

## Deterministic thermostats, theories of nonequilibrium systems and parallels with the ergodic condition

This article has been downloaded from IOPscience. Please scroll down to see the full text article.

2010 J. Phys. A: Math. Theor. 43 133001

(<http://iopscience.iop.org/1751-8121/43/13/133001>)

View [the table of contents for this issue](#), or go to the [journal homepage](#) for more

Download details:

IP Address: 171.66.16.157

The article was downloaded on 03/06/2010 at 08:42

Please note that [terms and conditions apply](#).

## TOPICAL REVIEW

# Deterministic thermostats, theories of nonequilibrium systems and parallels with the ergodic condition

Owen G Jepps<sup>1</sup> and Lamberto Rondoni<sup>2</sup>

<sup>1</sup> School of Biomolecular and Physical Sciences and Queensland Micro- and Nanotechnology Centre, Griffith University, Brisbane QLD 4111, Australia

<sup>2</sup> Dipartimento di Matematica and INFN, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino, Italy

E-mail: [o.jepps@griffith.edu.au](mailto:o.jepps@griffith.edu.au) and [lamberto.rondoni@polito.it](mailto:lamberto.rondoni@polito.it)

Received 6 August 2009, in final form 8 January 2010

Published 10 March 2010

Online at [stacks.iop.org/JPhysA/43/133001](http://stacks.iop.org/JPhysA/43/133001)**Abstract**

Deterministic ‘thermostats’ are mathematical tools used to model nonequilibrium steady states of fluids. The resulting dynamical systems correctly represent the transport properties of these fluids and are easily simulated on modern computers. More recently, the connection between such thermostats and entropy production has been exploited in the development of nonequilibrium fluid theories. The purpose and limitations of deterministic thermostats are discussed in the context of irreversible thermodynamics and the development of theories of nonequilibrium phenomena. We draw parallels between the development of such nonequilibrium theories and the development of notions of ergodicity in equilibrium theories.

PACS numbers: 02.70.Ns, 05.40.-a, 05.45.-a, 05.70.Ln, 45.50.-j

**Abbreviations**

|               |   |
|---------------|---|
| EMD           | equilibrium molecular dynamics                                    |
| FR            | fluctuation relation  |
| IK            | isokinetic (i.e. with constant kinetic energy; thermostatted)     |
| IE            | isoenergetic (i.e. with constant internal energy; ergostatted)    |
| LTE           | local thermodynamic equilibrium                                   |
| MC            | Monte Carlo   |
| MD            | molecular dynamics  |
| NEMD          | non-equilibrium molecular dynamics                                |
| ODE           | ordinary differential equation                                    |
| SLLOD         | (not an acronym) it denotes one NEMD algorithm for Couette flows  |
| $\Lambda$ -FR | fluctuation relation for the phase space expansion rate $\Lambda$ |
| $\Omega$ -FR  | fluctuation relation for the dissipation function $\Omega$        |

## 1. Introduction

Statistical mechanics is a powerful framework for understanding equilibrium thermodynamics, notwithstanding the deep mathematical questions that it poses. These questions can arise from quite modest observations of real-life thermodynamic systems. For example, the behaviour of such systems must be independent of their microscopic initial conditions, since in reality we cannot know what those initial conditions are. The challenge is to turn the physical notion into an appropriate mathematical one: for this example, the effort has a well-documented history (see e.g. [1–3]), leading in the present day to the thorny issue of ergodicity. Ergodicity is a strong condition, requiring the time average of *all integrable functions* to be independent of initial conditions and equal to the function's phase average over a probability distribution whose existence is guaranteed. In reality, it is a stronger condition than is warranted by the physical observation, which is limited to only a handful of thermodynamic properties (even though it could hold for many more). It is also a condition that is notoriously difficult to prove—indeed, it is known *not* to hold in many systems whose behaviour is certainly thermodynamic. So while ergodicity is a mathematically rigorous notion, its connection with the physical phenomenon that motivates it is complicated.

An alternative to the ergodic approach was championed by Khinchin. Recognizing the limited class of functions for which independence from initial conditions was really motivated, Khinchin demonstrated that an ergodic-like relation held for such a function if and only if there was a sufficiently fast decay in its auto-correlation function. This is a much weaker result, but a more natural one from the physical viewpoint: not least because the existence of auto-correlation integrals is a requirement for linear and nonlinear response theories. This approach identifies an important requirement that relevant physical properties must (and do) satisfy, i.e. for theory to match reality. But Khinchin's approach does not specifically prescribe the entire universe of functions and systems for which one can expect such convenient physical behaviour—and this turns out to be a very difficult problem mathematically.

While both approaches provide powerful insights into the behaviour of the systems under examination, neither can fully satisfy both the physical and mathematical aspects of the discourse. The first approach invokes some heavy mathematical artillery, but does not explain why non-ergodic physical systems nevertheless display thermodynamic ('ergodic-like') behaviour. Worse still, the mathematical property that seemingly *is* required for thermodynamic behaviour is that of *mixing*: a property that implies a rapid decay of correlations, and is mathematically even stronger than 'mere' ergodicity. The second approach invokes known properties of physical systems, but only specifies a *condition* on the range of systems and properties to which the result applies (and not the range itself).

An equivalent theoretical framework for nonequilibrium systems has remained elusive. In recent years, observations of computer simulations of *nonequilibrium* systems led to the first of various fluctuation relations (FRs) in this vein. These FRs relate the nature of fluctuations forward and backward in time for nonequilibrium systems. Interestingly, a similar vexed dichotomy between physically and mathematically motivated approaches appears to have emerged from the efforts to provide a theoretical framework for the FRs. Beyond this point, the bridge between the mathematical and physical aspects of these theories has a further complication—the nonequilibrium requirement of entropy production.

The Hamiltonian formalism of classical mechanics provides suitable dynamical equations for equilibrium systems: how does one change this dynamics to incorporate the effects of entropy production? Fortunately, this question has already been addressed in the development of computer simulation algorithms for nonequilibrium systems. This approach augments the Hamiltonian equations of motion with fictitious driving forces used to represent the

thermodynamic forces driving the system away from equilibrium. Such forces introduce energy that must be dissipated if nonequilibrium steady states are to be obtained: one therefore introduces further terms, collectively called a *thermostat*, whose original application was to generate constant-temperature (as opposed to constant-energy) equilibrium ensembles. For nonequilibrium simulations, the thermostat dissipates the energy introduced by non-equilibrium driving forces, and therefore reflects the entropy they produce. Thermostats therefore allow a mathematical representation of the entropy production, and play a central role in the description of the dynamics at the heart of the FRs.

The point of departure for this review is an overview of the ergodic question, in anticipation of the analogous issues that emerge for the nonequilibrium FRs. We then present a review of thermostats, with particular attention to how they have developed from a simulation tool into a more formal mechanism for representing nonequilibrium phenomena. We examine the connection between the mathematics of the dynamical representation and the physical notions of entropy, which is fundamental to theories of nonequilibrium systems such as the FRs. From here, we examine the theories of nonequilibrium fluctuations, highlighting how the strengths and weaknesses of the various approaches bear a striking resemblance to those regarding the ergodic question, before drawing conclusions from this perspective.

## 2. The ergodic question

The goal of statistical mechanics is to understand the macroscopic properties of physical systems from the dynamics of their microscopic constituents, by means of suitable averaging procedures. For instance, consider a system comprising  $N$  classical particles in  $d$  dimensions, and let its dynamics be described by

$$\dot{\Gamma} = G(\Gamma); \quad \Gamma = (\mathbf{q}, \mathbf{p}) \in \mathcal{M} \subset \mathbb{R}^{2dN}, \quad (1)$$

where  $\mathcal{M}$  is the phase space, and the vector field  $G$  is determined by the forces acting on the system and by the particles' interactions. Denote by  $S^t\Gamma$ ,  $t \in \mathbb{R}$ , the solution of equation (1) with the initial condition  $\Gamma$ . The macroscopic quantity associated with an *observable*, i.e. with a function of phase  $\mathcal{O} : \mathcal{M} \rightarrow \mathbb{R}$ , is defined by

$$\bar{\mathcal{O}}(\Gamma) = \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t \mathcal{O}(S^s\Gamma) ds. \quad (2)$$

This reflects the fact that macroscopic observations occur on time scales which are long compared to the time scales of the microscopic dynamics (equation (1)), and that an observation amounts to a time average of the chosen observable  $\mathcal{O}$ . The function space  $\mathcal{F}(\mathcal{M})$  in which the limit (equation (2)) is required to exist is called the space of observables.

Equation (2) suffers from various difficulties of a practical nature. In the first place, computing the limit may not be a trivial task at all, because of the many degrees of freedom involved. Furthermore, equation (2) is not particularly suitable for further theoretical derivations, given its highly implicit form.

For systems in (or evolving towards) equilibrium<sup>3</sup>, the problem is commonly solved by assuming that the system is *ergodic*, i.e. that

$$\bar{\mathcal{O}}(\Gamma) = \frac{1}{\mu(\mathcal{M})} \int_{\mathcal{M}} \mathcal{O}(y) d\mu(y) = \langle \mathcal{O} \rangle_{\mu} \quad (3)$$

for a properly chosen probability measure  $\mu$  on  $\mathcal{M}$ , and for  $\mu$ , almost all  $\Gamma \in \mathcal{M}$ . In principle, ergodic theory should identify the dynamics verifying equation (3), and the corresponding

<sup>3</sup> A physical system is in an *equilibrium* state if all currents—of heat, momentum, etc—vanish, and the system is uniquely described by a set of *state variables* which do not change with time.

measures  $\mu$ . In practice, this is too difficult to carry out in most cases of physical interest, so instead we *hypothesize* that the system is ergodic. Therefore, one commonly *postulates* that  $\mu$  has a density  $\rho$ , specific to the ensemble at hand. Fermi’s work convinced physicists that ergodicity was not a problem for systems of physical interest [4]—this hypothesis was later challenged by systems such as the Fermi–Pasta–Ulam model [5], whose ergodicity immediately, unexpectedly and correctly appeared far from obvious [6].

One therefore checks *a posteriori* whether these assumptions are valid or not, and finds that the classical ensembles [2] describe most equilibrium situations very well, despite the fact that the mathematical notion of ergodicity is violated, for instance, by the ideal gas in a cubic box, by chains of harmonic oscillators and by blackbody radiation. Indeed, even in cases where ergodicity is clearly violated, the assumption of ergodicity leads to thermodynamic results.

The fact is that the ergodic hypothesis can be adopted for *practical* purposes, such as computing the averages of physically relevant quantities: this set of observables is often too small, and the observation times too short, to probe the finer differences among the ensemble probability distributions for the high dimensional systems of interest in *Statistical Mechanics* (cf chapter 1 of [2]). For these purposes, such differences are essentially undetectable, particularly as the classical ensembles become in some sense ‘equivalent’ in the thermodynamic limit (whereby the system volume grows to infinity while the number and energy densities tend towards constants).

Given our primary interest in this narrow set of properties, one could consider an alternative approach which focuses on demonstrating the ergodic property for a restricted range of functions. This is precisely the approach undertaken by Khinchin. In his celebrated book, *Mathematical Foundations of Statistical Mechanics*, Khinchin obtained important results concerning the ergodic and mixing problem in statistical mechanics, without recourse to the scarcely physically relevant concept of metric transitivity (as implied by the mathematical ergodic notion of Birkhoff’s theorem) [7]. His physical perspective is based on the following premises: that statistical mechanics concerns systems with a large number of degrees of freedom; that the physical observables are but a few and quite special functions; and that it is physically acceptable that ensemble averages do not coincide with time averages, for sets of initial conditions whose measure vanishes in the  $N \rightarrow \infty$  limit. In particular, Khinchin considered systems with a separable Hamiltonian, i.e. a Hamiltonian which is the sum of single particle contributions,

$$H = \sum_{n=1}^N H_n(\mathbf{q}_n, \mathbf{p}_n),$$

and took the *sum functions*, defined by a sum over single particle contributions,

$$f(\Gamma) = \sum_{n=1}^N f_n(\mathbf{q}_n, \mathbf{p}_n)$$

where  $f_n = O(1)$ , as the only observables. The pressure and the kinetic energy are examples of such functions.

Denoting by  $\langle \cdot \rangle$  the microcanonical ensemble average, Khinchin demonstrated the following under quite general hypotheses:

$$\text{Prob} \left( \frac{|\bar{f} - \langle f \rangle|}{|\langle f \rangle|} \geq K_1 N^{-1/4} \right) \leq K_2 N^{-1/4},$$

where  $K_1$  and  $K_2$  are  $O(1)$ . This means that ensemble averages of sum functions only differ from their time averages by more than a given tolerance along trajectories whose initial conditions have a vanishing measure in the  $N \rightarrow \infty$  limit.

Mazur and van der Linden [8] later extended this result to systems of particles interacting via short range potentials. In this theory, the good statistical properties required for the normal thermodynamic behaviour are basically ascribed only to the fact that  $N$  is very large. This leads to the conclusion that the details of the microscopic (phase space) dynamics are practically irrelevant for a notion of ergodicity that describes physical systems, and indeed they ought to be irrelevant, for thermodynamic behaviour to be as common as it is.

Now, consider an observable  $\mathcal{O}$  which depends on the coordinates of only a few particles, in a system with many degrees of freedom and weak particle interactions<sup>4</sup>. Khinchin shows that equation (3) holds for all  $\Gamma \in \mathcal{M}$  if its auto-correlation satisfies the relation

$$\lim_{t \rightarrow \infty} \langle \mathcal{O}(t)\mathcal{O}(0) \rangle = \langle \mathcal{O} \rangle^2 \tag{4}$$

for almost all initial conditions. This is the case if the memory of the initial state, carried by  $\mathcal{O}$ , is gradually lost, so that values of  $\mathcal{O}$  reached after sufficiently long times become decorrelated from the value in the initial state. In other words, the evolution of the system at hand, observed through the evolution of  $\mathcal{O}$ , is irreversible, because it is impossible to use the final values of  $\mathcal{O}$  to recover information about the initial state, or to return to it. We may then define those evolutions such that equation (4) holds for the set of *physically relevant* observables as *physically irreversible*.

Conversely, Kubo heuristically argues that the auto-correlation  $C_{\mathcal{O}}(t) = \langle \mathcal{O}(t)\mathcal{O}(0) \rangle$  satisfies (4) if equation (3) holds uniformly in  $\mathcal{M}$  (cf [9]): something, however, that seldom happens in dynamical systems<sup>5</sup>.

It must be stressed that these are much weaker results than the mathematical concept of ergodicity, which requires the validity of equation (3) for *all* elements of a large function space, irrespective of the system dimension. These results concern only specific observables for systems with many degrees of freedom. Furthermore, Khinchin’s approach was only developed for the constant-energy ensemble.

As correlation functions are closely related to the response function, it is quite natural to investigate the connection between ergodicity and linear response theory. This issue has been thoroughly examined by Lee [10–14]. The main result is a connection between ergodic behaviour and the response of a system to an external perturbation. One finds that physical irreversibility, as defined above, is a necessary but not sufficient condition for ergodicity, while the validity of

$$\int_0^\infty C_{\mathcal{O}}(t) dt < \infty \tag{5}$$

is a sufficient condition. Note that equation (5) is implied by the Green–Kubo formulae, when the static susceptibility

$$\chi = \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t \int_0^s R_{\mathcal{O}}(s-r) ds dr, \quad \text{where} \quad R_{\mathcal{O}}(s) = -\frac{d}{ds} C_{\mathcal{O}}(s),$$

is finite. In [10–12], criteria are given to determine the validity of (5).

<sup>4</sup> Khinchin considers this case, but it seems reasonable that the result should also hold in systems of particles interacting through a short range potential, like those of [8].

<sup>5</sup> Typically, there are periodic orbits (a set of vanishing measures), in which the time averages differ from those of other trajectories.

So what is the state of affairs for the nonequilibrium analogue? The situation is more complex in the case of nonequilibrium systems in a steady state<sup>6</sup>. The connection between a microscopic description and the phenomena described by irreversible thermodynamics appears harder to find than in the equilibrium case. Specifically, the nonequilibrium counterparts of the classical ensembles and of the ergodic hypotheses are not clearly identified yet: the form of a measure  $\mu$  to be used in equation (3) is not generally known, neither are the conditions under which a single realization of the nonequilibrium system would correspond to such an ensemble average. Consequently, the need arises for simple models on which various hypotheses can be tested. Various proposals have been made in the literature, and various models used to assess these proposals and to derive expressions that relate properties of practical interest. In the latter sections of this review, we focus on some of the models of nonequilibrium systems that have been devised over the past few decades, as well as their use in present day nonequilibrium statistical mechanics.

### 3. Thermostats

The use of computers to simulate molecular systems has grown from humble beginnings to become, in some fields, an indispensable research tool. Molecular simulation allows both qualitative and quantitative investigation of real-life systems, thanks to the progress across many disciplines (including technical and technological). Given that most real-world systems are in (thermal) equilibrium with their surroundings, Monte Carlo (MC) techniques provide a natural way of simulating such systems. MC relies upon ergodicity explicitly: the observable measurement  $\bar{O}$  is estimated as the weighted average  $\langle O \rangle$ , generated through an appropriate sampling of phase space. However, such ensemble averages cannot be calculated for systems where the ensemble distribution is not known (such as nonequilibrium systems) or where time-ordered behaviour is sought (such as the calculation of transport coefficients from autocorrelation functions). For these purposes, MD is the prevailing technique. Interestingly, equilibrium MD (EMD) also makes implicit use of an ergodic hypothesis: for the assumed equivalence of time averages from any initial condition and for the determination of thermodynamic properties such as entropy, temperature or heat capacity.

MD takes the mechanical approach by solving the evolution dynamics of particles, but is naturally suited to simulations of systems *isolated* from their surroundings, rather than in (thermal) equilibrium with them. The dynamics must therefore be manipulated in order to generate the relevant equilibrium ensemble, which Gibbs proposed as idealizations of collections of systems in a given identical thermodynamic state. In achieving thermal equilibrium, the resulting manipulation is termed a ‘thermostat’: analogous terms such as ‘barostat’ have been coined for other ensembles. The extension of MD to other ensembles, and to nonequilibrium systems, has focussed on manipulating the dynamics to create *behaviour consistent with the target system*. The claim is not that these equations are somehow a true representation of the underlying dynamics—only that the reported values be correct for properties we are interested in. Much of the theoretical work underpinning the success of MD has been dedicated to demonstrations along these lines.

This is particularly true of nonequilibrium MD (NEMD), whose original purpose was to compute transport coefficients. NEMD introduces artificial or ‘fictitious’ forces that provide a convenient mechanical substitute for the thermodynamic forces (gradients in density, temperature, shear rate) driving real nonequilibrium behaviour. The key advantage is that

<sup>6</sup> These are systems whose state parameters remain constant in time, but such that there is transport of mass, energy, entropy, etc.

simulations driven by these fictitious forces are generally more efficient than those performed using thermodynamic boundary conditions [15, 16]. In NEMD, the thermostat dissipates precisely that energy introduced into the system by the fictitious forces. This procedure appears to generate ensembles that are in some way the nonequilibrium counterparts of the equilibrium ensembles, in the sense that these states appear to depend only on the macroscopic boundary conditions. Unlike the equilibrium case, however, our understanding of these ensembles, their distributions and indeed their interpretation, is still limited. There is no distribution that we can hope to match, so we must turn to other criteria to assess their merits. We will consider the physical appropriateness and relevance of thermostatted dynamics in much greater detail in section 4. For now we note that, for practical purposes, assessment of their validity has focussed on equivalence of transport coefficients between the equilibrium and nonequilibrium systems. In some instances, there is a well-developed theory [15, 17–19]. In other cases, such as the simple algorithm of [20] for the calculation of the viscosity of a simple fluid, such a theory is not yet available (despite the algorithm’s success).

For the rest of this section, we outline the origin of thermostats in molecular simulation, focussing on thermostats of particular relevance to the studies of nonequilibrium theories.

### 3.1. Early thermostats

The aim of the earliest thermostats was to provide a means of simulating thermal equilibrium, rather than providing a comprehensive dynamical theory on the subject. Thermostatting was performed with recourse to the equipartition theorem: controlling the mean kinetic energy per degree of freedom was equivalent to controlling the system temperature. The earliest thermostats [21] involved the regular rescaling of each particle’s momentum over the course of a simulation—more sophisticated approaches soon followed [22, 23], although not without their problems. Most pertinent here are the effect of scaling on the time-correlation data, and the lack of time-reversibility. Correlation functions are required to calculate transport coefficients for equilibrium systems: continual rescaling leads to artefacts in their estimates that depend on the particular rescaling scheme [24]. This problem was overcome with the differential correction schemes, of which the most well known is arguably the Berendsen thermostat:

$$\dot{\mathbf{q}} = \frac{\mathbf{p}}{m}, \quad \dot{\mathbf{p}} = \mathbf{F} - \frac{1}{2\tau} \left(1 - \frac{T_0}{T}\right) \mathbf{p}. \quad (6)$$

Here the additional non-Hamiltonian term transforms the discrete rescaling process into a differential contribution in the equations of motion, governed by a time scale  $\tau$  and the target temperature  $T_0$  ( $T$  is the instantaneous kinetic temperature). This technique allows a ‘softer’ control of the temperature—the system is directed towards the target temperature, rather than periodically rescaled to it. However, the Berendsen scheme is still not time-reversible, as can be seen from equation (6).

**Remark.** The notion of time-reversibility considered here concerns the equations of motion, i.e. the microscopic dynamics. If  $S^t : \mathcal{M} \rightarrow \mathcal{M}$  is the time evolution operator for any time  $t \in \mathbb{R}$ , so that  $S^t \Gamma$  is the solution of the equations of motion with the initial condition  $\Gamma \in \mathcal{M}$ , the microscopic dynamics are called *time reversal invariant* if there exists a time reversal operator  $I : \mathcal{M} \rightarrow \mathcal{M}$ ,  $I^2 = \text{identity}$ , such that

$$S^t I \Gamma = I S^{-t} \Gamma \quad \text{for all } t \in \mathbb{R}, \quad \text{and all } \Gamma \in \mathcal{M}. \quad (7)$$

For instance,  $I$  is the reversal of momenta for equation (8) below, but is more complicated for the shearing system of [25]. This reversibility, however, has no direct relation with



the thermodynamic notion of reversibility of quasi-static transformations, which concern the macroscopic behaviour. As a matter of fact, a transformation of a macroscopic body can be irreversible and very far from local thermodynamic equilibrium (LTE), while its microscopic constituents still obey reversible equations of motion.

Time-reversibility is not a necessary property for the purposes of calculating dynamical properties of molecular systems, as the widespread use of the Berendsen thermostat attests. However, in addressing fundamental questions regarding dynamical systems, it is important that our models alter as little as possible the nature of the microscopic dynamics as we know it. As such, time-reversibility is an important property that dynamics arising from relevant simulations should obey, if answers to fundamental questions are sought. In the following subsections, we consider two important thermostating schemes that satisfy this criterion—the so-called Gaussian thermostats and the Nosé–Hoover thermostats.

### 3.2. Gaussian thermostats

Arguably the most elegant way of thermostating a particle system subjected to external driving is given by Gauss' principle of least constraint [26, 27]:

*Gauss' principle (1829).* Consider  $N$  point particles of mass  $m_i$ , subjected to frictionless bilateral constraints  $\Phi_i$  and to external forces  $\mathbf{F}_i$ . Among all motions allowed by the constraints, the *natural* one minimizes the *curvature*, defined by

$$C = \sum_{i=1}^N m_i \left( \dot{\mathbf{q}}_i - \frac{\mathbf{F}_i}{m_i} \right)^2 = \sum_{i=1}^N \frac{1}{m_i} \Phi_i^2.$$

Gauss' principle is equivalent to d'Alembert's principle, converting it into a minimal principle in the accelerations, based on the 'constraint'  $C$ . Gauss' principle suffers some important disadvantages when compared with the more commonly used extremal principles of variational mechanics (Hamilton's principle): it requires the calculation of accelerations, which are generally more complicated to evaluate numerically; and it is not independent of coordinate transformation and therefore, not as generally applicable as the Lagrangian and Hamiltonian formulations of mechanics. However, it has at least one significant advantage over these other approaches, relevant to the current topic—it applies equally for non-holonomic and holonomic constraints [26]. In the case of holonomic constraints, Gauss' principle is equivalent to the principle of least action, producing Hamiltonian equations of motion.

Evans, Hoover and collaborators [28] utilized Gauss' principle to develop a scheme for constraining the energy of a nonequilibrium molecular dynamics simulation. Such a constraint is non-holonomic, leading to non-Hamiltonian equations of motion. In particular, consider the isokinetic (IK) constraint [29] which fixes the kinetic energy of the system,  $K = \sum_i \mathbf{p}_i^2/2m$ , and the isoenergetic (IE) constraint [30] which fixes the internal energy  $H_0 = K + \Phi^{\text{int}}$  for the particle interaction potential  $\Phi^{\text{int}}$ . For an  $N$ -particle system subjected to an external field, these constraints yield the following dynamics for the  $i$ th particle:

$$\dot{\mathbf{q}}_i = \mathbf{p}_i/m \quad \dot{\mathbf{p}}_i = \mathbf{F}_i^{\text{int}}(\mathbf{q}) + \mathbf{F}_i^{\text{ext}}(\mathbf{q}) - \alpha(\Gamma)\mathbf{p}_i, \quad (8)$$

where  $\mathbf{F}_i^{\text{int}}(\mathbf{q})$  and  $\mathbf{F}_i^{\text{ext}}(\mathbf{q})$  denote inter-particle forces and external forces doing work on the system, respectively. The term  $-\alpha(\Gamma)\mathbf{p}_i$  (recalling that  $\Gamma = (\mathbf{q}, \mathbf{p})$ ) makes the dynamics dissipative (for non-zero external forces), allowing the system to reach a steady state in the long time limit. If  $\mathbf{F}_i^{\text{ext}}(\mathbf{q}) = C_i \mathbf{F}_e$ , where  $C_i$  is a coupling constant for the  $i$ th particle, and  $\mathbf{J} = \sum_{i=1}^N C_i \dot{\mathbf{q}}_i$ , then

$$\alpha_{\text{IK}}(\Gamma) = \frac{1}{2K} \left( \mathbf{J} \cdot \mathbf{F}_e + \sum_{i=1}^N \frac{\mathbf{p}_i}{m} \cdot \mathbf{F}_i^{\text{int}} \right) \quad \text{and} \quad (9)$$

$$\alpha_{\text{IE}}(\Gamma) = \frac{1}{2K} \mathbf{J} \cdot \mathbf{F}_e. \quad (10)$$

Importantly, we note that  $\alpha(\Gamma)$  contains terms involving the product of thermodynamic force and conjugate flux terms. Indeed, these terms provide the only non-zero contribution to the time-average  $\alpha$ , indicating that  $\alpha$  could be associated with the entropy production for the system. It is through this relationship that the thermostat ceases to be a mere simulation mechanism and becomes a potential component for nonequilibrium theories.

For equilibrium systems, the IE dynamics reduces to the usual Hamiltonian equations of motion, while the IK dynamics has zero mean entropy production. A coordinate transformation in (8) can be used to produce Hamiltonian versions for both IK [31] and IE [32] nonequilibrium dynamics.

IK and IE constraints are only two possible options in Gaussian thermostats, which have been extended to include a wide range of constraints, including isothermal-isobaric, isobaric-isoenthalpic and constant stress ensembles [15, see chapter 6]. In the IE systems of many interacting particles, the link between the phase space expansion rate, defined by

$$\Lambda \equiv \text{div}_\Gamma \dot{\Gamma} = \sum_{i=1}^N \left[ \frac{d}{d\mathbf{q}_i} \cdot \dot{\mathbf{q}}_i + \frac{d}{d\mathbf{p}_i} \cdot \dot{\mathbf{p}}_i \right], \quad (11)$$

and the entropy production rate persists, because  $\Lambda$  is proportional to  $\alpha_{\text{IE}}$ . However, it is also possible to develop Gaussian IE thermostats where the phase space expansion rate comprises additional terms beyond those related to entropy production, as has been demonstrated for the nonequilibrium Lorentz gas [33]. The opposite of the phase space expansion rate,  $-\Lambda$ , is often referred to as the phase space contraction rate.

### 3.3. Nosé–Hoover thermostat

Nosé and Hoover [34–36] devised a popular thermostat whose form is similar in nature to the Gaussian thermostat, but whose origins in theory are quite distinct. Consider a system of particles with internal energy  $H_0$  exchanging heat with a reservoir at temperature  $T$ . The interaction between particles and reservoir is represented by variables  $(s, p_s)$ , and the whole (Hamiltonian) system has energy

$$H = H_0 + 3NkT \ln s + \frac{p_s^2}{2Q} = \frac{\mathbf{p}^2}{2m} + \Psi(\mathbf{q}) + 3NkT \ln s + \frac{p_s^2}{2Q}. \quad (12)$$

For the variable set  $(\mathbf{q}, \mathbf{p}, s, p_s)$ , the system is Hamiltonian. The ingenuity of this approach is that, under the transformed variables

$$\tilde{\mathbf{q}} = \mathbf{q}, \quad \tilde{\mathbf{p}} = \frac{\mathbf{p}}{s}, \quad \tilde{t} = \int_0^t \frac{d\tau}{s}, \quad \tilde{s} = s, \quad \tilde{\xi} = \frac{p_s}{Q}, \quad (13)$$

the partition function for the new variables becomes

$$\begin{aligned} Z &= \int \frac{1}{s} \delta(H - E) \, d\mathbf{q} \, d\mathbf{p} \, ds \, dp_s \\ &= \left[ \int \frac{Q}{3NkT} e^{-\beta E} e^{-\beta Q \xi^2 / 2} \, d\xi \right] \int e^{-\beta H_0(\tilde{\mathbf{q}}, \tilde{\mathbf{p}})} \, d\tilde{\mathbf{q}} \, d\tilde{\mathbf{p}}, \end{aligned} \quad (14)$$

where  $\beta = 1/kT$ . The distribution is therefore *microcanonical* in the original frame  $(\mathbf{q}, \mathbf{p}, s, p_s)$ , but *canonical* with temperature  $T$  in the transformed frame  $(\tilde{\mathbf{q}}, \tilde{\mathbf{p}})$ . We can

therefore simulate a canonical system  $(\tilde{\mathbf{q}}, \tilde{\mathbf{p}})$  through the machinery of the Hamiltonian (super-)system  $(\mathbf{q}, \mathbf{p}, s, p_s)$ . Conveniently, the equations of motion in the new frame are quite straightforward:

$$\frac{d\tilde{\mathbf{q}}}{d\tilde{t}} = \frac{\tilde{\mathbf{p}}}{m}, \quad \frac{d\tilde{\mathbf{p}}}{d\tilde{t}} = \mathbf{F} - \xi \tilde{\mathbf{p}}, \quad \frac{d\xi}{d\tilde{t}} = \frac{1}{Q} \left( \frac{\tilde{\mathbf{p}}^2}{m} - gkT \right) \quad (15)$$

(with  $\tilde{s}$  redundant for simulation purposes).

The resemblance to the Gaussian equations is strong—however, the nature of the thermostat multipliers ( $\alpha$  for the Gaussian thermostat,  $\xi$  for the Nosé–Hoover) is subtly different. While  $\alpha$  is a phase *function* in the Gaussian case, here  $\xi$  is a phase *variable*, with its own equations of motion. The connection between phase space compression and entropy production remains the same as for the Gaussian thermostats, through the thermostat term  $\xi$ . To simulate nonequilibrium systems, the equations of motion are augmented with appropriate external fields, exactly as for the Gaussian case: under these conditions, the thermostat dissipates energy and produces entropy in an analogous fashion.

As with the Gaussian approach, Nosé–Hoover dynamics can be extended to other ensembles—Nosé’s original work [35] proposed both canonical and isothermal-isobaric ensembles. Furthermore, Nosé’s Hamiltonian is only one of a family of Hamiltonians which, coupled with an appropriate transformation to a new set of particle variables  $(\mathbf{q}, \mathbf{p}) \mapsto (\tilde{\mathbf{q}}, \tilde{\mathbf{p}})$ , can produce a canonical distribution [37]. In particular, while each of these couplings ensures that the average over the transformed variables is canonical, the manner in which a trajectory passes through the  $(\tilde{\mathbf{q}}, \tilde{\mathbf{p}})$  space depends strongly on the particular choice. For example, the canonical harmonic oscillator cannot be simulated using the usual Nosé–Hoover formulation, as the resulting dynamics are not ergodic [36]. However, by a judicious alternative choice of Hamiltonian  $H$ , together with a suitable alternative transformation of coordinates, the space  $(\tilde{\mathbf{q}}, \tilde{\mathbf{p}})$  can be ergodically spanned (for details, see [38, 39]). Finally, we note that the Nosé–Hoover formalism assumes the existence of a globally defined set of canonical variables: a condition satisfied for classical particle systems, but not for the classical limit of certain Lie algebras, such as the  $SU(2)$  algebra used to describe the dynamics of spin states. Bulgac and Kusnezov have extended the Nosé–Hoover approach to such systems as well [38], using their method for constructing classical canonical variables for any Lie algebra [40].

### 3.4. Configurational thermostats

All the early thermostats controlled the temperature by controlling the kinetic energy (and while Nosé’s thermostat was not designed with this method in mind, it is effectively the result). While we traditionally associate the kinetic energy with system temperature, statistical mechanics shows that there is in fact an infinite family of phase functions whose average at equilibrium is the system temperature—including functions of positions alone, and not momenta. Such an approach opens up the possibilities of a set of thermostats analogous to the Gaussian and Nosé–Hoover types, but that control different phase functions—the so-called configurational thermostats. In this section we briefly introduce the configurational temperatures, and then their application as thermostats.

In 1997, Rugh [41] derived a dynamical expression for the temperature of a microcanonical ensemble from first principles, by considering the rate of change in area of a surface of constant energy in the phase space, as the energy is changed. The expression resulting from Rugh’s elegant derivation

$$\frac{1}{kT} = \left\langle \nabla \cdot \frac{\nabla H}{\|\nabla H\|^2} \right\rangle \quad (16)$$

gives a (perhaps surprisingly) complicated expression for the temperature. The unusual combination of both kinetic and configurational contributions in the temperature expression led early investigators to a number of realizations. Isotropy and dimensionality arguments indicate that the contributions from either the configurational or kinetic terms *alone* should yield the temperature (to  $O(1/N)$ ). Furthermore, equipartition-style results also indicate that contributions from any individual phase variable should also yield the temperature. Generalized temperature expressions of the form

$$kT = \frac{\nabla H \cdot \mathbf{B}}{\nabla \cdot \mathbf{B}} \quad (17)$$

were developed for phase space vector fields  $\mathbf{B}$  matching particular criteria [42–44], of which the usual equipartition result can be seen as a special case ( $\mathbf{B} = (\mathbf{0}, \mathbf{p})$ ). The configurational temperature  $T_C$ , obtained from  $\mathbf{B} = (\mathbf{F}, \mathbf{0})$ , consists of only configurational contributions, and is thus a way of determining the temperature without explicit reference to the kinetics of the particles. Indeed, the first confirmation of this expression [45] involved comparison of the computed configurational temperature with the input temperature in a Monte Carlo simulation (which has no kinetic degrees of freedom).

Configurational temperatures thus provide an alternative phase function for thermostating a system. Gauss' principle (or indeed Hamilton's) is of little help in constraining the configurational temperature, which constitutes a *holonomic* constraint: this approach would constrain both the energy and the configurational temperature, which is not our objective. The first configurational thermostats [46, 47] considered equations of motion of the form

$$\frac{d\mathbf{q}}{dt} = \frac{\mathbf{p}}{m} - \xi \nabla_{\mathbf{q}} T_C(\mathbf{q}), \quad \frac{d\mathbf{p}}{dt} = \mathbf{F}. \quad (18)$$

As with the kinetic thermostats, one can apply 'hard' thermostats that keep the configurational temperature  $T_C(\mathbf{q})$  constant, or 'soft' thermostats that control their average value. However, various problems arise from the higher-order potential-energy derivatives required to calculate  $\nabla_{\mathbf{q}} T_C(\mathbf{q})$ : they are computationally expensive; the 'hard' thermostat requires impractically low timesteps; the 'soft' mechanism is not time-reversible; and holonomic constraints such as bond lengths greatly complicates the calculations.

A much more attractive alternative, inspired by the Nosé–Hoover approach, is to consider a thermostat of the form [48]

$$\frac{d\mathbf{q}}{dt} = \frac{\mathbf{p}}{m} - \xi \mathbf{F}, \quad \frac{d\mathbf{p}}{dt} = \mathbf{F}, \quad \frac{d\xi}{dt} = \frac{1}{Q} (\mathbf{F} \cdot \mathbf{F} - kT \nabla \cdot \mathbf{F}). \quad (19)$$

In the phase space  $(\mathbf{q}, \mathbf{p}, \xi)$ , the Liouville theorem can be used to show that the canonical distribution is preserved by this dynamics. This thermostat was first fully expounded and subsequently developed by Braga and Travis [48–51] (independently of an early version appearing in [52]; see also [53]). We note that the existence of the average value of  $\xi$  implies that

$$\overline{\xi} = 0 = \frac{1}{Q} (\overline{\mathbf{F} \cdot \mathbf{F}} - kT \overline{\nabla \cdot \mathbf{F}}) \Rightarrow kT = \frac{\overline{\mathbf{F} \cdot \mathbf{F}}}{\overline{\nabla \cdot \mathbf{F}}}; \quad (20)$$

in other words, the system temperature is identically the configurational temperature. A further advantage of this approach is that it naturally preserves the holonomic constraints (such as bond lengths) that would be preserved by the forces.

The configurational temperatures raise important issues regarding temperature and thermostating in nonequilibrium systems. They remind us that the association of temperature with kinetic energy, away from equilibrium, is somewhat arbitrary. The fact that these temperatures do not agree away from equilibrium [47] highlights the need for an *operational*

concept of nonequilibrium temperature that is appropriate for molecular simulations. Interestingly, if one were to define such an operational temperature as a Fourier-law potential to describe the heat flux, one set of early studies [54] indicated that Rugh's temperature (16) provides the best representation. One possible approach is through the use of temperatures based on degrees of freedom orthogonal to external driving fields, where they agree in the large- $N$  limit [55].

The configurational temperatures also remind us that the choice for thermostating vector  $B$  is somewhat arbitrary, to the extent that the temperature (17) is well defined. One can, for example, choose thermostat terms of the form  $-\alpha \mathbf{p}^n$ , for odd  $n$ , and successfully thermostat the system. Hoover *et al* argue that for physical reasons, the usual form  $-\alpha \mathbf{p}$  is to be preferred over higher-order alternatives, and that this usual form is more reliable for nonequilibrium conditions [56]. Bright and Evans similarly argue that the IK thermostat is the most natural of Gaussian kinetic thermostats [57].

Configurational thermostats have proven useful in identifying simulation artefacts induced by kinetic thermostats such as string phases and antisymmetric stress [58–61]. The use of relative particle positions in configurational thermostats provides a means of circumventing this problem: unlike the kinetic thermostats, their configurational counterparts are not affected by local streaming [62], although it has been recently suggested that an analogous problem may affect the rotational modes [63]. As a relatively recent development, however, the theory of configurational thermostats in nonequilibrium simulations has yet to be developed to the same level as that of the kinetic thermostats.

### 3.5. Other thermostating methods

Here we note some alternative thermostating techniques that rely on quite different methods to the aforementioned. Although devised through quite diverse approaches to thermostating, each offers its own particular insight.

**3.5.1. Dirac brackets.** Dirac's method for constraining Hamiltonian systems [64, 65], well known in quantum field theory, has also been applied in equilibrium MD simulations [66], although it does not appear to be successfully used in NEMD. As with the classical Hamiltonian systems, the formalism is not directly suited to non-holonomic constraints, but rather to paired constraints associated with a holonomic constraint and its time derivative (thus reducing the set of variables by one coordinate and one conjugate momentum). This symplectic structure of Dirac's method does not combine naturally with the NEMD methodology that requires a single non-holonomic constraint on the kinetic energy [67].

**3.5.2. Müller–Plathe approach.** Most NEMD techniques require an explicit thermostat to dissipate the energy introduced by the nonequilibrium field. The scheme introduced by Müller–Plathe, used to calculate shear viscosities [20, 68, 69], does not introduce an external field. Rather, it imposes a flux through molecular momentum exchange that induces the relevant field shear rate, from which the viscosity transport coefficient can be determined. This approach builds on an analogous technique used to calculate thermal conductivities and Soret transport coefficients [20].

The periodic shear profile can be divided into layers (of constant  $x$ , say) parallel to the direction of flow—for convenience we describe the layer with the lowest  $x$  streaming velocity as the slowest layer, and that with the highest  $x$  streaming velocity as the fastest layer. The algorithm exchanges the slowest molecules in the fastest layer with the fastest molecules in the slowest layer, conserving the total energy and momentum (although not angular momentum).

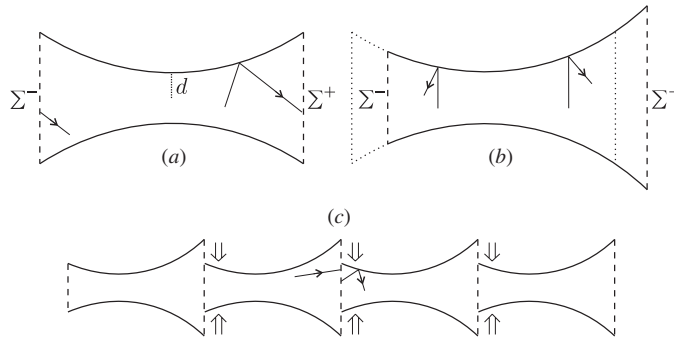


Figure 1. Pump-billiard equilibrium cell (a); nonequilibrium cell (b); chain of cells (c).

However, the temperature is measured with respect to the local streaming profile, meaning that this form of exchange will *reduce* the local temperature, countering the viscous heating (entropy production) introduced by momentum exchange due to the induced shear rate. This technique thus implicitly controls viscous heating through the momentum exchange algorithm itself.

The Müller–Plathe example raises the notion of equivalence among nonequilibrium ensembles. This concept is familiar in equilibrium statistical mechanics, where thermodynamic quantities are divided into conjugate pairs, and the ensembles defined in terms of which quantity for a given pair is held constant, and which allowed to vary. In the NEMD case, the thermodynamic force  $F$  is held constant while the conjugate flux  $J$  varies: for the Müller–Plathe algorithm,  $J$  is controlled while  $F$  varies. The success of both techniques, in terms of the calculation of transport coefficients, indicates that this concept should be extensible beyond the equilibrium realm. We explore this idea further in section 4.3.

3.5.3. *Pump-billiard thermostat.* Household heating and cooling devices often operate through a compression mechanism. As a final example, we note a new thermostat that is inspired by this form of temperature control—the so-called pump-billiard thermostat [70] that models a (rarefied) gas in a pump.

The system consists of a periodic chain of billiard cells as in figure 1. The equilibrium periodic cell is delimited by two equal vertical lines,  $\Sigma^-$  and  $\Sigma^+$ , and by two arcs of radius  $r$ , amplitude  $\phi = 2 \arcsin r^{-1}$  and minimum distance  $2d$  (cf figure 1(b)). Particles collide elastically with the arcs, or exit the cell through  $\Sigma^\pm$ . If a particle hits  $\Sigma^-$ , it re-enters from  $\Sigma^+$  and vice versa. At equilibrium, there is no transport of particles.

If the periodic cell is asymmetric (as in figure 1(b)), so that the ratio of the sizes of  $\Sigma^+$  and  $\Sigma^-$ ,  $\alpha = |\Sigma^+|/|\Sigma^-| > 1$ , then the system is no longer in equilibrium, and a net current is produced. The current then depends on the reinjection rule (figure 1(c)). The law

$$x' = x \mp \ell, \quad y' = \alpha^{\mp 1} y, \quad v'_x = c v_x, \quad v'_y = c \beta^{\pm 1} v_y, \quad c^2 = \frac{v_x^2 + v_y^2}{v_x^2 + \beta^{\pm 2} v_y^2} \tag{21}$$

represents in an idealized way the effect of a compressing mechanism and of a ‘thermostat’ conserving the kinetic energy [70]:  $\ell$  is the length of the cell, while  $\beta$  is a second system parameter. Analogous laws can be found in [71]. To ensure dissipation for the forward motion, as is expected of thermodynamic systems, it suffices to take  $\beta < \alpha$ . For  $\beta\alpha > 1$ , one obtains an average current  $\langle v_x \rangle$  proportional to  $\varepsilon = \log \alpha$ , the ‘driving field’, for small

$\varepsilon$ , while this proportionality is lost at larger values of  $\varepsilon$ . In this manner, a ‘toy-model’ of the thermodynamic linear regime is recovered.

#### 4. Physical suitability

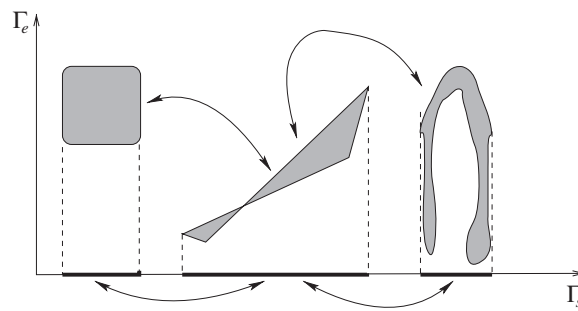
As described in the previous section, thermostats are mathematical tools, introduced to conveniently simulate thermodynamic phenomena from a microscopic, dynamical standpoint. For a particular system of interest, it is simply unfeasible to simulate the microscopic boundary interactions responsible for its thermodynamic state. This is as true of the exchange of energy through walls to maintain thermal equilibrium as it is of the transport processes that maintain a temperature gradient between two reservoirs. We know from experiment how such systems behave macroscopically—an initial transience that depends on initial conditions leads to a final state that depends only on the boundary conditions. For equilibrium systems, we avoid this problem by using the statistical mechanical theory of ensembles: either to invoke the equivalence of ensembles in the thermodynamic limit or in using thermostats to generate the appropriate ensemble distribution.

For nonequilibrium transport processes, one looks for methods to account for the coupling with the reservoirs, without actually simulating their microscopic dynamics. The methodology of NEMD, accomplished through the introduction of artificial forces and an accompanying thermostat, has been developed over the past decades. NEMD has proven to be a highly successful technique for modelling nonequilibrium thermodynamic phenomena including various rheological properties as well as other phenomena including fracturing, see e.g. [15, 72–75]. As with all models of the physical world, some aspects of the NEMD model represent real phenomena, while others are artefacts that do not reflect the behaviour of real systems. Examined under greater scrutiny, the extent to which NEMD model properties can be thought to describe real-world phenomena has been greatly clarified, and is the focus of this section. In particular, we will consider the use of non-Hamiltonian equations, the description of linear response, notions of equivalence of nonequilibrium ensembles, the replacement of thermodynamic boundary conditions, the use of deterministic equations and the form of the thermostat.

##### 4.1. Non-Hamiltonian equations

Given that the simplest and most common mechanical systems considered by physicists are naturally described using Hamiltonian equations of motion, one might object to the use of thermostatted equations of motion for the construction of a theory of nonequilibrium phenomena, on the basis that they contain artificial forces that lead to non-Hamiltonian dynamics.

There is strong justification, however, for the use of non-Hamiltonian dynamics to represent nonequilibrium dynamics [67, 76, 77]. A system that exchanges energy with its environment should not be modelled by Hamiltonian equations of motion if that environment is not included in the description [78]. Hamiltonian dynamics is characterized by conservation of phase space volumes—but the volume restricted to arbitrary subspaces is not preserved. As the Nosé–Hoover model demonstrates, the subsystem of interest in a universe is not bound to preserve its volume (so cannot be Hamiltonian), even if the universe to which it belongs is Hamiltonian. These ideas are schematically represented in figure 2 for a two-dimensional phase space with Hamiltonian dynamics, where the subsystem dynamics corresponds to the projection on the horizontal axis. While the subsystem dynamics retains the time reversal invariance of the whole system, it does not retain the preservation of volume.



**Figure 2.** Evolution of a phase space volume in the full  $(\Gamma_s, \Gamma_e)$ -space, and in the projected  $\Gamma_s$ -space. In the  $(\Gamma_s, \Gamma_e)$ -space, volumes are preserved, while in the  $\Gamma_s$ -space they fluctuate. On the other hand, if the backward evolution occurs in the full phase space, it also occurs in the reduced space.

If the system in question is a nonequilibrium system induced by some external driving, the environment will generate entropy while dissipating energy in reaching a steady state. Given that the entropy in statistical mechanics is intimately connected with the phase space volume, Hamiltonian dynamics would necessarily imply a preservation of entropy that seems fundamentally at odds with the notion of nonequilibrium. It is therefore quite reasonable to represent such dynamics using NEMD models that are time reversal invariant but that do not preserve phase space volume—and therefore could not be Hamiltonian in nature. The fact that they are not obtained through the ideal projection procedure described above, however, implies that they must be used *cum grano salis*.

The use of *artificial* forces is not a strong criticism of the technique, given that the equivalence with thermodynamic driving forces has been demonstrated [15]. Less certain is the relation between the microscopic dynamics of a real nonequilibrium steady state and the dynamics defined by NEMD. A key difficulty in this regard is the formal representation of specific thermodynamic concepts in nonequilibrium dynamical models. While the connections between temperature and kinetic energy, and between entropy and phase-space volume, are well understood for equilibrium systems, these relationships are less clear for nonequilibrium steady states. The lack of a comprehensive theory of nonequilibrium steady states makes it difficult to assess the general limit of applicability of NEMD techniques, and the interplay between the artificial forces and thermostats. However, the success of NEMD in reproducing nonequilibrium steady-state phenomena suggests that it may provide a sounding board for successful nonequilibrium theories.

#### 4.2. Linear and nonlinear response theory

As we noted earlier, we do not expect molecular simulation to produce a perfect replication of real-world microscopic behaviour—rather, we must assess its success against the simulation objectives. For NEMD, the goal is to measure nonequilibrium transport behaviour. Apart from numerical results demonstrating the success of the technique [15, 79–81], there is also a comprehensive body of theory supporting its use [15].

NEMD has primarily been used, and theories developed, for the so-called *linear regime*, where the response (induced flux) to an applied field varies linearly with that field, with the coefficient of linearity given by the transport coefficient. The linear response can be associated with equilibrium fluctuations (for example, through the Onsager regression hypothesis as a consequence of the fluctuation–dissipation relations), implying that transport coefficients



measured in the linear regime of a non-equilibrium system should be equal to those obtained from equilibrium calculations. The success of the NEMD method is well established for its capacity to reproduce equilibrium transport coefficients—often at greater computational efficiency.

Beyond numerical evidence, a rigorous theory has been developed to demonstrate the equivalence of the linear response to the equilibrium fluctuations. This linear response theory determines the response of an arbitrary phase function (usually denoted as  $B$ ) to the nonequilibrium dynamics imposed by NEMD—dynamics that include artificial external forces to generate the targeted nonequilibrium flux, as well as an appropriate thermostat (or equivalent mechanism, such as a barostat) to generate the targeted nonequilibrium steady state. It has then be shown, for a variety of dynamical equations (including Gaussian and Nosé–Hoover thermostating techniques), that the response of a system to a real thermodynamic gradient is identical to the response to the artificial field—we will consider this further in section 4.3.

As an example, the Evans–Gillan equations of motion [82, 83]

$$\begin{aligned} \dot{\mathbf{q}}_i &= \mathbf{p}_i/m \\ \dot{\mathbf{p}}_i &= \mathbf{F}_i(\mathbf{q}) + (E_i - \bar{E})\mathbf{F}_e(t) - \frac{1}{2} \sum_{j=1}^N \mathbf{F}_{ij}(\mathbf{q}_{ij} \cdot \mathbf{F}_e(t)) + \frac{1}{2N} \sum_{j,k}^N \mathbf{F}_{jk}(\mathbf{q}_{jk} \cdot \mathbf{F}_e(t)) \end{aligned} \quad (22)$$

(where  $E_i$  is the energy of particle  $i$  and  $\bar{E} = \sum_i^N E_i/N$ ) provide the most efficient presently known model for the heat flow in systems close to equilibrium. The external field  $\mathbf{F}_e$  is artificial, but induces a response that precisely mimics the effect of a real temperature gradient, transforming an inhomogeneous thermodynamic boundary condition that is very difficult to simulate into a homogeneous mechanical term in the equations of motion. The term  $(E_i - \bar{E})\mathbf{F}_e(t)$  pushes particles with above-average energy in the direction of  $\mathbf{F}_e(t)$  and pushes particles with below-average energy in the opposite direction. Consequently, this field induces an energy flow in the absence of temperature gradients and mass flows, efficiently simulating heat flow under a temperature gradient, provided the number of particles and the collisions among them suffice for LTE to be reasonably well approximated<sup>7</sup>.

A rigorous nonlinear response theory has also been developed by Evans, based on the pioneering work of Kawasaki [84–86]. For example, Kawasaki showed that the boundary condition that corresponds to planar Couette shear flow can be incorporated exactly into the SLLOD<sup>8</sup> equations of motion [84–86]. The result is that they give an exact description of the shearing motion of systems arbitrarily far from equilibrium (while, indeed, no Hamiltonian has been found for these equations of motion). A significant application of this work is in the study of transient (nonequilibrium) systems away from the steady state, which lies beyond the scope of this review.

#### 4.3. Equivalence of nonequilibrium ensembles

We have already touched on the idea that one may extend the equilibrium ensembles to the nonequilibrium steady states. In this section we explore the concepts of nonequilibrium ensembles, and of equivalence among them.

<sup>7</sup> For this to be the case, the artificial forces acting on the system must not dominate the other forces present, so that their impact on the system behaviour is as limited as possible beyond their purpose. This problem is also faced when deterministic (artificial) thermostating terms are added to the equations of motion to simulate the effect of real thermostats placed at the boundaries of the systems of interest, cf subsection 4.5 below.

<sup>8</sup> The name SLLOD is not an acronym. It denotes a modification (form of transpose) of the DOLLS algorithm, an NEMD algorithm developed (unsuccessfully) to compute the pressure tensor of Couette flow [168].

Analogously to the pairs of conjugate thermodynamic variables generating conjugate equilibrium ensembles, one may use NEMD to generate nonequilibrium ensembles. The pairs of thermodynamically conjugate variables for equilibrium systems are determined from the entropy contributions to the first law of thermodynamics: this idea is neatly extensible to the nonequilibrium case. For a thermodynamic force  $X$  and flux  $J$ , one could consider the  $(N, V, T, X)$  and the  $(N, p, T, X)$  ensembles as conjugate (for conjugate variables  $p$  and  $V$ ): one could also consider  $(N, V, T, X)$  and  $(N, V, T, J)$  as conjugate ensembles. In the case of shear stress, these two ensembles correspond to the SLLOD ensemble and the Müller–Plathe ensemble, respectively: in the first case, the field is imposed and the flux is the independent state variable, while this is reversed in the second case. In 1985, Evans and Morriss postulated a principle of equivalence of nonequilibrium ensembles, analogous to that concerning equilibrium ensembles [15]. The principle states that linear responses of conjugate nonequilibrium ensembles are equivalent. This principle, later generalized also to nonlinear response, may provide a more formal basis for analysing the suitability and range of applicability of the deterministic thermostatted dynamics, beyond the use of transport coefficients.

Given that the earliest thermostats were computational devices designed to facilitate EMD (and eventually NEMD) simulations at constant temperature, the first investigations into the equivalence of nonequilibrium ensembles considered equivalence with respect to transport coefficients measured for bulk fluids. Equivalence between the time correlation functions (required to calculate equilibrium transport coefficients) for Newtonian, Gaussian thermostatted and Nosé–Hoover thermostatted systems in the thermodynamic limit was demonstrated, and subsequently the equivalent response to the field in the linear regime for the same set of systems [15, 79, 80]. Later on, the equivalence of these thermostats under nonlinear response was also demonstrated, but under conditions of mixing as well as in the thermodynamic limit [18, 81].

The first result of this sort was related by Evans and Morriss to the equivalence of the Norton constant current (flux) electrical circuit and the corresponding Thévenin constant voltage (force) circuit. The voltage and the current being conjugate variables suggests conjugate ensembles  $(N, V, T, J)$  and  $(N, V, T, E)$  as equivalent. This is the content of the theory developed in [79–81]. Essentially, two Gaussian multipliers are introduced, to hold the temperature and the current constant. From the resultant equations of motion, an expression for the required applied field is obtained, whose fluctuations are consistent with the flux fluctuations of the constant force ensemble (at zero frequency, the ‘Thévenin’ fluctuation correlation function is proportional to the conductivity, while the ‘Norton’ fluctuation correlation function is proportional to its inverse, the resistivity). It is clear, then, that this result does not imply equivalence in every respect (just as it does not for the equilibrium ensembles). Interestingly, the zero-frequency condition appears in some sense to be the nonequilibrium counterpart to the thermodynamic-limit condition—the zero-frequency result is representative of the system behaviour when the non-zero-frequency contributions are less significant, which becomes the case as the number of degrees of freedom goes to infinity.

The theory formulated in [79–81] has been recently reconsidered and understood in rigorous mathematical terms by Gallavotti and Presutti, who address the problem of equivalence of ensembles for systems in contact with different kinds of thermostats, placed at their boundaries [19, 87]. In particular, the following main result is stated in [87].

*In the thermodynamic limit, the thermostatted evolution, within any prefixed time interval  $[0, \Theta]$ , becomes identical to the frictionless evolution at least on a set of configurations which have  $\mu_0$ -probability 1 with respect to the initial distribution  $\mu_0$ , in spite of the*

*non-stationarity of the latter. In the same limit, also the frictionless evolutions with open or elastic regularization become identical.*

This result holds under the technical assumption of ‘local dynamics’, proven for the systems of [87], which implies that in a finite time the external forces can only perform a finite amount of work. Here, thermostatted evolution means Gaussian IE, while frictionless refers to Hamiltonian heat baths with infinitely many degrees of freedom. Open and elastic regularizations represent two different conditions on the process of taking the thermodynamic limit.

A feature common to these theories and to the equilibrium theories is the need for large  $N$ . In equilibrium statistical mechanics, large  $N$  is an essential ingredient not only for the equivalence of ensembles but also for legitimate representation of thermodynamic concepts such as entropy and temperature within the statistical mechanical framework. The large- $N$  limit plays a similar role for nonequilibrium theories as well.

As an example, we note that the phase space expansion rate of thermostatted systems cannot *in general* be thought of as proportional to the entropy production rate of the phenomenon under consideration [88, 89]. Consider Gaussian IE and IK thermostatted dynamics, whose phase space expansion rate,  $\Lambda$ , is simply related to the term  $\alpha$  of equation (8). In this case,  $\alpha$  is only a Lagrange multiplier, introduced in order to implement Gauss’ principle of least constraint, and there is no reason, as discussed above, to believe that it correctly represents a real heat reservoir or any other physical observable. Similarly,  $\Lambda$  does not necessarily have any physical meaning, apart from the fact that, in  $d$  dimensions, its time average obeys

$$\bar{\Lambda} = \sum_{i=1}^{2dN} \lambda_i, \tag{23}$$

where  $\{\lambda_i\}_{i=1}^{2dN}$  are the Lyapunov exponents of the system.

In [88], the values of  $\bar{\Lambda}$  for the IK and IE systems were compared. Equation (9) yields

$$\begin{aligned} \bar{\Lambda}_{\text{IK}} = \langle \Lambda_{\text{IK}} \rangle &= -(dN - 1) \left[ \left\langle \frac{\sum_{i=1}^N \frac{\mathbf{p}_i}{m} \cdot \mathbf{F}_i^{\text{int}}}{\sum_{i=1}^N \frac{\mathbf{p}_i^2}{m}} \right\rangle + \left\langle \frac{\sum_{i=1}^N \frac{\mathbf{p}_i}{m} \cdot \mathbf{F}^{\text{ext}}}{\sum_{i=1}^N \frac{\mathbf{p}_i^2}{m}} \right\rangle \right] \\ &= -\frac{1}{m} \frac{\langle \sum_{i=1}^N \mathbf{p}_i \cdot \mathbf{F}^{\text{ext}} \rangle}{k_B T} \end{aligned} \tag{24}$$

for an IK ergodic system, where the brackets  $\langle \cdot \rangle$  represent the average with respect to the steady-state probability distribution, and where the inclusion of the temperature (with the usual equipartition definition) implies near-equilibrium conditions. Introducing the particle current density,  $\mathbf{I} = \langle \sum_{i=1}^N \frac{\mathbf{p}_i}{m} \rangle / V$ , and dividing  $\Lambda_{\text{IK}}$  by the volume  $V$  of the system, to compare dynamical averages with macroscopic quantities, one obtains

$$-\hat{\Lambda}_{\text{IK}} = \frac{\mathbf{I} \cdot \mathbf{F}^{\text{ext}}}{k_B T}, \tag{25}$$

for the phase space expansion rate, where the right-hand side of equation (25) is *formally* the expression for the entropy production rate in irreversible thermodynamics. Thus, a formal connection between  $\Lambda$  and a physical observable has been found.

While equation (25) may not appear to depend on large  $N$ , in reality there is no concept of a thermodynamic quantity away from this limit. For small  $N$  we can only speak of a connection between  $\Lambda$ , the kinetic energy and the dissipated energy in *statistical* rather than *thermodynamic* terms (cf section 5.4.4 in [77]). Indeed, irreversible thermodynamics is based

on local thermodynamic equilibrium (LTE), in which the extensive properties are proportional to  $N$  and depend further only on the temperature and on the number density  $n = N/V$ . Without *many interacting* particles, that dependence is impossible. Finally, we note that relaxation times for the values of observables turn out to be unphysically long for systems consisting of a small number of particles.

What about the isoenergetic case, which imposes no constraint on the kinetic energy? The temperature can again be defined by

$$\left\langle \sum_{i=1}^N \frac{\mathbf{p}_i^2}{m} \right\rangle = dNk_B T, \quad (26)$$

while equation (10) yields

$$\hat{\Lambda}_{\text{IE}} \equiv \frac{\langle \Lambda_{\text{IE}} \rangle}{V} = -(dN - 1) \left\langle \frac{\sum_{i=1}^N \frac{\mathbf{p}_i}{mV} \cdot \mathbf{F}^{\text{ext}}}{\sum_{i=1}^N \frac{\mathbf{p}_i^2}{m}} \right\rangle. \quad (27)$$

As is common in equilibrium statistical mechanics, where  $N$  is very large and fluctuations are negligible with respect to the mean values, one may argue that the average of the ratio on the right-hand side of (27) can be replaced by the ratio of the averages. Indeed, in order to speak sensibly of temperatures or entropies, the conditions of LTE must be met, which imply large  $N$ . We can thus neglect terms of order  $O(1/N)$  to obtain

$$-\hat{\Lambda}_{\text{IE}} = \frac{\mathbf{I} \cdot \mathbf{F}^{\text{ext}}}{k_B T}, \quad (28)$$

which shows that entropy production and phase space expansion rates may be simply related only for large  $N$ , even in the IE case. Away from large  $N$ , in all cases, one may only speak of energy dissipation.

Thus, only for large  $N$  and small drivings (so that the equipartition temperature is a reasonable approximation) may the average phase space expansion rates of the IK and IE systems be considered to be related to the entropy production rates of a given physical system, and hence related to each other. Nevertheless, the above calculations lead to  $\hat{\Lambda}_{\text{IE}} = \hat{\Lambda}_{\text{IK}} + O(1/N)$  which constitutes one example of dynamical (if not thermodynamical) equivalence of nonequilibrium models, analogous to that of equilibrium ensembles but not restricted to near-equilibrium states. The result may be stated as follows.

*In the thermodynamic limit, i.e.  $N, V \rightarrow \infty$ , with density and the energy density converging to a common finite value, the mean phase space expansion rates of IK and IE dynamics coincide.*

Like the equivalence of Norton and Thevénin ensembles, this provides one explanation of why different equations of motion, reaching steady states which differ from a microscopic point of view, may nevertheless describe the same physical situation, at least in part, in the thermodynamic limit. The issues considered in this subsection contribute to our understanding of the fact that various aspects of real physical phenomena may be correctly described by microscopic evolution laws that may appear artificial.

#### 4.4. Equivalence versus non-equivalence

An equivalence principle has been proposed by Gallavotti in the context of turbulence in [90], where a modified version of the forced Navier–Stokes equation, for a Newtonian incompressible fluid, was introduced. Consider the equation

$$\dot{\mathbf{u}} + (\mathbf{u} \cdot \nabla)\mathbf{u} = -\frac{1}{\rho}\nabla p + \mathbf{g} + \nu\Delta\mathbf{u}, \quad \nabla \cdot \mathbf{u} = 0. \quad (29)$$

Here,  $\mathbf{u}$  is the velocity field,  $\rho$  is the fluid density,  $p$  the pressure,  $\mathbf{g}$  is a constant forcing term and  $\nu$  can either be the constant viscosity, in which case (29) is the Navier–Stokes equation, or can be defined by

$$\nu(\mathbf{u}, \omega, \mathbf{f}) = \frac{\int [\omega \cdot \mathbf{f} + \omega \cdot (\omega \cdot \nabla) \mathbf{u}] \, d\mathbf{x}}{\int (\nabla \times \omega)^2 \, d\mathbf{x}}, \quad (30)$$

with  $\omega = \nabla \times \mathbf{u}$  and  $\mathbf{f} = \nabla \times \mathbf{g}$ . In this case (29) is called the Gauss–Navier–Stokes equation, is time-reversible and the enstrophy  $Q = \int \omega^2 \, d\mathbf{x}$  is a constant of motion. The Navier–Stokes and Gauss–Navier–Stokes equations thus represent the dynamics of potentially conjugate ensembles, depending on which of the pair  $(\nu, Q)$  is a constant of the motion.

With periodic boundary conditions, one can expand  $\mathbf{u}$  in Fourier modes, obtaining an infinite system of ODEs from (29). Truncating the expansion of  $\mathbf{u}$  at a finite number of modes yields a dynamical system in a finite-dimensional phase space, characterized by a given phase space expansion rate. The equivalence principle then states [90] the following.

*Equivalence principle.* The stationary probability distributions of the Navier–Stokes and of the Gauss–Navier–Stokes equations are *equivalent* in the limit of large Reynolds number, provided the enstrophy  $Q$  and the viscosity  $\nu$  are so related that the constant phase space contraction rate of the Navier–Stokes equation and the average of the fluctuating phase space contraction rate of the Gauss–Navier–Stokes equation are equal.

As with the equivalence of equilibrium ensembles, this principle is intended to hold for local variables, while the large Reynolds number is invoked for the fluctuations of  $\beta$  to be fast on the observation time scales. In that case, if the average of  $\beta$  equals  $\nu$ , the macroscopic behaviour of the Navier–Stokes equation and of the Gauss–Navier–Stokes equation should be the same. The idea has been later generalized in [91] and verified to a certain extent in [91, 92] for an infinite hierarchy of hydrodynamic equations and of relevant observables, under low resolution approximations. However, further verifications at higher resolutions have not been possible thus far [76]. This is probably due to the greater complications encountered in solving partial differential equations, as opposed to ordinary differential equations, but it also indicates that the equivalence of nonequilibrium ensembles needs to be further investigated.

Indeed, various instances of non-equivalence of both equilibrium and nonequilibrium ensembles have been pointed out in various papers [93–96]. The non-equivalence of canonical and microcanonical ensembles is often found in systems with an ‘anti-thermodynamic’ nature—such as systems with negative specific heats, long range interactions or non-interacting particles under an external field—whose consequences for thermodynamics, such as a non-extensive ‘entropy’, are not fully understood [93–95]. Analogously, non-equivalence of nonequilibrium ensembles has been associated with a form of long-range interaction, as observed for a model of electric conduction whose electrons interact with each other via a Gaussian IK thermostat [96]. Interaction with the ions of a crystal are further modelled by elastic collisions with hard scatterers. The authors of [96] conclude generating nonequilibrium steady states though the use of deterministic thermostats requires some caution, at least when there is no interaction between the particles other than that induced by the thermostat. A similar conclusion can be reached observing that systems of particles which do not interact with each other, and are subjected to external fields and thermostats, meet difficulties when used to represent thermodynamic features, such as the validity of linear response [97–100]. Similarly, long-range interactions, especially if attractive, appear to generate macroscopic behaviours at variance with the thermodynamic ones. This may be the reason for the difficulty in observing the equivalence of the turbulence models discussed above, since the thermostating

mechanisms couple all their modes in a similar manner to the cases of (for example) [96, 97]. But this matter needs to be clarified.

In general, the necessary conditions for ensemble equivalence appear closely related to the conditions for establishing LTE, which itself is a prerequisite for thermodynamic behaviour. The equivalence and non-equivalence of nonequilibrium ensembles certainly merits more thorough investigation, especially in relation to the presence or absence of LTE.

#### 4.5. Deterministic bulk and boundary thermostats

The thermostats we have considered are all deterministic, reversible *bulk* thermostats: the thermostat is coupled to each particle in the bulk of the system. This property runs at odds with our everyday experience of temperature control, whereby a particle's temperature (however we might define it) would be influenced primarily by its neighbours. Instead, the bulk thermostat acts universally while not being part of the same universe as the particles. This can lead to strange phenomena, for example, when interactions between particles are discrete rather than continuous [101, 102]. While response theory demonstrates that these techniques generate correct transport coefficients, this does not address the issue of *why* one would choose such an approach, over other techniques with more overt connections to physical reality.

As an example, consider the study of mass transport. The mechanically driven system uses the colour field technique—one could alternatively implement a thermodynamically driven nonequilibrium simulation by maintaining a concentration profile along the length of the simulation cell. This can be achieved via the dual control volume method [103], whereby a Hamiltonian MD cell is bounded by two cells maintained at different densities via MC. Typically, the advantage of the mechanical force approach is one of efficiency, which is equally true of microporous systems as it is of bulk systems (e.g. [16]). The difficulty and computational cost in maintaining the thermodynamic boundary condition is a clear disadvantage.

To implement more realistic thermodynamic boundary conditions requires yet further computational cost. As alluded to in the introduction, the way in which a system of interest is maintained in a steady state can involve an enormous number of degrees of freedom. We know that these interactions will vary from system to system: yet the irreversible thermodynamics we are interested in appears largely independent of the microscopic details of such interactions. We thus arrive at the idea that the majority of thermodynamic phenomena taking place in the system of interest are not influenced by the details of the coupling with its environment. For instance, if the system of interest is sufficiently large itself, then most of it is physically far removed from the reservoirs, and interacts with them only indirectly, via its boundary regions. Rieder, Lebowitz and Lieb put it in these terms: 'the properties of a 'long' metal bar should not depend on whether its ends are in contact with water or with wine 'heat reservoirs' at temperature  $T_1$  and  $T_2$ ' [104]. This suggests that different forms of thermostats may lead to the same thermodynamic phenomena within the system of interest. It thus seems legitimate to choose the most convenient description of the effect of the reservoirs on the system, although the question of which properties of the system of interest may be properly described by which models of the reservoirs remains to be satisfactorily addressed.

A key point to note here, however, is that the theory of equivalence of thermostatted responses, to which we refer, is based on the assumption that a condition similar to LTE is satisfied. For instance, [18] states that the equivalence of steady-state averages and time correlation functions under either Gaussian IE, Gaussian IK and Nosé–Hoover thermostats holds even in the far-from-equilibrium nonlinear regime, as long as the system is mixing and the quantities involved are local and not trivially related to constants of the motion. The

absence of such a condition can lead to peculiarities preventing a fully general theory of equivalence of responses. For instance, in the problem of heat conduction in one-dimensional systems, thermostating all particles induces ‘normal’ behaviour, while for generical systems, with boundary thermostats and in the absence of external forces, one obtains anomalous behaviours. We add that, for LTE to be established, the thermostating artificial terms should not dominate, as may occur in systems with few or noninteracting particles [97].

#### 4.6. *Deterministic versus stochastic thermostats*

Among boundary thermostats, one often finds the implementation of Maxwellian boundary conditions, meaning that a particle hitting a wall will be reflected with a Maxwellian distribution of velocities corresponding to the temperature of that wall. With two opposite walls at different temperatures, one expects a steady state to be established and that is largely independent of whether a stochastic or deterministic boundary thermostat is used. But a complete equivalence is far from established and is probably not reasonable to expect. While the Gibbs entropy of deterministically thermostatted nonequilibrium systems does not exist (tending to  $-\infty$  over time [105, 106]), it typically does exist for stochastically modelled nonequilibrium steady states [107]. This is due to the fact that deterministically thermostatted steady states are described by a singular phase space probability distribution, while the stochastically driven systems are generally thought to be characterized by regular distributions. However, the Gibbs entropy is a rather peculiar object in nonequilibrium systems, and this difference need not impair the equivalence picture.

At the same time, stochastic thermostats have various advantages over the deterministic ones, from the perspective of mathematical tractability. The stochastic description, commonly assumed to be a reduced (mesoscopic) representation of the ‘chaotic’ microscopic dynamics, is free from the intricate fractal structures of deterministic dynamics [108]. Considering that, at times, the mathematical approach to deterministic dynamics in physical contexts makes assumptions which can hardly be directly verified (cf below and [109]), the stochastic approach is thought by some authors to be preferable to the deterministic one [108, 110]. This has to be balanced by the observation that the identification of physical observables in stochastic processes is often affected by ambiguities [76, 111–114]. Even from the mathematical standpoint, the use of deterministic or of stochastic approaches has to be dictated by convenience. Hence, there is no absolutely superior method.

### 5. Recent applications to nonequilibrium theories

Many of the observations in the previous section have been drawn from investigations of nonequilibrium phenomena, where thermostats have been applied under nonequilibrium conditions. For example, the need for sufficient degrees of freedom in order to equate statistical and thermodynamic properties has been recognized through investigations over a range of systems, including the highly idealized Lorentz gas model, microporous media and bulk media. In this section, however, we focus on the role of thermostats in the development of *theories* of nonequilibrium phenomena.

#### 5.1. *Phase space expansion rate and fluctuation relations*

In the space of relevant variables, neither the Gaussian thermostatted equations of motion nor the Nosé–Hoover equations of motion can be derived from a Hamiltonian. This is true even

in the absence of external perturbing fields, which implies that the usual form of the Liouville equation for the  $N$ -particle distribution function  $f$

$$\frac{df}{dt} = 0,$$

which is related to the conservation of phase space volumes, does not hold in general. Volumes do indeed fluctuate and the corresponding generalized Liouville equation takes the form

$$\frac{df}{dt} = -f \nabla_{\Gamma} \cdot \dot{\Gamma} = -f \Lambda,$$

where  $\Lambda$ , the *phase space expansion rate*, is the divergence of  $\dot{\Gamma}$  in the  $2dN$ -dimensional phase space ( $d$  being the dimension of the physical space) defined by equation (11).

Typically, thermostatted equations of motion lead to  $\langle \Lambda \rangle < 0$ , implying that volumes contract on average and that, in the steady state, the whole probability is distributed on a set of vanishing phase space volume. This reveals that the Gibbs entropy for thermostatted equations of motion is not fit for the purpose of describing the entropy of a nonequilibrium steady state, since it diverges [105, 106]; formally one may write:

$$S_G = -k_B \int f(\Gamma) \ln f(\Gamma) d\Gamma = -\infty. \quad (31)$$

This result is independent of how close to equilibrium the system might be; hence, it is not a consequence of lack of local equilibrium, without which it certainly makes no sense to speak of thermodynamic entropy<sup>9</sup>. Nevertheless, as we have seen above, the phase space expansion rate of certain Gaussian thermostatted IE systems is simply related to the entropy production rate, and this has led to [25], a milestone of contemporary nonequilibrium statistical mechanics.

*5.1.1. History.* In 1993, Evans, Cohen and Morriss published a seminal paper [25] on the fluctuations of the dissipated power, or the entropy production rate  $\sigma$ , in macroscopic systems close to equilibrium. In the model of [25], this observable equals the dissipation function  $\Omega$  [115] (defined in section 5.1.3) as well as the phase space contraction rate  $-\Lambda$ . The relation proposed and tested in [25] is

$$\frac{P_{\tau}(A)}{P_{\tau}(-A)} = e^{\tau A}, \quad (32)$$

where  $A$  and  $-A$  are averages of the dissipated power, divided by  $k_B T$ , on evolution segments of duration  $\tau$  and  $P_{\tau}$  is their steady-state probability. To obtain this remarkable relation, which does not contain any adjustable parameter, the authors were inspired by the periodic orbit theory being developed in those years [116, 117].

In 1994, Evans and Searles obtained the first transient  $\Omega$ -FR, which is only formally similar to equation (32), because it concerns an ensemble of experiments all beginning in the same initial state, in which  $\Omega$  is measured [115, 118–123]. Differently, equation (32) concerns a single steady-state experiment, which is observed as it evolves in time. The only requirement for the transient  $\Omega$ -FRs to hold is the reversibility of the microscopic dynamics, while the steady-state relation needs further conditions to be met. The transient relations have been experimentally verified [124].

<sup>9</sup> A solution to this difficulty is proposed in [161], where a microscopic definition of entropy is presented, based on a truncation of the Green's expansion of  $f$  (cf below).



In 1995, Gallavotti and Cohen made the use of the Lyapunov weights of [25] mathematically precise, introducing the chaotic hypothesis [109, 125–127], which states the following.

**Chaotic hypothesis.** *A reversible many-particle system in a stationary state can be regarded as a transitive Anosov system for the purpose of computing its macroscopic properties.*

The result is a steady-state relation, which we call  $\Lambda$ -FR, as it concerns the fluctuations of  $\Lambda$ . As systems of physical interest cannot be expected to be of Anosov type, Gallavotti and Cohen assumed that their deviations from the ideal situation are not relevant in practice, i.e. for their observable properties.

This motivated further studies of the Evans–Searles approach [128] and eventually led to the identification of the roles, differences and contact points of the theories developed by Gallavotti and Cohen and by Evans and Searles [76]. It was recognized that the Gallavotti and Cohen theory is more concerned about mathematical aspects, like the existence of systems for which the  $\Lambda$ -FR can be rigorously obtained, independent of its physical relevance, while the Evans and Searles theory is mostly interested in physical aspects, such as the time scales and the physical mechanisms involved in the verification of the  $\Omega$ -FR in systems which do verify it. These two approaches contribute together to our understanding of nonequilibrium phenomena, even if they mostly proceed along distinct paths (see [76] for open problems met in both approaches). A similar parallel can be made between the ergodic hypothesis of physical interest, championed by Khinchin [7], and the mathematical ergodic theory, which may be traced back to the works of Birkhoff [129]. Even in the case of the ergodic hypothesis, indeed, one is faced with the fact that most systems of physical interest are not mathematically ergodic, and with the problem of time scales: yet the ergodic hypothesis is almost universally accepted.

*5.1.2. The Gallavotti–Cohen approach.* The idea proposed by Gallavotti and Cohen [109] is that dissipative, reversible, transitive Anosov maps,  $S : \mathcal{M} \rightarrow \mathcal{M}$ , are idealizations of nonequilibrium particle systems and hence that properties like the  $\Lambda$ -FR are enjoyed by physical systems as well. That the system evolves with discrete or continuous time was thought to be a side issue in [109]. To prove the validity of the  $\Lambda$ -FR for Anosov maps, one should first observe that they admit a *Markovian* partitioning of the phase space  $\mathcal{M}$  [130]. This is a subdivision of  $\mathcal{M}$  into cells whose interiors are disjoint from each other, and whose boundaries are invariant sets, which in two dimensions are constructed using pieces of the stable and unstable manifolds. Consequently, the interior of a cell is mapped by  $S$  into the interior of other cells, and not across two cells, which would include a piece of their boundary. Furthermore, arbitrarily fine partitions can be constructed, exploiting the time-reversibility of the dynamics. Gallavotti and Cohen further assume that the dynamics is transitive, i.e. that a typical trajectory explores all regions of  $\mathcal{M}$ , as finely as one wishes. It is this structure that guarantees that the probability (Lyapunov) weights of equation (1) in [25], from which the  $\Lambda$ -FR follows, can be assigned to the cells of a Markov partition.

More precisely, let  $\Lambda(X) = \log J(X)$ , where  $J$  is the Jacobian determinant of  $S^{10}$  and consider the steady-state probability of the dimensionless phase space contraction rate  $e_\tau$ , obtained along a trajectory segment  $w_{X,\tau}$ , of origin  $X \in \mathcal{M}$  and duration  $\tau$ :

<sup>10</sup> If the point  $X$  has  $d$  coordinates,  $X_i, i = 1, \dots, d$ , we can write  $X_i(k+1) = f_i(X(k))$ , where  $f_i$  is a suitable function determined by  $S$ . Then  $J(X)$  is the absolute value of the determinant of the matrix  $(\partial f_i / \partial X_j)_X$ .

$$e_\tau(X) = \frac{1}{\tau \langle \Lambda \rangle} \sum_{k=-\tau/2}^{\tau/2-1} \Lambda(S^k X), \tag{33}$$

where  $\langle \cdot \rangle$  is the steady-state phase space average and  $S^k X$  denotes the evolution that  $S$  generates from the initial condition  $X(0) = X$ . Let  $J^u$  be the Jacobian determinant of  $S$  restricted to the unstable manifold  $V^+$ , i.e. the product of the asymptotic factors of separation of nearby points, along the directions in which distances asymptotically grow at an exponential rate. If the system is Anosov, the probability of the event  $e_\tau(X) \in B_{p,\epsilon} \equiv (p - \epsilon, p + \epsilon)$  coincides, in the limit of fine Markov partitions and long  $\tau$ 's, with the sum of the weights

$$w_{X,\tau} = \prod_{k=-\tau/2}^{\tau/2-1} \frac{1}{J^u(S^k X)} \tag{34}$$

of the cells containing the points  $X$  such that  $e_\tau(X) \in B_{p,\epsilon}$ . Then, if  $\pi_\tau(B_{p,\epsilon})$  is the corresponding probability, one can write

$$\pi_\tau(e_\tau(X) \in B_{p,\epsilon}) \approx \frac{1}{M} \sum_{X, e_\tau(X) \in B_{p,\epsilon}} w_{X,\tau}, \tag{35}$$

where  $M$  is a normalization constant. If the support of the physical measure is  $\mathcal{M}$ , which is the case if the dissipation is not exceedingly high [131], time-reversibility guarantees that the support of  $\pi_\tau$  includes an interval  $[-p^*, p^*]$ ,  $p^* > 0$ , and one can consider the ratio

$$\frac{\pi_\tau(B_{p,\epsilon})}{\pi_\tau(B_{-p,\epsilon})} \approx \frac{\sum_{X, e_\tau(X) \in B_{p,\epsilon}} w_{X,\tau}}{\sum_{X, e_\tau(X) \in B_{-p,\epsilon}} w_{X,\tau}}, \tag{36}$$

where each  $X$  in the numerator has a counterpart in the denominator. Denoting by  $I$  the involution which replaces the initial condition of one trajectory with the initial condition of the reversed trajectory, time-reversibility yields

$$\Lambda(X) = -\Lambda(IX), \quad w_{IX,\tau} = w_{X,\tau}^{-1} \quad \text{and} \quad \frac{w_{X,\tau}}{w_{IX,\tau}} = e^{-\tau \langle \Lambda \rangle p} \tag{37}$$

if  $e_\tau(X) = p$ . Taking small  $\epsilon$  in  $B_{p,\epsilon}$ , the division of each term in the numerator of (36) by its counterpart in the denominator approximately equals  $e^{-\tau \langle \Lambda \rangle p}$ , which then equals the ratio in (36). In the limit of small  $\epsilon$ , infinitely fine Markov partition and large  $\tau$ , the authors of [109] obtain the following theorem.

**Gallavotti–Cohen theorem (1995).** *Let  $(\mathcal{M}, S)$  be dissipative (i.e.  $\langle \Lambda \rangle < 0$ ), reversible and assume that the chaotic hypothesis holds. Then,*

$$\frac{\pi_\tau(B_{p,\epsilon})}{\pi_\tau(B_{-p,\epsilon})} = e^{-\tau \langle \Lambda \rangle p} \tag{38}$$

*with an error in the argument of the exponential which can be estimated to be  $p$ - and  $\tau$ -independent.*

If the  $\Lambda$ -FR (hence the chaotic hypothesis on which it is based) holds, the function  $C(p; \tau, \epsilon) = (1/\tau \langle -\Lambda \rangle) \log[\pi_\tau(B_{p,\epsilon})/\pi_\tau(B_{-p,\epsilon})]$  tends to a straight line of slope 1 for growing  $\tau$ , apart from small errors. If  $\Lambda$  can be identified with a physical observable, the  $\Lambda$ -FR is a parameter-free statement about the physics of nonequilibrium systems.

Under the assumption that  $-\Lambda$  coincides with the entropy production rate, the  $\Lambda$ -FR can be used to obtain the Green–Kubo relations and the Onsager reciprocal relations, in the limit of small external drivings [132]. This way, the  $\Lambda$ -FR appears to be an extension of such relations to nonequilibrium systems. In order to obtain this result, Gallavotti assumes that

the (Gaussian isokinetic, continuous time) system is driven by  $\ell$  fields  $F = (F_1, F_2, \dots, F_\ell)$ , that the phase space contraction rate vanishes when all drivings vanish, and that it can be expanded as  $-\Lambda(X) = \sum_{i=1}^{\ell} F_i J_i^0(X) + O(F^2)$ . In this way, the linear ‘currents’  $J_i^0$ , which are proportional to the forces  $F_i$ , are defined. If the decay of the  $\Lambda$ -autocorrelation function is sufficiently fast, one then has

$$\zeta(p) \equiv - \lim_{\tau \rightarrow \infty} \frac{\log \pi_{\tau}(p)}{\tau} = \frac{\langle \Lambda \rangle^2}{2C_2} (p-1)^2 + O((p-1)^3 F^3),$$

$$C_2 = \int_{-\infty}^{\infty} \langle \Lambda(S^t X) \Lambda(X) \rangle_C dt,$$
(39)

where  $\langle \cdot \rangle_C$  denotes the cumulant. Thus, the validity of the  $\Lambda$ -FR, equation (38), implies  $\langle -\Lambda \rangle = C_2/2 + O(F^3)$ . Now, let the full (nonlinear) ‘currents’ be defined by  $J_i(X) = -\partial_{F_i} \Lambda(X)$ , and the transport coefficients be  $L_{ij} = \partial_{F_j} \langle J_i \rangle|_{F=0}$ . The derivatives with respect to the parameters  $F$  require a property of differentiability of SRB measures, which has been proven by Ruelle [133]. Assuming this property, the validity of the  $\Lambda$ -FR and time-reversibility, one can write

$$\langle -\Lambda \rangle = \frac{1}{2} \sum_{i,j=1}^{\ell} (\partial_{F_j} \langle J_i \rangle + \partial_{F_i} \langle J_j \rangle)|_{F=0} F_i F_j = \frac{1}{2} \sum_{i,j=1}^{\ell} (L_{ij} + L_{ji}) F_i F_j$$
(40)

to second order in the forces. Equating this with  $C_2/2$  and considering  $(L_{ij} + L_{ji})/2$  with  $i = j$ , one recovers the Green–Kubo relations. To obtain the symmetry  $L_{ij} = L_{ji}$ , Gallavotti extends the  $\Lambda$ -FR to consider the joint distribution of  $-\Lambda$  and its derivatives. He introduces the dimensionless current  $q$ , averaged over a long time  $\tau$ , through the relation

$$\frac{1}{\tau} \int_{-\tau/2}^{\tau/2} F_j \partial_{F_j} \Lambda(S^t X) dt = F_j \langle \partial_{F_j} \Lambda \rangle q(X)$$
(41)

and considers the joint distribution  $\pi_{\tau}(p, q)$ , with the corresponding large deviation functional  $\zeta(p, q) = - \lim_{\tau \rightarrow \infty} \frac{1}{\tau} \log \pi_{\tau}(p, q)$ . The result is a relation similar to the  $\Lambda$ -FR:

$$\lim_{\tau \rightarrow \infty} \frac{1}{\tau \langle -\Lambda \rangle p} \log \frac{\pi_{\tau}(p, q)}{\pi_{\tau}(-p, -q)} = 1.$$
(42)

This makes the difference  $(\zeta(p, q) - \zeta(-p, -q))$  independent of  $q$ , which leads to the desired result,  $L_{ij} = L_{ji}$ , in the limit of small  $F$ . This work was refined in [134]; for related results, based on orbital measures, see [67, 99].

Assuming that the currents and transport coefficients, defined here in terms of the phase space contraction rate, do represent physical quantities, these results show that the FRs are consistent with irreversible thermodynamics, close to equilibrium. Hence they may be considered in the construction of a comprehensive nonequilibrium theory. However, some difficulties affect the present approach, the first being that  $\Lambda$ , which is directly related to the thermostating term  $\alpha_{th}$ , is only proportional to the energy dissipation rate divided by the kinetic energy in very special cases [76, 135].

Because global fluctuations are not observable in macroscopic systems, local fluctuation relations have been devised, see e.g. [136–138], and in [92] a local version the  $\Lambda$ -FR was tested numerically. In particular, the local  $\Lambda$ -FR of [138] concerns an infinite chain of weakly interacting chaotic maps. Let  $V_0$  be a finite region of the chain centred at the origin,  $T_0 > 0$  be a time interval and define

$$\langle \Lambda \rangle = \lim_{V_0, T_0 \rightarrow \infty} \frac{1}{|V_0| T_0} \sum_{j=0}^{T_0-1} \Lambda_{V_0}(S^j X), \quad p = \frac{1}{\langle \Lambda \rangle |V|} \sum_{j=-T_0/2}^{T_0/2} \Lambda_{V_0}(S^j X),$$
(43)

where  $V = V_0 \times T_0$ ,  $\Lambda_{V_0}(X)$  is the contribution to  $\Lambda$  given by  $V_0$ , and  $|E|$  denotes the volume of the set  $E$ . Then, one obtains

$$\pi_V(p) = e^{\zeta(p)|V|+O(|\partial V|)} \quad \text{with} \quad \frac{\zeta(p) - \zeta(-p)}{p\langle\Lambda\rangle} = 1 \quad \text{and} \quad |p| < p^*, \quad (44)$$

where  $|\partial V|$  is the size of the boundary of  $V$ ,  $p^* \geq 1$  and  $\zeta$  is analytic in  $p$ . The contribution of the boundary term  $|\partial V|$  should decrease with growing  $V$ , leading to the  $\Lambda$ -FR in the limit of large (compared to microscopic scales) volume  $V_0$  and long times  $T_0$ .

This naturally leads to a possible extension of the Onsager–Machlup theory (see section 5.2) to nonequilibrium systems [139, 140], again taking the entropy production rate as proportional to  $\Lambda$ .

Because  $\Lambda$  differs from the dissipated power, in general, some interesting problems have arisen in the physical interpretation of the Gallavotti–Cohen theorem [135]. For instance, it was found that the steady-state  $\Lambda$ -FR is difficult, if not impossible, to verify in non-isoenergetic systems with singular  $\Lambda$ , close to equilibrium [121, 141, 142]. To explain these facts, [135] investigates Gaussian isokinetic systems, whose  $\Lambda$  is the sum of a dissipative term  $\Omega$  and a conservative singular term. In that case,  $\Omega$  obeys the FR, while the conservative term does not, but its averages over long time intervals are expected to become negligible with respect to the averages of  $\Omega$  as the length of the interval grows [135, 142]. Thus, in the long time limit, the  $\Lambda$ -FR should hold as a consequence of the validity of the  $\Omega$ -FR, while the convergence times of the  $\Lambda$ -FR would diverge when equilibrium is approached, because  $\Omega$  vanishes as the square of the driving forces. These observations eventually led to the conclusion that, in some cases,  $\Lambda$  describes heat fluxes, not entropy productions [135, 143, 144], and hence that in those cases the  $\Lambda$ -FR has to be modified, to mimic the heat FR of Van Zon and Cohen for stochastic systems [145]. This amounts also to saying that the white noise present in the systems studied by Van Zon and Cohen is reproduced to some extent by the deterministic chaos of the uniformly hyperbolic dynamical systems of [144], and that the same may be expected in sufficiently chaotic particle systems, like the typical NEMD models. Therefore, we have one example of equivalence between stochastic and deterministic evolutions, as far as the FRs are concerned.

*5.1.3. The Evans–Searles approach.* Having realized that  $\Lambda$  has no obvious physical meaning, in general, and that the  $\Lambda$ -FR is based on strong assumptions which are not strictly met by systems of physical interest, one may legitimately ask whether a steady-state relation for a physically relevant observable, like the dissipated power, may be obtained under physically meaningful assumptions. Alternatively, one may ask which physical mechanisms are at work when a given observable obeys a fluctuation relation. This goal is the subject of [128], in which the original approach of Evans and Searles, first proposed in [118], is developed.

Reference [128] considers time reversal invariant dynamics  $S^\tau : \mathcal{M} \rightarrow \mathcal{M}$  on a phase space  $\mathcal{M}$ , with the time reversal involution operation  $I : \mathcal{M} \rightarrow \mathcal{M}$ . Time integrals and time averages of observables  $\phi : \mathcal{M} \rightarrow \mathbb{R}$  are denoted as

$$\bar{\phi}_{t_0, t_0+\tau}(\Gamma) \equiv \frac{1}{\tau} \int_{t_0}^{t_0+\tau} \phi(S^s \Gamma) ds \equiv \frac{1}{\tau} \phi_{t_0, t_0+\tau}(\Gamma). \quad (45)$$

For the purpose of deriving FRs,  $\phi : \mathcal{M} \rightarrow \mathbb{R}$  is assumed to be odd with respect to time reversal, i.e.  $\phi(I\Gamma) = -\phi(\Gamma)$ .

An absolutely continuous probability measure  $d\mu(\Gamma) = f(\Gamma) d\Gamma$  on  $\mathcal{M}$ , with time even density [ $f(I\Gamma) = f(\Gamma)$ ], is introduced, in order to define the *Dissipation Function*

$$\Omega(\Gamma) = - \frac{d}{d\Gamma} \log f \Big|_{\Gamma} \cdot \dot{\Gamma} - \Lambda(\Gamma), \quad \text{or} \quad \bar{\Omega}_{t_0, t_0+\tau}(\Gamma) = \frac{1}{\tau} \left[ \ln \frac{f(S^{t_0}\Gamma)}{f(S^{t_0+\tau}\Gamma)} - \Lambda_{t_0, t_0+\tau} \right]. \quad (46)$$

It is a remarkable fact that  $\Omega$  equals the dissipated power (up to  $O(1/N)$  corrections, for  $N$ -particle systems) if  $f$  is the equilibrium probability density for the given system, or if it is appropriately obtained from the equilibrium probability density [128]. The dissipation function does not encompass any non-dissipative terms, and therefore vanishes at equilibrium [146].

The existence of the logarithmic term of equation (46) has been called *ergodic consistency* [115]. For  $\delta > 0$ , denote  $(A)_\delta \equiv (A - \delta, A + \delta)$ ,  $A \in \mathbb{R}$ , and let  $E(\phi \in (a, b))$  be the set of points  $\Gamma$  such that  $\phi(\Gamma) \in (a, b)$ . Then,  $E(\bar{\Omega}_{0,\tau} \in (-A)_\delta) = IS^t E(\bar{\Omega}_{0,\tau} \in (A)_\delta)$ , and the transformation  $\Gamma = IS^t X$  has Jacobian

$$\left| \frac{d\Gamma}{dX} \right| = \exp \left( - \int_0^\tau \Lambda(S^s X) ds \right) = e^{-\Lambda_{0,\tau}(X)}, \quad (47)$$

which leads to

$$\frac{\int_{E(\bar{\Omega}_{0,\tau} \in (A)_\delta)} f(\Gamma) d\Gamma}{\int_{E(\bar{\Omega}_{0,\tau} \in (-A)_\delta)} f(X) dX} = \frac{\int_{E(\bar{\Omega}_{0,\tau} \in (A)_\delta)} f(\Gamma) d\Gamma}{\int_{E(\bar{\Omega}_{0,\tau} \in (A)_\delta)} \exp[-\Omega_{0,\tau}(X)] f(X) dX} \quad (48)$$

i.e. to the *transient*  $\Omega$ -FR

$$\frac{\mu(E(\bar{\Omega}_{0,\tau} \in (A)_\delta))}{\mu(E(\bar{\Omega}_{0,\tau} \in (-A)_\delta))} = e^{[A + \epsilon(\delta, A, \tau)]\tau}, \quad |\epsilon| \leq \delta \quad (49)$$

which concerns the non-invariant probability measure  $\mu$  of density  $f$ . Time-reversibility is essentially the only ingredient in the above derivation. This relation is valid for any  $f$ , and hence any  $\Omega$  that is obtained from  $f$  as in equation (46), but it acquires the meaning of a relation for a physically relevant quantity only if  $f$  is properly chosen, as mentioned above.

To obtain the steady-state  $\Omega$ -FR, consider instead the ratio

$$\frac{\mu(E(\bar{\phi}_{t_0, t_0+\tau} \in (A)_\delta))}{\mu(E(\bar{\phi}_{t_0, t_0+\tau} \in (-A)_\delta))} = \frac{\int_{E(\bar{\phi}_{t_0, t_0+\tau} \in (A)_\delta)} f(\Gamma) d\Gamma}{\int_{E(\bar{\phi}_{t_0, t_0+\tau} \in (-A)_\delta)} f(\Gamma) d\Gamma} \quad (50)$$

and take  $t = 2t_0 + \tau$ . Then

$$E(\bar{\phi}_{t_0, t_0+\tau} \in (-A)_\delta) = IS^t E(\bar{\phi}_{t_0, t_0+\tau} \in (A)_\delta) \quad (51)$$

so that setting  $\Gamma = IS^t W$  and performing the same algebra leads to

$$\frac{\mu(E(\bar{\phi}_{t_0, t_0+\tau} \in (A)_\delta))}{\mu(E(\bar{\phi}_{t_0, t_0+\tau} \in (-A)_\delta))} = \langle \exp(-\bar{\Omega}_{0, t_0} t_0) \rangle_{\bar{\phi}_{t_0, t_0+\tau} \in (A)_\delta}^{-1}. \quad (52)$$

Using  $\mu(E) = \mu_{t_0}(S^{t_0}E) = \int_{S^{t_0}E} f_{t_0}(W) dW$ , where  $\mu_{t_0}$  is the evolved measure up to time  $t_0$ , and  $f_{t_0}$  its density, one obtains

$$\frac{\mu_{t_0}(E(\bar{\phi}_{0,\tau} \in (A)_\delta))}{\mu_{t_0}(E(\bar{\phi}_{0,\tau} \in (-A)_\delta))} = \frac{\mu_{t_0}(S^{t_0}E(\bar{\phi}_{t_0, t_0+\tau} \in (A)_\delta))}{\mu_{t_0}(S^{t_0}E(\bar{\phi}_{t_0, t_0+\tau} \in (-A)_\delta))} \quad (53)$$

$$= \frac{\mu(E(\bar{\phi}_{t_0, t_0+\tau} \in (A)_\delta))}{\mu(E(\bar{\phi}_{t_0, t_0+\tau} \in (-A)_\delta))} = \frac{1}{\langle \exp(-\bar{\Omega}_{0,t} t) \rangle_{\bar{\phi}_{t_0, t_0+\tau} \in (A)_\delta}}. \quad (54)$$

This has been called the  $\phi$ -FR relation [128] and for  $\bar{\phi}_{t_0, t_0+\tau} = \bar{\Omega}_{t_0, t_0+\tau}$ , it may be rewritten as

$$\frac{1}{\tau} \ln \frac{\mu_{t_0}(E(\bar{\Omega}_{0,\tau} \in (A)_\delta))}{\mu_{t_0}(E(\bar{\Omega}_{0,\tau} \in (-A)_\delta))} = A + \epsilon(\delta, t_0, A, \tau) - \frac{1}{\tau} \ln \langle e^{-t_0(\bar{\Omega}_{0,t_0} + \bar{\Omega}_{t_0+\tau, 2t_0+\tau})} \rangle_{\bar{\Omega}_{t_0, t_0+\tau} \in (A)_\delta}. \quad (55)$$

This is an exact relation for  $\Omega$ , which may yield the steady-state  $\Omega$ -FR if  $\mu_{t_0}$  tends to a steady-state distribution  $\mu_\infty$  for  $t_0 \rightarrow \infty$  and certain conditions are met. In the  $\tau \rightarrow \infty$  limit, this would turn (55) from a statement on the ensemble  $f_{t_0}$  into a statement concerning also the time statistics of a single typical trajectory. One delicate point in this framework is that  $t_0$  tends to infinity before  $\tau$  does and the growth of  $t_0$  could make the conditional average in (55) diverge. However, the decay of the auto-correlation of  $\Omega$ , with respect to the initial probability measure  $\mu_0$ , suffices. Indeed, the conservation of probability yields

$$\langle e^{-s\bar{\Omega}_{0,s}} \rangle = 1 \quad \text{for every } s \in \mathbb{R}, \quad (56)$$

and if the  $\Omega$ -autocorrelation time vanishes instantaneously, one can write

$$\begin{aligned} \langle e^{-t_0\bar{\Omega}_{0,t_0}} e^{-t_0\bar{\Omega}_{t_0+\tau,2t_0+\tau}} \rangle_{\bar{\Omega}_{t_0,t_0+\tau} \in (A)_\delta} &= \langle e^{-t_0(\bar{\Omega}_{0,t_0} + \bar{\Omega}_{t_0+\tau,2t_0+\tau})} \rangle \\ &= \langle e^{-t_0\bar{\Omega}_{0,t_0}} \rangle \langle e^{-t_0\bar{\Omega}_{t_0+\tau,2t_0+\tau}} \rangle \end{aligned} \quad (57)$$

and

$$1 = \langle e^{-s\bar{\Omega}_{0,s} - (t-s)\bar{\Omega}_{s,t}} \rangle = \langle e^{-\Omega_{0,s}} \rangle \langle e^{-\Omega_{s,t}} \rangle, \quad \text{i.e. } \langle e^{-(t-s)\bar{\Omega}_{s,t}} \rangle = 1 \quad \text{for all } s, t. \quad (58)$$

Hence, the logarithmic correction term in (55) identically vanishes for all  $t_0, \tau$ , and the  $\Omega$ -FR is verified even at short  $\tau$ 's. Of course, this idealized situation does not need to be realized, but a constant  $K$ , such that

$$0 < \frac{1}{K} \leq \langle e^{-t_0(\bar{\Omega}_{0,t_0} + \bar{\Omega}_{t_0+\tau,2t_0+\tau})} \rangle_{\bar{\Omega}_{t_0,t_0+\tau} \in (A)_\delta} \leq K \quad (59)$$

is likely to exist in standard NEMD models [128]. As a matter of fact, the decorrelation (or Maxwell) time  $t_M$  expresses a physical property of the system; thus, it does not depend on  $t_0$  or  $\tau$ , and depends only mildly on the external field (usually,  $t_M(F_e) = t_M(0) + O(F_e^2)$ ). Its order of magnitude is that of the mean free time. If these scenarios are realized, equation (59) follows and the logarithmic correction term of equation (55) vanishes as  $1/\tau$ , with a characteristic scale of order  $O(t_M)$ .

Therefore, the steady-state  $\Omega$ -FR can be obtained only from time-reversibility and from the decay of the  $\Omega$ -autocorrelation. One may then write

$$\frac{1}{\tau} \ln \frac{\mu_\infty(E(\bar{\Omega}_{0,\tau} \in (A)_\delta))}{\mu_\infty(E(\bar{\Omega}_{0,\tau} \in (-A)_\delta))} = A + \text{correction}, \quad (60)$$

where the correction term can be made arbitrarily small taking sufficiently large  $\tau$  and sufficiently small  $\delta$ . This explains that the relevant convergence times are functions of material properties of the systems and do not diverge in the equilibrium limit, in systems of thermodynamic interest. The decorrelation times of  $\Omega$  are indeed material properties, which are approximately constant around equilibrium. Also, the required decay of correlations is needed for the convergence to a steady state, without which no steady-state relation makes sense. If the steady-state  $\Omega$ -FR does not hold, equation (55) remains valid and describes a property of the ensemble  $f_{t_0}$ .

Various other relations can now be obtained, like

$$\langle \exp(-t\bar{\Omega}_{0,t}) \rangle_{\bar{\phi}_{t_0,t_0+\tau} \in (-\delta,\delta)} = 1 \quad (61)$$

for any odd  $\phi$ , any  $\delta > 0$ , any  $t_0$  and any  $\tau$ , which, in the  $\delta \rightarrow \infty$  limit, leads to the so-called nonequilibrium partition identity (56). Similarly, one obtains the dissipation relation [147]

$$\langle \phi \rangle_t = \int_0^t ds \langle \Omega(0)\phi(s) \rangle, \quad (62)$$

where  $\langle \cdot \rangle_t$  is the average with respect to  $\mu_t$  and represents the time-dependent response to the driving forces.

That this theory is consistent with irreversible thermodynamics close to equilibrium is shown, for instance, by its relation with the Green–Kubo formulae [135]. Differently from [132], which deals with  $\Lambda$  and with asymptotic times, [135] deals with  $\Omega$  and stresses the role of the physical time scales, as illustrated by a Nosé–Hoover thermostatted system (equation (15)), with the equilibrium state given by the extended canonical density

$$f_c(\Gamma, \xi) = \frac{e^{-\beta(H_0+Q\xi^2/2)}}{\int d\xi d\Gamma e^{-\beta(H_0+Q\xi^2/2)}} \quad (63)$$

with  $Q = 2K_0\theta$  and internal energy  $H_0$  [15]. In the case that  $\mathbf{F}_i^{\text{ext}} = c_i \mathbf{F}_e$  and  $\mathbf{J} = \sum_{i=1}^N c_i (d\mathbf{q}_i/dt)$ , one has  $\xi = -\frac{1}{2K} [\frac{dH_0}{dt} + \mathbf{J} \cdot \mathbf{F}_e V]$  and  $\Lambda = \nabla_\Gamma \cdot (\frac{d\Gamma}{dt}) + \partial_\xi (\frac{d\xi}{dt}) = dN\xi$ , where  $d$  is the spatial dimension. Therefore,  $\Lambda$  and its fluctuations are not directly related to the dissipation rate  $\Omega = \mathbf{J} \cdot \mathbf{F}_e V/2K$ , while its average is, because  $\langle H_0 \rangle = 0$ . Also, one has

$$f_c(\xi) \equiv \int d\Gamma f_c(\Gamma, \xi) = \sqrt{\frac{\beta Q}{2\pi}} \exp[-\beta Q \xi^2/2], \quad (64)$$

and the distribution of  $\bar{\xi}_{0,t}$  is Gaussian in equilibrium, with variance proportional to that of the average internal energy

$$\bar{h}_{0,t} \equiv \frac{1}{t} \int_0^t \dot{H}_0 d\tau.$$

Thus, near equilibrium, it can be assumed to remain Gaussian about its mean, for large  $t$ , as a consequence of the central limit theorem (CLT). To use the  $\Lambda$ -FR together with the CLT, the values  $A$  and  $-A$  of the random variable  $\bar{\xi}_{0,t}$  must be a small number of standard deviations away from  $\langle \Omega \rangle = \langle \mathbf{J} \cdot \mathbf{F}_e V/2K$ , where  $K$  is the kinetic energy. However, as explained in [135, 142, 144] and briefly recalled in section 5.1.2, the distribution of  $dN\bar{\xi}_{0,t}$  verifies the  $\Lambda$ -FR only for times  $t$  sufficiently large that the standard deviation of the random variable  $\bar{h}_{0,t}$  is negligible with respect to that of  $\bar{\Omega}_{0,t}$ ,  $\sigma_{\bar{\Omega}_{0,t}}$  say. Because the variance of the average current obeys  $t\sigma_{\bar{\Omega}_{0,t}}^2(F_e) = \frac{2L(F_e)k_B T}{V} + O(\frac{F_e^2}{tN})$ , [121], that of  $\bar{\Omega}_{0,t}$  obeys

$$t\sigma_{\bar{\Omega}_{0,t}}^2(F_e) = \left(\frac{F_e V}{2K}\right)^2 \left[ \frac{2L(F_e)k_B T}{V} + O\left(\frac{F_e^2}{tN}\right) \right], \quad (65)$$

where

$$L(F_e) = \beta V \int_0^\infty dt \langle (J(t) - \langle J \rangle)(J(0) - \langle J \rangle) \rangle$$

and  $L(0) = \lim_{F_e \rightarrow 0} L(F_e)$  is the associated linear transport coefficient. As the standard deviation of the average current decreases when  $t$  grows at fixed  $F_e$ , while it tends to a positive constant, when  $F_e$  decreases at fixed  $t$ , the standard deviation of  $\bar{\Omega}_{0,t}$  tends to zero when  $F_e \rightarrow 0$  at fixed  $t$ . Differently, the variance of  $\bar{h}_{0,t}$  tends to a constant when  $F_e$  tends to zero at fixed  $t$ . Therefore, the smaller the  $F_e$  the longer the time  $t$  needed for  $\sigma_{\bar{\Omega}_{0,t}}$  to dominate over the variance of  $\bar{h}_{0,t}$ , and this time grows without bounds when  $F_e$  tends to zero. Assume for simplicity that the variations of these standard deviations are monotonic when either  $F_e$  or  $t$  varies and the other variable is fixed. Then, given  $F_e$  and  $A$ , there is  $t_\sigma(F_e, A)$  such that the standard deviation of  $\bar{\Omega}_{0,t}$  is sufficiently large that  $A$  and  $-A$  are within a few standard deviations from the mean only if  $t < t_\sigma(F_e, A)$ . At the same time, let  $t_\delta(F_e, A)$  be sufficiently large that the steady-state  $\Lambda$ -FR applies to the values  $A$  and  $-A$ , with accuracy  $\delta$ . To derive the Green–Kubo relations from the  $\Lambda$ -FR, it is necessary (but not sufficient, [135]) that  $t_\delta(F_e, A) < t < t_\sigma(F_e, A)$ . But this does not happen if  $t_\delta(F_e, A)$  grows too fast with decreasing  $F_e$ , which makes problematic the derivation of the Green–Kubo formula from the

$\Lambda$ -FR for systems with singular  $\Lambda$ , because of the diverging times for the verification of their  $\Lambda$ -FR.

Differently, the Green–Kubo relation is conveniently derived from the  $\Omega$ -FR. Indeed, the corresponding convergence times are material properties which do not diverge when  $F_e$  tends to zero, they simply tend to their equilibrium value, which amounts to several Maxwell times. Furthermore, equating equation (33) and (34) in [135], one obtains  $\beta V F_e = \langle J \rangle / t \sigma_{J_{0,t}}^2$ . Then, the fact that  $\Omega = \beta V F_e J$  implies  $\beta V F_e = 2 \langle \Omega \rangle / t \sigma_{J_{0,t}}$ , from which one obtains

$$\langle \Omega \rangle = \frac{t}{2} \sigma_{\Omega_{0,t}}^2 \quad \text{or} \quad L(0) = \lim_{F_e \rightarrow 0} \frac{\langle J \rangle}{F_e} = \beta V \int_0^\infty ds \langle J(0) J(s) \rangle \quad (66)$$

for  $t$  of the order of the Maxwell time, where the first equality is due to the validity of the  $\Omega$ -FR and the second to (65). Observing that  $\Omega$  is the observable of interest, the proper Green–Kubo formula has been recovered.

*5.1.4. Relaxation to the equilibrium state.* One recent result, based on the Evans–Searles approach to the transient fluctuation relation and the dissipation theorem, concerns a first-principles proof of convergence to a steady state or to the microcanonical distribution [148, 149]. The authors show, under very general conditions, that initial distributions will relax (not necessarily monotonically) to the canonical distribution for a system in thermal contact with its environment.

The system considered in [148] is a closed system of particles surrounded by a heat bath. Particles away from the extremities of the system obey the usual Hamiltonian dynamics, whereas particles near the boundary come into contact with the heat bath and are acted upon by a thermostat. In this fashion, the dynamics replicates the influence of an external heat bath on a closed but not energetically isolated system of particles—a common choice for motivating derivations of the canonical distribution. A Nosé–Hoover thermostat was chosen, although the nature of the thermostat was not expected to influence the conclusions of [148].

By invoking the transient  $\Omega$ -FR (equation (49)), the authors show that any initial phase space distribution  $f(\Gamma, t)$  (required only to be *time-even*, i.e. having an even distribution in the momenta, as is usual for molecular systems) evolves under the system dynamics as

$$f(\Gamma, t) = \left[ \exp \int_{-t}^0 ds \Omega(S^s \Gamma) \right] f(\Gamma, 0). \quad (67)$$

Only those distributions that are uniformly dissipationless ( $\Omega(\Gamma) \equiv 0$ ) will be time-independent ( $f(\Gamma, t) \equiv f(\Gamma)$ ). Such dissipationless, time-independent distributions are none other than the *equilibrium* distributions. The authors of [148] show that the canonical distribution<sup>11</sup> is the unique distribution that is uniformly dissipationless under these dynamics—for *any other time-even distribution*,  $\Omega(\Gamma) \neq 0$ . Most importantly, they demonstrate via the transient  $\Omega$ -FR that the average dissipation for any initial distribution must be non-negative, and can only be zero for the equilibrium, canonical distribution. Consequently, any initial (time-even) distribution will converge to the canonical distribution on average, although the monotonicity of this convergence is only guaranteed for certain classes of initial distributions. Monotonic relaxation towards equilibrium is not generally observed in real systems, however, and the authors of [148] thus note that their relaxation theorem ‘allows for much more complex dynamical behaviour as seen experimentally’ than earlier results such as Boltzmann’s *H*-theorem.

<sup>11</sup> Subject to a conservation of momentum constraint in each Cartesian direction.



5.2. Extensions of Onsager–Machlup theory and deterministic thermostats

Various attempts have been made to extend Onsager–Machlup theory [150, 151] to the large fluctuations of physical systems in nonequilibrium steady states [152]. Among them, the work by Jona-Lasinio and collaborators [153–155] is particularly relevant to our discussion. Their approach generalizes those of Derrida *et al* [152], and in addition it leads to an independent derivation of the FR. The theory of [153–155] begins from the assumption that a hydrodynamic-like description of the system at hand is possible: hence, physically, it can be applied as far from equilibrium as the validity of the local thermodynamic equilibrium allows<sup>12</sup>.

The theory leads to the conclusion that the nonequilibrium entropy functional, which generalizes the Onsager–Machlup entropy to extended systems, is a non-local functional of the thermodynamic variables, and hence that correlations are present over macroscopic scales. This result was observed experimentally in nonequilibrium systems in the 1970s (see [156] for a review on the subject): mathematically, it is expressed for toy models such as the open symmetric simple exclusion process by the fact that the corresponding large deviation functional is given by an integral over all points of the system (cf equation (2.2) of [152]). As an illustration, consider stochastic models of interacting particles, whose number is locally conserved, in contact with particle reservoirs. Assume that these systems admit the hydrodynamic description

$$\partial_t \varrho = \nabla \cdot \left[ \frac{1}{2} D(\varrho) \nabla \varrho \right] \equiv \mathcal{D}(\varrho), \quad \varrho = \varrho(\mathbf{x}, t), \tag{68}$$

where  $\varrho$  is the vector of macroscopic observables,  $\mathbf{x}$  is the macroscopic space variable,  $t$  is the macroscopic time,  $D$  is the diffusion matrix.

The mathematical theory is necessarily developed for very idealized models, such as the simple exclusion or the zero range stochastic processes<sup>13</sup>. Nevertheless, the assumptions under which the theory holds are thought to be valid much more generally than in these cases, and are the following.

**Assumptions**

- (1) *The mesoscopic evolution is given by a Markov process  $X_t$ , which represents the configuration of the system at time  $t$ . The nonequilibrium steady state is described by a probability measure  $\text{Prob}$  over the trajectories of  $X_t$ .*
- (2) *The macroscopic description is given in terms of fields  $\varrho$  which constitute the local thermodynamic variables, whose evolution is described by (68), which has a unique stationary solution  $\tilde{\varrho}$ , under the given nonequilibrium boundary conditions.*
- (3) *Denoting by  $I$  the time inversion operator defined by  $I X_t = X_{-t}$ , the probability measure  $\text{Prob}^*$ , describing the evolution of the time-reversed process  $X_t^*$ , and  $\text{Prob}$  are related by*

$$\text{Prob}^*(X_t^* = \phi_t, t \in [t_1, t_2]) = \text{Prob}(X_t = \phi_{-t}, t \in [-t_2, -t_1]). \tag{69}$$

*Moreover, if  $L$  is the generator of  $X_t$ , the adjoint dynamics is generated by the adjoint (with respect to the invariant measure  $\mu$ ) operator  $L^*$ , which admits the adjoint hydrodynamic description*

$$\partial_t \varrho = \mathcal{D}^*(\varrho). \tag{70}$$

<sup>12</sup> This, of course, includes a very wide range of phenomena, well beyond the linear regime, which has quite a wide applicability by itself.

<sup>13</sup> These models are usually one-dimensional. Their hydrodynamic limit consists of a scaling of the microscopic space and time variables,  $\tau$  and  $\mathbf{r}$ , with the macroscopic space and time variables given by  $t = \tau/N^2$  and  $\mathbf{x} = \mathbf{r}/N$ , and the number of particles per unit length  $N$  tending to infinity.

- (4) *The measure Prob admits a large deviation principle describing the fluctuations of  $\varrho$ , i.e. the probability for a large number  $N$  of particles that the evolution of the random variable  $\varrho_N$  deviates from the solution of (68), to follow a given path  $\hat{\varrho}(t)$ , in the interval  $[t_i, t_f]$ , goes like*

$$\text{Prob}(\varrho_N(X_{N^2t}) \sim \hat{\varrho}(t), [t_i, t_f]) \approx e^{-N^d[\mathcal{S}(\hat{\varrho}(t_i)+J_{[t_i,t_f]}(\hat{\varrho}))]}, \quad (71)$$

where  $d$  is the spatial dimension,  $\mathcal{S}(\hat{\varrho}(t_i))$  is the entropy cost to produce  $\hat{\varrho}(t_i)$  and  $J$  is the extra cost required to follow the given path (taking  $\mathcal{S}(\tilde{\varrho}) = 0$ ,  $J$  vanishes at  $\varrho = \tilde{\varrho}$ ).

This machinery leads to a number of interesting results, among which the generalization of the Onsager–Machlup theory and the introduction of the ‘adjoint hydrodynamic’ equation (70), for the spontaneous fluctuations around nonequilibrium steady states. Assuming that  $\mathcal{D}$  can be decomposed as

$$\mathcal{D}(\varrho) = \frac{1}{2} \nabla \cdot \left( \chi(\varrho) \nabla \frac{\delta \mathcal{S}}{\delta \varrho} \right) + \mathcal{A}, \quad (72)$$

where  $\mathcal{A}$  is a vector field orthogonal to the thermodynamic force  $\delta \mathcal{S} / \delta \varrho$  (the functional derivative of the entropy with respect to the state), a temporal asymmetry arises in the fluctuation–relaxation paths. Indeed, a spontaneous fluctuation out of a nonequilibrium steady state follows a trajectory which is the *time reversal* of the relaxation path, according to the adjoint hydrodynamics, i.e. it solves

$$\partial_t \varrho = -\mathcal{D}^*(\varrho) = -\mathcal{D}(\varrho) + 2\mathcal{A}, \quad (73)$$

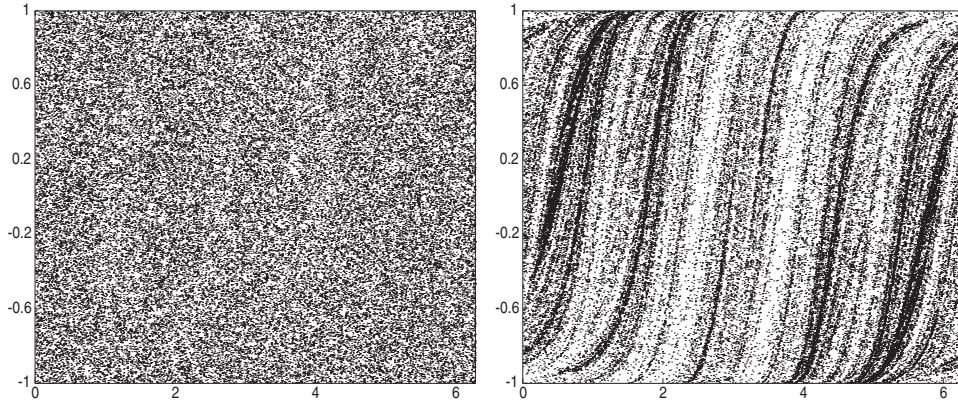
which is not merely the time reversal of the hydrodynamic equation. Being orthogonal to the thermodynamic force, the term that breaks the time symmetry,  $\mathcal{A}$ , is not the one that contributes to the entropy production. Note that equation (70) describes the adjoint hydrodynamics, while equation (73) refers to the evolution of spontaneous fluctuations.

The question arises as to which aspects of the large deviation theory described above may be verified experimentally or numerically tested in different models, particularly in deterministic chaotic systems. As observed in subsection 4.6, a direct comparison between the predictions of deterministic and stochastic models may however be problematic, although desired [157]. Therefore, in [111, 112, 158, 159] tests of the prediction of the temporal asymmetries of fluctuations are attempted. In [158] the fluctuations of the current of the nonequilibrium Lorentz gas are considered for large numbers of noninteracting particles, and they have been found to be symmetric in time. As explained in [112], this is due to the lack of interactions among the particles, and hence to the lack of correlations among them, which is more clearly manifested in the large  $N$  limit. In [111, 112, 159, 160], it is then argued that sufficiently chaotic systems of interacting particles should typically have asymmetric fluctuation paths, as long as particle correlations do not identically vanish. The role of correlations has been extensively investigated in [160].

### 5.3. Nonequilibrium entropy

Far from equilibrium the concept of entropy fades, but near equilibrium it does make sense; hence, one may ask whether the difficulty of the divergence of  $S_G$  for the NEMD systems, mentioned at the beginning of this section, could be overcome. The paper [161] has been intended to offer a possible solution, extending the ideas of [15, 162, 163].

Although there seems to be no smooth transition between the regularity of the equilibrium distributions and the singularity of the nonequilibrium ones, no matter how small the driving might be, for molecular dynamics models, numerous indicators of the state of the system



**Figure 3.** Representation of the phase space distribution, for the nonequilibrium Lorentz gas, in the Birkhoff coordinates. At equilibrium ( $\mathbf{E} = 0$ ) the distribution is uniform; for  $\mathbf{E} > 0$  it is fractal, but the fractality is not apparent to the eye, unless  $\mathbf{E}$  is sufficiently large. The left panel corresponds to  $\mathbf{E} = (0.1, 0)$ ; the right panel corresponds to  $\mathbf{E} = (1, 0)$ .

do not lose their meaning in the nonequilibrium steady states, and gradually shift from their equilibrium values, varying with the driving forces.

For instance, consider the periodic nonequilibrium Lorentz gas, whose pointlike particles obey

$$\dot{\mathbf{q}} = \mathbf{p}; \quad \dot{\mathbf{p}} = \mathbf{E} - \frac{\mathbf{p} \cdot \mathbf{E}}{\mathbf{p}^2} \mathbf{p} \tag{74}$$

and suffer from elastic collisions with scatterers, placed on a triangular lattice [164]. As  $\mathbf{E} \rightarrow 0$ , the equations of motion reduce to the equilibrium form,  $\mathbf{E} = 0$  implies zero current and small  $\mathbf{E}$  implies a small current [165]. Of course, the current alone affords only a very coarse description of the state. But other observables which do not explicitly depend on the details of the phase space probability distribution (unlike  $S_G$ ) behave the same. Indeed, the distribution itself does not seem to undergo the abrupt change which we have described, as figure 3 indicates. There, the distribution at  $\mathbf{E} = (0.1, 0)$  is *practically* uniform, indistinguishable to the eye from the equilibrium distribution. This is despite 0.1 not being a particularly small field, while 1 is definitely large (the linear regime should not extend beyond  $|\mathbf{E}| < 10^{-6}$  [88]). As discussed in [161], these are not numerical artefacts and are consistent with other works on  $N$  particle systems, such as [131].

Therefore, while  $S_G$  equals the ‘entropy’ at null  $\mathbf{E}$ , it does not even exist at  $\mathbf{E} = (\epsilon, 0)$ , for arbitrarily small  $\epsilon$ . The problem rests with the fact that  $S_G$  is defined as the mean of the log of the phase space probability distribution, which is a microscopic property of the dynamics and not a thermodynamic, observable, quantity. The attempt in [161] to overcome this difficulty relies on the possibility of expressing  $S_G$  in terms of physically relevant properties. The perspective is that of systems of many identical interacting particles, i.e. the perspective of the partial  $n$ -body distributions  $f_n$ ,  $n < N$  adopted by H S Green’s for dense liquids at equilibrium [162, 163]. First of all, the  $z$ -functions are introduced as follows:

$$\ln f_1(x_1) \equiv z_1(x_1)$$

$$\ln f_2(x_1, x_2) \equiv z_2(x_1, x_2) + z_1(x_1) + z_1(x_2)$$

$$\ln f_3(x_1, x_2, x_3) \equiv z_3(x_1, x_2, x_3) + z_2(x_1, x_2) + z_2(x_2, x_3) + z_2(x_3, x_1) + \sum_{i=1}^3 z_1(x_i)$$

...

so that the Gibbs entropy can be expressed as

$$S_G = -Nk_B \left\{ \int \frac{dx_1}{1!} f_1 z_1 + \int \frac{dx_1 dx_2}{2!} f_2 z_2 + \dots \right\}.$$

Here  $f_1$  gives the density of particles,  $f_2$  is related to the radial distribution function, which can be measured from experiment (e.g. via x-ray scattering). It is known that equilibrium properties such as energy, pressure, specific heat and temperature are obtained to a good degree of accuracy expanding the equilibrium probability density  $f$  only up to the terms that contain  $f_1$  and  $f_2$  [15, 166].

Away from equilibrium, the full phase space distribution  $f = f_N$  is singular, but  $f_n$  with  $n \ll N$  is expected to be regular (if particles interact), because  $f_n$  is a projection on a much lower dimensional phase space of the full  $f_N$ . This common-sense expectation is not guaranteed from a rigorous mathematical point of view: but it agrees, for example, with the results for  $f_2$  of Evans and Morriss, for 32 soft discs at high shear rates [15], as well as for results obtained for lattices of coupled maps [167].

Furthermore, close to equilibrium we expect the various observables to differ little from their equilibrium values; hence, we expect the above series to converge quite rapidly, even if not necessarily as rapidly as in equilibrium. Considering that tiny dimensional reduction (e.g. of order  $O(1)$  in  $10^{23}$ ) is expected for large drivings for many particle systems [131], it seems plausible that the expansion expressing  $S_G$  in terms of partial distributions, truncated before the singularities are manifested, should work. This argument is substantiated by an accurate analysis of the behaviour of the very singular Lorentz gas [161].

Therefore, for large  $N$  particle systems of physical interest, one can make the following.

**Assumption.** *Even for non-vanishing drivings, partial distributions are regular, i.e. the functions  $f_n, z_n$  are integrable functions of phase, up to some  $n_c$  such that  $1 \ll n_c \ll N$*

And, inspired by [131], one may introduce the

**Definition.** *For (interacting)  $N$ -particle systems, the Green's dimension  $N_G$  is defined by  $N_G = 2dm_G$ , if it exists, where  $m_G$  is defined by*

$$\left. \begin{aligned} \lim_{t \rightarrow \infty} |S_m(t)| < \infty, m \leq m_G \\ \lim_{t \rightarrow \infty} |S_m(t)| = \infty, m > m_G \end{aligned} \right\} S_m = \int \frac{dx_1 \dots dx_m}{m!} f_m z_m.$$

With the usual interpretation of the Kaplan–Yorke dimension  $D_{KY}$ , according to which volumes of dimension lower than  $D_{KY}$  expand, on average, while volumes of higher dimension contract, on average,  $N_G$  is meant to exploit the possibility that partial probability distributions of dimension smaller than  $D_{KY}$  should remain regular, while only those of higher dimension should be singular. As  $N$  grows, the Lyapunov spectrum should tend towards a continuous distribution, and there should then be volumes of integer dimension which neither expand nor contract on average. Then the following seems plausible.

**Conjecture.**  $N_G/D_{KY} \rightarrow 1$  as  $N \rightarrow \infty$

This may be combined with the following.

**Assumption.** *For small driving  $\mathbf{E}$  and large  $N$ ,  $|S_m(0) - S_m(\mathbf{E})|$  is small if  $m \ll m_G(\mathbf{E})$ . Moreover, if  $|S(0) - [S_1(0) + \dots + S_m(0)]| < \delta$ , there is  $M, m \leq M \ll m_G(\mathbf{E})$ , such that*

$$|S(\mathbf{E}) - [S_1(\mathbf{E}) + \dots + S_M(\mathbf{E})]| < \delta.$$

*In the  $N \rightarrow \infty$  limit,  $\delta$  can be made arbitrarily small.*

One may indeed expect  $|S_m(0) - S_m(\mathbf{E})| \sim O(\mathbf{E}^2)$ , for small  $m$  [131], and for sufficiently small  $\mathbf{E}$  that LTE is established. Then, for  $m$  of order  $O(10^{23})$ , the divergence of the contributions to the Gibbs entropy should become irrelevant, since they belong to non-observable scales. It is important to note that this ‘non-observability’ requires both very large  $N$ , otherwise LTE cannot be established, and particle interactions, otherwise  $f_N$  is merely a product of  $f_1$  distributions; hence, its projections remain singular if it is already singular.

The physical entropy is now denoted by  $S$ , without the subscript  $G$ , because  $S_G$  does not represent the physical entropy away from equilibrium. One should also note that the definition of  $S$  as a truncated series in the partial distributions  $f_n$  is not peculiar of the NEMD models, for which it has been conceived. Indeed, as the  $f_n$  can be considered in different contexts and approaches, the proposal may apply much more generally.

## 6. Conclusion

Woodcock’s efforts in the early 1970s to generate a canonical ensemble by altering the dynamics of a constant energy simulation have proven to be the first steps towards a much more ambitious goal. Thermostats have evolved from these early simple mechanisms into much more sophisticated forms: partly to maintain particular forms of equilibrium with the external universe, and partly as a means of dissipating the energy introduced by entropy-generating force fields in the NEMD technique. This evolution was initially motivated by a desire to make thermostats and the simulations that use them, more ‘realistic’. But what exactly does such an objective mean? Partly out of the realization that this question requires profound consideration, a body of theory has developed alongside the evolution in thermostats and their related simulation methods.

In this paper we have focused on the nonequilibrium simulation methods and NEMD in particular. NEMD has proven itself to be an efficient and reliable technique for studying nonequilibrium transport processes. Clearly, as with any model of a physical system, there are limits to its suitability as a general model of nonequilibrium behaviour. In its application to studying nonequilibrium steady states, these limits appear analogous to those of the ensembles of equilibrium statistical mechanics. Macroscopic properties related to the definition of the nonequilibrium steady state appear to be reliably reproduced, to the extent that one can postulate an equivalence principle (again analogous to the equivalence of equilibrium ensembles).

While conditions for this equivalence to hold have not been definitively elucidated, the validity of local thermodynamic equilibrium appears to be a good first approximation. Certainly, there is good evidence to suggest that the presence of LTE implies equivalent behaviour, while its absence implies non-thermodynamic behaviour. Coincidentally, this condition also appears to be a requirement for the association of thermodynamic quantities with the statistical values obtained from the theory—without LTE, phase space expansion rates may in no circumstance be associated with the rates of entropy production. Again, this result is strongly analogous to the equilibrium case.

Under such conditions, then, the thermostat formalism provides a theoretical framework for describing the key thermodynamic properties of the nonequilibrium steady state and has an important role to play in the development of current nonequilibrium theories. Recent years have seen important developments in the theories describing the fluctuations of nonequilibrium systems, based around notions of entropy and changes in phase space volume. It is usually constructive to review the context in which such development takes place: this additional

perspective can often highlight parallels with other topics that may provide useful insight. In the current case, there is a clear analogy between diverse approaches that have been taken in the study of nonequilibrium fluctuations, and the approaches taken to address the equilibrium requirement of ergodicity.

On the one hand is a physically motivated approach that requires the properties of the system in question to obey certain mathematical criteria (such as decay of autocorrelations). These properties are typical of thermodynamic systems: however, a broader understanding of what systems and properties meet these criteria is far from trivial. On the other hand is a mathematically motivated approach that presents a set of mathematical conditions under which a nonequilibrium system's fluctuations match those observed from (computational) experiment. This elegant approach allows valuable insight into the nature of nonequilibrium fluctuations: however, many physical systems that obey the fluctuation relations are known *not* to meet these mathematical conditions.

At the beginning of this paper, we identified a dichotomy between the mathematical notion of ergodicity and the physically motivated approach of Khinchin. It is apparent that there is more than a passing similarity between this dichotomy and that of the nonequilibrium fluctuation theorems. Interestingly, the ergodic dichotomy remains unresolved. While ergodicity seems an unnecessarily strong condition, whose benefits cannot be reaped on the timescales over which we observe thermodynamic systems, the bottom line is that it remains a linchpin of modern statistical mechanics because, as a model, it proves to be physically suitable. The more physically realistic model instigated by Khinchin almost certainly provides a more likely explanation for ergodicity as we observe it, but is at present too limited in scope to support a comprehensive statistical mechanical theory. Neither situation is entirely satisfactory—and yet we would be much poorer without either approach. On this final point, the analogy with the nonequilibrium theories is excellent.

## Acknowledgments

We would like to thank Denis Evans, Stefano Lepri, Bill Hoover and Angelo Vulpiani for helpful discussions and feedback. We also thank the anonymous referee, who raised the example mentioned in the final paragraph of section 4.5. OGJ thanks the ISI Foundation and the Australian Research Council for financial support throughout this work. LR acknowledges the contribution of the European Research Council within the 7th Framework Programme (FP7) of the European Community (EC), ERC Grant Agreement no. 202680. The EC is not responsible for any use that might be made of the data appearing herein.

## References

- [1] Gallavotti G 1995 Ergodicity, ensembles, irreversibility in Boltzmann and beyond *J. Stat. Phys.* **78** 1571
- [2] Gallavotti G 2000 *Statistical Mechanics: A Short Treatise* (Berlin: Springer)
- [3] Castiglione P, Falcioni M, Lesne A and Vulpiani A 2008 *Chaos and Coarse Graining in Statistical Mechanics* (Cambridge: Cambridge University Press)
- [4] Fermi E 1923 Dimostrazione che in generale un sistema meccanico normale è quasi ergodico *Nuovo Cimento* **25** 267
- [5] Fermi E, Pasta J and Ulam S 1955 *Studies of Nonlinear Problems (Los Alamos Report LA-1940)* (reprinted in Segre E (ed) 1965 *Collected Papers of Enrico Fermi* (Chicago: University of Chicago Press))
- [6] Gallavotti E G 2008 *The Fermi-Pasta-Ulam Problem (Lecture Notes in Physics vol 728)* (Berlin: Springer)
- [7] Khinchin A I 1949 *Mathematical Foundations of Statistical Mechanics* (New York: Dover)
- [8] Mazur P and van der Linden J 1963 Asymptotic form of the structure function for real systems *J. Math. Phys.* **4** 271

- [9] Kubo R 1957 Statistical-mechanical theory of irreversible processes: I. General theory and simple applications to magnetic and conduction problems *J. Phys. Soc. Japan* **12** 570
- [10] Lee M H 2001 Ergodic theory, infinite products, and long time behavior in Hermitian models *Phys. Rev. Lett.* **87** 250601
- [11] Lee M H 2002 Ergodicity in simple and not so simple systems and Kubo's condition *Physica A* **314** 583
- [12] Lee M H 2006 Why does Boltzmann's ergodic hypothesis work and when does it fail *Physica A* **365** 150
- [13] Lee M H 2007 Why irreversibility is not a sufficient condition for ergodicity *Phys. Rev. Lett.* **98** 190601
- [14] Lee M H 2007 Birkhoff's theorem, many-body response functions, and the ergodic condition *Phys. Rev. Lett.* **98** 110403
- [15] Evans D J and Morriss G P 2008 *Statistical Mechanics of Nonequilibrium Liquids* (Cambridge: Cambridge University Press)
- [16] Bhatia S K and Nicholson D 2003 Molecular transport in nanopores *J. Chem. Phys.* **119** 1719
- [17] Hoover W G 1991 *Computational Statistical Mechanics* (Amsterdam: Elsevier)
- [18] Sarman S, Evans D J and Cummings P T 1998 Recent developments in non-Newtonian molecular dynamics *Phys. Rep.* **305** 1
- [19] Gallavotti G and Presutti E 2009 Thermodynamic limit of isoenergetic and Hamiltonian Thermostats arXiv:0903.3316
- [20] Müller-Plathe F 1999 Reversing the perturbation in nonequilibrium molecular dynamics: An easy way to calculate the shear viscosity of fluids *Phys. Rev. E* **59** 4894
- [21] Woodcock L V 1971 Isothermal molecular dynamics calculations for liquid salts *Chem. Phys. Lett.* **10** 257
- [22] Andersen H C 1980 Molecular dynamics simulations at constant pressure and/or temperature *J. Chem. Phys.* **72** 2384
- [23] Haile J M and Gupta S 1983 Extensions of the molecular dynamics simulation method. II. Isothermal systems *J. Chem. Phys.* **79** 3067
- [24] Frenkel D and Smit B 1996 *Understanding Molecular Simulation* (San Diego: Academic)
- [25] Evans D J, Cohen E G D and Morriss G P 1993 Probability of second law violations in shearing steady flows *Phys. Rev. Lett.* **71** 2401
- [26] Lanczos C 1979 *The Variational Principles of Mechanics* (New York: Dover)
- [27] Gauss K F 1829 Über ein neues allgemeines Grundgesetz der Mechanik *J. Reine Angew. Math.* **4** 232
- [28] Evans D J, Hoover W J, Failor B H, Moran B and Ladd A J C 1983 Nonequilibrium molecular dynamics via Gauss' principle of least constraint *Phys. Rev. A* **28** 1016
- [29] Hoover W J, Ladd A J C and Moran B 1982 High-strain-rate plastic flow studied via nonequilibrium molecular dynamics *Phys. Rev. Lett.* **48** 1818
- [30] Evans D J 1983 Computer 'experiment' for nonlinear thermodynamics of Couette flow *J. Chem. Phys.* **78** 3297
- [31] Dettmann C P and Morriss G P 1996 Hamiltonian formulation of the Gaussian isokinetic thermostat *Phys. Rev. E* **54** 2495
- [32] Dettmann C P 1999 Hamiltonian for a restricted isoenergetic thermostat *Phys. Rev. E* **60** 7576
- [33] Klages R 2003 Microscopic chaos and transport in thermostated dynamical systems arXiv:nlin/0309069v1 [nlin.CD]
- [34] Nosé S 1984 A unified formulation of the constant temperature molecular-dynamics methods *J. Chem. Phys.* **81** 511
- [35] Nosé S 1984 A molecular-dynamics method for the simulations in the canonical ensemble *Mol. Phys.* **52** 255
- [36] Hoover W G 1985 Canonical dynamics: equilibrium phase-space distributions *Phys. Rev. A* **31** 1695
- [37] Jellinek J 1989 Dynamics for nonconservative systems: ergodicity beyond the microcanonical ensemble *J. Phys. Chem.* **92** 3163
- [38] Bulgac A and Kusnezov D 1990 Canonical ensemble average from pseudomicrocanonical dynamics *Phys. Rev. A* **42** 5045
- [39] Winkler R G 1992 Extended phase-space isothermal molecular dynamics: canonical harmonic oscillator *Phys. Rev. A* **45** 2250
- [40] Bulgac A and Kusnezov D 1990 Classical Limit for Lie Algebras *Ann. Phys.* **199** 187
- [41] Rugh H H 1997 A dynamical approach to temperature *Phys. Rev. Lett.* **78** 772
- [42] Rugh H H 1998 Geometric, dynamical approach to thermodynamics *J. Phys. A: Math. Gen.* **31** 7761
- [43] Jepps O G, Ayton G and Evans D J 2000 Microscopic expressions for the thermodynamic temperature *Phys. Rev. E* **62** 4757
- [44] Rickayzen G and Powles J G 2001 Temperature in the classical microcanonical ensemble *J. Chem. Phys.* **114** 4333
- [45] Butler B D, Ayton G, Jepps O G and Evans D J 1998 Configurational temperature: verification of Monte Carlo simulations *J. Chem. Phys.* **109** 6519

- [46] Delhommelle J and Evans D J 2001 Configurational temperature thermostat for fluids undergoing shear flow: application to liquid chlorine *Mol. Phys.* **99** 1825
- [47] Lue L, Jepps O G, Delhommelle J and Evans D J 2002 Configurational thermostats for molecular systems *Mol. Phys.* **100** 2387
- [48] Braga C and Travis K P 2005 A configurational temperature Nosé-Hoover thermostat *J. Chem. Phys.* **123** 134101
- [49] Braga C and Travis K P 2006 Configurational constant pressure molecular dynamics *J. Chem. Phys.* **124** 104102
- [50] Travis K P and Braga C 2006 Configurational temperature and pressure molecular dynamics: Review of current methodology and applications to the shear flow of a simple fluid *Mol. Phys.* **104** 3735
- [51] Travis K P and Braga C 2008 Configurational temperature control for atomic and molecular systems *J. Chem. Phys.* **128** 014111
- [52] Jepps O G 2001 Microscopic expressions for the thermodynamic temperature *PhD Thesis* Australian National University, Canberra
- [53] Samoletov A A, Dettmann C P and Chaplain M A J 2007 Thermostats for 'slow' configurational modes *J. Stat. Phys.* **128** 1321
- [54] Ayton G, Jepps O G and Evans D J 1999 On the validity of Fourier's law in systems with spatially varying strain rates *Mol. Phys.* **96** 915
- [55] Morriss G P and Rondoni L 1999 Definition of temperature in equilibrium and nonequilibrium systems *Phys. Rev. E* **59** R5
- [56] Hoover W G, Aoki K, Hoover C G and De Groot S V 2004 Time-reversible deterministic thermostats *Physica D* **187** 253
- [57] Bright J N and Evans D J 2005 New observations regarding deterministic, time-reversible thermostats and Gauss's principle of least constraint *J. Chem. Phys.* **122** 194106
- [58] Travis K P, Davis P J and Evans D J 1995 Computer simulation algorithms for molecules undergoing planar Couette flow: A nonequilibrium molecular dynamics study *J. Chem. Phys.* **103** 1109
- [59] Travis K P, Davis P J and Evans D J 1995 Thermostats for molecular fluids undergoing shear flow: application to liquid chlorine *J. Chem. Phys.* **103** 10638
- [60] Delhommelle J, Petracic J and Evans D J 2003 On the effects of assuming flow profiles in nonequilibrium simulations *J. Chem. Phys.* **119** 11005
- [61] Delhommelle J 2005 Should 'lane formation' occur systematically in driven liquids and colloids? *Phys. Rev. E* **71** 016705
- [62] Delhommelle J, Petracic J and Evans D J 2003 Reexamination of string phase and shear thickening in simple fluids *Phys. Rev. E* **68** 031201
- [63] Hoover W G and Hoover C G 2009 Nonlinear stresses and temperatures in transient adiabatic and shear flows via nonequilibrium molecular dynamics: three definitions of temperature *Phys. Rev. E* **79** 046705
- [64] Dirac P A M 1950 Generalized Hamiltonian dynamics *Can. J. Math.* **2** 129
- [65] Dirac P A M 1958 Generalized Hamiltonian dynamics *Proc. R. Soc. A* **246** 326
- [66] Bulgac A and Roman H E 2005 Equations of motion for polymer chains in a thermostat *New J. Phys.* **7** 2
- [67] Rondoni L 2002 Deterministic thermostats and fluctuation relations *Dynamics of Dissipation* ed P Garbaczewski and R Olkiewicz (Berlin: Springer)
- [68] Müller-Plathe F and Reith D 1999 Cause and effect reversed in non-equilibrium molecular dynamics: an easy route to transport coefficients *Comput. Theor. Polym. Sci.* **9** 203
- [69] Bordat P and Müller-Plathe F 2002 The shear viscosity of molecular fluids: a calculation by reverse nonequilibrium molecular dynamics *J. Chem. Phys.* **116** 3362
- [70] Benettin G and Rondoni L 2001 A new model for the transport of particles in a thermostatted system *Math. Phys. Electron. J.* **7** 3
- [71] Chernov N I and Lebowitz J L 1997 Stationary nonequilibrium states in boundary-driven Hamiltonian systems: shear flow *J. Stat. Phys.* **86** 953
- [72] Allen M P and Tildesley D J 1987 *Computer Simulation of Liquids* (Oxford: Clarendon)
- [73] Holian B L, Lomdahl P S and Zhou S J 1997 Fracture simulations via large-scale nonequilibrium molecular dynamic *Physica A* **240** 340
- [74] Kröger M 2005 *Models for Polymeric and Anisotropic Liquids (Lecture Notes in Physics vol 675)* (New York: Springer)
- [75] Rapaport D C 2004 *The Art of Molecular Dynamics Simulation* (Cambridge: Cambridge University Press)
- [76] Rondoni L and Mejia-Monasterio C 2007 Fluctuation theorems in nonequilibrium statistical mechanics *Nonlinearity* **20** R1



- [77] Marconi U M B, Puglisi A, Rondoni L and Vulpiani A 2008 Fluctuation-dissipation: response theory in statistical physics *Phys. Rep.* **461** 111
- [78] Ruelle D 2004 Conversations on nonequilibrium physics with an extraterrestrial *Phys. Today* **57** 48
- [79] Evans D J and Morriss G P 1984 Equilibrium time correlation functions under Gaussian isothermal dynamics *Chem. Phys.* **87** 451
- [80] Evans D J and Holian B L 1985 The Nosé–Hoover thermostat *J. Chem. Phys.* **83** 4069
- [81] Evans D J and Sarman S 1993 Equivalence of thermostatted nonlinear responses *Phys. Rev. E* **48** 65
- [82] Evans D J 1982 Homogeneous NEMD algorithm for thermal conductivity-application of non-canonical linear response theory *Phys. Lett. A* **91** 457
- [83] Gillan M J and Dixon M 1983 The calculation of thermal conductivities by perturbed molecular dynamics simulation *J. Phys. C: Solid State Phys.* **16** 869
- [84] Kawasaki K and Gunton J D 1973 Theory of nonlinear transport processes: nonlinear shear viscosity and normal stress effects *Phys. Rev. A* **8** 2048
- [85] Yamada T and Kawasaki K 1975 Applications of mode coupling theory to the nonlinear stress tensor in fluids *Prog. Theor. Phys.* **53** 111
- [86] Yamada T and Kawasaki K 1975 Contributions to statistical mechanics far from equilibrium: IV. Improved and simplified treatment of non-steady states *Prog. Theor. Phys.* **53** 437
- [87] Gallavotti G and Presutti E 2009 Nonequilibrium, thermostats and thermodynamic limit arXiv:0905.3150v1
- [88] Cohen E G D and Rondoni L 1998 Note on phase space contraction and entropy production in thermostatted Hamiltonian systems *Chaos* **8** 357
- [89] Ruelle D 2000 A remark on the equivalence of isokinetic and isoenergetic thermostats in the thermodynamic limit *J. Stat. Phys.* **100** 757
- [90] Gallavotti G 1997 Dynamical ensemble equivalence in fluid mechanics *Physica D* **105** 163
- [91] Rondoni L and Segre E 1999 Fluctuations in two-dimensional reversibly damped turbulence *Nonlinearity* **12** 1471
- [92] Gallavotti G, Rondoni L and Segre E 2004 Lyapunov spectra and nonequilibrium ensembles equivalence in 2D fluid mechanics *Physica D* **187** 338
- [93] Hertel P and Thirring W 1971 A soluble model for a system with negative specific heat *Ann. Phys.* **63** 520
- [94] Kiessling M K-H and Lebowitz J L 1997 The micro-canonical point vortex ensemble: beyond equivalence *Lett. Math. Phys.* **42** 43
- [95] Dauxois T, Holdsworth P and Ruffo S 2000 Violation of ensemble equivalence in the antiferromagnetic mean-field XY model *Eur. Phys. J. B* **16** 659
- [96] Bonetto F, Daems D, Lebowitz J L and Ricci V 2002 Properties of stationary nonequilibrium states in the thermostatted periodic Lorentz gas: the multiparticle system *Phys. Rev. E* **65** 051204
- [97] Jepps O G and Rondoni L 2006 Thermodynamics and complexity of simple transport phenomena *J. Phys. A: Math. Gen.* **39** 1311
- [98] Klages R 2007 *Microscopic Chaos, Fractals and Transport in Nonequilibrium Statistical Mechanics* (Singapore: World Scientific)
- [99] Rondoni L and Cohen E G D 1998 Orbital measures in non-equilibrium statistical mechanics: the Onsager relations *Nonlinearity* **11** 1395
- [100] Rondoni L and Cohen E G D 2002 On some derivations of irreversible thermodynamics from dynamical systems theory *Physica D* **168–169** 341
- [101] Petracic J and Jepps O G 2003 Homogeneous shear flow of hard-sphere fluid: analytic solutions *Phys. Rev. E* **67** 021105
- [102] Jepps O G and Petracic J 2004 Colour conductivity of hard spheres *Mol. Phys.* **102** 513
- [103] Heffelfinger G S and van Swol F 1994 Diffusion in Lennard–Jones fluids using dual control-volume grand-canonical molecular dynamics simulation (DCV-GCMD) *J. Chem. Phys.* **100** 7548
- [104] Rieder Z, Lebowitz J L and Lieb E 1967 Properties of a harmonic crystal in a stationary nonequilibrium state *J. Math. Phys.* **8** 1073
- [105] Rondoni L and Cohen E G D 2000 Gibbs entropy and irreversible thermodynamics *Nonlinearity* **13** 1905
- [106] Lebowitz J L, Goldstein S and Sinai Y G 1998 Remark on the (non)convergence of densities in mixing systems *Chaos* **8** 393
- [107] Goldstein S, Kipnis C and Ianiro N 1985 Stationary states for a mechanical system with stochastic boundary conditions *J. Stat. Phys.* **41** 915
- [108] Kurchan J 2007 Nonequilibrium work relations *J. Stat. Mech.: Theory Exp.* **2007** P07005 (arXiv:cond-mat/0511073v1)
- [109] Gallavotti G and Cohen E G D 1995 Dynamical ensembles in stationary states *J. Stat. Phys.* **80** 931

- [110] Attard P 2009 Statistical mechanical theory for non-equilibrium systems: IX. Stochastic molecular dynamics *J. Chem. Phys.* **130** 194113
- [111] Giberti C, Rondoni L and Vernia C 2006 Asymmetric fluctuation–relaxation paths in FPU models *Physica A* **365** 229
- [112] Giberti C, Rondoni L and Vernia C 2007 Temporal asymmetry of fluctuations in the nonequilibrium FPU model *Physica D* **228** 64
- [113] Bonaldi M, Conti L, de Gregorio P, Rondoni L, Vedovato G, Vinante A, Bignotto M, Cerdonio M, Falferi P and Liguori N *et al* 2009 Nonequilibrium steady-state fluctuations In actively cooled resonators *Phys. Rev. Lett.* **103** P010601
- [114] de Gregorio P, Rondoni L, Bonaldi M and Conti L 2009 Harmonic damped oscillators with feedback: a Langevin study *J. Stat. Mech.: Theory Exp.* P10016
- [115] Evans D J and Searles D J 2002 The fluctuation theorem *Adv. Phys.* **52** 1529
- [116] Parry W 1986 Synchronisation of canonical measures for hyperbolic attractors *Commun. Math. Phys.* **106** 267
- [117] Vance W N 1992 Unstable periodic orbits, transport properties of nonequilibrium steady states *Phys. Rev. Lett.* **69** 1356
- [118] Evans D J and Searles D J 1994 Equilibrium microstates which generate second law violating steady states *Phys. Rev. E* **50** 1645
- [119] Evans D J and Searles D J 1995 Steady states, invariant measures, response theory *Phys. Rev. E* **52** 5839
- [120] Searles D J, Ayton G and Evans D J 2000 Generalised fluctuation formula *AIP Conf. Ser.* **519** 271
- [121] Searles D J and Evans D J 2000 Ensemble dependence of the transient fluctuation theorem *J. Chem. Phys.* **113** 3503
- [122] Williams S R, Searles D J and Evans D J 2004 Thermostat invariance of the transient fluctuation theorem *Phys. Rev. E* **70** 066113
- [123] Williams S R, Searles D J and Evans D J 2006 Numerical study of the steady state fluctuation relations far from equilibrium *J. Chem. Phys.* **124** 194102
- [124] Wang G M, Sevcik E M, Mittag E, Searles D J and Evans D J 2002 Experimental demonstration of violations of the second law of thermodynamics for small systems and short time scales *Phys. Rev. Lett.* **89** 050601
- [125] Gallavotti G and Cohen E G D 1995 Dynamical ensembles in nonequilibrium statistical mechanics *Phys. Rev. Lett.* **94** 2694
- [126] Gallavotti G 1995 Reversible Anosov diffeomorphisms, large deviations *Math. Phys. Electron. J.* **1** 1
- [127] Gallavotti G 2004 Fluctuation theorem revisited arXiv:cond-mat/0402676v1
- [128] Searles D J, Rondoni L and Evans D J 2007 The steady state fluctuation relation for the dissipation function *J. Stat. Phys.* **128** 1337
- [129] Birkhoff G D 1931 Proof of the ergodic theorem *Proc. Natl Acad. Sci. USA* **17** 656
- [130] Sinai Ya G 1977 *Lectures in Ergodic Theory (Lecture Notes in Mathematics)* (Princeton, NJ: Princeton University Press)
- [131] Evans D J, Cohen E G D, Searles D J and Bonetto F 2000 Note on the Kaplan–Yorke dimension and linear transport coefficients *J. Stat. Phys.* **101** 17
- [132] Gallavotti G 1996 Extension of Onsager’s reciprocity to large fields, the chaotic hypothesis *Phys. Rev. Lett.* **77** 4334
- [133] Ruelle D 1997 Differentiation of SRB states *Commun. Math. Phys.* **187** 227
- [134] Gallavotti G and Ruelle D 1997 SRB states and nonequilibrium statistical mechanics close to equilibrium *Commun. Math. Phys.* **190** 279
- [135] Evans D J, Searles D J and Rondoni L 2005 On the application of the Gallavotti–Cohen fluctuation relation to thermostatted steady states near equilibrium *Phys. Rev. E* **71** 056120
- [136] Rondoni L, Tél T and Vollmer J 2000 Fluctuation theorems for entropy production in open systems *Phys. Rev. E* **61** R4679
- [137] Maes C, Redig F and Verschuere M 2001 From global to local fluctuation theorems *Mosc. Math. J.* **1** 421
- [138] Gallavotti G 1999 A local fluctuation theorem *Physica A* **263** 39
- [139] Gallavotti G 1999 Fluctuation patterns and conditional reversibility in nonequilibrium system *Ann. Inst. H. Poincaré* **70** 429
- [140] Gallavotti G 2002 Large deviations, fluctuation theorem, Onsager–Machlup theory in nonequilibrium statistical mechanics <http://ipparco.roma1.infn.it/>
- [141] Dolowschiak M and Kovacs Z 2005 Fluctuation formula in the Nosé–Hoover thermostatted Lorentz gas *Phys. Rev. E* **71** 025202
- [142] Zamponi F, Ruocco G and Angelani L 2004 Fluctuations of entropy production in the isokinetic ensemble *J. Stat. Phys.* **115** 1655
- [143] Evans D J 2003 A non-equilibrium free energy theorem for deterministic systems *Mol. Phys.* **101** 1551

- [144] Bonetto F, Gallavotti G, Giuliani A and Zamponi F 2006 Chaotic hypothesis, fluctuation theorem, singularities *J. Stat. Phys.* **123** 39
- [145] van Zon R and Cohen E G D 2003 Stationary and transient work-fluctuation theorems for a dragged Brownian particle *Phys. Rev. E* **67** 046102
- [146] Evans D J 2009 private communication
- [147] Evans D J, Searles D J and Williams S R 2008 On the fluctuation theorem for the dissipation function and its connection with response theory *J. Chem. Phys.* **128** 014504
- [148] Evans D J, Searles D J and Williams S R 2009 Dissipation and the relaxation to equilibrium *J. Stat. Mech.: Theory Exp.* **07** P07029
- [149] Evans D J, Searles D J and Williams S R 2009 A simple mathematical proof of Boltzmann's equal *a priori* probability hypothesis *diffusion-fundamentals.org* **11** 1–18
- [150] Onsager L and Machlup S 1953 Fluctuations and irreversible processes *Phys. Rev.* **91** 1505
- [151] Machlup S and Onsager L 1953 Fluctuations and irreversible process: II. Systems with kinetic energy *Phys. Rev.* **91** 1512
- [152] Derrida B, Lebowitz J L and Speer E R 2002 Large deviation of the density profile in the steady state of the open symmetric simple exclusion process *J. Stat. Phys.* **107** 599
- [153] Bertini L, De Sole A, Gabrielli D, Jona-Lasinio G and Landim C 2001 Fluctuations in stationary nonequilibrium states of irreversible processes *Phys. Rev. Lett.* **87** 040601
- [154] Bertini L, De Sole A, Gabrielli D, Jona-Lasinio G and Landim C 2002 Macroscopic fluctuation theory for stationary non-equilibrium states *J. Stat. Phys.* **107** 635
- [155] Bertini L, De Sole A, Gabrielli D, Jona-Lasinio G and Landim C 2005 Current fluctuations in stochastic lattice gases *Phys. Rev. Lett.* **94** 030601
- [156] Sengers J V, Dorfman J R and Kirkpatrick T R 1994 Generic long-range correlations in molecular fluids *Annu. Rev. Phys. Chem.* **45** 213
- [157] Bodineau T and Derrida B 2004 Current fluctuations in nonequilibrium diffusive systems: an additivity principle *Phys. Rev. Lett.* **92** 180601
- [158] Gamba A and Rondoni L 2004 Current fluctuations in the nonequilibrium Lorentz gas *Physica A* **340** 274
- [159] Paneni C, Searles D J and Rondoni L 2006 Temporal asymmetry of fluctuations in nonequilibrium states *J. Chem. Phys.* **124** 114109
- [160] Paneni C, Searles D J and Rondoni L 2008 Temporal asymmetry of fluctuations in nonequilibrium steady states: fluctuation paths, correlation functions and nonlinear response theory *J. Chem. Phys.* **128** 164515
- [161] Evans D J and Rondoni L 2002 Comments on the entropy of nonequilibrium states *J. Stat. Phys.* **109** 895
- [162] Green H S 1952 *The Molecular Theory of Fluids* (Amsterdam: North-Holland)
- [163] Raveché H J 1971 Entropy and molecular correlation functions in open systems: I. Derivation *J. Chem. Phys.* **55** 2242
- [164] Lloyd J, Niemeyer M, Rondoni L and Morriss G P 1995 The nonequilibrium Lorentz gas *Chaos* **5** 536
- [165] Chernov N I, Eyink G L, Lebowitz J L and Sinai Ya G 1993 Steady state electric conductivity in the periodic Lorentz gas *Commun. Math. Phys.* **154** 569
- [166] Baranyai A and Evans D J 1990 Three-particle contribution to the configurational entropy of simple fluids *Phys. Rev. A* **42** 849
- [167] Bonetto F, Kupiainen A and Lebowitz J L 2005 Absolute continuity of projected SRB measures of coupled Arnold cat map lattices *Ergodic Theory Dyn. Syst.* **25** 59
- [168] Evans D J and Morriss G P 1984 Nonlinear-response theory for steady planar Couette flow *Phys. Rev. A* **30** 1528