Information thermodynamics for a multi-feedback process with time delay

2017 EPL 117 10011

(http://iopscience.iop.org/0295-5075/117/1/10011)

A LETTERS JOURNAL EXPLORING
THE FRONTIERS OF PHYSICS

January 2017

EPL, **117** (2017) 10011
doi: 10.1209/0295-5075/117/10011

www.epljournal.org

Information thermodynamics for a multi-feedback process with time delay

CHULAN KWON[1,2], JAEGON UM[2,4] and HYUNGGYU PARK[2,3]

[1] *Department of Physics, Myongji University - Yongin, Gyeonggi-Do 17058, Korea*
[2] *Quantum Universe Center, Korea Institute for Advanced Study - Seoul 02455, Korea*
[3] *School of Physics, Korea Institute for Advanced Study - Seoul 02455, Korea*
[4] *CCSS, CTP and Department of Physics and Astronomy, Seoul National University - Seoul 08826, Korea*

received 16 June 2016; accepted in final form 27 January 2017
published online 2 March 2017

PACS 05.40.-a – Fluctuation phenomena, random processes, noise, and Brownian motion
PACS 05.10.Gg – Stochastic analysis methods (Fokker-Planck, Langevin, etc.)

Abstract – We investigate a measurement-feedback process of repeated operations with time delay. During a finite-time interval, measurement on the system is performed and the feedback protocol derived from the measurement outcome is applied with time delay. This protocol is maintained into the next interval until a new protocol from the next measurement is applied. Unlike a feedback process without delay, both memories associated with previous and present measurement outcomes are involved in the system dynamics, which naturally brings forth a joint system described by a system state and two memory states. The thermodynamic second law provides a lower bound for heat flow into a thermal reservoir by the (3-state) Shannon entropy change of the joint system. However, as the feedback protocol depends on memory states sequentially, we can deduce a tighter bound for heat flow by integrating out irrelevant memory states during dynamics. As a simple example, we consider the so-called cold damping feedback process where the velocity of a particle is measured and a dissipative feedback protocol is applied to decelerate the particle. We confirm that the heat flow is well above the tightest bound. We also examine the long-time limit of this feedback process, which turns out to exhibit an interesting instability transition as well as heating by controlling parameters such as measurement errors, time interval, protocol strength, and time delay length. We discuss the underlying mechanism for instability and heating, which might be unavoidable in reality.

The recent information thermodynamics has been proven to resolve the paradox of Maxwell's demon [1] which was a long-lived problem in spite of enormous research works [1–5]. Replacing Maxwell's demon by a physical memory device, that was refined by Landauer [4], one is able to describe measurement inside a memory device and feedback after measurement acting on the system (engine) as thermodynamic processes. In the measurement process, information acquisition is realized as mutual information gain in the entropy of the joint system (system and memory device). In the subsequent feedback process, mutual information is expended through relaxation out of the initial state producing work outside. The work production is balanced energetically by heat dissipation into the reservoir, which may be negative like in the Szilard engine [2], resulting in entropy loss in the reservoir.

It was shown that such entropy loss in the reservoir, if any, be compensated sufficiently by the entropy gain of the joint system through mutual information decrease so as to satisfy the second law of thermodynamics. Hence the paradox of Maxwell's demon is resolved. It is the main feature of the information thermodynamics developed by Sagawa and Ueda [6–9]. The increase of the total entropy of the joint system and reservoir was proven with the aid of the fluctuation theorem (FT), which was discovered about two decades ago and has been regarded as a principle of nonequilibrium statistical mechanics [10–14]. The role of mutual information in feedback processes has also been confirmed in experiments [15,16]. Measurement and feedback can reasonably be regarded as processes with separated time periods [9,17] and the fluctuation theorem for the total entropy production was shown to hold separately for the two bipartite periods [18].

Recent studies of repeated feedback processes showed a generalized fluctuation theorem for work modified by mutual information [19,20]. The cold-damping problem [21,22] was revisited from the viewpoint of the feedback process in the continuous measurement limit [23]

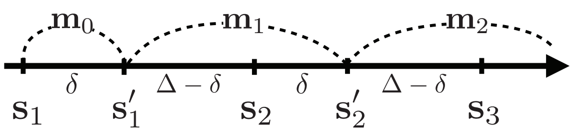

Fig. 1: A schematic picture for repeated measurement-feedback processes. $\mathbf{m}_i$ is a measurement outcome for an initial state $\mathbf{s}_i$ of step i, which is applied as a protocol with time delay δ. This protocol is maintained into the next step until the next protocol is applied. $\mathbf{m}_0 = \mathbf{0}$ when there is no previous measurement.

where the memory state is considered to evolve fluctuating around the system state. Time-delayed feedback was also studied in a simple situation where a feedback protocol is given by a past state or the history of past states weighted by the memory kernel [24,25] in the continuous and exact measurement limit.

In the present work, we consider a realistic feedback process, combining and generalizing frameworks previously studied [19,20,23–25], which is composed of multiple steps repeated in a finite-time interval, in each of which a feedback protocol from erratic measurement is applied with time delay. As an example, we revisit the cold-damping problem, which becomes highly complicated in this general setup. Beyond the verification of a generalized second law, we study important characteristics applicable to other feedback models and real experiments such as the optimal mutual information for the lowest bound of entropy production among various possible choices for coexisting memories due to time delay, instability of repeated feedback, phase diagram in parameter space, overshooting, etc.

Consider that both system state $\mathbf{s}(t)$ and memory state $\mathbf{m}(t)$ in d dimensions coevolve in time t by their own dynamics. At measurement time $t = t_i$, memory starts to measure or copy $\mathbf{s}(t_i) = \mathbf{s}_i$ that acts as a protocol to drive memory into a copied state. One may think of the Langevin dynamics for such a process: $\dot{\mathbf{m}} = -\tau_{\rm m}^{-1}(\mathbf{m} - \mathbf{s}_i) + \boldsymbol{\xi}(t)$ where $\langle \xi_a(t)\xi_b(t')\rangle = 2\tau_{\rm m}^{-1}\sigma\delta_{ab}\delta(t-t')$ with component indices $a, b = 1, \ldots, d$ and temperature $T_{\rm m} = \tau_{\rm m}^{-1}\sigma$ of the reservoir surrounding memory. The Boltzmann constant is set to unity here and also in the following. Waiting for a long enough time δ compared to relaxation time $\tau_{\rm m}$, the memory reaches a local equilibrium with the probability density function (PDF) given as

$$p_{\rm eq}(\mathbf{m}_i) = (2\pi\sigma)^{-d/2} e^{-(\mathbf{m}_i - \mathbf{s}_i)^2/(2\sigma)}. \tag{1}$$

For this period, the system undergoes a transition to state $\mathbf{s}'_i$ at $t = t_i + \delta$ under a previous protocol $\mathbf{m}_{i-1}$. A new protocol $\mathbf{m}_i$ chosen from the distribution $p_{\rm eq}(\mathbf{m}_i)$ is applied, in turn, to the dynamics of the system for $t_i + \delta < t < t_{i+1} = t_i + \Delta$. In fig. 1, the corresponding path of $\mathbf{s}(t)$ is shown with $\mathbf{m}_{i-1}$ and $\mathbf{m}_i$ coexisting in step i.

We introduce an adjoint dynamics with time-reverse protocols in which the probability of the system tracing the time-reverse path conjugate to a given (forward) path will be considered. The time-reversed path is defined as $\bar{\mathbf{s}}(t) = \varepsilon\mathbf{s}(t_N + t_1 - t)$ conjugate to a (forward) path $\mathbf{s}(t)$, where ε is the parity operator giving $+1$ (-1) if it is applied to an even (odd) parity state in time reversal such as position (momentum). The time-reverse protocols are defined as $\bar{\mathbf{m}}_i = \varepsilon\mathbf{m}_{N-i}$. For each of time-reverse protocols, the parity is multiplied, copying a time-reverse state.

Let $\Pi^{\mathbf{m}_{i-1}}_{\mathbf{s}_i,\mathbf{s}'_i}[\mathbf{s}(t)]$ $(\Pi^{\mathbf{m}_i}_{\mathbf{s}'_i,\mathbf{s}_{i+1}}[\mathbf{s}(t)])$ be the conditional probability for a partial path from $\mathbf{s}_i$ $(\mathbf{s}'_i)$ to $\mathbf{s}'_i$ $(\mathbf{s}_{i+1})$ under a protocol $\mathbf{m}_{i-1}$ $(\mathbf{m}_i)$ for $t_i \le t < t_i + \delta$ $(t_i + \delta \le t < t_{i+1})$ in step i. Similarly, we define the conditional path probabilities for time-reverse paths and protocols as $\Pi^{\bar{\mathbf{m}}_i}_{\bar{\mathbf{s}}_i,\bar{\mathbf{s}}'_i}[\bar{\mathbf{s}}(t)]$ and $\Pi^{\bar{\mathbf{m}}_{i+1}}_{\bar{\mathbf{s}}'_i,\bar{\mathbf{s}}_{i+1}}[\bar{\mathbf{s}}(t)]$ for $t_i \le t \le t_i + (\Delta - \delta)$ and $t_i + (\Delta - \delta) \le t \le t_{i+1}$, respectively. For the usual thermodynamic process without feedback, the change in the total entropy of system and reservoir is known as the log-ratio of the path probabilities of the forward and time-reverse path. Extending to the joint system of system and memory, the corresponding *total entropy* change may be written as

$$\begin{aligned}\sum_{i=1}^{N-1} \Delta S_{{\rm tot},i} &= \prod_{i=1}^{N-1} \ln\left[\frac{\rho_i(\mathbf{s}_i)\rho_i(\mathbf{m}_{i-1}|\mathbf{s}_i)\rho_i(\mathbf{m}_i|\mathbf{s}_i)}{\rho_{i+1}(\mathbf{s}_{i+1})\bar{\rho}(\varepsilon\mathbf{m}_i, \varepsilon\mathbf{m}_{i-1}|\bar{\mathbf{s}}(t))}\right. \\ &\qquad\left.\times\frac{\Pi^{\mathbf{m}_{i-1}}_{\mathbf{s}_i,\mathbf{s}'_i}[\mathbf{s}(t)]\Pi^{\mathbf{m}_i}_{\mathbf{s}'_i,\mathbf{s}_{i+1}}[\mathbf{s}(t)]}{\Pi^{\varepsilon\mathbf{m}_i}_{\varepsilon\mathbf{s}_{i+1},\varepsilon\mathbf{s}'_i}[\bar{\mathbf{s}}(t)]\Pi^{\varepsilon\mathbf{m}_{i-1}}_{\varepsilon\mathbf{s}'_i,\varepsilon\mathbf{s}_i}[\bar{\mathbf{s}}(t)]}\right] \\ &= \sum_{i=1}^{N-1}\left[\Delta S_{{\rm sm},i} + \Delta S_{{\rm env},i}\right],\end{aligned} \tag{2}$$

where $\Delta S_{{\rm tot},i}$ denotes the contribution from step i and ρ_i denotes a PDF at $t = t_i$, obtained from the given dynamics. $\rho_i(\mathbf{m}_i|\mathbf{s}_i)$ is the measurement probability of outcome $\mathbf{m}_i$, which is equal to the local equilibrium distribution $p_{\rm eq}(\mathbf{m}_i)$ in eq. (1). A conditional probability $\bar{\rho}$ for time-reverse protocols in the adjoint dynamics can be chosen in various ways, which will be discussed later.

The environmental entropy production for step i is defined as

$$\Delta S_{{\rm env},i} = \ln\left[\frac{\Pi^{\mathbf{m}_{i-1}}_{\mathbf{s}_i,\mathbf{s}'_i}[\mathbf{s}(t)]\Pi^{\mathbf{m}_i}_{\mathbf{s}'_i,\mathbf{s}_{i+1}}[\mathbf{s}(t)]}{\Pi^{\varepsilon\mathbf{m}_i}_{\varepsilon\mathbf{s}_{i+1},\varepsilon\mathbf{s}'_i}[\bar{\mathbf{s}}(t)]\Pi^{\varepsilon\mathbf{m}_{i-1}}_{\varepsilon\mathbf{s}'_i,\varepsilon\mathbf{s}_i}[\bar{\mathbf{s}}(t)]}\right]. \tag{3}$$

In the absence of odd-parity states, $\Delta S_{{\rm env},i}$ is equal to Q_i/T for heat production Q_i into the reservoir at temperature T. However, it may contain an unconventional contribution due to an odd-parity force induced by an odd-parity protocol [26]. We will encounter this situation for a cold-damping problem where the velocity of a particle is measured.

$\Delta S_{{\rm sm},i}$ is the entropy change of the joint system for step i, which reads $\Delta S_{{\rm sys},i} - \Delta I_i$. Here ΔI_i is the mutual information change between system and memory. Note that the memory state does not change during each step. We find the first term to be the Shannon entropy change of the system, given as

$$\Delta S_{\text{sys},i} = -[\ln \rho_{i+1}(\mathbf{s}_{i+1}) - \ln \rho_i(\mathbf{s}_i)], \tag{4}$$

resulting from choosing the initial PDF of the time-reverse dynamics to be the final PDF $\rho_{i+1}(\mathbf{s}_{i+1})$ of the given dynamics.

Mutual information is a measure for the correlation between system and memory states. Due to time delay, there are two memories coexisting in each step. There are three ways for examining the correlation: i) correlation by joint memory $\{\mathbf{m}_{i-1}, \mathbf{m}_i\}$ for the whole period, ii) correlation by joint memory for time delay and correlation by new memory for the remaining period, iii) correlation by old memory for time delay and correlation by new memory for the remaining period. Mathematically the three choices depend on how $\bar{\rho}(\varepsilon\mathbf{m}_i, \varepsilon\mathbf{m}_{i-1}|\bar{\mathbf{s}}(t))$ is expressed in the time-reverse dynamics.

The first choice is given by

$$\begin{aligned}\bar{\rho}(\varepsilon\mathbf{m}_i, \varepsilon\mathbf{m}_{i-1}|\bar{\mathbf{s}}(t)) &= \rho_{i+1}(\mathbf{s}_{i+1}, \mathbf{m}_{i-1}, \mathbf{m}_i)/\rho_{i+1}(\mathbf{s}_{i+1}) \\ &\equiv \rho_{i+1}(\mathbf{m}_{i-1}, \mathbf{m}_i|\mathbf{s}_{i+1}) \ ,\end{aligned} \tag{5}$$

which is the conditional PDF of the joint system at time t_{i+1} for the given dynamics. One can define mutual information I_i at time t_i between state $\mathbf{s}_i$ and coexisting memories $\{\mathbf{m}_{i-1}, \mathbf{m}_i\}$ as $\ln[\rho_i(\mathbf{m}_{i-1}, \mathbf{m}_i|\mathbf{s}_i)/\rho_i(\mathbf{m}_{i-1}, \mathbf{m}_i)]$, where the conditional probability is given by $\rho_i(\mathbf{m}_{i-1}, \mathbf{m}_i|\mathbf{s}_i) = \rho_i(\mathbf{s}_i, \mathbf{m}_{i-1})\rho_i(\mathbf{m}_i|\mathbf{s}_i)/\rho_i(\mathbf{s}_i) = \rho_i(\mathbf{m}_{i-1}|\mathbf{s}_i)\rho_i(\mathbf{m}_i|\mathbf{s}_i)$. Similarly, I_{i+1} can be defined using eq. (5). Noting that the PDF $\rho_i(\mathbf{m}_{i-1}, \mathbf{m}_i)$ of the two memories are fixed in step i, we get mutual information change of the first kind as

$$\Delta I_i^{(1)} = \ln \frac{\rho_{i+1}(\mathbf{m}_{i-1}, \mathbf{m}_i|\mathbf{s}_{i+1})}{\rho_i(\mathbf{m}_{i-1}|\mathbf{s}_i)\rho_i(\mathbf{m}_i|\mathbf{s}_i)} \ . \tag{6}$$

The second choice is given by

$$\bar{\rho}(\varepsilon\mathbf{m}_i, \varepsilon\mathbf{m}_{i-1}|\bar{\mathbf{s}}(t)) = \rho_{i+1}(\mathbf{m}_i|\mathbf{s}_{i+1})\rho_{i'}(\mathbf{m}_{i-1}|\mathbf{s}'_i, \mathbf{m}_i), \tag{7}$$

where the first (second) factor determines the distribution of $\varepsilon\mathbf{m}_i$ ($\varepsilon\mathbf{m}_{i-1}$) for the period $\Delta - \delta$ (δ) of step $N - i$ in the time-reverse dynamics. Then, we have

$$\Delta I_i^{(2)} = \ln \frac{\rho_{i'}(\mathbf{m}_{i-1}, \mathbf{m}_i|\mathbf{s}'_i)}{\rho_i(\mathbf{m}_{i-1}|\mathbf{s}_i)\rho_i(\mathbf{m}_i|\mathbf{s}_i)} + \ln \frac{\rho_{i+1}(\mathbf{m}_i|\mathbf{s}_{i+1})}{\rho_{i'}(\mathbf{m}_i|\mathbf{s}'_i)}, \tag{8}$$

where $\rho_{i'}$ is the PDF at $t = t_i + \delta$ and $\rho_{i'}(\mathbf{m}_{i-1}, \mathbf{m}_i|\mathbf{s}'_i)/\rho_{i'}(\mathbf{m}_i|\mathbf{s}'_i) = \rho_{i'}(\mathbf{m}_{i-1}|\mathbf{s}'_i, \mathbf{m}_i)$ is used. The first term is the change in mutual information between system and two-state memory coexisting in the delay period, and the second is that between system and new memory in the remaining period. Writing $\Delta S_{\text{tot},i}^{(1,2)} = \Delta S_{\text{sm},i}^{(1,2)} + \Delta S_{\text{env},i}$ with $\Delta S_{\text{sm},i}^{(1,2)} = \Delta S_{\text{sys},i} - \Delta I_i^{(1,2)}$. We can show that both satisfy the integral FT (IFT) such that $\langle e^{-\sum_i \Delta S_{\text{tot},i}^{(1,2)}}\rangle = 1$ and also $\langle e^{-\Delta S_{\text{tot},i}^{(1,2)}}\rangle = 1$, leading to the inequality $\langle \Delta S_{\text{tot},i}^{(1,2)}\rangle \geq 0$, the generalized thermodynamic second law.

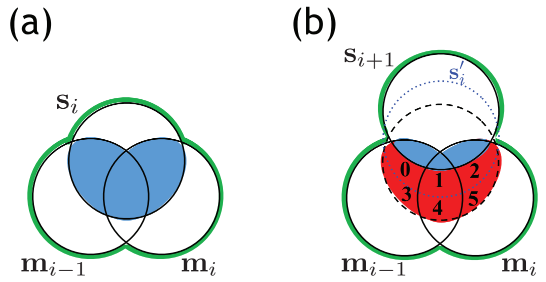

Fig. 2: (Color online) Venn diagrams for Shannon entropies (discs) and mutual information (intersections). $I_i^{(1)}$ is presented by blue areas in (a) at $t = t_i$ and in (b) at $t = t_{i+1}$. Panel (b) presents that the initial state $\mathbf{s}_i$ evolves to $\mathbf{s}'_i$ and subsequently to $\mathbf{s}_{i+1}$. $-\Delta I_i^{(1)}$ is represented by the whole red area, $-\Delta I_i^{(2)}$ by the areas labeled by 1, 2, 3, 4, 5, and $-\Delta I_i^{(3)}$ by those labeled by 1, 2, 3, 4. In this figure, we assume that the correlation between system and memory is decreasing in time. A general proof is given below eq. (12).

The third choice is given by replacing $\bar{\rho}(\varepsilon\mathbf{m}_i, \varepsilon\mathbf{m}_{i-1}|\bar{\mathbf{s}}(t))$ with

$$\rho_{i+1}(\mathbf{m}_i|\mathbf{s}_{i+1})\frac{\rho_{i'}(\mathbf{m}_{i-1}|\mathbf{s}'_i)}{\rho_{i'}(\mathbf{m}_i|\mathbf{s}'_i)}\rho_i(\mathbf{m}_i|\mathbf{s}_i), \tag{9}$$

which is no longer the conditional probability required to make the adjoint dynamics stochastic. Therefore, the IFT does not hold, $\langle e^{-\Delta S_{\text{tot},i}^{(3)}}\rangle \neq 1$. However, we can divide $\Delta S_{\text{tot},i}$ into two parts:

$$\Delta S_{\text{tot},i}^{\delta(3)} = \ln \left[\frac{\rho_i(\mathbf{s}_i)\rho_i(\mathbf{m}_{i-1}|\mathbf{s}_i)\Pi_{\mathbf{s}_i,\mathbf{s}'_i}^{\mathbf{m}_{i-1}}[\mathbf{s}(t)]}{\rho_{i'}(\mathbf{s}'_i)\rho_{i'}(\mathbf{m}_{i-1}|\mathbf{s}'_i)\Pi_{\varepsilon\mathbf{s}'_i,\varepsilon\mathbf{s}_i}^{\varepsilon\mathbf{m}_{i-1}}[\bar{\mathbf{s}}(t)]}\right], \tag{10}$$

$$\Delta S_{\text{tot},i}^{\Delta-\delta(3)} = \ln \left[\frac{\rho_{i'}(\mathbf{s}'_i)\rho_{i'}(\mathbf{m}_i|\mathbf{s}'_i)\Pi_{\mathbf{s}'_i,\mathbf{s}_{i+1}}^{\mathbf{m}_i}[\mathbf{s}(t)]}{\rho_{i+1}(\mathbf{s}_{i+1})\rho_{i+1}(\mathbf{m}_i|\mathbf{s}_{i+1})\Pi_{\varepsilon\mathbf{s}_{i+1},\varepsilon\mathbf{s}'_i}^{\varepsilon\mathbf{m}_i}[\bar{\mathbf{s}}(t)]}\right], \tag{11}$$

which are defined for $t_i \leq t \leq t_i + \delta$ and $t_i + \delta \leq t \leq t_{i+1}$, respectively. Then, the IFT can be shown to hold separately for the two as $\langle e^{-\Delta S_{\text{tot},i}^{\delta(3)}}\rangle = 1$ and $\langle e^{-\Delta S_{\text{tot},i}^{\Delta-\delta(3)}}\rangle = 1$. As a result, the inequality still holds for the sum of the two, which is the total entropy change $\Delta S_{\text{tot},i}^{(3)}$ of the third kind having $\langle \Delta S_{\text{tot},i}^{(3)}\rangle \geq 0$. We can similarly write the Shannon entropy change of the joint system as $\Delta S_{\text{sm},i}^{(3)} = \Delta S_{\text{sys},i} - \Delta I_i^{(3)}$ where

$$\Delta I_i^{(3)} = \ln \frac{\rho_{i'}(\mathbf{m}_{i-1}|\mathbf{s}'_i)}{\rho_i(\mathbf{m}_{i-1}|\mathbf{s}_i)} + \ln \frac{\rho_{i+1}(\mathbf{m}_i|\mathbf{s}_{i+1})}{\rho_{i'}(\mathbf{m}_i|\mathbf{s}'_i)}. \tag{12}$$

As presented in fig. 2, $-\Delta I_i^{(3)}$ is found to have the lowest bound to the change in total entropy among the three representations. It can be shown rigorously by using the technique for the IFT: $\langle e^{-R}\rangle = 1 \to \langle R\rangle \geq 1$. For

$t_i \le t \le t_i+\delta$, we get

$$\left\langle e^{\Delta I_i^{\delta(1,2)}-\Delta I_i^{\delta(3)}}\right\rangle = \left\langle \frac{\rho_{i'}(\mathbf{m}_{i-1},\mathbf{m}_i|\mathbf{s}'_i)}{\rho_i(\mathbf{m}_{i-1}|\mathbf{s}_i)p_{\rm m}(\mathbf{m}_i|\mathbf{s}_i)}\cdot\frac{\rho_i(\mathbf{m}_{i-1}|\mathbf{s}_i)}{\rho_{i'}(\mathbf{m}_{i-1}|\mathbf{s}'_i)}\right\rangle = 1. \tag{13}$$

For $t_i+\delta \le t \le t_i+\Delta$, we get

$$\left\langle e^{\Delta I_i^{\Delta-\delta(1)}-\Delta I_i^{\Delta-\delta(2,3)}}\right\rangle = \left\langle \frac{\rho_{i+1}(\mathbf{m}_{i-1},\mathbf{m}_i|\mathbf{s}_{i+1})}{\rho_{i'}(\mathbf{m}_{i-1},\mathbf{m}_i|\mathbf{s}'_i)}\cdot\frac{\rho_{i'}(\mathbf{m}_i|\mathbf{s}'_i)}{\rho_{i+1}(\mathbf{m}_i|\mathbf{s}_{i+1})}\right\rangle = 1. \tag{14}$$

In the above equations, the brackets denote the averages by path probabilities for the two periods[1]. Hence $-\Delta I_i^{(3)} = -\Delta I_i^{\delta(3)} - \Delta I_i^{\Delta\delta(3)}$ is the lowest.

One can say that the total entropy change is overestimated as mutual information between system and protocol having no influence on the dynamics is considered. Overestimated is mutual information due to a new protocol in time delay and that due to a past protocol in the subsequent period, labeled by 0 and 5 in the figure, respectively.

We apply our theory to a cold-damping problem where a feedback force is applied in the opposite direction to the measured velocity [21,22,27]. From now on, we investigate the problem within a single step, say for $t_1 \le t \le t_2$. We consider the one-dimensional motion of a particle described by the Langevin equation for the velocity v,

$$\dot v = -\gamma v - \tilde\gamma y_i + \xi(t), \tag{15}$$

where the mass is set to unity. Then, $\mathbf{s}=v$ and $\mathbf{m}_i=y_i$ where $i=0$ ($i=1$) denotes the past (new) protocol. y_0 is applied for $t_1 \le t \le t_1+\delta$ and y_1 for the remaining period. This feedback process can be realized in experiment for a colloidal particle with charge q where $\tilde\gamma$ is a control parameter for an electric field $E=\tilde\gamma y/q$. ξ is a usual stochastic force with mean zero and variance $\langle\xi(t)\xi(t')\rangle = 2T\gamma\delta(t-t')$. $\tilde\gamma>0$ is used for the purpose of cold damping.

One can find various PDFs and moments recursively given the initial PDF $\rho_1(v_1,y_0)$ with initial moments,

$$T_1=\langle v_1^2\rangle,\quad P_1=\langle y_0^2\rangle,\quad R_1=\langle v_1y_0\rangle. \tag{16}$$

It is convenient to consider composite states at $t=t_1$ and $t=t_2$, given as $\mathbf{c}_1=(v_1,y_0,y_1)$ and $\mathbf{c}_2=(v_2,y_0,y_1)$. Then, $\rho_1(\mathbf{c}_1)$ is equal to the product of $\rho_1(v_1,y_0)$ and $\rho_1(y_1|v_1)=(2\pi\sigma)^{-1/2}e^{-(y_1-v_1)^2/(2\sigma)}$. The Onsager-Machlup theory [28] gives the conditional probability for path $v(t)$ from $v(\tau)=v$ to $v(\tau')=v'$ as $\Pi^{y_i}_{v,v'}[v(t)] \propto \exp[-(4\gamma T)^{-1}\int_\tau^{\tau'}{\rm d}t(\dot u+\gamma u)^2]$, where $u(t)=v(t)+(\tilde\gamma/\gamma)y_i$. Then, the path integral of $\Pi^{y_0}_{v_1,v'_1}[v(t)]\Pi^{y_1}_{v'_1,v_2}[v(t)]$ over all paths gives rise to the transition probability of $v(t_2)=v_2$ given a composite state $\mathbf{c}_1$. We find

$$\rho(v_2|\mathbf{c}_1) = \frac{1}{\sqrt{2\pi w_\Delta}}e^{-\left(v_2-e^{-\gamma\Delta}v_1+(\tilde\gamma/\gamma)f\right)^2/(2w_\Delta)}, \tag{17}$$

where $w_\Delta = T(1-e^{-2\gamma\Delta})$ and

$$f=\left(e^{-\gamma(\Delta-\delta)}-e^{-\gamma\Delta}\right)y_0+\left(1-e^{-\gamma(\Delta-\delta)}\right)y_1. \tag{18}$$

Using this, the PDF of $\mathbf{c}_2$ is given as

$$\rho_2(\mathbf{c}_2)=\int{\rm d}v_1\,\rho_1(\mathbf{c}_1)\rho(v_2|\mathbf{c}_1)=\sqrt{\frac{\det\mathsf{D}_2}{(2\pi)^3}}e^{-\mathbf{c}_2\mathsf{D}_2\mathbf{c}_2^{\rm t}/2}, \tag{19}$$

where the superscript "t" denotes the transpose. Using the property of multi-variate Gaussian integral, the inversion of the matrix D_2 yields six moments such that

$$\mathsf{D}_2^{-1}=\begin{pmatrix}\langle v_2^2\rangle & \langle v_2y_0\rangle & \langle v_2y_1\rangle\\ \langle v_2y_0\rangle & \langle y_0^2\rangle & \langle y_0y_1\rangle\\ \langle v_2y_1\rangle & \langle y_0y_1\rangle & \langle y_1^2\rangle\end{pmatrix}, \tag{20}$$

which can be found in terms of T_1, P_1, and R_1 given in eq. (16).

In particular, $T_2=\langle v_2^2\rangle$, $P_2=\langle y_1^2\rangle$, and $R_2=\langle v_2y_1\rangle$ are found to satisfy the linear recursion relation:

$$\begin{aligned}T_2&=w_\Delta+\sigma h^2+K^2T_1+L^2P_1-2KLR_1,\\ P_2&=\sigma+T_1,\\ R_2&=-\sigma H+KT_1-LR_1,\end{aligned} \tag{21}$$

where $K=e^{-\gamma\Delta}-H$ with $H=(\tilde\gamma/\gamma)(1-e^{-\gamma(\Delta-\delta)})$ and $L=(\tilde\gamma/\gamma)e^{-\gamma\Delta}(e^{\gamma\delta}-1)$. The recursion relation can be rewritten as $\mathbf{Z}_2=\mathsf{G}\cdot\mathbf{Z}_1+\mathbf{A}$ for $\mathbf{Z}_i=(T_i,P_i,R_i)^{\rm t}$ where the matrix G and the vector $\mathbf{A}$ are given from eq. (21). $T_i=\langle v_i^2\rangle$ is defined as the effective temperature at $t=t_i$ and is updated through feedback steps as $T_1\to T_2\to T_3\to\cdots$. The recursion relation will lead to a fixed value T_∞ only if $|\lambda_a|<1$ for eigenvalues λ_a of G for $a=1,2,3$. The average effective temperature at step i can be found as $T_i^{\rm av}=\Delta^{-1}\int_{t_i}^{t_{i+1}}{\rm d}t\langle v(t)^2\rangle$. Cold damping will be successful if $T_\infty^{\rm av}<T$. In fig. 3, C (cold) stands for the region for $T_\infty^{\rm av}<T$, W (warm) for $T_\infty^{\rm av}>T$, and I (instability) for the instability region with $|\lambda_a|\ge 1$.

We can compute the parts of the total entropy change. The Shannon entropy for $\rho_2(\mathbf{c}_2)$ in eq. (19) can be written as $-\langle\ln\rho_2(\mathbf{c}_2)\rangle = (1/2)(-\ln\det\mathsf{D}_2+3\ln 2\pi+3)$, and similarly for $\rho_1(\mathbf{c}_1)$. Then, we obtain $\langle\Delta S^{(1)}_{\rm sm}\rangle = \langle\ln[\rho_1(\mathbf{c}_1)/\rho_2(\mathbf{c}_2)]\rangle$. By integrating $\rho_t(v(t),y_0,y_1)$ over y_0 or y_1, one can find $\rho_t(v(t),y_i)$. Then, we find

$$\left\langle\ln\frac{\rho_1(v_1,y_0)}{\rho_{1'}(v'_1,y_0)}\right\rangle=\frac12\ln\left[e^{-2\gamma\delta}+\frac{w_\delta P_1}{T_1P_1-R_1^2}\right], \tag{22}$$

and

$$\left\langle\ln\frac{\rho_{1'}(v'_1,y_1)}{\rho_2(v_2,y_1)}\right\rangle=\frac12\ln\Big[e^{-2\gamma(\Delta-\delta)}+\frac{w_{\Delta-\delta}(T_1+\sigma)}{w_\delta T_1+\sigma\langle v_1'^2\rangle+H_\delta^2(T_1P_1-R_1^2)}\Big], \tag{23}$$

[1]The path probability for time delay is given by $\rho_i(\mathbf{s}_i,\mathbf{m}_{i-1})\rho_i(\mathbf{m}_i|\mathbf{s}_i)\Pi^{\mathbf{m}_{i-1}}_{\mathbf{s}_i,\mathbf{s}'_i}[\mathbf{s}(t)]$ and that for the remaining period by $\rho_{i'}(\mathbf{s}'_i,\mathbf{m}_{i-1},\mathbf{m}_i)\Pi^{\mathbf{m}_i}_{\mathbf{s}'_i,\mathbf{s}_{i+1}}[\mathbf{s}(t)]$.

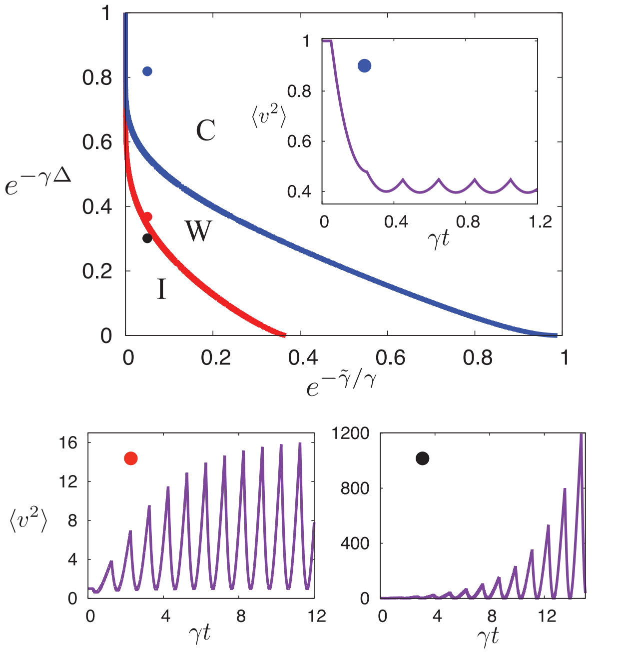

Fig. 3: (Color online) The diagram is drawn for $\delta/\Delta = 0.25$, $\sigma = 0.1$. C denotes the region for $T_\infty^{\text{av}} < T$, W for $T < T_\infty^{\text{av}} < \infty$, and I for $T_\infty^{\text{av}} = \infty$. The three points are picked from the three regions, for which the plots for $\langle v(t)^2\rangle$ *vs.* γt are shown.

where $H_\delta = (\tilde{\gamma}/\gamma)(1 - e^{-\gamma\delta})$. Adding eqs. (22) and (23) leads to $\langle \Delta S_{\text{sm}}^{(3)}\rangle$. $\langle \Delta S_{\text{sm}}^{(2)}\rangle$ in eq. (8) can be found by adding the one in eq. (23) and $\langle \ln \rho_1(\mathbf{c}_1) - \ln \rho_{1'}(\mathbf{c}'_1)\rangle$. The three representations of the Shannon entropy change for the joint system are shown in fig. 4.

The average environmental entropy production in eq. (3) is given as

$$\langle \Delta S_{\text{env}}\rangle = \left\langle \ln\left[\frac{\Pi_{v_1,v'_1}^{y_0}[v(t)]\Pi_{v'_1,v_2}^{y_1}[v(t)]}{\Pi_{-v_2,-v'_1}^{-y_1}[-v(t)]\Pi_{-v'_1,-v_1}^{-y_0}[-v(t)]}\right]\right\rangle$$
$$= \frac{T_1 - T_2}{2T} - \frac{\tilde{\gamma}}{\gamma T}\left[\langle v_2 y_1\rangle - \langle v'_1 y_1\rangle\right] - \frac{\tilde{\gamma}}{\gamma T}\left[\langle v'_1 y_0\rangle - \langle v_1 y_0\rangle\right]. \tag{24}$$

$\langle v'_1 y_1\rangle$, and $\langle v'_1 y_0\rangle$ can be obtained from $\langle v_2 y_1\rangle$, and $\langle v_2 y_0\rangle$ in eq. (20) by putting $\Delta = \delta$.

The average heat production is found from $\int_{t_1}^{t_2} \mathrm{d}t\langle[\gamma v(t) - \xi(t)] \circ v(t)\rangle$ with $\circ$ denoting the Stratonovich calculus[2]. We find $\langle Q\rangle = \gamma\Delta(T^{\text{av}} - T)$. When the average effective temperature is lower than the reservoir temperature, meeting the need of cold damping, the average heat becomes negative, which is the situation in which the paradox of Maxwell's demon is raised.

$\langle \Delta S_{\text{uc}}\rangle = \langle \Delta S_{\text{env}}\rangle - \langle Q/T\rangle$ is an unconventional entropy production which is known to appear in the presence of an odd-parity force; $-\tilde{\gamma}y_i$ in our case [26]. Without feedback control, $\langle \Delta S_{\text{tot}}\rangle$ maintains positivity even for a negative $\langle Q/T\rangle$ thanks to $\langle \Delta S_{\text{uc}}\rangle$. $-\langle \Delta I\rangle$ plays an additive role in compensating entropy loss in reservoir together with $\langle \Delta S_{\text{uc}}\rangle$.

[2] For the Wiener process $\mathrm{d}W$ defined by $\int_t^{t+\epsilon} \mathrm{d}s\xi(s)$ in $\epsilon \to 0$ limit, $\langle \mathrm{d}W \circ v(t)\rangle = \langle \mathrm{d}W[v(t) + v(t+\epsilon)]/2\rangle = \langle \mathrm{d}W \cdot \mathrm{d}v/2\rangle$, where $\langle \mathrm{d}W \cdot v(t)\rangle = 0$ is used. Using $\mathrm{d}v \simeq \mathrm{d}W$ from eq. (15), $\langle \mathrm{d}W \circ v(t)\rangle \simeq \langle(\mathrm{d}W)^2\rangle/2 = \epsilon\gamma T$.

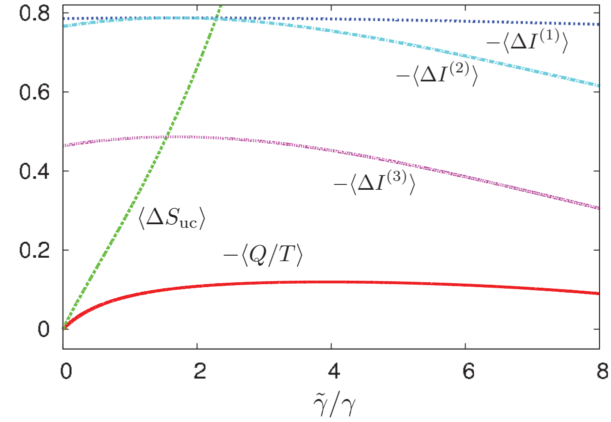

Fig. 4: (Color online) The components of $\langle \Delta S_{\text{tot}}^{(\alpha)}\rangle$ as functions of $\tilde{\gamma}/\gamma$ for the $i \to \infty$ step in the recursion procedure where $T_i = T_{i+1}$ and the Shannon entropy change of the system vanishes. Here, the plot is drawn for $\sigma = 0.1$, $\gamma\Delta = 0.2$, $\gamma\delta = 0.05$, and $T = 1$.

In fig. 4, we display the components comprising the total entropy change in the long-time limit of the recursive feedback process. Note that $\langle \Delta S_{\text{sm}}^{(\alpha)}\rangle = -\langle \Delta I^{(\alpha)}\rangle$ for zero Shannon entropy change of the system in the $i \to \infty$ step. $-\langle \Delta I^{(\alpha)}\rangle + \langle \Delta S_{\text{uc}}\rangle$ is shown to be greater than $-\langle Q/T\rangle$ for all α, which confirms the generalized second law of thermodynamics. As expected from fig. 2, $\langle \Delta S_{\text{tot}}^{(3)}\rangle$ is shown to yield the tightest bound.

In the feedback-driven process, there are two competing mechanisms for mutual information change: decrease due to thermal relaxation *vs.* increase due to induced correlation by memory-dependent protocol. The former occurs in the first stage after measurement. On the other hand, the latter may occur in a later stage for large duration (large Δ), strong drive (large $\tilde{\gamma}$), and large measurement error (large σ), which is the origin of overshooting. In our case, the increase of mutual information is found to result from *negative* correlation, as seen from the time-dependent behavior of $\langle v_2 y_1\rangle$ by replacing Δ by t in eq. (21). The relation of mutual information with correlation was recently investigated in detail [29]. In fig. 4, mutual information change in every type is not monotonous, but decreases for large $\tilde{\gamma}$, signalling overshooting.

In fig. 5(a), the reduction of mutual information between system and new memory is drawn in time. Note that the reduction increases if mutual information decreases or vice versa. For the delay period, the reduction increases due to simple thermal relaxation insensitively to $\tilde{\gamma}$ since the new memory is not yet used for the protocol in dynamics. For the period after delay, it shows the expected behavior: monotonous increase for small $\tilde{\gamma}$ and non-monotonous change for large $\tilde{\gamma}$ due to overshooting.

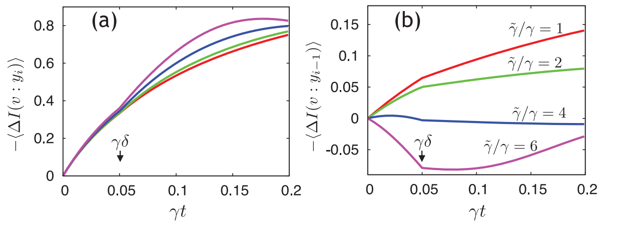

Fig. 5: (Color online) The reduction of mutual information in time within a step for various $\tilde{\gamma}/\gamma$. (a) $-\langle \Delta I(v : y_i) \rangle$ between system and new memory. (b) $-\langle \Delta I(v : y_{i-1}) \rangle$ between system and old memory. Here, the plots are drawn for $\sigma = 0.1$, $\gamma\Delta = 0.2$, $\gamma\delta = 0.05$, and $T = 1$ in the $i \to \infty$ limit. In (a), the same values of $\tilde{\gamma}/\gamma$ are used for the curves with the same colors as in (b).

The difference of $-\langle \Delta I^{(3)} \rangle$ from $-\langle \Delta I^{(2)} \rangle$ in fig. 4 mainly comes from the reduction of mutual information by the new memory for the delay period, which is shown to be insensitive to $\tilde{\gamma}$ in fig. 5(a). This explains well an almost uniform gap between the second and third curves from the top in fig. 4. In fig. 5(b), the drop of mutual information between system and old memory is drawn. For the delay period, it shows typical overshooting behaviors for large $\tilde{\gamma}$: non-monotonous or even decreasing. However, after delay, it shows monotonous increase since the old memory is no longer used for the protocol in the later dynamics. A tiny decrease occurs near $\tilde{\gamma}/\gamma = 4$, which might be due to an overlap between new and old memories. The difference of $-\langle \Delta I^{(2)} \rangle$ from $-\langle \Delta I^{(1)} \rangle$ comes form the drop of mutual information between system and old memory for the period after the delay. From the figure, one can see that the drop relative to the value at $t = \delta$ is decreasing for small $\tilde{\gamma}$, but is increasing for large $\tilde{\gamma}$. It also explains well the same behavior of the gap between the first and second curves from the top in fig. 4.

We examine the generalized thermodynamic second law in the presence of coexisting past and present memories. We show the total entropy change to have the tightest bound as only mutual information influencing the dynamics is considered, which is confirmed in the cold-damping problem. For the cold damping using a multi-step feedback, the effective temperature can be reduced below the reservoir temperature for a certain range of parameters, while it can reach a higher value or even diverge unlimitedly due to overshooting caused by large $\tilde{\gamma}$ and Δ, as shown in fig. 3. We derive the stability condition for the convergence of feedback. An intriguing role of δ to enhance the stability for large Δ will be further investigated in a future study [30]. We expect overshooting and instability to take place in general feedback processes for finite δ and Δ [31], which are unavoidable in reality.

* * *

This research was supported by the NRF Grant Nos. 2016R1D1A1A09918020 (CK) and 2013R1A1A2A10009722 (HP).

REFERENCES

[1] Maxwell J. C., *Theory of Heat* (Appleton, London) 1871.
[2] Szilard L., *Z. Phys.*, **53** (1929) 840; *Behav. Sci.*, **9** (1964) 301 (translated in English).
[3] Brillouin L., *J. Appl. Phys.*, **22** (1951) 334.
[4] Landauer R., *IBM J. Res. Dev.*, **5** (1961) 183; *Phys. Today*, **44**, issue No. 5 (1991) 23; *Science*, **272** (1996) 1914.
[5] Leff H. S. and Rex A. F. (Editors), *Maxwell's Demon 2: Entropy, Classical and Quantum Information, Computing* (IOP Publishing) 2003.
[6] Sagawa T. and Ueda M., *Phys. Rev. Lett.*, **100** (2008) 080403; **102** (2009) 250602; **104** (2010) 090602; *Phys. Rev. E*, **85** (2012) 021104; Klages R., Just W. and Jarzynski C. (Editors), *Nonequilibrium Statistical Physics of Small Systems: Fluctuation Relations and Beyond* (Wiley-VCH, Weinheim) 2012.
[7] Sagawa T. and Ueda M., *Phys. Rev. Lett.*, **109** (2012) 180602.
[8] Ito S. and Sagawa T., *Phys. Rev. Lett.*, **110** (2013) 180603.
[9] Sagawa T. and Ueda M., *New J. Phys.*, **15** (2013) 125012.
[10] Evans D. J., Cohen E. G. D. and Morriss G. P., *Phys. Rev. Lett.*, **71** (1993) 2401.
[11] Jarzynski C., *Phys. Rev. Lett.*, **78** (1997) 2690.
[12] Crooks G. E., *J. Stat. Phys.*, **90** (1998) 1481.
[13] Kurchan J., *J. Phys. A*, **31** (1998) 3719.
[14] Lebowitz J. L. and Spohn H., *J. Stat. Phys.*, **95** (1999) 333.
[15] Toyabe S., Sagawa T., Ueda M., Muneyuki E. and Sano M., *Nat. Phys.*, **6** (2010) 988.
[16] Koski J. V., Maisi V. F., Sagawa T. and Pekola J. P., *Phys. Rev. Lett.*, **113** (2014) 030601.
[17] Um J., Hinrichsen H., Kwon C. and Park H., *New J. Phys.*, **17** (2015) 085001.
[18] Shiraishi N. and Sagawa T., *Phys. Rev. E*, **91** (2015) 012130.
[19] Horowitz J. M. and Vaikuntanathan S., *Phys. Rev. E*, **82** (2010) 061120.
[20] Ponmurugan M., *Phys. Rev. E*, **82** (2010) 031129.
[21] Kim K. H. and Qian H., *Phys. Rev. E*, **75** (2007) 022102.
[22] Ito S. and Sano M., *Phys. Rev. E*, **84** (2011) 021123.
[23] Munakata T. and Rosinberg M. L., *J. Stat. Mech.* (2013) P06014.
[24] Munakata T. and Rosinberg M. L., *Phys. Rev. Lett.*, **112** (2014) 180601.
[25] Rosinberg M. L., Munakata T. and Tarjus G., *Phys. Rev. E*, **91** (2015) 042114.
[26] Kwon C., Yeo J. H., Lee H. and Park H., *J. Korean Phys. Soc.*, **68** (2016) 633.
[27] Jourdan G., Torricelli G., Chevrier J. and Comin F., *Nanotechnology*, **18** (2007) 475502.
[28] Onsager L. and Machlup S., *Phys. Rev.*, **91** (1953) 1505; 1512.
[29] Kwon C., arXiv:1608.04557v2 (2016).
[30] Um J., Noh J. D., Kwon C. and Park H., unpublished.
[31] Jun Y. and Bechhoefer J., *Phys. Rev. E*, **86** (2012) 061106.