Nonequilibrium work energy relation for non-Hamiltonian dynamics

Dibyendu Mandal[1] and Michael R. DeWeese[1,2]

[1]*Department of Physics, University of California, Berkeley, California 94720, USA*
[2]*Redwood Center for Theoretical Neuroscience and Helen Wills Neuroscience Institute,
University of California, Berkeley, California 94720, USA*

Recent years have witnessed major advances in our understanding of nonequilibrium processes. The Jarzynski equality, for example, provides a link between equilibrium free energy differences and finite-time nonequilibrium dynamics. We propose a generalization of this relation to non-Hamiltonian dynamics, relevant for active matter systems, continuous feedback, and computer simulation. Surprisingly, this relation allows us to calculate the free energy difference between the desired initial and final equilibrium states using arbitrary dynamics. As a practical matter, this dissociation between the dynamics and the initial and final states promises to facilitate a range of techniques for free energy estimation in a single universal expression.

I. INTRODUCTION

Free energy determination lies at the heart of nearly any application of statistical mechanics [1,2], the conventional methods being based on either the calculation of a partition function or the determination of work in a transition from one equilibrium state to another [3]. In the latter case, the Helmholtz free energy $F_0(A,\beta)$ of the initial equilibrium state with probability density $\rho_A^{\mathrm{eq}} \propto \exp{-\beta\mathcal{H}_A}$ is assumed to be known for a given Hamiltonian $\mathcal{H}_A$ with external parameters A and inverse temperature $\beta = 1/k_{\mathrm{B}}T$, where k_{B} denotes the Boltzmann constant. Then the free energy difference $\Delta F_0 \equiv [F_0(B,\beta) - F_0(A,\beta)]$ corresponding to the transition to another Hamiltonian $\mathcal{H}_B$ is estimated from the work as the external parameters are switched from A to B. To get an exact relation, the switching protocol is often assumed to be either very fast, as in free energy perturbation theory, or adiabatically slow, as in thermodynamic integration theory [3]. A major breakthrough was achieved with the introduction of the Jarzynski equality (JE) [4–6]

$$e^{-\beta\Delta F_0} = \langle e^{-\beta W_0^{\mathrm{in}}}\rangle, \tag{1}$$

whereby the free energy difference ΔF_0 could be calculated from the exponential average of the work W_0^{in} for any switching protocol *of arbitrary speed*. Here the angular brackets denote averaging over many repetitions of the switching protocol. The JE has lead to a plethora of new results in the context of nonequilibrium thermodynamics and statistical mechanics [7–14] and there have been recent advances in the thermodynamics of control [15–19], prediction [20], self-replication [21], and information processing [22–29].

Before the introduction of the JE, Bochkov and Kuzovlev had derived a similar relation (BKR) [30–32]

$$1 = \langle e^{-\beta W_1^{\mathrm{ex}}}\rangle, \tag{2}$$

where, surprisingly, W_1^{ex} is not equal to $W_0^{\mathrm{in}} - \Delta F_0$, as one might expect from the JE [Eq. (1)]. [Note the definitions of W_0^{in} and W_1^{ex} given below in Eqs. (7) and (8), respectively.] The apparent discrepancy between Eqs. (1) and (2) was resolved in [33,34] by showing that these two relations correspond to two different conventions for defining internal energy, each leading to its own definition of work. In fact, by considering the dynamics under two different time-dependent conservative forces f_0 and f_1, Ref. [34] presented a unified expression

$$e^{-\beta\Delta F_0} = \langle e^{-\beta(W_0^{\mathrm{in}}+W_1^{\mathrm{ex}})}\rangle \tag{3}$$

relating the free energy difference ΔF_0 to the different measures of work W_0^{in} and W_1^{ex} for the two forces f_0 and f_1, respectively. The 0 subscript on ΔF_0 is meant to indicate that this free energy difference actually depends only on changes in f_0; ΔF_0 is insensitive to changes in f_1. We note that the detailed form of Eq. (3) differs from the corresponding formula presented in Ref. [34], which assumes that the Hamiltonian is linear in f_1.

We point out here the surprising fact that Eq. (3) remains valid even if the system dynamics during the switching are not related to the two Hamiltonians $\mathcal{H}_A$ and $\mathcal{H}_B$. Unlike the JE, where the dynamics during switching are derived from a time-dependent Hamiltonian $\mathcal{H}(t) : \mathcal{H}_A \to \mathcal{H}_B$, connecting the initial and final Hamiltonians, the combined JE and BKR formula [Eq. (3)] implies that the intermediate Hamiltonian $\mathcal{H}(t)$ can be independent of them. More specifically, as long as the system is initiated in the equilibrium condition ρ_A^{eq}, the dynamics during the switching can be governed by any modified Hamiltonian $\mathcal{H}'(t) = \mathcal{H}(t) + \mathcal{H}_1(t)$, for arbitrary $\mathcal{H}_1(t)$.

Detailed studies have revealed the statistical quality of free energy estimation based on the JE [35–38]. While slow driving protocols produce an effectively unbiased estimator, fast driving protocols induce far-from-equilibrium dynamics that often result in a bias. The convergence of a free energy estimator with respect to the number of independent samples is slow whenever the final phase space distribution of the system at the end of the protocol has a poor overlap with the final equilibrium distribution ρ_B^{eq}. Several strategies have been employed to improve the convergence of the JE estimator for fast driving protocols, including modifying the system dynamics to enhance the overlap between the actual distribution of the system and ρ_B^{eq} and employing bidirectional protocols. However, while some previous studies have exploited specific forms of non-Hamiltonian dynamics in order to improve free energy estimation, it has not been clear what the optimal strategy should be, and a general framework unifying and extending previous results has been lacking.

The main contribution of this paper is to introduce a generalization of the combined JE and BKR [Eq. (3)], given by Eq. (37) below, that is compatible with non-Hamiltonian dynamics during switching. In particular, we extend the former strategy to *arbitrary dynamics*, whereby the averaging in Eq. (1) is performed with the same initial equilibrium condition $\rho_A^{\rm eq}$, but the subsequent evolution is determined by potentially non-Hamiltonian dynamics, completely unrelated to the Hamiltonians $\mathcal{H}_A$ and $\mathcal{H}_B$.

Even though our central result (37) is valid for arbitrary dynamics, for optimal estimation, the dynamics must be tailored such that the actual distribution of the system at the end of the dynamics is the same as the equilibrium distribution $\rho_B^{\rm eq}$. Achieving this condition can make it possible to obtain accurate estimates after as few as one simulated transition. We derive the equation satisfied by such optimal modified dynamics and point out its relation to the so-called escorted dynamics [35]. We emphasize that, whereas the optimal estimation strategy for a broad class of systems involves the use of escorted dynamics, it is typically not the case that one can compute the specific form these dynamics take for interesting nonequilibrium systems, pointing to the need for a more general framework such as ours.

The paper is organized as follows. In Sec. II we illustrate our derivation of the modified JE compatible with arbitrary dynamics [Eq. (37)] for the simple case of the one-dimensional Langevin equation with only position-dependent forces. We first consider the underdamped case and then we describe some important subtleties of the overdamped limit associated with different stochastic integration schemes. In Sec. III we give a general proof of Eq. (37). By construction, our general proof applies to situations involving many interacting Brownian particles. We show that, just like the JE, our result applies even when the dynamics take place in the absence of a thermal reservoir. In Sec. IV we emphasize how there is a clear separation between the dynamics and the end-point Hamiltonians $\mathcal{H}_A$ and $\mathcal{H}_B$ in our relation. In Sec. V we derive an expression for optimal dynamics for free energy estimation based on Eq. (37) and discuss its relation to the so-called escorted dynamics. We conclude by comparing our result to some recent studies on work-fluctuation relations in the absence of detailed balance.

II. ONE-DIMENSIONAL EXAMPLE

Consider the isothermal dynamics of a particle of mass m constrained to move on a circle of circumference l. Let x and p denote its position and momentum, respectively, with the identification $x + l \equiv x$. There are two forces from the heat reservoir, damping $-\gamma p/m$ and noise ξ_t, the latter having the following statistical properties:

$$\langle \xi_t \rangle = 0, \quad \langle \xi_t \xi_{t'} \rangle = \frac{2\gamma}{\beta} \delta(t - t'). \tag{4}$$

In addition, we consider two other forces: $f_0(x; \lambda) = -\partial_x V(x; \lambda_0)$, derived from some potential $V(x; \lambda_0)$ with external parameter λ_0, and $f_1(x; \lambda_1)$, with external parameter λ_1, which is not necessarily derivable from a potential. The

dynamics of the particle are given by

$$\dot{x} = \frac{p}{m},$$

$$\dot{p} = f_0(x; \lambda_0) + f_1(x; \lambda_1) - \gamma \frac{p}{m} + \xi_t, \tag{5}$$

where the dots over the variables x and p denote their time derivatives.

An equivalent way to describe the dynamics of the particle is via the Fokker-Planck equation for the phase-space probability density $\rho(\mathbf{z}, t)$, with $\mathbf{z} = (x, p)$:

$$\frac{\partial \rho}{\partial t} = \hat{\mathcal{L}} \rho = -\nabla_{\mathbf{z}} \cdot \mathbf{J}, \tag{6}$$

where $\mathbf{J} = [(\frac{p}{m}, f_0 + f_1 - \gamma p - \frac{\gamma}{\beta} \partial_p) \rho]$ denotes the phase space probability current. If the parameters $\{\lambda_0, \lambda_1\}$ are held fixed in time, it can be shown that any initial distribution $\rho(\mathbf{z}, 0)$ relaxes to a unique stationary distribution $\rho^{\rm s}(\mathbf{z})$ with $\hat{\mathcal{L}} \rho^{\rm s} = 0$ [39]. Note that, if the force f_1 is zero, the stationary distribution $\rho^{\rm s}$ is the equilibrium distribution $\rho^{\rm eq}(\mathbf{z}; \lambda_0, \beta) \propto \exp -\beta \mathcal{H}_{\lambda_0}$, with respect to the Hamiltonian $\mathcal{H}_{\lambda_0}(\mathbf{z}) = p^2/2m + V(x; \lambda_0)$.

Consider now a switching protocol specified by time-varying parameters $\{\lambda_0(t), \lambda_1(t)\}$ with $\lambda_0(0) = A$, $\lambda_0(\tau) = B$, and arbitrary $\lambda_1(t)$. Following Refs. [33,34], we now introduce two different notions of work, *inclusive* and *exclusive*. Inclusive work is applicable to only conservative forces, while exclusive work is applicable to both conservative and nonconservative forces. Inclusive work done by the conservative force $f_0(x; \lambda)$ for a given protocol $\{\lambda_0(t), \lambda_1(t)\}$ and over a trajectory $\{x(t), p(t)\}$ is

$$W_0^{\rm in} = \int dt\, \dot{\lambda}_0(t) \frac{\partial V(x; \lambda_0)}{\partial \lambda_0} \bigg|_{x(t), \lambda_0(t)}. \tag{7}$$

The exclusive work done by the force $f_1(x; \lambda_1)$ is

$$W_1^{\rm ex} = \int dt\, \frac{p(t)}{m} \circ f_1[x(t); \lambda_1(t)], \tag{8}$$

where the circle on the right-hand side denotes Stratonovich multiplication [40]. According to the Feynman-Kac theorem [35,41], the solution to the sink equation

$$\frac{\partial g}{\partial t} = \hat{\mathcal{L}} g - hg, \tag{9}$$

with the initial condition $g(\mathbf{z}, 0) = \rho^{\rm eq}(\mathbf{z}, \lambda_0, \beta)$ and arbitrary phase space function $h(\mathbf{z}, t)$, is given by the average

$$g(\mathbf{z}, t) = \left\langle \delta(\mathbf{z}(t) - \mathbf{z}) \exp\left(- \int_0^t dt' h(t') \right) \right\rangle \tag{10}$$

with $h(t) = h(\mathbf{z}(t), t)$ and δ denoting the Dirac delta function. While Eq. (10) is true of any h, if we consider the particular form

$$h = \beta \dot{\lambda}_0 \partial_{\lambda_0} V + \beta \frac{p}{m} f_1, \tag{11}$$

its time integral $\int dt\, h$ is equivalent to the sum $W \equiv (W_0^{\rm in} + W_1^{\rm ex})$, in units of $(1/\beta)$, despite the fact that we have not used Stratonovich multiplication in the last term of Eq. (11) (Appendix A). By direct substitution we can show that the

unnormalized, time-dependent Boltzmann distribution

$$g(\mathbf{z},t) = \frac{1}{Z_0(A,\beta)} e^{-\beta \mathcal{H}_{\lambda_0(t)}(\mathbf{z})} \qquad (12)$$

is also a solution of Eq. (9) for the special choice of h given in Eq. (11). Combining Eqs. (10) and (12) and integrating with respect to $\mathbf{z}$ at time $t = \tau$, we get the equation

$$e^{-\beta \Delta F_0} = \langle e^{-\beta(W_0^{\text{in}} + W_1^{\text{ex}})} \rangle, \qquad (13)$$

which is a special case of our more general result (37) with $\Delta F_0 = F_0(B,\beta) - F_0(A,\beta)$. Note that our approach, applying the Feynman-Kac theorem to the original protocol, as opposed to applying a Crooks-like fluctuation theorem to forward and reverse trajectories as an intermediate step [7–9,14], gives a much quicker derivation of Eq. (13) compared to previous approaches that have been brought to bear on this one-dimensional example [42].

A. Relation to fluctuation theorem for entropy production

There is a close connection between Eq. (13) and the integral fluctuation theorem for entropy production in the framework of stochastic thermodynamics [14],

$$\left\langle \frac{\rho_1(\mathbf{z}(\tau))}{\rho_2(\mathbf{z}(0))} e^{\beta Q} \right\rangle = 1, \qquad (14)$$

where ρ_1 and ρ_2 are any two normalized distributions and Q is the heat supplied to the system

$$Q = \int dt \frac{p(t)}{m} \circ \left[-\gamma \frac{p(t)}{m} + \xi_t \right]. \qquad (15)$$

Here Q can also be thought of as the exclusive work done by the reservoir forces [40]. If we consider the forms for ρ_i,

$$\rho_1(\mathbf{z}) = \rho^{\text{eq}}(\mathbf{z}; B, \beta), \quad \rho_2(\mathbf{z}) = \rho^{\text{eq}}(\mathbf{z}; A, \beta), \qquad (16)$$

then the fluctuation theorem (14) reduces to Eq. (13). One just needs to use conservation of energy at the level of each trajectory

$$\Delta \mathcal{H}_{\lambda_0} = W_0^{\text{in}} + W_1^{\text{ex}} + Q, \qquad (17)$$

derivable from the Langevin equation (5).

B. Subtlety in the overdamped limit

Interestingly, our approach leads to a different integral fluctuation theorem than the entropy production fluctuation theorem in the overdamped limit, described by the following dynamics:[1]

$$\dot{x} = \gamma^{-1}(f_0 + f_1) + \gamma^{-1}\xi_t, \qquad (18)$$

often useful in molecular simulations. In this limit, the two definitions of work W_0^{in} and W_1^{ex} remain essentially unchanged; only p/m in Eq. (8) needs to be replaced by $\dot{x}$. From the fluctuation theorem for entropy production (14), valid also in the overdamped limit, one can show the validity of Eq. (13)

---

[1]We have assumed uniform temperature and friction coefficient in the medium so that there is no anomalous contribution to the entropy production.

(Appendix B). However, using the Feynman-Kac approach described above, one can derive the relation (Appendix C)

$$e^{-\beta \Delta F_0^c} = \left\langle \exp\left(-\int dt \, h_{\text{od}} \right) \right\rangle, \qquad (19)$$

$$h_{\text{od}} = \beta \dot{\lambda}_0 \partial_{\lambda_0} V - \frac{\beta}{\gamma} \left(f_0 f_1 + \frac{1}{\beta} \partial_x f_1 \right), \qquad (20)$$

where, unlike the underdamped scenario, the quantity in the exponent on the right-hand side of Eq. (19) is not equal to the sum $\beta W = \beta(W_0^{\text{in}} + W_1^{\text{ex}})$, i.e., Eqs. (13) and (19) are not the same. Note that the free energy $F_0^c(\lambda_0, \beta)$ in this context is the configurational free energy

$$e^{-\beta F_0^c(A,\beta)} = Z_0^c(\lambda_0, \beta) = \int dx \, e^{-\beta V(x;\lambda_0)}. \qquad (21)$$

This duality of integral fluctuation theorems, observable only in the overdamped limit, has been reported before [43], but only in the context of the Bochkov-Kuzovlev relation (2). In contrast, our treatment proves the existence of this duality in the broader context we consider here. Our preliminary investigation suggests that the duality stems from the effects of the transformation $f_1 \to -f_1$ on the probabilities of trajectories. We reserve the detailed investigation for a future study.

We note that, from the perspective of numerical computation of ΔF_0^c, it is advantageous to use Eq. (19) instead of Eq. (13), because the latter involves cancellation of large integrals, which is numerically costly. The requirement of the cancellation is easily seen via the Itô representation of the sum $W = (W_0^{\text{in}} + W_1^{\text{ex}})$:

$$W = \int dt \left[\dot{\lambda}_0 \partial_{\lambda_0} V + \frac{1}{\gamma} \left(f_0 f_1 + \frac{1}{\beta} \partial_x f_1 + f_1^2 + \xi_t f_1 \right) \right]. \qquad (22)$$

The fourth term on the right-hand side is non-negative and its integral grows with time. Convergence of the right-hand side of Eq. (13) can be achieved only if this growing integral is canceled at each power.

III. FLUCTUATION THEOREM VALID FOR GENERAL DYNAMICS

We now derive a fluctuation theorem that is valid even for general non-Hamiltonian dynamics (37). Consider a system described by the phase space coordinates $\mathbf{z} = \{\mathbf{x}, \mathbf{p}\}$ and the Hamiltonian

$$\mathcal{H}_{\lambda_0}(\mathbf{z}) = \frac{p^2}{2m} + V(\mathbf{x}; \lambda_0), \qquad (23)$$

where p is the magnitude of momentum $\mathbf{p}$ and $\lambda_0 = \{\lambda_{01}, \lambda_{02}, \ldots\}$ is a set of external parameters. For a system with many particles, all of mass m, we have $\mathbf{p} = (\mathbf{p}_1, \mathbf{p}_2, \ldots)$ and $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \ldots)$ for particles 1, 2, and so on. If we couple the system to a thermal reservoir of inverse temperature β, quite generally, we can write down the equation to motion to

be of the following form:[2]

$$\dot{\mathbf{x}} = \frac{\mathbf{p}}{m},$$

$$\dot{\mathbf{p}} = -\nabla_{\mathbf{x}}V(\mathbf{x};\lambda_0) - \Gamma\frac{\mathbf{p}}{m} + \Xi_t, \qquad (24)$$

with $\langle \Xi_t \rangle = 0$ and $\langle \Xi_t \Xi_{t'}^{\mathrm{T}} \rangle = 2(\Gamma/\beta)\delta(t-t')$. Here Γ is a positive-definite matrix denoting the damping coefficient matrix and Ξ is the noise vector. The phase space distribution $\rho(\mathbf{z},t)$ evolves according to the Fokker-Planck equation

$$\frac{\partial\rho}{\partial t} = \hat{\mathcal{L}}_0\rho = -\nabla_{\mathbf{z}}\cdot\mathbf{J}, \qquad (25)$$

with $\mathbf{J} = [(\mathbf{p}/m, -\nabla_{\mathbf{x}}V(\mathbf{x};\lambda_0) - \Gamma\mathbf{p} - (\Gamma/\beta)\nabla_{\mathbf{p}})\rho]$. As in the one-dimensional example, the asymptotic solution is the Boltzmann distribution

$$\rho^{\mathrm{eq}}(\mathbf{z};\lambda_0,\beta) = e^{-\beta[\mathcal{H}_{\lambda_0}-F_0(\lambda_0,\beta)]}, \qquad (26)$$

$$e^{-\beta F_0(\lambda_0,\beta)} = Z_0(\lambda_0,\beta) = \int d\mathbf{z}\, e^{-\beta\mathcal{H}_{\lambda_0}}. \qquad (27)$$

We now add an arbitrary phase space velocity vector

$$\mathbf{v}_1 = [\mathbf{f}_{1\mathbf{x}}(\mathbf{z};\lambda_1), \mathbf{f}_{1\mathbf{p}}(\mathbf{z};\lambda_1)] \qquad (28)$$

to the dynamics of the system with external parameters $\lambda_1 = \{\lambda_{11},\lambda_{12},\ldots\}$, leading to the modified dynamics

$$\dot{\mathbf{x}} = \frac{\mathbf{p}}{m} + \mathbf{f}_{1\mathbf{x}},$$

$$\dot{\mathbf{p}} = \mathbf{f}_0 + \mathbf{f}_{1\mathbf{p}} - \Gamma\frac{\mathbf{p}}{m} + \Xi_t, \qquad (29)$$

where we have defined

$$\mathbf{f}_0 = -\nabla_{\mathbf{x}}V(\mathbf{x};\lambda_0). \qquad (30)$$

Such additional phase space velocity vectors arise in many different contexts: (i) for velocity-dependent feedback control, with $\mathbf{v}_1 = (\mathbf{0}, -\Gamma\mathbf{p}/m)$ for some stable matrix Γ [44,45]; (ii) for self-propelled active particles, with $\mathbf{v}_1 = [\mathbf{0},\mathbf{F}(\mathbf{p}/m) + \mathbf{f}(t)]$ for some odd function $\mathbf{F}$ and a generic function $\mathbf{f}$ [46]; and (iii) for escorted simulation dynamics, with $\mathbf{v}_1 = \sum_i \dot{\lambda}_{0i}\mathbf{u}_i(\mathbf{z};\lambda_0)$ for arbitrary, continuous phase space vector fields $\mathbf{u}_i$ [35,47]. Note that $\mathbf{v}_1$ can arise either from real physical forces, as in cases (i) and (ii) above, or from artificial dynamics intended to facilitate computer simulation and sampling of a system, as in case (iii). Note also that $\mathbf{v}_1$ does not generally follow from any Hamiltonian. The addition of $\mathbf{v}_1$ leads to a modified Fokker-Planck operator

$$\hat{\mathcal{L}} = \hat{\mathcal{L}}_0 + \hat{\mathcal{L}}_1, \quad \hat{\mathcal{L}}_1\rho = -\nabla_{\mathbf{z}}\cdot(\mathbf{v}_1\rho), \qquad (31)$$

leading to a modified stationary distribution $\rho^{\mathrm{s}}(\mathbf{z})$, $\hat{\mathcal{L}}\rho^{\mathrm{s}} = 0$. Even when $\mathbf{v}_1$ has a physical origin, i.e., it is of the form $\mathbf{v}_1 = [\mathbf{0},\mathbf{f}_{1\mathbf{p}}(\mathbf{z};\lambda_1)]$ for some physical force $\mathbf{f}_{1\mathbf{p}}(\mathbf{z};\lambda_1)$, the stationary distribution ρ^{s} may be unknown if the force is not derivable from a potential.

Consider now initiating the system at the equilibrium distribution $\rho^{\mathrm{eq}}(\mathbf{z};\lambda_0,\beta)$ and driving the system according to

some protocol $\lambda(t) = \{\lambda_0(t),\lambda_1(t)\}$, as λ_0 varies from $\mathbf{A}$ to $\mathbf{B}$. We wish to calculate the free energy difference $\Delta F_0 = [F_0(\mathbf{B},\beta) - F_0(\mathbf{A},\beta)]$. At any point along any trajectory $\mathbf{z}(t)$, the inclusive power by the original conservative forces is given by

$$\dot{W}_0^{\mathrm{in}} = \dot{\lambda}_0 \cdot \nabla_{\lambda_0}V(\mathbf{z}(t);\lambda_0) \qquad (32)$$

and the exclusive power by the additional force $\mathbf{f}_{1\mathbf{p}}(\mathbf{z}(t);\lambda_1)$ is given by

$$\dot{W}_1^{\mathrm{ex}} = \mathbf{f}_{1\mathbf{p}}(\mathbf{z}(t);\lambda_1) \circ \mathbf{p}(t)/m. \qquad (33)$$

However, unlike the one-dimensional case, the average of the exponential of minus the sum $\beta(W_0^{\mathrm{in}} + W_1^{\mathrm{ex}})$ does not give the free energy change ΔF_0 in this general case. We need to consider additional terms. To see this, let us begin with the unnormalized distribution

$$g(\mathbf{z},t) = \frac{1}{Z_0(\mathbf{A},\beta)}e^{-\beta H_{\lambda_0(t)}(\mathbf{z})}. \qquad (34)$$

By substituting it into the sink equation

$$\frac{\partial g}{\partial t} = \hat{\mathcal{L}}g - hg \qquad (35)$$

and requiring that the right-hand side of Eq. (34) is a solution of Eq. (35), we obtain the following expression for h:

$$h = \beta(\dot{W}_0^{\mathrm{in}} + \dot{W}_1^{\mathrm{ex}}) - \nabla_{\mathbf{z}}\cdot\mathbf{v}_1 + \beta\mathbf{f}_{1\mathbf{x}}\cdot\nabla_{\mathbf{x}}V. \qquad (36)$$

Finally, applying the Feynman-Kac theorem to the sink equation (35) with h as defined in Eq. (36) and combining with Eq. (34), we get

$$e^{-\beta\Delta F_0} = \left\langle \exp\left(-\beta(W_0^{\mathrm{in}} + W_1^{\mathrm{ex}}) + \int dt(\nabla_{\mathbf{z}}\cdot\mathbf{v}_1 - \beta\mathbf{f}_{1\mathbf{x}}\cdot\nabla_{\mathbf{x}}V)\right)\right\rangle, \qquad (37)$$

which is our main result.

Equation (37) generalizes Eq. (13) to the situation where the additional field $\mathbf{v}_1 = (\mathbf{f}_{1\mathbf{x}},\mathbf{f}_{1\mathbf{p}})$ need not involve solely position-dependent forces. In particular, Eq. (37) includes as a special case Brownian dynamics under electromagnetic forces [48]. In this case, surprisingly, the integral in the exponent on the right-hand side of Eq. (37) drops out, leading to Eq. (13), as observed in [48]. In a general scenario, both terms in the integral are nonvanishing. The first term $\int dt\, \nabla_{\mathbf{z}}\cdot\mathbf{v}_1$ accounts for the phase space contraction if the additional velocity term $\mathbf{v}_1$ is dissipative. In the continuous feedback literature, this term has been referred to as entropic pumping [45]. The meaning of the second additional term $-\int dt\, \mathbf{f}_{1\mathbf{x}}\cdot\nabla_{\mathbf{x}}V$ is less transparent, probably because such terms do not appear to have any physical origin, though they can arise in artificial, simulation dynamics [35].

Thermally isolated dynamics

Just like the original JE, a feature of Eq. (37) is that it is valid even when the dynamics during the switching are thermally isolated [4]. In this case, the system is initiated in the same equilibrium distribution $\rho^{\mathrm{eq}}(\mathbf{z};\mathbf{A},\beta)$, but the subsequent evolution does not involve the reservoir terms in Eq. (29), i.e., the system evolves according to the following dynamics:

$$\dot{\mathbf{x}} = \frac{\mathbf{p}}{m} + \mathbf{f}_{1\mathbf{x}}, \quad \dot{\mathbf{p}} = \mathbf{f}_0 + \mathbf{f}_{1\mathbf{p}}. \qquad (38)$$

---

[2]One can use different mass values m_i for different particles without changing the final result.

The proof based on the Feynman-Kac theorem still applies with a modified Fokker-Planck operator [35]. However, the following derivation provides more insight. We evaluate the average $\langle \exp(-\int dt\, h) \rangle$ [with h given by Eq. (36)] over many repetitions of the protocol $\lambda(t) = \{\lambda_0(t), \lambda_1(t)\}$. We can rewrite the integral $(1/\beta)\int dt\, h$ along any phase space trajectory $\mathbf{z}(t)$ as

$$
\begin{aligned}
&\frac{1}{\beta}\int dt\, h \\
&= \int dt[(\boldsymbol{\lambda}_0 \cdot \boldsymbol{\nabla}_{\lambda_0} + \mathbf{f}_{1\mathbf{p}} \cdot \boldsymbol{\nabla}_{\mathbf{p}} + \mathbf{f}_{1\mathbf{x}} \cdot \boldsymbol{\nabla}_{\mathbf{x}})\mathcal{H}_{\lambda_0} - (1/\beta)\boldsymbol{\nabla}_{\mathbf{z}} \cdot \mathbf{v}_1] \\
&= \int dt\left(\frac{d}{dt}\mathcal{H}_{\lambda_0} - (1/\beta)\boldsymbol{\nabla}_{\mathbf{z}} \cdot \dot{\mathbf{z}}\right) \\
&= \mathcal{H}_{\mathbf{B}}(\mathbf{z}(\tau)) - \mathcal{H}_{\mathbf{A}}(\mathbf{z}(0)) - \int dt\,(1/\beta)\boldsymbol{\nabla}_{\mathbf{z}} \cdot \dot{\mathbf{z}}, \qquad (39)
\end{aligned}
$$

where we have used Eqs. (36), (32), and (33) in the first line and Eq. (38) in the second line. Note that, because the evolution of the system takes place in the absence of a thermal reservoir, the evolution is deterministic: If we know the initial phase space coordinate $\mathbf{z}(0)$, we know the future trajectory $\mathbf{z}(t > 0)$ for any given protocol $\{\boldsymbol{\lambda}_0(t), \boldsymbol{\lambda}_1(t)\}$. As a result, the integral $(1/\beta)\int dt\, h$ can be treated as a function of just $\mathbf{z}(0)$. In particular, we can rewrite the average $\langle \exp(-\int dt\, h)\rangle$ as

$$
\left\langle \exp\left(-\int dt\, h\right)\right\rangle = \int d\mathbf{z}(0)\rho^{\mathrm{eq}}(\mathbf{z}(0); \mathbf{A}, \beta)\exp\left(-\int dt\, h\right). \qquad (40)
$$

We can simplify Eq. (40) further:

$$
\begin{aligned}
&\left\langle \exp\left(-\int dt\, h\right)\right\rangle \\
&= \int d\mathbf{z}(0)\frac{\exp\left(-\beta\mathcal{H}_{\mathbf{A}}(\mathbf{z}(0)) - \int dt\, h\right)}{Z_0(\mathbf{A}, \beta)} \qquad (41\mathrm{a}) \\
&= \int d\mathbf{z}(0)\frac{e^{-\beta\mathcal{H}_{\mathbf{B}}(\mathbf{z}(\tau))}}{Z_0(\mathbf{A}, \beta)}\exp\left(\int dt\, \boldsymbol{\nabla}_{\mathbf{z}} \cdot \dot{\mathbf{z}}\right) \qquad (41\mathrm{b}) \\
&= \int d\mathbf{z}(\tau)\frac{e^{-\beta\mathcal{H}_{\mathbf{B}}(\mathbf{z}(\tau))}}{Z_0(\mathbf{A}, \beta)} \qquad (41\mathrm{c}) \\
&= \frac{Z_0(\mathbf{B}; \beta)}{Z_0(\mathbf{A}, \beta)} \qquad (41\mathrm{d}) \\
&= e^{-\beta[F_0(\mathbf{B}, \beta) - F_0(\mathbf{A}, \beta)]}. \qquad (41\mathrm{e})
\end{aligned}
$$

In Eq. (41a), we have rewritten Eq. (40); in Eq. (41b), we have used Eq. (39) for the integral $\int dt\, h$; in Eq. (41c), we have used the fact that the Jacobian $|\partial\mathbf{z}(\tau)/\partial\mathbf{z}(0)|$ is given by $\exp(\int dt\, \boldsymbol{\nabla}_{\mathbf{z}} \cdot \dot{\mathbf{z}})$; and in Eqs. (41d) and (41e) we have used the definitions in Eq. (27). This completes our alternate derivation of Eq. (37) for thermally isolated dynamics.

IV. ARBITRARY DYNAMICS

We want to emphasize that the dynamics during the protocol $\boldsymbol{\lambda}_0(t)$ can be completely independent of $\mathcal{H}_{\lambda_0}$ and we will still

have Eq. (37). Consider an additional field $\mathbf{v}_1$ of the form

$$
\mathbf{f}_{1\mathbf{x}} = -\frac{\mathbf{p}}{m} + \tilde{\mathbf{f}}_{1\mathbf{x}}, \quad \mathbf{f}_{1\mathbf{p}} = \frac{\partial V(\mathbf{x}; \lambda_0)}{\partial\mathbf{x}} + \tilde{\mathbf{f}}_{1\mathbf{p}} \qquad (42)
$$

for any given $\boldsymbol{\lambda}_0(t)$ and arbitrary $\tilde{\mathbf{v}}_1 = (\tilde{\mathbf{f}}_{1\mathbf{x}}, \tilde{\mathbf{f}}_{1\mathbf{p}})$. The system then evolves according to

$$
\dot{\mathbf{x}} = \tilde{\mathbf{f}}_{1\mathbf{x}}, \quad \dot{\mathbf{p}} = \tilde{\mathbf{f}}_{1\mathbf{p}} - \Gamma\frac{\mathbf{p}}{m} + \Xi_t, \qquad (43)
$$

without any term related to the Hamiltonian $\mathcal{H}_{\lambda_0(t)}$, and yet we will still recover the free energy difference ΔF_0 from Eq. (37). (The reservoir terms $\Gamma\frac{\mathbf{p}}{m}$ and Ξ_t will be missing in context of thermally isolated evolution of Sec. III A.) This indicates an interplay between the dynamics and the quantity to average on the right-hand side of Eq. (37), which keeps the left-hand side intact. This level of flexibility in choosing the dynamics seems not to have been appreciated before. Another benefit of the current approach is that we can quickly derive Eq. (37) without going into detailed considerations of path integrals and conjugate processes [14,34,42].

V. OPTIMAL DYNAMICS

The dissociation between the dynamics and the initial and final equilibrium states promises to facilitate a range of techniques for free energy estimation in a single universal expression. Indeed, such an instance has been seen before [35] for a special class of additional phase space velocity vector $\mathbf{v}_1$ referred to as escorted dynamics. In the presence of a single time-dependent parameter $\lambda_0(t)$, the following form of $\mathbf{v}_1$ was chosen:

$$
\mathbf{v}_1 = \dot{\lambda}_0\mathbf{u}(\mathbf{z}; \lambda_0) \qquad (44)
$$

and it was shown that with an appropriate choice of $\mathbf{u}(\mathbf{z}; \lambda_0)$ it is possible to vastly improve the statistical quality of the free energy estimator based on Jarzynski-like relations. The essential idea behind choosing the appropriate dynamics was to ensure that the distribution of the system under the modified dynamics $\rho(\mathbf{z}; t)$ evolves close to the time-dependent equilibrium distribution $\rho^{\mathrm{eq}}(\mathbf{z}; \lambda_0(t))$. In fact, an exact equation was proposed for the optimal choice of $\mathbf{u}(\mathbf{z}; \lambda_0)$ by requiring that the time-dependent distribution is exactly the same as $\rho^{\mathrm{eq}}(\mathbf{z}; \lambda_0(t))$. In this section we consider the case of more than one time-dependent parameter. We see that the optimal dynamics for free energy estimation can be recast in a generalized version of Eq. (44) in an extremely general setting.

In order to obtain the optimal dynamics, such that a single instantiation is sufficient to yield an accurate estimate of the free energy difference, we can impose the condition that $\rho^{\mathrm{eq}}(\mathbf{z}; \lambda_0(t))$ is a solution of the modified Fokker-Planck equation

$$
\frac{\partial\rho}{\partial t} = \hat{\mathcal{L}}\rho \qquad (45)
$$

[see Eqs. (25) and (31)] and then, after some algebra, we get the following equation[3] for optimal $\mathbf{v}_1(\mathbf{z}, t)$, denoted by $\mathbf{v}_1^*$:

$$\nabla_{\mathbf{z}} \cdot \mathbf{v}_1^* - \beta \mathbf{v}_1^* \cdot \nabla_{\mathbf{z}} \mathcal{H}_{\lambda_0} = -\beta \dot{\boldsymbol{\lambda}}_0(t) \cdot \nabla_{\lambda_0} [\mathcal{H}_{\lambda_0} - F_0(\boldsymbol{\lambda}_0; \beta)]. \tag{46}$$

In this case, we can obtain a solution of a form similar to Eq. (44) by considering an additional phase space velocity field $\mathbf{v}_{1i}^*$ for each external parameter λ_{0i}, i.e., by considering

$$\mathbf{v}_1^* = \sum_i \dot{\lambda}_{0i} \mathbf{u}_i^*(\mathbf{z}; \boldsymbol{\lambda}_0), \tag{47}$$

with each additional field $\mathbf{u}_i^*$ now satisfying the following equation [under the assumption that all of the $\dot{\lambda}_{0i}(t)$ are nonzero]:

$$\nabla_{\mathbf{z}} \cdot \mathbf{u}_i^* - \beta \mathbf{u}_i^* \cdot \nabla_{\mathbf{z}} \mathcal{H}_{\lambda_0} = -\beta \partial_{\lambda_{0i}} [\mathcal{H}_{\lambda_0} - F_0(\boldsymbol{\lambda}_0; \beta)]. \tag{48}$$

Equation (48) can be simplified further due to the form of the Hamiltonian given in Eq. (23). Consider the notation $\mathbf{u}_1^* = (\mathbf{u}_{1\mathbf{x}}^*, \mathbf{u}_{1\mathbf{p}}^*)$. We can consistently assume that the momentum components $\mathbf{u}_{i\mathbf{p}}^*$ are zero and get the following equation for $\mathbf{u}_{i\mathbf{x}}^*(\mathbf{x}; \boldsymbol{\lambda})$:

$$\nabla_{\mathbf{x}} \cdot \mathbf{u}_{i\mathbf{x}}^* - \beta \mathbf{u}_{i\mathbf{x}}^* \cdot \nabla_{\mathbf{x}} V(\mathbf{x}; \boldsymbol{\lambda}_0) = -\beta \partial_{\lambda_{0i}} (V - F_0). \tag{49}$$

As may be seen from Eqs. (46), (48), and (49), equations for the optimal dynamics are complicated and they even involve the free energy itself that we are trying to calculate. Clearly, it is extremely unlikely that one could derive the optimal dynamics in all but the simplest cases. Nonetheless, as for the case of escorted dynamics, the current approach provides insight into how to choose $\mathbf{v}_1$ such that free energy estimation is enhanced. The fact that escorted dynamics are already sufficiently powerful to provide the optimal dynamics for estimation might seem to suggest that there would be no practical benefit to developing tools compatible with a broader class of dynamics. However, note that we would already need to know the free energy difference, the very quantity we are trying to estimate, in order to solve Eq. (46) in nearly any physical system of interest, which is why our more general relation (37) promises to be useful for efficient free energy estimation.

VI. DISCUSSION

The current work should be contrasted with the studies described in Refs. [42,46,49–57]. In [46] only velocity-dependent additional forces were considered, whereas in Refs. [42,49–51] only position-dependent forces were considered. In [52,53], only those additional forces were considered for which the steady state distribution was of the Boltzmann form, which is not the case in the current work. In [54,55], the authors started with a generic Langevin equation, not derived from a Hamiltonian, and tried to build a thermodynamic theory for the dynamics. Their approach was based on the decomposition of their abstract dynamics into reversible and

irreversible components. Our approach is complementary to theirs as we start from a given Hamiltonian and then add forces that may be nonconservative. Finally, in [56] the authors propose a speed-up of the calculation of free energy differences by utilizing the fact that violation of detailed balance can be used to accelerate the relaxation to steady states [57]. Given that Eq. (37) is valid even in the absence of detailed balance, it will be interesting to investigate whether our relation will lead to yet faster algorithms for calculating changes in free energy.

Another interesting direction for future research concerns the further generalization of our results to include the determination of free energy profiles along reaction coordinates, as opposed to the free energy difference just between two given equilibrium states. Also, it remains to be seen to what extent the framework of bidirectional protocols developed for the Jarzynski relation and its generalization by Hummer and Szabo [41] may be developed for our relation [58,59]. This could yield even more efficient approaches for calculating changes in free energy from limited data or simulations.

In addition to the benefits of more efficient free energy estimation techniques for physical systems that actually obey Hamiltonian dynamics, the importance of developing a general framework compatible with non-Hamiltonian dynamics is highlighted by a recent example [60] of an active matter, nonequilibrium system that cannot be handled by the JE, the combined JE and BKR, or any other previous generalizations that we are aware of. This particular system consists of a colloidal particle in contact with an active bath containing bacteria that themselves dissipate heat and create microscopic structure in the solution, but other types of active matter systems have been observed [61,62]. We hope and expect that our framework will facilitate the study of active matter systems, as well as systems subject to continuous feedback, which are not correctly described by any previous generalization of the JE or the combined JE and BKR.

APPENDIX A: IRRELEVANCE OF THE STRATONOVICH SCHEME IN UNDERDAMPED DYNAMICS

Here we show the equivalence of the following stochastic integrals in the context of underdamped Langevin dynamics:

$$W_0^{\text{in}} + W_1^{\text{ex}} = \int dt \left(\dot{\lambda}_0 \partial_{\lambda_0} V + f_1 \circ \frac{p}{m} \right), \tag{A1}$$

$$\frac{1}{\beta} \int dt\, h = \int dt \left(\dot{\lambda}_0 \partial_{\lambda_0} V + f_1 \frac{p}{m} \right). \tag{A2}$$

The circle on the right-hand side of Eq. (A1) denotes Stratonovich multiplication [40]. The Itô representation of the

---

[3]As before, we are assuming isothermal dynamics. However, we might also consider time-dependent temperature by considering β as a parameter in the set $\boldsymbol{\lambda}_0$.

integral (A1) can be obtained through the following steps:

$$\int dt \left(\dot{\lambda}_0 \partial_{\lambda_0} V + f_1 \circ \frac{p}{m} \right)$$

$$= \int dt \left[\dot{\lambda}_0 \partial_{\lambda_0} V + \frac{f_1(t) + f_1(t+dt)}{2} \frac{p}{m} \right] \quad \text{(A3a)}$$

$$= \int dt \left[\dot{\lambda}_0 \partial_{\lambda_0} V + f_1(t) \frac{p}{m} + \frac{f_1(t+dt) - f_1(t)}{2} \frac{p}{m} \right] \quad \text{(A3b)}$$

$$\approx \int dt \left(\dot{\lambda}_0 \partial_{\lambda_0} V + f_1(t) \frac{p}{m} \right)$$
$$+ \int dt \frac{[(\partial_x f_1)_t dx + (\partial_{\lambda_1} f_1)_t d\lambda_1]}{2} \frac{p}{m}. \quad \text{(A3c)}$$

In Eq. (A3a) we have used the definition of Stratonovich integration, in Eq. (A3b) we have simply rearranged terms, and in Eq. (A3c) we have used the Taylor expansion of $f_1(t + dt)$ to first order in dt; f_1 is a function of x and λ_1, both of which depend on time. We have neglected the higher-order terms in the Taylor expansion for the reasons given below.

The first integral in Eq. (A3b) is equal to the integral in Eq. (A2). In the following, we argue that the other terms in Eq. (A3c) are negligible. According to the underdamped Langevin equation (5), the differential dx varies as dt. If we assume the given protocol $\lambda_1(t)$ to be continuous and smooth, the differential $d\lambda_1$ also varies as dt. We may conclude that the whole integrand in the last integral of Eq. (A3b) varies as dt and therefore the integral is zero. The higher-order terms in the Taylor expansion of Eq. (A3a) vary as higher powers of dt and their contributions are zero as well. Combining Eqs. (A2), (A3a), and (A3b), we find that the two integrals in Eqs. (A1) and (A2) are equivalent.

APPENDIX B: DERIVATION OF EQ. (13) FOR ONE-DIMENSIONAL OVERDAMPED DYNAMICS

Using the overdamped Langevin equation (18), we can derive the following form of the first law of thermodynamics valid at the level of each realization $x(t)$:

$$\Delta E = Q + W_0^{\text{in}} + W_1^{\text{ex}}, \quad \text{(B1)}$$

where the energy E is equal to the potential energy V; heat Q is given by Eq. (15), with p/m being replaced by $\dot{x}$; and the two types of work W_0^{in} and W_1^{ex} are given by Eqs. (7) and (8), respectively, again with the replacement $p/m \to \dot{x}$. If we use

Eq. (B1) in Eq. (14), with ρ_1 and ρ_2 having the forms

$$\rho_1 \propto e^{-\beta V(x;B)}, \quad \rho_2 \propto e^{-\beta V(x;A)}, \quad \text{(B2)}$$

after some cancellation of terms we arrive at Eq. (13).

APPENDIX C: DERIVATION OF EQ. (19)

The Fokker-Planck equation corresponding to the overdamped Langevin equation (18) is given by

$$\frac{\partial \rho}{\partial t} = \hat{\mathcal{L}}_{\text{od}} = -\partial_x J_{\text{od}}, \quad \text{(C1)}$$

where $\rho(x,t)$ is the probability density at position x at time t, $\hat{\mathcal{L}}_{\text{od}}$ is the overdamped Fokker-Planck operator, and $J_{\text{od}}(x,t) = [\gamma^{-1}(f_0 + f_1) - (\gamma\beta)^{-1}\partial_x]\rho$ is the probability current density at position x at time t. (Other symbols have the same meaning as in the main text.) According to Feynman-Kac theorem, the solution of the sink equation

$$\frac{\partial g}{\partial t} = \hat{\mathcal{L}}_{\text{od}} g - hg \quad \text{(C2)}$$

for an arbitrary function $h(x,t)$ and initial condition $g(x,0) = \exp[-\beta V(x,\lambda_0)]/Z^c(\lambda_0(0),\beta)$ is given by the expression

$$g(x,t) = \left\langle \delta(x(t) - x) \exp\left(-\int_0^\tau dt\, h(t) \right) \right\rangle, \quad \text{(C3)}$$

with $h(t) = h(x(t),t)$. Consider now the following expression of h:

$$h = h_{\text{od}} = \beta \dot{\lambda}_0 \partial_{\lambda_0} V - \frac{\beta}{\gamma} \left(f_0 f_1 + \frac{1}{\beta} \partial_x f_1 \right). \quad \text{(C4)}$$

For this particular choice of h, by direct substitution, we can show that the following expression of g also solves the sink equation (C2):

$$g(x,t) = \frac{e^{-\beta V(x;\lambda_0)}}{Z_0^c(A,\beta)}, \quad Z_0^c(A,\beta) = \int dx. \quad \text{(C5)}$$

Combining Eqs. (C3) and (C5), we get

$$\frac{e^{-\beta V(x;\lambda_0)}}{Z_0^c(A,\beta)} = \left\langle \delta(x(t) - x) \exp\left(-\int_0^\tau dt\, h_{\text{od}}(t) \right) \right\rangle. \quad \text{(C6)}$$

Integrating both sides of this equation with respect to x at time $t = \tau$ we get

$$e^{-\beta \Delta F_0^c} = \frac{Z_0^c(B,\beta)}{Z_0^c(A,\beta)} = \left\langle \exp\left(-\int_0^\tau dt\, h_{\text{od}}(t) \right) \right\rangle, \quad \text{(C7)}$$

with $\Delta F_0^c = F_0^c(B,\beta) - F_0^c(B,\beta)$. Equation (C7) is the same as Eq. (19).

[1] P. M. Chaikin and T. C. Lubensky, *Principles of Condensed Matter Physics* (Cambridge University Press, Cambridge, 1995).

[2] D. Chandler, *Introduction to Modern Statistical Mechanics* (Oxford University Press, New York, 1987).

[3] *Free Energy Calculations—Theory and Applications in Chemistry and Biology*, edited by C. Chipot and A. Pohorille (Springer-Verlag, Berlin, 2007).

[4] C. Jarzynski, Nonequilibrium Equality for Free Energy Differences, Phys. Rev. Lett. **78**, 2690 (1997).

[5] C. Jarzynski, Equilibrium free-energy differences from nonequilibrium measurements: A master-equation approach, Phys. Rev. E **56**, 5018 (1997).

[6] C. Jarzynski, Nonequilibrium work theorem for a system strongly coupled to a thermal environment, J. Stat. Mech. (2004) P09005.

[7] G. E. Crooks, Nonequilibrium measurements of free energy differences for microscopically reversible Markovian systems, J. Stat. Phys. **90**, 1481 (1998).

[8] G. E. Crooks, Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences, Phys. Rev. E **60**, 2721 (1999).

[9] G. E. Crooks, Path-ensemble averages in systems driven far from equilibrium, Phys. Rev. E **61**, 2361 (2000).

[10] T. Hatano and S.-I. Sasa, Steady-State Thermodynamics of Langevin Systems, Phys. Rev. Lett. **86**, 3463 (2001).

[11] T. Speck and U. Seifert, Integral fluctuation theorem for the housekeeping heat, J. Phys. A: Math. Gen. **38**, 581 (2005).

[12] U. Seifert, Entropy Production Along a Stochastic Trajectory and An Integral Fluctuation Theorem, Phys. Rev. Lett. **95**, 040602 (2005).

[13] M. Esposito and C. Van den Broeck, Three Detailed Fluctuation Theorems, Phys. Rev. Lett. **104**, 090601 (2010).

[14] U. Seifert, Stochastic thermodynamics, fluctuation theorems, and molecular machines, Rep. Prog. Phys. **75**, 126001 (2012).

[15] D. A. Sivak and G. E. Crooks, Thermodynamic Metrics and Optimal Paths, Phys. Rev. Lett. **108**, 190602 (2012).

[16] P. R. Zulkowski, D. A. Sivak, G. E. Crooks, and M. R. DeWeese, Geometry of thermodynamic control, Phys. Rev. E **86**, 041148 (2012).

[17] P. R. Zulkowski and M. R. DeWeese, Optimal finite-time erasure of a classical bit, Phys. Rev. E **89**, 052140 (2014).

[18] P. R. Zulkowski and M. R. DeWeese, Optimal protocols for slowly driven quantum systems, Phys. Rev. E **92**, 032113 (2015).

[19] P. R. Zulkowski and M. R. DeWeese, Optimal control of overdamped systems, Phys. Rev. E **92**, 032117 (2015).

[20] S. Still, D. A. Sivak, A. J. Bell, and G. E. Crooks, Thermodynamics of Prediction, Phys. Rev. Lett. **109**, 120604 (2012).

[21] J. L. England, Statistical physics of self-replication, J. Chem. Phys. **139**, 121923 (2013).

[22] T. Sagawa and M. Ueda, Generalized Jarzynski Equality Under Nonequilibrium Feedback Control, Phys. Rev. Lett. **104**, 090602 (2010).

[23] J. M. Horowitz and S. Vaikuntanathan, Nonequilibrium detailed fluctuation theorem for repeated discrete feedback, Phys. Rev. E **82**, 061120 (2010).

[24] T. Sagawa and M. Ueda, Fluctuation Theorem with Information Exchange: Role of Correlations in Stochastic Thermodynamics, Phys. Rev. Lett. **109**, 180602 (2012).

[25] D. Mandal and C. Jarzynski, Work and information processing in a solvable model of Maxwell's demon, Proc. Natl. Acad. Sci. USA **109**, 11641 (2012).

[26] S. Deffner and C. Jarzynski, Information Processing and the Second Law of Thermodynamics: An Inclusive, Hamiltonian Approach, Phys. Rev. X **3**, 041003 (2013).

[27] S. Ito and T. Sagawa, Information Thermodynamics on Causal Networks, Phys. Rev. Lett. **111**, 180603 (2013).

[28] J. M. Horowitz and M. Esposito, Thermodynamics with Continuous Information Flow, Phys. Rev. X **4**, 031015 (2014).

[29] J. M. R. Parrondo, J. M. Horowitz, and T. Sagawa, Thermodynamics of information, Nat. Phys. **11**, 131 (2015).

[30] G. N. Bochkov and Y. E. Kuzovlev, General theory of thermal fluctuations in nonlinear systems, Sov. Phys. JETP **45**, 125 (1977).

[31] G. N. Bochkov and Y. E. Kuzovlev, Nonlinear fluctuation-dissipation relations and stochastic models in nonequilibrium thermodynamics. I. Generalized fluctuation-dissipation theorem, Physica A **106**, 443 (1981).

[32] G. N. Bochkov and Y. E. Kuzovlev, Nonlinear fluctuation-dissipation relations and stochastic models in nonequilibrium thermodynamics. II. Kinetic potential and variational principles for nonlinear irreversible processes, Physica A **106**, 480 (1981).

[33] C. Jarzynski, Comparison of far-from-equilibrium work relations, C. R. Phys. **8**, 495 (2007).

[34] J. Horowitz and C. Jarzynski, Comparison of work fluctuation relations, J. Stat. Mech. (2007) P11002.

[35] S. Vaikuntanathan and C. Jarzynski, Escorted Free Energy Simulations: Improving Convergence by Reducing Dissipation, Phys. Rev. Lett. **100**, 190601 (2008).

[36] S. X. Sun, Equilibrium free energies from path sampling of nonequilibrium trajectories, J. Chem. Phys. **118**, 5769 (2003).

[37] H. Oberhofer, C. Dellago, and P. L. Geissler, Biased sampling of nonequilibrium trajectories: Can fast switching simulations outperform conventional free energy calculation methods? J. Phys. Chem. B **109**, 6902 (2005).

[38] A. Suárez, R. Silbey, and I. Oppenheim, Phase transition in the Jarzynski estimator of free energy difference, Phys. Rev. E **85**, 051108 (2012).

[39] H. Risken, *The Fokker-Planck Equation* (Springer, Berlin, 1984).

[40] K. Sekimoto, *Stochastic Energetics* (Springer, Heidelberg, 2010).

[41] G. Hummer and A. Szabo, Free energy reconstruction from nonequilibrium single-molecule pulling experiments, Proc. Natl. Acad. Sci. USA **98**, 3658 (2001).

[42] R. Spinney and I. Ford, in *Nonequilibrium Statistical Physics of Small Systems: Fluctuation Relations and Beyond*, edited by R. Klages, W. Just, and C. Jarzynski (Wiley-VCH, Weinheim, 2012).

[43] F. Liu, H.-C. Xie, and Z.-H. Lu, Generalized integral fluctuation relation with feedback control for diffusion processes, Commun. Theor. Phys. **62**, 571 (2014).

[44] K. Y. Kim and H. Qian, Entropy Production of Brownian Molecules with Inertia, Phys. Rev. Lett. **93**, 120602 (2004).

[45] K. H. Kim and H. Qian, Fluctuation theorems for a molecular refrigerator, Phys. Rev. E **75**, 022102 (2007).

[46] C. Ganguly and D. Chaudhuri, Stochastic thermodynamics of active Brownian particles, Phys. Rev. E **88**, 032102 (2013).

[47] M. A. Miller and W. P. Reinhardt, Efficient free energy calculations by variationally optimized metric scaling: Concepts and applications to the volume dependence of cluster free energies and solid-solid phase transitions, J. Chem. Phys. **113**, 7035 (2000).

[48] P. Pradhan and U. Seifert, Nonexistence of classical diamagnetism and nonequilibrium fluctuation theorems for charged particles on a curved surface, Europhys. Lett. **89**, 37001 (2010).

[49] V. Chernyak, M. Chertkov, and C. Jarzynski, Path-integral analysis of fluctuation theorems for general Langevin processes, J. Stat. Mech. (2006) P08001.

[50] A. Imparato and L. Peliti, Fluctuation relations for a driven Brownian particle, Phys. Rev. E **74**, 026106 (2006).

[51] J. Kurchan, Non-equilibrium work relations, J. Stat. Mech. (2007) P07005.

[52] P. Pradhan, Nonequilibrium fluctuation theorems in the presence of a time-reversal symmetry-breaking field and nonconservative forces, Phys. Rev. E **81**, 021122 (2010).

[53] M. J. de Oliveira, Irreversible models with Boltzmann-Gibbs probability distribution and entropy production, J. Stat. Mech. (2011) P12012.

[54] L. Yin and P. Ao, Existence and construction of dynamical potential in nonequilibrium processes without detailed balance, J. Phys. A: Math. Gen. **39**, 8593 (2006).

[55] Y. Tang, R. Yuan, and P. Ao, Work relations connecting nonequilibrium steady states without detailed balance, Phys. Rev. E **91**, 042108 (2015).

[56] M. Ohzeki and A. Ichiki, Langevin dynamics neglecting detailed balance condition, Phys. Rev. E **92**, 012105 (2015).

[57] A. Ichiki and M. Ohzeki, Violation of detailed balance accelerates relaxation, Phys. Rev. E **88**, 020101 (2013).

[58] M. R. Shirts, E. Bair, G. Hooker, and V. Pande, Equilibrium Free Energies from Nonequilibrium Measurements Using Maximum-Likelihood Methods, Phys. Rev. Lett. **91**, 140601 (2003).

[59] D. D. L. Minh and A. B. Adib, Optimized Free Energies from Bidirectional Single-Molecule Force Spectroscopy, Phys. Rev. Lett. **100**, 180602 (2008).

[60] A. Argun, A.-R. Moradi, E. Pince, G. B. Bagci, and G. Volpe, Experimental evidence of the failure of Jarzynski equality in active baths, arXiv:1601.01123v1.

[61] S. Ramaswamy, The mechanics and statistics of active matter, Annu. Rev. Condens. Matter Phys. **1**, 323 (2010).

[62] M. C. Marcetti, J. F. Joanny, S. Ramaswamy, T. B. Liverpool, J. Prost, M. Rao, and R. A. Simha, Hydrodynamics of soft active matter, Rev. Mod. Phys. **85**, 1143 (2012).