

Path-integral analysis of fluctuation theorems for general Langevin processes

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

J. Stat. Mech. (2006) P08001

(<http://iopscience.iop.org/1742-5468/2006/08/P08001>)

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

Download details:

IP Address: 130.102.42.98

This content was downloaded on 11/11/2014 at 20:31

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

Path-integral analysis of fluctuation theorems for general Langevin processes

Vladimir Y Chernyak¹, Michael Chertkov^{2,3} and Christopher Jarzynski²

¹ Department of Chemistry, Wayne State University, 5101 Cass Ave, Detroit, MI 48202, USA

² T-13, Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA

³ CNLS, Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA

E-mail: chernyak@chem.wayne.edu, chertkov1@lanl.gov and chrisj@lanl.gov

Received 22 May 2006

Accepted 6 July 2006

Published 2 August 2006

Online at stacks.iop.org/JSTAT/2006/P08001

[doi:10.1088/1742-5468/2006/08/P08001](https://doi.org/10.1088/1742-5468/2006/08/P08001)

Abstract. We examine classical, transient fluctuation theorems within the unifying framework of Langevin dynamics. We explicitly distinguish between the effects of non-conservative forces that violate detailed balance, and non-autonomous dynamics arising from the variation of an external parameter. When both these sources of nonequilibrium behaviour are present, there naturally arise two distinct fluctuation theorems.

Keywords: rigorous results in statistical mechanics, driven diffusive systems (theory), exact results

ArXiv ePrint: [cond-mat/0605471](https://arxiv.org/abs/cond-mat/0605471)

Contents

1. Introduction	2
2. Stochastic modelling and definitions	3
3. Path integral formalism	6
4. Fluctuation theorem for reversed protocol	7
5. Fluctuation theorem for reversed protocol and dynamics	9
6. Physical interpretations	10
7. Illustrative example	12
8. Integrated fluctuation theorems and further extensions	14
9. Discussions and conclusions	15
Acknowledgments	16
Appendix: Discretization scheme	16
References	18

1. Introduction

The *fluctuation theorem* refers to a set of exact relations describing the statistical mechanics of systems away from equilibrium, generically expressed by the formula

$$\frac{p(+\Sigma)}{p(-\Sigma)} = e^{\Sigma}. \quad (1)$$

Here $p(\Sigma)$ is the distribution of observed values of a quantity representing dissipation or entropy production. The fluctuation theorem was originally formulated for non-conservative forces but autonomous dynamics [1]–[5], such as a fluid subject to constant shear, or a charged particle pushed through a thermal environment by a constant external field. Related results have also been derived for the case of non-autonomous dynamics [6]–[14], in which the system of interest is driven away from a stationary state by the external forcing of a work parameter, as when a piston is pushed into a gas of particles. For further extensions and unifying frameworks see [15]–[23], and for reviews, see [24]–[27].

In the current paper we present an exposition of transient fluctuation theorems within the path-integral formalism of Langevin dynamics. We consider two distinct mechanisms for achieving nonequilibrium behaviour: non-conservative forces (explicit violation of detailed balance), and non-autonomous dynamics (external forcing), and we distinguish the contributions that each of these factors makes to the fluctuation theorem. We will show that when both mechanisms are present, the definition of Σ is not unique, and there arise two distinct fluctuation theorems.

In the following two sections, we specify the general class of Langevin dynamics we consider in this paper, first using the Fokker–Planck formalism (section 2), then in the

path-integral representation (section 3). In section 4 we derive a fluctuation theorem, equation (43), by comparing the evolution of our system during a *forward* process (as an external parameter is varied from an initial value A to a final value B) to its evolution during the corresponding *reverse* process ($B \rightarrow A$). In section 5 we obtain a different fluctuation theorem, equation (51), by considering the effect of reversing not only the protocol for varying the external parameter, but also the underlying dynamics. In section 6 we discuss physical interpretations of the quantities appearing in our results, and in section 7 we illustrate these results using a simple model system with two degrees of freedom. Finally, in section 8 we discuss the *integrated* fluctuation theorems that follow immediately from equations (43) and (51); we present an alternative derivation of these integrated results, which in turn leads to an extension of the fluctuation theorems derived in sections 4 and 5.

2. Stochastic modelling and definitions

Consider an overdamped classical system described by the stochastic differential equation,

$$\frac{d}{dt}x_i = F_i(\mathbf{x}; \lambda) + \xi_i(t; \mathbf{x}; \lambda), \quad (2)$$

where x_i , with $i = 1, \dots, N$, denote a set of dynamical variables, and λ represents an externally controlled parameter. $\mathbf{F}$ represents the deterministic component of the dynamics; the stochastic component $\boldsymbol{\xi}$ is a δ -correlated noise field, whose mean is zero and whose pair correlation function is

$$\langle \xi_i(t) \xi_j(t') \rangle = G_{ij} \delta(t - t'), \quad (3)$$

where $\mathbf{G}(\mathbf{x}; \lambda)$ is a symmetric, positive definite matrix. Equation (2) can be viewed as the $\Delta \rightarrow 0$ limit of a Markov process with discrete time steps $t_{k+1} - t_k = \Delta$. In the appendix, we specify this limiting procedure, and show that an ensemble of systems governed by equation (2) is described by a probability density $p(\mathbf{x}, t)$, evolving under the Fokker–Planck equation

$$\frac{\partial p}{\partial t} = -\partial^i (F_i p) + \frac{1}{2} \partial^i [G_{ij} (\partial^j p)] \equiv \mathcal{L}_\lambda p. \quad (4)$$

Summation over repeated indices is implied. As the notation indicates, the Fokker–Planck operator $\mathcal{L}_\lambda$ depends explicitly on the parameter λ . Throughout this paper we will assume that when λ is held fixed, p relaxes exponentially to a unique stationary distribution

$$p^S(\mathbf{x}; \lambda) = \exp[-\varphi(\mathbf{x}; \lambda)], \quad (5)$$

hence

$$\mathcal{L}_\lambda \exp[-\varphi(\mathbf{x}; \lambda)] = 0 \quad (6)$$

and all other eigenvalues of $\mathcal{L}_\lambda$ are negative and bounded away from zero.

For fixed λ , the dynamics are specified by the vector field $\mathbf{F}$ and the matrix field $\mathbf{G}$. Now let $\boldsymbol{\Gamma}(\mathbf{x}; \lambda)$ denote the matrix inverse of $\mathbf{G}(\mathbf{x}; \lambda)$, i.e. $\Gamma^{ij} G_{jk} = \delta_k^i$, and consider two

vector fields, $\mathbf{v}(\mathbf{x}; \lambda)$ and $\mathbf{A}(\mathbf{x}; \lambda)$:

$$v^i = 2\Gamma^{ij}F_j \quad (7)$$

$$A^i = v^i + \partial^i \varphi. \quad (8)$$

These will play an important role in the following analysis. In terms of $\mathbf{v}$, equation (4) becomes

$$\frac{\partial p}{\partial t} = -\frac{1}{2}\partial^i [G_{ij}(v^j - \partial^j)p] = -\partial^i J_i, \quad (9)$$

where $\mathbf{J}(\mathbf{x}, t) = (1/2)\mathbf{G}(\mathbf{v} - \nabla)p$ is the current density. Substituting $p = p^S = e^{-\varphi}$, we get the stationary current density⁴:

$$J_i^S(\mathbf{x}; \lambda) = \frac{1}{2}G_{ij}A^j e^{-\varphi} = [F_i + \frac{1}{2}G_{ij}(\partial^j \varphi)] e^{-\varphi}. \quad (10)$$

This current is divergenceless:

$$\partial^i J_i^S = 0. \quad (11)$$

If $\mathbf{v}$ can be written as the gradient of a scalar field, $\mathbf{v}(\mathbf{x}; \lambda) = -\nabla U(\mathbf{x}; \lambda)$, then the stationary state is $p^S \propto e^{-U}$, as seen by inspection of equation (9). It then follows from equation (8) that $\mathbf{A} = \mathbf{0}$, which in turn implies a vanishing stationary current, $\mathbf{J}^S = \mathbf{0}$ (equation (10)). In this situation we say that the forces acting on the system are *conservative*, or equivalently that detailed balance is satisfied, and we interpret $p^S \propto e^{-U}$ as an equilibrium (canonical) distribution. By contrast, if $\mathbf{v} \neq -\nabla U$, then the forces are *non-conservative*, detailed balance is violated, and $\mathbf{A}, \mathbf{J}^S \neq \mathbf{0}$. We will thus view $\mathbf{A}$ as an indicator that distinguishes between conservative ($\mathbf{A} = \mathbf{0}$) and non-conservative ($\mathbf{A} \neq \mathbf{0}$) forces.

Another important distinction is that between *autonomous* and *non-autonomous* dynamics. The former refers to the situation in which we observe the evolution of the system with λ held fixed, while the latter denotes the case when the parameter is varied externally according to some schedule λ_t .

Throughout this paper we will consider processes during which the system evolves over a time interval $-\tau \leq t \leq +\tau$, and we will generally assume that initial conditions are sampled from the stationary state:

$$p(\mathbf{x}, -\tau) = \exp[-\varphi(\mathbf{x}; \lambda_{-\tau})]. \quad (12)$$

(See however the end of section 6 as well as [17, 18] for discussions of more general initial conditions.) If the forces are conservative and the dynamics autonomous, then the system remains in equilibrium over the interval of observation: $p(\mathbf{x}, t) \propto \exp[-U(\mathbf{x}; \lambda_{\text{fixed}})]$. However, if we have either non-conservative forces or non-autonomous dynamics, or both, then we achieve nonequilibrium behaviour, for which fluctuation theorems can be derived.

Finally, rearranging equation (10) to express $\mathbf{F}$ as a function of $\mathbf{G}$, φ , and $\mathbf{J}^S$, we can rewrite the Fokker–Planck operator explicitly in terms of the stationary density and current:

$$\mathcal{L}_\lambda p = -\partial^i (J_i^S e^\varphi p) - \frac{1}{2}\partial^i [G_{ij}e^{-\varphi}\partial^j (e^\varphi p)]. \quad (13)$$

⁴ For a Brownian particle in one dimension, analogues of our equations (10) and (8) are given by equations (2) and (3) of [28].

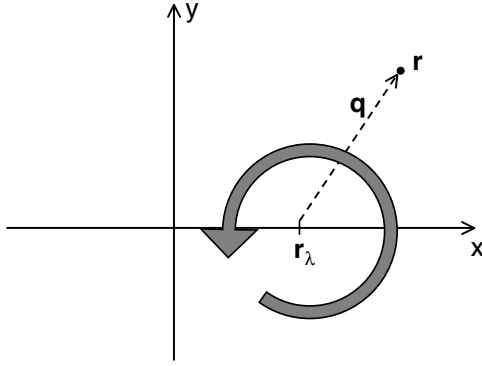


Figure 1. Depiction of the illustrative model described in the text. The point $\mathbf{r}$ is the location of a Brownian particle inside a harmonic well, $kq^2/2$, centred at $\mathbf{r}_\lambda$. The grey arrow represents the angular drift force $f(q)\hat{\boldsymbol{\theta}}$, also centred at $\mathbf{r}_\lambda$.

By considering various choices of (divergenceless) $\mathbf{J}^S$, while keeping $\mathbf{G}$ and φ fixed, we explore a family of stochastic dynamics with a common stationary density but different stationary currents. If $\mathcal{L}_\lambda$ corresponds to a given choice of $\{\mathbf{J}^S, \varphi, \mathbf{G}\}$, then we will use the notation $\hat{\mathcal{L}}_\lambda$ to denote the Fokker–Planck operator corresponding to $\{-\mathbf{J}^S, \varphi, \mathbf{G}\}$. By equation (10), the deterministic component of the dynamics associated with $\hat{\mathcal{L}}_\lambda$ is given by

$$\hat{F}_i = -F_i - G_{ij}\partial^j\varphi. \quad (14)$$

Using the path-integral formalism discussed in the following sections (equation (48) in particular) it is straightforward to establish that

$$P_\lambda(\mathbf{x}', \Delta t | \mathbf{x}) p^S(\mathbf{x}) = \hat{P}_\lambda(\mathbf{x}, \Delta t | \mathbf{x}') p^S(\mathbf{x}') \quad (15)$$

for any $\Delta t > 0$, where P_λ and $\hat{P}_\lambda$ are transition probabilities associated with the dynamics generated by $\mathcal{L}_\lambda$ and $\hat{\mathcal{L}}_\lambda$, respectively. Recognizing each side as a joint probability for observing a pair of events (separated by a time interval Δt) in the stationary state, equation (15) is interpreted as follows: if we generate an infinitely long trajectory using the dynamics $\mathcal{L}_\lambda$, and we then replace this trajectory by its time-reversed image, $\mathbf{x}_t \rightarrow \mathbf{x}_{-t}$, then the new trajectory will be statistically indistinguishable from a trajectory generated by $\hat{\mathcal{L}}_\lambda$. In particular the two trajectories will be characterized by the same stationary density, p^S , but opposite currents, $\pm\mathbf{J}^S$. When two stochastic dynamics are related by equation (15), we say that the one is the *reversal* [29], or the *p^S -dual* [30], of the other. This natural pairing of Fokker–Planck operators $(\mathcal{L}_\lambda, \hat{\mathcal{L}}_\lambda)$ will play an important role in section 5, where the analysis is very similar to that carried out by Crooks for discrete-time Markov processes [12]. Note that $\mathcal{L}_\lambda = \hat{\mathcal{L}}_\lambda$ when $\mathbf{J}^S = \mathbf{0}$; in this case equation (15) is just the familiar statement of detailed balance associated with conservative forces.

As a simple illustrative model, shown in figure 1, consider a particle in two dimensions, $\mathbf{r} = (x, y)$, with

$$\mathbf{F}(\mathbf{r}; \lambda) = -\frac{k}{\gamma}\mathbf{q} + \frac{1}{\gamma}f(q)\hat{\boldsymbol{\theta}}, \quad G_{ij}(\mathbf{r}; \lambda) = \frac{2}{\gamma}\delta_{ij}. \quad (16)$$

Here $\mathbf{q} = \mathbf{r} - \mathbf{r}_\lambda$ is the displacement of the particle relative to a point $\mathbf{r}_\lambda = (\lambda, 0)$; and the unit vector $\hat{\boldsymbol{\theta}}$ is normal (oriented 90° counterclockwise) to the unit vector $\hat{\mathbf{q}} = \mathbf{q}/q$. When

$f = 0$, equation (16) describes a Brownian particle in a harmonic well centred at $\mathbf{r}_\lambda$, with spring constant $k > 0$, friction coefficient $\gamma > 0$, and temperature $k_B T = 1$. When $f(q)$ is positive (negative), this particle experiences an additional counterclockwise (clockwise) drift around the point $\mathbf{r}_\lambda$. For this model, we have⁵:

$$p^S \propto e^{-kq^2/2}, \quad \mathbf{A} = f(q)\hat{\boldsymbol{\theta}}, \quad \mathbf{J}^S = \frac{1}{\gamma}f(q)\hat{\boldsymbol{\theta}}e^{-\varphi}. \quad (17)$$

Thus the stationary state is characterized by a Gaussian distribution, p^S , and an average angular drift around the point $\mathbf{r}_\lambda$. If $\mathcal{L}_\lambda$ denotes the Fokker–Planck operator for a given choice of k , γ , and $f(q)$, then $\hat{\mathcal{L}}_\lambda$ is obtained by reversing the direction of the angular drift:

$$\mathcal{L}_\lambda \leftrightarrow \{k, \gamma, +f(q)\} \Rightarrow \hat{\mathcal{L}}_\lambda \leftrightarrow \{k, \gamma, -f(q)\}. \quad (18)$$

We will use this model in section 7 to illustrate the central results of this paper.

3. Path integral formalism

The main theoretical tool that we will use in this paper is the path-integral representation of Langevin dynamics [31, 32]. Let $X \equiv \{\mathbf{x}_t\}_{-\tau}^{+\tau}$ denote a trajectory that specifies the evolution of the system from $-\tau$ to $+\tau$. When λ is held fixed, the conditional probability of observing this trajectory, given the initial microstate $\mathbf{x}_{-\tau}$, is

$$\mathcal{P}_\lambda[X|\mathbf{x}_{-\tau}] = \mathcal{N} \exp \left[- \int_{-\tau}^{+\tau} dt \mathcal{S}_+(\mathbf{x}_t, \dot{\mathbf{x}}_t; \lambda) \right] \quad (19)$$

$$\mathcal{S}_+(\mathbf{x}, \dot{\mathbf{x}}; \lambda) = \frac{1}{2}(\dot{\mathbf{x}}_i - F_i)\Gamma^{ij}(\dot{\mathbf{x}}_i - F_i) + \frac{1}{2}\partial^i F_i. \quad (20)$$

As discussed in the appendix, the continuous-time integral in equation (19) is interpreted as the limit of a discrete sum, using mid-point (Stratonovich) discretization.

Imagine that the system evolves as λ is varied externally from an initial value $\lambda_{-\tau}^F = A$ to a final value $\lambda_{+\tau}^F = B$, according to a schedule, or *protocol*, λ_t^F . We refer to this as the *forward process*, indicated by the superscript F. During this process, the system satisfies

$$\frac{d}{dt}x_i = F_i(\mathbf{x}; \lambda_t^F) + \xi_i(t; \mathbf{x}; \lambda_t^F). \quad (21)$$

The conditional probability of observing a trajectory X is obtained by a straightforward generalization of equation (19):

$$\mathcal{P}^F[X|\mathbf{x}_{-\tau}] = \mathcal{N} \exp \left[- \int_{-\tau}^{+\tau} dt \mathcal{S}_+(\mathbf{x}_t, \dot{\mathbf{x}}_t; \lambda_t^F) \right]. \quad (22)$$

If we sample $\mathbf{x}_{-\tau}$ from the stationary distribution $p^S(\mathbf{x}; A)$ (equation (12)), the net (unconditional) probability of observing the trajectory X is

$$\mathcal{P}^F[X] = p_A^S(\mathbf{x}_{-\tau})\mathcal{P}^F[X|\mathbf{x}_{-\tau}], \quad (23)$$

where $p_A^S(\mathbf{x}_{-\tau}) \equiv p^S(\mathbf{x}_{-\tau}; A)$.

⁵ Note that p^S is unaffected by the angular drift force $f(q)\hat{\boldsymbol{\theta}}$. This non-generic feature of our model arises from the fact that the flow $f(q)\hat{\boldsymbol{\theta}}/\gamma$ is everywhere parallel to the contours of the harmonic potential $kq^2/2$.

Along with the forward process, we will consider a *reverse process*, during which the parameter is manipulated from B to A . Specifically, during the reverse process we have

$$\frac{d}{dt}x_i = F_i(\mathbf{x}; \lambda_t^R) + \xi_i(t; \mathbf{x}; \lambda_t^R), \quad (24)$$

where

$$\lambda_t^R = \lambda_{-t}^F. \quad (25)$$

The conditional and unconditional probabilities of observing a trajectory X during this process are given by the analogues of equations (22) and (23):

$$\mathcal{P}^R[X|\mathbf{x}_{-\tau}] = \mathcal{N} \exp \left[- \int_{-\tau}^{+\tau} dt \mathcal{S}_+(\mathbf{x}_t, \dot{\mathbf{x}}_t; \lambda_t^R) \right] \quad (26)$$

$$\mathcal{P}^R[X] = p_B^S(\mathbf{x}_{-\tau}) \mathcal{P}^R[X|\mathbf{x}_{-\tau}]. \quad (27)$$

Here we have assumed the same underlying stochastic dynamics—that is, the same family of Fokker–Planck operators $\mathcal{L}_\lambda$ —for the reverse process as for the forward process; the only distinction between the two processes is the protocol for varying λ .

4. Fluctuation theorem for reversed protocol

Now let $X^\dagger \equiv \{\mathbf{x}_t^\dagger\}_{-\tau}^{+\tau}$ denote the time-reversed ‘conjugate twin’ of a trajectory X :

$$\mathbf{x}_t^\dagger = \mathbf{x}_{-t}, \quad (28)$$

and let us compare the probability of observing a trajectory X during the forward process with that of its twin $X^\dagger$ during the reverse process. Using equations (25), (26) and (28), we get

$$\mathcal{P}^R[X^\dagger|\mathbf{x}_{-\tau}^\dagger] = \mathcal{N} \exp \left[- \int_{-\tau}^{+\tau} dt \mathcal{S}_+(\mathbf{x}_t^\dagger, \dot{\mathbf{x}}_t^\dagger; \lambda_t^R) \right] = \mathcal{N} \exp \left[- \int_{-\tau}^{+\tau} dt \mathcal{S}_-(\mathbf{x}_t, \dot{\mathbf{x}}_t; \lambda_t^F) \right], \quad (29)$$

where

$$\mathcal{S}_-(\mathbf{x}, \dot{\mathbf{x}}; \lambda) \equiv \mathcal{S}_+(\mathbf{x}, -\dot{\mathbf{x}}; \lambda). \quad (30)$$

The definitions of $\mathbf{v}$, $\mathbf{A}$, and $\mathcal{S}_\pm$ then give us

$$\mathcal{S}_+ - \mathcal{S}_- = -\dot{x}_j v^j = \dot{x}_j (\partial^j \varphi - A^j), \quad (31)$$

which combines with equations (22) and (29) to yield

$$\frac{\mathcal{P}^F[X|\mathbf{x}_{-\tau}]}{\mathcal{P}^R[X^\dagger|\mathbf{x}_{-\tau}^\dagger]} = \exp \left(\int^F dt \dot{x}_j v^j \right), \quad (32)$$

where $\int^F$ denotes evaluation along the forward protocol and trajectory X . When the diffusion tensor is a constant, $\mathbf{G}(\mathbf{x}; \lambda) = 2D\delta_{ij}$, this ratio is equivalent to a result derived previously by Seifert (equation (14) of [18]), and extended to inertial systems by Imparato

and Peliti (equation (34) of [22]). From equation (32), we obtain

$$\frac{\mathcal{P}^F[X]}{\mathcal{P}^R[X^\dagger]} = \frac{p_A^S(\mathbf{x}_{-\tau})\mathcal{P}^F[X|\mathbf{x}_{-\tau}]}{p_B^S(\mathbf{x}_{-\tau}^\dagger)\mathcal{P}^R[X^\dagger|\mathbf{x}_{-\tau}^\dagger]} = \exp\left(\Delta\varphi + \int^F dt \dot{x}_j v^j\right), \quad (33)$$

where

$$\Delta\varphi \equiv \varphi(\mathbf{x}_\tau; B) - \varphi(\mathbf{x}_{-\tau}; A) = \int^F dt \left(\dot{\lambda}_t^F \frac{\partial\varphi}{\partial\lambda} + \dot{x}_j \partial^j \varphi \right). \quad (34)$$

Let us now rewrite equation (33) so that the quantity in the exponent is manifestly a sum of contributions representing non-autonomous dynamics and non-conservative forces. Defining

$$Y^F \equiv \int^F dt \dot{\lambda}_t^F \frac{\partial\varphi}{\partial\lambda}, \quad (35)$$

we obtain

$$\frac{\mathcal{P}^F[X]}{\mathcal{P}^R[X^\dagger]} = \exp\left(Y^F + \int^F d\mathbf{x} \cdot \mathbf{A}\right). \quad (36)$$

A non-zero value of Y^F is a signature of non-autonomous dynamics (equation (35)), whereas $\mathbf{A} \neq 0$ indicates non-conservative forces (section 2). Thus we associate the two terms, Y^F and $\int^F d\mathbf{x} \cdot \mathbf{A}$, with the two mechanisms for achieving nonequilibrium behaviour identified at the end of section 2. Let R denote the sum of these two terms:

$$R^F[X] \equiv Y^F + \int^F d\mathbf{x} \cdot \mathbf{A}. \quad (37)$$

When the dynamics are autonomous and the forces conservative, both terms are equal to zero, hence $\mathcal{P}[X] = \mathcal{P}[X^\dagger]$: in equilibrium, any sequence of events is as likely as the reverse sequence; see for example the discussion following equation (13) of [33].

For the reverse process we similarly define

$$Y^R \equiv \int^R dt \dot{\lambda}_t^R \frac{\partial\varphi}{\partial\lambda}, \quad R^R \equiv Y^R + \int^R d\mathbf{x} \cdot \mathbf{A}. \quad (38)$$

The quantities Y and R are odd under time-reversal, in the following sense:

$$Y^R[X^\dagger] = -Y^F[X], \quad R^R[X^\dagger] = -R^F[X]. \quad (39)$$

With this formalism in place, we now derive a fluctuation theorem for $R = Y + \int d\mathbf{x} \cdot \mathbf{A}$. Let $\rho^F(R)$ denote the distribution of R values for an ensemble of realizations of the forward process, and define $\rho^R(R)$ analogously for the reverse process. Then

$$\rho^F(R) = \int \mathcal{D}X \mathcal{P}^F[X] \delta(R - R^F[X]) \quad (40)$$

$$= \int \mathcal{D}X \mathcal{P}^R[X^\dagger] \exp(R^F[X]) \delta(R - R^F[X]) \quad (41)$$

$$= \exp(R) \int \mathcal{D}X^\dagger \mathcal{P}^R[X^\dagger] \delta(R + R^R[X^\dagger]), \quad (42)$$

where $\mathcal{D}X = \prod_{k=0}^K d^N x_k$ specifies an integral over all possible trajectories X (see appendix). We have used equation (36) to get from the first line to the second, and equation (39) to get to the third. Note also the change of variables, $\mathcal{D}X \rightarrow \mathcal{D}X^\dagger$. Recognizing the final integral as $\rho^R(-R)$, we obtain the desired fluctuation theorem [18]:

$$\frac{\rho^F(+R)}{\rho^R(-R)} = \exp(R). \quad (43)$$

5. Fluctuation theorem for reversed protocol and dynamics

The evolution of the system is influenced by both the protocol for varying λ , and the stochastic dynamics that define the Langevin transition rates. In the previous section we assumed that the forward and reverse processes are defined by conjugate protocols but the same underlying dynamics (equations (21), (24)). Thus the reverse process was defined relative to the forward process by the replacement

$$\{\mathcal{L}_\lambda, \lambda_t^F\} \rightarrow \{\mathcal{L}_\lambda, \lambda_t^R\}. \quad (44)$$

In this section we obtain a different fluctuation theorem by imagining that the reverse process is characterized by a reversal of both the protocol and the underlying dynamics.

Specifically, we imagine that during the forward process the system satisfies equation (21), as in the previous section. However, for the reverse process, we take

$$\frac{d}{dt}x_i = \hat{F}_i(\mathbf{x}; \lambda_t^R) + \xi_i(t; \mathbf{x}; \lambda_t^R), \quad (45)$$

where $\hat{\mathbf{F}} = -\mathbf{F} - \mathbf{G}\nabla\varphi$ (equation (14)), rather than equation (24). Thus the reverse process is now defined by the replacement

$$\{\mathcal{L}_\lambda, \lambda_t^F\} \rightarrow \{\hat{\mathcal{L}}_\lambda, \lambda_t^R\}. \quad (46)$$

In this situation the conditional probability of a trajectory X during the reverse process is

$$\hat{\mathcal{P}}^R[X|\mathbf{x}_{-\tau}] = \mathcal{N} \exp \left[- \int_{-\tau}^{+\tau} dt \hat{\mathcal{S}}_+(\mathbf{x}_t, \dot{\mathbf{x}}_t; \lambda_t^R) \right], \quad (47)$$

where $\hat{\mathcal{S}}_+$ is defined as its counterpart $\mathcal{S}_+$, but with $\mathbf{F}$ replaced by $\hat{\mathbf{F}}$. Let us similarly define $\hat{\mathcal{S}}_-(\mathbf{x}, \dot{\mathbf{x}}; \lambda) \equiv \hat{\mathcal{S}}_+(\mathbf{x}, -\dot{\mathbf{x}}; \lambda)$ (as in equation (30)). By direct evaluation we obtain

$$\mathcal{S}_+ - \hat{\mathcal{S}}_- = \dot{x}_i \partial^i \varphi + e^\varphi \partial^i \left[\left(F_i - \frac{1}{2} G_{ij} \partial^j \right) e^{-\varphi} \right] = \dot{x}_i \partial^i \varphi, \quad (48)$$

since the quantity in square brackets is just the i th component of the divergenceless stationary current (equation (10)).

Repeating the steps of section 4, but with equation (48) in place of equation (31), we get

$$\frac{\mathcal{P}^F[X|\mathbf{x}_{-\tau}]}{\hat{\mathcal{P}}^R[X^\dagger|\mathbf{x}_{-\tau}^\dagger]} = \exp \left(- \int^F dt \dot{x}_j \partial^j \varphi \right), \quad (49)$$

and in turn

$$\frac{\mathcal{P}^F[X]}{\hat{\mathcal{P}}^R[X^\dagger]} = \exp \left(\Delta\varphi - \int^F dt \dot{x}_i \partial^i \varphi \right) = \exp \left(\int^F dt \dot{\lambda}_t^F \frac{\partial \varphi}{\partial \lambda} \right) = \exp(Y^F). \quad (50)$$

This is identical to equation (36), except that the term $\int^F d\mathbf{x} \cdot \mathbf{A}$ no longer appears inside the exponent on the right-hand side. In effect, by using $\hat{\mathcal{L}}_\lambda$ for the reverse process, equation (45), we have ‘gauged away’ the non-conservative contribution arising from $\mathbf{A}$ (compare equations (31) and (48)), leaving only the non-autonomous contribution associated with the variation of λ .

Equation (50) leads to the analogue of equation (43):

$$\frac{\rho^F(+Y)}{\hat{\rho}^R(-Y)} = \exp(Y), \quad (51)$$

where ρ^F is the distribution of Y values for the forward process $(\mathcal{L}_\lambda, \lambda_t^F)$, and $\hat{\rho}^R$ is defined similarly for the reverse process $(\hat{\mathcal{L}}_\lambda, \lambda_t^R)$. This is a continuous-time analogue of the fluctuation theorem obtained by Crooks for discrete-time processes (our equations (49) and (51) correspond to equations (13) and (20) of [12]).

We thus see that two different fluctuation theorems (equations (43), (51)) emerge naturally in the Langevin formalism that we have considered. The relationship between these two results can be appreciated by separately considering the three different combinations for achieving nonequilibrium behaviour.

- *Non-conservative forces, autonomous dynamics* ($\mathbf{A} \neq \mathbf{0}$, $\dot{\lambda} = 0$). In this case $Y = 0$ by definition, hence equation (51) has no meaning. Thus we get a unique fluctuation theorem, equation (43), for a quantity $R = \int d\mathbf{x} \cdot \mathbf{A}$ associated with the violation of detailed balance.
- *Conservative forces, non-autonomous dynamics* ($\mathbf{A} = \mathbf{0}$, $\dot{\lambda} \neq 0$). In this case $\mathcal{L}_\lambda = \hat{\mathcal{L}}_\lambda$ and $Y = R$. Thus equation (43) is equivalent to equation (51), so we again obtain a unique fluctuation theorem, this time for the quantity $Y = R = \int dt \dot{\lambda} \partial \varphi / \partial \lambda$ associated with external forcing.
- *Non-conservative forces, non-autonomous dynamics* ($\mathbf{A} \neq \mathbf{0}$, $\dot{\lambda} \neq 0$). Equations (43) and (51) now represent two different, but equally valid, fluctuation theorems, for the distinct quantities $Y = \int dt \dot{\lambda} \partial \varphi / \partial \lambda$ and $R = Y + \int d\mathbf{x} \cdot \mathbf{A}$.

The non-uniqueness of the fluctuation theorem in the third case listed above is analogous to a result obtained by Fingerle for a relativistic, cosmological model: in that situation two distinct fluctuation theorems arise, depending on whether or not the cosmic expansion is inverted for the reverse process [34].

6. Physical interpretations

To this point our analysis has been mostly abstract and mathematical. For the Langevin process defined by equation (2), we have derived two fluctuation theorems, and in section 8 below, closely related *integrated* fluctuation theorems are obtained. In the present section we briefly discuss physical interpretations of the quantities appearing in these results.

Equation (2) can be used to model the microscopic evolution of a system in contact with a thermal reservoir at temperature T , in the overdamped limit, with λ playing the role of an externally manipulated work parameter. The stationary state is then given by the Boltzmann–Gibbs distribution $p^S = e^{\beta(F-U)}$, with $\beta = (k_B T)^{-1}$; hence

$$\varphi(\mathbf{x}; \lambda) = \beta U(\mathbf{x}; \lambda) - \beta F(\lambda), \quad (52)$$

where $U(\mathbf{x}; \lambda)$ represents the internal energy of the system, and $F(\lambda)$ is the parameter-dependent free energy. (We leave implicit the temperature dependence of F .) We generically expect detailed balance to hold in such an equilibrium state, hence the dynamics are conservative: $\mathbf{A}, \mathbf{J}^S = 0$. Under these circumstances, we get

$$Y = R = \beta \left(\int dt \dot{\lambda} \frac{\partial U}{\partial \lambda} - \Delta F \right) = \beta W_{\text{diss}}, \quad (53)$$

where $\Delta F \equiv F(B) - F(A)$, and $W_{\text{diss}} = W - \Delta F$ is physically interpreted as *dissipated work* [7]. There is no distinction between equations (43) and (51) in this situation; both are equivalent to the Crooks fluctuation theorem [10], and the corresponding integrated result (equation (64) below) is the nonequilibrium work theorem,

$$\langle e^{-\beta W} \rangle = e^{-\beta \Delta F}. \quad (54)$$

When $\mathbf{A} \neq 0$, the dynamics are non-conservative. To gain insight into the physical meaning of the vector field $\mathbf{A}$, consider an overdamped, one-dimensional Brownian particle at temperature T , with periodic boundary conditions (see for example figure 1 of [21]):

$$\gamma \dot{x} = -\frac{\partial U}{\partial x} + f + \zeta(t), \quad (55)$$

where $\gamma > 0$ and f are constants; $U(x; \lambda)$ is a periodic potential; and the noise term satisfies $\langle \zeta(0)\zeta(t) \rangle = 2\gamma k_B T \delta(t)$. When λ is held fixed the system relaxes to $p^S = e^{-\varphi}$. Comparing with equation (2), and using the definition of $\mathbf{A}$ (which for this example is a scalar field), we obtain

$$A = \beta (-U' + f) + \varphi', \quad (56)$$

where the primes denote $\partial/\partial x$. In previous studies of this example by Hatano and Sasa [14] and Speck and Seifert [21], the quantity $Q_{\text{hk}} \equiv \beta^{-1} \int dx A$ was identified as the ‘house-keeping heat’, a concept introduced earlier by Oono and Paniconi [35]. In the stationary state, Q_{hk} represents the heat absorbed by the external reservoir; this fluctuating quantity grows linearly with time, on average, and can be viewed as the thermodynamic price that must continually be paid to maintain the system away from equilibrium. Equivalently, $\beta Q_{\text{hk}} = \int dx A$ is the increase in the entropy of the reservoir. We speculate that this interpretation remains valid more generally: when equation (2) models a system in contact with thermal surroundings, perhaps including multiple heat baths, $\int d\mathbf{x} \cdot \mathbf{A}$ is the entropy generated in these surroundings, in the stationary state. This interpretation emphasizes the physical connection between entropy generation and the violation of detailed balance.

We see that Y can be interpreted as dissipated work (in units of $k_B T$) when the dynamics are conservative, and $\int d\mathbf{x} \cdot \mathbf{A}$ as the entropy generation needed to maintain a nonequilibrium stationary state. For the non-autonomous, non-conservative case, we do not have simple thermodynamic interpretations of Y and R . However, we can provide some intuition regarding the difference between these quantities by considering a quasi-static process, with λ varied slowly from A to B . Then $Y \rightarrow 0$ [14], whereas $\int d\mathbf{x} \cdot \mathbf{A}$ can be expected to grow diffusively, with slowly changing drift and diffusion constants. Hence as $\tau \rightarrow \infty$, the distribution of Y values tends to a delta-function, while the mean and variance of the distribution of R values scale like τ .

Finally, let us briefly consider a generalization of equation (43). The quantity $\Delta\varphi$ appearing in equation (33) is essentially a boundary term [18], arising from the ratio of

the probabilities of sampling the initial conditions of the conjugate pair of trajectories, X and $X^\dagger$. If we choose to sample initial conditions from distributions other than the stationary distribution, p^S , then this term must correspondingly be modified. Specifically, consider a family of normalized distributions $q_\lambda(\mathbf{x}) = \exp[-\psi(\mathbf{x}; \lambda)]$, where $\psi(\mathbf{x}; \lambda)$ is arbitrary (apart from the normalization condition). Now imagine that we define our forward (reverse) process by sampling the initial conditions from q_A (q_B) rather than p_A^S (p_B^S). The boundary term in equation (33) then changes from $\Delta\varphi$ to $\Delta\psi$, which ultimately leads to the result

$$\frac{\rho^F(+R^\psi)}{\rho^R(-R^\psi)} = \exp(R^\psi), \quad (57)$$

where

$$R^{\psi,F} = \int^F dt \dot{\lambda}_t^F \frac{\partial \psi}{\partial \lambda} + \int^F d\mathbf{x} \cdot \mathbf{A}^\psi, \quad \mathbf{A}^\psi = \mathbf{v} + \nabla \psi, \quad (58)$$

and similarly for the reverse process. Equation (57) thus generalizes equation (43) to allow for initial conditions sampled from arbitrary distributions.

As a specific example, suppose there exists a natural decomposition $\beta\mathbf{v} = -\nabla U + \mathbf{f}$, where $\mathbf{f}$ cannot be expressed as the gradient of a scalar field; for example, imagine a Brownian particle exploring a potential landscape $U(\mathbf{x}; \lambda)$, but also subject to a non-conservative force $\mathbf{f}(\mathbf{x}; \lambda)$. This is a generalization of the one-dimensional example discussed above (equation (55)). Now suppose we sample initial conditions from the canonical distribution $q \propto \exp(-\beta U)$, rather than the stationary distribution $p^S = \exp(-\varphi)$. This corresponds to the choice $\psi = \beta U$, where for simplicity we have incorporated the free energy $F(\lambda)$ into the definition of $U(\mathbf{x}; \lambda)$. The quantity R^ψ appearing in equation (57) then becomes

$$R^{\psi,F} = \beta \int^F dt \dot{\lambda}_t^F \frac{\partial U}{\partial \lambda} + \beta \int^F d\mathbf{x} \cdot \mathbf{f}. \quad (59)$$

We thus get a fluctuation theorem for a quantity R^ψ that is physically interpreted as the sum of contributions due to the external variation of a conservative potential, U , and the path integral of a non-conservative force, $\mathbf{f}$. Neither term depends on the stationary distribution φ . This result was originally obtained by Kurchan, first for autonomous dynamics [4] ($R^\psi = \beta \int d\mathbf{x} \cdot \mathbf{f}$), and more recently for non-autonomous dynamics [19], assuming a spatially independent diffusion coefficient.

7. Illustrative example

In this section we illustrate the fluctuation theorems derived above, using the example introduced at the end of section 2. We begin with the autonomous case. If $f(q) = 0$, then we simply have a Brownian particle fluctuating in equilibrium in a two-dimensional harmonic well. When $f(q) > 0$, the particle is subject to an additional angular drift, oriented counterclockwise. Let us picture this drift to be the result of a vortex in the surrounding thermal medium, centred at $\mathbf{r}_\lambda$. In this case the integral

$$\int d\mathbf{x} \cdot \mathbf{A} = \int dt \dot{\mathbf{r}}_t \cdot \hat{\boldsymbol{\theta}} f(\mathbf{q}_t) \equiv \mathcal{C}[X] \quad (60)$$

provides a measure of the counterclockwise motion of the particle. We will refer to $\mathcal{C}$ as the *circulation* associated with a given trajectory X . Since we are considering autonomous dynamics, we have $R = \mathcal{C}$ (see equation (37)), and equation (43) becomes

$$\frac{\rho(+\mathcal{C})}{\rho(-\mathcal{C})} = \exp(\mathcal{C}). \quad (61)$$

Here $\rho(\mathcal{C})$ is the probability distribution of observing a circulation $\mathcal{C}$, over the interval of observation, assuming initial conditions sampled from the stationary distribution (equation (17)). Since λ is fixed, there is no distinction between the forward and reverse process. Equation (61) implies that positive values of circulation are more likely than negative values, as expected: the particle tends to flow with, rather than against, the vortex.

Now consider the case of non-autonomous dynamics, but conservative forces, $f(q) = 0$. For specificity, imagine that during the forward process we vary λ at a constant rate from A to B ; for example, we move the point $\mathbf{r}_\lambda$ rightwards along the $\mathbf{x}$ -axis. Thus we drag the Brownian particle through a thermal medium, by moving the harmonic potential in which it is trapped. During the reverse process we go from B back to A , at the same speed. Since the forces are conservative ($\mathbf{A}, \mathbf{J}^S = 0$, hence $\mathcal{L}_\lambda = \hat{\mathcal{L}}_\lambda$), there is no distinction between the predictions of section 4 and those of section 5; moreover there is no tendency for the particle to circulate in one direction or the other. We have

$$R = Y \equiv \int dt \dot{\lambda} \frac{\partial \varphi}{\partial \lambda} = -k \int d\lambda (x_t - \lambda_t), \quad (62)$$

where x_t is the x -component of the particle's position at time t . Physically, $-k(x - \lambda) d\lambda$ is the incremental work required to displace the potential well by an amount $d\lambda$; hence Y is equal to the total external work performed during a realization of either the forward or the reverse process (in units of temperature, since $k_B T = 1$ for this example). Equations (43) and (51) are identical in this situation, and are expressed as

$$\frac{\rho^F(+Y)}{\rho^R(-Y)} = \exp(Y). \quad (63)$$

Positive values of work are more likely than negative values, in agreement with the second law of thermodynamics.

Having obtained one prediction for the case of autonomous dynamics and non-conservative forces (equation (61) for $\mathcal{C}$), and another for conservative forces and non-autonomous dynamics (equation (63) for Y), we now consider the combination of non-autonomous dynamics and non-conservative forces. Thus $\mathbf{r}_\lambda$ now specifies the moving centre of both a harmonic well and a vortex in the thermal medium. Let $\mathcal{L}_\lambda$ denote the stochastic dynamics for a given choice $f(q) > 0$, describing a counterclockwise vortex; then $\hat{\mathcal{L}}_\lambda$ corresponds to the replacement $f \rightarrow -f$, describing a clockwise vortex.

If we use $\mathcal{L}_\lambda$ for both the forward and the reverse processes (equations (21), (24)), then this amounts to dragging the particle either rightwards ($A \rightarrow B$) or leftwards ($B \rightarrow A$), but in either case the particle tends to circulate counterclockwise around the moving point $\mathbf{r}_\lambda$. The fluctuation theorem in this situation, equation (43), involves the quantity $R = Y + \mathcal{C}$, with external work Y and circulation $\mathcal{C}$ as defined by equations (62) and (60) above. On the other hand, if we use $\hat{\mathcal{L}}_\lambda$ for the reverse process (equation (45)), then the

particle tends to circulate counterclockwise as it is dragged rightwards during the forward process, but clockwise as it is dragged leftwards during the reverse process. In this case the applicable fluctuation theorem is equation (51) for the work Y ; the circulation $\mathcal{C}$ no longer contributes to the quantity appearing in the fluctuation theorem.

8. Integrated fluctuation theorems and further extensions

Multiplying both sides of equation (43) by $\rho^R(-R) \exp(-R)$, then integrating with respect to R , and performing similar manipulations on equation (51), we are led to the *integrated* results,

$$\langle \exp(-R) \rangle^F = 1, \quad \langle \exp(-Y) \rangle^F = 1, \quad (64)$$

where the angular brackets denote an average over realizations of the forward process. For the general case of non-autonomous and non-conservative dynamics, with initial conditions sampled from the stationary state, these results have been discussed by Hatano and Sasa [14] (for Y), and by Speck and Seifert [21] (for R). Unlike equations (43) and (51), the integrated fluctuation theorems can be stated without any reference to the reverse process, suggesting that they can be derived by some means other than a comparison between forward and reverse processes. We now sketch such a derivation, which in turn leads to a further generalization of both integrated and non-integrated results (equations (70), (71)).

We assume our system obeys the equation of motion

$$\frac{d}{dt}x_i = F_i(\mathbf{x}; \lambda_t) + \xi_i(t; \mathbf{x}; \lambda_t), \quad (65)$$

where λ_t denotes a protocol for varying the parameter λ . For a trajectory generated during this process, consider the variable

$$\omega_t = \int_{-\tau}^t dt' \left[\dot{\lambda} \frac{\partial \varphi}{\partial \lambda} + \alpha \dot{\mathbf{x}} \cdot \mathbf{A} + \frac{1}{2} \alpha (\alpha - 1) G_{ij} A^i A^j \right], \quad (66)$$

where α is a constant, and the integrand is evaluated along the trajectory. For $\alpha = 0$ and 1, we get $\omega_\tau = W$ and R , respectively.

Now consider a function

$$\mathcal{G}(\mathbf{x}, t) = \langle \delta(\mathbf{x} - \mathbf{x}_t) \exp(-\omega_t) \rangle, \quad (67)$$

where the average is taken over an ensemble of realizations of the process, and $\mathbf{x}_t$ is the microstate at time t during a given realization. This function can be viewed as a ‘weighted’ phase space density, in which each realization carries a time-dependent statistical weight $\exp(-\omega_t)$; similar constructions have been considered in [8, 13, 17, 22]. Following a procedure analogous to the derivation of equation (4) in the appendix, we obtain an evolution equation for this density:

$$\frac{\partial}{\partial t} \mathcal{G} = \mathcal{L}_\lambda \mathcal{G} - \dot{\lambda} \frac{\partial \varphi}{\partial \lambda} \mathcal{G} + \alpha \partial^i (G_{ij} A^j \mathcal{G}). \quad (68)$$

The first term on the right represents the dynamical component of the evolution (equation (4)), while the other two arise from the weight $\exp(-\omega_t)$ assigned to each trajectory in the ensemble. Using equations (6), (10), and (11), we see by inspection that the *ansatz* $\mathcal{G}(\mathbf{x}, t) = \exp[-\varphi(\mathbf{x}; \lambda_t)]$ solves equation (68). Thus when initial conditions

are sampled from the stationary distribution $p^S(\mathbf{x}; \lambda_{-\tau})$, we have

$$\langle \delta(\mathbf{x} - \mathbf{x}_t) \exp(-\omega_t) \rangle = \exp[-\varphi(\mathbf{x}; \lambda_t)]. \quad (69)$$

Setting $t = +\tau$ and integrating both sides with respect to $\mathbf{x}$, we get

$$\langle \exp(-\omega_\tau) \rangle = 1. \quad (70)$$

This represents a family of predictions—parameterized by the value of α —that are all satisfied for the same process, equation (65). For the choices $\alpha = 1, 0$, equation (70) reduces to equation (64).

Since equation (70) generalizes equation (64) to arbitrary α , it is natural to search for a corresponding extension of equations (43) and (51). Such a result indeed exists, and is given by

$$\frac{\rho^F(+\omega_\tau)}{\rho_\alpha^R(-\omega_\tau)} = \exp(\omega_\tau). \quad (71)$$

Here ρ^F and ρ_α^R denote ensemble distributions of ω_τ for a forward and a reverse process. The forward process is defined as earlier, for the protocol λ_t^F and a family of Fokker–Planck operators $\mathcal{L}_\lambda \leftrightarrow \{\mathbf{J}^S, \varphi, \mathbf{G}\}$. The reverse process uses λ_t^R and

$$\mathcal{L}_\lambda^\alpha \leftrightarrow \{\mathbf{J}_\alpha^S, \varphi, \mathbf{G}\}, \quad (72)$$

where $\mathbf{J}_\alpha^S \equiv (2\alpha - 1)\mathbf{J}^S$. The derivation of equation (71) follows the same steps as in sections 4 and 5, starting from a generalization of equations (31) and (48),

$$\mathcal{S}_+ - \mathcal{S}_-^\alpha = \dot{x}_i (\partial^i \varphi - \alpha A^i) - \frac{1}{2} \alpha (\alpha - 1) G_{ij} A^i A^j, \quad (73)$$

with $\mathcal{S}_-^\alpha$ defined for $\mathcal{L}_\lambda^\alpha$, as $\mathcal{S}_-$ and $\hat{\mathcal{S}}_-$ were defined for $\mathcal{L}_\lambda$ and $\hat{\mathcal{L}}_\lambda$. For $\alpha = 1, 0$ we get $\mathbf{J}_\alpha^S = \pm \mathbf{J}^S$, and equation (71) reduces to the fluctuation theorems derived earlier (equations (43), (51)). Further generalizations are relatively obvious—for instance, replacing equation (72) by $\mathcal{L}_\lambda^K \leftrightarrow \{\mathbf{K}, \varphi, \mathbf{G}\}$, where $\mathbf{K}(\mathbf{x}; \lambda)$ is an arbitrary divergenceless vector field—and will not be pursued here.

9. Discussions and conclusions

There has been considerable recent interest in understanding fluctuation theorems within a Langevin framework [18, 19, 22], [36]–[41]. This interest has been motivated in part by experiments involving systems that are naturally modelled using Langevin dynamics, such as externally manipulated biomolecules [42]–[44], optically trapped colloids [45]–[49], or a torsional pendulum [50, 51]. In the present paper, using the path-integral framework, we have emphasized the separate contributions of two distinct sources of nonequilibrium behaviour: non-conservative forces and non-autonomous dynamics.

We conclude with a specific physical example where the exact relations discussed above apply. Consider a macromolecule, such as a polymer, that is subject to thermal noise and is manipulated externally (for instance using optical tweezers or atomic-force microscopy) on a timescale comparable to its relaxation scale. If such an experiment is carried out with the molecule immersed in a stationary, equilibrium solution, then we have the situation of conservative forces and non-autonomous dynamics. As pointed out by Hummer and Szabo [13] and subsequently verified in [42, 43], such experiments can be used to reconstruct the free energy landscape of the macromolecule.

We can imagine broadening the scope of these investigations to include non-conservative forces, for instance by manipulating a molecule subject to shear [52, 53] and/or chaotic [54, 55] flows. If such flows are accurately modeled as Markovian noise—that is, if the correlations of the flow decay on a timescale that is very short in comparison with the relaxation time of the macromolecule—then the results we have derived in this paper ought to apply directly. When the shear or chaotic flows are non-Markovian, then a suitable modification of our formalism would be needed. A natural starting point for such a modification is the recent progress achieved in the understanding of polymer stretching and particle separation in chaotic flows [56, 57].

The feasibility of any realistic experiment along these lines will be constrained by certain general considerations. While the theoretical predictions are expressed in terms of infinitely many realizations, in a real experiment the number of measurements is necessarily finite. At the same time, if the parameter λ is varied rapidly, then averages such as those in equation (64) might be dominated by realizations that are extremely atypical. The faster the protocol for varying λ , the more realizations we are likely to need. This introduces an important consideration when choosing a protocol for externally manipulating the system. Furthermore, in experiments with complex molecules we inevitably have access to only a small fraction of the system's degrees of freedom, such as its end-to-end extension. Thus we must design an experimental setup in which the quantities that we want to measure (Y , R) can be determined uniquely from the data to which we have access. This might significantly limit the scope of the predictions that we could hope to test.

Acknowledgments

This research was supported by the Department of Energy, under contract W-7405-ENG-36 and WSR program at LANL (MC, CJ), and by startup funds at WSU (VYC). CJ thanks R Dean Astumian and Udo Seifert for stimulating correspondence and discussions on this topic.

Appendix: Discretization scheme

A stochastic differential equation such as equation (2) remains ambiguous until we specify the limiting procedure associated with the discretization of time ($\Delta \rightarrow 0$). In this appendix we describe the discretization scheme that we adopt in this paper, and we sketch the steps then taken to derive equation (4). To avoid a proliferation of vector and tensor indices, we restrict ourselves to the one-dimensional case.

We imagine that the time interval $[-\tau, \tau]$ is cut into $K \gg 1$ equal segments of duration Δ , delimited by the sequence $\{t_0, t_1, \dots, t_K\}$, where $t_0 = -\tau$, $t_K = \tau$, and $t_{k+1} - t_k = \Delta$. The evolution of the dynamical variable $x(t)$ and external parameter $\lambda(t)$ is represented by a discrete sequence: $(x(t), \lambda(t)) \rightarrow (x(t_k), \lambda(t_k)) \equiv (x_k, \lambda_k)$. The conditional probability of the trajectory $X = \{x_0, \dots, x_K\}$ is given by a product of transition rates:

$$\mathcal{P}_\lambda[X|x_0] = \prod_k W_k(x_{k+1}, x_k), \quad (\text{A.1})$$

$$W_k(x_{k+1}, x_k) = \tilde{n}(\bar{x}_k) \exp \left[-\frac{\Delta}{2} \left(\frac{\varepsilon_k}{\Delta} - F(\bar{x}_k) \right)^2 \Gamma(\bar{x}_k) \right] \quad (\text{A.2})$$

where $\bar{x}_k \equiv (x_k + x_{k+1})/2$, $\varepsilon_k \equiv x_{k+1} - x_k$, and primes denote derivatives. The prefactor $\tilde{n}$ is given by the function

$$\tilde{n}(x) = \sqrt{\frac{\Gamma(x)}{2\pi\Delta}} \left(1 - \frac{\Delta}{\Gamma} \left[\frac{\Gamma'(x)^2}{4\Gamma(x)^2} - \frac{\Gamma''(x)}{8\Gamma(x)} \right] \right) \left(1 - \frac{\Delta}{2} F'(x) \right), \quad (\text{A.3})$$

which guarantees normalization to first order in Δ :

$$\int dx_{k+1} W_k(x_{k+1}, x_k) = 1 + \mathcal{O}(\Delta^2). \quad (\text{A.4})$$

In the limit $\Delta \rightarrow 0$, equations (A.1)–(A.3) define the Markov process that we have denoted by equation (2). Our task is now to derive the master equation associated with this Markov process.

Equation (A.2) uses a mid-point regularization scheme not only for the functions $F(x)$ and $\Gamma(x)$ inside the exponent, but also for the prefactor $\tilde{n}(x)$. Let us rewrite this prefactor,

$$\tilde{n}(x) \equiv n(x) \left(1 - \frac{\Delta}{2} F' \right) \approx n(x) e^{-\Delta F'/2}, \quad (\text{A.5})$$

and incorporate the factor $e^{-\Delta F'/2}$ inside the exponent in equation (A.2). Thus, to first order in Δ ,

$$W_k(x_{k+1}, x_k) = n(\bar{x}_k) \exp \left[-\frac{\Delta}{2} \left(\frac{\varepsilon_k}{\Delta} - F(\bar{x}_k) \right)^2 \Gamma(\bar{x}_k) - \frac{\Delta}{2} F'(\bar{x}_k) \right], \quad (\text{A.6})$$

with $n(x)$ as defined by equations (A.3), (A.5). This expression, combined with equation (A.1), specifies the sense in which equation (19) is interpreted, with $\mathcal{N} = \prod_{k=0}^{K-1} n(\bar{x}_k)$. Note that $\mathcal{N}[X] = \mathcal{N}[X^\dagger]$ (because we use the mid-point rule), and $\mathcal{N}$ does not depend on the function $F(x)$. These properties lead to the cancellation of normalization factors in equations (32) and (49), respectively.

To arrive at equation (4), we must evaluate

$$p_{k+1}(x) \equiv \int dx' W_k(x, x') p_k(x') = \int d\varepsilon W_k(x, x - \varepsilon) p_k(x - \varepsilon), \quad (\text{A.7})$$

where p_k , p_{k+1} denotes the probability distribution at time t_k , t_{k+1} . Substituting equation (A.6) into equation (A.7), and changing the variable of integration from ε to

$$z \equiv \sqrt{\frac{\Gamma(x)}{\Delta}} \varepsilon, \quad (\text{A.8})$$

we get

$$p_{k+1}(x) = \sqrt{\frac{1}{2\pi}} \int dz \exp \left(-\frac{z^2}{2} \right) p_k(x) H(x, z, \sqrt{\Delta}). \quad (\text{A.9})$$

We will not give the explicit, cumbersome expression for the function H defined by this procedure, but we note that $\lim_{\Delta \rightarrow 0} H(x, z, \sqrt{\Delta}) = 1$. Thus we have factored out the dominant Gaussian contribution to the integrand in equation (A.7), leaving a slowly

varying function of z . Expanding H in powers of $\sqrt{\Delta}$, to $O(\Delta)$, and evaluating the resulting Gaussian integrals (of the form $\int dz e^{-z^2/2} z^m$), we eventually obtain

$$p_{k+1} = p_k - \Delta \left(F' p_k + F p'_k + \frac{\Gamma'}{2\Gamma^2} p'_k - \frac{1}{2\Gamma} p''_k \right), \quad (\text{A.10})$$

where the p 's, F , and Γ , and their derivatives, are all evaluated at x . In the limit $\Delta \rightarrow 0$, this becomes equation (4).

We emphasize that with a different choice of discretization—for instance, if we had used a prefactor that is a function of x_k rather than the mid-point $\bar{x}_k$ —we would have obtained a different Fokker–Planck equation.

References

- [1] Evans D J, Cohen E G D and Morris G P, 1993 *Phys. Rev. Lett.* **71** 2401
- [2] Evans D and Searles D, 1994 *Phys. Rev. E* **50** 1645
- [3] Gallavotti G and Cohen E G D, 1995 *Phys. Rev. Lett.* **74** 2694
- [4] Kurchan J, 1998 *J. Phys. A: Math. Gen.* **31** 3719
- [5] Lebowitz J L and Spohn H, 1999 *J. Stat. Phys.* **95** 333
- [6] Bochkov G N and Kuzovlev Yu E, 1977 *Zh. Eksp. Teor. Fiz.* **72** 238
Bochkov G N and Kuzovlev Yu E, 1977 *Sov. Phys.—JETP* **45** 125 (translation)
- [7] Jarzynski C, 1997 *Phys. Rev. Lett.* **78** 2690
- [8] Jarzynski C, 1997 *Phys. Rev. E* **56** 5018
- [9] Crooks G E, 1998 *J. Stat. Phys.* **90** 1481
- [10] Crooks G E, 1999 *Phys. Rev. E* **60** 2721
- [11] Hatano T, 1999 *Phys. Rev. E* **60** R5017
- [12] Crooks G E, 2000 *Phys. Rev. E* **61** 2361
- [13] Hummer G and Szabo A, 2001 *Proc. Nat. Acad. Sci.* **98** 3658
- [14] Hatano T and Sasa S, 2001 *Phys. Rev. Lett.* **86** 3463
- [15] Maes C, 2003 *Sém. Poincaré* **2** 29
- [16] Maes C and Netocny K, 2003 *J. Stat. Phys.* **110** 269
- [17] Chernyak V, Chertkov M and Jarzynski C, 2005 *Phys. Rev. E* **71** 025102(R)
- [18] Seifert U, 2005 *Phys. Rev. Lett.* **95** 040602
- [19] Kurchan J, 2005 *Preprint* [cond-mat/0511073](#)
- [20] Reid J C, Sevick E M and Evans D J, 2005 *Europhys. Lett.* **72** 726
- [21] Speck T and Seifert U, 2005 *J. Phys. A: Math. Gen.* **38** L581
- [22] Imparato A and Peliti L, 2006 *Preprint* [cond-mat/0603506](#)
- [23] Maes C and van Wieren M H, 2006 *Phys. Rev. Lett.* **96** 240601
- [24] Evans D J and Searles D, 2002 *Adv. Phys.* **51** 1529
- [25] Ritort F, 2003 *Sém. Poincaré* **2** 193
- [26] Park S and Schulten K, 2004 *J. Chem. Phys.* **120** 5946
- [27] Bustamante C, Liphardt J and Ritort F, 2005 *Phys. Today* **58** 43
- [28] Speck T and Seifert U, 2006 *Europhys. Lett.* **74** 391
- [29] Norris J R, 1997 *Markov Chains* (Cambridge: Cambridge University Press)
- [30] Kemeny J G, Snell J L and Knapp A W, 1976 *Denumerable Markov Chains* 2nd edn (New York: Springer)
- [31] Onsager L and Machlup S, 1953 *Phys. Rev.* **91** 1505
Onsager L and Machlup S, 1953 *Phys. Rev.* **91** 1512
- [32] Wigel F W, 1986 *Introduction to Path-Integral Methods in Physics and Polymer Science* (Philadelphia, PA: World Scientific)
- [33] Bier M, Derényi I, Kostur M and Astumian R D, 1999 *Phys. Rev. E* **59** 6422
- [34] Fingerle A, unpublished work
- [35] Oono Y and Paniconi M, 1998 *Prog. Theor. Phys. Suppl.* **130** 29
- [36] van Zon R and Cohen E G D, 2003 *Phys. Rev. E* **67** 046102
van Zon R and Cohen E G D, 2003 *Phys. Rev. Lett.* **91** 110601
- [37] Reid J C, Carberry D M, Wang G M, Sevick E M, Evans D J and Searles D J, 2004 *Phys. Rev. E* **70** 016111
- [38] Kim K H and Qian H, 2004 *Phys. Rev. Lett.* **93** 120602
- [39] Adjanor G and Athènes M, 2005 *J. Chem. Phys.* **123** 234104

- [40] Astumian R D, 2006 *Am. J. Phys.* **74** 683
- [41] Kim K H and Qian H, 2006 *Preprint* [physics/0601085](#)
- [42] Liphardt J, Dumont S, Smith S B, Tinoco I and Bustamante C, 2002 *Science* **296** 1832
- [43] Collin D, Ritort F, Jarzynski C, Smith S B, Tinoco I and Bustamante C, 2005 *Nature* **437** 231
- [44] Harris N C, Song Y and Kiang C-H, *Direct reconstruction of the free energy surface for mechanical unfolding of proteins*, submitted for publication
- [45] Wang G M, Sevick E M, Mittag E, Searles D J and Evans D J, 2002 *Phys. Rev. Lett.* **89** 050601
- [46] Carberry D M, Reid J C, Wang G M, Sevick E M, Searles D J and Evans D J, 2004 *Phys. Rev. Lett.* **92** 140601
- [47] Trepagnier E H, Jarzynski C, Ritort F, Crooks G E, Bustamante C J and Liphardt J, 2004 *Proc. Nat. Acad. Sci.* **101** 15038
- [48] Wang G M, Reid J C, Carberry D M, Williams D R, Sevick E M, Searles D J and Evans D J, 2005 *Phys. Rev. E* **71** 046141
- [49] Blickle V, Speck T, Helden L, Seifert U and Bechinger C, 2006 *Phys. Rev. Lett.* **96** 070603
- [50] Douarche F, Ciliberto S, Petrosyan A and Rabbiosi I, 2005 *Europhys. Lett.* **70** 593
- [51] Douarche F, Ciliberto S and Petrosyan A, 2005 *J. Stat. Mech.* [P09011](#)
- [52] Smith D E, Babcock H P and Chu S, 1999 *Science* **283** 1724
- [53] Hur J S, Shaqfeh E and Larson R G, 2000 *J. Rheol.* **44** 713
- [54] Groisman A and Steinberg V, 2000 *Nature* **405** 53
- Groisman A and Steinberg V, 2001 *Phys. Rev. Lett.* **86** 934
- [55] Groisman A and Steinberg V, 2004 *New J. Phys.* **6** 29
- [56] Balkovsky E, Fouxon A and Lebedev V, 2000 *Phys. Rev. Lett.* **84** 4765
- Balkovsky E, Fouxon A and Lebedev V, 2001 *Phys. Rev. E* **64** 056301
- [57] Chertkov M, 2000 *Phys. Rev. Lett.* **84** 4761