Nonequilibrium detailed fluctuation theorem for repeated discrete feedback

Jordan M. Horowitz[1] and Suriyanarayanan Vaikuntanathan[2]
[1]*Departamento de Física Atómica, Molecular y Nuclear, Universidad Complutense de Madrid, 28040 Madrid, Spain*
[2]*Institute for Physical Science and Technology, University of Maryland, College Park, Maryland 20742, USA*
(Received 16 July 2010; revised manuscript received 19 October 2010; published 13 December 2010)

We extend the framework of forward and reverse processes commonly utilized in the derivation and analysis of the nonequilibrium work relations to thermodynamic processes with repeated discrete feedback. Within this framework, we derive a generalization of the detailed fluctuation theorem, which is modified by the addition of a term that quantifies the change in uncertainty about the microscopic state of the system upon making measurements of physical observables during feedback. As an application, we extend two nonequilibrium work relations: the nonequilibrium work fluctuation theorem and the relative-entropy work relation.

DOI: 10.1103/PhysRevE.82.061120 PACS number(s): 05.70.Ln, 05.20.−y

I. INTRODUCTION

The *nonequilibrium work relations* are a family of predictions concerning the fluctuations in the work performed on a microscopic system driven far from equilibrium [1–8]. They have been important for the study of fundamental issues in the thermodynamics of small systems and have proven to be powerful tools for calculating equilibrium free-energy differences from nonequilibrium processes, both in experiments [9–11] as well as in computer simulations [12].

At the heart of the nonequilibrium work relations is a statement about the time-reversal symmetry of the microscopic dynamics termed as the *detailed fluctuation theorem* [13–19] (also called microscopic reversibility [3] or the generalized fluctuation-dissipation theorem [1,20,21]). The detailed fluctuation theorem relates the probability to observe microscopic trajectories of the system through phase space during two thermodynamic processes related by time reversal: the forward process and the reverse process. This framework of forward and reverse processes has been beneficial for investigating the role of irreversibility at the microscopic scale [6].

However, the nonequilibrium work relations and the detailed fluctuation theorem are not applicable to systems manipulated using *feedback*—a procedure in which microscopic information about a system is utilized to manipulate or control its evolution. Given the frequency with which feedback occurs in physics, biology, and engineering [22], it is important to extend the work fluctuation relations to include feedback. This will clarify the thermodynamics of feedback [23–26], as well as the thermodynamics of computation [27–29], and possibly elucidate the role of information processing in control theory [30,31].

Feedback can be implemented *discretely* through a series of feedback loops initiated at a sequence of predetermined times or *continuously* at every instant of time. Initial investigations into the work fluctuation relations in the presence of continuous feedback were made by Kim and Qian in the context of molecular refrigerators driven by velocity-dependent feedback control [32]. The first work relation extended to include discrete feedback was the nonequilibrium work fluctuation theorem [2], recently reported by Sagawa and Ueda [33]. They demonstrated that when a system is manipulated using one feedback loop the nonequilibrium work fluctuation theorem is modified by the addition of a term that accounts for the microscopic information gained during feedback. In this paper, we develop a framework of forward and reverse processes for *repeated* discrete feedback in order to analyze and extend Sagawa and Ueda's result. Moreover, we generalize the detailed fluctuation theorem to include repeated discrete feedback. We find that the information gained during feedback must be incorporated into the work relations. As an application, we extend the nonequilibrium work fluctuation theorem [2,33] as well as the relative-entropy work relation [4,5] in the presence of repeated feedback. (While this paper was under consideration, similar results were published [34]. We postpone a discussion comparing Ref. [34] with the present work until the conclusion.)

Our central result [Eq. (1) below] can be summarized as follows. Consider a classical thermodynamic system initially in equilibrium at inverse temperature β. Imagine driving this system away from equilibrium from time $t=0$ to τ by implementing a series of feedback loops at N predetermined times t_k, with $k=1,\ldots,N$. At each t_k, a physical observable M_k is measured. Based on the outcome of this measurement we drive the system by varying a set of external parameters λ with time. In each repetition or realization of this entire process, which we call the *forward* process, the system will trace out a different microscopic trajectory $\gamma_{\tau,0}$ through phase space. Furthermore, the protocol Λ_t used to vary the external parameters λ will differ in each realization due to fluctuations in the measurement process. We are interested in comparing the statistics of $\gamma_{\tau,0}$ and Λ_t in the forward process to those of the time-reversed conjugate pairs $\tilde{\gamma}_{\tau,0}$ and $\tilde{\Lambda}_t$ in the time-reversed process, which we call the *reverse* process. There is no feedback in the reverse process (no measurements are made). Instead, an ensemble of realizations of the reverse process is generated by executing each external parameter protocol observed in the forward process in reverse. Our main result is that the ratio of the probability to observe $\gamma_{\tau,0}$ and Λ_t in the forward process $\mathcal{P}[\gamma_{\tau,0};\Lambda_t]$ to the probability to observe $\tilde{\gamma}_{\tau,0}$ and $\tilde{\Lambda}_t$ in the reverse process $\tilde{\mathcal{P}}[\tilde{\gamma}_{\tau,0};\tilde{\Lambda}_t]$ satisfies a *detailed fluctuation theorem for discrete feedback*:

 ©2010 The American Physical Society

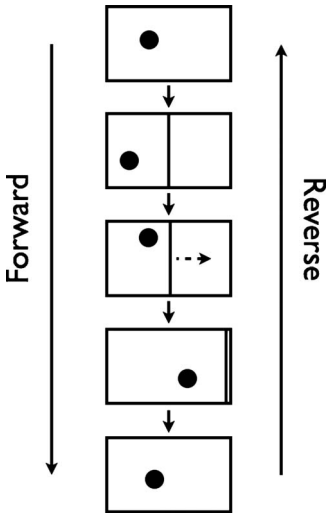

FIG. 1. Illustration of a realization of the forward and the reverse processes for the Szilard engine in which the particle is measured to be in the left half of the box. The forward process is depicted by the sequence of illustrations running from top to bottom. The reverse process is the time-reversed forward process; as such, time flows from bottom to top.

$$\frac{\mathcal{P}[\gamma_{\tau,0};\Lambda_t]}{\tilde{\mathcal{P}}[\tilde{\gamma}_{\tau,0};\tilde{\Lambda}_t]} = e^{\beta W_d[\gamma_{\tau,0};\Lambda_t]+I[\gamma_{\tau,0};\Lambda_t]}, \tag{1}$$

where W_d is the dissipated work. The new quantity appearing in Eq. (1), I, quantifies the change in our uncertainty about the microscopic state of the system upon measuring the physical observables $M_1, M_2, \ldots, M_N$ in each realization. The average of I over many realizations $\langle I \rangle$ is the mutual information [30], which is an information theoretic measure of the reduction in our uncertainty about the microscopic state of the system upon making measurements. Moreover, $\langle I \rangle$ naturally appears in the thermodynamics of feedback developed by Cao and Feito [26], where $\langle I \rangle$ is equal to the reduction in the Shannon entropy of a thermodynamic system under feedback control (see Eq. (7) of Ref. [26]).

Our analysis begins in Sec. II by motivating the definitions of the forward and reverse processes using the Szilard engine as a pedagogical example. Our main result [Eq. (1)] is then derived in Sec. III. In Sec. IV, the interpretation of I is developed in detail using ideas from Bayesian inference. Equation (1) is then exploited in Sec. V to generalize two work fluctuation relations: the nonequilibrium work fluctuation theorem [2,33] and the relative-entropy work relation [4,5]. Finally, we conclude in Sec. VI with an outlook toward future research directions.

II. MOTIVATION AND DEFINITIONS

Before deriving Eq. (1), it is instructive to first motivate and establish the definitions of the forward and reverse processes in the context of the Szilard engine [35], depicted in Fig. 1. This will generalize the usual notions of forward and reverse processes common in the study of the work relations [3,4]. The Szilard engine is composed of a single ideal-gas particle in a box of volume V in thermal contact with a heat bath at inverse temperature β. We begin by describing the forward process, which is illustrated in Fig. 1 by the sequence of snapshots proceeding from top to bottom. Initially, the engine is allowed to relax to equilibrium. A partition is then inserted in the center of the box, isolating the particle in either the left or right half of the box. Feedback begins by measuring in which half of the box the particle is located. The position of the partition is then shifted in a manner that depends on the measurement outcome: if the particle is found in the left (right) half of the box, the partition is slide all the way to the right (left). Finally, the partition is removed and the particle is allowed to relax back to equilibrium. Imagine repeating this process a number of times, each time equilibrating the particle, implementing the feedback loop, and finally allowing the engine to relax back to equilibrium. This generates an ensemble of realizations of the forward process.

Within the framework of the work relations, the reverse process is implemented by carrying out each step (or each macroscopic control action) of the forward process in the reverse order. For feedback processes the external parameter protocols are implemented in response to the outcomes of measurements. The naive time reversal of this procedure—implementing a protocol and then making the measurement used to determine this protocol—would be acausal, because we would have to implement protocols in response to measurements made after the protocol was executed. Instead, we generate an ensemble of realizations of the reverse process by first generating an ensemble of realizations of the forward process and then implementing the reverse of each protocol, which was observed in the forward process. For example, suppose that we observe a realization of the Szilard engine in which the particle was found in the left half of the box and the partition was moved to the right. Having observed this realization of the forward process, we generate a realization of the reverse process by actuating each action of the forward process in reverse, which is depicted in Fig. 1 by reading the images from bottom to top. The particle is first equilibrated at inverse temperature β. The partition is then inserted on the right side of the box and then slide to the center. Finally, the partition is removed and the engine is allowed to relax back to equilibrium. Repeating this process a number of times, each time reversing an observed realization of the forward process, generates an ensemble of realizations of the reverse process.

Observe that in the reverse process no measurements are performed. Instead, in an ensemble of realizations of the reverse process the protocols are implemented randomly according to the distribution in which they occur in the forward process. The reverse process cannot be executed independently of the forward process; one must first perform the forward process. This reliance of the reverse process on the forward process is a consequence of the time-reversal asymmetry of feedback and is an essential difference between thermodynamic processes with and without feedback.

III. DERIVATION

We are now in a position to derive Eq. (1). Let us begin by fixing notation. Consider a classical system, whose posi-

tion in phase space (or microscopic configuration) is $z=(\mathbf{x},\mathbf{p})$, where $\mathbf{x}$ denotes the system's coordinates and $\mathbf{p}$ denotes its momentum. The energy of the system $E(z,\lambda)$ is parametrized by a vector of controllable external parameters λ and is assumed to be time-reversal invariant for each fixed λ, $E(z,\lambda)=E(z^*,\lambda)$, where $z^*=(\mathbf{x},-\mathbf{p})$. The dynamics are assumed to be Markovian dynamics (which includes deterministic dynamics) that preserve the canonical equilibrium distribution for each fixed λ:

$$P^{eq}(z|\lambda) = e^{\beta[F(\lambda)-E(z,\lambda)]}, \tag{2}$$

where $F(\lambda)$ is the free energy. The position of the system at time t will be denoted as z_t. The collection of phase-space points visited by the system during the course of its evolution from $t=r$ to s will be termed as a *microscopic trajectory* and will be labeled $\gamma_{s,r}=\{z_t\}_{t=r}^{s}$.

The *forward process* is defined as the following sequence of events. The system is initially equilibrated with a thermal reservoir at inverse temperature β with the external parameters fixed at $\lambda=A_0$. Consequently, the initial statistical state of the system is $P^{eq}(z|A_0)$ [Eq. (2)]. From $t=0$ to t_1 the system is driven away from equilibrium by varying λ with time using a predetermined initial protocol λ_t^0, from $\lambda_0^0=A_0$ to $\lambda_{t_1}^0=B_0$. Then at subsequent times t_k, with $k=1,\ldots,N$, feedback loops are implemented. At each t_k a physical observable M_k is measured with (possibly continuous) outcomes m_k. Each measurement outcome m_k occurs with a probability that depends on the phase-space position of the system at the time of measurement, $P_k(m_k|z_{t_k})$, and is independent of the previous measurements. We collect all the measurement outcomes up to and including time t_k into a vector $\mu_k=\{m_1,\ldots,m_k\}$, which we call the *measurement trajectory* up to t_k. During each time interval from $t=t_k$ to t_{k+1} $(t_{N+1}=\tau)$ the external parameters are varied using a protocol which depends on the outcomes of all measurements up to t_k, $\lambda_t^k(\mu_k)$, from $\lambda_{t_k}^k(\mu_k)=A^k(\mu_k)$ to $\lambda_{t_{k+1}}^k(\mu_k)=B^k(\mu_k)$. Additionally, we assume that each μ_k is associated to a unique protocol [i.e., $\lambda_t^k(\mu_k)\neq\lambda_t^k(\mu_k')$ for all $\mu_k\neq\mu_k'$] and that $A^k(\mu_k)=B^{k-1}(\mu_{k-1})$ to ensure that the protocol is continuous at each measurement time t_k. The microscopic trajectory γ_{t_{k+1},t_k} taken by the system during this time interval occurs with probability $P[\gamma_{t_{k+1},t_k}|z_k,\lambda_t^k(\mu_k)]$, which is conditioned only on the position of the system at time t_k, z_{t_k}—since the dynamics are Markovian—and depends on the protocol executed $\lambda_t^k(\mu_k)$. The *complete protocol* executed from $t=0$ to τ is represented by collecting the individual protocols used in each feedback loop into a vector, $\Lambda_t(\mu_N)=\{\lambda_t^0,\ldots,\lambda_t^N(\mu_N)\}$. The probability to observe a realization of the entire forward process with trajectory $\gamma_{\tau,0}$ and protocol $\Lambda_t(\mu_N)$ is

$$\begin{aligned}\mathcal{P}[\gamma_{\tau,0};\Lambda_t] = {} & P[\gamma_{\tau,t_N}|z_{t_N},\lambda_t^N(\mu_N)]P_N(m_N|z_{t_N})\cdots \\ & \times P[\gamma_{t_2,t_1}|z_{t_1},\lambda_t^1(\mu_1)]P_1(m_1|z_{t_1}) \\ & \times P[\gamma_{t_1,0}|z_0,\lambda_t^0]P^{eq}(z_0|A_0).\end{aligned} \tag{3}$$

The work done on the system along this trajectory is

$$\begin{aligned}W[\gamma_{\tau,0};\Lambda_t] &= \sum_{k=0}^{N} W^k[\gamma_{t_{k+1},t_k};\lambda_t^k(\mu_k)] \\ &= \sum_{k=0}^{N}\int_{t_k}^{t_{k+1}} ds\dot{\lambda}_s^k(\mu_k)\frac{\partial}{\partial\lambda}E[z_s,\lambda_s^k(\mu_k)],\end{aligned} \tag{4}$$

the heat flow into the system is

$$\begin{aligned}Q[\gamma_{\tau,0};\Lambda_t] &= \sum_{k=0}^{N} Q^k[\gamma_{t_{k+1},t_k};\lambda_t^k(\mu_k)] \\ &= \sum_{k=0}^{N}\int_{t_k}^{t_{k+1}} ds\dot{z}_s\frac{\partial}{\partial z}E[z_s,\lambda_s^k(\mu_k)0],\end{aligned} \tag{5}$$

and the change in energy satisfies the first law of thermodynamics:

$$\begin{aligned}\Delta E[\gamma_{\tau,0};\Lambda_t] &= E[z_\tau,B^N(\mu_N)] - E(z_0,A_0) \\ &= W[\gamma_{\tau,0};\Lambda_t] + Q[\gamma_{\tau,0};\Lambda_t],\end{aligned} \tag{6}$$

where $t_0=0$ and $\lambda_t^0(\mu_0)=\lambda_t^0$. Since the protocols depend on the measurement outcomes, the free-energy difference is realization dependent:

$$\Delta F[\Lambda_t] = F[\lambda_\tau^N(\mu_N)] - F[\lambda_0^0] = F[B^N(\mu_N)] - F(A_0). \tag{7}$$

Likewise, the dissipated work is

$$W_d[\gamma_{\tau,0};\Lambda_t] = W[\gamma_{\tau,0};\Lambda_t] - \Delta F[\Lambda_t]. \tag{8}$$

As discussed in Sec. II, we generate an ensemble of realizations of the reverse process by carrying out each observed realization of the forward process backward in time. Take, for example, the time reversal of a realization of the forward process with protocol $\Lambda_t(\mu_N)=\{\lambda_t^0,\ldots,\lambda_t^N(\mu_N)\}$. The system is first equilibrated at inverse temperature β with external parameters fixed at $\lambda_\tau^N(\mu_N)=B^N(\mu_N)$, so that the initial statistical state of the reverse process is $P^{eq}[z|B^N(\mu_N)]$ [Eq. (2)]. Then from time $t=0$ to τ the external parameters are varied according to the time-reversed individual protocols executed in the reverse order: for each time interval $t=\tau-t_{k+1}$ to $\tau-t_k$, with $k=0,\ldots,N$, the external parameters are varied according to the reverse individual protocol $\tilde{\lambda}_t^{N-k}(\mu_k)=\lambda_{\tau-t}^k(\mu_k)$. The *reverse complete protocol* $\tilde{\Lambda}_t=\Lambda_{\tau-t}$ is $\tilde{\Lambda}_t(\mu_N)=\{\tilde{\lambda}_t^0(\mu_N),\ldots,\tilde{\lambda}_t^N\}$. Observe that in an ensemble of realizations of the reverse process the probability to observe reverse complete protocol $\tilde{\Lambda}_t(\mu_N)$, $\tilde{\pi}[\tilde{\Lambda}_t(\mu_N)]$, is independent of the microscopic trajectory and is equal to the probability that the conjugate forward complete protocol $\Lambda_t(\mu_N)=\tilde{\Lambda}_{\tau-t}(\mu_N)$ occurs in the forward process, $\pi[\Lambda_t(\mu_N)]$:

$$\tilde{\pi}[\tilde{\Lambda}_t] = \pi[\Lambda_t] = \int d\gamma_{\tau,0}\mathcal{P}[\gamma_{\tau,0};\Lambda_t], \tag{9}$$

where $d\gamma_{\tau,0}$ is a measure on the space of microscopic trajectories. Moreover, due to the assumed one-to-one correspondence between measurement trajectories and protocols, the probability distribution $\pi[\Lambda_t(\mu_N)]$ is equal to the probability distribution of measurement trajectories in the forward process

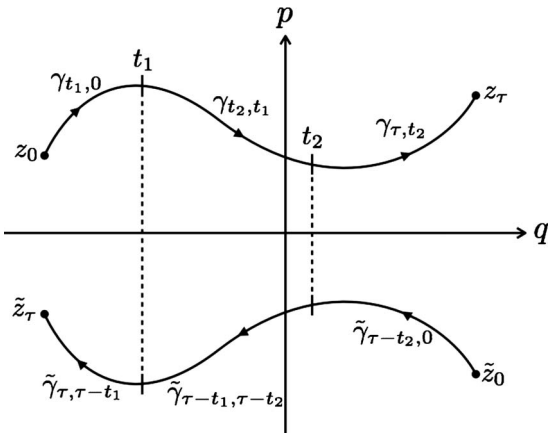

FIG. 2. Illustration of the forward trajectory $\gamma_{\tau,0}$ and reverse trajectory $\tilde{\gamma}_{\tau,0}$ with two feedback loops implemented at times t_1 and t_2.

$$P_N(\mu_N) = P_N(m_N|\mu_{N-1}) \cdots P_2(m_2|\mu_1)P_1(m_1), \tag{10}$$

where $P_k(m_k|\mu_{k-1})$ is the conditional probability to observe measurement outcome m_k in the forward process conditioned on the measurement trajectory μ_{k-1}, and the equality follows from the product rule of conditional probabilities [36]. Combining Eqs. (9) and (10), the probability to implement reverse complete protocol $\tilde{\Lambda}_t$ in the reverse process is

$$\tilde{\pi}[\tilde{\Lambda}_t(\mu_N)] = P_N(\mu_N). \tag{11}$$

For every trajectory from time $t=s$ to r, $\gamma_{r,s}=\{z_t\}_{t=s}^{r}$ there is a conjugate reverse trajectory $\tilde{\gamma}_{\tau-s,\tau-r}=\{\tilde{z}_t\}_{t=\tau-r}^{\tau-s}=\{z_t^*\}_{t=r}^{s}$, where $\tilde{z}_t=z_{\tau-t}^*$ (see Fig. 2).

The probability to observe reverse trajectory $\tilde{\gamma}_{\tau,0}$ and reverse complete protocol $\tilde{\Lambda}_t$ in the reverse process is

$$\tilde{\mathcal{P}}[\tilde{\gamma}_{\tau,0};\tilde{\Lambda}_t] = P[\tilde{\gamma}_{\tau,0}|\tilde{\Lambda}_t]\tilde{\pi}[\tilde{\Lambda}_t], \tag{12}$$

where $P[\tilde{\gamma}_{\tau,0}|\tilde{\Lambda}_t]$ is the conditional probability to observe $\tilde{\gamma}_{\tau,0}$ conditioned on executing protocol $\tilde{\Lambda}_t$. Substituting in Eqs. (10) and (11), and expanding $P[\tilde{\gamma}_{\tau,0}|\tilde{\Lambda}_t]$ in conditional probabilities using the product rule of conditional probabilities [36], allows us to express $\tilde{\mathcal{P}}$ as

$$\begin{aligned}\tilde{\mathcal{P}}[\tilde{\gamma}_{\tau,0};\tilde{\Lambda}_t] = {} & P[\tilde{\gamma}_{\tau,\tau-t_1}|\tilde{z}_{\tau-t_1},\tilde{\lambda}_t^N]P[\tilde{\gamma}_{\tau-t_1,\tau-t_2}|\tilde{z}_{\tau-t_2},\tilde{\lambda}_t^{N-1}(\mu_1)] \\ & \times P_1(\mu_1)\cdots P[\tilde{\gamma}_{\tau-t_N,0}|\tilde{z}_0,\tilde{\lambda}_t^0(\mu_N)]P_N(m_N|\mu_{N-1}) \\ & \times P^{eq}[\tilde{z}_0|B^N(\mu_N)].\end{aligned} \tag{13}$$

The structure of Eq. (12) [Eq. (13)] suggests an alternative method for implementing the reverse process. We randomly select a reverse protocol $\tilde{\Lambda}_t(\mu_N)$ according to the distribution $\tilde{\pi}[\tilde{\Lambda}_t(\mu_N)]$ [Eq. (11)]. Next, we equilibrate the system at inverse temperature β with external parameters fixed at $\tilde{\Lambda}_0(\mu_N)=B^N(\mu_N)$, drive the system away from equilibrium according to $\tilde{\Lambda}_t(\mu_N)$, and finally allow the system relax back to equilibrium at inverse temperature β with external parameters fixed at $\tilde{\Lambda}_\tau(\mu_N)=A^0$.

With this setup, we can now derive Eq. (1) as a consequence of the detailed fluctuation theorem [1,3,5,13–21],

$$\frac{P[\gamma_{t_{k+1},t_k}|z_{t_k},\lambda_t^k(\mu_k)]}{P[\tilde{\gamma}_{\tau-t_k,\tau-t_{k+1}}|\tilde{z}_{\tau-t_{k+1}},\tilde{\Lambda}_t^{N-k}(\mu_k)]} = e^{-\beta Q^k[\gamma_{t_{k+1},t_k};\lambda_t^k(\mu_k)]}, \tag{14}$$

where Q^k is defined in Eq. (5). Equation (14) has been derived for a wide class of dynamics and is a consequence of the time-reversal symmetry of the microscopic dynamics—the energy is time-reversal invariant [see the discussion preceding Eq. (2)].

To derive Eq. (1), we take the ratio of Eqs. (3) and (13), then substitute in Eqs. (2), (5), (6), (8), and (14), and the definition of the change in uncertainty:

$$I[\gamma_{\tau,0};\Lambda_t] = \ln\left[\frac{P_N(m_N|z_{t_N})\cdots P_2(m_2|z_{t_2})P_1(m_1|z_{t_1})}{P_N(m_N|\mu_{N-1})\cdots P_2(m_2|\mu_1)P_1(m_1)}\right], \tag{15}$$

which, after a short manipulation, leads to Eq. (1), reprinted here for convenience,

$$\frac{\mathcal{P}[\gamma_{\tau,0};\Lambda_t]}{\tilde{\mathcal{P}}[\tilde{\gamma}_{\tau,0};\tilde{\Lambda}_t]} = e^{\beta W_d[\gamma_{\tau,0};\Lambda_t]+I[\gamma_{\tau,0};\Lambda_t]}. \tag{16}$$

Equation (16) [Eq. (1)] is an extension of the detailed fluctuation theorem [Eq. (14)] for systems driven away from equilibrium by repeated discrete feedback. However, there is a fundamental difference between Eqs. (16) and (14) due to the inherent time-reversal asymmetry of feedback. Since no measurements are made in the reverse process, there are microscopic trajectories and reverse complete protocols whose time-reversed conjugates do *not* occur together in the forward process; that is, there exists a $\gamma_{\tau,0}$ with conjugate reverse trajectory $\tilde{\gamma}_{\tau,0}$, and $\Lambda_t=\tilde{\Lambda}_{\tau-t}$, such that $\mathcal{P}[\gamma_{\tau,0};\Lambda_t]=0$ and $\tilde{\mathcal{P}}[\tilde{\gamma}_{\tau,0};\tilde{\Lambda}_t]\neq 0$. Consequently, the ratio $\mathcal{P}/\tilde{\mathcal{P}}$, appearing in Eq. (16), is well defined, but the reciprocal $\tilde{\mathcal{P}}/\mathcal{P}$ is not well defined—mathematically, we say that $\mathcal{P}$ is absolutely continuous with respect to $\tilde{\mathcal{P}}$ ($\mathcal{P}\ll\tilde{\mathcal{P}}$) [37]; however, the reverse is not true. For example, consider a Hamiltonian system in which we implement feedback by making an error-free measurement at $t=0$ of whether the system is in a region of phase space δ. If the system is found in δ, we drive the system with external parameter protocol λ_t^δ. The region of phase space δ then evolves deterministically to the region of phase space δ', as illustrated in Fig. 3. In the reverse process the initial system state is sampled from a canonical distribution over all phase space. Consequently, when the protocol $\tilde{\lambda}_t^\delta$ is executed, the system may evolve along a trajectory $\tilde{\Gamma}$—the conjugate trajectory of Γ depicted in Fig. 3, which begins outside δ and terminates outside δ'. Clearly, the conjugate trajectory Γ can never be observed in the forward process simultaneously with λ_t^δ; $\mathcal{P}[\Gamma;\lambda_t^\delta]=0$, while $\tilde{\mathcal{P}}[\tilde{\Gamma};\tilde{\lambda}_t^\delta]\neq 0$.

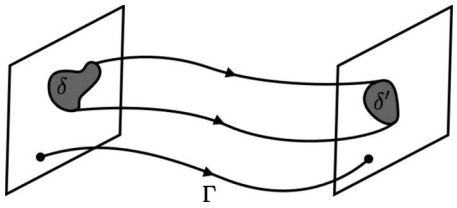

FIG. 3. Illustration of the tube of trajectories of the forward process evolving from phase-space region δ and to region δ' under Hamiltonian dynamics driven by external parameter protocol λ_t^{δ} associated to measuring the initial state of the system inside region δ. $\tilde{\Gamma}$, the conjugate trajectory of Γ, exemplifies a trajectory of the reverse process whose conjugate trajectory Γ cannot be realized in the forward process simultaneously with λ_t^{δ}, since Γ begins outside phase-space region δ.

IV. INTERPRETATION OF *I*

We have mentioned that I [Eq. (15)] quantifies a change in uncertainty about the microscopic state of the system upon making measurements. We now provide an argument supporting that assertion using methods of Bayesian inference. Our analysis begins by using Bayes' theorem [36] to rewrite Eq. (15) in terms of the conditional probability distributions $\rho(z_{t_k}|\mu_k)$ to find the system at z_{t_k} conditioned on the sequence of measurement outcomes in μ_k,

$$I[\gamma_{\tau,0};\Lambda_t(\mu_N)] = \sum_{k=1}^{N} \ln\left[\frac{\rho(z_{t_k}|\mu_k)}{\rho(z_{t_k}|\mu_{k-1})}\right]. \tag{17}$$

To interpret Eq. (17), recall that probability distributions measure the degree of belief a *rational* person has in the truth of a proposition, i.e., they quantify our uncertainty [36]. For example, as rational statistical physicists, our uncertainty in the state of our system at time t_1, just prior to the first measurement, is

$$\rho(z_{t_1}) = \int^{z_{t_1}} d\gamma_{t_1,0} P[\gamma_{t_1,0}|z_0,\lambda_t^0] P^{eq}(z_0|A_0). \tag{18}$$

Upon making a measurement, we gain information altering our beliefs and forcing us to update (or change) the probability distribution describing our state of knowledge about the system. For example, suppose that at time t_1 we measured M_1 and obtained outcome m_1. We have gained some information and as rational beings we *must* update our uncertainty $\rho(z_{t_1})$. Bayesian inference tells us that the new probability distribution describing our uncertainty—the posterior probability distribution—is obtained from Bayes' theorem and is simply $\rho(z_{t_1}|m_1)$, the conditional probability for the system to be at z_{t_1} given that the outcome of the measurement was m_1 [36]. Comparing with Eq. (17), we see that the $k=1$ term in the sum is $\ln[\rho(z_{t_1}|m_1)/\rho(z_{t_1})]$, the logarithm of the ratio of the probability distributions before and after the measurement; hence, it is a measure of how our uncertainty changes upon making a measurement. Repeating this argument, we find that each term in the sum in Eq. (17) represents a change in our uncertainty upon making each new measurement. Notice that I can be positive or negative in any given realization; our uncertainty can decrease or increase. However, the average of I over many realizations $\langle I\rangle$ is always positive [30], reflecting that on average gaining information lowers our uncertainty.

V. APPLICATIONS

Equation (1) immediately leads to two work relations [Eqs. (19) and (20) below]. It is a straightforward exercise using Eq. (1) to show that

$$\langle e^{-\beta W_d - I}\rangle = 1, \tag{19}$$

where the angular brackets denote an average of an ensemble of realizations of the forward process. Equation (19) is a generalization of the nonequilibrium work fluctuation relation of Sagawa and Ueda [33] for multiple feedback loops. Similarly, Eq. (1) implies a generalization of the relative-entropy work fluctuation relation [4,5]:

$$D[\mathcal{P}\,\|\,\tilde{\mathcal{P}}] = \beta\langle W_d\rangle + \langle I\rangle, \tag{20}$$

where $D(f\|g)=\int dx f(x)\ln[f(x)/g(x)]$ is the relative entropy, an information theoretic measure of the distinguishability of two probability distributions [30]. Furthermore, applying Jensen's inequality [30] to Eq. (19) or exploiting the positivity of the relative entropy [30] in Eq. (20), one finds that

$$\beta\langle W_d\rangle + \langle I\rangle \geq 0, \tag{21}$$

which can be viewed as a generalization of the second law of thermodynamics in the presence of feedback [33].

VI. CONCLUSION

For systems driven by repeated discrete feedback, we have introduced a framework of forward and reverse processes. We defined a reverse process in which the steps of the forward process are carried out backward in time. As a consequence, we found that the change in uncertainty I [Eq. (15)] during each feedback loop must be incorporated when analyzing the thermodynamics of feedback. I is a natural generalization to repeated discrete feedback of the information measure utilized by Sagawa and Ueda in Ref. [33]. Cao and Feito also observed that the ensemble average $\langle I\rangle$ naturally occurs in their thermodynamics of feedback [26]. These observations support the conclusion that analyzing feedback using the framework of forward and reverse processes developed here may be beneficial to understanding the thermodynamics of feedback. Exploiting Eq. (1), we generalized the detailed fluctuation theorem [Eq. (1)], the nonequilibrium work fluctuation theorem [Eq. (19)], and the relative-entropy work relation [Eq. (20)] to systems manipulated by repeated feedback.

The next step in understanding the thermodynamics of feedback is to incorporate feedback into the fluctuation relations [13–19,38–40] which are predictions about the fluctuations of thermodynamic quantities in far from equilibrium systems. A first step in this regard has already been taken by Kim and Qian [32], who analyzed the fluctuation relations in the presence of velocity-dependent feedback control.

While this paper was under consideration, another paper proposing a detailed fluctuation theorem in the presence of feedback was published [34]. Although the results are similar, our analysis contains a number of additional important elements not found in Ref. [34]. Our central result [Eq. (1)] applies to processes with multiple feedback loops, while Ref. [34] considers only a single loop. We also provide a detailed physical interpretation of the reverse process, including a description of the procedure for executing that process, and a discussion of the asymmetry between the forward and reverse processes. Finally, we give a physical interpretation for the change in uncertainty along a microscopic trajectory.

Moreover, we believe that the main conclusions [Eqs. (11) and (13)] of Ref. [34] suffer from physical inconsistencies. While Eq. (11) of Ref. [34] assumes that feedback is implemented in both the forward and reverse processes, the protocol employed in the reverse process is acausal: it is executed in response to a measurement made in the future (cf. Eq. (8) of Ref. [34]). Reference [34] also investigates forward and reverse processes identical to those discussed in the present paper. In this context Ref. [34] proposes a Crooks-type fluctuation relation for the joint distribution of dissipated work and change in uncertainty, $p(W_d, I)$. This result is problematic: the change in uncertainty in the reverse process is ill defined since no measurements are made in the reverse process [41].

ACKNOWLEDGMENTS

We are grateful to Chris Jarzynski for many stimulating discussions as well as to J.M.R. Parrondo for his helpful suggestions. We would also like to thank the anonymous referees for their insightful comments. J.M.H. was supported by the American Recovery and Reinvestment Act (ARRA) funds through Grant No. ECCS 0925365 from the National Science Foundation and by Grant MOSAICO (Spain). S.V. acknowledges support from the National Science Foundation under Grant No. CHE-0841557.

[1] G. N. Bochkov and Y. E. Kuzovlev, Zh. Eksp. Teor. Fiz. **72**, 238 (1977) [Sov. Phys. JETP **45**, 125 (1977)].
[2] C. Jarzynski, Phys. Rev. Lett. **78**, 2690 (1997).
[3] G. E. Crooks, J. Stat. Phys. **90**, 1481 (1998); Phys. Rev. E **61**, 2361 (2000).
[4] R. Kawai, J. M. R. Parrondo, and C. Van den Broeck, Phys. Rev. Lett. **98**, 080602 (2007).
[5] C. Jarzynski, Phys. Rev. E **73**, 046105 (2006).
[6] C. Jarzynski, Eur. Phys. J. B **64**, 331 (2008).
[7] A. Gomez-Marin, J. M. R. Parrondo, and C. Van den Broeck, EPL **82**, 50002 (2008).
[8] S. Vaikuntanathan and C. Jarzynski, EPL **87**, 60005 (2009).
[9] D. Collin, F. Ritort, C. Jarzynski, S. B. Smith, I. Tinoco, Jr., and C. Bustamante, Nature (London) **437**, 231 (2005).
[10] J. Liphardt, S. Dumont, S. B. Smith, I. Tinoco, Jr., and C. Bustamante, Science **296**, 1832 (2002).
[11] F. Douarche, S. Ciliberto, A. Petrosyan, and I. Rabbiosi, EPL **70**, 593 (2005).
[12] C. Chipot and A. Pohorille, *Free Energy Calculations* (Springer, Berlin, 2007).
[13] G. Gallavotti and E. G. D. Cohen, Phys. Rev. Lett. **74**, 2694 (1995).
[14] D. J. Evans and D. J. Searles, Adv. Phys. **51**, 1529 (2002).
[15] J. L. Lebowitz and H. Spohn, J. Stat. Phys. **95**, 333 (1999).
[16] T. Hatano and S. I. Sasa, Phys. Rev. Lett. **86**, 3463 (2001).
[17] C. Maes, Seminaire Poincare **2**, 29 (2003).
[18] U. Seifert, Phys. Rev. Lett. **95**, 040602 (2005).
[19] R. J. Harris and G. M. Schütz, J. Stat. Mech.: Theory Exp. (2007), P07020.
[20] G. N. Bochkov and Y. E. Kuzovlev, Physica A **106**, 443 (1981).
[21] R. D. Astumian, Phys. Rev. E **76**, 020102 (2007).
[22] J. Bechhoefer, Rev. Mod. Phys. **77**, 783 (2005).
[23] K. H. Kim and H. Qian, Phys. Rev. Lett. **93**, 120602 (2004).
[24] A. E. Allahverdyan and D. B. Saakian, EPL **81**, 30003 (2008).
[25] T. Sagawa and M. Ueda, Phys. Rev. Lett. **100**, 080403 (2008).
[26] F. J. Cao and M. Feito, Phys. Rev. E **79**, 041118 (2009).
[27] C. H. Bennett, in *Maxwell's Demon: Entropy, Information, Computing*, edited by H. S. Leff and A. F. Rex (Princeton University Press, Princeton, NJ, 1990).
[28] B. Piechocinska, Phys. Rev. A **61**, 062314 (2000).
[29] T. Sagawa and M. Ueda, Phys. Rev. Lett. **102**, 250602 (2009).
[30] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. (Wiley-Interscience, Hoboken, NJ, 2006).
[31] H. Touchette and S. Lloyd, Phys. Rev. Lett. **84**, 1156 (2000).
[32] K. H. Kim and H. Qian, Phys. Rev. E **75**, 022102 (2007).
[33] T. Sagawa and M. Ueda, Phys. Rev. Lett. **104**, 090602 (2010).
[34] M. Ponmurugan, Phys. Rev. E **82**, 031129 (2010).
[35] L. Szilard, in *Maxwell's Demon: Entropy, Information, Computing*, edited by H. S. Leff and A. F. Rex (Princeton University Press, Princeton, NJ, 1990).
[36] A. Caticha, e-print arXiv:0808.0012.
[37] L. B. Koralov and Y. G. Sinai, *Theory of Probability and Random Processes* (Springer-Verlag, Berlin, 2007).
[38] T. Speck and U. Seifert, J. Phys. A **38**, L581 (2005).
[39] J. Kurchan, J. Phys. A **31**, 3719 (1998).
[40] G. E. Crooks, Phys. Rev. E **60**, 2721 (1999).
[41] Anonymous referee (private communication).