}\beta_2^2 t^2.$$

For these values of β_2, ω_2 we therefore have

$$\begin{aligned} U_2 = {} & \partial_t g_2 - J\nabla H \cdot \nabla g_2 - J\nabla(\psi * \rho_t) \cdot \nabla g_2 + \nabla_p H \cdot \nabla_p g_2 - \Delta_p g_2 \\ & - \frac{g_2}{2}\left(J\nabla H \cdot \nabla\psi * \rho_t - \Psi\right) \geq 0. \end{aligned}$$

Using the comparison principle of Lemma C.9 we then obtain

$$g \leq \alpha_2 \exp\left(\beta_2 t\sqrt{\omega_2 + H}\right).$$

Proceeding similarly it also follows that $g_1 := \alpha_1 \exp(-\beta_1 t\sqrt{\omega_1 + H})$ is a subsolution for (96) for appropriately chosen β_1 and ω_1, and the first inequality in Proposition (C.10) follows.

In the next result we make a specific choice for Ψ (which corresponds to the Fisher Information for the VFP equation) and show that with this choice, the $L^2(\mathrm{e}^{-H})$ norm of the solution of (96) decreases in time.

Proposition C.11 *The solution g for the variational problem (117) (in the sense of Corollary C.7) with $U = 0$ and*

$$\Psi = -\Big(\Delta_p\varphi - \nabla_p\varphi \cdot \nabla_p H - \frac{1}{2}|\nabla_p\varphi|^2\Big), \tag{121}$$

for some $\varphi \in C_c^\infty([0, T] \times \mathbb{R}^{2d})$, satisfies

$$\int_{\mathbb{R}^{2d}} g^2\Big|_0^T \mathrm{e}^{-H} \leq 0.$$

Proof Let $g \in Y$ be the solution given by Corollary C.7. Since $g \in X$, there exists a sequence $\phi_n \in C_c^\infty((0, T) \times \mathbb{R}^{2d})$ such that $\phi_n \to g$ in X. Furthermore $\partial_t g - J\nabla H \cdot \nabla g \in X'$ and we have

$$\langle \partial_t g - J\nabla H \cdot \nabla g, g\rangle_{X',X} = \lim_{n\to\infty} \langle \partial_t g - J\nabla H \cdot \nabla g, \phi_n\rangle_{X',X}.$$

Using the same approximation arguments as in the proof of the comparison principle we find

$$\begin{aligned} \frac{1}{2}\int_{\mathbb{R}^{2d}} g_t^2 \mathrm{e}^{-H}\Big|_{t=0}^{t=T} &= \langle \partial_t g - J\nabla H \cdot \nabla g, g\rangle_{X',X} \\ &= \lim_{n\to\infty} \int_0^T \int_{\mathbb{R}^{2d}} \Big(\frac{1}{2}\phi_n J\nabla\psi * \rho_t \cdot \nabla g - \frac{1}{2} g J\nabla\psi * \rho_t \cdot \nabla\phi_n \\ &\quad - \nabla_p g \nabla_p \phi_n - \frac{1}{2} g\phi_n \Psi\Big)\mathrm{e}^{-H} \end{aligned}$$

$$= \int_0^T \int_{\mathbb{R}^{2d}} \Big(-|\nabla_p g|^2 - \frac{1}{2} g^2 \Psi \Big) \mathrm{e}^{-H}.$$

Using Lemma C.4 and substituting (121) into this relation we find

$$\begin{aligned} \frac{1}{2} \int_{\mathbb{R}^{2d}} g^2 \mathrm{e}^{-H} \Big|_0^T &= \int_0^T \int_{\mathbb{R}^{2d}} \Big(-|\nabla_p g|^2 + \frac{1}{2} g^2 \Big[\Delta_p \varphi - \nabla_p \varphi \cdot \nabla_p H - \frac{1}{2} |\nabla_p \varphi|^2 \Big] \Big) \mathrm{e}^{-H} \\ &= - \int_0^T \int_{\mathbb{R}^{2d}} \Big(|\nabla_p g|^2 + g \nabla_p \varphi \cdot \nabla_p g + \frac{1}{4} g^2 |\nabla_p \varphi|^2 \Big) \mathrm{e}^{-H} \leq 0. \end{aligned}$$

where the second equality follows by applying integration by parts to the $\Delta_p \varphi$ term. This completes the proof. □

C.2.2 Regularity

In this section we prove certain regularity properties for the solution of Eq. (96). We first present a general result regarding regularity of kinetic equations. This result is a combination of Theorems 1.5 and @@1.6 [14]. The main difference is that we assume more control on the second derivative with respect to momentum, which also gives us a stronger regularity in the position variable.

Proposition C.12 *Assume that*

$$\partial_t f + p \cdot \nabla_q f - \sigma \Delta_p f = g \quad \text{in } \mathbb{R} \times \mathbb{R}^{2d} \tag{122}$$

holds with $\sigma > 0$ *and*

$$f, g \in L^2(\mathbb{R} \times \mathbb{R}^{2d}), \quad \nabla_p f, \nabla_p g \in L^2(\mathbb{R} \times \mathbb{R}^{2d}).$$

Then $\Delta_p f, \nabla_q f \in L^2(\mathbb{R} \times \mathbb{R}^{2d})$, $\partial_t f \in L^2_{\mathrm{loc}}(\mathbb{R} \times \mathbb{R}^{2d})$ *and*

$$\|\nabla_q f\|_{L^2} \leq C \Big(\|\nabla_p g\|_{L^2} + \|f\|_{L^2} \Big).$$

Proof From [14, Theorem 1.5] it follows that $\Delta_p f \in L^2(\mathbb{R} \times \mathbb{R}^{2d})$ with

$$\sigma \|\Delta_p f\|_{L^2} \leq C_d \|g\|_{L^2},$$

for a constant C_d that only depends on the dimension d. This implies that the Hessian in the p-variable satisfies $D_p^2 f \in L^2(\mathbb{R} \times \mathbb{R}^{2d})$ as well.

To prove the Proposition, we first assume that $f, g \in C_c^\infty([0, T] \times \mathbb{R}^{2d})$. We will later extend the results to the low-regularity situation via regularization arguments.

Writing $(f, g) = \int_{\mathbb{R} \times \mathbb{R}^{2d}} fg$ and using integration by parts we have

$$\begin{aligned} \|\partial_{q_j} f\|_{L^2}^2 &= \big(\partial_{q_j} f, \partial_{q_j} f \big) = \big(\partial_{q_j} f, \partial_{p_j} (\partial_t + p \nabla_q) f - (\partial_t + p \nabla_q) \partial_{p_j} f \big) \\ &= \big(\partial_{q_j} f, \partial_{p_j} (\partial_t + p \nabla_q) f \big) + \big(\partial_{q_j} (\partial_t + p \nabla_q) f, \partial_{p_j} f \big) = 2 \big(\partial_{q_j} \partial_{p_j} f, \sigma \Delta_p f \big) \\ &\quad + 2 \big(\partial_{q_j} f, \partial_{p_j} g \big) \\ &\leq 0 + 2 \|\partial_{q_j} f\|_{L^2} \|\partial_{p_j} g\|_{L^2} \end{aligned} \tag{123}$$

Here we have used the (hypoelliptic) relation $\partial_{q_j} = \partial_{p_j} (\partial_t + p \nabla_q) - (\partial_t + p \nabla_q) \partial_{p_j}$ to arrive at the second equality. The final inequality follows since f is real-valued, which implies that $|\hat{f}|^2$ is an even function and therefore

$$\left(\partial_{q_j}\partial_{v_j} f_{\delta,R}, \Delta_p f\right) = \int_{\mathbb{R}^{2d}} \zeta_j \eta_j |\eta|^2 |\hat{f}|^2 = 0,$$

where ζ, η are the Fourier variables corresponding to q, p.

Inequality (123) gives

$$\|\partial_{q_j} f\|_{L^2} \le 2\|\partial_{p_j} g\|_{L^2}. \tag{124}$$

Since $\nabla_q f, \Delta_p f, g \in L^2(\mathbb{R}\times\mathbb{R}^{2d})$, using (122) we have $\partial_t f \in L^2_{\text{loc}}(\mathbb{R}\times\mathbb{R}^{2d})$. This proves the result for smooth and compactly supported f and g.

Let us now consider general $f, g \in L^2(\mathbb{R}\times\mathbb{R}^{2d})$ as in the Proposition, and define $f_\delta := \nu_\delta * f$ and $g_\delta := \nu_\delta * g$, where ν_δ is a regularizing sequence in $\mathbb{R}\times\mathbb{R}^{2d}$. Then we have

$$\partial_t f_\delta + p\cdot\nabla_q f_\delta - \Delta_p f_\delta = g_\delta + \bar{g}_\delta,$$

where $\bar{g}_\delta = p\cdot\nabla_q f_\delta - \nu_\delta * (p\nabla_q f)$. Next we define $f_{\delta,R} := f_\delta \chi_R$ and $g_{\delta,R} := g_\delta \chi_R$, where

$$\chi_R(x) = \chi_1\Big(\frac{x}{R}\Big), \text{ where } \chi_1 \in C_c^\infty(\mathbb{R}^{2d}),\ \chi_1(x) = 1 \text{ for } |x|\le 1,\ \chi_1(x) = 0 \text{ for } |x|\ge 2.$$

Then we have

$$\partial_t f_{\delta,R} + p\cdot\nabla_q f_{\delta,R} - \Delta_p f_{\delta,R} = (g_\delta + \bar{g}_\delta)\chi_R + \bar{g}_{\delta,R} =: g_{\delta,R},$$

where

$$\bar{g}_{\delta,R} = f_\delta p\cdot\nabla_q \chi_R - f_\delta \Delta_p \chi_R + \nabla_p f_\delta \cdot \nabla_p \chi_R. \tag{125}$$

Note that $f_{\delta,R}, g_{\delta,R} \in C_c^\infty(\mathbb{R}\times\mathbb{R}^{2d})$. To apply (124) we need to show that $g_{\delta,R}, \nabla_p g_{\delta,R} \in L^2(\mathbb{R}\times\mathbb{R}^{2d})$. In fact we will show that $g_{\delta,R}, \nabla_p g_{\delta,R}$ are bounded in $L^2(\mathbb{R}\times\mathbb{R}^{2d})$ independently of δ and R with

$$\|\nabla_p g_{\delta,R}\|_{L^2} \le C\left(\|\nabla_p g\|_{L^2} + \|f\|_{L^2} + \|\nabla_p f\|_{L^2}\right). \tag{126}$$

Combining with estimate (124), we have $\nabla_q f \in L^2(\mathbb{R}\times\mathbb{R}^{2d})$ with

$$\|\nabla_q f\|_{L^2} = \lim_{\delta\to 0, R\to\infty} \|\nabla_q f_{\delta,R}\|_{L^2} \le C\left(\|\nabla_p g\|_{L^2} + \|f\|_{L^2} + \|\nabla_p f\|_{L^2}\right).$$

Now we prove that $g_{\delta,R}$ satisfies inequality (126). Since the equations are defined in a distributional sense, for any $\phi \in C_c^\infty(\mathbb{R}\times\mathbb{R}^{2d})$ we have

$$\begin{aligned}
\int_{\mathbb{R}^{1+2d}} \bar{g}_\delta \phi &= \int_{\mathbb{R}^{1+2d}} \left[-f_\delta p\cdot\nabla_q\phi + f p\cdot\nabla_q \nu_\delta * \phi\right] \\
&= \int_{\mathbb{R}^{1+2d}} \left[-f\nu_\delta * (p\cdot\nabla_q\phi) + f p\cdot\nabla_q\nu_\delta * \phi\right] \\
&\le \|f\|_{L^2}\|\nu_\delta * (p\cdot\nabla_q\phi) + p\cdot\nabla_q\nu_\delta * \phi\|_{L^2} \\
&\le \|f\|_{L^2}\|\kappa_\delta\|_{L^1}\|\phi\|_{L^2} \le C\|f\|_{L^2}\|\phi\|_{L^2}.
\end{aligned}$$

where $\kappa_\delta(q,p) = |p||\nabla_q\nu_\delta(q,p)|$. Here the final inequality follows from Lemma C.2 since $\|\kappa_\delta\|_{L^1} \le C$ independent of δ (recall (105)). As a result of this calculation it follows that

$$\|\bar{g}_\delta\|_{L^2} \le C\|f\|_{L^2},$$

where C is independent of δ.

A similar calculation for $\nabla_p \bar{g}_\delta$ gives, using implicit summation over repeated indices,

$$\begin{aligned}\int_{\mathbb{R}^{2d}} \bar{g}_\delta \partial_{p_j}\phi &= \int_{\mathbb{R}^{2d}} \left[-(\nu_\delta * f)(p_i \partial_{q_i p_j}\phi) + f p_i \partial_{q_i}(\nu_\delta * \partial_{p_j}\phi)\right] \\ &= \int_{\mathbb{R}^{2d}} \left[\partial_{q_i}\phi\, \partial_{p_j}(p_i \nu_\delta * f) + f p_i \partial_{p_j}(\nu_\delta * \partial_{q_i}\phi)\right] \\ &= \int_{\mathbb{R}^{2d}} \left[\partial_{q_i}\phi\, \partial_{p_j}(p_i \nu_\delta * f) - \nu_\delta * \partial_{q_i}\phi\, \partial_{p_j}(f p_i)\right] \\ &= \int_{\mathbb{R}^{2d}} \left[\partial_{q_i}\phi \left(p_i \nu_\delta * \partial_{p_j} f + \delta_{ij}\nu_\delta * f\right) - \nu_\delta * \partial_{q_i}\phi \left(p_i \partial_{p_j} f + \delta_{ij} f\right)\right] \\ &= \int_{\mathbb{R}^{2d}} \partial_{p_j} f \left[\nu_\delta * (p_i \partial_{q_i}\phi) - p_i \nu_\delta * \partial_{q_i}\phi\right] \le C \|\partial_{p_j} f\|_{L^2} \|\nabla p_i\|_\infty \|\phi\|_{L^2},\end{aligned}$$

where C is independent of δ, implying

$$\|\partial_{p_j} \bar{g}_\delta\| \le C \|\nabla_p f\|_{L^2}.$$

Now let us consider $\bar{g}_{\delta,R}$ (defined in (125)). Since $|\nabla_p \chi_R| \le 1/R$ and $|\Delta_p \chi_R| \le 1/R^2$, it follows that

$$\|\bar{g}_{\delta,R}\|_{L^2} \le C\|f_\delta p \cdot \nabla_q \chi_R\|_{L^2} + \frac{C}{R^2}\|f_\delta\| + \frac{C}{R}\|\nabla_p f_\delta\| \le C\|f\|_{L^2} + \frac{C}{R^2}\|f\| + \frac{C}{R}\|\nabla_p f\|,$$

i.e. $\bar{g}_{\delta,R}$ is bounded in $L^2(\mathbb{R}\times\mathbb{R}^{2d})$ independent of δ, R. A similar calculation shows that

$$\|\partial_{p_j} \bar{g}_{\delta,R}\|_{L^2} \le C(R) \xrightarrow{R\to\infty} 0.$$

This completes the proof. □

We now use Proposition C.12 to prove regularity properties of Eq. (96).

Proposition C.13 *Let g be the solution of the variational problem (117) (in the sense of Corollary C.7) with $U = 0$ and with initial datum $g^0 \in X$. If $g^0 \in C^3(\mathbb{R}^{2d}) \cap X$, then g satisfies*

$$\partial_t g, \nabla g, \Delta_p g \in L^2_{\mathrm{loc}}([0,T]\times\mathbb{R}^{2d}).$$

Proof Let g be the solution of the variational problem (117) in the sense of Corollary C.7, but on the time interval $[0,\infty)$; since Corollary C.7 guarantees existence and uniqueness on any finite interval, this g is well defined. We extend g to all t by setting

$$g(t) := \begin{cases} g^0 & t \le 0 \\ g(t) & t > 0 \end{cases}$$

We next recast the variational problem (117) in the form used in Proposition C.12. Changing p to $-p$ and rearranging (117) we find, also using Remark C.6, for all $\phi \in C_c^\infty(\mathbb{R}\times\mathbb{R}^{2d})$

$$\begin{aligned}&\int_0^T \int_{\mathbb{R}^{2d}} \left\{g\left(-\partial_t\phi - p\cdot\nabla_q\phi\right) + \nabla_p g \cdot \nabla_p\phi\right\} \mathrm{e}^{-H} \\ &= \int_0^T \int_{\mathbb{R}^{2d}} \left\{-g\nabla_q V\cdot\nabla_p\phi - \frac{1}{2} g \nabla_q\psi * \rho_t \cdot \nabla_p\phi - \frac{1}{2} g \Psi\phi + \frac{1}{2}\phi\nabla_q\psi * \rho_t \cdot \nabla_p g\right\} \mathrm{e}^{-H} \\ &\quad - \int_{\mathbb{R}^{2d}} g\phi\Big|_{t=0}^{t=T} \mathrm{e}^{-H}.\end{aligned}$$

With the choice $\phi = \tilde{\phi} \mathrm{e}^H$, where $\tilde{\phi} \in C_c^\infty(\mathbb{R} \times \mathbb{R}^{2d})$ we rewrite this as

$$\begin{aligned}
&\int_0^T \int_{\mathbb{R}^{2d}} \left\{ g \left(-\partial_t \tilde{\phi} - p \cdot \nabla_q \tilde{\phi} \right) + \nabla_p g \cdot \nabla_p \tilde{\phi} \right\} \\
&= \int_0^T \int_{\mathbb{R}^{2d}} \Big\{ \nabla_p g \cdot \nabla_p H \tilde{\phi} - g \nabla_q V \cdot \nabla_p \tilde{\phi} - \frac{1}{2} g \nabla_q \psi * \rho_t \cdot \nabla_p \tilde{\phi} - \frac{1}{2} g \nabla_q \psi * \rho_t \cdot \nabla_p H \tilde{\phi} \\
&\quad - \frac{1}{2} g \tilde{\Psi} \phi + \frac{1}{2} \tilde{\phi} \nabla_q \psi * \rho_t \cdot \nabla_p g \Big\} - \int_{\mathbb{R}^{2d}} g \tilde{\phi} \Big|_{t=0}^{t=T}. \qquad (127)
\end{aligned}$$

After combining this expression with similar expressions for the regions $t > T$ and $t < 0$, we find that these expressions form the distributional version of the equation

$$\partial_t g - p \nabla_q g - \Delta_p g = G \qquad \text{in } \mathbb{R} \times \mathbb{R}^{2d}, \qquad (128)$$

where

$$G = \begin{cases} -p \nabla_q g^0 - \Delta_p g^0 & t < 0 \\ \nabla_p g \cdot \nabla_q V - \nabla_q \psi * \rho_t \cdot \nabla_p g - \nabla_p g \cdot \nabla_p H - \frac{1}{2} g \left(\nabla_q \psi * \rho_t \cdot \nabla_p H + \Psi \right) & t > 0. \end{cases} \qquad (129)$$

Since $g, \nabla_p g \in L^2(0, T; L^2(\mathrm{e}^{-H})) \subset L^2_{\mathrm{loc}}(\mathbb{R} \times \mathbb{R}^{2d})$ and by assumption $g^0 \in C^3(\mathbb{R}^{2d})$, it follows that $G \in L^2_{\mathrm{loc}}(\mathbb{R} \times \mathbb{R}^{2d})$. After a smooth truncation, Theorem 1.5 of [14] implies that $\Delta_p g \in L^2_{\mathrm{loc}}(\mathbb{R} \times \mathbb{R}^{2d})$. Using this additional regularity in the definition of G (129), it then follows that $\nabla_p G \in L^2_{\mathrm{loc}}(\mathbb{R} \times \mathbb{R}^{2d})$. Applying Proposition C.12 to a truncated version of (128) then implies the result. □

Remark C.14 From Proposition C.13 it follows that the solution for the variational problem (117) satisfies the original equation (96) (with the choice $U = 0$)

$$\partial_t g - J \nabla H \cdot \nabla g - J \nabla (\psi * \rho_t) \cdot \nabla g + \nabla_p H \cdot \nabla_p g - \Delta_p g - \frac{g}{2} \left(J \nabla H \cdot \nabla \psi * \rho_t - \Psi \right) = 0,$$

$$g|_{t=0} = g^0,$$

in $L^1_{\mathrm{loc}}([0, T] \times \mathbb{R}^{2d})$ (i.e. all derivatives are in L^1_{loc}).

D Proof of Theorem 3.1

In this section we prove Theorem 3.1. We will use the following alternative definition of the rate functional

$$I(\rho) = \begin{cases} \frac{1}{2} \int_0^T \int_{\mathbb{R}^{2d}} |h_t|^2 \, d\rho_t dt & \text{if } \partial_t \rho_t = \varepsilon^{-1} \operatorname{div}(\rho J \nabla H) + \Delta_p \rho - \operatorname{div}_p(\rho_t h_t), \text{ for } h \in L^2(0, T; L^2_\nabla(\rho)), \\ & \quad \text{and } \rho|_{t=0} = \rho_0, \\ +\infty & \text{otherwise,} \end{cases} \qquad (130)$$

where $\varepsilon > 0$ is fixed.

Proof of Theorem 3.1 We first show that the estimate (49) holds. Since ρ satisfies $I(\rho) < C$, using the defintion (130) of the rate functional we find that there exists $h \in L^2(0, T; L^2_\nabla(\rho))$ such that for any $f \in C_c^2(\mathbb{R}^2)$

$$\frac{d}{dt}\int_{\mathbb{R}^2} f\,d\rho_t = \int_{\mathbb{R}^2}\Big(\frac{1}{\varepsilon}J\nabla H\cdot\nabla f + \Delta_p f + \nabla_p f\cdot h_t\Big)d\rho_t. \tag{131}$$

Formally substituting $f = H$ in (131) and using the growth conditions on H (see (A2)) we find

$$\begin{aligned}\partial_t\int_{\mathbb{R}^2} H\,d\rho_t = \int_{\mathbb{R}^2}\big(\Delta_p H + \nabla_p H\cdot h_t\big)\,d\rho_t &\le C + \frac{1}{2}\int_{\mathbb{R}^2}|\nabla_p H|^2\,d\rho_t + \frac{1}{2}\int_{\mathbb{R}^2}|h_t|^2\,d\rho_t\\ &\le C + C\int_{\mathbb{R}^2} H\,d\rho_t + \frac{1}{2}\int_{\mathbb{R}^2}|h_t|^2\,d\rho_t.\end{aligned}$$

The bound $\int H\rho_t^\varepsilon < C$ then follows by applying a Gronwall-type estimate, integrating in time over $[0, T]$, and using the fact that $h \in L^2(0, T; L^2_\nabla(\rho))$. To make the choice $f = H$ admissible in the definition (130) of the rate functional we use a two-step approximating argument. We first extend the class of admissible functions from $C_c^2(\mathbb{R}^2)$ to

$$\mathcal{A} := \Big\{F \in C_b^2(\mathbb{R}^2) : \sup_{x\in\mathbb{R}^2}(1+|x|)|F(x)| < \infty\Big\}.$$

For a given $F \in \mathcal{A}$, define the sequence $f_k(x) = F(x)\xi_k(x) \in C_c^2(\mathbb{R}^2)$, where $\xi_k \in C_c^\infty(\mathbb{R})$ is a sequence of smoothed characteristic functions converging pointwise to one, with $0 \le \xi_k \le 1$, $|\nabla\xi_k| \le 1/k$, and $|d^2\xi_k| \le 1/k^2$. Then $|\nabla H\cdot\nabla f_k|$, $\Delta_p f_k$, and $|\nabla_p f|^2$ are bounded uniformly and converge pointwise to the corresponding terms with f_k replaced by f; convergence follows by the Dominated Convergence Theorem. In the second step, we extend $\mathcal{A}$ to include $H(q, p)$ by using an approximating sequence $\mathcal{A} \ni g_k(q, p) = H(q, p)\psi_k(H(q, p))$ where $\psi_k : \mathbb{R} \to \mathbb{R}$ is defined as $\psi_k(s) := (1 + |s|/k)^{-2}$. Note that $\psi_k \to 1$ pointwise as $k \to \infty$. Proceeding as described in the formal calculations above we find

$$\partial_t\Big(\int g_k d\rho_t\Big) \le C\Big(1 + \int g_k d\rho_t + \int |h_t|^2 d\rho_t\Big),$$

where C is independent of k and ε. Using a Gronwall-type estimate, integrating in time over $[0, T]$ and applying the monotone convergence theorem we find (49).

Next we prove (50). The main idea of the proof is to consider a modified equation for which an estimate of the type (50) holds, and then arrive at (50) by passing to an appropriate limit.

We consider the following modification of Eq. (41),

$$\partial_t\rho = -\frac{1}{\varepsilon}\operatorname{div}(\rho J\nabla H) + \alpha\operatorname{div}_p(\rho\nabla_p H) + \Delta_p\rho^\varepsilon, \tag{132}$$

where $\alpha > 0$. Essentially, we have added a friction term to Eq. (41), as a result of which $\mu^\alpha(dqdp) = Z_\alpha^{-1}\mathrm{e}^{-\alpha H(q,p)}dqdp$ is a stationary measure for (132) (Z_α is the normalization constant).

The rate functional corresponding to (132) is

$$I_\alpha(\rho) = \begin{cases}\dfrac{1}{2}\displaystyle\int_0^T\int_{\mathbb{R}^2}|h_t^\alpha|^2\,d\rho_t dt & \text{if } \partial_t\rho_t = -\varepsilon^{-1}\operatorname{div}(\rho J\nabla H) + \Delta_p\rho + \operatorname{div}_p(\rho[\alpha\nabla_p H - h_t^\alpha]),\\ & \quad \text{for } h^\alpha \in L^2(0, T; L^2_\nabla(\rho_t)), \text{ and } \rho|_{t=0} = \rho_0,\\ +\infty & \text{otherwise.}\end{cases} \tag{133}$$

Note that Eq. (132) is a special case of the VFP equation (with the choice $\psi = 0$) and therefore the proof of Theorem 2.3 also applies to this case. We follow the proof up to (95) (adding a constant α to the friction) to find for any $\tau \in [0, T]$

$$\mathcal{H}(\rho_\tau|\mu^\alpha) + \int_0^\tau \int_{\mathbb{R}^2} \Big(\Delta_p\varphi - \alpha\nabla_p H\cdot\nabla_p\varphi - \frac{1}{2}|\nabla_p\varphi|^2\Big)d\rho_t dt \le I_\alpha(\rho) + \mathcal{H}(\rho_0|\mu^\alpha),$$

for any $\varphi \in C_c^\infty(\mathbb{R}\times\mathbb{R}^2)$. Using the definition of relative entropy we have

$$\begin{aligned}\mathcal{F}(\rho_\tau) + \int_0^\tau \int_{\mathbb{R}^2} \Big(\Delta_p\varphi - \alpha\nabla_p H\cdot\nabla_p\varphi - \frac{1}{2}|\nabla_p\varphi|^2\Big)d\rho_t dt \le I_\alpha(\rho) + \mathcal{F}(\rho_0)\\ +\alpha\int_{\mathbb{R}^2} H\rho_\tau - \alpha\int_{\mathbb{R}^2} H\rho_0. \qquad (134)\end{aligned}$$

Below we show that $I_\alpha(\rho) \to I(\rho)$ as $\alpha \to 0$. Then passing to the limit $\alpha \to 0$ in (134) we find

$$\mathcal{F}(\rho_\tau) + \int_0^\tau \int_{\mathbb{R}^2} \Big(\Delta_p\varphi - \frac{1}{2}|\nabla_p\varphi|^2\Big)d\rho_t dt \le I(\rho) + \mathcal{F}(\rho_0),$$

where we have used $|\nabla_p H|^2 \le C(1+H)$ along with the estimate (49). The required inequality (50) then follows by taking the supremum over $\varphi \in C_c^\infty(\mathbb{R}\times\mathbb{R}^2)$.

To complete the proof we show that $I_\alpha(\rho) \to I(\rho)$ as $\alpha \to 0$. Using the definition of the rate functionals for the original Eq. (130) and the modified Eq. (133), we write the rate functional for the modified equation as

$$\begin{aligned}I_\alpha(\rho) &= \frac{1}{2}\int_0^T \int_{\mathbb{R}^2} |h_t^\alpha|^2 d\rho_t dt = \frac{1}{2}\int_0^T \int_{\mathbb{R}^2} |h_t - \alpha\nabla_p H|^2 d\rho_t dt\\ &= \frac{1}{2}\int_0^T \int_{\mathbb{R}^2} \Big(|h_t|^2 + \alpha^2|\nabla_p H|^2 - 2\alpha\nabla_p H\cdot h_t\Big)d\rho_t dt \xrightarrow{\alpha\to 0} I(\rho),\end{aligned}$$

where we have used $|\nabla_p H|^2 \le C(1+H)$ and the estimate (49) to arrive at the convergence statement. Note that (49) along with the definition of the rate functionals implies that $I(\rho) < \infty$ iff $I_\alpha(\rho) < \infty$. □

References

1. Aronson, D., Crandall, M.G., Peletier, L.A.: Stabilization of solutions of a degenerate nonlinear diffusion problem. Nonlinear Anal. **6**(10), 1001–1022 (1982)
2. Adams, S., Dirr, N., Peletier, M.A., Zimmer, J.: From a large-deviations principle to the Wasserstein gradient flow: a new micro–macro passage. Commun. Math. Phys. **307**, 791–815 (2011)
3. Adams, S., Dirr, N., Peletier, M.A., Zimmer, J.: Large deviations and gradient flows. Philos. Trans. R. Soc. A Math. Phys. Eng. Sci. **371**(2005), 20120341 (2013)
4. Ambrosio, L., Gigli, N., Savaré, G.: Gradient Flows in Metric Spaces and in the Space of Probability Measures. Lectures in Mathematics ETH Zürich, Birkhäuser (2008)
5. Allaire, G.: Homogenization and two-scale convergence. SIAM J. Math. Anal. **23**, 1482 (1992)
6. Arnrich, S., Mielke, A., Peletier, M.A., Savaré, G., Veneroni, M.: Passing to the limit in a Wasserstein gradient flow: from diffusion to reaction. Calc. Var. Part. Differ. Eq. **44**, 419–454 (2012)
7. Ambrosio, L., Savaré, G., Zambotti, L.: Existence and stability for Fokker–Planck equations with log-concave reference measure. Probab. Theory Relat. Fields **145**(3), 517–564 (2009)
8. Bonilla, L.L., Carrillo, J.A., Soler, J.: Asymptotic behavior of an initial-boundary value problem for the Vlasov–Poisson–Fokker–Planck system. SIAM J. Appl. Math. **57**(5), 1343–1372 (1997)

9. Budhiraja, A., Dupuis, P., Fischer, M.: Large deviation properties of weakly interacting processes via weak convergence methods. Ann. Prob. **40**(1), 74–102 (2012)
10. Brezis, H., Ekeland, I.: Un principe variationnel associé à certaines equations paraboliques. Le cas indépendant du temps. Comptes Rendus de l'Acad. des Sci. de Paris Série A **282**, 971–974 (1976)
11. Bakry, D., Gentil, I., Ledoux, M., et al.: Analysis and Geometry of Markov Diffusion Operators. Grundlehren der Mathematischen Wissenschaften 348. Springer, Cham (2014)
12. Bertsch, M., Kersner, R., Peletier, L.A.: Positivity versus localization in degenerate diffusion equations. Nonlinear Anal. **9**(9), 987–1008 (1985)
13. Bogachev, V.I., Krylov, N.V., Röckner, M., Shaposhnikov, S.V.: Fokker-Planck-Kolmogorov Equations, vol. 207. American Mathematical Soc, Rhode Island, Providence (2015)
14. Bouchut, F.: Hypoelliptic regularity in kinetic equations. J. de Math. Pures et Appl. **81**(11), 1135–1159 (2002)
15. Barret, F., von Renesse, M.: Averaging principle for diffusion processes via Dirichlet forms. Potential Anal. **41**(4), 1033–1063 (2014)
16. Cioranescu, D., Damlamian, A., Griso, G.: Periodic unfolding and homogenization. Comptes Rendus Math. **335**(1), 99–104 (2002)
17. Cioranescu, D., Damlamian, A., Griso, G.: The periodic unfolding method in homogenization. SIAM J. Math. Anal. **40**, 1585 (2008)
18. Cerrai, S., Freidlin, M.: On the Smoluchowski–Kramers approximation for a system with an infinite number of degrees of freedom. Probab. Theory Relat. Fields **135**(3), 363–394 (2006)
19. Crandall, M., Ishii, H., Lions, P.: User's guide to viscosity solutions of second order partial differential equations. Am. Math. Soc **27**, 1–67 (1992)
20. Dupuis, P., Ellis, R.S.: A Weak Convergence Approach to the Theory of Large Deviations, vol. 902. John Wiley and & Sons, New York (1997)
21. Degond, P.: Global existence of smooth solutions for the Vlasov–Fokker–Planck equation in 1 and 2 space dimensions. Ann. Sci. de l'École Normale Supérieure **19**(4), 519–542 (1986)
22. Dawson, D.A., Gartner, J.: Large deviations from the McKean–Vlasov limit for weakly interacting diffusions. Stochastics **20**(4), 247–308 (1987)
23. Duong, M.H., Lamacz, A., Peletier, M.A., Schlichting, A., Sharma, U.: Quantification of coarse-graining error in Langevin and overdamped Langevin dynamics. (in preparation)
24. Dirr, N., Laschos, V., Zimmer, J.: Upscaling from particle models to entropic gradient flows. J. Math. Phys. **53**(6), 063704 (2012)
25. de Pablo, J.J., Curtin, W.A.: Multiscale modeling in advanced materials research: challenges, novel methods, and emerging applications. Mrs Bull. **32**(11), 905–911 (2007)
26. Duong, M.H., Peletier, M.A., Zimmer, J.: GENERIC formalism of a Vlasov–Fokker–Planck equation and connection to large-deviation principles. Nonlinearity **26**, 2951–2971 (2013)
27. Duong, M.H., Peletier, M.A., Zimmer, J.: Conservative–dissipative approximation schemes for a generalized Kramers equation. Math. Methods Appl. Sci. **37**(16), 2517–2540 (2014)
28. Daneri, S., Savaré, G.: Lecture notes on gradient flows and optimal transport. arXiv preprint arXiv:1009.3737, (2010)
29. Eidus, D.: The Cauchy problem for the non-linear filtration equation in an inhomogeneous medium. J. Differ. Equ. **84**, 309–318 (1990)
30. Feller, W.: The parabolic differential equations and the associated semi-groups of transformations. Ann. Math. **55**(3), 468–519 (1952)
31. Frank, J., Gottwald, G.A.: The Langevin limit of the Nosé–Hoover–Langevin thermostat. J. Stat. Phys. **143**(4), 715–724 (2011)
32. Feng, J., Kurtz, T.G.: Large Deviations for Stochastic Processes, Volume 131 of Mathematical Surveys and Monographs. American Mathematical Society, Rhode Island, Providence (2006)
33. Fleming, G., Ratner, M. (eds). Directing Matter and Energy: Five Challenges for Science and the Imagination. Basic Energy Sciences Advisory Committee (2007)
34. Freidlin, M.: Some remarks on the Smoluchowski–Kramers approximation. J. Stat. Phys. **117**(3–4), 617–634 (2004)
35. Funaki, T.: A certain class of diffusion processes associated with nonlinear parabolic equations. Z. Wahrscheinlichkeitstheorie und Verwandte Gebiete **67**(3), 331–348 (1984)
36. Freidlin, M.I., Wentzell, A.D.: Diffusion processes on graphs and the averaging principle. Ann. Prob. **21**(4), 2215–2245 (1993)
37. Freidlin, M.I., Wentzell, A.D.: Random perturbations of Hamiltonian systems. Mem. Am. Math. Soc. **109**, 523 (1994)
38. Freidlin, M., Weber, M.: Random perturbations of nonlinear oscillators. Ann. Prob. **26**(3), 925–967 (1998)

39. Freidlin, M., Weber, M.: On random perturbations of Hamiltonian systems with many degrees of freedom. Stoch. Process. Appl. **94**(2), 199–239 (2001)
40. Freidlin, M.I., Wentzell, A.D.: Diffusion processes on an open book and the averaging principle. Stoch. Process. Appl. **113**(1), 101–126 (2004)
41. Ghoussoub, N.: Self-Dual Partial Differential Systems and Their Variational Principles. Springer, New York (2009)
42. Grunewald, N., Otto, F., Villani, C., Westdickenberg, M.G.: A two-scale approach to logarithmic Sobolev inequalities and the hydrodynamic limit. Ann. Inst. H. Poincaré Probab. Stat. **45**(2), 302–351 (2009)
43. Hottovy, S., Volpe, G., Wehr, J.: Noise-induced drift in stochastic differential equations with arbitrary friction and diffusion in the Smoluchowski–Kramers limit. J. Stat. Phys. **146**(4), 762–773 (2012)
44. Ishii, H., Souganidis, P.E.: A pde approach to small stochastic perturbations of Hamiltonian flows. J. Differ. Eq. **252**(2), 1748–1775 (2012)
45. Kramers, H.: Brownian motion in a field of force and the diffusion model of chemical reactions. Physica **7**(4), 284–304 (1940)
46. Kružkov, S.N.: First order quasilinear equations in several independent variables. Mat. USSR Sb. **10**(2), 217–243 (1970)
47. Kinderlehrer, D., Stampacchia, G.: An Introduction to Variational Inequalities and Their Applications. SIAM, Classics in Applied Mathematics (2000)
48. Lions, J.L.: Équations différentielles opérationnelles et problèmes aux limites. Die Grundlehren der mathematischen Wissenschaften, Bd. 111. Springer, Berlin (1961)
49. Legoll, F., Lelièvre, T.: Effective dynamics using conditional expectations. Nonlinearity **23**(9), 2131 (2010)
50. Mandl, P.: Analytical Treatment of One-Dimensional Markov Processes. Academia, Publishing House of the Czechoslovak Academy of Sciences (1968)
51. Mielke, A.: On evolutionary gamma-convergence for gradient systems. Technical Report 1915, WIAS, Berlin (2014)
52. Mielke, A., Peletier, M.A., Renger, D.R.M.: On the relation between gradient flows and the large-deviation principle, with applications to Markov chains and diffusion. Potential Anal. **41**(4), 1293–1327 (2014)
53. Mielke, A., Roubíček, T., Stefanelli, U.: Γ-limits and relaxations for rate-independent evolutionary problems. Cal. Var. Partial Differ. Equ. **31**(3), 387–416 (2008)
54. Mielke, A., Rossi, R., Savaré, G.: Variational convergence of gradient flows and rate-independent evolutions in metric spaces. Milan J. Math. **80**(2), 381–410 (2012)
55. Murat, F.: A survey on compensated compactness. Contrib. Modern Cal. Var. **148**, 145–183 (1987)
56. Narita, K.: Asymptotic behavior of fluctuation and deviation from limit system in the Smoluchowski–Kramers approximation for SDE. Yokohama Math. J. **42**(1), 41–76 (1994)
57. Nayroles, B.: Deux théoremes de minimum pour certains systèmes dissipatifs. C. R. Acad. Sci. Paris Ser. A B **282**, A1035–A1038 (1976)
58. Nelson, E.: Dynamical Theories of Brownian Motion, vol. 17. Princeton University Press, Princeton (1967)
59. Nicolis, G., Nicolis, C.: Foundations of Complex Systems: Emergence, Information and Predicition. World Scientific, Singapore (2012)
60. Oelschläger, K.: A martingale approach to the law of large numbers for weakly interacting stochastic processes. Ann. Prob. **12**(2), 458–479 (1984)
61. Ottobre, M., Pavliotis, G.A.: Asymptotic analysis for the generalized Langevin equation. Nonlinearity **24**(5), 1629–1653 (2011)
62. Öttinger, H.: Beyond Equilibrium Thermodynamics. Wiley, Hoboken (2005)
63. Peletier, M.A., Duong, M.H., Sharma, U.: Coarse-graining and fluctuations: two birds with one stone. In: Oberwolfach Reports, vol. 10(4) (2013)
64. Pennacchio, M., Savaré, G., Colli Franzone, P.: Multiscale modeling for the bioelectric activity of the heart. SIAM J. Math. Anal. **37**(4), 1333–1370 (2005)
65. Rosenau, P., Kamin, S.: Non-linear diffusion in a finite mass medium. Commun. Pure Appl. Math. **35**, 113–127 (1982)
66. Stainforth, D.A., Allen, M.R., Tredger, E.R., Smith, L.A.: Confidence, uncertainty and decision-support relevance in climate predictions. Philos. Trans. A **365**(1857), 2145 (2007)
67. Serfaty, S.: Gamma-convergence of gradient flows on Hilbert and metric spaces and applications. Discrete Contin. Dyn. Syst. A **31**(4), 1427–1451 (2011)
68. Sharma, U.: Coarse-Graining of Fokker–Planck Equations. Ph.D. thesis, Eindhoven University of Technology (2017)
69. Smoller, J.: Shock Waves and Reaction-Diffusion Equations. Springer, New York (1994)

70. Sandier, E., Serfaty, S.: Gamma-convergence of gradient flows with applications to Ginzburg–Landau. Commun. Pure Appl. Math. **57**(12), 1627–1672 (2004)
71. Stefanelli, U.: The Brezis–Ekeland principle for doubly nonlinear equations. SIAM J. Control Optim. **47**, 1615 (2008)
72. Tartar, L.: Compensated compactness and applications to partial differential equations. Nonlinear Anal. Mech. Heriot Watt Symp. **4**, 136–212 (1979)