GENERIC formalism of a Vlasov–Fokker–Planck equation and connection to large-deviation principles

Manh Hong Duong[1]**, Mark A Peletier**[2] **and Johannes Zimmer**[3]

[1] Department of Mathematics and Computer Science, Eindhoven University of Technology, Den Dolech 2, PO Box 513, 5600 MB Eindhoven, The Netherlands
[2] Department of Mathematics and Computer Science and Institute for Complex Molecular Systems, Eindhoven University of Technology, Den Dolech 2, PO Box 513, 5600 MB Eindhoven, The Netherlands
[3] Department of Mathematical Sciences, University of Bath, Bath BA2 7AY, UK

E-mail: m.h.duong@tue.nl, m.a.peletier@tue.nl and zimmer@maths.bath.ac.uk

Abstract

In this paper we discuss the connections between a Vlasov–Fokker–Planck equation and an underlying microscopic particle system, and we interpret those connections in the context of the GENERIC framework (Öttinger 2005 *Beyond Equilibrium Thermodynamics* (New York: Wiley-Interscience)). This interpretation provides (a) a variational formulation for GENERIC systems, (b) insight into the origin of this variational formulation, and (c) an explanation of the origins of the conditions that GENERIC places on its constitutive elements, notably the so-called degeneracy or non-interaction conditions. This work shows how the general connection between large-deviation principles on one hand and gradient-flow structures on the other hand extends to non-reversible particle systems.

Mathematics Subject Classification: 60F10, 49S99, 35Q99

1. Introduction

1.1. Overview

The framework GENERIC [34] provides a systematic method to derive thermodynamically consistent evolution equations. It was originally introduced in the context of complex fluids [41, 42], and more recently has been applied to anisotropic inelastic solids [31], to viscoplastic solids [30], and thermoelastic dissipative materials [32]. The key ingredients of GENERIC are its building blocks: a Poisson operator L, a dissipative operator M, an energy

functional E, and an entropy functional S, which are required to satisfy certain properties. Although many equations have been shown to have a GENERIC structure, two important aspects have not been addressed.

The first one is the relationship between the GENERIC framework on one hand and large deviations of underlying microscopic particle systems on the other. It is well known that many deterministic evolution equations can be derived as hydrodynamic limits of a stochastic particle system. More recently it has become clear that the connection between particle systems and their upscaled limits runs deeper: gradient-flow structures of the limit equations arise as characterizations of the large-deviation behaviour of the stochastic particle systems, thus explaining amongst other things the origin of the Wasserstein gradient flows [15, 18, 39, 40, 48, 49]. In this paper we generalize this relationship beyond gradient flows to an example from the class of GENERIC systems.

The second aspect is a variational structure for GENERIC systems. The study of variational structure has important consequences for the analysis of an evolution equation. It provides general methods for proving well-posedness [43] and characterizing large-time behaviour (e.g., [38]), gives rise to natural numerical discretizations (e.g. [37]), and creates handles for the analysis of singular limits (e.g. [27–29]). The appearance of the concepts of energy and entropy in the formulation of GENERIC suggests a strong variational connection, but to date this has not been made explicit. In this paper we exhibit such a variational structure, and as in the case of the gradient flows, this structure is intimately tied to the large-deviation behaviour of an underlying system.

In this paper we treat some of these questions in full generality, that is, for a general, abstract GENERIC system. Because of this generality the treatment is necessarily formal. We illustrate the abstract features with a specific system, that of the *Vlasov–Fokker–Planck (VFP) equation* , for which the large-deviation behaviour has been proved rigorously. This gives a specific case in which the impact of the abstract arguments can be recognized. We first introduce the specific example and then explain the GENERIC framework in detail.

1.2. A VFP equation and its generalisation

The central example of this paper will be the following VFP equation,

$$\partial_t \rho = -\text{div}_q\left(\rho \frac{p}{m}\right) + \text{div}_p \rho\left(\nabla_q V + \nabla_q \psi * \rho + \gamma \frac{p}{m}\right) + \gamma\theta\Delta_p\rho. \tag{1}$$

The spatial domain is $\mathbf{R}^{2d}$ with coordinates (q, p), with q and p each in $\mathbf{R}^d$. We use subscripts as in div_q and Δ_p to indicate that the differential operators act only on those variables. The unknown is a time-dependent probability measure $\rho\colon [0, T] \to \mathcal{P}(\mathbf{R}^{2d})$; the functions $V = V(q)$ and $\psi = \psi(q)$ are given, as are the positive constants γ, m, and θ. The convolution $\psi * \rho$ is defined by $(\psi * \rho)(q) = \int_{\mathbf{R}^{2d}} \psi(q - q')\rho(q', p')\,\mathrm{d}q'\,\mathrm{d}p'$.

Equation (1) arises as the many-particle limit of a collection of *interacting Brownian particles with inertia*, given by the following stochastic differential equation

$$\mathrm{d}Q_i(t) = \frac{P_i(t)}{m}\,\mathrm{d}t, \tag{2a}$$

$$\mathrm{d}P_i(t) = -\nabla V(Q_i(t))\,\mathrm{d}t - \sum_{j=1}^{n}\nabla\psi(Q_i(t) - Q_j(t)) - \frac{\gamma}{m}P_i(t)\,\mathrm{d}t + \sqrt{2\gamma\theta}\,\mathrm{d}W_i(t). \tag{2b}$$

Here Q_i and P_i are the position and momentum of particle $i = 1, \ldots, n$, with mass m, and the equations describe the movement of this particle under a fixed potential V, an interaction potential ψ, a friction force (the drift term $-\gamma P_i\mathrm{d}t/m$) and a stochastic forcing described by the n independent d-dimensional Wiener measures W_i.

Both the friction force and the noise term arise from collisions with the solvent, and the parameter γ in both terms characterizes the intensity of these collisions. 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 the 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 or gravitational forces[4].

Equation (1) is the many-particle limit of the SDE (2), also known as the hydrodynamic limit, in the sense that as $n \to \infty$, the *empirical measure*

$$\rho_n(t) := \frac{1}{n} \sum_{i=1}^{n} \delta_{(Q_i(t), P_i(t))}$$

converges almost surely to the solution of (1) with appropriate initial data. Equation (1) has been extensively studied, especially in the case in which ψ is a Coulomb or gravitational potential. The central difficulty in these works is the singularity of ψ. For our purposes, this issue is not important, and we will simply assume that ψ is bounded, thus eliminating difficulties in proving existence and uniqueness.

Although we prove a rigorous result, theorem 2.5, the main statement of this paper is not theorem 2.5; the main statement is the general structure that theorem 2.5 strongly suggests, which extends much further than the example above, and which connects to the GENERIC structure that we describe below. Because of this suggestion of a general structure, we now describe a generalized version of the Vlasov–Fokker–Planck equation in somewhat more abstract terms.

Let $\mathcal{H}, \mathcal{S} : \mathcal{P}(\boldsymbol{R}^{2d}) \to \boldsymbol{R}$ be two functionals on $\mathcal{P}(\boldsymbol{R}^{2d})$. Denote by grad $\mathcal{H}$ and grad $\mathcal{S}$ the L^2-gradient of $\mathcal{H}$ and $\mathcal{S}$, otherwise known as the variational derivative.

The following equation we call a *generalized VFP equation*,

$$\partial_t \rho = \operatorname{div}(\rho \boldsymbol{J} \nabla \operatorname{grad} \mathcal{H}) + \operatorname{div}\big(\rho \sigma \sigma^{\mathrm{T}} \nabla \operatorname{grad} (\mathcal{H} + \mathcal{S})\big), \qquad (3)$$

where ∇ and div are the gradient and divergence operators with respect to the full spatial variable $x = (q, p) \in \boldsymbol{R}^{2d}$, and $\boldsymbol{J}$ is the $2d \times 2d$ skew symmetric block matrix

$$\boldsymbol{J} = \begin{pmatrix} 0 & -I_d \\ I_d & 0 \end{pmatrix}, \qquad (4)$$

where I_d is the $\boldsymbol{R}^{d \times d}$-identity matrix.

The VFP equation (1) is an example of this abstract equation, in which

$$x = (q, p)^{\mathrm{T}} \in \boldsymbol{R}^{2d}, \qquad \mathcal{H}(\rho) = \int_{\boldsymbol{R}^{2d}} \left(\frac{p^2}{2m} + V(q) + \frac{1}{2}(\psi * \rho)(q) \right) \rho \, \mathrm{d}q \, \mathrm{d}p, \qquad (5a)$$

$$\sigma = \sqrt{\gamma} \begin{pmatrix} 0 & 0 \\ 0 & I_d \end{pmatrix}, \qquad \mathcal{S}(\rho) = \theta \int_{\boldsymbol{R}^{2d}} \rho \log \rho \, \mathrm{d}q \, \mathrm{d}p. \qquad (5b)$$

Other well-known equations are of the same form; the *Kramers equation* [44] is equation (1) with $\psi \equiv 0$, and Wasserstein gradient flows [16] are of the form (3) with $\sigma = I_{2d}$ and $\mathcal{H} = 0$. As a final example, when

$$\sigma = I_{2d}, \qquad \mathcal{E}(\rho) = \frac{1}{2} \int_{\boldsymbol{R}^{2d}} \rho(\psi * \rho) \, \mathrm{d}x, \qquad \mathcal{S}(\rho) = \theta \int_{\boldsymbol{R}^{2d}} \rho \log \rho \, \mathrm{d}x,$$

---

[4] This is only one of several ways of representing the macroscopic consequences of many microscopic collisions. Other possibilities include stochastic momentum reversals (e.g., [1]), momentum mixing (e.g., [3]), or generalized Langevin equations involving history kernels [2].

equation (3) becomes

$$\partial_t \rho = \mathrm{div}(\rho \boldsymbol{J} \nabla \psi * \rho) + \theta \Delta \rho + \mathrm{div}(\rho \nabla \psi * \rho).$$

This equation describes the relaxation of a point vortex towards statistical equilibrium, that arises in the kinetic theory of point vortices. It is closely related to the two-dimensional Navier–Stokes equation [21–24].

1.3. GENERIC

We now switch gears and introduce an abstract equation structure. Later we will connect the example above with this structure. A GENERIC equation (general equation for non-equilibrium reversible-irreversible coupling [34]) for an unknown z in a state space Z is a mixture of both reversible and dissipative dynamics:

$$\partial_t z = \mathsf{L}\,\mathrm{d}\mathsf{E} + \mathsf{M}\,\mathrm{d}\mathsf{S}. \tag{6}$$

Here

- $\mathsf{E}, \mathsf{S} \colon Z \to \boldsymbol{R}$ are interpreted as energy and entropy functionals,
- $\mathrm{d}\mathsf{E}, \mathrm{d}\mathsf{S}$ are appropriate derivatives of E and S (such as either the Fréchet derivative or a gradient with respect to some inner product);
- $\mathsf{L} = \mathsf{L}(z)$ is for each z an antisymmetric operator satisfying the Jacobi identity

$$\{\{\mathsf{F}_1, \mathsf{F}_2\}_\mathsf{L}, \mathsf{F}_3\}_\mathsf{L} + \{\{\mathsf{F}_2, \mathsf{F}_3\}_\mathsf{L}, \mathsf{F}_1\}_\mathsf{L} + \{\{\mathsf{F}_3, \mathsf{F}_1\}_\mathsf{L}, \mathsf{F}_2\}_\mathsf{L} = 0, \tag{7}$$

  for all functions $\mathsf{F}_i \colon Z \to \boldsymbol{R}$, $i = 1, 2, 3$, where the Poisson bracket $\{\cdot, \cdot\}_\mathsf{L}$ is defined via

$$\{\mathsf{F}, \mathsf{G}\}_\mathsf{L} := \mathrm{d}\mathsf{F} \cdot \mathsf{L}\,\mathrm{d}\mathsf{G} \tag{8}$$

  (see remark 1.1 for a discussion of the meaning of the 'dot' here).
- $\mathsf{M} = \mathsf{M}(z)$ is symmetric and positive semidefinite.

Moreover, the building blocks $\{\mathsf{L}, \mathsf{M}, \mathsf{E}, \mathsf{S}\}$ are required to fulfil the *degeneracy conditions*: for all $z \in Z$,

$$\mathsf{L}\,\mathrm{d}\mathsf{S} = 0, \qquad \mathsf{M}\,\mathrm{d}\mathsf{E} = 0. \tag{9}$$

As a consequence of these properties, energy is conserved along a solution, and entropy is non-decreasing:

$$\frac{\mathrm{d}\mathsf{E}(z(t))}{\mathrm{d}t} = \mathrm{d}\mathsf{E} \cdot \frac{\mathrm{d}z}{\mathrm{d}t} = \mathrm{d}\mathsf{E} \cdot (\mathsf{L}\,\mathrm{d}\mathsf{E} + \mathsf{M}\,\mathrm{d}\mathsf{S}) = 0,$$

$$\frac{\mathrm{d}\mathsf{S}(z(t))}{\mathrm{d}t} = \mathrm{d}\mathsf{S} \cdot \frac{\mathrm{d}z}{\mathrm{d}t} = \mathrm{d}\mathsf{S} \cdot (\mathsf{L}\,\mathrm{d}\mathsf{E} + \mathsf{M}\,\mathrm{d}\mathsf{S}) = \mathrm{d}\mathsf{S} \cdot \mathsf{M}\,\mathrm{d}\mathsf{S} \geqslant 0.$$

A GENERIC system is then fully characterized by $\{Z, \mathsf{E}, \mathsf{S}, \mathsf{L}, \mathsf{M}\}$.

Remark 1.1. In equation (6) we implicitly have assumed that Z is a space with a differentiable structure, in which time derivatives $\partial_t z$ and state-space derivatives $\mathrm{d}\mathsf{S}$ and $\mathrm{d}\mathsf{E}$ exist. In many cases of importance, including the main example of this paper, this is not true, and then generalizations are necessary; the book by Ambrosio, Gigli and Savaré [43] is an example of such generalizations in the case of gradient flows. Nonetheless, we feel that the formal differentiable way of writing provides the right intuition, and therefore in this formal part of the paper we maintain this way of writing the system.

Even in the smooth setting, we have not made specific exactly which derivative $\mathrm{d}\mathsf{E}$ and $\mathrm{d}\mathsf{S}$ should be, and let us briefly make the situation concrete. Derivatives of the functionals E and S are naturally defined as covectors, i.e. elements of the cotangent space (they are then

called differentials) or dual space (called Fréchet derivatives). Since $\partial_t \mathsf{z}$ is an element of the tangent or primal space, L and M should be duality maps, mapping cotangent to tangent spaces, or equivalently dual to primal spaces. In this case the meaning of the dot in (8) is that of the duality pairing.

In practice, however, it often is more convenient to use gradients rather than differentials: then the covectorial derivative is mapped to a tangent vector by some fixed duality mapping, associated with an inner product, often only formally. In all of the explicit calculations in this paper this will be the case; for instance, we already used the $L^2(\mathbf{R}^{2d})$ structure as a formal inner product on the space of measures $\mathcal{P}(\mathbf{R}^{2d})$ to define ' grad $\mathcal{H}$' in equation (3). In this situation L and M map vectors to vectors, and the dot in (8) is that of the formal inner product. $\qquad\square$

1.4. Overview

As described in the introduction, the aim of this paper is twofold: to connect the GENERIC structure with large deviations of stochastic processes, and to construct a useful variational formulation for an abstract GENERIC equation.

In this context, the role of the VFP equation (1) is that of a guiding example. In brief, the story runs as follows: with some modification, the VFP equation can be written as a GENERIC system. In addition, the VFP equation has a particle background, and a recent large-deviation result allows us to connect the large deviations of the particle system with the GENERIC structure. Finally, this same connection shows how the VFP equation can be given a variational formulation.

The first part of this story is told in section 2, in which we construct a large-deviation principle for the SDE (2) associated with the VFP equation. Next, in section 3 we construct a GENERIC structure for the VFP equation and reformulate the large-deviation rate function in this context. Finally, in section 4 we deduce from the large-deviation result a variational formulation for the VFP equation and more generally for any GENERIC system.

Having connected the GENERIC structure with particle systems and large deviations, in section 6 we use this connection to understand the origin and interpretation of the various properties of GENERIC listed in section 1.3. Section 7 is devoted to the generalization (3).

2. Main results 1: large deviations for the VFP equation

For many gradient-flow systems it is now understood that the gradient-flow structure itself arises from the fluctuation behaviour of an underlying stochastic process [15, 18, 36, 39, 40, 48, 49]. The theory of large deviations allows one to make this statement precise. We now apply the same ideas to the VFP equation.

We first specify our conditions on the functions ψ and V. Since we are interested in presenting ideas rather than obtaining the most general results, we choose fairly restrictive conditions on V and ψ to eliminate technical complications:

$$V \in C^2(\mathbf{R}^d) \text{ with globally bounded second derivatives, and } V \geqslant 0; \tag{10a}$$

$$\psi \in C^2(\mathbf{R}^d) \text{ with globally bounded first and second derivatives, and } \psi \geqslant 0. \tag{10b}$$

In addition, we assume that the initial datum ρ^0 satisfies

$$\rho^0 \in \mathcal{P}(\mathbf{R}^{2d}) \qquad \text{with } \mathcal{H}(\rho^0) < \infty, \tag{10c}$$

where $\mathcal{H}$ is defined in (5a). With these assumptions,

- Given a deterministic starting position, the stochastic differential equation (2) has strong solutions that are weakly unique (see e.g. [9, chapter 3]) and non-explosive (e.g. [6]);
- The VFP equation (1) is well defined in the distributional sense and has a unique distributional solution with initial datum ρ^0 that has finite second moment [4].

Given a realization $\{(Q_i, P_i)_{i=1}^n\}$ of the particle system (2), we define the *empirical measure*

$$\rho_n: [0, \infty) \to \mathcal{P}(\boldsymbol{R}^{2d}), \qquad \rho_n(t) := \frac{1}{n} \sum_{i=1}^n \delta_{(Q_i, P_i)(t)}.$$

Theorem 2.5 states that the random variable ρ_n satisfies a *large-deviation principle* as $n \to \infty$.

Definition 2.1 (A large-deviation principle [19, 20, 35]). Let $\mathcal{M}$ be a complete separable metric space and $\{\mu_n\}$ be a sequence of probability measures on $\mathcal{M}$. We say that $\{\mu_n\}$ satisfy a large-deviation principle with a rate functional $I: \mathcal{M} \to [0, \infty)$ if

(i) For each open set $A \subset \mathcal{M}$, $\liminf_{n\to\infty} \frac{1}{n} \log \mu_n(A) \geqslant -\inf_{x\in A} I(x)$;
(ii) For each closed set $B \subset \mathcal{M}$, $\limsup_{n\to\infty} \frac{1}{n} \log \mu_n(B) \leqslant -\inf_{x\in B} I(x)$.

The rate functional I is said to be *good* if its sub-level sets $\{x \in \mathcal{M} \,|\, I(x) \leqslant a\}$ are compact for all $a \geqslant 0$.

Morally, this definition describes the property that

$$\mu_n(A) \sim \exp\left(-n \inf_A I\right) \qquad \text{as } n \to \infty.$$

We refer to [19, 20, 35] for more information on large-deviation theory.

For the theorem below we equip $\mathcal{P}(\boldsymbol{R}^{2d})$ with the weak or narrow topology, generated by the duality with $C_b(\boldsymbol{R}^{2d})$, so that the space $C([0, T]; \mathcal{P}(\boldsymbol{R}^{2d}))$ consists of narrowly continuous curves in $\mathcal{P}(\boldsymbol{R}^{2d})$.

Define for $\nu \in \mathcal{P}(\boldsymbol{R}^{2d})$ the parametrized generator

$$A_\nu \; : D(A_\nu) \subset C_b(\boldsymbol{R}^{2d}) \to C_b(\boldsymbol{R}^{2d}),$$
$$A_\nu f := \frac{p}{m} \cdot \nabla_q f - \left[\nabla_q V + \nabla_q \psi * \nu + \gamma \frac{p}{m}\right] \cdot \nabla_p f + \gamma\theta \Delta_p f.$$

Note that equation (1) can be written in terms of the transpose A^τ as

$$\partial_t \rho_t = A^\tau_{\rho_t} \rho_t.$$

For the formulation of the rate function we will also need the concept of absolute continuity in distributional sense. For a compact set $K \subset \boldsymbol{R}^{2d}$, the space $\mathcal{D}_K$ is the set of all $f \in C_c^\infty(\boldsymbol{R}^{2d})$ with supp $f \subset K$; the set $\mathcal{D}$ is the union of all $\mathcal{D}_K$, with the usual test-function topology.

Definition 2.2. A curve $[0, T] \ni t \mapsto \rho_t \in \mathcal{P}(\boldsymbol{R}^{2d})$ is called *absolutely continuous in distributional sense* if it has the following property: for each compact $K \subset \boldsymbol{R}^{2d}$ there exists a neighbourhood U_K of 0 in $\mathcal{D}_K$ and an absolutely continuous function $G_K : [0, T] \to \boldsymbol{R}$ such that

$$\forall 0 \leqslant t_1 \leqslant t_2 \leqslant T, \; \forall f \in U_K : \qquad |\langle \rho_{t_1}, f\rangle - \langle \rho_{t_2}, f\rangle| \leqslant |G_K(t_1) - G_K(t_2)|.$$

The set of all such curves is denoted $AC([0, T]; \mathcal{P}(\boldsymbol{R}^{2d}))$.

If ρ is absolutely continuous, then for almost all $t \in [0, T]$ the time derivative $\partial_t \rho_t$ exists in $\mathcal{D}'(\boldsymbol{R}^{2d})$. The proof of this and other properties of this concept can be found in [17, section 4].

Finally, we define the norm that will measure the magnitude of fluctuations:

Definition 2.3. Fix $\rho \in \mathcal{P}(\mathbf{R}^{2d})$. For any distribution $\mathcal{T} \in \mathcal{D}'(\mathbf{R}^{2d})$ define

$$\|\mathcal{T}\|_{-1,\rho}^2 := \sup_{f \in C_c^\infty(\mathbf{R}^{2d})} 2\langle \mathcal{T}, f \rangle - \int_{\mathbf{R}^{2d}} |\nabla_p f|^2 \, \mathrm{d}\rho. \tag{11}$$

Define $L_\nabla^2(\rho)$ as the completion of $\{\nabla_p f : f \in C_c^\infty(\mathbf{R}^{2d})\}$ with respect to the norm

$$\| \cdot \|_\rho^2 := \int_{\mathbf{R}^{2d}} |\cdot|^2 \, \mathrm{d}\rho.$$

Note that, depending on ρ, $\| \cdot \|_\rho$ may be only a seminorm and not a norm; but since the completion identifies elements that have zero distance in this seminorm, $L_\nabla^2(\rho)$ is a well-defined Hilbert space. Its elements are equivalence classes of measurable functions that are ρ-a.e. equal. Also note that whenever $\mathcal{H}(\rho) < \infty$, the function $(q, p) \mapsto p$ belongs to $L_\nabla^2(\rho)$.

The dual norm $\| \cdot \|_{-1,\rho}$ has an explicit representation.

Lemma 2.4. *It holds that*

$$\|\mathcal{T}\|_{-1,\rho}^2 = \begin{cases} \displaystyle\int_{\mathbf{R}^{2d}} |h|^2 \, \mathrm{d}\rho & \text{if } \mathcal{T} = -\mathrm{div}_p(\rho h) \text{ with } h \in L_\nabla^2(\rho), \\ +\infty & \text{otherwise.} \end{cases}$$

Proof. Results of this type are common; this argument is adapted from [17].

Since there is a one-to-one correspondence between $f \in C_c^\infty(\mathbf{R}^{2d})$ and $\nabla_p f \in L :=$ $\{\nabla_p f : f \in C_c^\infty(\mathbf{R}^{2d})\}$, $\mathcal{T}$ can be considered to be a linear functional on L. If $\|\mathcal{T}\|_{-1,\rho} < \infty$, we can replace f by λf and optimize with respect to $\lambda \in \mathbf{R}$ in (11). We then find that

$$|\langle \mathcal{T}, f \rangle| \leqslant \|\mathcal{T}\|_{-1,\rho} \|\nabla_p f\|_\rho.$$

Therefore $\mathcal{T}$ is bounded with respect to the $L_\nabla^2(\rho)$-norm; it can be uniquely extended to a bounded linear functional on the whole of $L_\nabla^2(\rho)$, and Riesz' representation theorem implies the assertion of the lemma. $\square$

We can now state the large-deviation principle.

Theorem 2.5. *Assume that the initial data $(Q_i(0), P_i(0))$, $i = 1, \ldots, n$ are deterministic and chosen such that $\rho_n(0) \rightharpoonup \rho^0$ for some $\rho^0 \in \mathcal{P}(\mathbf{R}^{2d})$. Then the empirical process $\{\rho_n\}$ satisfies a large-deviation principle in the space $C([0, T], \mathcal{P}(\mathbf{R}^{2d}))$, with good rate function*

$$I(\rho) = \begin{cases} \dfrac{1}{4\gamma\theta} \displaystyle\int_0^T \left\| \partial_t \rho_t - A_{\rho_t}^\tau \rho_t \right\|_{-1,\rho_t}^2 \, \mathrm{d}t & \text{if } \rho \in AC([0, T]; \mathcal{P}(\mathbf{R}^{2d})) \text{ and } \rho|_{t=0} = \rho^0, \\ +\infty & \text{otherwise.} \end{cases} \tag{12}$$

The rate function I can also be written as

$$I(\rho) = \begin{cases} \dfrac{1}{4\gamma\theta} \displaystyle\int_0^T \int_{\mathbf{R}^{2d}} |h_t|^2 \, \mathrm{d}\rho_t \, \mathrm{d}t & \text{if } \partial_t \rho_t = A_{\rho_t}^\tau \rho_t - \mathrm{div}_p(\rho_t h_t), \\ & \qquad \text{for } h \in L^2(0, T; L_\nabla^2(\rho_t)) \\ & \qquad \text{and } \rho|_{t=0} = \rho^0, \\ +\infty & \text{otherwise.} \end{cases} \tag{13}$$

Proof. We set $x = (q, p)$ and $b(x, \nu) = \big(p/m, -\nabla V(q) - (\nabla\psi * \nu)(q) - \gamma p/m\big)$ for $\nu \in \mathcal{P}(\mathbf{R}^{2d})$. Then $b \colon \mathbf{R}^{2d} \times \mathcal{P}(\mathbf{R}^{2d}) \to \mathbf{R}^{2d}$ is continuous and, by the assumptions (10), satisfies the estimate

$$|b(x, \nu) \cdot x| \leqslant C(1 + |x|^2) \qquad \text{for all } x \in \mathbf{R}^{2d} \quad \text{and } \nu \in \mathcal{P}(\mathbf{R}^{2d}).$$

The system (2) can be written as system of weakly interacting diffusions

$$dX_i(t) = b(X_i(t), \rho_n(t)) \, dt + \sigma \, dW_i(t), \tag{14}$$

where W_i are d-dimensional standard Wiener processes and for the length of this proof, σ is the $2d \times d$ matrix

$$\sigma = \sqrt{2\gamma\theta} \begin{pmatrix} 0 \\ I_d \end{pmatrix}.$$

Theorem 3.1 and remark 3.2 of [33] implies that ρ_n satisfies a large-deviation principle with rate function

$$\widetilde{I}(\rho) := \inf \mathbb{E}\left[\frac{1}{2}\int_0^T |U_t|^2 \, dt\right],$$

where the infimum is taken over all processes $(\overline{X}, U, W)$ taking values in $\mathbf{R}^{2d} \times \mathbf{R}^d \times \mathbf{R}^d$ that solve

$$d\overline{X}_t = b(\overline{X}_t, \rho_t) \, dt + \sigma U_t \, dt + \sigma \, dW_t, \tag{15a}$$

$$W \text{ is a standard } d\text{-dimensional Wiener : process}, \tag{15b}$$

$$\text{law } \overline{X}_t = \rho_t \quad \text{for all } t. \tag{15c}$$

For each such triple, for any $f \in C_c^\infty(\mathbf{R} \times \mathbf{R}^{2d})$ the process

$$M_t := f_t(\overline{X}_t) - f_0(\overline{X}_0) - \int_0^t \left[(\partial_s + A_{\rho_s} + (\sigma U_s) \cdot \nabla) f_s\right](\overline{X}_s) \, ds$$

is a martingale, and therefore $\mathbb{E}M_t = \mathbb{E}M_0 = 0$ for every $t > 0$.

We now show (12) by showing that $\widetilde{I} = I$. Define for any $\rho \in C([0,T]; \mathcal{P}(\mathbf{R}^{2d}))$ and $f \in C_c^\infty(\mathbf{R} \times \mathbf{R}^{2d})$,

$$J(\rho, f) := \int_{\mathbf{R}^{2d}} f_T \, d\rho_T - \int_{\mathbf{R}^{2d}} f_0 \, d\rho_0 - \int_0^T \int_{\mathbf{R}^{2d}} \left[(\partial_s + A_{\rho_s}) f_s\right] d\rho_s \, ds$$

$$- \gamma\theta \int_0^T \int_{\mathbf{R}^{2d}} |\nabla_p f_t|^2 \, d\rho_t \, dt.$$

It is well known (see e.g. [17, lemma 4.8]) that

$$I(\rho) = \sup_{f \in C_c^\infty(\mathbf{R} \times \mathbf{R}^{2d})} J(\rho, f).$$

We have for any $f \in C_c^\infty(\mathbf{R} \times \mathbf{R}^{2d})$ and for any solution $(\overline{X}, U, W)$ of (15),

$$\mathbb{E}\left[\frac{1}{2}\int_0^T |U_t|^2 \, dt\right] = \mathbb{E}\left[\int_0^T \left(U_t \nabla_p f_t(\overline{X}_t) - \frac{1}{2}|\nabla_p f_t(\overline{X}_t)|^2\right) dt\right]$$

$$+ \mathbb{E}\left[\frac{1}{2}\int_0^T |U_t - \nabla_p f_t(\overline{X}_t)|^2 \, dt\right].$$

Using $\mathbb{E}M_T = 0$ we rewrite this as

$$\mathbb{E}\left[f_T(\overline{X}_T) - f_0(\overline{X}_0) - \int_0^T \left[(\partial_s + A_{\rho_s}) f_s\right](\overline{X}_s) \, ds - \frac{1}{2}\int_0^T |\nabla_p f_s(\overline{X}_s)|^2 \, ds\right]$$

$$+ \mathbb{E}\left[\frac{1}{2}\int_0^T |U_t - \nabla_p f_t(\overline{X}_t)|^2 \, dt\right]$$

$$= J\left(\rho, \frac{f}{\sqrt{2\gamma\theta}}\right) + \mathbb{E}\left[\frac{1}{2}\int_0^T |U_t - \nabla_p f_t(\overline{X}_t)|^2 \, dt\right]. \tag{16}$$

Therefore

$$\widetilde{I}(\rho) = \inf \mathbb{E}\left[\frac{1}{2}\int_0^T |U_t|^2 \, dt\right] \geqslant \sup_f J(\rho, f) = I(\rho).$$

To prove the converse inequality, assume without loss of generality that $I(\rho) < \infty$. Using a reasoning similar to the proof of lemma 2.4 we find that there exists an $h \in L^2(0, T; L^2_\nabla(\rho_t))$ such that

$$\partial_t \rho_t - A^\tau_{\rho_t} \rho_t = -\sqrt{2\gamma\theta}\,\mathrm{div}_p\,\rho_t h_t \qquad \text{in the sense of distributions.} \qquad (17)$$

Here the space $L^2(0, T; L^2_\nabla(\rho_t))$ is the Hilbert space obtained by closing $C_c^\infty(\mathbf{R} \times \mathbf{R}^{2d})$ with respect to the (semi-)norm

$$\|f\|^2_{\rho, T} := \int_0^T \int_{\mathbf{R}^{2d}} |f(x, t)|^2 \, \rho_t(x) \, dt. \qquad (18)$$

We now construct a specific solution of (15). Let $(\widetilde{X}, W)$ be a solution of (15a) with $U = 0$ and law $\widetilde{X}_0 = \rho^0$; let P be the law of $(\widetilde{X}, W)$ on $C([0, T]; \mathbf{R}^{2d}) \times C([0, T]; \mathbf{R}^d)$. Since $\|h\|_{\rho, T} < \infty$, the process

$$N_t := \sigma \int_0^t h_s(\widetilde{X}_s) \, dW_s$$

is a P-square integrable continuous martingale with quadratic variation $\langle N \rangle_t = 2\gamma\theta t$.

Define P_h as the modified law on $C([0, T]; \mathbf{R}^{2d}) \times C([0, T]; \mathbf{R}^d)$ given by

$$P_h := \exp\left[N_T - \frac{1}{2}\langle N \rangle_T\right] P.$$

By the Girsanov theorem (e.g. [7, section IV.4]) P_h is the law of the unique solution (X, W) of equation (15a) with $U_t = h_t(X_t)$, and since equation (17) is the corresponding Fokker–Planck equation, it follows that the law of X_t is equal to ρ_t. Therefore $(X, h \circ X, W)$ is a solution of (15). Using (16) for this solution, we find for all f that

$$\widetilde{I}(\rho) \leqslant J\left(\rho, \frac{f}{2\gamma\theta}\right) + \frac{1}{4\gamma\theta}\mathbb{E}\left[\int_0^T |h_t(X_t) - \nabla_p f_t(X_t)|^2 \, dt\right]$$

$$\leqslant I(\rho) + \frac{1}{4\gamma\theta}\mathbb{E}\left[\int_0^T |h_t(X_t) - \nabla_p f_t(X_t)|^2 \, dt\right]$$

$$= I(\rho) + \frac{1}{4\gamma\theta}\int_0^T \int_{\mathbf{R}^{2d}} |h_t(\xi) - \nabla_p f_t(\xi)|^2 \, \rho_t(d\xi) \, dt.$$

Since $L^2(0, T; L^2_\nabla(\rho_t))$ is the closure of C_c^∞ under the norm (18),

$$\inf_{f \in C_c^\infty(\mathbf{R} \times \mathbf{R}^{2d})} \int_0^T \int_{\mathbf{R}^{2d}} |h_t(\xi) - \nabla_p f_t(\xi)|^2 \, \rho_t(d\xi) \, dt = 0.$$

Hence $\widetilde{I}(\rho) \leqslant I(\rho)$ and this concludes the proof of (12). The form as in (13) of I then follows from (12) and lemma 2.4. □

Remark 2.6. The structure of the large-deviation result of theorem 2.5 reflects a number of properties of the stochastic particle system (2). To start with, the rate function is only finite if $\partial_t \rho - A^\tau_\rho \rho$ only has a perturbation in the p-direction, not in the q-direction; this reflects the fact in (2) that the noise is confined to the P-equation. In addition, the perturbation can only be in divergence form; this reflects the deterministic conservation of particles. Finally, the flux is of the form ρh where h is in the closure $L^2_\nabla(\rho)$ of p-gradients; this property is also seen in the characterization of absolutely continuous curves in the Wasserstein metric [43, theorem 8.3.2]. □

Remark 2.7. There is a large literature on large-deviation principles for stochastic particle systems; here we just mention a few results. Dawson and Gärtner [17] prove a large-deviations result for systems of interacting particles with non-degenerate diffusion, i.e. for non-singular mobilities σ with range $\boldsymbol{R}^{2d}$. Cattiaux and Léonard [45, 46] generalize the method of Dawson and Gärtner to singular mobilities, but for independent particles. In a separate paper [8], Cattiaux and Léonard also discuss the identification question treated in the proof of theorem 2.5 in more generality. Fischer [25] also proves identification results on related systems.

In the proof above we used the large-deviation result by Budhiraja *et al* [33] above to obtain the large-deviation principle itself and a first characterization of the rate functional. The methods by which we identified $\widetilde{I}$ with I are standard, but we did not find a theorem that suited our needs, and therefore we gave a separate proof. $\square$

Remark 2.8. If the initial datum for the particle system is not deterministic, as in the case of theorem 2.5, then we expect that the sequence $\{\rho_n\}$ then satisfies a large-deviation principle with rate function $I(\rho) + I_0(\rho|_{t=0})$, where I_0 is the rate function of the initial data $\rho_n|_{t=0}$. $\square$

For the sequel it will be useful to have a regularity result on the Hamiltonian $\mathcal{H}$ (see (5a)) associated with those curves ρ for which $I(\rho)$ is finite:

Lemma 2.9. *If $I(\rho) < \infty$ and $\mathcal{H}(\rho_0) < \infty$, then the function $t \mapsto \mathcal{H}(\rho_t)$ is an element of $W^{1,2}(0, T)$, and $\int_{\boldsymbol{R}^{2d}} p^2 \, d\rho_t \in L^\infty(0, T)$.*

Proof. By (10), $\mathcal{H}(\rho)$ bounds the integral $\int p^2/m^2 \, d\rho$ from above. Using the characterization of I in (13), we formally calculate that

$$\partial_t \mathcal{H}(\rho_t) = \frac{\gamma \theta \mathrm{d}}{m} - \gamma \int \frac{p^2}{m^2} \, \mathrm{d}\rho_t + \int \frac{p}{m} \cdot h_t \, \mathrm{d}\rho_t \tag{19}$$

$$\leqslant \frac{\gamma \theta \mathrm{d}}{m} - \gamma \int \frac{p^2}{m^2} \, \mathrm{d}\rho_t + \gamma \int \frac{p^2}{m^2} \, \mathrm{d}\rho_t + \frac{1}{4\gamma} \int |h_t|^2 \, \mathrm{d}\rho_t$$

$$= \frac{\gamma \theta \mathrm{d}}{m} + \frac{1}{4\gamma} \int |h_t|^2 \, \mathrm{d}\rho_t.$$

This calculation can be made rigorous in its time-integrated form by approximating $p^2/m^2 + V(q)$ by a sequence of smooth functions $f_n \in C_c^\infty(\boldsymbol{R}^{2d})$, and using f_n in the distributional form of the equation $\partial_t \rho_t = A_{\rho_t}^{\mathrm{T}} \rho_t - \mathrm{div}_p(\rho_t h_t)$. Continuing with the proof, it follows that

$$\sup_{t \in [0,T]} \mathcal{H}(\rho_t) \leqslant \mathcal{H}(\rho_0) + \frac{\gamma \theta \mathrm{d}}{m} T + \frac{1}{4\gamma} \int_0^T \int |h_t|^2 \, \mathrm{d}\rho_t = \mathcal{H}(\rho_0) + \frac{\gamma \theta \mathrm{d}}{m} T + \theta \, I(\rho) < \infty,$$

and consequently $\int p^2 \, d\rho_t$ is also uniformly bounded. We conclude by remarking that the right-hand side of (19), as a function of time t, is an element of $L^2(0, T)$. $\square$

Remark 2.10. Note that a solution ρ of (1) satisfies $I(\rho) = 0$, and therefore lemma 2.9 also applies to solutions of (1). $\square$

3. Main results 2: The VFP equation and the large deviations in GENERIC form

In this section we reformulate both the VFP equation and the large-deviation rate functional of the previous section in terms of the GENERIC structure. It will become apparent that the large-deviation behaviour respects the GENERIC structure, in the sense that the rate function for this system can be formulated in an abstract form, using only the GENERIC building blocks. This will suggest in section 4 a variational formulation for a very general GENERIC system.

3.1. Making the VFP equation conserve energy

As it stands, the VFP equation (1) does not satisfy the conditions of GENERIC, since there is no conserved functional E. The reason for this is physical: the SDE (2) models a system of particles in interaction with a heat bath, and this interaction causes fluctuations of the natural energy (the Hamiltonian) of the particle system,

$$H_n(Q_1, \ldots, Q_n, P_1, \ldots, P_n) := \frac{1}{n} \sum_{i=1}^{n} \left[\frac{P_i^2}{2m} + V(Q_i) \right] + \frac{1}{2n^2} \sum_{i,j=1}^{n} \psi(Q_i - Q_j). \tag{20}$$

Indeed, combining (2) with Itô's lemma the derivative of the expression above is

$$-\frac{1}{n} \sum_{i=1}^{n} \left[\frac{\gamma}{m^2} P_i^2 \, dt - \frac{\gamma\theta d}{m} \, dt + \frac{\sqrt{2\gamma\theta}}{m} P_i \, dW_i \right],$$

which has no reason to vanish. There is a simple remedy for this: we add a single scalar unknown e_n and define its evolution by the negative of the above, leading to the extended particle system

$$dQ_i = \frac{P_i}{m} \, dt, \tag{21a}$$

$$dP_i = -\nabla V(Q_i) \, dt - \sum_{j=1}^{n} \nabla \psi(Q_i - Q_j) - \frac{\gamma}{m} P_i \, dt + \sqrt{2\gamma\theta} \, dW_i, \tag{21b}$$

$$de_n = \frac{1}{n} \sum_{i=1}^{n} \left[\frac{\gamma}{m^2} P_i^2 \, dt - \frac{\gamma\theta d}{m} \, dt + \frac{\sqrt{2\gamma\theta}}{m} P_i \, dW_i \right], \tag{21c}$$

with which $H_n + e_n$ becomes deterministically constant. Note that e_n can be interpreted as the energy of the heat bath; the flow of energy between the particle system and the heat bath is described by the flow of energy between H_n and e_n.

Exactly the same arguments apply to the VFP equation (1). At this level the analogue of the Hamiltonian H_n is the functional $\mathcal{H}$ defined in (5a), and indeed $\mathcal{H}$ is not constant along a solution, as can be directly verified. We mirror the arguments above and add a new variable e, *depending only on time*, so that the solution space becomes $(\rho, e) \in \mathcal{P}(\mathbf{R}^{2d}) \times \mathbf{R}$. The full system is now defined by the VFP equation (1) plus the equation $de/dt = -(d/dt)\mathcal{H}(\rho)$, that guarantees that $\mathcal{H}(\rho) + e$ is conserved. When writing this equation in full, it becomes

$$\partial_t \rho = -\text{div}_q \left(\rho \frac{p}{m} \right) + \text{div}_p \rho \left(\nabla_q V + \nabla_q \psi * \rho + \gamma \frac{p}{m} \right) + \gamma\theta\Delta_p \rho, \tag{22a}$$

$$\frac{d}{dt} e = \gamma \int_{\mathbf{R}^{2d}} \frac{p^2}{m^2} \rho(dq \, dp) - \frac{\gamma\theta d}{m}. \tag{22b}$$

We stress that this system is coupled only in one direction: the second equation is slaved to the first one. Note that equation (22b) is well defined: if $\mathcal{H}(\rho_0) < \infty$, then by lemma 2.9 and remark 2.10 $\mathcal{H}(\rho_t)$ is bounded for all t; therefore $\int p^2 d\rho_t$ is finite for all t.

By this simple mechanism a non-conserving system can be made conserving. Although mathematically this is no more than a trick, for this system it has physical meaning, as we argued above: the additional variable keeps track of the movement of energy between the particle system and the heat bath. We next show that the remaining conditions of GENERIC can also be verified.

3.2. The VFP equation as a GENERIC system

With the extension of the previous section, the VFP equation is formally a GENERIC system with the following building blocks:

$$\mathsf{Z} = \mathcal{P}_2(\boldsymbol{R}^{2d}) \times \boldsymbol{R}, \qquad \mathsf{E}(\rho, e) = \mathcal{H}(\rho) + e, \qquad \mathsf{L} = \mathsf{L}(\rho, e) = \begin{pmatrix} \mathsf{L}_{\rho\rho} & 0 \\ 0 & 0 \end{pmatrix},$$

$$\mathsf{z} = (\rho, e), \quad \mathsf{S}(\rho, e) = \mathcal{S}(\rho) + e, \qquad \mathsf{M} = \mathsf{M}(\rho, e) = \gamma \begin{pmatrix} \mathsf{M}_{\rho\rho} & \mathsf{M}_{\rho e} \\ \mathsf{M}_{e\rho} & \mathsf{M}_{ee} \end{pmatrix}, \tag{23}$$

where the operators defining L and M are given, upon applying them to a vector (ξ, r) at (ρ, e), by

$$\mathsf{L}_{\rho\rho}\xi = \operatorname{div}\rho\boldsymbol{J}\nabla\xi, \qquad \mathsf{M}_{\rho\rho}\xi = -\operatorname{div}_p\rho\nabla_p\xi, \qquad \mathsf{M}_{\rho e}r = r\operatorname{div}_p\left(\rho\frac{p}{m}\right),$$

$$\mathsf{M}_{e\rho}\xi = -\int_{\boldsymbol{R}^{2d}} \frac{p}{m} \cdot \nabla_p\xi \, \rho(\mathrm{d}q\mathrm{d}p), \qquad \mathsf{M}_{ee}r = r\int_{\boldsymbol{R}^{2d}} \frac{p^2}{m^2} \rho(\mathrm{d}q\mathrm{d}p).$$

The space $\mathcal{P}_2(\boldsymbol{R}^{2d})$ is the subset of $\mathcal{P}(\boldsymbol{R}^{2d})$ with bounded second p-moments:

$$\mathcal{P}_2(\boldsymbol{R}^{2d}) := \left\{ \rho \in \mathcal{P}(\boldsymbol{R}^{2d}) : \int_{\boldsymbol{R}^{2d}} p^2\rho(\mathrm{d}p\mathrm{d}q) < \infty \right\}.$$

We equip $\mathcal{P}_2(\boldsymbol{R}^{2d})$ with the same weak topology as $\mathcal{P}(\boldsymbol{R}^{2d})$. Finally, the entropy $\mathcal{S}$ is defined as

$$\mathcal{S}(\rho) := -\theta \int_{\boldsymbol{R}^{2d}} f(x) \log f(x) \, \mathrm{d}x \qquad \text{whenever } \rho \text{ has Lebesgue density } f.$$

With these definitions, equation (1) can be written as

$$\partial_t \mathsf{z}_t = \mathsf{L}(\mathsf{z}_t) \operatorname{grad} \mathsf{E}(\mathsf{z}_t) + \mathsf{M}(\mathsf{z}_t) \operatorname{grad} \mathsf{S}(\mathsf{z}_t), \tag{24}$$

where the gradient operators are to be interpreted as L^2-gradients. At this stage, however, this equation is formal, since the sense in which this equation holds has not been specified. Rather than going into detail here, we defer this discussion to after the introduction of the variational structure in section 4.

The operators L and M can readily be seen to be antisymmetric and symmetric (with respect to the L^2-inner-product, since we use L^2-gradients as derivatives); for instance, in the case of L, we have for any vectors (ξ_1, r_1) and (ξ_2, r_2) at (ρ, e) by partial integration that

$$\langle (\xi_1, r_1), \mathsf{L}(\rho, e)(\xi_2, r_2) \rangle = \langle \xi_1, \mathsf{L}_{\rho\rho}(\rho)\xi_2 \rangle = \int_{\boldsymbol{R}^{2d}} \xi_1 \operatorname{div}\rho\boldsymbol{J}\nabla\xi_2 = -\int_{\boldsymbol{R}^{2d}} \nabla\xi_2 \cdot \boldsymbol{J}^T\nabla\xi_1 \, \rho,$$

which is antisymmetric since $\boldsymbol{J}$ is antisymmetric (see (4)). The verification of the symmetry of M is similar; the verification of the Jacobi identity (7) is a tedious but elementary calculation, which hinges on the fact that $\boldsymbol{J}$ is constant and antisymmetric. Finally, the verification of the degeneracy conditions (9) is again straightforward.

3.3. Large deviations for the VFP equation in GENERIC form

We now reformulate the large-deviations rate functional of theorem 2.5 in terms of the GENERIC building blocks above, and therefore in terms of the extended unknown $\mathsf{z} = (\rho, e) \in \mathsf{Z}$. To do this, we also generalize the concepts of absolute continuity and introduce the appropriate norms.

Definition 3.1. The function $[0, T] \ni t \mapsto \mathsf{z}(t) = (\rho(t), e(t)) \in \mathsf{Z}$ is *absolutely continuous* if $\rho \in AC([0, T]; \mathcal{P}_2(\boldsymbol{R}^{2d}))$ and $e \in AC([0, T]; \boldsymbol{R})$.

Again, if z is absolutely continous, then $\partial_t z$ exists for almost all t as an element of $\mathcal{D}'(\mathbf{R}^{2d}) \times \mathbf{R}$.

The 'matrix' M generates a natural pair of semi-inner-products and seminorms.

Definition 3.2. Fix $z = (\rho, e) \in \mathsf{Z}$. The seminorms $\| \cdot \|_{\mathsf{M}(z)}$ and $\| \cdot \|_{\mathsf{M}(z)^{-1}}$ are defined as follows. For $(\xi, r) \in C_c^\infty(\mathbf{R}^{2d}) \times \mathbf{R}$,

$$\|(\xi, r)\|_{\mathsf{M}(z)}^2 := \gamma \int_{\mathbf{R}^{2d}} \left[\xi \mathsf{M}_{\rho\rho} \xi + \xi \mathsf{M}_{\rho e} r + r \mathsf{M}_{e\rho} \xi + r \mathsf{M}_{ee} r \right] \mathrm{d}x$$

$$= \gamma \int_{\mathbf{R}^{2d}} \left| \nabla_p \xi - r \frac{p}{m} \right|^2 \mathrm{d}\rho = \gamma \left\| \nabla_p \xi - r \frac{p}{m} \right\|_\rho^2.$$

For $(\mathcal{T}, s) \in \mathcal{D}'(\mathbf{R}^{2d}) \times \mathbf{R}$,

$$\|(\mathcal{T}, s)\|_{\mathsf{M}(z)^{-1}}^2 = \sup_{\xi \in C_c^\infty(\mathbf{R}^{2d}) r \in \mathbf{R}} 2\langle \mathcal{T}, \xi \rangle + 2sr - \|(\xi, r)\|_{\mathsf{M}(z)}^2. \tag{25}$$

The inner products $(\cdot, \cdot)_{\mathsf{M}}$ and $(\cdot, \cdot)_{\mathsf{M}^{-1}}$ are then defined through the expression $4(a, b) = \|a + b\|^2 - \|a - b\|^2$.

As in the case of $L_\nabla^2(\rho)$, the M-seminorm is degenerate: there exist ρ, ξ, and r for which it vanishes. Let $\mathscr{H}_{\mathsf{M}}$ be the set of equivalence classes of elements of $C_c^\infty(\mathbf{R}^{2d}) \times \mathbf{R}$ with zero distance in this norm. On $\mathscr{H}_{\mathsf{M}}$, the M-seminorm is a norm, and we define H_{M} as the completion of $\mathscr{H}_{\mathsf{M}}$ with respect to this norm. Note that H_{M} can be identified with the space $L_\nabla^2(\rho)$, as follows. On one hand, if (η_n, s_n) is a Cauchy sequence in $\mathscr{H}_{\mathsf{M}}$, then

$$\left\| (\eta_n, s_n) - (\eta_{n'}, s_{n'}) \right\|_{\mathsf{M}} = \sqrt{\gamma} \left\| \nabla_p (\eta_n - \eta_{n'}) - (s_n - s_{n'}) \frac{p}{m} \right\|_\rho \longrightarrow 0 \qquad \text{as } n, n' \to \infty,$$

so that $\nabla_p \eta_n - s_n p/m$ is a Cauchy sequence in $L_\nabla^2(\rho)$ and thus converges to some $h \in L_\nabla^2(\rho)$; vice versa, for each $h \in L_\nabla^2(\rho)$ by definition there exists a sequence $\eta_n \in C_c^\infty$ such that $\nabla_p \eta_n \to h$ in $L_\nabla^2(\rho)$, and therefore $(\eta_n, 0)$ is a Cauchy sequence in $\mathscr{H}_{\mathsf{M}}$ corresponding to h.

Since the M-seminorm is degenerate, the M^{-1}-seminorm is singular. Indeed, lemma 2.4 implies the following

Lemma 3.3. Assume that $\int_{\mathbf{R}^{2d}} p^2 \, \mathrm{d}\rho < \infty$. Then

$$\|(\mathcal{T}, s)\|_{\mathsf{M}(z)^{-1}}^2 = \begin{cases} \dfrac{1}{\gamma} \displaystyle\int_{\mathbf{R}^{2d}} |h|^2 \, \mathrm{d}\rho & \text{if } \mathcal{T} = -\mathrm{div}_p \rho h \text{ with } h \in L_\nabla^2(\rho) \\ & \text{and } s = -\int_{\mathbf{R}^{2d}} \frac{p}{m} \cdot h \, \mathrm{d}\rho, \\ +\infty & \text{otherwise.} \end{cases}$$

Proof. As in the case of lemma 2.4, $\|(\mathcal{T}, s)\|_{\mathsf{M}(z)^{-1}} < \infty$ implies that $(\mathcal{T}, s)$ is a linear functional on $C_c^\infty \times \mathbf{R}$, and by the assumption $\int p^2 \, \mathrm{d}\rho < \infty$ it is bounded with respect to the M-seminorm. Because of the identification with $L_\nabla^2(\rho)$ we can consider it as a bounded linear functional on $L_\nabla^2(\rho)$. By the Riesz representation theorem there exists an element $h \in L_\nabla^2(\rho)$ such that for all ξ and r

$$\langle \mathcal{T}, \xi \rangle + rs = \int_{\mathbf{R}^{2d}} h \left(\nabla_p \xi - r \frac{p}{m} \right) \mathrm{d}\rho = \int_{\mathbf{R}^{2d}} h \cdot \nabla_p \xi \, \mathrm{d}\rho - r \int_{\mathbf{R}^{2d}} h \cdot \frac{p}{m} \, \mathrm{d}\rho.$$

From this identity the claim follows. $\qquad\square$

The rate function of theorem 2.5 now has a reformulation in terms of the objects that we have just defined.

Lemma 3.4. *The rate function I of theorem 2.5 can be written in terms of z as*

$$J(z) = \begin{cases} \int_0^T \frac{1}{4\theta} \left\| \partial_t z_t - \mathsf{L}(z_t) \operatorname{grad} \mathsf{E}(z_t) - \mathsf{M}(z_t) \operatorname{grad} \mathsf{S}(z_t) \right\|^2_{\mathsf{M}(z_t)^{-1}} \mathrm{d}t, \\ \qquad \text{if } z = (\rho, e) \in AC([0, T]; \mathsf{Z}) \quad \text{and} \quad \rho_{t=0} = \rho^0, \\ +\infty \qquad \text{otherwise,} \end{cases} \tag{26}$$

in the sense that

$$J\big((\rho, e)\big) = \begin{cases} I(\rho) & \text{provided } t \mapsto \mathcal{H}(\rho_t) + e_t \text{ is constant} \\ +\infty & \text{otherwise.} \end{cases}$$

Proof. First assume that $I(\rho) < \infty$. By (12) and lemma 2.4 we have

$$\partial_t \rho_t - A^\tau_{\rho_t} \rho_t = -\operatorname{div} \rho_t h_t,$$

where $h \in L^2(0, T; L^2_\nabla(\rho_t))$. Define e by

$$e_0 := 0 \qquad \text{and} \qquad \partial_t e_t = \gamma \int_{R^{2d}} \frac{p^2}{m^2} \rho_t(\mathrm{d}q\,\mathrm{d}p) - \frac{\gamma\theta\mathrm{d}}{m} + \int_{R^{2d}} \frac{p}{m} h_t \, \rho_t(\mathrm{d}q\,\mathrm{d}p).$$

By lemma 2.9 the function $t \mapsto \int p^2 \, d\rho_t$ is in $L^\infty(0, T)$, and since $h \in L^2(0, T; L^2_\nabla(\rho_t))$ the last term is in $L^1(0, T)$; therefore e is well defined, and an element of $AC([0, T]; R)$. By construction the function $t \mapsto \mathcal{H}(\rho_t) + e_t$ is constant. Upon setting $z := (\rho, e)$, an explicit calculation shows that $I(\rho)$ and $J(z)$ are both equal to $(4\gamma\theta)^{-1} \int_0^T \int_{R^{2d}} |h_t|^2 \, \mathrm{d}\rho_t \, \mathrm{d}t$.

A similar argument starts by assuming $J(z) < \infty$ for $z = (\rho, e)$ and showing that $I(\rho)$ and $J(z)$ are again equal. □

Remark 3.5. Note how the condition of constant energy $\mathcal{H} + e$ is contained in (26) through the defintion of the seminorm $\| \cdot \|_{\mathsf{M}^{-1}}$. □

4. Main results 3: a variational formulation for GENERIC systems

The functional J in (26) has the interesting property that it only depends on the GENERIC building blocks, and therefore makes sense, at least formally, for an arbitrary GENERIC system. We now explore the consequences of this observation for general GENERIC systems. The discussion in this section is therefore necessarily formal.

First, we note that the functional J can be written in a different way by using one of the degeneracy conditions (9). As above, we associate a formal inner product with M and M^{-1} by

$$(a, b)_\mathsf{M} := a \cdot \mathsf{M} b \qquad \text{and} \qquad (a, b)_{\mathsf{M}^{-1}} := a \cdot \mathsf{M}^{-1} b.$$

(See remark 1.1 for a discussion of the dot in these expressions). Then the antisymmetry of L and the first degeneracy condition in (9) imply that

$$\big(\mathsf{L} \operatorname{grad} \mathsf{E}, \mathsf{M} \operatorname{grad} \mathsf{S}\big)_{\mathsf{M}^{-1}} = \mathsf{L} \operatorname{grad} \mathsf{E} \cdot \operatorname{grad} \mathsf{S} = -\operatorname{grad} \mathsf{E} \cdot \mathsf{L} \operatorname{grad} \mathsf{S} = 0.$$

Therefore

$$\left\| \partial_t z - \mathsf{L} \operatorname{grad} \mathsf{E} - \mathsf{M} \operatorname{grad} \mathsf{S} \right\|^2_{\mathsf{M}^{-1}} = \left\| \partial_t z - \mathsf{L} \operatorname{grad} \mathsf{E} \right\|^2_{\mathsf{M}^{-1}} + \left\| \mathsf{M} \operatorname{grad} \mathsf{S} \right\|^2_{\mathsf{M}^{-1}}$$
$$+ 2\big(\partial_t z, \mathsf{M} \operatorname{grad} \mathsf{S}\big)_{\mathsf{M}^{-1}}$$
$$= \left\| \partial_t z - \mathsf{L} \operatorname{grad} \mathsf{E} \right\|^2_{\mathsf{M}^{-1}} + \left\| \operatorname{grad} \mathsf{S} \right\|^2_{\mathsf{M}} + 2 \, \partial_t z \cdot \operatorname{grad} \mathsf{S},$$

so that

$$2\theta J(z) = \mathsf{S}(z(T)) - \mathsf{S}(z(0)) + \frac{1}{2} \int_0^T \left[\left\| \partial_t z - \mathsf{L} \operatorname{grad} \mathsf{E} \right\|^2_{\mathsf{M}^{-1}} + \left\| \operatorname{grad} \mathsf{S} \right\|^2_{\mathsf{M}} \right] \mathrm{d}t. \tag{27}$$

This discussion suggests a general variational formulation for any GENERIC system, as follows:

Variational formulation of a GENERIC system: Given a GENERIC system {Z, E, S, L, M}, define J as in (27). A function $z: [0, T] \to Z$ is a solution of the GENERIC equation (6) iff $J(z) = 0$.

In full generality, this characterization is formal; no details about the functional setting are stated. In the example of the VFP equation, however, this formulation is exact, as described by lemma 3.4.

Indeed, let us now come back to the question in which sense the VFP equation satisfies the GENERIC equation (24). The discussion above suggests that this variational formulation could be a natural solution concept. Indeed, for any $z = (\rho, e) \in AC([0, T]; Z)$ with finite $S(z(0))$ each of the terms in (27) makes sense as an element of $(-\infty, \infty]$:

- $S(z(T)) \in (-\infty, \infty]$ by definition;
- The assumption that $z \in AC([0, T]; Z)$ implies that for almost all t, $\partial_t \rho$ is a distribution on $\mathbf{R}^{2d}$ and $\partial_t e$ exists in $\mathbf{R}$;
- Under reasonable assumptions on V and ψ, $\mathsf{L}\,\mathrm{grad}\,\mathsf{E} = -\mathrm{div}_q \rho p/m + \mathrm{div}_p \rho \big[\nabla_q V + \nabla_q \psi *$ $\rho \big]$ is well defined in the sense of distributions;
- Therefore the seminorm $\|\partial_t z - \mathsf{L}\,\mathrm{grad}\,\mathsf{E}\|_{\mathsf{M}^{-1}}^2$ is well defined in $[0, \infty]$;
- The seminorm $\|\cdot\|_{\mathsf{M}}^2$ can be assumed well defined in $[0, \infty]$ for any argument, by extending it by $+\infty$ outside of H_{M}.

For the VFP equation there are several other solution concepts that are natural for different reasons and have various advantages; examples are distributional solutions and solutions in the sense of semigroups (since the first and last terms on the right-hand side of (1) form a hypoelliptic operator with a smooth and strictly positive fundamental solution). The relevance of this discussion therefore lies not so much in the specific case of the VFP equation, but more in the potential application to general GENERIC systems.

Remark 4.1. Gradient flows are GENERIC systems with $\mathsf{E} = 0$. For this class of systems, this variational formulation is well known and has been put to good use. For instance, Sandier and Serfarty [28] (see also e.g. [13, 26, 27, 29]) showed how the variational form can be used to pass to limits in parameters in the equation. We expect something similar might be possible for these GENERIC variational formulations, and will return to this in a future publication.   □

5. Synthesis

Let us recapitulate what we have just seen.

- The VFP equation has a variational formulation of the type '$J(z) \geqslant 0$, and $J(z) = 0$ iff z is a solution';
- This variational formulation, the functional J, is identical to the large-deviation rate functional for the stochastic particle system (2) for the case of fixed energy;
- The equation and the variational formulation can both be written in terms of only the GENERIC building blocks;
- This suggests a variational formulation for an arbitrary GENERIC system.

In the remainder of this paper we discuss a number of consequences. In section 6 we use the connection between the VFP equation, large deviations, and the GENERIC structure to shed some light on the properties of GENERIC as formulated in section 1.3. Section 7 is devoted to the generalization mentioned in section 1.2.

6. Interpretation of the GENERIC properties

The GENERIC structure of the VFP equation, introduced in section 3.2, does raise some questions. Why are these bulding blocks the 'right' ones, from a philosophical, or modelling point of view? Is it clear why E and S should be what they are defined to be in (23)? Is it clear why L and M are what they are? Why they do indeed satisfy the various conditions described above?

In addition, the origin of the GENERIC properties themselves, as described in section 1.3, is somewhat obscure. Why should 'every' thermodynamic system satisfy these properties? We now show how the connection with large deviations of the underlying particle system gives us some answers to these questions.

The reversible operator L *and the Hamiltonian* $\mathcal{H}$. First consider the simpler case when $\psi = 0$. Then the only non-zero component of the operator L, which is $\mathsf{L}_{\rho\rho} = -\mathrm{div}\rho J\nabla$, is the Liouville operator for the Hamiltonian flow on $\mathbf{R}^{2d}$ generated by the symplectic matrix J and the Hamiltonian $H(q, p) = p^2/2m + V(q)$. Indeed, $x(t) = (q(t), p(t))$ solves the Hamiltonian equation

$$\frac{\mathrm{d}}{\mathrm{d}t}x = -J\nabla H(x)$$

if and only if $\rho(t) := \delta_{x(t)}$ solves

$$\partial_t\rho - \mathrm{div}(\rho J\nabla H) = 0.$$

Therefore L is the natural embedding of the symplectic geometry of J in $\mathbf{R}^{2d}$ into the space of measures $\mathcal{P}(\mathbf{R}^{2d})$; and when $\psi = 0$, $\mathcal{H}(\delta_x) = H(x)$, and therefore $\mathcal{H}$ similarly is the natural embedding of the $\mathbf{R}^{2d}$-space Hamiltonian H into the space of measures. The antisymmetry and Jacobi identity properties of L follow directly from that of the matrix J.

When ψ is non-zero, a similar interpretation of $\mathcal{H}$ is possible, since with the notation of (20) we have

$$\mathcal{H}\big(\eta_n(x_1, \ldots, x_n)\big) = H_n(x_1, \ldots, x_n), \qquad \text{where} \quad \eta_n(x_1, \ldots, x_n) := \frac{1}{n}\sum_{i=1}^{n}\delta_{x_i}.$$

Similarly, L can be interpreted as the embedding into $\mathcal{P}(\mathbf{R}^{2d})$ of the Hamiltonian flow on $\mathbf{R}^{2nd}$ generated by a symplectic matrix J_n consisting of n copies of J.

The entropy functional S. The functional S in (23) is defined as $e + \mathcal{S}(\rho) = e - \theta \int \rho \log \rho \, \mathrm{d}x$. The second term in this sum is the usual entropy of ρ, multiplied by temperature θ. Its form arises from the *loss of information* in the mapping η_n defined above. We explain it now for the case of finite state $S = \{1, \cdots, r\}$; the general case can be handled using the characterization of the relative entropy as a supremum over finite partitions [14, lemma 1.4.3]. Let $X_1, \cdots, X_n$ be independent identically distributed S-valued random variables with common law μ on a probability space $(\Omega, \Sigma, \mathbb{P})$. Define the (random) empirical measure

$$L_n := \frac{1}{n}\sum_{i=1}^{n}\delta_{X_i}.$$

There is a loss of information in going from $X_1, \cdots, X_n$ to the empirical measure L_n: $L_n(\omega)$ characterizes the observed frequencies of $\{1, \ldots, r\}$ among $X_1(\omega), \ldots, X_n(\omega)$, but does not tell us exactly what values they take. The *degree of degeneracy*, the number of possible ways

that $X_1(\omega), \ldots, X_n(\omega)$ can be such that $L_n(\omega)$ is equal to a given $\rho = (\rho_i)_{i=1}^n = (\frac{k_1}{n}, \ldots, \frac{k_r}{n})$, where $(k_1, \ldots, k_r) \in \mathbb{N}^r$, $\sum_{i=1}^r k_i = n$, is $\frac{n}{k_1 \cdots k_r}$. We have

$$\mathrm{Prob}(L_n = \rho) = \frac{n}{k_1 \cdots k_r} \prod_{i=1}^r \mu_i^{k_i},$$

where $\mu_i = \mu(i)$ for $i = 1, \ldots, r$. Hence

$$\frac{1}{n} \log \mathrm{Prob}(L_n = \rho) = \frac{1}{n} \left(\log n - \sum_{i=1}^r k_i + \sum_{i=1}^r k_i \log \mu_i \right).$$

Using Stirling's formula in the form

$$\log m = m \log m - m + o(m) \qquad \text{as } m \to \infty,$$

we find

$$\frac{1}{n} \log \mathbb{P}(L_n = \rho) \approx \frac{1}{n} \left[n \log n - n - \sum_{i=1}^r (k_i \log k_i - k_i) + \sum_{i=1}^r k_i \log \mu_i \right]$$

$$= \log n - \sum_{i=1}^r \frac{k_i}{n} \log k_i + \sum_{i=1}^r \frac{k_i}{n} \log \mu_i \qquad \left(\text{since } \sum_{i=1}^r k_i = n \right)$$

$$= \sum_{i=1}^r \rho_i (\log n - \log k_i + \log \mu_i) \quad \left(\text{since } \rho_i = \frac{k_i}{n} \text{ and } \sum_{i=1}^r \rho_i = 1 \right)$$

$$= \sum_{i=1}^r \rho_i (-\log \rho_i + \log \mu_i) = -\sum_{i=1}^r \rho_i \log \frac{\rho_i}{\mu_i}.$$

Retracing the steps in this computation we see that the term $\sum_{i=1}^r \rho_i \log \rho_i$ originates from the degree of degeneracy $\frac{n}{k_1 \cdots k_r}$.

The degeneracy condition $\mathsf{L}\,\mathrm{grad}\,\mathsf{S} = 0$. In the case of the VFP equation, this property holds true for any functional which depends locally on ρ, i.e., any functional of the form

$$F(\rho, e) = e + \int f(\rho) \, \mathrm{d}x.$$

The functional S indeed has this form with $f(\rho) = \rho \log \rho$. Therefore the degeneracy $\mathsf{L}\,\mathrm{grad}\,\mathsf{S} = 0$ holds exactly because the entropy is a local functional—and this locality is closely connected to the fact that the entropy characterizes the loss of information encountered when taking a limit and representing the system in terms of (limits of) empirical measures, as described above.

The irreversible operator M *and its properties.* To understand the operator M we use an argument that we learned from Alexander Mielke. We transform the co ordinates $\mathsf{z} = (\rho, e)$ to $\widetilde{\mathsf{z}} = (\widetilde{\rho}, \widetilde{e})$, where

$$\widetilde{\rho} := \rho, \qquad \widetilde{e} := e + \int \mathrm{grad}\,\mathcal{H} \, \mathrm{d}\rho.$$

Then the new variable $\widetilde{\mathsf{z}}$ again solves a GENERIC equation, with new building blocks $\widetilde{\mathsf{L}}$, $\widetilde{\mathsf{M}}$, $\widetilde{\mathsf{E}}$, and $\widetilde{\mathsf{S}}$. Using the change-of-variable formula [41], the operator $\widetilde{\mathsf{M}}$ is given by

$$\widetilde{\mathsf{M}} = \frac{\partial(\widetilde{\mathsf{z}})}{\partial(\mathsf{z})} \mathsf{M} \left[\frac{\partial(\widetilde{\mathsf{z}})}{\partial(\mathsf{z})} \right]^{\mathrm{T}}, \tag{28}$$

where

$$\frac{\partial(\widetilde{z})}{\partial(z)} = \begin{pmatrix} \dfrac{\partial\widetilde{\rho}}{\partial\rho} & \dfrac{\partial\widetilde{\rho}}{\partial e} \\[2mm] \dfrac{\partial\widetilde{e}}{\partial\rho} & \dfrac{\partial\widetilde{e}}{\partial e} \end{pmatrix} = \begin{pmatrix} \mathrm{id} & 0 \\[2mm] \int \Box\,\mathrm{grad}\,\mathcal{H} & \mathrm{id} \end{pmatrix}$$

is the transformation matrix. This formula should be read as operator composition; we write id for the identity operator, both for functions on $\boldsymbol{R}^{2d}$ and for elements of $\boldsymbol{R}$, and we use the notation

$$\int \Box\,\mathrm{grad}\,\mathcal{H} \qquad \text{for the operator} \quad \xi \mapsto \int \xi\,\mathrm{grad}\,\mathcal{H}.$$

Hence

$\widetilde{\mathsf{M}}(\widetilde{z})$

$$= \begin{pmatrix} \mathrm{id} & 0 \\[2mm] \int \Box\,\mathrm{grad}\,\mathcal{H} & \mathrm{id} \end{pmatrix} \begin{pmatrix} -\mathrm{div}_p(\rho\nabla_p\Box) & \Box\,\mathrm{div}_p(\nabla_p\,\mathrm{grad}\,\mathcal{H}) \\[2mm] -\int \nabla_p\,\mathrm{grad}\,\mathcal{H}\cdot\nabla_p\Box\,\mathrm{d}\rho & \Box\int |\nabla_p\,\mathrm{grad}\,\mathcal{H}|^2\,\mathrm{d}\rho \end{pmatrix} \begin{pmatrix} \mathrm{id} & \Box\,\mathrm{grad}\,\mathcal{H} \\[2mm] 0 & \mathrm{id} \end{pmatrix}$$

$$= \begin{pmatrix} \mathrm{id} & 0 \\[2mm] \int \Box\,\mathrm{grad}\,\mathcal{H} & \mathrm{id} \end{pmatrix} \begin{pmatrix} -\mathrm{div}_p(\rho\nabla_p\Box) & 0 \\[2mm] -\int \nabla_p\,\mathrm{grad}\,\mathcal{H}\cdot\nabla_p\Box\,\mathrm{d}\rho & 0 \end{pmatrix}$$

$$= \begin{pmatrix} -\mathrm{div}_p(\rho\nabla_p\Box) & 0 \\[2mm] 0 & 0 \end{pmatrix}.$$

These remarks now enable us to comment on the form of M. First, the transformation to a different set of variables has the effect of 'cleaning up' the operator M: in the new variables $\widetilde{z}$, the operator only acts on the ρ variable. Also, The operator $\widetilde{\mathsf{M}}$ is clearly symmetric and positive semidefinite. The same properties for M then follow as a consequence of (28).

The operator $-\mathrm{div}_p(\rho\nabla_p\Box)$ that appears in $\widetilde{\mathsf{M}}$ is a familiar figure. It also appears in the characterization of Wasserstein gradient flows [40], and originates in the fluctuation behaviour of the Brownian noise in those systems—as is the case in theorem 2.5. In the SDE (2), however, the noise only appears in the P-variable, and as a consequence the operator $-\mathrm{div}_p(\rho\nabla_p\Box)$ also only operates on the p-variables. The symmetry of this operator is a consequence of Itō's formula: in this formula for the stochastic evolution of functions $f(X_t)$ of a stochastic variable X_t, the second derivative $\mathrm{d}^2 f$ appears, and this second derivative gives rise to the second-order derivative in $-\mathrm{div}_p(\rho\nabla_p\Box)$. The symmetry of this expression therefore has the same origin as the symmetry of second-derivative matrices of functions.

In the new variables, the degeneracy condition $\widetilde{\mathsf{M}}\,\mathrm{grad}\,\widetilde{\mathsf{E}}$ is natural; indeed, $\widetilde{\mathsf{E}}(\widetilde{z}) = \widetilde{\mathsf{E}}((\widetilde{\rho},\widetilde{e})) = \widetilde{e}$. Therefore $\widetilde{\mathrm{grad}\,\mathsf{E}} = (0,1)$, and the degeneracy condition coincides with the property that only $\widetilde{\mathsf{M}}_{\rho\rho}$ is non-zero.

To conclude, the connection between large deviations and the GENERIC structure in the case of the VFP equation allows us to understand and explain where the various properties of the GENERIC formalism come from:

- The antisymmetry and the Jacobi identity of L follow from the same properties of the underlying Hamiltonian system;
- The symmetry of M follows from the symmetry of second derivatives, as they appear in Itō's formula;
- The energy E is (an extended version of) the Hamiltonian of the underlying system, after embedding into the space of measures;

- The entropy S characterizes the loss of information upon passing to empirical measures, in the sense of large deviations;
- The degeneracy condition $L \operatorname{grad} S = 0$ arises from the fact that S is a local functional;
- The degeneracy condition $M \operatorname{grad} E = 0$ arises as a consequence of energy conservation.

7. GENERIC formulation of the generalized VFP equation and its variational structure

Once the variational structure of the VFP equation (1) has been recognized, a natural generalization of the VFP equation presents itself. By replacing the various terms by their equivalents in terms of S and H one arrives at equation (3). In this section, we show that this equation, after extension, also is a GENERIC system for arbitrary S and H, and we compute the corresponding functional J explicitly. This section is necessarily formal.

By computing the derivative $\partial H(\rho_t)$ for a solution ρ of (3) we construct the extended version of (3):

$$\partial_t \rho = \operatorname{div}(\rho J \nabla \operatorname{grad} H) + \operatorname{div}\big(\mathcal{D}(\rho)\nabla \operatorname{grad}(H + S)\big), \tag{29a}$$

$$\frac{\mathrm{d}}{\mathrm{d}t} e = \int_{R^{2d}} \nabla \operatorname{grad} H \cdot \mathcal{D}(\rho) \cdot \nabla \operatorname{grad}(H + S). \tag{29b}$$

Here $\mathcal{D}(\rho) := \rho \sigma \sigma^{\mathrm{T}}$. The corresponding GENERIC building blocks are

$$Z = \mathcal{P}_2(R^{2d}) \times R, \qquad E(\rho, e) = H(\rho) + e, \qquad L = L(\rho, e) = \begin{pmatrix} L_{\rho\rho} & 0 \\ 0 & 0 \end{pmatrix},$$

$$z = (\rho, e), \qquad S(\rho, e) = S(\rho) + e, \qquad M = M(\rho, e) = \gamma \begin{pmatrix} M_{\rho\rho} & M_{\rho e} \\ M_{e\rho} & M_{ee} \end{pmatrix}, \tag{30}$$

where the components of L and M are given by

$$L_{\rho\rho}\xi = \operatorname{div}\rho J \nabla \xi, \qquad M_{\rho\rho}\xi = -\operatorname{div}\big(\mathcal{D}(\rho)\nabla \xi\big), \qquad M_{\rho e} r = r \operatorname{div}\big(\mathcal{D}(\rho)\nabla \operatorname{grad} H\big),$$

$$M_{e\rho}\xi = -\int_{R^{2d}} \nabla \xi^{\mathrm{T}} \cdot \mathcal{D}(\rho) \cdot \nabla \operatorname{grad} H, \qquad M_{ee} r = r \int_{R^{2d}} (\nabla \operatorname{grad} H)^{\mathrm{T}} \cdot \mathcal{D}(\rho) \cdot \nabla \operatorname{grad} H.$$

Most of the GENERIC properties of section 1.3 follow immediately from this setup, such as the antisymmetry and symmetry of L and M, the Jacobi identity, the positive semidefiniteness of M. The degeneracy condition $M \operatorname{grad} E = 0$ can be checked explicitly, but it can also be understood in the same way as in section 6, by first transforming the system to a new set of variables.

Finally, the degeneracy condition $L \operatorname{grad} S$ requires a specific assumption, as we already encountered above:

Lemma 7.1. *If $S(\rho) = \int f(\rho)$ for some function f, then system (7) is a GENERIC system with the building blocks (7).*

The proof consists of simple verification.

By following the same arguments as in section 2, we find a variational formulation of exactly the same type: a curve $z \in AC([0, T]; Z)$ is a variational solution if $J(z) = 0$, where J is defined by (27) with building blocks (7). We have the following characterization:

Lemma 7.2. *For equation (7) the functional J, defined in (27), can be characterized as follows: If*

$$\frac{\mathrm{d}}{\mathrm{d}t} \begin{pmatrix} \rho \\ e \end{pmatrix} = \mathrm{VFPg}(\rho, e) + \begin{pmatrix} \operatorname{div}(\mathcal{D}(\rho)\nabla \eta) \\ \int \mathcal{D}(\rho)\nabla \eta \cdot \nabla \operatorname{grad} H \end{pmatrix},$$

then

$$J(\rho, e) = \frac{1}{2} \int_0^T \int_{\mathbf{R}^{2d}} \nabla \eta^T \cdot \mathcal{D}(\rho) \cdot \nabla \eta \, \mathrm{d}x \, \mathrm{d}t.$$

Here $\mathrm{VFPg}(\rho, e)$ *is the right-hand side of (7).*

The proof follows the same lines as as lemmas 3.3 and 3.4.

Acknowledgments

References

[1] Bernardin C and Olla S 2005 Fourier's law for a microscopic model of heat conduction *J. Stat. Phys.* **121** 271–89
[2] Ottobre M and Pavliotis G A 2011 Asymptotic analysis for the generalized Langevin equation *Nonlinearity* **24** 1629–53
[3] Kipnis C, Marchioro C and Presutti E 1982 Heat flow in an exactly solvable model *J. Statist. Phys.* **27** 65–74
[4] Méléard S 1996 Asymptotic behaviour of some interacting particle systems; McKean–Vlasov and Boltzmann models *Probabilistic Models for Nonlinear Partial Differential Equations (Montecatini Terme, 1995)* (*Lecture Notes in Mathematics* vol 1627) (Berlin: Springer) pp 42–95
[5] Bolley F, Guillin A and Malrieu F 2010 Trend to equilibrium and particle approximation for a weakly selfconsistent Vlasov–Fokker–Planck equation *M2AN Math. Model. Numer. Anal.* **44** 867–84
[6] Wu L 2001 Large and moderate deviations and exponential convergence for stochastic damping Hamiltonian systems *Stoch. Proc. Appl.* **2** 205–38
[7] Ikeda N and Watanabe S 1981 *Stochastic Differential Equations and Diffusion Processes* (Amsterdam: North-Holland)
[8] Cattiaux P and Léonard C 1995 Large deviations and Nelson processes *Forum Mathematicum* **7** 95–116
[9] Karatzas I and Shreve S E 1991 *Brownian Motion and Stochastic Calculus* vol 113 (Berlin: Springer)
[10] Carrillo J A and Soler J 1997 On the Vlasov–Poisson–Fokker–Planck equations with measures in Morrey spaces as initial data *J. Math. Anal. Appl.* **207** 475–95
[11] Victory H D Jr and O'Dwyer B P 1990 On classical solutions of Vlasov–Poisson–Fokker–Planck systems *Indiana Univ. Math. J.* **39** 105–56
[12] Chandrasekhar S 1943 Stochastic problems in physics and astronomy *Rev. Mod. Phys.* **15** 1–89
[13] Serfaty S 2009 Gamma-convergence of gradient flows on Hilbert and metric spaces and applications www.math.nyu.edu/faculty/serfaty/gcv-erice2.pdf
[14] Dupuis P and Ellis R S 1997 *A Weak Convergence Approach to the Theory of Large Deviations* vol 902 (New York: Wiley-Interscience)
[15] Renger D R M 2013 Microscopic interpretation of Wasserstein gradient flows *PhD Thesis* TU Eindhoven
[16] Otto F 2001 The geometry of dissipative evolution equations: the porous medium equation *Commun. Partial. Diff. Eqns* **26** 101–74
[17] Dawson D A and Gärtner J 1987 Large deviations from the McKean–Vlasov limit for weakly interacting diffusions *Stochastics* **20** 247–308
[18] Dawson D A and Gärtner J 1989 Large deviations, free energy functional and quasi-potential for a mean field model of interacting diffusions *Mem. Amer. Math. Soc.* **78** iv+94
[19] Feng J and Kurtz T G 2006 Large deviations for stochastic processes (*Mathematical Surveys and Monographs* vol 131) (Providence, RI: American Mathematical Society)
[20] Hollander F d 2000 *Large Deviations* (Providence, RI: American Mathematical Society)
[21] Feng J and Świech A 2013 Optimal control for a mixed flow of Hamiltonian and gradient type in space of probability measures *Trans. Am. Math. Soc.* **365** 3987–4039 (With an appendix by Atanas Stefanov)
[22] Chavanis P H 2001 Optimal control for a mixed flow of Hamiltonian and gradient type in space of probability measures *Phys. Rev.* E **64** 026309
[23] Caglioti E, Pulvirenti M and Rousset F 2008 The 2D constrained Navier–Stokes equation and intermediate asymptotics *J. Phys. A. Math. Theor.* **41** 344001

[24] Caglioti E, Pulvirenti M and Rousset F 2009 On a constrained 2-D Navier–Stokes equation *Commun. Math. Phys.* **290** 651–77

[25] Fischer M, On the form of the large deviation rate function for the empirical measures of weakly interacting systems *Bernoulli* submitted (arXiv:1208.0472)

[26] Le N Q 2008 A gamma-convergence approach to the Cahn–Hilliard equation *Cal. Var. Partial. Diff. Eqns* **32** 499–522

[27] Stefanelli U 2008 The Brezis–Ekeland principle for doubly nonlinear equations *SIAM J. Control Optim.* **47** 1615

[28] Sandier E and Serfaty S 2004 Gamma-convergence of gradient flows with applications to Ginzburg–Landau *Commun. Pure Appl. Math.* **57** 1627–72

[29] Arnrichm S, Mielke A, Peletier M A, Savaré G and Veneroni M 2012 Passing to the limit in a Wasserstein gradient flow: from diffusion to reaction *Calc. Var. Partial Diff. Eqns* **44** 419–54

[30] Hütter M and Tervoort T A 2008 Thermodynamic considerations on non-isothermal finite anisotropic elasto-viscoplasticity *J. Nonnewton Fluid Mech.* **152** 53–65

[31] Hütter M and Tervoort T A 2008 Finite anisotropic elasticity and material frame indifference from a nonequilibrium thermodynamics perspective *J. Nonnewton Fluid Mech.* **152** 45–52

[32] Mielke A 2011 Formulation of thermoelastic dissipative material behavior using GENERIC *Continuum Mech. Therm.* **23** 233–56

[33] Budhiraja A, Dupuis P and Fischer M 2012 Large deviation properties of weakly interacting processes via weak convergence methods *Ann. Probab.* **40** 74–102

[34] Öttinger H C 2005 *Beyond Equilibrium Thermodynamics* (New York: Wiley-Interscience)

[35] Dembo A and Zeitouni O 1998 Large deviations techniques and applications (*Stochastic Modelling and Applied Probability* vol 38) 2nd edn (New York: Springer)

[36] Duong M H, Laschos V and Renger M 2013 Wasserstein gradient flows from large deviations of many-particle limits *ESAIM: Control, Optim. Cal. Var.* eFirst

[37] Düring B, Matthes D and Milišić J 2010 A gradient flow scheme for nonlinear fourth order equations *Discrete Contin. Dyn. Syst. Ser.* B **14** 935–59

[38] Carrillo J A, McCann R J and Villani C 2003 Kinetic equilibration rates for granular media and related equations: entropy dissipation and mass transportation estimates *Rev. Mat. Iberoamericana* **19** 971–1018

[39] Adams S, Dirr N, Peletier M A and Zimmer J 2011 From a large-deviations principle to the Wasserstein gradient flow: a new micro-macro passage *Commun. Math. Phys.* **307** 791–815

[40] Adams S, Dirr N, Peletier M A and Zimmer J 2012 Large deviations and gradient flows arXiv:1201.4601

[41] Grmela M and Öttinger H C 1997 Dynamics and thermodynamics of complex fluids: I. Development of a general formalism *Phys. Rev.* E **56** 6620–32

[42] Grmela M and Öttinger H C 1997 Dynamics and thermodynamics of complex fluids: II. Illustrations of a general formalism *Phys. Rev.* E **56** 6633–55

[43] Ambrosio L, Gigli N and Savaré G 2008 Gradient flows in metric spaces and in the space of probability measures *Lectures in Mathematics ETH Zürich* 2nd edn (Basel: Birkhauser)

[44] Kramers H A 1940 Brownian motion in a field of force and the diffusion model of chemical reactions *Physica* **7** 284–304

[45] Cattiaux P and Léonard C 1994 Minimization of the Kullback information of diffusion processes *Ann. Inst. Henri Poincaré Probab. Statist.* **30** 83–132

[46] Cattiaux P and Léonard C 1995 Correction to: 'Minimization of the Kullback information of diffusion processes' *Ann. Inst. Henri Poincaré Probab. Statist.* **31** 705–7

[47] Cattiaux P and Léonard C 1995 Minimization of the Kullback information of diffusion processes *Forum Mathematicum* **7** 95–115

[48] Dirr N, Laschos V and Zimmer J 2012 Upscaling from particle models to entropic gradient flows *J. Math. Phys.* **53** 063–074

[49] Peletier M A and Renger D R M Variational formulation of the Fokker–Planck equation with decay: a particle approach *Commun. Contemp. Math.* submitted (arXiv:1108.3181)