Illustrative example of the relationship between dissipation and relative entropy

Jordan Horowitz[1] and Christopher Jarzynski[2]

[1]*Department of Physics, University of Maryland, College Park, Maryland 20742, USA*

[2]*Department of Chemistry and Biochemistry and Institute for Physical Science and Technology, University of Maryland, College Park, Maryland 20742, USA*

Kawai, Parrondo, and Van den Broeck [Phys. Rev. Lett. **98**, 080602 (2007)] have recently established a quantitative relationship between dissipated work and a microscopic, information-theoretic measure of irreversibility. We illustrate this result using the exactly solvable system of a Brownian particle in a dragged harmonic trap.

I. INTRODUCTION

During a thermodynamic process in which a system, in contact with a thermal reservoir, evolves from one state of thermal equilibrium A, to another B, the average work performed on the system must exceed the free energy difference: $\langle W \rangle \geq \Delta F = F_B - F_A$ [1]. The work in excess of ΔF, i.e., the *dissipated work* $\langle W_{\text{diss}} \rangle = \langle W \rangle - \Delta F$, quantifies the amount of energy irretrievably lost to the surrounding thermal environment. Recently, Kawai, Parrondo, and Van den Broeck [2] have related $\langle W_{\text{diss}} \rangle$ to another measure of irreversibility, roughly speaking the distinction between the forward and reverse directions of the "arrow of time." Specifically, they have shown that

$$\beta \langle W_{\text{diss}} \rangle \geq D(\rho_F \| \rho_R), \tag{1}$$

where ρ_F and ρ_R are time-dependent phase-space densities describing the evolution of the system from A to B, and from B to A, respectively; and $D(\rho \| \rho') = \int \rho \ln(\rho / \rho')$ denotes the *relative entropy* [3], a measure of the distinguishability between two distributions. When the system in question evolves deterministically, under Hamilton's equations, then Eq. (1) is an equality. However, if the system description is coarse grained, or explicitly stochastic, then the relative entropy only provides a lower bound on the dissipated work.

Connections between dissipation and temporal asymmetry similar to Eq. (1) have also been established by other authors. Maes [4] and later Maes and Netočný [5] have obtained a correspondence between thermodynamic measures of dissipation and relative entropies between forward and reverse distributions in path space; and Gaspard [6] has connected this result to the dynamical randomness that characterizes nonequilibrium steady states. In the context described in the previous paragraph, Jarzynski [7] has derived a relation analogous to Eq. (1), but expressed in terms of distributions in path space [see Eq. (4)].

The goal of the present paper is to illustrate Eq. (1) using the simple model of a particle inside a one-dimensional, moving harmonic well. This model is both analytically tractable and experimentally relevant [8–14]. After briefly reviewing the central result of Ref. [2] in Sec. II, we introduce our model at the beginning of Sec. III. We then solve the model and illustrate Eq. (1) when the system evolves under Hamilton's equations (Sec. III A) and both overdamped and inertial Langevin dynamics (Secs. III B and III C, respectively). In Sec. IV we extend our analysis and show that the bound in Eq. (1) can be improved by specifying the phase-space density at two times, rather than one. Gomez-Marin *et al.* [15] have also studied this model in the context of Eq. (1) and in Sec. IV we briefly discuss the relationship between our results and theirs.

II. BACKGROUND

Consider a classical system with N degrees of freedom, described by the coordinates $\mathbf{x} = \{x_1, \ldots, x_N\}$ and conjugate momenta $\mathbf{p} = \{p_1, \ldots, p_N\}$. Let $z = (\mathbf{x}, \mathbf{p})$ denote a point in phase space, and let $H(z, \lambda)$ denote a parameter-dependent Hamiltonian for this system. We assume that this Hamiltonian is time-reversal invariant: $H(z^*, \lambda) = H(z, \lambda)$, where the asterisk denotes the reversal of momenta, $\mathbf{p} \rightarrow -\mathbf{p}$.

The term *thermodynamic process* will indicate a sequence of events whereby the system evolves in phase space as the external parameter λ is varied according to an arbitrary schedule, or *protocol* λ_t [also labeled as $\lambda(t)$], from an initial time $t = 0$ to a final time $t = \tau$. During this interval of time, the microscopic evolution of the system is specified by a trajectory z_t, or $z(t)$, and $\rho(z, t)$ will denote the time-dependent phase-space density describing an ensemble of such trajectories. As in Ref. [2] we will explicitly consider two such processes: one defined by a *forward* protocol λ_t^F, during which the parameter is varied from an initial value $\lambda_0^F = A$ to a final value $\lambda_\tau^F = B$, the other defined by a *reverse* protocol $\lambda_t^R = \lambda_{\tau-t}^F$, from $\lambda_0^R = B$ to $\lambda_\tau^R = A$.

Prior to the start of either the forward or the reverse process ($t < 0$), the system is brought to equilibrium by weak contact with a thermal reservoir at inverse temperature β. As a result, the initial phase-space density is a canonical distribution at $\lambda = A$ or B. The system might subsequently be thermally isolated from the reservoir, or else it might remain in contact with the reservoir. In the former case, Hamilton's equations govern its evolution from $t = 0$ to τ, while in the latter case we will use stochastic dynamics to model the random effects of the environment. In either situation the *work* performed on the system during this interval of time is given by the following functional of the trajectory:

$$W[z_t] = \int_0^\tau dt \, \dot{\lambda} \frac{\partial H}{\partial \lambda}(z_t, \lambda_t). \tag{2}$$

For the case of Hamiltonian dynamics, the average dissipated work can be expressed as (see Ref. [2] for the details of the derivation)

$$\beta\langle W_{\mathrm{diss}}\rangle = \int dz \rho_F(z,t) \ln\left[\frac{\rho_F(z,t)}{\rho_R(z^*,\tau-t)}\right]$$

$$= D[\rho_F(z,t) \| \rho_R(z^*,\tau-t)]. \qquad (3)$$

Since relative entropy is a measure of the distinguishability of two probability distributions [3], Eq. (3) relates the dissipation of energy to the ease with which a process can be distinguished from its time-reversal [2,15,16].

Equation (3) is an *equality* because the underlying Hamiltonian dynamics are deterministic. If the system remains in contact with the reservoir during the process, we instead obtain an inequality, Eq. (1). There are two ways to argue this, both of which make use of a key property of the relative entropy between two distributions, namely that it decreases when the distributions are projected onto a smaller set of variables [3]. When the system remains in contact with a reservoir, then we can view the system and the reservoir as two subsystems that together form a large, isolated "supersystem" to which Eq. (3) applies: the dissipated work gives the relative entropy between the two distributions in the full phase space. Upon projecting out the reservoir variables, the relative entropy decreases, and we obtain Eq. (1), as discussed in Ref. [2]. Alternatively, we note that an equality analogous to Eq. (3) can be formulated for distributions in *path space*, rather than phase space [7]:

$$\beta\langle W_{\mathrm{diss}}\rangle = D[\mathcal{P}_F(\gamma^F) \| \mathcal{P}_R(\gamma^R)]. \qquad (4)$$

Here the trajectory γ^F describes the evolution of the system during a given realization of the forward process, and $\mathcal{P}_F(\gamma^F)$ is the probability distribution of such trajectories; γ^R and $\mathcal{P}_R$ are defined in a similar manner for the reverse process. Equation (4) holds for both deterministic and stochastic evolution. If we now project from path space onto phase space, e.g., $\mathcal{P}_F(\gamma^F) \to \rho_F(z,t)$, then the relative entropy decreases, and we again obtain the inequality, Eq. (1). (Reference [15] refers to this projection procedure as "coarse-graining in time.") In either case, the decrease of relative entropy has a simple interpretation: when we discard microscopic information—either about the state of the reservoir, or about states of the system itself at times other than the specified instant, t—then we diminish our ability to distinguish between the forward and the reverse process. This suggests that the more microscopic information we retain, the closer the value of D is to its upper bound $\beta\langle W_{\mathrm{diss}}\rangle$. We investigate this in Sec. IV, where we consider a generalization of Eq. (1) in which the microstate of the system is specified at two instants in time, rather than one. References [2,15,16] contain a related analysis specifically addressing the loss of microscopic information during coarse graining and how this loss affects the value of the relative entropy.

III. PARTICLE IN A MOVING HARMONIC WELL

In this section we illustrate Eq. (1) through an explicit calculation of the average dissipated work and relative entropy for a particle in a one-dimensional, moving harmonic well. We begin by specifying the model and obtaining useful preliminary results. Then in Sec. III A we evaluate the average dissipated work and relative entropy for a system following Hamiltonian dynamics. In Secs. III B and III C we repeat the calculation with the system modeled by Langevin dynamics, in different limiting regimes.

Our system is a particle of mass m trapped in a harmonic well with spring constant k:

$$H(x,p,\lambda) = \frac{p^2}{2m} + \frac{k}{2}(x-\lambda)^2. \qquad (5)$$

We will consider processes during which the center of the well is moved either rightward or leftward at constant speed u. These correspond to forward and reverse protocols, $\lambda_F(t) = ut$ and $\lambda_R(t) = u(\tau-t)$. Explicit expressions for the initial equilibrium densities are given by the following Gaussians:

$$\rho_F(z,0) = \frac{\beta}{2\pi}\sqrt{\frac{k}{m}}\exp\left[-\beta\left(\frac{p^2}{2m} + \frac{kx^2}{2}\right)\right], \qquad (6a)$$

$$\rho_R(z,0) = \frac{\beta}{2\pi}\sqrt{\frac{k}{m}}\exp\left[-\beta\left(\frac{p^2}{2m} + \frac{k}{2}(x-u\tau)^2\right)\right]. \qquad (6b)$$

For this system $\Delta F = 0$ by translational symmetry, therefore for the forward process

$$\langle W_{\mathrm{diss}}\rangle = -uk\int_0^\tau dt[\bar{x}_F(t) - ut], \qquad (7)$$

where $\bar{x}_F(t)$ is the average position during the forward process.

Since the time-dependent densities $\rho_F(z,t)$ and $\rho_R(z,t)$ will prove to be Gaussians for all the cases considered in this paper (see Secs. III A–III C), it is useful here to establish uniform notation for the description of two-dimensional Gaussian distributions $f_G(z) = f_G(x,p)$. Such a distribution is uniquely determined by the moments (means and covariances),

$$\bar{x} = \int dz f_G(z)x, \qquad (8a)$$

$$\bar{p} = \int dz f_G(z)p, \qquad (8b)$$

$$\sigma_x^2 = \int dz f_G(z)(x-\bar{x})^2, \qquad (8c)$$

$$\sigma_p^2 = \int dz f_G(z)(p-\bar{p})^2, \qquad (8d)$$

$$\sigma_{xp} = \int dz f_G(z)(x-\bar{x})(p-\bar{p}). \qquad (8e)$$

An explicit expression for $f_G(z)$ in terms of these moments is

$$f_G(z) = \frac{1}{2\pi\sqrt{\det\sigma}} \exp\left[-\frac{1}{2}(z-\bar{z})^T\sigma^{-1}(z-\bar{z})\right], \quad (9)$$

where z is a vector in phase space and σ is the covariance matrix:

$$z = \begin{pmatrix} x \\ p \end{pmatrix}, \quad \sigma = \begin{pmatrix} \sigma_x^2 & \sigma_{xp} \\ \sigma_{xp} & \sigma_p^2 \end{pmatrix}. \quad (10)$$

Lastly, the relative entropy between two Gaussian distributions, $f_G(z)$ and $g_G(z^*)$, is [see Eq. (3)]

$$D[f_G(z) \| g_G(z^*)] = -1 + \frac{1}{2}\left[\ln\left(\frac{\det\sigma_g}{\det\sigma_f}\right) + \mathrm{Tr}(\sigma_g^{-1}\sigma_f^*)\right]$$
$$+ \frac{1}{2}(\bar{z}_f^* - \bar{z}_g)^T\sigma_g^{-1}(\bar{z}_f^* - \bar{z}_g), \quad (11)$$

where $\sigma_{xp}^* = -\sigma_{xp}$ and all elements of σ^* are unaltered [17].

A. Hamiltonian dynamics

We now consider the case in which the system is thermally isolated from the reservoir after the initial equilibration stage, and thereafter evolves under Hamilton's equations as the harmonic well is translated leftward or rightward,

$$\dot{x}_F = p_F/m, \quad \dot{p}_F = -k(x_F - ut), \quad (12a)$$

$$\dot{x}_R = p_R/m, \quad \dot{p}_R = -k(x_R - u\tau + ut). \quad (12b)$$

To solve for the evolution of the phase-space densities ρ_F and ρ_R, we observe that since the initial density is Gaussian [Eqs. (6)] and the equations of motion are linear [Eqs. (12)], the distribution remains Gaussian for all times [18]. Thus we need only determine the means and (co)variances as functions of time. From Eqs. (6) we have the initial means

$$\bar{x}_F(0) = 0, \quad \bar{p}_F(0) = 0, \quad \bar{x}_R(0) = u\tau, \quad \bar{p}_R(0) = 0, \quad (13a)$$

and the initial variances

$$\sigma_x^2(0) = \frac{m}{\beta}, \quad \sigma_p^2(0) = \frac{1}{\beta k}, \quad \sigma_{xp}^2(0) = 0, \quad (13b)$$

which are the same for the forward and reverse processes. Solving Eqs. (12) leads to a set of linear equations for the positions and momenta in terms of their initial conditions. Combining these solutions with Eqs. (8), (13a), and (13b) leads to the solutions

$$\bar{x}_F(t) = u\left[t - \frac{\sin(\omega t)}{\omega}\right], \quad \bar{p}_F(t) = mu[1 - \cos(\omega t)], \quad (14a)$$

$$\bar{x}_R(t) = u\left[\tau - t + \frac{\sin(\omega t)}{\omega}\right], \quad \bar{p}_R(t) = mu[\cos(\omega t) - 1], \quad (14b)$$

$$\sigma_x^2(t) = \frac{m}{\beta}, \quad \sigma_p^2(t) = \frac{1}{\beta k}, \quad \sigma_{xp}(t) = 0, \quad (14c)$$

where $\omega^2 = k/m$.

We can now explicitly verify Eq. (1). The relative entropy between the forward phase-space density $\rho_F(z,t)$ and the reverse phase-space density $\rho_R(z^*, \tau - t)$ is determined by plugging Eqs. (14) into Eq. (11),

$$D[\rho_F(z,t) \| \rho_R(z^*, \tau - t)] = \beta mu^2[1 - \cos(\omega\tau)]. \quad (15)$$

Comparing this with

$$\beta\langle W_{\mathrm{diss}}\rangle = \beta mu^2[1 - \cos(\omega\tau)], \quad (16)$$

computed from Eqs. (14a) and (7), we find the predicted result $\beta\langle W_{\mathrm{diss}}\rangle = D(\rho_F\|\rho_R)$.

B. Overdamped Langevin dynamics

We now imagine that the system remains in contact with the thermal reservoir throughout the process, and we will use Langevin dynamics to model the presence of the reservoir. As discussed in Sec. II, Eq. (1) should apply as a strict inequality in this case.

In this section we consider the overdamped limit, in which the momentum effectively equilibrates instantaneously, and as a result the momentum does not contribute to the relative entropy. We therefore focus on the position dynamics. For the forward process the Fokker-Planck equation for $\rho_F(x,t)$ is

$$\frac{\partial}{\partial t}\rho_F(x,t) = \frac{k}{\gamma}\frac{\partial}{\partial x}[(x - ut)\rho_F(x,t)] + \frac{1}{\gamma\beta}\frac{\partial^2}{\partial x^2}\rho_F(x,t), \quad (17)$$

where γ is the friction coefficient.

To solve Eq. (17), we recognize that an initially Gaussian distribution will remain Gaussian for all time under the evolution of Eq. (17), as can be checked by substitution (see, for example, [8,19]). Thus $\rho_F(x,t)$ is Gaussian with mean and variance,

$$\bar{x}_F(t) = ut - \frac{\gamma u}{k}(1 - e^{-kt/\gamma}), \quad \sigma_F^2 = \frac{1}{\beta k}. \quad (18)$$

For the reverse process we replace t with $\tau - t$ in Eq. (17). The solution is a Gaussian distribution with mean and variance,

$$\bar{x}_R(t) = u(\tau - t) + \frac{\gamma u}{k}(1 - e^{-k(\tau - t)/\gamma}), \quad \sigma_R^2 = \frac{1}{\beta k}. \quad (19)$$

We now use these results to calculate the relative entropy and average dissipated work. Using Eq. (11) we obtain

$$D[\rho_F(x,t) \| \rho_R(x, \tau - t)]$$
$$= \frac{2\beta\gamma^2 u^2}{k}\left\{1 - e^{-k\tau/2\gamma}\cosh\left[\frac{k}{\gamma}\left(\frac{\tau}{2} - t\right)\right]\right\}^2. \quad (20)$$

The average dissipated work is obtained by substituting Eq. (18) into Eq. (7) and evaluating the integral:

$$\beta\langle W_{\mathrm{diss}}\rangle = \beta\gamma u^2\left[\tau - \frac{\gamma}{k}(1 - e^{-k\tau/\gamma})\right]. \quad (21)$$
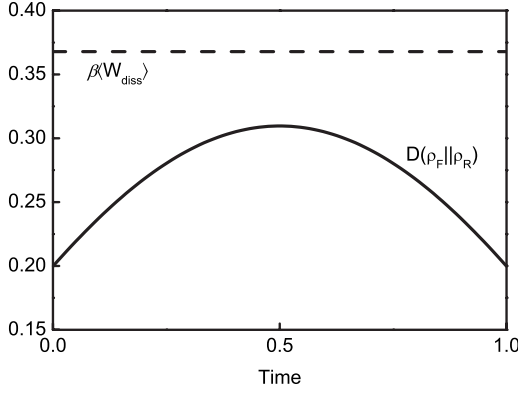
FIG. 1. Comparison of $\beta \langle W_{\text{diss}} \rangle$ (dashed line) and $D(\rho_F \| \rho_R)$ (solid line) in the overdamped limit. All parameters (k, γ, u, and β) have been set to one in their respective units.

Verifying Eq. (1) requires demonstrating that the average dissipated work is always greater than the relative entropy for any values of system parameters. To begin, we note that $\beta \langle W_{\text{diss}} \rangle$ does not depend on t and $D(\rho_F \| \rho_R)$ obtains its maximum value D_{max} at time $t = \tau/2$. Combining Eqs. (20) and (21), we get

$$\beta \langle W_{\text{diss}} \rangle - D(\rho_F \| \rho_R) \geqslant \beta \langle W_{\text{diss}} \rangle - D_{\text{max}}$$
$$= \frac{\beta \gamma^2 u^2}{k} [\zeta - (3 - e^{-\zeta/2})(1 - e^{-\zeta/2})] \geqslant 0,$$

$$(22)$$

where $\zeta = k\tau/\gamma$ is the scaled time. Using introductory calculus techniques, it is easy to verify that the bracketed quantity in Eq. (22) is non-negative for any $\zeta \geqslant 0$. As an example, we plot $\beta \langle W_{\text{diss}} \rangle$ and $D(\rho_F \| \rho_R)$ in Fig. 1 as functions of time for k, γ, u, and β all set to one.

In the quasistatic limit, namely $u \to 0$ with $u\tau$ fixed, both the dissipation and relative entropy approach zero, although at different rates. Specifically, from Eqs. (20) and (21), $\beta \langle W_{\text{diss}} \rangle \sim u$ and $D(\rho_F \| \rho_R) \sim u^2$. This limiting behavior can be justified on general grounds since we are considering continuous Markovian stochastic processes [20]. A heuristic argument is as follows: consider such a continuous Markovian stochastic process perturbed by varying an external parameter λ according to a specified protocol $\lambda(t)$. In the quasistatic limit, i.e., $\dot{\lambda} \sim \epsilon \ll 1$, the phase-space density is approximately

$$P(z, t) \sim P_{\lambda(t)}^{ss}(z) + \epsilon \delta P(z, t), \quad (23)$$

where $P_{\lambda}^{ss}(z)$ is the unique stationary state with fixed external parameter λ and $\delta P(z, t)$ is the first-order correction to the phase-space density [21]. Combining Eq. (23) with the definitions of average dissipated work [cf. Eq. (2)] and relative entropy, and taking the limit $\epsilon \to 0$, leads to $\beta \langle W_{\text{diss}} \rangle \sim \epsilon$ and $D(\rho_F \| \rho_R) \sim \epsilon^2$. In the above model, $\lambda(t) = ut$ with $\epsilon = u$.

C. Full phase-space Langevin dynamics

In this section we analyze the same stochastic system without assuming the overdamped limit. (We do assume that
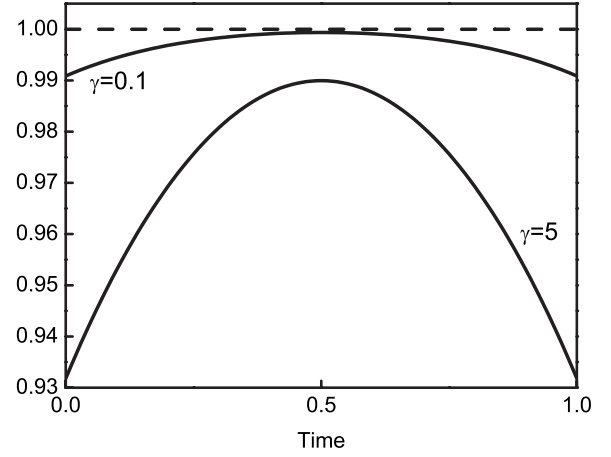

FIG. 2. Comparison of $\beta \langle W_{\text{diss}} \rangle / D(\rho_F \| \rho_R)$ for various values of the friction coefficient γ; the solid lines are, from bottom to top, $\gamma = 5$ and 0.1. The dashed line represents Hamiltonian dynamics ($\gamma = 0$) with $\beta \langle W_{\text{diss}} \rangle = D(\rho_F \| \rho_R)$. All other system parameters (m, k, u, and β) have been set to one in their respective units.

the motion is not critically damped, i.e., $\gamma^2 \neq 4mk$.) Using the same notation as the previous section III B the Fokker-Planck equation for the phase-space density of the forward process is

$$\frac{\partial}{\partial t} \rho_F(z, t) = -\frac{p}{m} \frac{\partial}{\partial x} \rho_F(z, t) + \frac{\partial}{\partial p} \left[\left(k(x - ut) + \frac{\gamma p}{m} \right) \rho_F(z, t) \right]$$
$$+ \frac{\gamma}{\beta} \frac{\partial^2}{\partial p^2} \rho_F(z, t). \quad (24)$$

While Eq. (24) can be solved exactly, the solution is complicated and unilluminating. An analytic solution is presented in the Appendix; here, we illustrate Eq. (1) by plotting the ratio $\beta \langle W_{\text{diss}} \rangle / D(\rho_F \| \rho_R)$ in Fig. 2 for various values of the friction coefficient γ.

Roughly speaking, γ measures the coupling between the system and the reservoir: for larger γ, the reservoir and system interact more strongly; whereas, when $\gamma \to 0$, the reservoir and the system decouple, with the result that the system evolves independently of the reservoir, under deterministic Hamiltonian dynamics. Figure 2 clearly shows that as γ decreases, the relative entropy approaches the average dissipated work. Decreasing γ weakens the interaction between the system and the reservoir, resulting in less microscopic information leaking into the reservoir variables. The result is more information is contained in the system variables, reflected by an increase in the relative entropy. The flow of microscopic information between system and reservoir degrees of freedom and its relation to dissipation has previously been discussed in the context of other models [16].

IV. TWO-TIME PHASE-SPACE DENSITIES

In the previous sections, the phase-space densities were evaluated at one particular time. For Hamiltonian dynamics this is enough information to determine the average dissipated work. In contrast, when the dynamics are stochastic the

phase-space densities alone do not contain enough information to determine the average dissipated work. In this section we address this issue by investigating the implications of specifying the phase-space density at two times. Let $P(z_1, t_1; z_0, t_0)$ be the "two-time" probability for the system to be at z_0 at time t_0 *and* z_1 at a later time t_1. We will explicitly demonstrate that two-time distributions provide a better bound for the average dissipated work: $\beta\langle W_{\mathrm{diss}}\rangle \geq D(P_F\|P_R) \geq D(\rho_F\|\rho_R)$.

For clarity we consider the overdamped limit as in Sec. III B. The goal is to determine the pairwise probability $P(x_1, t_1; x_0, t_0)$. For the forward process, we can decompose the joint probability distribution into a conditional probability to be at x_1 at t_1 *given* the system was at x_0 at t_0, $\rho_F(x_1, t_1|x_0, t_0)$, and a one-time probability $\rho_F(x_0, t_0)$, i.e., $P_F(x_1, t_1; x_0, t_0) = \rho_F(x_1, t_1|x_0, t_0)\rho_F(x_0, t_0)$. The expression for $\rho_F(x_0, t_0)$ was determined in Sec. III B [see Eq. (18)]. The conditional probability is computed by solving the Fokker-Planck equation [see Eq. (17)] with the initial condition $\delta(x-x_0)$ at t_0, i.e., the initial density is Gaussian with $\bar{x}(t_0) = x_0$ and $\sigma_x^2(t_0) = 0$. At time t_1, the mean and variance for the forward process conditional probability are

$$\bar{x}(t_1) = e^{-k(t_1-t_0)/\gamma}x_0 + u\left(t_1 - t_0 e^{-k(t_1-t_0)/\gamma}\right) - \frac{\gamma u}{k}(1 - e^{-k(t_1-t_0)/\gamma}),$$
(25a)

$$\sigma_x^2(t_1) = \frac{1}{\beta k}(1 - e^{-2k(t_1-t_0)/\gamma}).$$
(25b)

The pairwise distribution is the product of Gaussian distributions; as such, it is Gaussian as well. A simple, yet lengthy calculation shows that the means and variances of the pairwise distribution for the forward process are

$$\bar{\mathbf{x}}_F = \begin{pmatrix} \bar{x}_0 \\ \bar{x}_1 \end{pmatrix} = \begin{pmatrix} ut_0 - \dfrac{\gamma u}{k}(1 - e^{-kt_0/\gamma}) \\[2mm] ut_1 - \dfrac{\gamma u}{k}(1 - e^{-kt_1/\gamma}) \end{pmatrix},$$
(26a)

$$\sigma_F = \begin{pmatrix} \sigma_{x_0}^2 & \sigma_{x_0 x_1} \\ \sigma_{x_0 x_1} & \sigma_{x_1}^2 \end{pmatrix} = \frac{1}{\beta k}\begin{pmatrix} 1 & e^{-k(t_1-t_0)/\gamma} \\ e^{-k(t_1-t_0)/\gamma} & 1 \end{pmatrix}.$$
(26b)

Repeating the above calculation for the reverse joint distribution leads to similar results.

The next step is to determine the relative entropy. We are interested in comparing the pairwise probability for the forward process $P_F(x_1, t_1; x_0, t_0)$, with the *time-reversed* pairwise probability for the reverse process $P_R(x_0, \tau - t_0; x_1, \tau - t_1)$. If we allow t_0 and t_1 to be arbitrary, then from Eq. (11) the relative entropy is

$$D(P_F\|P_R) = \frac{2\beta\gamma^2 u^2/k}{1 - e^{-2k|t_1-t_0|/\gamma}}(\alpha_0^2 + \alpha_1^2 - 2e^{-k|t_1-t_0|/\gamma}\alpha_0\alpha_1),$$
(27)

where
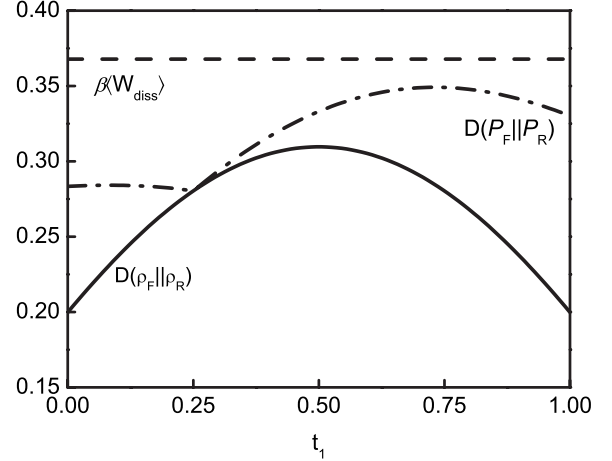

FIG. 3. Comparison of $D(P_F\|P_R)$ (dot-dashed line), $D(\rho_F\|\rho_R)$ (solid line), and $\beta\langle W_{\mathrm{diss}}\rangle$ (dashed line) in the overdamped limit. All system parameters have been set to one in their respective units.

$$\alpha_i = 1 - e^{-k\tau/2\gamma}\cosh\left[\frac{k}{\gamma}\left(\frac{\tau}{2} - t_i\right)\right], \quad i = 0, 1. \quad (28)$$

Figure 3 compares the "two-time" relative entropy $D(P_F\|P_R)$, "one-time" relative entropy $D(\rho_F\|\rho_R)$, and the average dissipated work. For $D(P_F\|P_R)$, we fix $t_0 = \tau/4$. We see that $\beta\langle W_{\mathrm{diss}}\rangle \geq D(P_F\|P_R) \geq D(\rho_F\|\rho_R)$ for all times t_1. Thus by specifying an additional time in our probability distributions we have improved the bound in Eq. (1).

This result is an instance of the coarse-graining procedure outlined in Sec. II [see the discussion below Eq. (4)], where the path-space probability distribution $\mathcal{P}_F(\gamma^F)$ has been projected down onto a two-time phase-space density. Projecting onto a two-time phase-space density eliminates less microscopic information than projecting onto a one-time phase-space density; hence the two-time relative entropy is greater than the one-time relative entropy.

In a related independent analysis, Gomez-Marin *et al.* [15] used the present model to investigate the behavior of the "N-time" relative entropy D_N. By evaluating D_N at N equally spaced times, i.e., $t_{i+1} - t_i = \tau/N$, they were able to demonstrate that $\beta\langle W_{\mathrm{diss}}\rangle - D_N \sim 1/N^2$ as $N \to \infty$. We note that our two-time relative entropy agrees with theirs: choosing $t_1 = \tau$ and $t_0 = 0$ in Eq. (27), we recover Eq. (22) (with $n = 1$) of Ref. [15].

V. CONCLUSION

The average energy dissipated by a thermodynamic process is related to how distinguishable the process is from its time reversal, i.e., how well one can discern the "arrow of time." This paper provides a pedagogical illustration of this idea using an exactly solvable model. For deterministic Hamiltonian systems, complete knowledge of the forward and reverse phase-space density at any *one* time is enough to determine the dissipation. This is not true for stochastic systems. Calculating the dissipation from microscopic

information in stochastic systems requires knowledge of the entire evolution of the system, both for the forward and reverse processes. Partial knowledge gives only a lower bound on the dissipation. We have seen that the tightness of the lower bound is correlated with the amount of information known about the system's evolution. Specifically, when the system is loosely coupled to a reservoir, very little information is lost to the reservoir. As a result the relative entropy between the system's forward and reverse phase-space densities reasonably approximates the dissipation in this limit. The lower bound can also be tightened using a multitime relative entropy where the microscopic state of the system is specified many times along the system's trajectory.

ACKNOWLEDGMENTS

APPENDIX: PHASE-SPACE DENSITY FOR SEC. III C

In this appendix, we solve Eq. (24) for the full phase-space density of Sec. III C. To begin, the coefficients of Eq. (24) are linear in x and p. Consequently, as in the previous sections, an initial Gaussian distribution will remain Gaussian. Again we must determine the means, $\bar{x}$ and $\bar{p}$, and covariance matrix σ as functions of time. Combining Eq. (24) with the derivatives of Eqs. (8) we find the equations of motion for the means and variances of the forward process

$$\frac{d\bar{x}_F}{dt} = \frac{\bar{p}_F}{m}, \quad \frac{d\bar{p}_F}{dt} = -k(\bar{x}_F - ut) - \frac{\gamma}{m}\bar{p}_F, \quad \text{(A1a)}$$

$$\frac{d\sigma_x^2}{dt} = \frac{2}{m}\sigma_{xp}, \quad \frac{d\sigma_p^2}{dt} = -\frac{2\gamma}{m}\sigma_p^2 - 2k\sigma_{xp} + \frac{2\gamma}{\beta},$$

$$\frac{d\sigma_{xp}}{dt} = -\frac{\gamma}{m}\sigma_{xp} - k\sigma_x^2 + \frac{1}{m}\sigma_p^2. \quad \text{(A1b)}$$

The reverse process phase-space density is described by a similar set of equations.

The solutions of Eqs. (A1a) and their reverse process counterparts are

$$\bar{x}_F(t) = Ae^{r_+t} + Be^{r_-t} + u(t - \gamma/k), \quad \text{(A2a)}$$

$$\bar{p}_F(t) = mr_+Ae^{r_+t} + mr_-Be^{r_-t} + mu, \quad \text{(A2b)}$$

$$\bar{x}_R(t) = -Ae^{r_+t} - Be^{r_-t} + u(\tau - t + \gamma/k), \quad \text{(A2c)}$$

$$\bar{p}_R(t) = -mr_+Ae^{r_+t} - mr_-Be^{r_-t} - mu, \quad \text{(A2d)}$$

where

$$r_\pm = \frac{-\gamma \pm \sqrt{\gamma^2 - 4mk}}{2m} \quad \text{(A3)}$$

and

$$\binom{A}{B} = \frac{-mu}{\sqrt{\gamma^2 - 4mk}} \binom{\gamma r_+/k + 1}{-\gamma r_-/k - 1} \quad \text{(A4)}$$

are determined by the initial conditions [Eq. (13a)]. The solutions to Eq. (A1b) with initial conditions given by Eq. (13b) are

$$\sigma_x^2(t) = \frac{m}{\beta}, \quad \sigma_p^2(t) = \frac{1}{\beta k}, \quad \sigma_{xp}(t) = 0. \quad \text{(A5)}$$

Using Eq. (7), the average dissipated work is

$$\langle W_{\text{diss}} \rangle = \gamma u^2 \tau + uk\left[\frac{A}{r_+}(1 - e^{r_+t}) + \frac{B}{r_-}(1 - e^{r_-t})\right]. \quad \text{(A6)}$$

From Eq. (11), the relative entropy is

$$D(\rho_F \| \rho_R) = \frac{1}{2}\{\beta k[A(e^{r_+t} + e^{r_+(\tau-t)}) + B(e^{r_-t} + e^{r_-(\tau-t)})]^2$$

$$+ \beta m[r_+A(e^{r_+t} - e^{r_+(\tau-t)}) + r_-B(e^{r_-t} - e^{r_-(\tau-t)})]^2\}. \quad \text{(A7)}$$

[1] L. Landau and E. M. Lifshitz, *Statistical Physics: Part 1*, Course of Theoretical Physics Vol. 5 (Elsevier, New York, 1980).

[2] R. Kawai, J. M. R. Parrondo, and C. Van den Broeck, Phys. Rev. Lett. **98**, 080602 (2007).

[3] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. (Wiley-Interscience, New York, 2006).

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

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

[6] P. Gaspard, J. Stat. Phys. **117**, 599 (2004).

[7] C. Jarzynski, Phys. Rev. E **73**, 046105 (2006).

[8] O. Mazonka and C. Jarzynski, e-print arXiv:cond-mat/991212.

[9] R. van Zon, S. Ciliberto, and E. G. D. Cohen, Phys. Rev. Lett. **92**, 130601 (2004).

[10] N. Garnier and S. Ciliberto, Phys. Rev. E **71**, 060101(R) (2005).

[11] G. M. Wang, J. C. Reid, D. M. Carberry, D. R. M. Williams, E. M. Sevick, and D. J. Evans, Phys. Rev. E **71**, 046142 (2005).

[12] G. M. Wang, E. M. Sevick, E. Mittag, D. J. Searles, and D. J. Evans, Phys. Rev. Lett. **89**, 050601 (2002).

[13] D. M. Carberry, J. C. Reid, G. M. Wang, E. M. Sevick, D. J. Searles, and D. J. Evans, Phys. Rev. Lett. **92**, 140601 (2004).

[14] F. Douarche, S. Joubaud, N. B. Garnier, A. Petrosyan, and S. Ciliberto, Phys. Rev. Lett. **97**, 140603 (2006).

[15] A. Gomez-Marin, J. M. R. Parrondo, and C. Van den Broeck, Phys. Rev. E **78**, 011107 (2008).

[16] A. Gomez-Marin, J. M. R. Parrondo, and C. Van den Broeck, Europhys. Lett. **82**, 50002 (2008).

[17] S. Ihara, *Information Theory For Continuous Systems* (World Scientific, Singapore, 1993).

[18] R. Kubo, M. Toda, and N. Hashitsume, *Statistical Physics II: Nonequilibrium Statistical Mechanics* (Springer-Verlag, Berlin, 1985).

[19] R. van Zon and E. G. D. Cohen, Phys. Rev. E **67**, 046102 (2003).

[20] S. Vaikuntanathan, J. Horowitz, S. Rahav, and C. Jarzynski (unpublished).

[21] Equation (23) can be derived from the Fokker-Planck equation by employing a two-time scale analysis.