Journal of Statistical Physics (2025) 192:167
https://doi.org/10.1007/s10955-025-03530-w

Martingale Properties of Entropy Production and a Generalized Work Theorem with Decoupled Forward and Backward Processes

Xiangting Li[1] · Tom Chou[1,2]

Received: 4 April 2025 / Accepted: 6 October 2025 / Published online: 18 November 2025

Abstract
By decoupling forward and backward stochastic trajectories, we construct a family of martingales and work theorems for both overdamped and underdamped Langevin dynamics. Our results are made possible by an alternative derivation of work theorems that uses tools from stochastic calculus instead of path-integration. We further strengthen the equality in work theorems by evaluating expectations conditioned on an arbitrary initial state value. These generalizations extend the applicability of work theorems and offer new interpretations of entropy production in stochastic systems. Lastly, we discuss the violation of work theorems in far-from-equilibrium systems.

1 Background and Introduction

A fundamental relationship in nonequilibrium physics is the Jarzynski equality [1–3] which relates the free energy difference between two states of a system to the work required to force the system from one state to the other. The work done during a nonequilibrium process is described by the time-integral over $\dot{\lambda}_t \partial_\lambda H$, where $H = H(\lambda_t)$ is a λ-dependent Hamiltonian and λ_t is a time-dependent control parameter as depicted in Fig. 1(a). The Jarzynski equality states that starting the system from an equilibrium distribution, the expectation of the exponential of the negative work performed on the system is equal to the exponential of the negative free energy difference between the two states:

$$\mathbb{E}\big[e^{-\beta W}\big] = e^{-\beta \Delta F}. \tag{1}$$

Here, $W_t = \int_0^t \partial_\lambda H(z_s, \lambda_s)\dot{\lambda}_s \mathrm{d}s$ is the work performed, ΔF is the free energy difference, and $\beta = 1/(k_\mathrm{B} T)$ is the inverse temperature.

Jarzynski also extended the equality to stochastic trajectories. His original result was then found to be a consequence of a more general fluctuation theorem proposed by Crooks [4]. In this work, Crooks considers the Markovian dynamics of a system that can be influ-

Communicated by Udo Seifert.

✉ Tom Chou
tomchou@ucla.edu

1 Department of Computational Medicine, University of California, Los Angeles, CA 90095, USA

2 Department of Mathematics, University of California, Los Angeles, CA 90095, USA

enced through a time-dependent control parameter λ_t and that satisfies the microreversibility condition

$$\frac{\mathcal{P}[x_{+t} \mid \lambda_{+t}]}{\mathcal{P}[\bar{x}_{-t} \mid \bar{\lambda}_{-t}]} = \exp\big(-\beta Q[x_{+t}, \lambda_{+t}]\big). \tag{2}$$

Here, $\mathcal{P}[x(+t) \mid \lambda_{+t}]$ is the probability density of the forward trajectory x_{+t}, given the control parameter λ_{+t}, and $\mathcal{P}[\bar{x}_{-t} \mid \bar{\lambda}_{-t}]$ is the probability density of the time-reversed trajectory $\bar{x}_{-t}$, given the time-reversed control parameter $\bar{\lambda}_{-t}$. Starting from Eq. (2), Crooks found

$$\frac{\mathcal{P}_F(W)}{\mathcal{P}_R(-W)} = e^{\beta(W-\Delta F)}, \tag{3}$$

where $\mathcal{P}_F(W)$ and $\mathcal{P}_R(-W)$ are the probabilities of observing work W in the forward and reverse processes, respectively. This result generalizes the Jarzynski equality to a probability density over work. As with the Jarzynski equality, derivation of the Crooks fluctuation theorem seems to require an ensemble of states sampled from equilibrium at the start of the process. However, from Jarzynski's work, it was not entirely clear how the usual concepts of heat, entropy, and free energy in thermodynamics can be defined in a general stochastic system. In particular, for what kind of system can the microreversibility condition in Eq. (2) be satisfied?

By introducing the concept of *stochastic energetics*, Sekimoto [5] developed a framework connecting thermodynamics to overdamped Langevin dynamics with diffusion subject to the fluctuation-dissipation relation. Specifically, the heat can be computed according to the first law of thermodynamics as $Q = \Delta U - W$, where ΔU is the change in internal energy.

Later, Seifert [6] showed that microreversibility holds in overdamped Langevin dynamics obeying the fluctuation-dissipation relation. Most importantly, Seifert observed that Eq. (2) and Eq. (3) do not require the system to start from equilibrium. Instead, they hold as long as the initial distribution ρ_0 is non-singular, *i.e.*, can be written as an integrable function on the phase space, *excluding* the Dirac delta distribution. To achieve this, he explicitly introduced a trajectory-dependent, information-theoretic formulation of entropy production for nonequilibrium thermodynamics, which was also used implicitly in [4] and discussed in [7]. The trajectory-dependent entropy was defined as

$$S(x_t, t) \equiv -k_{\mathrm{B}} \ln \rho(x_t, t), \tag{4}$$

where $\rho(x_t, t)$ is the probability density in state coordinate x evaluated at the value x_t of the stochastic trajectory $\{x_t : t \geq 0\}$ at time t. In Eq. (4), the probability density is implicitly defined relative to the uniform density (i.e., the Lebesgue measure) in the phase space such that the density is dimensionless and the logarithm of probability density is well-defined. Seifert then considered overdamped Langevin dynamics under the influence of a Hamiltonian (potential) $U(x, \lambda)$ and an external force $f(x, \lambda)$, both subject to a time-dependent control parameter λ_t. In the stochastic differential equation (SDE) formulation, the associated dynamics obey

$$\mathrm{d}x_t = \mu\big[-\partial_x U(x, \lambda_t) + f(x_t, \lambda_t)\big]\mathrm{d}t + \sqrt{2D}\,\mathrm{d}B_t, \tag{5}$$

where μ is the mobility, D is the diffusion coefficient, and B_t is a Wiener process that generates Brownian motion. Using the path integral formulation of overdamped Langevin dynamics [8], Seifert validated Eq. (2) and Eq. (3) for initial distributions $\rho_0(x)$ where $\ln(\rho_0(x))$ is well defined over the whole phase space. Eq. (1) was then generalized to the following form:

$$\mathbb{E}\big[e^{-\beta(W_t-\Delta F_t)}\big] = 1. \tag{6}$$

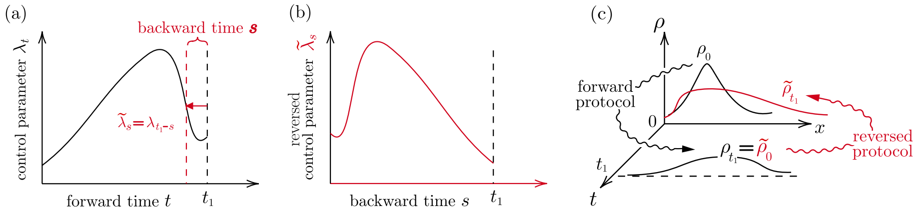

Fig. 1 Schematic of standard backward driving protocols and the associated probability densities. (a) The forward driving protocol λ_t as a function of forward time t. (b) The backward driving protocol $\tilde{\lambda}_s$ is obtained by counting time backwards from the terminal time t_1. (c) Probability densities of the forward (ρ) and the standard backward ($\tilde{\rho}$) processes described in previous work [9]. With an initial distribution $\tilde{\rho}(x_0, 0) = \rho(t_1)$, $\tilde{\rho}(s)$ indexed by the backward time s evolves under the time-reversed driving protocol $\tilde{\lambda}_s$

Here, and in the rest of the paper, $\mathbb{E}[X_t]$ is the expectation of X_t over all trajectories starting from a given initial distribution $\rho_0(x)$ up to time t. The subscript t indicates the process sampled at time t. W_t is the cumulative work up to time t and $\Delta F_t = F_t - F_0$ is the free energy difference between time t and time 0. Note that we have adapted previous notation [6] to be consistent with the context and derivations in the rest of the paper. In the original paper [6], the heat exchange between the system and the environment is understood in terms of entropy change in the environment. Then, considering the total entropy change in the system and the environment, we have $\Delta S_{\text{tot}} = (W - \Delta F)/T$.

Subsequently, Sagawa and Ueda [10] introduced the concept of *feedback control* and *measurement* into nonequilibrium thermodynamics, generalizing the Jarzynski equality to account for systems where a feedback controller influences the dynamics. They further generalized Jarzynski's equality starting from equilibrium to

$$\mathbb{E}\left[e^{-\beta(W_t - \Delta F_t) - I_t}\right] = 1, \tag{7}$$

where I_t is the (trajectorywise version of) mutual information between the actual system state x and the measurement outcome y, defined to be $I_t = I\big(x_t, y_t, t\big) = -\ln\big[(\rho\big(x_t, y_t, t\big)/(\rho(x_t, t)\rho(y_t, t))\big]$. This is the first result that explicitly connects the Jarzynski equality to information theory.

More recently, martingale properties of entropy production in stochastic systems have been explored by Neri [11] and Manzano et al. [9], who extended the Jarzynski equality from a fixed time t to a stopping time $\tau \leq t$. Through the use of path probability densities from a path integral formulation of an overdamped process, they derived the following martingale identity that is associated with expectations of entropy production evaluated over all trajectories of duration τ after the initial time at which the distribution is ρ_0:

$$\mathbb{E}\left[e^{-\beta(W_\tau - \Delta F_\tau) - \delta_\tau}\right] = 1, \tag{8}$$

where the stochastic distinguishability δ_τ is defined as

$$\delta_\tau = \begin{cases} \ln\left[\dfrac{\rho_{\text{eq}}(x_\tau, \lambda_\tau)}{\tilde{\rho}(x_\tau, t - \tau)}\right] & \text{(Neri [11])} \\ \ln\left[\dfrac{\rho(x_\tau, \tau)}{\tilde{\rho}(x_\tau, t - \tau)}\right] & \text{(Manzano [9]).} \end{cases} \tag{9}$$

Here, $\rho_{\text{eq}}(x_\tau, \lambda_\tau)$ refers to the equilibrium probability density of the system at x_τ with parameter λ_τ, $\rho(x_\tau, \tau)$ is the probability density of the system at x, evaluated at x_τ at time τ, and $\tilde{\rho}$ is the time-reversed probability density under a time-reversed driving protocol as shown in Fig. 1(b,c). In the setting of Neri [11], the initial condition is set to equilibrium, rendering his results equivalent to Jarzynski's equality at random times. On the other hand, Manzano et al. works with entropy defined from nonequilibrium distributions, which has been further generalized by Yang and Ge [12] to decoupled auxiliary processes.

Nearly all previous results have been developed using a path integral formulation [8] for overdamped Langevin dynamics or quantum systems [9, 13]. While the path integral formulation is a convenient tool in many areas of physics, gaps in its mathematical rigor may preclude certain desirable directions of analysis. For example, the path integral integrates over the space of continuously differentiable functions but solutions to stochastic differential equations (SDEs) are nowhere differentiable. There have been several efforts to formulate the path integral in a mathematically rigorous way [14–16]; however, these approaches were primarily focused on quantum path integrals and typically assigned a different interpretation of the probability density in the path integral. A mathematically satisfying formulation of the path integral for classical stochastic systems can arise through Girsanov's theorem, and was treated previously in [17]. Even though Girsanov's theorem is a powerful tool in the theory of stochastic calculus, dealing with the path integral with both forward and backward paths is challenging in terms of precise interpretation of the probability density in the forward and backward paths. Moreover, derivations using path integrals rely on microreversibility, precluding treatment of far-from-equilibrium systems where microreversibility does not hold.

On the other hand, solving the corresponding Fokker-Planck equation to find the trajectorywise entropy $S(x_t, t)$ at sufficient numerical precision requires significant computational resources. This is especially challenging for high-dimensional systems or systems with complex potentials. While biological systems can often be described as Maxwell's demons that convert information into work [18, 19], the aforementioned computational demands limit application of the generalized work theorem to biological systems. Thus, experimental verification of the Jarzynski equality has been restricted to relatively simple artificial systems [20–22].

In this paper, we side-step path integration by providing an alternative mathematical proof of the martingale property of entropy production. While our method mirrors that described in a recent treatise on martingale methods for physicists [23], we further show that this proof reveals a generalization to the work theorem, extending it to a family of equations that hold for the same stochastic process but using different choices for the backward process. Our proof also strengthens the equality in Eq. (8) by explicitly evaluating the conditional expectation of the same exponential given the initial value x_0 for *any* initial distribution ρ_0. In particular, the initial condition can be singular, such as the Dirac delta distribution. Thus, it is not necessary to average over all trajectories sampled from the initial distribution $\rho(x_0, 0)$. Moreover, our new "forward-backward-decoupled" work theorem can be developed using underdamped dynamics described by position x and velocity v. Our generalized work theorem can be applied to high-dimensional out-of-equilibrium systems or systems with complex potentials with lower computational costs. Specifically, while the initial condition of the forward process can be arbitrary, we can set the backward process to be the flux associated with a nonequilibrium steady state. In this way, the only computations needed are solving for a stationary distribution and forward simulations. The need for solving a PDE in time is no longer required.

2 Analysis and Results

We formally derive and describe a number of results below.

2.1 Mathematical Approach

We provide the mathematical intuition behind our approach. Consider a stochastic process x that evolves according to a stochastic differential equation (SDE) of the form $\mathrm{d}x_t = b_t \mathrm{d}t + \sigma_t \cdot \mathrm{d}B$. If $b_t \equiv 0$, the process x_t is purely driven by diffusion and has no drift bias in any direction. Once some regularity conditions are satisfied, the mean of the process x_t is the same as that of the initial condition, *i.e.*, $\mathbb{E}[x_t] = \mathbb{E}[x_0]$.

We formalize the intuition developed above by introducing the concept of an exponential martingale. Consider a predictable process θ_t adapted to the filtration of a standard Wiener process B_t. The **exponential martingale** $(M_t)_{t\geq 0}$ associated with $\varphi(x_t)$ is defined by:

$$M_t \equiv \exp\Big(\int_0^t \varphi(x_s) \cdot \mathrm{d}B_s - \frac{1}{2} \int_0^t \|\varphi(x_s)\|^2 \, \mathrm{d}s \Big), \tag{10}$$

where $\int_0^t \varphi(x_s) \cdot \mathrm{d}B_s$ denotes the stochastic integral with respect to the Wiener process that drives x_t and s is the integrated-over dummy time variable. The term $\frac{1}{2} \int_0^t \|\varphi(x_s)\|^2 \, \mathrm{d}s$ is the compensating drift term, ensuring that M_t has zero mean drift.

To guarantee that $\big[M_t\big]_{t\geq 0}$ is indeed a true martingale, a sufficient condition is provided by the well-known **Novikov Condition**, which states that if

$$\mathbb{E}\Big[\exp\Big(\frac{1}{2} \int_0^t \|\varphi(x_s)\|^2 \, \mathrm{d}s \Big) \Big] < \infty \quad \forall t > 0, \tag{11}$$

holds then the process M_t defined in (10) is a true martingale. Throughout this paper, we will use Itô calculus rules for stochastic integrals. Here, the terms "predictable process" and "filtration" are mathematical definitions necessary for the Itô integral to be defined. In the usual context of stochastic thermodynamics, they can be thought of as being automatically satisfied when the processes of interest are functions of some "fundamental" processes (e.g., coordinates of the particle) at current time points. For example, the potential $H(x_t, t)$ of x_t at time t is predictable with respect to the natural filtration of x_t or the underlying Brownian motion B_t. Interested readers can refer to [24] for a comprehensive and pedagogical formulation of stochastic calculus.

2.2 Overdamped Dynamics

Consider overdamped Langevin dynamics defined by

$$\gamma \mathrm{d}x_t = -\nabla H(x_t, t) \, \mathrm{d}t + f(x_t, t) \, \mathrm{d}t + \sqrt{\frac{2\gamma}{\beta}} \, \mathrm{d}B_t. \tag{12}$$

Here, x is the state of the system, $H(x, t)$ is the Hamiltonian, $f(x, t)$ is the external force, and γ is the friction coefficient. The time dependence of $H(x, t)$ and $f(x, t)$ can include that of the control parameter λ_t in the original formulation.

The corresponding Fokker-Planck equation is given by

$$\partial_t \rho(x, t) = \frac{1}{\gamma} \nabla \cdot \big[\big(\nabla H(x, t) - f(x, t) \big) \rho(x, t) \big] + \frac{1}{\beta\gamma} \Delta \rho(x, t), \tag{13}$$

where $\rho(x, t)$ is the probability density of the system at time t, and ∇ and Δ are the gradient and Laplacian operators, respectively, with respect to the state variable x.

Along a given trajectory $x_{s\leq t}$, the work performed on the system up to time t is given by [7]

$$\begin{aligned} W_t &= \int_0^t f(x_s, s)\circ \mathrm{d}x_s + \int_0^t \partial_t H(x_s, s)\,\mathrm{d}s, \\ &= \int_0^t f(x_s, s)\,\mathrm{d}x_s + \frac{1}{\gamma\beta}\int_0^t \nabla\cdot f(x_s, s)\,\mathrm{d}s + \int_0^t \partial_t H(x_s, s)\,\mathrm{d}s, \end{aligned} \tag{14}$$

where $\circ$ denotes the Stratonovich integral. The entropy $S(x_t)$ specific to the trajectory x_t is given by Eq. (4).

Our backward process. While the identity in Eq. (6) suggests that $\exp\big[-\beta(W-\Delta F)\big]$ is a martingale, it is actually not if $F = H - TS$. In the following, we will define a general version of entropy, $\Sigma(x_t, t) = -k_{\mathrm{B}}\ln\psi(x_t, t)$, and show that $\exp[-\beta(W_t - \Delta F_t)]$ is a martingale when Σ replaces S. Here $\psi(x, t)$ solves a backward Fokker-Planck equation, which can be interpreted as a probability distribution for a generalized backward process, as detailed below.

Consider a time-reversed driving protocol described by $\tilde{\lambda}_s$ depicted in Fig. 1(b). In our analysis we will implicitly define forces, Hamiltonians, and densities under the reversed protocol by reversing the time direction of the Fokker-Planck equation by switching the sign of the time-derivative term and do not need to explicitly invoke $\tilde{\lambda}_s$. Specifically, the Hamiltonian and external force are replaced by their time-reversed forms, $\widetilde{H}(x, s) := H(x, t_1 - s)$ and $\widetilde{f}(x, t) := f(x, t_1 - s)$ [9]. Here, $s = t_1 - t$ represents how far back in time the current time is compared to the terminal time t_1. A new "time-reversed" probability density $\tilde{\rho}(x, s)$ with initial condition $\tilde{\rho}(x, 0) = \rho(x, t_1)$, depicted in Fig. 1(c), evolves according to the "backward" Fokker-Planck equation

$$\partial_t\tilde{\rho} = \frac{1}{\gamma}\nabla\cdot\left[\big(\nabla\widetilde{H}(x, t) - \widetilde{f}(x, t)\big)\tilde{\rho}\right] + \frac{1}{\beta\gamma}\Delta\tilde{\rho}. \tag{15}$$

When the potential $H(x, t)$ and the force $f(x, t)$ are time-asymmetric, i.e., $H(x, t) \neq \widetilde{H}(x, t)$ and $f(x, t) \neq \widetilde{f}(x, t)$, the time-reversed process for $\tilde{\rho}(x, t)$ is different from the original process for $\rho(x, t)$ in the sense that $\tilde{\rho}(x, t) \neq \rho(x, t_1 - t)$ for $t \in (0, t_1)$. Note that the Fokker-Planck equation is also known as the Kolmogorov forward equation, but here, the backward Fokker-Planck equation differs from the Kolmogorov backward equation which is simply the adjoint of the Kolmogorov forward equation [25].

We define our backward process by its probability density $\psi(x, t)$ which obeys a specific type of backward Fokker-Planck equation

$$-\partial_t\psi(x, t) = \frac{1}{\gamma}\nabla\cdot\Big[\big(\nabla H(x, t) - f(x, t)\big)\psi(x, t)\Big] + \frac{1}{\beta\gamma}\Delta\psi(x, t). \tag{16}$$

Eq. (16) differs from the Fokker-Planck equation (13) by an extra minus sign in front of the time derivative. It is straightforward to verify that $\psi(x, t) = \tilde{\rho}(x, t_1 - t)$ and that $\psi(x, t)$ is associated with the time-reversed probability $\tilde{\rho}$ used in [9].

In the setting of Crooks' fluctuation theorem [4], a specific terminal time t_1 was *a priori* chosen, and $\psi(x, t) := \tilde{\rho}(x, t_1 - t)$ solves Eq. (16) with initial condition $\psi(x, t_1) = \rho(x, t_1)$. The specific time-reversed probability density $\tilde{\rho}(x, s)$ is used in Eq. (9) to define the stochastic distinguishability δ. $\psi(x, t_1 - s) = \tilde{\rho}(x, s)$ maps the time-reversed probability density at the backward time s to the time-reversed probability density at the forward time $t = t_1 - s$, as illustrated in Figs. 1(c).

To derive the generalized work theorem, $\psi(x, t)$ need only satisfy Eq. (16) *regardless of* initial condition. Consequently, choosing different initial conditions results in different martingales and corresponding identities. Our generalization of the work theorem is based

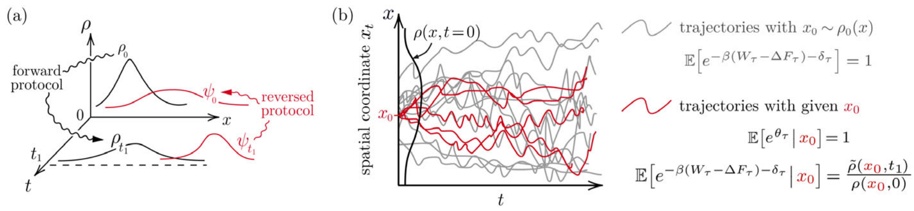

Fig. 2 Decoupling of the forward and backward processes. (a) In our derivations, we employ the probability density $\psi(x, t)$ of a backward process indexed by forward time t. The initial condition is arbitrary so that in general $\psi(x, t_1) \neq \rho(x, t_1)$. (b) Trajectories of the forward processes x_t sampled from the initial distribution $\rho(x_0, 0)$ are shown in grey, while trajectories with a specific initial value x_0 are shown in red. The original work theorem uses averages over the grey trajectories, while our generalized work theorem considers averages over the red trajectories. See Eqs. (24) and (27)

on the observation that the backward process $\psi(x, t)$ can be defined with an arbitrary initial condition, not necessarily $\rho(x, t_1)$, as shown in Fig. 2(a). Broadly speaking, our approach does not require specifying a terminal time t_1.

Analogous to the entropy $S(x_t)$ defined in Eq. (4), we define the "entropy" of the backward process along the trajectory x_t as

$$\Sigma(x_t, t) = -k_{\mathrm{B}} \ln \psi(x_t, t). \tag{17}$$

As with the trajectorywise entropy, $\Sigma(x_t, t)$ implicitly requires a reference probability density $\psi(x_0, 0)$. Throughout this work, we use A_t to denote the trajectory-specific value of $A(x, t)$ evaluated at $x = x_t$, e.g., $\Sigma_t \equiv \Sigma(x_t, t)$.

Martingale and generalized work theorem. We now consider the quantity defined by

$$\theta_t = -\beta W_t + \beta\big(\Delta H(x_t, t) - T \Delta\Sigma(x_t, t)\big). \tag{18}$$

This process represents the nondimensionalized "entropy production" along the trajectory x_t when the entropy S is replaced by that of the backward process Σ. Upon using Itô's formula to expand terms, the time derivative of θ_t is given by

$$\begin{aligned} \mathrm{d}\theta_t &= -\beta f \cdot \mathrm{d}x_t - \frac{1}{\gamma}\nabla \cdot f \,\mathrm{d}t - \beta \partial_t H \,\mathrm{d}t + \beta \nabla H \cdot \mathrm{d}x_t + \frac{\beta}{2}\Delta H \left[\mathrm{d}x_t\right]^2 + \beta \partial_t H \,\mathrm{d}t \\ &\quad + \frac{\mathrm{d}\psi}{\psi} - \frac{[\mathrm{d}\psi]^2}{2\psi^2} \\ &= \beta(\nabla H - f) \cdot \mathrm{d}x_t + \frac{1}{\gamma}(\Delta H - \nabla \cdot f)\,\mathrm{d}t + \frac{\mathrm{d}\psi}{\psi} - \frac{[\mathrm{d}\psi]^2}{2\psi^2}. \end{aligned} \tag{19}$$

Our goal is to express $\mathrm{d}\theta_t$ in terms of $\mathrm{d}B_t$ and $\mathrm{d}t$. To achieve this, we need to evaluate $\mathrm{d}\psi$ as $\mathrm{d}\psi(x_t, t)$. Using the chain rule and Eq. (16), we find

$$\begin{aligned} \mathrm{d}\psi &= \partial_t \psi \,\mathrm{d}t + \nabla\psi \cdot \mathrm{d}x_t + \frac{1}{2}\Delta\psi \,[\mathrm{d}x_t]^2 \\ &= -\frac{1}{\gamma}\nabla \cdot \big[(\nabla H - f)\psi\big]\mathrm{d}t - \frac{1}{\beta\gamma}\Delta\psi \,\mathrm{d}t + \nabla\psi \cdot \mathrm{d}x_t + \frac{1}{2}\Delta\psi[\mathrm{d}x_t]^2 \\ &= -\frac{\psi}{\gamma}(\Delta H - \nabla \cdot f)\,\mathrm{d}t - \frac{1}{\gamma}\big[(\nabla H - f) \cdot \nabla\psi\big]\mathrm{d}t + \nabla\psi \cdot \mathrm{d}x_t \end{aligned} \tag{20}$$

Consequently, $\frac{\mathrm{d}\psi}{\psi}$ and $\frac{1}{2\psi^2}[\mathrm{d}\psi]^2$ are given by

$$\begin{aligned}\frac{1}{\psi}\,\mathrm{d}\psi &= -\frac{1}{\gamma}(\Delta H - \nabla\cdot f)\,\mathrm{d}t - \frac{1}{\gamma}\big[(\nabla H - f)\cdot\nabla\ln\psi\big]\mathrm{d}t + \nabla\ln\psi\cdot\mathrm{d}x_t\\ \frac{1}{2\psi^2}[\mathrm{d}\psi]^2 &= \frac{1}{\gamma\beta}\frac{\|\nabla\psi\|^2}{\psi^2}\mathrm{d}t.\end{aligned} \tag{21}$$

Substituting Eqs. (21) into Eq. (19), we find

$$\begin{aligned}\mathrm{d}\theta_t &= -\frac{1}{\gamma}(\nabla\ln\psi)\cdot(\nabla H - f)\,\mathrm{d}t - \frac{1}{\gamma\beta}\big\|\nabla\ln\psi\big\|^2\mathrm{d}t + \big(\beta\nabla H - \beta f + \nabla\ln\psi\big)\cdot\mathrm{d}x_t\\ &= -\frac{1}{\gamma\beta}\big\|\beta(\nabla H - f) + \nabla\ln\psi\big\|^2\mathrm{d}t + \sqrt{\frac{2}{\gamma\beta}}\Big[\beta(\nabla H - f) + \nabla\ln\psi\Big]\cdot\mathrm{d}B_t.\end{aligned} \tag{22}$$

The exponential of θ_t can now be expressed as

$$\begin{aligned}\mathrm{d}e^{\theta_t} &= e^{\theta_t}\Big[\mathrm{d}\theta_t + \frac{1}{2}(\mathrm{d}\theta_t)^2\Big]\\ &= e^{\theta_t}\sqrt{\frac{2}{\gamma\beta}}\Big[\beta(\nabla H - f) + \nabla\ln\psi\Big]\cdot\mathrm{d}B_t.\end{aligned} \tag{23}$$

Thus, as long as the Novikov condition holds, the process e^{θ_t} is a martingale with initial value one.

By the optional stopping theorem for martingales, we have

$$\mathbb{E}\left[e^{-\beta(W_\tau - \Delta H_\tau + T\Delta\Sigma_\tau)}\,\Big|\,x_0\right] = \mathbb{E}\left[e^{\theta_\tau}\mid x_0\right] = e^{\theta_0} = 1,\ \text{a.s.},\ \forall\text{ bounded stopping time }\tau. \tag{24}$$

Eq. (24) is the main result of this paper; note that Eq. (24) is stronger than $\mathbb{E}\left[e^{\theta_\tau}\right] = 1$ since Eq. (24) is an average over trajectories starting from an arbitrary initial value x_0, independent of the initial distribution $\rho_0(x)$, while the latter is an average over all trajectories starting from the initial distribution $\rho_0(x)$, shown by the red and grey trajectories in Fig. 2(b), respectively. Pigolotti et al. [26] first studied a special case, the nonequilibrium stationary system, of Eq. (23) and then Neri et al. proposed the stationary version of Eq. (24) in [27].

Yang and Ge [12] generalized the fluctuation relation using the path integral method by introducing another stochastic process y_t mutually absolutely continuous with the process of interest x_t and a third process z_t driven by the time-reversed protocol of y_t with an arbitrary initial condition. The logarithm of the ratio of the forward path probability under x to the backward path probability under z plays a similar role to the free energy change ΔF. When this functional is compensated by a proper log probability ratio between distributions of z_0 and z'_{t_1-t} (where t_1 is a chosen fixed time point) and exponentiated, a martingale is formally constructed. Here, z'_t is another process derived in the same way as z_t but may have a different initial condition.

In terms of overdamped dynamics, our result coincides with that in [12], if we restrict our generalized backward process to the same fixed finite interval $[0, t_1]$ and restrict z_t and z'_t to be identically distributed as our generalized backward process ψ_{t_1-t}. However, there are two important differences. First, while [12] enjoys more degrees of freedom in choosing the auxiliary processes z_t, their construction still requires a fixed time interval to be chosen in advance, as this interval is intrinsic to the definition of their functional and compensation. By contrast, our construction does not rely on this fixed time interval and can be extended to the whole time axis. Second, the path integral or path probability method employed in prior work formally only requires that x_t and y_t to be mutually absolutely continuous. This is satisfied automatically if y_t is chosen to be x_t or effectively z_t is chosen to be the generalized backward process, as in our case. Our Itô calculus approach requires an additional regularity condition to be satisfied in order for the local martingale to be a martingale. In other words, predictions made by the path integral methods may fail for some unforeseeable singular cases.

Manzano's result. We now show that Eq. (24) is a generalization of Eq. (8). The key observation is that in the definition of the backward process $\psi(x,t)$, we have the freedom to choose the initial condition. For different choices of the initial condition, we will arrive at different forms of the martingale. Specifically, in the setting of [9], the backward process is chosen by fixing a final time t_1, and choosing the "initial condition" of the backward process to be

$$\psi(x,t_1) = \rho(x,t_1), \ \forall x. \tag{25}$$

Then, $\psi(x,t)$ evolves backward in time according to Eq. (16).

Consider a stopping time τ such that $0 \le \tau \le t_1$ almost surely. Note that $-\beta T \Delta\Sigma(\tau) = \ln\psi(x_\tau,\tau) - \ln\psi(x_0,0) = \ln\tilde{\rho}(x_\tau,t_1-\tau) - \ln\tilde{\rho}(x_0,t_1)$, while $-\beta T\Delta S = \ln\rho(x_\tau,\tau) - \ln\rho(x_0,0)$. Recalling the definition of δ_τ in Eq. (9), we find

$$\mathbb{E}\Big[e^{-\beta\left[W_\tau-\Delta F_\tau\right]-\delta_\tau}\Big] = \mathbb{E}\Big[e^{\theta_\tau}\frac{\tilde{\rho}(x_0,t_1)}{\rho(x_0,0)}\Big] = \int \tilde{\rho}(x_0,t_1)\mathbb{E}\left[e^{\theta_t} \mid x_0\right]\mathrm{d}x_0 = 1, \tag{26}$$

which is Eq. (8). Here, the second equality follows from conditioning on x_0 and the last equality follows from Eq. (24). Additionally, by conditioning on x_0, we have

$$\mathbb{E}\Big[e^{-\beta\left[W_\tau-\Delta F_\tau\right]-\delta_\tau}\Big| x_0\Big] = \mathbb{E}\Big[e^{\theta_t}\frac{\tilde{\rho}(x_0,t_1)}{\rho(x_0,0)}\Big| x_0\Big] = \frac{\tilde{\rho}(x_0,t_1)}{\rho(x_0,0)}, \quad \forall x_0. \tag{27}$$

Eq. (27) is stronger than Eq. (8) since Eq. (8) only holds for all trajectories sampled according to the initial distribution $\rho_0(x)$, while our result holds for trajectories starting from an arbitrary initial value x_0, independent of the initial distribution $\rho_0(x)$, as shown in Fig. 2(b).

This result is particularly helpful when one wants to consider the work theorem for a system where the initial distribution cannot be written as a density function over the state space, such as the Dirac delta distribution. In such cases, the trajectorywise entropy at the initial time $S(0) = -k_\mathrm{B}\ln\rho(x_0,0)$ is not well-defined. Our result provides a way to bypass this issue by decoupling the initial sample x_0 from its initial distribution ρ_0. Energy changes and external work can be evaluated by the conditional work theorem Eq. (27) for trajectories starting at x_0 with an arbitrarily chosen well-behaved initial distribution ρ_0.

Stationary Hamiltonian. There are other interesting choices for the backward process. For example, in the case of time-independent potentials and forces, we can choose $\psi(x,0) = \rho_\mathrm{ss}(x)$, the stationary distribution of Eq. (13). In this case, $\psi(x,t) \equiv \psi(x,0) = \rho_\mathrm{ss}(x)$ and

$$\mathbb{E}\Big[e^{-\beta\left(W_\tau-\Delta H_\tau+T\Delta S_\mathrm{ss}(\tau)\right)}\Big] = 1, \ \forall \text{ bounded stopping time } \tau, \tag{28}$$

where $S_\mathrm{ss}(t) := -k_\mathrm{B}\ln\rho_\mathrm{ss}(x_t)$ and the initial distribution $\rho_0(x)$ of x_0 can be different from $\rho_\mathrm{ss}(x)$.

The stationary case is of interest when f is a dissipative force. Such a system can be used to model nonequilibrium stochastic chemical reactions commonly found in biological systems, including kinetic proofreading [28, 29] and chemotaxis [30]. While evaluating Eq. (8) requires solution to the $d+1$-dimensional time-dependent PDE for $\rho(x,t)$, computing Eq. (28) requires only the solution to the d-dimensional time-independent PDE for $\rho_\mathrm{ss}(x)$, leading to an easier computational evaluation. A similar identity was derived in [23] assuming that the initial distribution $\rho_0(x)$ is the stationary (backward) distribution $\rho_\mathrm{ss}(x)$. Our result relaxes this assumption and shows that the identity holds for any initial distribution $\rho_0(x)$.

Fluctuation-dissipation relation. In our formulation of the Langevin dynamics in Eq. (12), we related the diffusion coefficient $D = 1/(\beta\gamma)$ to the friction coefficient γ as required by the fluctuation-dissipation theorem. Specifically, a general overdamped process is governed by

$$\mathrm{d}x_t = \frac{1}{\gamma}\big[-\nabla H(x_t,t) + f(x_t,t)\big]\,\mathrm{d}t + \sqrt{2D}\,\mathrm{d}B_t. \tag{29}$$

Eq. (12) is obtained from Eq. (29) by setting $D = 1/(\beta\gamma)$.

Starting from Eq. (29), using the backward process $\psi(x,t)$ associated with the time-reversed Fokker-Planck equation (29)

$$-\partial_t\psi(x,t) = \frac{1}{\gamma}\nabla\cdot\Big[\big(\nabla H(x,t) - f(x,t)\big)\psi(x,t)\Big] + D\Delta\psi(x,t), \tag{30}$$

and defining Σ_t, θ_t through Eqs. (17) and (18), we find (see Appendix A.1 for a detailed derivation)

$$\begin{aligned}\mathrm{d}e^{\theta_t} = e^{\theta_t}\bigg[&\Big(D - \frac{1}{\beta\gamma}\Big)\Big(\beta^2\|\nabla H - f\|^2 + 2\beta\big(\nabla H - f\big)\cdot\nabla\ln\psi + \beta\big(\Delta H - \nabla\cdot f\big)\Big)\mathrm{d}t\\ &+ \sqrt{2D}\Big(\beta\big(\nabla H - f\big) + \nabla\ln\psi\Big)\cdot\mathrm{d}B_t\bigg].\end{aligned} \tag{31}$$

When the assumption $D = 1/(\beta\gamma)$ is violated, $\mathrm{d}e^{\theta_t}$ carries a non-zero drift and is no longer a martingale, unless $f = \nabla H$ cancelling out the energy gradient everywhere.

2.3 Generalization to underdamped Langevin dynamics

Our method of introducing a backward process is general and can be applied to *underdamped* Langevin dynamics that obey

$$\begin{aligned}\mathrm{d}x_t &= v_t\,\mathrm{d}t\\ m\mathrm{d}v_t &= \big[-\gamma v_t + (-\nabla_x U + f)\big]\mathrm{d}t + \sqrt{\frac{2\gamma}{\beta}}\,\mathrm{d}B_t,\end{aligned} \tag{32}$$

where m is an effective mass and γ represents a velocity decay rate, or the decay rate of the velocity-velocity correlation function under fluctuation-dissipation conditions.

The main differences in the analysis between the underdamped and overdamped cases are in the definition of the Hamiltonian

$$H(x,v,t) = \frac{1}{2}m\|v\|^2 + U(x,t), \tag{33}$$

and the evaluation of the work performed

$$W_t = \int_0^t f(x_s,s)\,\mathrm{d}x_s + \int_0^t \partial_t U(x_s,s)\,\mathrm{d}s. \tag{34}$$

Here, because x_s is now a continuously differentiable function of s, the Stratonovich integral with respect to x_s is equivalent to the normal Lebesgue-Stieltjes integral. The evolution of the probability density $\rho(x,v,t)$ is then given by

$$\partial_t\rho(x,v,t) = -v\cdot\nabla_x\rho - \frac{1}{m}\nabla_v\cdot\Big[\big(-\gamma v - \nabla_x U + f\big)\rho\Big] + \frac{\gamma}{\beta m^2}\Delta_v\rho. \tag{35}$$

where we use ∇_x and ∇_v to denote the gradient with respect to x and v, respectively, and Δ_v to represent the Laplacian with respect to v. Correspondingly, we define a backward process $\psi(x,v,t)$ by

$$-\partial_t\psi(x,v,t) = v\cdot\nabla_x\psi - \frac{1}{m}\nabla_v\cdot\Big[\big(-\gamma v + \nabla_x U - f\big)\psi\Big] + \frac{\gamma}{\beta m^2}\Delta_v\psi. \tag{36}$$

Eq. (36) is obtained by applying the transformations $v \to -v$, $\nabla_v \to -\nabla_v$ and $\partial_t \to -\partial_t$ to the forward Kramers equation (35). $\psi(x,v,t)$ represents a probability density in the sense that $\tilde{\rho}(x,v,t) := \psi(-x,v,-t)$ is the probability density of the Langevin dynamics under an inverted and time-reversed potential and force $\widetilde{U}(x,t) = U(-x,-t)$ and $\tilde{f}(x,t) = -f(-x,-t)$. Specifically, we do not impose constraints on the initial condition of $\psi(x,v,0)$.

The backward process $\psi(x,v,t)$ or $\tilde{\rho}(-x,v,-t)$ is a generalization of the previous "time-reversed" processes discussed in [9] for underdamped Langevin dynamics. To make the connection clearer, we can choose a specific final time t_1 and shift the time argument to $t_1 - t$ in the backward process, *i.e.*, $\tilde{\rho}(x,v,t) = \psi(-x,v,t_1-t)$. If we let $\tilde{\rho}(x,v,0) = \psi(-x,v,t_1)$, $\tilde{\rho}$ evolves forward in time under the space- and time-reversed potential $\widetilde{U}$ and force $\tilde{f}$ associated with the time-reversed driving protocol, analogous to that shown in Fig. 1(b).

Ignoring their dependence on v_t, the evolution of distributions for forward and backward processes is analogous to that of the overdamped case depicted in Fig. 2(a). Note that we also impose space inversion which is different from the standard backward process in which only the time and velocity are reversed. In the latter case of velocity inversion, the term $\gamma(\nabla_v\cdot v)/m$ carries the opposite sign, rendering the exponential of the "entropy" production (defined below) no longer a martingale. Moreover, in our generalization of the work theorem, we do not impose constraints on the initial or final condition of $\psi(x,v,t)$. An arbitrary choice of the initial condition enables decoupling of the forward and backward processes.

Extending the quantities in Eqs. (17) and (18), we define the entropy production along trajectories (x_t, v_t) as $\theta_t = -\beta W_t + \beta H(x_t,v_t,t) + \ln\psi(x_t,v_t,t)$. Chain-rule-expanding all terms, we find

$$\begin{aligned} \mathrm{d}\theta_t =& \beta v\cdot(\nabla_x H - f)\mathrm{d}t + \beta\nabla_v H\cdot\mathrm{d}v_t + \frac{\gamma}{m^2}\Delta_v H\mathrm{d}t + v\cdot\nabla_x\ln\psi\,\mathrm{d}t \\ &+ \nabla_v\ln\psi\cdot\mathrm{d}v_t + \frac{\gamma}{\beta m^2}\frac{\Delta_v\psi}{\psi}\mathrm{d}t + \frac{\partial_t\psi}{\psi}\mathrm{d}t - \frac{\gamma}{\beta m^2}\big\|\nabla_v\ln\psi\big\|^2\mathrm{d}t. \end{aligned} \tag{37}$$

If U and f are independent of v, substituting $\partial_t\psi$ from Eq. (36) into Eq. (37), we find

$$\begin{aligned} \mathrm{d}\theta_t =& \beta v\cdot(\nabla_x U - f)\mathrm{d}t + \beta\nabla_v H\cdot\mathrm{d}v_t + \frac{n\gamma}{m}\mathrm{d}t \\ &+ \nabla_v\ln\psi\cdot\mathrm{d}v_t - \frac{n\gamma}{m}\mathrm{d}t + \frac{1}{m}(\nabla_x U - f)\cdot\nabla_v\ln\psi\,\mathrm{d}t - \frac{\gamma}{\beta m^2}\big\|\nabla_v\ln\psi\big\|^2\mathrm{d}t \end{aligned} \tag{38}$$

Here, n is the physical dimension of v_t, the first $n\gamma/m$ term results from $\frac{\gamma\Delta_v H}{m^2} = \frac{\gamma\Delta_v\|v\|^2}{2m}$, and the second $n\gamma/m$ term derives from the $\gamma(\nabla_v\cdot v)/m$ term on the RHS of Eq. (36). After cancelling terms and using Eq. (32) to express $\mathrm{d}v_t$ in terms of $\mathrm{d}t$ and $\mathrm{d}B_t$, we find

$$\begin{aligned} \mathrm{d}\theta_t &= -\gamma\beta\|v\|^2\mathrm{d}t + \sqrt{2\gamma\beta}\,v\cdot\mathrm{d}B_t - \frac{\gamma}{m}v\cdot\nabla_v\ln\psi\,\mathrm{d}t + \sqrt{\frac{2\gamma}{\beta m^2}}\nabla_v\ln\psi\,\mathrm{d}B_t - \frac{\gamma}{\beta m^2}\big\|\nabla_v\ln\psi\big\|^2\mathrm{d}t \\ &= -\frac{1}{2}\bigg\|\sqrt{2\gamma\beta}v + \sqrt{\frac{2\gamma}{\beta m^2}}\nabla_v\ln\psi\bigg\|^2\mathrm{d}t + \bigg[\sqrt{2\gamma\beta}v + \sqrt{\frac{2\gamma}{\beta m^2}}\,\nabla_v\ln\psi\bigg]\cdot\mathrm{d}B_t. \end{aligned} \tag{39}$$

In the last equality, we rearrange the terms to show that $\mathrm{d}\theta_t = a_t\cdot\mathrm{d}B_t - \frac{1}{2}\|a_t\|^2\mathrm{d}t$ for some process a_t (in this case, $a_t \equiv \sqrt{2\gamma\beta}v_t + \sqrt{\frac{2\gamma}{\beta m^2}}\nabla_v\ln\psi_t$). As with Eq. (40), this form immediately implies that $\exp(\theta_t)$ is an exponential martingale:

$$\begin{aligned} \mathrm{d}e^{\theta_t} &= e^{\theta_t}\left[\mathrm{d}\theta_t + \frac{1}{2}\big(\mathrm{d}\theta_t\big)^2\right] \\ &= e^{\theta_t}\left[a_t\cdot \mathrm{d}B_t - \frac{1}{2}\|a_t\|^2\mathrm{d}t + \frac{1}{2}\|a_t\|^2\mathrm{d}t\right] \\ &= e^{\theta_t}a_t\cdot \mathrm{d}B_t. \end{aligned} \tag{40}$$

Consequently, provided that the Novikov condition holds, we find the martingale property

$$\mathbb{E}\left[e^{-\beta(W_t-\Delta H_t+T\Delta\Sigma_t)}\right] = 1. \tag{41}$$

Extension of Eq. (41) to a bounded stopping time τ follows immediately from the optional stopping theorem for martingales.

Manzano's result for underdamped Langevin dynamics. In order to obtain the classical work theorem, we choose a specific time t_1, and set $\psi(x,v,t_1) = \rho(x,v,t_1)$. It can be shown that $\psi(-x,v,t_1-t)$ follows a canonical Fokker-Planck equation; thus $\int\psi(x,v,t)\mathrm{d}x\mathrm{d}v = 1$. Similarly, one can derive

$$\mathbb{E}\left[e^{-\beta(W_\tau-\Delta F_\tau)-\delta_\tau}\right] = 1,\ \ \forall \text{ bounded stopping time } \tau\le t_1, \tag{42}$$

where $\delta_\tau \equiv \ln\rho(x_\tau,v_\tau,\tau) - \ln\psi(x_\tau,v_\tau,\tau)$ when ψ is chosen such that $\psi(x,v,t_1) = \rho(x,v,t_1)$.

Deterministic limits. Our derivation also carries through in the deterministic limits of zero-friction ($\gamma\to 0$) and zero-temperature ($T\to 0$). In the zero-friction limit, the noise term in Eq. (32) vanishes and the acceleration follows deterministic Hamiltonian dynamics *without friction*. We can directly compute the dimensionless entropy production of the system: $\sigma_t \equiv \beta(H_t - W_t - TS_t) = \beta(H_t - W_t) + \ln\rho(x_t,v_t,t)$. According to Liouville's theorem, $\frac{\mathrm{d}}{\mathrm{d}t}\rho(x_t,v_t,t) = 0$ in the deterministic limit. Thus,

$$\begin{aligned} \mathrm{d}\sigma_t &= \beta\mathrm{d}(H_t - W_t) \\ &= \frac{\gamma}{m}\big(n - \beta\|v\|^2\big)\mathrm{d}t + \sqrt{2\gamma\beta}\, v\cdot \mathrm{d}B_t, \end{aligned} \tag{43}$$

and in the $\gamma\to 0$ limit, we have $\sigma_t = \sigma_0$, almost surely. In other words, when the entropy is evaluated by the information entropy of the probability density, there is no entropy production in the deterministic limit even for a non-autonomous non-conservative Hamiltonian system.

In the zero-temperature limit ($\beta\to+\infty$), there is no contribution from the entropy term to the free energy difference and we have

$$\mathrm{d}(H_t - W_t) = -\frac{\gamma}{m}\|v\|^2\mathrm{d}t,$$

which shows that energy is dissipated into the environment at rate $\gamma\|v\|^2/m$.

Fluctuation-dissipation relation – underdamped limit. Similar to the overdamped case, we can explicitly compute $\mathrm{d}e^{\theta_t}$ in cases where the fluctuation-dissipation relation is not assumed. The general underdamped Langevin equation is given by

$$\begin{aligned} \mathrm{d}x_t &= v_t\,\mathrm{d}t, \\ m\,\mathrm{d}v_t &= \Big[-\gamma\, v_t - \nabla_x U(x_t,t) + f(x_t,t)\Big]\mathrm{d}t + \gamma\sqrt{2D}\,\mathrm{d}B_t, \end{aligned} \tag{44}$$

and the corresponding backward equation for ψ is given by

$$-\partial_t\psi(x,v,t) = v\cdot\nabla_x\psi + \frac{\gamma^2 D}{m^2}\Delta_v\psi - \frac{1}{m}\nabla_v\cdot\Big[\big(-\gamma v + \nabla_x U - f\big)\psi\Big]. \tag{45}$$

Following the same procedure as in Eqs. (32)–(40) and detailed in Appendix A.2, we can find

$$\mathrm{d}e^{\theta_t} = e^{\theta_t}\Big[\Big(D - \frac{1}{\beta\gamma}\Big)\frac{\gamma^2\beta}{m}\Big(\beta m\|v\|^2 + 2v_t \cdot \nabla_v \ln\psi_t + n\Big)\,\mathrm{d}t + \gamma\sqrt{2D}\Big(\beta\, v + \frac{1}{m}\nabla_v \ln\psi\Big)\cdot \mathrm{d}B_t\Big]. \tag{46}$$

When $D \neq 1/(\beta\gamma)$, the drift term is typically nonzero as the factor $\beta m\|v\|^2$ depends on the current velocity v of the particle. In general, $\mathbb{E}\left[\beta m\|v\|^2 + 2v_t \cdot \nabla_v \ln\psi_t + n\right]$ measures the deviation of velocity distribution from the Maxwell-Boltzmann distribution where $v \propto \exp(-\frac{\beta m}{2}\|v\|^2)$. When v follows the Maxwell-Boltzmann distribution, $\mathbb{E}\left[\beta m\|v\|^2 + 2v_t \cdot \nabla_v \ln\psi_t + n\right] = 0$.

3 Numerical examples

Drift-diffusion process with a singular initial distribution. We first provide a simple but useful numerical example on which to apply our work theorem. In this case, the initial distribution of the system is a delta distribution, making the trajectorywise entropy $S(x_0, 0)$ undefined at $t = 0$ since $\ln\delta(x)$ is undefined. The classical work theorem is invalid in this case. Our example is inspired by the continuous-time version of the kinetic proofreading (KPR) mechanism [28]. The KPR mechanism is a nonequilibrium process typically invoked in DNA replication or cell signaling that amplifies differences in the unbinding rates of different ligands in order to increase the specificity of ligand recognition. The original kinetic proofreading mechanism relies on multiple discrete activation steps on the ligand-receptor complex. In our previous work [29], we generalized the discrete activation process to a continuum process in the limit of large number of activation steps. In this continuum limit, can the work theorem provide a fundamental bound on the energy-information trade-off? Answering this question can provide additional geometric insight into the speed-energy-accuracy trade-off in nonequilibrium systems [31].

A simple schematic of a continuum kinetic proofreading process is shown in Fig. 3(a). Let $(x, \alpha) \in \mathbb{R} \times \{0, 1\}$ be the state of the receptor. x represents the level of activation or "reaction coordinate" and $\alpha = 1$ indicates the presence of ligand. Without ligand ($\alpha = 0$), the Hamiltonian $H(x)$ is minimized at $x = 0$. With bound ligand ($\alpha = 1$), an external force f arises such that $-\partial_x H(x) + f > 0$. Ligand-receptor binding and unbinding are assumed independent of the activation level x and follow exponential waiting time distributions, $\tau_+ \sim \mathrm{Exp}(k_+)$ and $\tau_- \sim \mathrm{Exp}(k_-)$, respectively. The complex is considered activated if $x > x^*$ for some threshold x^*. During the activation process ($\alpha = 1$), the activation probability $P(x > x^*)$ is roughly exponentially dependent on the unbinding rate k_-, thus enabling highly selective ligand recognition. The deactivation process ($\alpha = 0$) can be considered as erasure of memory (i.e., the activation level is reset to $x = 0$). Previous application of the work theorem leads to the Landauer principle, which suggests that erasing one bit of information requires $k_B T \ln 2$ work.

As a proof of concept, we further simplify the dynamics of the continuum proofreading process to a drift-diffusion process, as shown in Fig. 3(b). In the drift-diffusion process, we combine the Hamiltonian and force into a constant total force f so that Eq. (12) can be simplified to

$$\gamma \mathrm{d}x_t = f\,\mathrm{d}t + \sqrt{\frac{2\gamma}{\beta}}\,\mathrm{d}B_t. \tag{47}$$

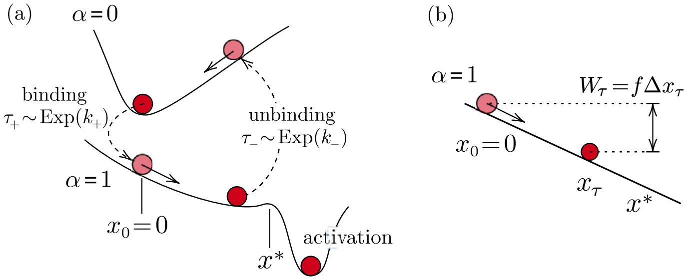

Fig. 3 (a) Schematic representation of the continuum kinetic proofreading process. The receptor state is defined by the activation level x and the ligand-binding status $\alpha \in \{0, 1\}$. When the ligand is absent ($\alpha = 0$), the system relaxes to the stable state at $x = 0$. Ligand binding ($\alpha = 1$) introduces an external force f that biases the activation dynamics, enabling transitions to higher activation levels. Binding and unbinding events follow exponential waiting time distributions, $\tau_+ \sim \mathrm{Exp}(k_+)$ and $\tau_- \sim \mathrm{Exp}(k_-)$, respectively. Activation occurs when $x > x^*$, with the probability of activation depending exponentially on the unbinding rate k_-, facilitating high selectivity in ligand recognition. Ligand unbinding ($\alpha = 0$) resets the activation level to $x = 0$, analogous to erasing memory as constrained by Landauer's principle. (b) A simplified drift-diffusion representation of the process with bound ligand ($\alpha = 1$). The dynamics reduce to an overdamped Langevin equation under a constant force f. Starting from $x_0 = 0$, the system evolves according to $\gamma \mathrm{d}x_t = f\mathrm{d}t + \sqrt{\frac{2\gamma}{\beta}}\mathrm{d}B_t$. The work performed by the force is given by $W = f\Delta x_\tau$, where Δx_τ is the displacement over time τ. This model captures the fundamental principles of energy expenditure and memory erasure in ligand-receptor systems

Given an initial position (state) $x_0 = 0$, the solution to Eq. (47) is

$$x_t = \frac{f}{\gamma}t + \sqrt{\frac{2\gamma}{\beta}}B_t. \tag{48}$$

The backward process $\psi(x, t)$ given by Eq. (16) now becomes

$$-\partial_t \psi = -\frac{f}{\gamma}\partial_x \psi + \frac{1}{\beta\gamma}\partial_x^2 \psi. \tag{49}$$

Using an initial δ-function distribution at $x = 0$, we can construct a family of solutions to Eq. (49) by time-reversing the solution of the corresponding Fokker-Planck equation using different terminal times t_2. Specifically,

$$\begin{aligned} p(x,t) &= \sqrt{\frac{\beta\gamma}{4\pi t}}\exp\left[-\frac{\beta\gamma\left(x - x_0 - \frac{f}{\gamma}t\right)^2}{4t}\right], \\ \psi(x, t \mid t_2) &= p(x, t_2 - t), \quad \text{for } t < t_2. \end{aligned} \tag{50}$$

In general, $\psi(x, t \mid t_2)$ solves Eq. (49) up to time t_2.

Now, consider the exponentially distributed stopping time $\tau_- \sim \mathrm{Exp}(k_-)$ for the unbinding process. In order to apply the optional sampling theorem, we bound the stopping time by the desired terminal time t_1: $\tau = \min\{\tau_-, t_1\}$. For concreteness, we choose $f = 0.5$, $\gamma = 1$, and $k_- = 1$ and numerically evaluate the expectations of normalized "entropy" production θ_τ, "entropy" change $\Delta\Sigma_\tau$, and e^{θ_τ} at the stopping time τ for different choices of $\psi(x, t \mid t_2 \geq t_1)$. The results are plotted in Fig. 4 and are consistent with predictions of our work theorem; the red curve in Fig. 4 shows that $\mathbb{E}[e^{\theta_\tau}] = 1$ for all choices of $t_2 \geq t_1$. By contrast, $\mathbb{E}[\theta_\tau]$ (blue curve) and $\mathbb{E}[\Delta\Sigma_\tau]$ (green curve) vary with t_2. It is noteworthy that this numerical example does not admit a classical work theorem counterpart, as the initial distribution of x_0 is not a density function and its trajectorywise entropy is not well-defined.

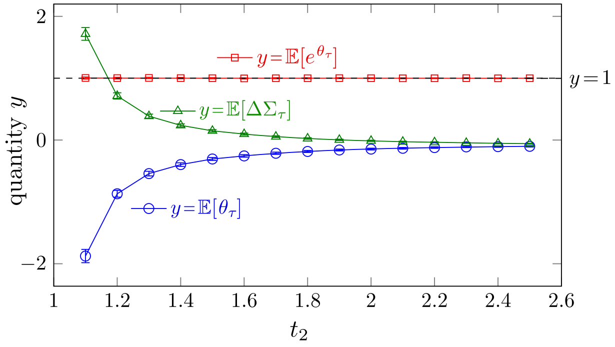

Fig. 4 The quantities $\mathbb{E}[e^{\theta_\tau}]$ (red squares), $\mathbb{E}[\theta_\tau]$ (blue circles), and $\mathbb{E}[\Delta\Sigma_\tau]$ (green triangles) for different choices of $t_2 \geq t_1 = 1$, constructed from simulated trajectories obeying Eq. (47), with $\gamma = 1$, $f = 0.5$, and $\beta = 1$. The red line is coincident with unity, consistent with our work theorem. By contrast, $\mathbb{E}[\theta_\tau]$ and $\mathbb{E}[\Delta\Sigma_\tau]$ vary with t_2, as indicated by the blue circles and green triangles along with the interpolating segments, respectively. We used a rate $k_- = 1$ to define the exponentially distributed time τ. 10^5 sample trajectories are used to evaluate the expectations for each t_2

This calculation is helpful for understanding how the "entropy" Σ changes with time for different choices of the backward process $\psi(x, t \mid t_2)$. In order to better understand the energy-information trade-off in the continuum proofreading mechanism, we need to relate the backward-process associated entropy $\Sigma(x, \tau \mid t_2)$ to the actual entropy $-k_\mathrm{B} \ln \rho(x_\tau)$ at the stopping time τ. We leave this as future work.

Violation of the fluctuation-dissipation relation in the Ornstein-Uhlenbeck process. We now consider the Ornstein-Uhlenbeck (OU) process defined by

$$\mathrm{d}x_t = -\frac{k}{\gamma} x_t \mathrm{d}t + \sqrt{2D}\,\mathrm{d}B_t, \tag{51}$$

which is Eq. (29) with $U(x) = \frac{k}{2}x^2$ and $f(x) = 0$. We choose a Gaussian initial distribution $\rho(x, 0) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}$ with variance σ^2 such that the classical work theorem can be applied to this scenario.

The probability density $\rho(x, t)$ of the OU process is explicitly given by

$$\rho(x, t) = \frac{1}{\sqrt{2\pi}\,\sigma(t)} e^{-\frac{x^2}{2\sigma^2(t)}}, \quad \sigma^2(t) = \sigma^2 e^{-kt/\gamma} + \frac{D\gamma}{k}\left(1 - e^{-kt/\gamma}\right). \tag{52}$$

To demonstrate the general applicability of our forward-backward-decoupled work theorem, we choose $\psi(x, t)$ to be the stationary distribution of the OU process:

$$\psi(x, t) = \sqrt{\frac{k}{2\pi D\gamma}}\, e^{-\frac{kx^2}{2D\gamma}}. \tag{53}$$

We again conduct numerical simulations of Eq. (51) with $\gamma = 1$, $k = 0.5$, and varying diffusion coefficient D up to a fixed time $t_1 = 2$. For each D, we simulate 10^7 trajectories to ensure convergence of the expectation. We evaluate both the classical exponentiated entropy production $\mathbb{E}\left[e^{\beta(W_t - \Delta F_t)}\right]$ and the generalized exponentiated entropy production $\mathbb{E}\left[e^{\theta_t}\right]$ at time t_1. Note that the exact expressions of $\rho(x, t)$ and $\psi(x, t)$ are used to compute the entropies in the two cases, respectively.

As is shown in Fig. 5, when $D = 1/(\beta\gamma) = 1$, both $\mathbb{E}\left[e^{\beta(W_t - \Delta F_t)}\right]$ and $\mathbb{E}\left[e^{\theta_t}\right]$ are equal to unity, consistent with the classical and generalized work theorems, respectively. As

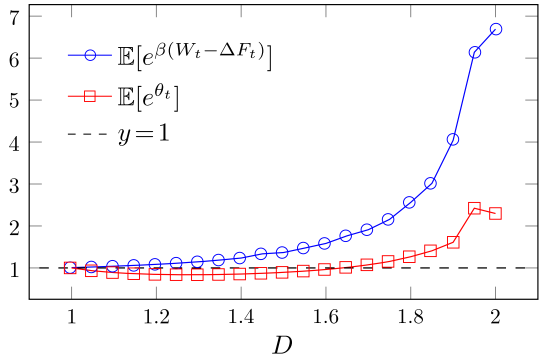

Fig. 5 The exponentiated entropy production $\mathbb{E}\left[e^{\beta(W_t-\Delta F_t)}\right]$ (blue circles) and $\mathbb{E}\left[e^{\theta_t}\right]$ (red squares) constructed from simulations of the OU process (Eq. 51) using different diffusion coefficients D. When $D = 1/(\beta\gamma) = 1$, both quantities are equal to unity. The stochastic simulations were performed using $\beta = 1$, $\gamma = 1$, $k = 0.5$, and $\sigma^2 = 10$ with a terminal time $t_1 = 2$. To ensure convergence, we used 10^7 trajectories

Table 1 Summary of variants of work theorems. Here, the fluctuation-dissipation relations is denoted FDR and "terminal time?" indicates whether a terminal time needs to be specified and whether a backward process has to be defined accordingly. The "identity?" column indicates whether averaging over trajectories up to a fixed time can lead to a Jarzynski identity. "Stopping time?" indicates whether there is a stopping-time form of the generalized Jarzynski identity, averaging over trajectories of different time durations

variant	initial cond.	FDR?	terminal time?	identity?	stopping time?
Jarzynski (Eq. 1)	Equilibrium only	Yes	Yes	Yes	No
Seifert (Eq. 6)	Non-singular	Yes	Yes	Yes	No
Sagawa and Ueda (Eq. 7)	Equilibrium only	Yes	Yes	Yes	No
Manzano, Neri (Eqs. 8-9)	Non-singular	Yes	Yes	Yes	Yes
Generalized (Eq. 24)	*Any*	Yes	No	Yes	Yes
Non-FDR (App. A)	–	No	No	No	No

D deviates from $1/(\beta\gamma) = 1$, $\mathbb{E}\left[e^{\beta(W_t-\Delta F_t)}\right]$ and $\mathbb{E}\left[e^{\theta_t}\right]$ can be either greater or less than unity, indicating a violation of the work theorem when the fluctuation-dissipation relation is not satisfied.

4 Discussion and Conclusions

Work theorems are fundamental because they characterize of entropy production in stochastic systems and generalize the second law of thermodynamics to nonequilibrium systems. In this work, we have shown that the work theorem can be further generalized to a broader class of stochastic systems by introducing a specific backward process that is essentially a time-reversed diffusion process with time-reversed Hamiltonian and force. However, unlike the path-integral formulation of Crooks' theorem, the backward process does not necessarily share a common distribution with the original "forward" process at any given time t. Specifically, when the Hamiltonian and force are time-independent, we provide a new equality that relates the work and energy change of the system to the stationary entropy at steady state. A comparison of different work theorem variants is provided in Table 1.

A key insight of our work is that forward and backward processes can be decoupled as shown in Fig. 2(a). The generalized work theorem also *decouples* the initial sample x_0 from the initial distribution $\rho(x,0)$, as shown in Fig. 2(b). Specifically, we have shown that $\mathbb{E}\big[e^{\theta_\tau} \mid x_0\big] = 1$ for any bounded stopping time τ and any initial sample x_0 in Eq. (24). The usual formulation of the work theorem involves an expectation or integral over the initial distribution $\rho(x,0)$, *i.e.*, $\int \mathbb{E}\big[e^{\beta(W_t-\Delta F_t)} \mid x_0\big]\rho(x_0,0)\,\mathrm{d}x_0 = 1$. Our analysis provides a stronger result: the work theorem holds for any initial sample x_0 regardless of the initial distribution. It also suggests new directions of investigation such as the physical interpretation of the backward process with different "initial" conditions. It is clear that when the backward process can be interpreted as a time-reversed process (with proper reflection in the case of underdamped dynamics), its probability distribution coincides with that of the forward process at a specific time. However, the martingale property and Eq. (24) do allow for different choices of the initial condition of the backward process. The simple stationary form of the generalized work theorem given in Eq. (28) is also of particular interest. Since the entropy term is time-independent, we do not need to compute the time evolution of $\rho(x,t)$, allowing easier analytical or numerical application of the generalized work theorem to complex systems.

We have also extended our methods to underdamped Langevin dynamics which is of particular interest in view of its natural connection to the deterministic Hamiltonian dynamics in the limit of vanishing friction. While the underdamped Langevin dynamics is widely anticipated to follow the same work theorem as the overdamped Langevin dynamics, our work provides a formal proof of this.

Nonequilibrium thermodynamics laws that are applicable to both underdamped and overdamped Langevin dynamics have been of recent interest [32, 33]. When measured by the information entropy of the probability density, the entropy production vanishes in the deterministic limit, even for a non-autonomous non-conservative Hamiltonian system. This result is expected but highlights the importance of understanding the definition of entropy in macroscopic systems, which may require coarse-graining measurement of the phase space [34]. On the other hand, because of the introduction of velocity in the underdamped dynamics, it is possible to separate the internal "temperature" of the system (defined by the kinetic energy of the system) from the external temperature (of the heat bath) [32]. Consequences of this separation are yet to be explored. Coupled thermal machines operating at different temperatures [19] is also another promising direction in which to extend work theorems.

Our Itô calculus-based derivation enables us to analyze the work theorem in cases where the fluctuation-dissipation relation is not satisfied. Our theoretical and numerical results emphasize that the work theorems require careful adaptation for nonequilibrium systems. This highlights how violation of the fluctuation-dissipation relation, common in active matter and biological systems, generates intrinsic entropy production that cannot be reconciled with equilibrium-based theorems.

Different choices of the backward process may shed light on important nonequilibrium processes. For example, from the original work theorem, one can derive the Landauer principle by setting $\Delta H = 0$, which states that the work required to erase one bit of information is $k_\mathrm{B}T \ln 2$; this has been experimentally verified and extended to include the effects of erasure speed [32, 35]. Faster erasure requires more total work. The original work theorem only provides the original bound of the Landauer principle. It would be worthwhile to investigate whether the generalized work theorem can incorporate the effects of erasure speed and provide a more accurate bound for the work required to erase information.

Lastly, from both theoretical and applied perspectives, one inconvenient aspect of the work theorem is that it requires working in unbounded spaces and the Novikov condition to hold. Additionally, experimental measurement may introduce discretization of the continuum distribution of states. Systems such as stochastic chemical reactions and other biological processes are more realistically described by bounded and/or discrete spaces. This has been numerically explored by Manzano *et al*. [9] using a simple two-state model. However, the theoretical understanding of the work theorem in general discrete spaces such as chemical reaction networks [36] remains relatively unexplored. Doing so could require using additional mathematical tools such as the cycle representation theory of Markov chains [37] and invoking insights such as the generalized Legendre-Frenchel transform and response theory for the steady state of nonequilibrium Markov chains [38, 38, 39].

Data Availability There were no data generated or analyzed in this study. The script for data generation and plotting is available at github.com/hsianktin/Jarzynski.

Declarations

Conflict of interest The authors have no conflicts of interest relevant to the content of this article.

A Mathematical Appendix

A.1 Overdamped Langevin dynamics with $D \neq 1/(\beta\gamma)$

In the main text, the derivation of $\mathrm{d}\theta_t$ and the martingale property of $\exp(\theta_t)$ relies on the specific choice of the diffusion coefficient $D = 1/(\beta\gamma)$. Here, we present a more general derivation that avoids this assumption. Instead, we consider the stochastic dynamics

$$\mathrm{d}x_t = \frac{1}{\gamma}\big[-\nabla H(x_t,t) + f(x_t,t)\big]\,\mathrm{d}t + \sqrt{2D}\,\mathrm{d}B_t, \tag{A1}$$

where D is a general diffusion coefficient that is no longer tied to the fluctuation-dissipation relation $D = 1/(\beta\gamma)$.

Setup We define the backward process by its probability density $\psi(x,t)$ that satisfies

$$-\partial_t\psi(x,t) = \frac{1}{\gamma}\nabla\cdot\Big[\big(\nabla H(x,t) - f(x,t)\big)\psi(x,t)\Big] + D\Delta\psi(x,t). \tag{A2}$$

We also recall the definitions:

$$W_t = \int_0^t f(x_s,s)\circ \mathrm{d}x_s + \int_0^t \partial_t H(x_s,s)\,\mathrm{d}s,$$

and

$$\Sigma(x_t,t) = -k_{\mathrm{B}}\ln\psi(x_t,t).$$

and consider Eq. 18

$$\theta_t = -\beta\big(W_t + H(x_t,t) - T\Sigma(x_t,t)\big) \tag{A3}$$

but with $\psi(x_t,t)$ in the definition of $\Sigma(x_t,t)$ determined by Eq. (A2) in which D is thus far arbitrary.

Derivation of $\mathrm{d}\theta_t$ First, we compute $\mathrm{d}H(x_t, t)$ using Itô's lemma:

$$\mathrm{d}H = \partial_t H\,\mathrm{d}t + \nabla H \cdot \mathrm{d}x_t + \frac{1}{2}\Delta H(\mathrm{d}x_t)^2. \tag{A4}$$

Since $(\mathrm{d}x_t)^2 = 2D\,\mathrm{d}t$, we find

$$\mathrm{d}H = \partial_t H\,\mathrm{d}t + \nabla H \cdot \mathrm{d}x_t + D\Delta H\,\mathrm{d}t. \tag{A5}$$

Then, from the definition of W_t, we have

$$\mathrm{d}W_t = f(x_t, t) \circ \mathrm{d}x_t + \partial_t H(x_t, t)\,\mathrm{d}t. \tag{A6}$$

Converting this Stratonovich integral to an Itô integral, we find

$$f(x_t, t) \circ \mathrm{d}x_t = f(x_t, t) \cdot \mathrm{d}x_t + D\nabla \cdot f(x_t, t)\,\mathrm{d}t \tag{A7}$$

and

$$\mathrm{d}W_t = f \cdot \mathrm{d}x_t + \partial_t H\,\mathrm{d}t + D(\nabla \cdot f)\,\mathrm{d}t. \tag{A8}$$

Finally, we decompose $\mathrm{d}\Sigma$ as

$$\mathrm{d}\Sigma = -k_\mathrm{B}\,\mathrm{d}\ln\psi(x_t, t) = -k_\mathrm{B}\left(\frac{\mathrm{d}\psi}{\psi} - \frac{1}{2}\left[\frac{\mathrm{d}\psi}{\psi}\right]^2\right). \tag{A9}$$

Upon using Eqs. (A5), (A8), and (A9) in Eq. (A3), we find

$$\begin{aligned}\mathrm{d}\theta_t &= -\beta\mathrm{d}W_t + \beta\mathrm{d}H - \beta\mathrm{d}\Sigma\\ &= -\beta\big(f \cdot \mathrm{d}x_t + \partial_t H\,\mathrm{d}t + D\nabla \cdot f\,\mathrm{d}t\big) + \beta\big(\partial_t H\,\mathrm{d}t + \nabla H \cdot \mathrm{d}x_t + D\Delta H\,\mathrm{d}t\big) + \frac{\mathrm{d}\psi}{\psi} - \frac{[\mathrm{d}\psi]^2}{2\psi^2}.\end{aligned} \tag{A10}$$

We now decompose the $\mathrm{d}\psi/\psi$ and $(\mathrm{d}\psi/\psi)^2$ by applying Itô's lemma to ψ,

$$\mathrm{d}\psi = \partial_t\psi\,\mathrm{d}t + \nabla\psi \cdot \mathrm{d}x_t + \frac{1}{2}\Delta\psi(\mathrm{d}x_t)^2, \tag{A11}$$

using Eq. (A2) to eliminate $\partial_t\psi$, and substituting $(\mathrm{d}x_t)^2 = 2D\,\mathrm{d}t$, to find

$$\mathrm{d}\psi = -\frac{1}{\gamma}\nabla \cdot \big[(\nabla H - f)\psi\big]\,\mathrm{d}t + \nabla\psi \cdot \mathrm{d}x_t. \tag{A12}$$

Using this to construct $\mathrm{d}\psi/\psi$ and $(\mathrm{d}\psi/\psi)^2$, we find

$$\begin{aligned}\frac{\mathrm{d}\psi}{\psi} &= -\frac{1}{\gamma}(\Delta H - \nabla \cdot f)\,\mathrm{d}t - \frac{1}{\gamma}(\nabla H - f) \cdot \nabla\ln\psi\,\mathrm{d}t + \nabla\ln\psi \cdot \mathrm{d}x_t\\ \frac{[\mathrm{d}\psi]^2}{2\psi^2} &= \frac{(\nabla\psi \cdot \mathrm{d}x_t)^2}{2\psi^2} = D\|\nabla\ln\psi\|^2\,\mathrm{d}t.\end{aligned} \tag{A13}$$

Substitution of Eqs. (A13) into Eq. (A10) leads to

$$\begin{aligned}a_t = &\left[\big(\beta D - \tfrac{1}{\gamma}\big)\big(\Delta H - \nabla \cdot f\big) - \frac{1}{\gamma}(\nabla H - f) \cdot \nabla\ln\psi - D\big\|\nabla\ln\psi\big\|^2\right]\mathrm{d}t\\ &+ \big(\beta(\nabla H - f) + \nabla\ln\psi\big) \cdot \mathrm{d}x_t\end{aligned} \tag{A14}$$

Since the forward SDE is $\mathrm{d}x_t = -\frac{1}{\gamma}(\nabla H - f)\,\mathrm{d}t + \sqrt{2D}\,\mathrm{d}B_t$ we can replace $\mathrm{d}x_t$ in Eq. (A14) to find

$$\mathrm{d}\theta_t = \Big(D - \frac{1}{\beta\gamma}\Big)\Big(\beta\big(\Delta H - \nabla\cdot f\big) - \big\|\nabla \ln\psi\big\|^2\Big)\mathrm{d}t - \frac{1}{\beta\gamma}\big\|\beta(\nabla H - f) + \nabla\ln\psi\big\|^2\mathrm{d}t + \sqrt{2D}\Big(\beta(\nabla H - f) + \nabla\ln\psi\Big)\cdot\,\mathrm{d}B_t. \tag{A15}$$

When $D = 1/(\beta\gamma)$, the last term in Eq. (A15) vanishes, rendering it equivalent to Eq. (22). In general,

$$\mathrm{d}e^{\theta_t} = e^{\theta_t}\Bigg[\Big(D - \frac{1}{\beta\gamma}\Big)\Big(\beta^2\big\|\nabla H - f\big\|^2 + 2\beta(\nabla H - f)\cdot\nabla\ln\psi + \beta(\Delta H - \nabla\cdot f)\Big)\mathrm{d}t + \sqrt{2D}\big(\beta(\nabla H - f) + \nabla\ln\psi\big)\cdot\mathrm{d}B_t\Bigg], \tag{A16}$$

from which it is apparent that $\mathrm{d}e^{\theta_t}$ is subject to non-zero drift and e^{θ_t} is not a martingale in general.

A.2 Underdamped Langevin dynamics with $D \neq 1/(\beta\gamma)$

We now outline how dynamics that deviate from the fluctuation-dissipation relation modify the work theorem. Consider an n-dimensional underdamped Langevin system described by

$$\begin{aligned} \mathrm{d}x_t &= v_t\,\mathrm{d}t, \\ m\,\mathrm{d}v_t &= \Big[-\gamma\, v_t - \nabla_x U(x_t, t) + f(x_t, t)\Big]\mathrm{d}t + \gamma\sqrt{2D}\,\mathrm{d}B_t, \end{aligned} \tag{A17}$$

where $x_t,\ v_t \in \mathbb{R}^n$, m, γ and D are constants, and B_t is a standard diagonal, n-dimensional Brownian motion. Let the backward process $\psi(x, v, t)$ be defined by the PDE

$$-\,\partial_t\psi(x, v, t) = v\cdot\nabla_x\psi + \frac{\gamma^2 D}{m^2}\Delta_v\psi - \frac{1}{m}\nabla_v\cdot\big[(-\gamma v + \nabla_x U - f)\,\psi\big]. \tag{A18}$$

Other quantities are defined in the same way as in the main text.

To compute $\mathrm{d}e^{\theta_t}$, first apply Itô's formula to $\mathrm{d}H_t$ to obtain

$$\mathrm{d}H_t = \Big[-\gamma\|v\|^2 + v\cdot f + \partial_t U + \frac{\gamma^2 D\, n}{m}\Big]\mathrm{d}t + \gamma\sqrt{2D}\, v\cdot\mathrm{d}B_t. \tag{A19}$$

From the definition of work and $\mathrm{d}x = v\mathrm{d}t$, we have

$$\mathrm{d}W_t = \Big[f(x_t, t)\cdot v_t + \partial_t U(x_t, t)\Big]\mathrm{d}t. \tag{A20}$$

Similarly, applying Itô's formula to $\mathrm{d}\psi$, inserting the PDE for ψ to eliminate $\partial_t\psi$, and assuming that f and U are v-independent, we obtain

$$\mathrm{d}\ln\psi = \Big[-\frac{2\gamma}{m}\, v\cdot\nabla_v\ln\psi - \frac{\gamma n}{m} - \frac{\gamma^2 D}{m^2}\big\|\nabla_v\ln\psi\big\|^2\Big]\mathrm{d}t + \frac{\gamma\sqrt{2D}}{m}\nabla_v\ln\psi\cdot\mathrm{d}B_t. \tag{A21}$$

Given the above results, we now derive the differential of $\theta_t = \beta(\Delta H_t - W_t) + \Delta\ln\psi_t$:

$$\mathrm{d}\theta_t = \Big[-\beta\gamma\|v\|^2 + \beta\frac{\gamma^2 D n}{m} - \frac{2\gamma}{m}\, v\cdot\nabla_v\ln\psi - \frac{\gamma n}{m} - \frac{\gamma^2 D}{m^2}\big\|\nabla_v\ln\psi\big\|^2\Big]\,\mathrm{d}t + \gamma\sqrt{2D}\Big[\beta\, v + \frac{1}{m}\nabla_v\ln\psi\Big]\cdot\mathrm{d}B_t. \tag{A22}$$

After expanding and collecting terms, $\mathrm{d}e^{\theta_t}$ becomes

$$\mathrm{d}e^{\theta_t} = e^{\theta_t}\Bigg[\Big(D - \frac{1}{\beta\gamma}\Big)\frac{\gamma^2\beta}{m}\Big(\beta m\|v\|^2 + 2v_t\cdot\nabla_v \ln\psi_t + n\Big)\,\mathrm{d}t + \gamma\sqrt{2D}\Big(\beta\, v + \frac{1}{m}\nabla_v \ln\psi\Big)\cdot \mathrm{d}B_t\Bigg]. \tag{A23}$$

Notice that $D = 1/(\beta\gamma)$ is a sufficient condition for the drift term to vanish and our generalized work theorem to hold. The term $\beta m\|v\|^2 + 2v\cdot\nabla_v \ln\psi + n$ can be a measure of the violation of the the equipartition theorem.

If the system is in equilibrium such that $\psi(v) \propto e^{-\frac{\beta m}{2}\|v\|^2}$ and the equipartition theorem holds ($\mathbb{E}[\beta m\|v\|^2] = n$), then on average

$$\mathbb{E}\left[\beta m\|v\|^2 + 2v\cdot\nabla_v \ln\psi + n\right] = n - 2n + n = 0. \tag{A24}$$

In this case, the generalized work theorem holds for underdamped Langevin dynamics even when $D \neq 1/(\beta\gamma)$ provided $\psi(v,t) \propto e^{-\frac{\beta m}{2}\|v\|^2}$ for all t.

References

1. Jarzynski, C.: Nonequilibrium equality for free energy differences. Phys. Rev. Lett. **78**(14), 2690–2693 (1997)
2. Jarzynski, C.: Equilibrium free-energy differences from nonequilibrium measurements: A master-equation approach. Phys. Rev. E **56**(5), 5018–5035 (1997)
3. Sevick, E.M., Prabhakar, R., Williams, Stephen R., Searles, Debra J.: Fluctuation theorems. Annual Review of Physical Chemistry **59**(1), 603–633 (2008)
4. Crooks, G.E.: Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences. Phys. Rev. E **60**(3), 2721–2726 (1999)
5. Sekimoto, K.: Langevin equation and thermodynamics. Prog. Theor. Phys. Suppl. **130**, 17–27 (1998)
6. Seifert, U.: Entropy production along a stochastic trajectory and an integral fluctuation theorem. Phys. Rev. Lett. **95**(4), 040602 (2005)
7. Qian, H.: Mesoscopic nonequilibrium thermodynamics of single macromolecules and dynamic entropy-energy compensation. Phys. Rev. E **65**(1), 016102 (2001)
8. Kurchan, J.: Fluctuation theorem for stochastic dynamics. J. Phys. A: Math. Gen. **31**(16), 3719–3729 (1998)
9. Manzano, G., Subero, D., Maillet, O., Fazio, R., Pekola, J.P., Roldán, É.: Thermodynamics of gambling demons. Phys. Rev. Lett. **126**(8), 080603 (2021)
10. Sagawa, T., Ueda, M.: Generalized Jarzynski equality under nonequilibrium feedback control. Phys. Rev. Lett. **104**(9), 090602 (2010)
11. Neri, I.: Second law of thermodynamics at stopping times. Phys. Rev. Lett. **124**(4), 040601 (2020)
12. Yang, H., Ge, H.: Fluctuation theorems and thermodynamic inequalities for nonequilibrium processes stopped at stochastic times. Phys. Rev. E. **108**(5), (2023)
13. Hernández-Gómez, S., Poggiali, F., Cappellaro, P., Cataliotti, F. S., Trombettoni, A., Fabbri, N., Gherardini, S.: Energy exchange statistics and fluctuation theorem for non-thermal asymptotic states. arXiv, quant-ph:2404.05310v1, (2024)
14. Albeverio, S.: Wiener and Feynman-path integrals and their applications. In: Proceedings of Symposia in Applied Mathematics, volume 52, pages 163–194. American Mathematical Society, (1997)
15. Mazzucchi, S.: Mathematical Feynman path integrals and their applications. World Scientific (2009)
16. Albeverio, S.: Mathematical aspects of Feynman path integrals, divergences, quantum fields and diagrams, and some more general reflections. In: When Form Becomes Substance: Power of Gestures, Diagrammatical Intuition and Phenomenology of Space, pages 267–282. Springer, (2022)
17. Dutta, A., Sarkar, S.: Fluctuation theorem as a special case of Girsanov theorem, (2023)
18. Brown, A.I., Sivak, D.A.: Theory of nonequilibrium free energy transduction by molecular machines. Chem. Rev. **120**(1), 434–459 (2019)
19. Leighton, M.P., Ehrich, J., Sivak, D.A.: Information arbitrage in bipartite heat engines. Phys. Rev. X **14**(4), 041038 (2024)

20. Toyabe, S., Sagawa, T., Ueda, M., Muneyuki, E., Sano, M.: Experimental demonstration of information-to-energy conversion and validation of the generalized Jarzynski equality. Nat. Phys. **6**(12), 988–992 (2010)
21. An, S., Zhang, J.-N., Um, M., Lv, D., Yao, L., Zhang, J., Zhang-Qi Yin, H.T., Quan, Kim, K.: Experimental test of the quantum Jarzynski equality with a trapped-ion system. Nat. Phys. **11**(2), 193–199 (2014)
22. Paneru, G., Dutta, S., Sagawa, T., Tlusty, T., Pak, H.K.: Efficiency fluctuations and noise induced refrigerator-to-heater transition in information engines. Nat. Commun. **11**(1), 1012 (2020)
23. Roldán, É., Neri, I., Chetrite, R., Gupta, S., Pigolotti, S., Jülicher, F., Sekimoto, K.: Martingales for physicists: a treatise on stochastic thermodynamics and beyond. Adv. Phys. **72**(1–2), 1–258 (2023)
24. Karatzas, I., Shreve, S.: Brownian motion and stochastic calculus, vol. 113. Springer (2014)
25. Anderson, Brian D. O.: Reverse-time diffusion equation models. Stoch. Process. Their Appl. **12**(3), 313–326 (1982)
26. Pigolotti, S., Neri, I., Roldán, É., Jülicher, F.: Generic properties of stochastic entropy production. Phys. Rev. Lett. **119**(14), (2017)
27. Neri, I., Roldán, É., Pigolotti, S., Jülicher, F.: Integral fluctuation relations for entropy production at stopping times. J. Stat. Mech: Theory Exp. **2019**(10), 104006 (2019)
28. Hopfield, J.J.: Kinetic proofreading: A new mechanism for reducing errors in biosynthetic processes requiring high specificity. Proc. Natl. Acad. Sci. **71**(10), 4135–4139 (1974)
29. Li, X.., Chou, T..: Reliable ligand discrimination in stochastic multistep: kinetic proofreading: First passage time vs. product counting strategies. PLoS Computational Biology **20**(6), e1012183 (2024)
30. Hathcock, D., Qiwei, Yu., Mello, B.A., Amin, D.N., Hazelbauer, G.L., Yuhai, T.: A nonequilibrium allosteric model for receptor-kinase complexes: The role of energy dissipation in chemotaxis signaling. Proc. Natl. Acad. Sci. **120**(42), e2303115120 (2023)
31. Klinger, J., Rotskoff, G. M.: Universal energy-speed-accuracy trade-offs in driven nonequilibrium systems. arXiv, cond-mat.stat-mech:2402.17931v2, (2024)
32. Dago, S., Bellon, L.: Dynamics of information erasure and extension of Landauer's bound to fast processes. Phys. Rev. Lett. **128**(7), 070604 (2022)
33. Lyu, J., Ray, K.J., Crutchfield, J.P.: Learning entropy production from underdamped Langevin trajectories. Phys. Rev. E **110**(6), 064151 (2024)
34. Barkan, C.O.: On the convergence of phase space distributions to microcanonical equilibrium: dynamical isometry and generalized coarse-graining. J. Phys. A: Math. Gen. **57**(47), 475001 (2024)
35. Proesmans, K., Ehrich, J., Bechhoefer, J.: Finite-time Landauer principle. Phys. Rev. Lett. **125**(10), 100602 (2020)
36. Schmiedl, T., Seifert, U.: Stochastic thermodynamics of chemical reaction networks. J. Chem. Phys. **126**(4), 044101 (2007)
37. Jiang, Y., Bingjie, W., Jia, C.: Large deviations and fluctuation theorems for cycle currents defined in the loop-erased and spanning tree manners: A comparative study. Physical Review Research **5**(1), 013207 (2023)
38. Yang, Y.-J., Dill, K. A.: Deriving the forces of nonequilibria from two laws. arXiv, cond-mat.stat-mech:2410.09277v1, (2024)
39. Zheng, J., Lu, Z.: Universal non-equilibrium response theory beyond steady states. arXiv, cond-mat.stat-mech:2403.10952v4, (2024)