Detailed fluctuation theorem for mesoscopic modeling

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

Detailed fluctuation theorem for mesoscopic modeling

E. A. J. F. Peters*

Department of Chemical Engineering, Technische Universiteit Eindhoven, Den Dolech 2, Postbus 513,
5600 MB Eindhoven, The Netherlands

The detailed fluctuation theorem is derived. The basic assumptions are phase space incompressibility (Liouville's theorem) and time reversibility on the microscopic level. The theorem relates the conditional probability to end up in a mesoscopic state Γ_B at time t_B, starting from Γ_A at time t_A, to the time-reversed process. The ratio of these two probability densities is related to the entropy difference of the two mesoscopic states. The fluctuation theorem remains valid even far from equilibrium as long as the local equilibrium condition is obeyed. It is shown that the theorem imposes constraints on the form mesoscopic equations can take. For stochastic differential equations a generalized kinetic form is derived. The fluctuation theorem can be used to derive thermodynamically consistent simulation techniques. At the end of this paper the relation with the GENERIC formalism is discussed.

I. INTRODUCTION

The (steady-state) fluctuation theorem is a relation for the probability density of measuring a certain average entropy production in a nonequilibrium steady-state experiment. It states that

$$\frac{p(\bar{\sigma})}{p(-\bar{\sigma})} = \exp(\bar{\sigma}\Delta t/k_B) \qquad (1)$$

for Δt large enough. Here $p(\bar{\sigma})$ is the probability density of measuring an average entropy production $\bar{\sigma}$ over a time Δt. The fluctuation theorem is valid in the limit $\Delta t \rightarrow 0$. This theorem is special because it is believed to be valid not only for the near-equilibrium situations, but also in the far from equilibrium stationary situations. It illustrates that the second law of thermodynamics is sometimes violated, because the probability of negative entropy production is finite when there is positive entropy production. On the other hand, it illustrates that the probability for measuring positive entropy production is exponentially more likely than negative entropy production.

Equation (1) was first found by Evans, Cohen, and Morriss [1] on the grounds of theoretical considerations and confirmed to be obeyed by a simulation of thermostated particles in a shear flow. It was put on a more rigorous footing by Evans and Searles [2] and Gallavotti and Cohen [3,4]. The theoretical considerations in these papers, such as the Sinai-Rowen-Bowen measures, are taken from the field of chaotic dynamics (see [5]). A review on the fluctuation theorem, including numerical and experimental verification, can be found in [6].

Besides the steady-state fluctuation theorem also transient fluctuation theorems have been derived [6–10]. In this paper I also give a derivation of a transient fluctuation theorem. My approach is closest related to those of Jarzynski [9] and Maes and Netočný [8]. The detailed fluctuation theorem derived here gives a relation for the conditional probability densities to end up in a (mesoscopic) state Γ_B at time t_B when the system starts out in state Γ_A at time t_A. The probability density is related to that of the time-reversed process. This kind of relation can be called a detailed fluctuation theorem since it generalizes the detailed balance condition.

The derivation is straightforward. The main ingredients are Liouville's theorem—i.e., conservation of microscopic phase space volume during time evolution—and microreversibility. A mesoscopic state is defined as an ensemble of microscopic states. The ratio of the conditional probabilities for the forward and time-reversed process is related to the entropy difference of the initial and final states. The definition we use for the mesoscopic entropy of a mesoscopic state is essentially a Boltzmann entropy. The same definition is used in, for example, the projection operator theory of Zwanzig [11] and the GENERIC formalism of Grmela and Öttinger [12,13] (GENERIC is the acronym for the general equation for the nonequilibrium reversible-irreversible coupling).

It will be argued that in coarse-grained theories the detailed fluctuation theorem remains valid, although Liouville's theorem itself is no longer valid. Next, I will show that this theorem imposes constraints on the equations that describe the mesoscopic dynamics. Specifically I will show that, when using stochastic differential equations for the modeling, the constraints give equations of the form of the GENERIC formalism. Since the detailed fluctuation theorem relates mesoscopic states separated in time, it can be a valuable tool for developing simulation algorithms.

One of the goals of the present paper is to create a bridge between two approaches to nonequilibrium thermodynamics: the fluctuation-theorem approach and the GENERIC approach. Two communities seem to have developed similar results without much interaction. The connection of the detailed fluctuation theorem to the GENERIC formalism will be discussed more extensively in the final discussion section.

---

*Electronic address: e.a.j.f.peters@tue.nl

II. DERIVATION OF THE FLUCTUATION THEOREM

Consider a mesoscopic system described by a states Γ. The phase space is of too low a dimension to fully specify the microscopic state of the system. The mesoscopic state Γ labels a subspace in the microscopic phase space. As an example let us choose one component Γ_i to represent the total momentum of a blob of particles. The corresponding microscopic subspace is the union of all points in microscopic phase space for which the total momenta of the blob have the specified total momentum Γ_i. The full mesoscopic state $\Gamma = (\Gamma_1, \ldots, \Gamma_n)$ corresponds to the intersection of the microscopic subspaces corresponding to the specific values for $\Gamma_1, \ldots, \Gamma_n$. If one considers a neighborhood around state Γ in mesoscopic space, this corresponds to a region in microscopic space. The volume in microscopic space of this region is

$$\delta V = \delta \Gamma \, \exp[S(\Gamma)/k_B], \qquad (2)$$

where $\delta \Gamma$ denotes the volume of the the mesoscopic space that is occupied. This equation defines the entropy $S(\Gamma)$ of the mesoscopic state Γ. The constant k_B is the Boltzmann constant. Note that the entropy as defined in this way is, partly, a coordinate-dependent definition. If one changes the parametrization of the mesoscopic space from Γ to Γ', the entropy will transform as

$$S'(\Gamma') = S(\Gamma) + k_B \ln \det\left(\frac{\partial \Gamma'}{\partial \Gamma}\right). \qquad (3)$$

The notion of a mesoscopic state is useful when the evolution of the mesoscopic quantity is a slow variable compared to the time evolution within the microscopic region defined by it. In this case, during a short enough time interval, much of the microscopic space corresponding to Γ has been explored while the value of Γ itself has changed only a little. Assuming ergodicity in the microscopic subspace, the mesoscopic state is well described by assuming that all microscopic states in Γ are equally likely to occur.

Now let us consider two times t_A and t_B and two states Γ_A and Γ_B. Let us assume that at time t_A the system is prepared in an ensemble of microstates corresponding to a neighborhood of volume $\delta \Gamma_A$ around the mesoscopic state Γ_A. All points in the microscopic subspace will be assumed to have the same statistical weight. The total volume of microscopic phase space is $\delta \Gamma_A \exp[S(\Gamma_A)/k_B]$. Let $p(\Gamma_B, t_B | \Gamma_A, t_A)$ be the conditional probability density that a microstate, which, at time t_A, is an element of the ensemble corresponding to mesostate Γ_A, after evolution to time t_B ends up in state Γ_B. The conditional probability $p(\Gamma_B, t_B | \Gamma_A, t_A) \delta \Gamma_B$ is the fraction of the original microspace volume—i.e., $\delta \Gamma_A \times \exp[S(\Gamma_A)/k_B]$—that ends up in the region in microscopic space defined by a neighborhood around the mesoscopic phase Γ_B with volume $\delta \Gamma_B$. If we define $\delta V_{\text{overlap}}$ as the overlap between the microscopic volume of the volume corresponding to the region with volume $\delta \Gamma_A$ around Γ_A (evolved to time t_B) and the region around Γ_B, then
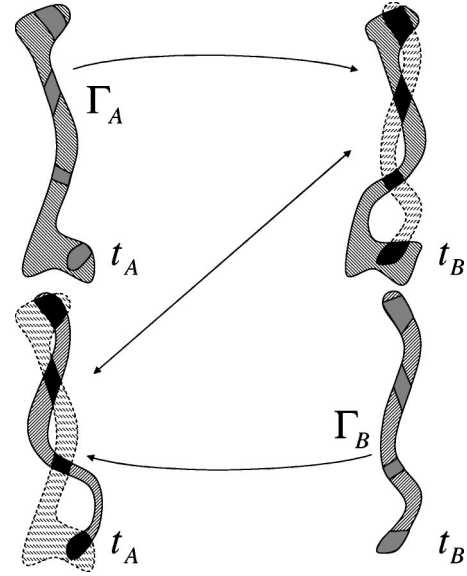

FIG. 1. The upper picture shows the microscopic time evolution of the region of microstates corresponding to a neighborhood around a mesoscopic state Γ_A. The solid black regions are the overlap of the evolved state with (a neighborhood around) the mesoscopic state Γ_B. The gray subdomains of the original state Γ_A are the corresponding states tracked backward to time t_A. The lower picture shows the microscopic evolution of Γ_B tracked backward in time to t_A. The black regions are the overlap between this state and state Γ_A. From microreversibility we conclude that the gray and black regions in the upper and lower figures at a fixed time are the same parts of microscopic space. From Liouville's theorem we conclude that the volumes of the gray and black regions on the left- and right-hand sides are equal. Combining these two observations gives that the volumes of the black regions in the lower left figure and the upper-right figure are equal. This elementary observation results into the fluctuation theorem.

$$p(\Gamma_B, t_B | \Gamma_A, t_A) \delta \Gamma_B \approx \frac{\delta V_{\text{overlap}}}{\delta \Gamma_A \exp[S(\Gamma_A)/k_B]}. \qquad (4)$$

Here the equality is approached when $\delta \Gamma_A$ and $\delta \Gamma_B$ approach zero. Alternatively we can consider state Γ_B at time t_B and determine to which subspace of microscopic space it corresponds at time t_A. The overlap between this subspace and the one corresponding to Γ_A is characterized by the conditional probability density $p(\Gamma_A, t_A | \Gamma_B, t_B)$:

$$p(\Gamma_A, t_A | \Gamma_B, t_B) \delta \Gamma_A \approx \frac{\delta \widetilde{V}_{\text{overlap}}}{\delta \Gamma_B \exp[S(\Gamma_B)/k_B]}. \qquad (5)$$

Here $\delta \widetilde{V}_{\text{overlap}}$ is the volume of the overlap region. The overlap region of (the region around) Γ_A evolved to t_B with (the region around) Γ_B and that of Γ_B evolved to t_A with Γ_A correspond to each other. This is illustrated in Fig. 1. Due to Liouville's theorem—i.e., the fact that the volume of microscopic phase space is conserved—the microscopic volume of the overlap region does not change with time. This means that in both expressions for $p(\Gamma_B, t_B | \Gamma_A, t_A)$ and $p(\Gamma_A, t_A | \Gamma_B, t_B)$ the overlap volumes are equal. From the equality $\delta \widetilde{V}_{\text{overlap}} = \delta V_{\text{overlap}}$ one finds that

$$p(\Gamma_B, t_B | \Gamma_A, t_A) = \exp\{[S(\Gamma_B) - S(\Gamma_A)]/k_B\} p(\Gamma_A, t_A | \Gamma_B, t_B).$$
(6)

This is the detailed fluctuation theorem.

For many purposes it is more convenient to put the initial time before the final time. This can be achieved by exploiting the time reversibility of the microscopic system. When applying the time reversal operator the arrow of time and all microscopic rates, such as momenta, change sign. The operation is an operation working on the microscopic space. On the mesoscopic space it is only well defined when the microscopic subspace corresponding to a state Γ is transformed into a subspace that corresponds to a single mesoscopic state. We will assume that the mesoscopic states are chosen such that this is the case. In this paper the time-reversed state corresponding to Γ will be written as Γ^*. Since the microscopic time-reversal operation leaves the volume of phase space invariant, the one-on-one relation also implies [similarly to Eq. (3)] that

$$\exp[S(\Gamma)/k_B] = \exp[S(\Gamma^*)/k_B] \det(\partial \Gamma^*/\partial \Gamma).$$
(7)

For the conditional transformation of conditional probabilities microreversibility implies that

$$p(\Gamma_A, t_A | \Gamma_B, t_B) = p(\Gamma_A^*, t_A^* | \Gamma_B^*, t_B^*) \det(\partial \Gamma_A^*/\partial \Gamma_A).$$
(8)

Using Eq. (6) we conclude that

$$p(\Gamma_B, t_B | \Gamma_A, t_A) = \exp\{[S(\Gamma_B) - S(\Gamma_A^*)]/k_B\} p(\Gamma_A^*, t_A^* | \Gamma_B^*, t_B^*).$$
(9)

If a system is not externally driven, then the conditional probabilities will be time-translation invariant. In this case the equality $\Delta t^* = -\Delta t$ suffices to specify the effect of time reversal on the time. If the system is externally driven at the time of time reversal, also all external driving forces have to reverse direction. At the time of reversal we have $t_r = t_r^*$.

The rationale for reversing external fluxes is that, although the derivation is made for a closed system, this closed system can encompass quite a lot. The distinction between open and closed depends on where one places the boundary of the system. Since we are not restricted to being close to equilibrium, there is no fundamental problem with incorporating a large part of the external world into the system and modeling it in a simple way. For example, one or more subsystems can behave as heat baths that are well characterized by one variable only (namely, its temperature). It is even possible that a subsystem is an experimental setup. When the time-reversal operation is applied to this subsystem the, so-called, external driving forces are reversed.

III. COARSE-GRAINING

In Eqs. (6) and (9) the states arising in the conditional part and those arising in the probability-density part of the conditional probabilities densities have a subtle different meaning. In the conditional part, state Γ indicates a microstate sampled uniformly from the ensemble corresponding to mesostate Γ. Let γ be the microscopic state; then, it is distributed according to

$$p_{\text{uniform}}(\gamma | \tilde{\Gamma}) = \exp[- S(\tilde{\Gamma})/k_B] \delta(\Gamma(\gamma) - \tilde{\Gamma}).$$
(10)

Here the entropy arises as a normalization factor of the conditional probability Eq. (10):

$$\exp[S(\tilde{\Gamma})/k_B] = \int d\gamma \, \delta(\Gamma(\gamma) - \tilde{\Gamma}).$$
(11)

This is the formal definition of the entropy as a measure for the volume of the microscopic phase space of a mesostate $\tilde{\Gamma}$ as used in Eq. (2).

The state that enters in the probability-density part of the conditional probability distributions—e.g., Γ_B in Eq. (4)—corresponds to a microstate that is originally distributed according to Eq. (10). Subsequently, this ensemble is transformed. The distribution of the microstates at the new time does not obey Eq. (10). The state Γ_B denotes the projection of the transformed microstate onto the mesoscopic phase space [by means of $\Gamma(\gamma)$]. It is not implied that the microstates are uniformly distributed.

However, we assume that states are chosen such that, after a short time γ, most of the microscopic phase space corresponding to the state Γ has been sampled while the mesoscopic state itself has evolved only little. Therefore we can consider an averaged microstate distribution

$$p_{\text{av}}(\gamma, t) = \frac{1}{\tau \, \delta V(\gamma)} \int_t^{t+\tau} dt' \int_{\delta V(\gamma)} d\gamma' p(\gamma', t').$$
(12)

Here $\delta V(\gamma)$ is a small volume in microscopic phase space around γ and τ is a small time. During a time τ the trajectory in microscopic phase space has crossed the volume $\delta V(\gamma)$ many times. For a large range of values τ and $\delta V(\gamma)$ the value of $p_{\text{av}}(\gamma, t)$ reaches a well-defined plateau value. In this plateau region the characteristic time scale for the evolution of mesoscopic state Γ is much larger than τ, but τ is much larger than the typical equilibration time. The characteristic distances in microscopic phase space corresponding to significant changes in Γ are larger than the width of the region $\delta V(\gamma)$. If one considers smaller volumes $\delta V(\gamma)$, one needs to consider longer times τ to obtain good enough statistics. When certain ergodic properties are obeyed the plateau value of $p_{\text{av}}(\gamma, t)$ is uniformly within the subspace corresponding to the mesoscopic state Γ:

$$p_{\text{av}}(\gamma, t) = \int d\tilde{\gamma} \, p_{\text{uniform}}(\gamma | \Gamma(\tilde{\gamma})) p(\tilde{\gamma}, t)$$

$$= \int d\Gamma \, p_{\text{uniform}}(\gamma | \Gamma) p(\Gamma, t)$$

$$= \exp[- S(\Gamma(\gamma))/k_B] p(\Gamma(\gamma), t),$$
(13)

with the conditional probability given by Eq. (10).

By means of the averaging procedure, Eq. (13), the fluctuation relations, Eqs. (6) and (9), remain unchanged. The only difference is that after the averaging, both the initial mesostate Γ_A and the final state Γ_B imply a uniform distributions of the microscopic space corresponding to these mesostates. After averaging the microscopic distribution is smeared out. Therefore the dynamics of the averaged prob-

ability distribution, Eq. (13), does not conserve phase space. However, the consequence of the conservation of phase space at the underlying level still is taken into account by means of the detailed fluctuation relation. This is the most important conclusion of this work.

The microscopic dynamics leaves the Gibbs entropy defined as

$$S_G(t) = -k_B \int d\boldsymbol{\gamma} \, p(\boldsymbol{\gamma},t) \ln p(\boldsymbol{\gamma},t) \qquad (14)$$

unchanged. This is a consequence of Liouville's theorem. When maximizing the Gibbs entropy with the constraint

$$p(\tilde{\boldsymbol{\Gamma}},t) = \int d\boldsymbol{\gamma} \, p(\boldsymbol{\gamma},t) \delta(\boldsymbol{\Gamma}(\boldsymbol{\gamma}) - \tilde{\boldsymbol{\Gamma}}), \qquad (15)$$

one will find that $p(\boldsymbol{\gamma},t) := p_{av}(\boldsymbol{\gamma},t)$ given by Eq. (13). The entropy for this locally equilibrated distribution is

$$S_{G,\text{course}}(t) = \int d\boldsymbol{\Gamma}[p(\boldsymbol{\Gamma},t)S(\boldsymbol{\Gamma}) - k_B p(\boldsymbol{\Gamma},t) \ln p(\boldsymbol{\Gamma},t)], \qquad (16)$$

where $S(\boldsymbol{\Gamma})$ is given by Eq. (11).

The picture that arises for the dynamics of mesoscopic models is as follows. Starting from a mesoscopic ensemble described by $p(\boldsymbol{\Gamma})$ and a uniform distribution within the mesostate given by Eq. (10), the microscopic evolution leaves the Gibbs entropy unchanged. Many microstates in the deformed mesostates are visited in a time τ. Because of the incompressibility of phase space, this means that the a fine structure is created. This structure is not smeared out in a smooth way but becomes finer and finer when time proceeds. The scales on which the structure changes are too small to be resolved by the resolution of the measurements of an experimenter or the level of description of a model builder. They do not measure $p(\boldsymbol{\gamma},t)$, but $p_{av}(\boldsymbol{\gamma},t)$ given by Eq. (12). Because they cannot follow the evolution in time down to the smallest scales, they continuously lose information about their system. If they were able to see the details, then the (Gibbs) entropy change during a time interval would be zero according to Eq. (14).

If the mesoscopic states are chosen well, $p_{av}(\boldsymbol{\gamma},t)$ is well defined for a wide plateau of time and spatial resolutions [characterized by τ and δV used in Eq. (12)]. This causes the mesoscopic entropy $S(\boldsymbol{\Gamma})$ to be an objective quantity and not merely be a mathematical construct. It is possible that different mechanisms with well-separated time scales are important to describe the dynamics of a system. If one wants to describe the system on a certain scale, one can then introduce an appropriate level of coarse-graining. The entropy one needs to use at this level can be obtained by maximizing the Gibbs entropy, Eq. (16), valid for any of the finer scales, while keeping the large scale variables fixed.

The detailed fluctuation relation as it is derived in this paper holds for both the microscopic reversible dynamics as well as the coarse-grained dynamics. It does therefore not predict the increase in Gibbs entropy [as expressed by Eq. (14)]. Situations where the mesoscopic Gibbs entropy de-

creases can be easily imagined. A straightforward way of theoretically creating these situations is by means of the microscopic time reversal of everyday events (playing movies backward). Also for these situations the detailed fluctuation relation is obeyed. The reason that some theories can only predict an increase in Gibbs entropy is ultimately a coarse level of description consistent with the detailed fluctuation theorem.

IV. CONSISTENCY OF STOCHASTIC DIFFERENTIAL EQUATIONS

The fluctuation theorem as derived in this paper gives restrictions on mesoscopic models. In this respect it is similar to the GENERIC formalism [12,13]. This formalism imposes constraints on the form of the stochastic differential equations that can be used to model mesoscopic systems. In this section I will derive the constraints that can be derived from the fluctuation theorem, Eq. (9).

The aim is to model mesoscopic dynamics in a thermodynamic consistent way by means of a stochastic differential equation

$$d\boldsymbol{X} = \boldsymbol{A}(\boldsymbol{X})dt + \det\left(\frac{\partial \boldsymbol{X}}{\partial \boldsymbol{X}_{\text{ref}}}\right)\frac{\partial}{\partial \boldsymbol{X}} \cdot \left[\boldsymbol{D}(\boldsymbol{X})\det\left(\frac{\partial \boldsymbol{X}_{\text{ref}}}{\partial \boldsymbol{X}}\right)\right]dt$$
$$+ \sqrt{2\boldsymbol{D}(\boldsymbol{X})} \cdot d\boldsymbol{W}. \qquad (17)$$

Here $\boldsymbol{W}$ is the Wiener process. Increments of the Wiener process are normally distributed stochastic variables. The mean of an increment is zero. Increments at nonoverlapping time intervals are statistically independent. The variance of an increment equals the width of the time interval. This can be summarized by

$$\langle \Delta W_i \rangle = 0, \quad \langle \Delta W_i \Delta W_j \rangle = \delta_{ij} \Delta t_i, \qquad (18)$$

where the lower index is a time stamp and the increments are assumed to be taken over time intervals $[t_i, t_{i+1}][\Delta W_i = W(t_{i+1}) - W(t_i)]$. Equation (17) should be interpreted in the Ito form (see [14,15]). This means that Eq. (17) is the limit of a finite-difference scheme where the integrands are evaluated at the beginning of the time interval. Since a Wiener increment is an (isotropic) Gaussian variable, also $\boldsymbol{B} \cdot \Delta \boldsymbol{W}$ is a Gaussian variable which is fully specified by its variance $\boldsymbol{B}^T \cdot \boldsymbol{B} \Delta t$. This means that there is some redundancy in $\boldsymbol{B}$. In Eq. (17) we therefore introduce the diffusion tensor $\boldsymbol{D} = \frac{1}{2}\boldsymbol{B}^T \cdot \boldsymbol{B}$, which is by construction positive symmetric.

The Wiener increments model forces that fluctuate on small time scales. The coarse-graining step is to model these fluctuating as white noise and so ignoring all small time scale correlations that exist.

The form of Eq. (17) might seem somewhat extravagant, because by redefining $\boldsymbol{A}(\boldsymbol{X})$ one can write down a much more compact and simple equation. However, only in the form written in Eq. (17) are the quantities $\boldsymbol{A}$ and $\boldsymbol{D}$ a true vector and tensor in the sense that they transform in the usual way upon coordinate transformation. This is important since in a coarse-grained description it is often not clear what the "canonical" coordinates are. It is therefore good to use quantities that have a meaning irrespective of the chosen coordi-

nate system. In Appendix A it is demonstrated that when A and D are required to be tensors the form as given by Eq. (17) naturally appears.

To derive the constraints that the fluctuation theorem imposes on the coefficients of the stochastic differential equation we look at the limit $t_B \to t_A = t$ of Eq. (9). Note that we should use this form and not Eq. (6) because stochastic differential equations are only well defined for integration forward in time. In this limit the first-order term of the expansion of the conditional probability in the small variable $t_{i+1} - t_i$ inserted in Eq. (9) gives

$$\frac{\partial}{\partial t_{i+1}} p(X_{i+1}, t_{i+1} | X_i, t_i)_{t_{i+1}=t_i} = -\exp\{[S(X_{i+1}) - S(X_i^*)]/k_B\}$$

$$\times \frac{\partial}{\partial t_{i+1}^*} p(X_i^*, t_i^* | X_{i+1}^*, t_{i+1}^*)_{t_{i+1}=t_i}.$$

$$(19)$$

Here the minus sign arises because $\Delta t = \Delta t^*$.

The conditional probability distribution of the stochastic variable X obeys the Fokker-Planck equation (see Appendix A)

$$\frac{\partial}{\partial t_{i+1}} p(X_{i+1}, t_{i+1} | X_i, t_i)$$

$$= -\frac{\partial}{\partial X_{i+1}} \cdot \left\{ A(X_{i+1}, t) p(X_{i+1}, t_{i+1} | X_i, t_i) - D(X_{i+1}) p(X_{i+1}, t_{i+1} | X_i, t_i) \cdot \frac{\partial}{\partial X_{i+1}} \left[\ln p(X_{i+1}, t_{i+1} | X_i, t_i) + \ln \det\left(\frac{\partial X_{i+1}}{\partial X_{\text{ref}}}\right) \right] \right\} = 0.$$

$$(20)$$

Because the coefficients are time-translation invariant, also the conditional probabilities are time-translation invariant:

$$\frac{\partial}{\partial t_{i+1}^*} p(X_i^*, t_i^* | X_{i+1}^*, t_{i+1}^*) = -\frac{\partial}{\partial t_i^*} p(X_i^*, t_i^* | X_{i+1}^*, t_{i+1}^*). \quad (21)$$

This last time derivative obeys Eq. (20) with all appearances of X_{i+1} replaced by X_i^* and X_i replaced by X_{i+1}^* (but X_{ref} remains unchanged). The equalities for the two time derivatives can be inserted into Eq. (19). This will give constraints on the allowed forms for the tensors A and D. The determination of these constraints is complicated by the fact that the conditional probability for $t_{i+1} = t_i$ is a δ function. A straightforward procedure to determine the constraints is to multiply both sides of the equality, Eq. (19), by $g(X_i)\exp[S(X_i^*)/k_B]$ and integrate over the variable X_i^*. By means of integration by parts and the use of Eq. (7) we find that

$$\frac{\partial}{\partial X} \cdot \left\{ g(X)\exp[S(X)/k_B] \left[A(X) - D(X) \cdot \frac{\partial}{\partial X} \ln \det\left(\frac{\partial X}{\partial X_{\text{ref}}}\right) \right] \right.$$

$$\left. - D(X) \cdot \frac{\partial}{\partial X}\left(g(X)\exp[S(X)/k_B]\right) \right\}$$

$$= -\exp[S(X)/k_B] \left\{ \frac{\partial g(X)}{\partial X} \cdot \left[A(X^*) \right. \right.$$

$$\left. - D(X^*) \cdot \frac{\partial}{\partial X^*} \ln \det\left(\frac{\partial X^*}{\partial X_{\text{ref}}}\right) \right]$$

$$\left. + \frac{\partial}{\partial X^*}\left(D(X^*) \cdot \frac{\partial g(X)}{\partial X^*}\right) \right\}.$$

$$(22)$$

The requirement that this equality hold for any function $g(X)$ results in the relations given below. First, one can decompose $A(X)$ as

$$A(X) = A^{\text{rev}}(X) + D(X) \cdot \frac{\partial}{\partial X}\left\{ \frac{1}{k_B} S(X) + \ln \det\left(\frac{\partial X}{\partial X_{\text{ref}}}\right) \right\}.$$

$$(23)$$

The $A^{\text{rev}}(X)$ term models the reversible part of the motion. This reversibility is apparent from the behavior when applying the time-reversal operator. The requirements that follow from Eq. (22) for this component of the motion are

$$A^{\text{rev}}(X^*) = -\frac{\partial X^*}{\partial X} \cdot A^{\text{rev}}(X),$$

$$\frac{\partial}{\partial X} \cdot \{\exp[S(X)/k_B] A^{\text{rev}}(X)\} = 0. \quad (24)$$

According to the second equality the Gibbs entropy is not increased due to the reversible term. If one follows the time evolution of an ensemble of nearby mesoscopic states, an increase in entropy $S(X)$ due to this term is compensated for by a decrease in volume of the ensemble (and thus an increase in the probability density). The net result is that the total microscopic phase-space volume corresponding to the mesoscopic space is unchanged.

The second part in Eq. (23) is the irreversible part. This part scales as

$$A^{\text{irr}}(X^*) = \frac{\partial X^*}{\partial X} \cdot A^{\text{irr}}(X), \quad (25)$$

which is related to the relation

$$D(X^*) = \frac{\partial X^*}{\partial X} \cdot D(X) \cdot \frac{\partial X^{*T}}{\partial X}. \quad (26)$$

If one inserts Eq. (23) into Eq. (17), one will find that the part that refers to X_{ref} cancels. This gives the most general expression for a "thermodynamically consistent" stochastic differential equation

$$dX = A^{\text{rev}}(X)dt + \exp[-S(X)/k_B]$$

$$\times \frac{\partial}{\partial X} \cdot \{D(X)\exp[S(X)/k_B]\}dt + \sqrt{2D(X)} \cdot dW$$

$$= A^{\text{rev}}(X)dt + \frac{1}{k_B}D(X) \cdot \frac{\partial}{\partial X}S(X)dt + \frac{\partial}{\partial X} \cdot D(X)dt$$

$$+ \sqrt{2D(X)} \cdot dW. \qquad (27)$$

Any choice for X_{ref} gives the same final result. In view of expression (27) a particular convenient choice of reference is the one that obeys

$$\det\left(\frac{\partial X_{\text{ref}}}{\partial X}\right) = C \exp[S(X)/k_B]. \qquad (28)$$

This means that any unit volume in X_{ref} corresponds to a fixed volume in microscopic phase space. For this choice the irreversible term in Eq. (23) is zero. This shows that the irreversible part can be viewed upon as an apparent contribution rather than a physical driving force. It expresses the fact that an "unnatural" reference state is used. (This is a little bit similar as trying to describe the dynamics of a system using rotating coordinate systems.) For the choice made in Eq. (28) the contribution disappears and the notion of volume is directly inherited from the underlying microscopic space. Using the notation developed in Appendix A, Eq. (29) can be compactly rewritten as

$$dX = A^{\text{rev}}(X)dt + \{\exp[S(X)/k_B]2D(X)\}$$

$$\circ \{\exp[S(X)/k_B]\sqrt{2D(X)}\}^{-1} \cdot dW. \qquad (29)$$

This result can be obtained by inserting Eq. (28) into Eq. (A8).

The change of the mesoscopic level Gibbs entropy is given by

$$\frac{d}{dt}S_{G,\text{coarse}} = \int dX\, p(X,t)D:$$

$$\times\left(\frac{\partial}{\partial X}[S - k_B\ln p(X,t)]\frac{\partial}{\partial X}[S - k_B\ln p(X,t)]\right). \qquad (30)$$

Here $p(X,t)$ is the probability density to be in state $p(X,t)$ at time t. As discussed above the increase of entropy is a consequence of coarse-graining. In the case of the stochastic differential equation processes with very small characteristic time scales are not resolved. They are modeled as processes with zero correlation time by means of white noise. The unresolved correlations integrated over a larger time τ give rise to the diffusion tensor. This diffusion tensor plays a prominent role in the expression for the entropy production, Eq. (30).

V. MESOSCOPIC SIMULATIONS

As shown in the previous section stochastic differential equations that describe nonequilibrium processes can always be split into a reversible and an irreversible part. A numerical solution for a single time step of the full problem can be formed by first considering the purely reversible and the purely irreversible problems individually (by putting the other term to zero). An approximation to the full problem can then be found by combining the partial solutions using a Trotter expansion (or a higher-order expansion). Usually the construction of numerical solutions to the subproblems concerns further splitting and Trotter expansion.

The fluctuation theorem such as expressed in Eq. (9) can be very helpful in creating good numerical approximations to the subproblems. The reason is that it is expressed for finite time differences. If one makes sure that the subproblems obey the fluctuation theorem, then by construction the full numerical solution also obeys the theorem.

For a purely reversible process one finds that

$$p^{\text{rev}}(\Gamma_{i+1},t_{i+1}|\Gamma_i,t_i)\delta\Gamma_{i+1} = p^{\text{rev}}(\Gamma_i^*,t_i^*|\Gamma_{i+1}^*,t_{i+1}^*)\delta\Gamma_i^*. \qquad (31)$$

There is a one-on-one relation between the initial and final states. This means that the conditional probabilities are always δ-function like. The volume of the mesoscopic phase space need not be conserved. Combining Eq. (31) with the fluctuation theorem, Eq. (9), one finds that

$$\delta\Gamma_{i+1}\exp[S(\Gamma_{i+1})] = \delta\Gamma_i\exp[S(\Gamma_i)]; \qquad (32)$$

i.e., the underlying microscopic phase-space volume should not change. This is the discrete equivalent of the second relation in Eq. (24).

For a purely irreversible process,

$$p^{\text{irr}}(\Gamma_{i+1},t_{i+1}|\Gamma_i,t_i) = p^{\text{irr}}(\Gamma_{i+1}^*,t_{i+1}|\Gamma_i^*,t_i)\det(\partial\Gamma_{i+1}^*/\partial\Gamma_{i+1}). \qquad (33)$$

Combining Eq. (33) with Eq. (9) gives

$$\frac{p^{\text{irr}}(\Gamma_{i+1},t_{i+1}|\Gamma_i,t_i)}{p^{\text{irr}}(\Gamma_i,t_i^*|\Gamma_{i+1},t_{i+1}^*)} = \exp\{[S(\Gamma_{i+1}) - S(\Gamma_i)]/k_B\}. \qquad (34)$$

For the derivation of Eqs. (32) and (34) we made use of Eq. (8).

In the case of the molecular dynamics simulation (of the N-V-E ensemble) there is no entropy. The dynamics obeys Eqs. (31) and (32) for this special case. A widely used integrator is the Verlet algorithm. Its success is usually explained by the fact that it is both time reversible and it conserves phase-space volume [16]. These are exactly the properties that have to be obeyed according to the fluctuation theorem for purely reversible motion.

In the case of a microscopic system in contact with a heat bath the total entropy equals the entropy of the bath:

$$S(\Gamma) = S_0 - E(\Gamma)/T, \qquad (35)$$

where E is the energy of the system and T the temperature of the heat bath. Coarse-grained techniques such as dissipative particle dynamics (DPD) also consider a system in contact

with a heat bath (but without further internal entropy). The most distinct feature of DPD is that it conserves momentum; i.e., it is Galilean invariant. The equation of change for the momentum of particle i is

$$d\boldsymbol{p}_i = \left[\boldsymbol{f}_i^C - \sum_j \gamma\boldsymbol{\omega}(\boldsymbol{r}_{ij})\cdot\boldsymbol{v}_{ij}\right]dt + \sum_j \sqrt{2k_BT\ \gamma\boldsymbol{\omega}(\boldsymbol{r}_{ij})}\cdot d\boldsymbol{W}_{ij},$$

(36)

where the indices ij indicate a pair of particles. The first term on the right-hand side is the conservative force which constitutes the reversible part of the motion. The last two terms are the dissipative and fluctuating terms that constitute the irreversible part. The full equation is in accordance with Eq. (27). The irreversible part can be split down to individual pair interactions. The numerical approximation for this part only considers the change in momentum of the two particles in a pair i and j. All the momenta of the other particles remain unchanged, and the positions of all particles including those of i and j do not change. Therefore only the kinetic contribution to the energy in Eq. (35) changes. The requirement (34) now reduces to

$$\frac{p(\tilde{\boldsymbol{p}}_i,\tilde{\boldsymbol{p}}_j|\boldsymbol{p}_i,\boldsymbol{p}_j)}{p(\boldsymbol{p}_i,\boldsymbol{p}_j|\tilde{\boldsymbol{p}}_i,\tilde{\boldsymbol{p}}_j)} = \exp\{-[(\tilde{\boldsymbol{p}}_i^2-\boldsymbol{p}_i^2)/2m_i+(\tilde{\boldsymbol{p}}_j^2-\boldsymbol{p}_j^2)/2m_j]/k_BT\},$$

(37)

where the tildes indicate the new values. If one requires that the individual interactions conserve momentum—i.e., $\tilde{\boldsymbol{p}}_i+\tilde{\boldsymbol{p}}_j=\boldsymbol{p}_i+\boldsymbol{p}_j$—one finds that Eq. (37) is obeyed for

$$p(\tilde{\boldsymbol{p}}_i,\tilde{\boldsymbol{p}}_j|\boldsymbol{p}_i,\boldsymbol{p}_j) \propto \exp\{-[\sqrt{1+\alpha}(\tilde{\boldsymbol{p}}_j-\tilde{\boldsymbol{p}}_i)$$
$$-\sqrt{\alpha}(\boldsymbol{p}_j-\boldsymbol{p}_i)]^2/4\mu k_BT\},$$

(38)

with μ the reduced mass of the two-particle system. In [17], I developed a DPD discretization that is consistent with Eq. (38) using a different reasoning. This scheme was shown to be superior to other discretization schemes, especially for determining equilibrium properties.

The reversible part of Eq. (36) can be derived from a Hamiltonian. This means that it does preserve the volume of phase space and energy. Since the entropy of the heat bath is a function of energy via Eq. (35), this is consistent with Eq. (24). The use of the Verlet algorithm seems to be a good choice to solve the reversible part. However, in the discretized case round-off errors are present. This means that energy is not exactly conserved. Therefore the entropy does change. Since the Verlet algorithm does rigorously obey conservation of phase-space volume, Eq. (32) is not rigorously obeyed for the numerical solution. It turns out that this deviation is the dominant discretization error at finite time step; see [17].

In the case of Monte Carlo simulations (of the N-V-T ensemble) one is not interested in the dynamics but only in the equilibrium statistics. In this case one has a lot of freedom to construct conditional probabilities. The only condition that has to be obeyed is Eq. (34) with Eq. (35) inserted. This gives the ordinary detailed balance condition. A straightforward way to achieve this is by means of the Metropolis procedure.

For models that are not only in contact with a heat bath but also have internal entropy the detailed balance condition can be easily generalized by using Eq. (34). The first step is the generation of a trial move according to some probability density. The next step is the acceptation step where a move is accepted or rejected with a certain probability. The total conditional probability is

$$p(\boldsymbol{\Gamma}_{i+1}|\boldsymbol{\Gamma}_i) = P^{\text{acc}}(\boldsymbol{\Gamma}_{i+1}|\boldsymbol{\Gamma}_i)p^{\text{trial}}(\boldsymbol{\Gamma}_{i+1}|\boldsymbol{\Gamma}_i).$$

(39)

When one chooses

$$P^{\text{acc}}(\boldsymbol{\Gamma}_{i+1}|\boldsymbol{\Gamma}_i) = \min\left(1, \frac{p^{\text{trial}}(\boldsymbol{\Gamma}_i|\boldsymbol{\Gamma}_{i+1})e^{-S(\boldsymbol{\Gamma}_i)/k_B}}{p^{\text{trial}}(\boldsymbol{\Gamma}_{i+1}|\boldsymbol{\Gamma}_i)e^{-S(\boldsymbol{\Gamma}_{i+1})/k_B}}\right),$$

(40)

Eq. (34) is obeyed.

This scheme can also be used to generate thermodynamically consistent discretizations for the irreversible part of the dynamical equation (27). To generate such a scheme one has to ensure that one generates a stochastic trial step with the correct variance (for the limit $\Delta t\to 0$). For small enough time steps the stochastic term in the differential equation is always dominant. Also for small enough time steps P^{acc} will be very close to one. This means that Eq. (40) gives a perturbation on the stochastic step. As we have shown in Sec. IV the fluctuation theorem applied to the irreversible part results in the (irreversible) deterministic term in Eq. (27). Therefore application of Eq. (40), which obeys the fluctuation theorem applied to purely irreversible motion, will result in a perturbation on the stochastic part that in the limit $\Delta t\to 0$ is equal to the deterministic term in Eq. (27).

Therefore, generating a trial move with the correct variance and applying the generalized Metropolis scheme gives a valid discretization of the irreversible part of a thermodynamically consistent stochastic differential equation. Moreover, the discretization itself is also thermodynamically consistent.

One can think of a wide variety of ways of generating trial moves. Acceptance rates can be increased by introducing biasing. Everything is allowed as long as the variance of stochastic part is correct up to $\mathcal{O}(\Delta t)$. If one uses more advanced discretization (e.g., with a predictor step or partly implicit), then also the computation of the trial probability densities will become more computationally expensive.

A. Time discretization

The matter of time step dependence raises some important issues that have to be considered. When investigating a simple Euler forward discretization of the irreversible parts of Eq. (27) one will find that for small time steps the stochastic term is much larger than the deterministic term. Since ΔW is proportional to $\sqrt{\Delta t}$, the entropy gradient terms and the stochastic term become comparable for

$$\Delta t \approx k_B^2(|\boldsymbol{D}||\partial S/\partial \boldsymbol{X}|^2)^{-1}.$$

(41)

This corresponds to a displacement

$$\Delta X \approx k_B\sqrt{|\boldsymbol{D}|\Delta t} \approx |\partial S/\partial \boldsymbol{X}|^{-1}.$$

(42)

For such a displacement the change in entropy during one time step is of the order of k_B.

When increasing the time step still further the stochastic term becomes small compared to the deterministic term. If on the corresponding length scales the variation in the entropy gradient (and in the diffusion) is small, a Euler discretization of Eq. (27) using this large time step is still an accurate approximation. This is the case when

$$\left| \frac{\partial S}{\partial \boldsymbol{X}} \right|^2 \gg k_B \left| \frac{\partial^2 S}{\partial \boldsymbol{X} \, \partial \boldsymbol{X}} \right|. \tag{43}$$

This can be considered a far-from-equilibrium condition. If it is obeyed, fluctuations are negligible compared to the irreversible deterministic terms for large enough time steps. If it is not obeyed, the system is in the mesoscopic regime. Here significant changes occur on time scales smaller than that given by Eq. (41) or, equivalently, for entropy changes of the order of k_B. Fluctuations always play a dominant role and cannot be neglected even if one considers longer time scales.

Equation (29) is an alternative formulation of Eq. (27). A straightforward discretization (of the irreversible part) is

$$\Delta \boldsymbol{X}_{\text{pred}} = \sqrt{2\boldsymbol{D}(\boldsymbol{X})} \cdot \Delta \boldsymbol{W}$$

$$\Delta \boldsymbol{X} = \exp\{S([\boldsymbol{X} + \boldsymbol{X}_{\text{pred}}]/2)/k_B\} 2\boldsymbol{D}([\boldsymbol{X} + \boldsymbol{X}_{\text{pred}}]/2)$$

$$\times \{\exp[S(\boldsymbol{X})/k_B] \sqrt{2\boldsymbol{D}(\boldsymbol{X})}\}^{-1} \cdot \Delta \boldsymbol{W}. \tag{44}$$

(The discretized equation does not obey the detailed fluctuation theorem.) An advantage of this discretization is that no derivatives need to be determined. It is, however, only a good approximation for time steps smaller than that given by Eq. (41). The deterministic term arises as a perturbation of the stochastic term. If, however, the time step is too large, the deterministic term can no longer be treated as a perturbation.

Similar conclusions can be drawn for the thermodynamically consistent schemes given above. The correct equilibrium statistics will be sampled for any chosen time step. This does not mean that the schemes yield a satisfactory discretization for any finite Δt. The reason is that dynamical properties will be time step dependent.

For example, in the DPD scheme the typical velocity relaxation time (for a dense enough fluid) is m/γ. Since the deviation of the particle velocity from the center-of-mass velocity of a blob is of the order of the thermal velocity—i.e., $v_{\text{th}} = \sqrt{k_B T/m}$—the self-diffusion coefficient in space is proportional to $v_{\text{th}}^2 m/\gamma = kT/\gamma$. When time steps are chosen larger than the characteristic time, the particle velocity relaxes almost completely within one time step Δt. Therefore the self-diffusion found in the simulation will be proportional to $v_{\text{th}}^2 \Delta t$. This is larger than the true self-diffusion.

In the case of the generalized Metropolis scheme a rejection means that a dynamical variable is not updated during the current time step. Many rejections cause the dynamics in the simulated system to slow down. This happens when entropy changes during a time step are comparable or larger than k_B—i.e., for time steps larger than the one in Eq. (41).

VI. DISCUSSION

In this paper I derived a detailed fluctuation theorem. The main ingredients for the derivation are microscopic time re-

versibility and Liouville's theorem. If one assumes local equilibrium—i.e., fast degrees of freedom are in equilibrium and slow variables can be far from equilibrium—the detailed fluctuation theorem remains valid. This means that we have a relation that is valid far a large class of far-from-equilibrium models. This is especially valuable for the application to mesoscopic situations.

Most of the literature on the fluctuation theorem has a dynamical-systems point of perspective. The program seems to be to characterize nonequilibrium stationary states—e.g., by means of the Sinai-Rowen-Bowen (SRB) measure—and use its properties to derive relations such as the stationary-state fluctuation theorem [5]. Systems that are considered are typically deterministic dissipative systems.

The entropy production is defined as minus the phase-space contraction rate. The change of phase-space volume is inversely proportional to the probability density of a state tracked in time. When introducing the entropy change as an integral over time of the entropy production rate, the detailed fluctuation theorem, Eq. (6), is obeyed.

In Appendix B, I show how to derive the steady-state fluctuation theorem from the detailed fluctuation theorem by making some ergodic assumptions. A feature of this derivation is that detailed knowledge of the stationary state (such as an assumed SRB measure) is not needed. On the other hand, if a stationary state can be shown to have a SRB measure, this should be consistent with the stationary-state fluctuation theorem. The existence of a SRB measure is, however, not a prerequisite for the theorem to be valid. In the case of, for example, a system described by a stochastic differential equation the stationary distribution is not expected to be a SRB measure but the fluctuation theorem is valid.

The derivation given in this paper is essentially the same as the one given by Maes and Netočný [8], but less formal. The same expression was also derived before by Jarzynski [9] although he considered a special case—namely, a microscopic system in contact with one or more heat baths. The main contribution of the present paper is that it demonstrates the link with mesoscopic theories such as the GENERIC formalism [12,13].

In Sec. IV, I showed the constraints the detailed fluctuation theorem imposes on the form of stochastic differential equations. As a starting point I used the general stochastic differential equation written in the generalized kinetic form developed in Appendix A. It is interesting to note that this form is dictated by the fact that the primary variables in the stochastic differential equation are taken to be tensorial (i.e., they behave well upon coordinate transformation). The fact that this form is found to be most suitable for physical theories is a consequence of this observation.

The irreversible part of the motion of Eq. (27) is exactly of the form given by the GENERIC formalism. Up to now the form was motivated by reference to the fluctuation-dissipation theorem (of the second kind) [18]. In the present paper it was proved that, as long as the local equilibrium assumption holds, the irreversible part is of the form described by Eq. (27).

Öttinger and Grmela assume that the reversible phase-space velocity $\boldsymbol{A}^{\text{rev}}$ is perpendicular to the entropy gradient. This is consistent with Eq. (24) for the case that $\partial/\partial\boldsymbol{X} \cdot \boldsymbol{A}^{\text{rev}}$

=0. However, due to the transformation behavior of the entropy, Eq. (3), Eq. (24) is valid in any coordinate frame, but the orthogonality of A^{rev} and the entropy flux is not. In [18] de Pablo and Öttinger realize this and redefine the entropy as

$$S^{\mathrm{dPO}}(X) = S(X) - k_B \ln \det\left(\frac{\partial X_{\mathrm{ref}}}{\partial X}\right). \qquad (45)$$

This redefined entropy transforms as a scalar. Therefore the orthogonality condition can be maintained when changing coordinates. To be consistent with Eq. (24), A^{rev} should be incompressible for X_{ref}. The redefinition, Eq. (45), however, has a few consequences that are not taken into account consistently in [18]. In expressions such as Eq. (10), $S(X)$ is the proper entropy to be used in the normalization factor for the microcanonical ensemble, not $S^{\mathrm{dPO}}(X)$. Next, in Eq. (23), $S^{\mathrm{dPO}}(X)$ can be directly substituted. This has as a consequence that in the final expression of the stochastic differential equation a term referring to the reference state, as in Eq. (A6), appears. This means that, when using $S^{\mathrm{dPO}}(X)$, the general form of the stochastic differential equation has an extra term compared to the proposed GENERIC form. Last, also the expression for the Gibbs entropy, Eq. (16), acquires an extra term.

One has to conclude that in the GENERIC formalism a preferred coordinate system is used. It is implicitly assumed that there is no (mesoscopic) phase-space contraction due to reversible motion. This excludes the treatment of problems that are deterministic, reversible, and dissipative, such as the Slodd system that was so important in the development of the fluctuation theorem (see Appendix B). I therefore propose to use the entropy definition as proposed in this paper and take for granted that it does not transform as a usual scalar quantity.

The GENERIC formalism poses a more restrictive form on the reversible motion than is found in the present paper—namely, a Poisson structure. The fact that I do not find this is because I did not consider the detailed structure of the microscopic dynamics. The only part that was used was Liouville's theorem. It is in fact not proved from first principles that the reversible part has to have a Poisson structure [18]. I suspect that it cannot be proved. I think that the Poisson structure only holds for special choices of the mesoscopic variables. Once a good choice for the mesoscopic variables is made, any coordinate transformation of the mesoscopic space will leave the structure invariant. Probably this choice of variables is usually the natural choice to make. The observation that most (maybe all) known macroscopic equations are consistent with this structure [12,13] is explained by this. A fundamental proof is needed.

Since the form we propose in this paper is less restrictive than the GENERIC form, it has less predictive power. When using the GENERIC form, however, one should keep in mind that the extra structure has not been proved from first principles.

Besides the fact that the detailed fluctuation theorem imposes constraints on the form mesoscopic and macroscopic equations can have, it can also be used to create numerical approximations. In Sec. V, I gave some examples of this. The

use of the detailed fluctuation theorem ensures that the schemes are thermodynamically consistent. It was noted in Sec. V A that this is not necessarily equivalent to higher accuracy. In situations far from equilibrium, systematic irreversible driving forces due to entropy gradients are dominant compared to thermal fluctuations. In this case imposing thermodynamic consistency usually inversely affects the simulation accuracy of the dynamical behavior.

ACKNOWLEDGMENT

APPENDIX A: A GENERALIZATION OF THE KINETIC STOCHASTIC INTEGRAL

In this appendix we will investigate what the general form of a stochastic differential equation is when one requires that the primary variables in the equation be tensor quantities. Consider the simple stochastic differential equation

$$dX = A(X)dt + C(X)dt + \sqrt{2D(X)} \cdot dW \qquad (A1)$$

and interpret this equation in the Ito form [14,15]. Here the vector notation and dot product are just used as a shorthand for an index notation. Let us require that $A(X)$ and $D(X)$ are tensor quantities. This means that upon coordinate transformation $X \to X'$,

$$A'(X') = \frac{\partial X'}{\partial X} \cdot A(X),$$

$$D'(X') = \frac{\partial X'}{\partial X} \cdot D(X) \cdot \left(\frac{\partial X'}{\partial X}\right)^T. \qquad (A2)$$

Using Ito calculus [14,15] one finds that Eq. (A1) transforms to

$$dX' = \frac{\partial X'}{\partial X} \cdot [A(X) + C(X)]dt + \frac{\partial^2 X'}{\partial X \, \partial X} : D(X)dt$$

$$+ \sqrt{2\frac{\partial X'}{\partial X} \cdot D(X) \cdot \left(\frac{\partial X'}{\partial X}\right)^T} \cdot dW. \qquad (A3)$$

This gives for $C(X)$ the transformation rule

$$C'(X') = \frac{\partial X'}{\partial X} \cdot C(X) + \frac{\partial^2 X'}{\partial X \, \partial X} : D(X). \qquad (A4)$$

A choice for $C(X)$ that obeys this transformation rule is

$$C(X) = \det\left(\frac{\partial X_{\mathrm{ref}}}{\partial X}\right) \frac{\partial}{\partial X} \cdot \left[D(X)\det\left(\frac{\partial X}{\partial X_{\mathrm{ref}}}\right)\right]. \qquad (A5)$$

For any other $\widetilde{C}(X)$ that obeys Eq. (A4) the difference $\widetilde{C}(X) - C(X)$ scales as a vector and can therefore be included in $A(X)$. The full expression for Eq. (A1) becomes

$$dX = A(X)dt + \det\left(\frac{\partial X}{\partial X_{\text{ref}}}\right)\frac{\partial}{\partial X} \cdot \left[D(X)\det\left(\frac{\partial X_{\text{ref}}}{\partial X}\right)\right]dt$$
$$+ \sqrt{2D(X)} \cdot dW. \tag{A6}$$

The importance of this equation is that both quantities $A(X)$ and $D(X)$ are tensors. This is different from the ordinary Ito form, Eq. (A1), with $C(X)=0$. In physical theories one usually uses tensorial quantities, since this guarantees independence of the coordinate system. Therefore equations of the form of Eq. (A6) appear naturally in applications.

The occurrence of the reference coordinate X_{ref} might be a surprise to some. If one uses the expression, Eq. (A5), for $C(X)$ a change in reference, X_{ref} gives a contribution to $A(X)$. To understand the role of the reference it is useful to investigate the Fokker-Planck equation corresponding to Eq. (A6):

$$\frac{\partial}{\partial t}p(X,t) + \frac{\partial}{\partial X} \cdot \left\{ A(X,t)p(X,t) - D(X,t)p(X,t)\right.$$
$$\left. \times \frac{\partial}{\partial X}\left[\ln p(X,t) + \ln\det\left(\frac{\partial X}{\partial X_{\text{ref}}}\right)\right]\right\} = 0. \tag{A7}$$

From this form one can see that for $A(X)=0$ the equilibrium probability obeys $p(X)\det(\partial X/\partial X_{\text{ref}})=$const. Using the transformation rules for probability densities this corresponds to $p(X_{\text{ref}})=$const. This means that, in a nondriven system, all unit volumes defined by means of the coordinates X_{ref} are equally likely to be visited.

Equation (A6) can also be written as

$$dX = A(X)dt + \left[\det\left(\frac{\partial X_{\text{ref}}}{\partial X}\right)2D(X)\right]$$
$$\circ \left[\det\left(\frac{\partial X_{\text{ref}}}{\partial X}\right)\sqrt{2D(X)}\right]^{-1} \cdot dW. \tag{A8}$$

It is a generalization of the kinetic integral [19]. The $\circ$ symbol indicates the Stratonovich dot product. It here denotes that, in a finite-difference approximation, the part to the left of the dot should be evaluated centrally. The part to the right of the dot should be evaluated at the initial point of a time interval. Thus an equation

$$dX = f(X) \circ g(X)dW \tag{A9}$$

is the limit to $\Delta t \to 0$ of

$$X(t_{i+1}) = X(t_i) + \frac{1}{2}[f(X(t_i)) + f(X(t_{i+1}))]$$
$$\times g(X(t_i))[W(t_{i+1}) - W(t_i)]. \tag{A10}$$

Expanding this expression for $f()$ around X_i and using the fact that $\Delta W^2 \to \Delta t$ one can write down the equivalent Ito form. Doing this exercise for Eq. (A8) shows it is equivalent to Eq. (A6).

From a numerical point of view interpretation (A8) can be convenient since no evaluation of gradients is needed. Furthermore, different evaluations of the midpoint term give the same result in the $\Delta t \to 0$ limit. A particular practical evaluation is to use a predictor-corrector algorithm.

APPENDIX B: DERIVATION OF THE STATIONARY FLUCTUATION THEOREM

The stationary fluctuation theorem is a consequence of the detailed fluctuation theorem applied to stationary states of driven systems. It is valid for long times only. Some ergodicity properties have to be obeyed for it to apply.

The standard example used is a system of interacting particles driven by a shear flow. The equations of motion used are the Sllod equations (so named because of its close relationship to the Dolls tensor algorithm)

$$\dot{x}_i = p_i + \dot{\gamma} y_i e_x, \quad \dot{p}_i = F_i - \dot{\gamma} p_{y,i} e_x - \alpha p_i. \tag{B1}$$

This is Newton's equation for the particles with interparticle forces, but expressed using so-called peculiar momenta. The peculiar momentum is defined by means of the relative velocity with respect to the applied shear, with shear rate $\dot{\gamma}$. The driven system is thermostated by means of a Gaussian thermostat. This can be done by determining α (as a function of all coordinates) such that either the internal energy or the kinetic energy remains constant. In both cases the kinetic energy is defined using the peculiar momenta (i.e., locally with respect to the applied flow).

The system of equations is reversible, but phase-space volume is not conserved. Here by reversible we mean that the trajectories are traced back when both momenta and the driving force $\dot{\gamma}$ are reversed.

The phase-space contraction is

$$\sigma/k_B = -\frac{\partial}{\partial \Gamma} \cdot \dot{\Gamma} = -\sum_i \frac{\partial}{\partial x_i} \cdot \dot{x}_i + \frac{\partial}{\partial p_i} \cdot \dot{p}_i. \tag{B2}$$

The assumption that the system is an approximate description of some underlying microscopic system induces an entropy definition. Since the equations of motion are reversible, the total underlying microscopic phase-space volume should not change during the time evolution of the system. This means that the entropy should be such that $\delta V \exp(S/k_B)$ is constant in time. Here δV is the volume of a (deforming) neighborhood around a point in phase space. This implies the definition

$$\frac{dS}{dt} = \sigma. \tag{B3}$$

Since the entropy cannot be expressed in terms of the phase-space coordinates, it should be introduced as extra state variable. With Eq. (B3) to supplement the equations of motion the full equations obey the conditions for reversible motion as stated in Sec. IV.

When applying a constant shear the evolution of the system, Eq. (B2), will ultimately reach a stationary state. Note that the notion of stationarity always assumes some level of coarse-graining. In the case of microscopic incompressible motion the phase-space density keeps evolving and never approaches an attractor. Coarse-graining by means of time and spatial averaging combined with ergodic properties are necessary to define stationary (and equilibrium) states. In the case of deterministic systems like the Sllod equations the equations are already coarse grained to some degree. Since on average the phase-space contraction will be larger than

zero, the phase-space density will converge toward a strange attractor.

One can conclude that, for deterministic equations such as the Sllod equations, an entropy can always be determined by means of Eq. (B3). This entropy is such that the associated microscopic phase-space volume is a constant of the motion or, equivalently, that the generalized Gibbs entropy, Eq. (16), is conserved. An entropy that is consistent with the conservation of microscopic phase space for reversible motion causes the detailed fluctuation theorem as introduced in this paper to be valid.

The detailed fluctuation theorem leads to the stationary fluctuation theorem. When applying the detailed fluctuation theorem, Eq. (9), to the extended set of variables this gives

$$p_{\dot{\gamma}}(\Gamma_B, S_B, t_B | \Gamma_A, S_A, t_A) = \exp[(S_B - S_A^*)/k_B]$$
$$\times p_{\dot{\gamma}^*}(\Gamma_A^*, S_A^*, t_A^* | \Gamma_B^*, S_B^*, t_B^*).$$
(B4)

Note that we labeled the probability densities by means of the driving force. In the time-reversed case also the driving force is time reversed. In the case of deterministic equations such as the Sllod equations the conditional properties are all δ-function like.

We will assume that the states in Eq. (B4) are coarse grained to such an extent that stationarity is defined in a nonpathological way. Note that this is not strictly the case for deterministic systems like the Slodd equations. This will be discussed below.

In the stationary state only relative differences in time and entropy are important, so

$$p_{\dot{\gamma}}(\Gamma_B, S_B, t_B | \Gamma_A, S_A, t_A) = \frac{1}{\Delta t} p_{\dot{\gamma}}(\Gamma_B, \bar{\sigma}, \Delta t | \Gamma_A), \quad \text{(B5)}$$

where $\bar{\sigma}$ is the mean entropy production $(S_B - S_A)/\Delta t$. We can rewrite this conditional probability as

$$p_{\dot{\gamma}}(\Gamma_B, \bar{\sigma}, \Delta t | \Gamma_A) = p_{\dot{\gamma}}(\bar{\sigma}, \Delta t | \Gamma_B, \Gamma_A) p_{\dot{\gamma}}(\Gamma_B, \Delta t | \Gamma_A). \quad \text{(B6)}$$

This means that the conditional probability density to end up in state B and have a mean entropy production $\bar{\sigma}$ equals the probability density to have an entropy production $\bar{\sigma}$ for paths that are known to end up in state B times the probability density that the path indeeds ends up in state B. When substituting the relations outlined above into Eq. (B4) and taking the logarithm on both sides one finds that

$$\ln \frac{p_{\dot{\gamma}}(\bar{\sigma}, \Delta t | \Gamma_B, \Gamma_A)}{p_{\dot{\gamma}^*}(\bar{\sigma}^*, \Delta t | \Gamma_A^*, \Gamma_B^*)} + \ln \frac{p_{\dot{\gamma}}(\Gamma_B, \Delta t | \Gamma_A)}{p_{\dot{\gamma}^*}(\Gamma_A^*, \Delta t | \Gamma_B^*)}$$
$$= \bar{\sigma} \Delta t / k_B + (S_A - S_A^*)/k_B. \quad \text{(B7)}$$

The reason to make this splitting is then, when Δt is larger than a few correlation times, the conditional probabilities

will become independent of the initial and final states:

$$\lim_{\Delta t \to \infty} p_{\dot{\gamma}}(\bar{\sigma}, \Delta t | \Gamma_B, \Gamma_A) = p_{\dot{\gamma}}(\bar{\sigma}, \Delta t),$$

$$\lim_{\Delta t \to \infty} p_{\dot{\gamma}}(\Gamma_B, \Delta t | \Gamma_A) = p_{\dot{\gamma}}(\Gamma_B). \quad \text{(B8)}$$

If we choose Γ_A and Γ_B to be part if the attractor (and therefore Γ_A^* and Γ_B^* part of the attractor of the reversed process), then $p_{\dot{\gamma}}(\Gamma_B)$ and $p_{\dot{\gamma}^*}(\Gamma_A^*)$ are nonzero.

In the limit many $\Delta t \to \infty$ terms in Eq. (B7) become constant. Dividing both sides by Δt and using the limiting behavior as given by Eq. (B8) gives the stationary fluctuation theorem

$$\lim_{\Delta t \to \infty} \frac{1}{\Delta t} \ln \frac{p_{\dot{\gamma}}(\bar{\sigma}, \Delta t)}{p_{\dot{\gamma}^*}(-\bar{\sigma}, \Delta t)} = \frac{\bar{\sigma}}{k_B}. \quad \text{(B9)}$$

Here we used that $\bar{\sigma}* = -\bar{\sigma} + (S_A - S_A^*)/\Delta t$. Note that in the denominator still the reversed shear rate $\dot{\gamma}^*$ is used. However, for the stationary situation one expects that the probability to find a certain entropy production does not depend on the direction of the stationary shear rate. Therefore $\dot{\gamma}^*$ can be replaced by $\dot{\gamma}$. With this substitution Eq. (B9) becomes the famous fluctuation theorem. It states that for large times positive entropy production is exponentially more likely than negative entropy production.

For a deterministic system, such as the Slodd system, the analysis as outlined above has some weak points. All conditional probabilities are δ-function like. The attractor (if it exists) will be a fractal. Equation (B8) is not valid since there is always a one-on-one correspondence between initial and final states. This problem in the analysis can be resolved by looking at such a deterministic system on a somewhat coarser level. From this point of view a state denoted by Γ includes a small neighborhood. If one then assumes a strong dependency on initial conditions—i.e., a chaotic hypothesis [4]—points that start out near each other will end up in very different points in phase space. Therefore Eq. (B8) will be valid. In numerical experiments the coarse-graining usually occurs because $p_{\dot{\gamma}}(\bar{\sigma}, \Delta t)$ is determined using a finite resolution (by means of binning).

Recently it has been shown that the stationary fluctuation theorem is not valid for many systems [20–23]. A simple example is a Brownian particle dragged through a fluid by means of optical tweezers. As demonstrated in [23], in the case of the Brownian particle, the stationary fluctuation theorem holds for the work but not for the entropy. The point where the current derivation breaks down for this example is Eq. (B8). In the case of the Brownian particle, the work done on the system over a long time becomes statistically independent of the initial and final positions. The entropy difference, however, is the work minus the potential energy difference of the optical trap. Since this energy difference is dependent on the initial and final positions, the first limit in

Eq. (B8) is not valid. If one uses the work instead of the entropy as extra variable, besides the position of the Brownian particle, in the conditional probability, the limits of Eq. (B8) are allowed.

For a reversible system, like the Slodd system, the Gibbs entropy remains constant. In Eq. (16) the first term will increase with time. However, the probability density obtains more and more fine structure as time proceeds. Therefore the $p \ln p$ term will give a negative contribution to the Gibbs entropy. This means that the two contributions cancel out. If one looks on a coarser level, this process stops when the small scales can no longer be resolved. On this coarse-grained level $\bar{\sigma}$ indeed gives the Gibbs entropy production for long times.

Note that the stationary fluctuation theorem seems to imply the second law of thermo-dynamics for large times (but also deviations for smaller times). We earlier found that the transient fluctuation theorem itself does not imply a direction of the entropy production. The fact that the stationary state fluctuation theorem derived here predicts a entropy production that is positive on average depends on the assumptions made. The main assumption is that a stationary state exists. Such state can only be defined if one looks to a system on a somewhat coarse-grained level.

[1] D. J. Evans, E. G. D. Cohen, and G. P. Morriss, Phys. Rev. Lett. **71**, 2401 (1993).

[2] D. J. Evans and D. J. Searles, Phys. Rev. E **50**, 1645 (1994).

[3] G. Gallavotti and E. G. D. Cohen, Phys. Rev. Lett. **74**, 2694 (1995).

[4] G. Gallavotti and E. G. D. Cohen, J. Stat. Phys. **80**, 931 (1995).

[5] D. Ruelle, J. Stat. Phys. **95**, 393 (1999).

[6] D. J. Evans and D. J. Searles, Adv. Phys. **51**, 1529 (2002).

[7] C. Maes, J. Stat. Phys. **95**, 367 (1999).

[8] C. Maes and K. Netočný, J. Stat. Phys. **110**, 269 (2003).

[9] C. Jarzynski, J. Stat. Phys. **98**, 77 (2000).

[10] E. Mittag and D. J. Evans, Phys. Rev. E **67**, 026113 (2003).

[11] R. Zwanzig, Phys. Rev. **124**, 983 (1961).

[12] M. Grmela and H. C. Öttinger, Phys. Rev. E **56**, 6620 (1997).

[13] H. C. Öttinger and M. Grmela, Phys. Rev. E **56**, 6633 (1997).

[14] C. W. Gardiner, *Handbook of Stochastic Methods* (Springer, Berlin, 1997).

[15] H. Öttinger, *Stochastic Processes in Polymeric Fluids* (Springer-Verlag, Berlin, 1996).

[16] D. Frenkel and B. Smit, *Understanding Molecular Simulation (From Algorithms to Applications)* (Academic, San Diego, 2002).

[17] E. A. J. F. Peters, Europhys. Lett. **66**, 311 (2004).

[18] J. J. de Pablo and H. C. Öttinger, J. Non-Newtonian Fluid Mech. **96**, 137 (2001).

[19] M. Hütter and H. C. Öttinger, J. Chem. Soc., Faraday Trans. **94**, 1403 (1998).

[20] R. v. Zon and E. G. D. Cohen, Phys. Rev. Lett. **91**, 110601 (2003).

[21] R. v. Zon and E. G. D. Cohen, Phys. Rev. E **69**, 056121 (2004).

[22] D. J. Evans, D. J. Searles, and L. Rondoni, e-print cond-mat/0312353.

[23] D. J. Evans, e-print cond-mat/0408195.