

Entropy and Thermodynamic Second Laws: New Perspective - Stochastic Thermodynamics and Fluctuation Theorems

Hyunggyu PARK*

School of Physics and QUC, Korea Institute for Advanced Study, Seoul 02455, Korea

(Received 9 May 2018)

Recently, there has been a considerable progress on the issue of the thermodynamic second law, which is known as the law of entropy increase or irreversibility. In particular, a novel symmetry known as the Gallavotti-Cohen symmetry is found in nonequilibrium (NEQ) fluctuations, which leads to so-called fluctuation theorems. The thermodynamic second law is a simple corollary of fluctuation theorems, from which one can predict quantitatively how often NEQ processes violate the law of entropy increase. Violations disappear in the thermodynamic limit, but can be observed reasonably well in small systems. In this article, we will briefly introduce the stochastic thermodynamics and derive various fluctuation theorems, including the total entropy production (EP), the work-free-energy relation, the excess and house-keeping EP, and the information entropy.

PACS numbers: 05.70.Ln, 02.50.-r, 05.40.-a

Keywords: Entropy, Thermodynamic second law, Stochastic thermodynamics, Fluctuation theorems

DOI: 10.3938/jkps.72.1413

I. INTRODUCTION

Macroscopic systems tend to reach a stable and quiescent state without any external impulse, which is known as equilibrium (EQ). Equilibrium has been well explored over more than a century since the key thermodynamic quantity known as *entropy* could be microscopically evaluated by statistical mechanics, thanks mostly to Ludwig Boltzmann. However, most dynamic processes in nature are non-equilibrium (NEQ) processes and even some steady states (NESS) cannot be described by the EQ Boltzmann distribution. Until recently, not much things are known in general for these NEQ processes and states, except that the EQ Boltzmann entropy in the total system can not decrease between two dynamically connected EQ states (traditional thermodynamic second law) and the fluctuation dissipation theorems from the linear response theory near EQ.

The first and most striking breakthrough appeared in 1990's by Evans, Cohen, and Morriss [1], followed by a precise mathematical proof by Gallavotti and Cohen [2] on the fluctuation theorem (FT) of the entropy production in a NEQ steady state. As implied by the name of the FT, it provides an important and universal information on the fluctuation and distribution of the entropy production and, moreover, the thermodynamic second law follows automatically as a corollary of the FT.

Right after this breakthrough, various forms of the

FT's for general stochastic Markovian NEQ processes have been rigorously derived by many researchers, including Jarzynsky [3], Kurchan [4], Crooks [5], Lebowitz and Spohn [6]. Sources for NEQ processes are also diverse, including time-dependent Hamiltonians and non-conservative driving forces. Since then, there have been many interesting generalizations such as FT's for parts of the entropy production [7,8], inclusion of information entropy [9], odd-parity problem [10–13], and so forth. These theoretical results have been confirmed by experiments on small systems where fluctuations can be observed reasonably well [14–18]. We emphasize that all FT's hold regardless of how far the process is from EQ.

There were also attempts to generalize these ideas to systems with a strong coupling to environment [19], non-Markovian processes [20], non-Gaussian/correlated/multiplicative noises [21], and open (and also isolated) quantum systems [22], but there still remain many unresolved issues in these generalizations.

In this article, we will introduce the *stochastic thermodynamics* (ST) [23–25], which provides a key framework for fluctuation calculations with a new perspective on entropy. It will be seen that many fluctuation theorems can be derived almost trivially in this ST framework.

II. STOCHASTIC THERMODYNAMICS

Description of microscopic many-body deterministic systems in terms of a few macroscopic variables is one

*E-mail: hgpark@kias.re.kr

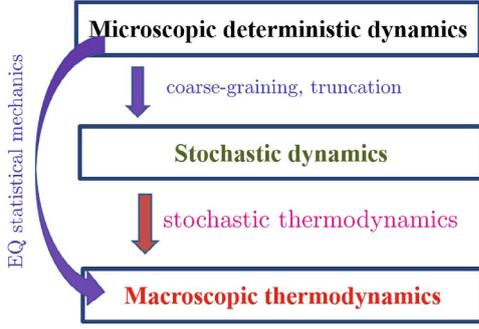


Fig. 1. (Color online) Conceptual flow diagram to macroscopic thermodynamics in EQ and NEQ. Stochastic thermodynamics bridges from stochastic dynamics described by Langevin or Master equations to macroscopic thermodynamics in general NEQ processes.

of the formidable task in physics. However, this is possible in EQ by utilizing the equilibrium statistical mechanics assuming the so-called *equally likely* postulate or the entropy maximization. Thus, the EQ statistical mechanics provides a direct link between microscopic deterministic systems and macroscopic thermodynamics. In contrast, there is no general postulate in NEQ, which is consistent with experimental findings. Thus, one has to go through *coarse-graining* procedures in order to reduce numerous uninterested degrees of freedom in microscopic systems and describe the system at the mesoscopic level with much less degrees of freedom. With some approximations or truncations, one can usually end up with stochastic dynamic equations such as Langevin equations or Master equations, which embraces time irreversibility in themselves.

Stochastic thermodynamics (ST) is a new framework to describe a general NEQ process in terms of the ensemble of dynamic trajectories of a stochastic equation. In particular, thermodynamic quantities such as heat and entropy production along with energy and work are *defined* for each dynamic trajectory [23,24]. These thermodynamic quantities fluctuate over the ensemble of trajectories and their ensemble averages become the conventional macroscopic thermodynamic observables. Thus, the ST provides a link between mesoscopic stochastic dynamics to macroscopic thermodynamics. A simple conceptual flow diagram to macroscopic thermodynamics starting from microscopic deterministic dynamics is shown in Fig. 1.

It is useful to look for an analogy between the EQ statistical mechanics and the ST. For convenience, we consider a system in contact with a single heat reservoir characterized by temperature T . EQ concerns only about statics because any macroscopic thermodynamic observable does not change over time by definition. Nevertheless, system state q fluctuates in the state space $\{q\}$ with the probability distribution function (PDF) $p(q)$. Note that state q is in principle a very high dimensional object, describing all degrees of freedom of the many-

body system. In this article, we only consider $q = (x, v)$ with position x and velocity v of a single particle in one dimension for simplicity. Generalization to more complex situations is mostly straightforward.

For a given Hamiltonian $H(q)$, it is well known that the PDF is given as the distribution $p(q) = e^{(F-H(q))/T}$ (canonical ensemble) with the free energy $F = \sum_q e^{-H(q)/T}$ in the Boltzmann unit ($k_B = 1$). Then, any state-dependent observable $A(q)$ such as the energy $H(q)$ can be averaged as $\langle A \rangle = \sum_q p(q)A(q)$. For an isolated system, we have $p(q) = 1/\Omega$ (microcanonical ensemble) with the number of *accessible* states Ω for a given macroscopic constraint, which is known as the equally likely postulate.

In the conventional EQ thermodynamics, the system entropy is defined usually by the ensemble-averaged quantity. One of the important ingredients of the ST is to put the system entropy on the same footing as the energetic quantities [24], *i.e.* as a fluctuating quantity defined as

$$S(q) \equiv -\ln p(q), \quad \langle S \rangle = -\sum_q p(q) \ln p(q), \quad (1)$$

where $\langle S \rangle$ is equivalent to the Shannon entropy in the information theory. With the PDF $p(q)$ inserted for EQ, $\langle S \rangle = \ln \Omega$ for the microcanonical ensemble, which is identical to the Boltzmann entropy, and $\langle S \rangle = -(F - \langle H \rangle)/T$ for the canonical ensemble, consistent with the EQ thermodynamics. The ST takes the same definition of the system entropy as above even for NEQ.

Now, consider the NEQ process where the dynamics really matters. To describe the dynamics, we consider dynamic trajectories $\mathbf{q} = \{q_t\}$ for $t_1 \leq t \leq t_2$ with the trajectory PDF $\mathcal{P}(\mathbf{q})$. Any trajectory-dependent observable $A(\mathbf{q})$ such as the energy change $\Delta H(\mathbf{q}) = H(q_{t_2}) - H(q_{t_1})$, work $W(\mathbf{q})$ done on the system by an external agent (force), and heat $Q(\mathbf{q})$ dissipated into the reservoir can be averaged over the trajectory ensemble as $\langle A \rangle = \sum_{\mathbf{q}} \mathcal{P}(\mathbf{q})A(\mathbf{q})$. Note that fluctuating heat $Q(\mathbf{q})$ in the trajectory space $\{\mathbf{q}\}$ [23] is related to fluctuating work $W(\mathbf{q})$ by the microscopic energy conservation as $W(\mathbf{q}) = \Delta H + Q(\mathbf{q})$.

To discuss the fluctuation theorem and the associated thermodynamic second law, we also need to treat the entropy production (EP) as a fluctuating quantity. In the ST, the EP or equivalently the entropy change of the environment (reservoir) $\Delta S_r(\mathbf{q})$ is defined as

$$\Delta S_r(\mathbf{q}) = Q(\mathbf{q})/T, \quad \langle \Delta S_r \rangle = \langle Q \rangle/T, \quad (2)$$

where $\langle \Delta S_r \rangle$ is equivalent to the Clausius EP in the conventional *reversible* thermodynamics. As the heat reservoir is assumed to be always in EQ even when the system undergoes a NEQ process, this result is exactly what is expected. The total entropy change is the sum of the system and reservoir entropy change for a weak coupling to the reservoir as

$$\Delta S_{\text{tot}}(\mathbf{q}) = \Delta S(\mathbf{q}) + \Delta S_r(\mathbf{q}), \quad (3)$$

where $\Delta S(\mathbf{q}) = S(q_{t_2}) - S(q_{t_1}) = -\ln p_{t_2}(q_{t_2}) + \ln p_{t_1}(q_{t_1})$. We expect that $\langle \Delta S_{\text{tot}} \rangle \geq 0$ for any natural process (thermodynamic second law), but $\Delta S_{\text{tot}}(\mathbf{q})$ fluctuates over dynamic trajectories. So there may be a chance to observe negative total EP in real experiments. In the next section, we investigate the universal feature of its distribution $P(\Delta S_{\text{tot}})$ for general Markovian processes, which is called the fluctuation theorem. The thermodynamic second law is just its corollary.

We end this section with a short summary on the ST. The ST is an ensemble theory for dynamic trajectories with an arbitrary trajectory PDF, in contrast to the EQ ensemble theory for accessible states with a specific PDF corresponding to a given environment. In addition, the system entropy and the EP are defined as stochastic (trajectory-dependent) quantities through Eqs. (1) and (2) along with stochastic heat and work.

III. FLUCTUATION THEOREMS

Fluctuation theorems are based on a very simple probability theory. We start with this probability theory, introduced by Esposito and Van den Broeck [25]. Consider two arbitrary normalized PDF's such as $\mathcal{P}(\mathbf{q})$ and $\tilde{\mathcal{P}}(\tilde{\mathbf{q}})$ with $\sum_{\mathbf{q}} \mathcal{P}(\mathbf{q}) = 1$ and $\sum_{\tilde{\mathbf{q}}} \tilde{\mathcal{P}}(\tilde{\mathbf{q}}) = 1$, where variables are linked to each other by an area-preserving map of $\tilde{\mathbf{q}} = \pi(\mathbf{q})$ (unit Jacobian: $J(\pi) = 1$). And we define the so-called *relative entropy* as

$$R(\mathbf{q}) \equiv \ln \frac{\mathcal{P}(\mathbf{q})}{\tilde{\mathcal{P}}(\tilde{\mathbf{q}})}, \quad (4)$$

which measures the difference between two PDF's. Then, it is trivial to show that

$$\langle e^{-R} \rangle_{\mathcal{P}} = \sum_{\mathbf{q}} e^{-R(\mathbf{q})} \mathcal{P}(\mathbf{q}) = \sum_{\tilde{\mathbf{q}}} \tilde{\mathcal{P}}(\tilde{\mathbf{q}}) = 1, \quad (5)$$

where the unit Jacobian is crucial in the derivation. In terms of a variable R , this result can be rewritten as

$$\langle e^{-R} \rangle = \sum_R e^{-R} P(R) = 1, \quad (6)$$

which constrains the distribution function $P(R)$. This is the *integral* FT for the observable R . Utilizing the Jensen's inequality, the above result guarantees

$$\langle R \rangle \geq 0, \quad (7)$$

which is called the Kullback-Leibler divergence.

One may get more information on $P(R)$ when a certain condition is satisfied. Consider a functional mapping f such as $\tilde{\mathcal{P}}(\tilde{\mathbf{q}}) = f \circ \mathcal{P}(\mathbf{q})$, which satisfies the *involution* property: $f^2 = I$ (I : identity mapping). Then, it is trivial to show that

$$\begin{aligned} P(R) &= \sum_{\mathbf{q}} \delta(R - R(\mathbf{q})) \mathcal{P}(\mathbf{q}) \\ &= \sum_{\tilde{\mathbf{q}}} \delta(R + \tilde{R}(\tilde{\mathbf{q}})) e^{-\tilde{R}(\tilde{\mathbf{q}})} \tilde{\mathcal{P}}(\tilde{\mathbf{q}}) = \tilde{P}(-R) e^R, \quad (8) \end{aligned}$$

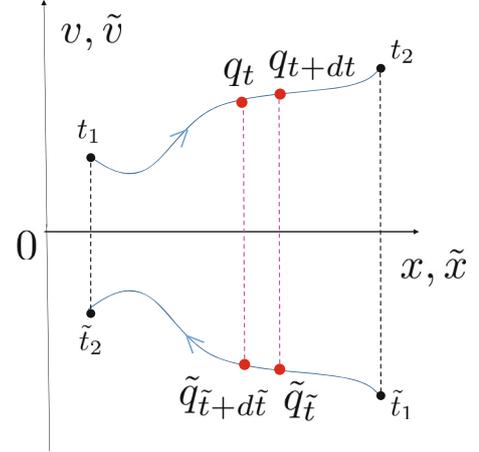


Fig. 2. (Color online) Dynamic trajectories. Curved lines denote a time-forward trajectory and its corresponding time-reverse trajectory with $\tilde{q}_{\tilde{t}} = \epsilon q_{t+dt}$ and $\tilde{q}_{\tilde{t}+d\tilde{t}} = \epsilon q_t$ with $\epsilon q = (x, -v)$.

where we used $\tilde{R}(\tilde{\mathbf{q}}) = -R(\mathbf{q})$ from the involution property and $\tilde{P}(R) = \sum_{\tilde{\mathbf{q}}} \delta(R - \tilde{R}(\tilde{\mathbf{q}})) \tilde{\mathcal{P}}(\tilde{\mathbf{q}})$. This relation links two PDF's by reflection with an extra exponential factor e^R , which is called the *detailed* FT.

1. Total EP and Irreversibility

The total EP measures irreversibility of a given process, so its average vanishes for a reversible process. In order to define the irreversibility, one should introduce a proper time-reverse process to compare with the original time-forward process. The standard time-reverse process should obey the *same* dynamic equation as in the time-forward process. A trajectory PDF in the time-reverse process is denoted by $\tilde{\mathcal{P}}(\tilde{\mathbf{q}})$ for trajectory $\tilde{\mathbf{q}} = \{\tilde{q}_{\tilde{t}}\}$ for $\tilde{t}_1 \leq \tilde{t} \leq \tilde{t}_2$, where all variables in the time-reverse process are denoted by ' $\tilde{\cdot}$ ' (Fig. 2).

It is convenient to consider a trajectory for an infinitesimal time duration between t and $t + dt$. Then, the trajectory PDF is given as

$$\mathcal{P}(\mathbf{q}) = p_t(q_t) \Pi_t(q_{t+dt}|q_t), \quad (9)$$

where $\Pi_t(q_{t+dt}|q_t)$ is the conditional probability (or propagator) to reach q_{t+dt} after time dt for a given initial state q_t . The propagator function Π_t may depend on time t with time-dependent Hamiltonians or driving forces. The trajectory PDF in the time-reverse process is given as

$$\tilde{\mathcal{P}}(\tilde{\mathbf{q}}) = \tilde{p}_{\tilde{t}}(\tilde{q}_{\tilde{t}}) \tilde{\Pi}_{\tilde{t}}(\tilde{q}_{\tilde{t}+d\tilde{t}}|\tilde{q}_{\tilde{t}}), \quad (10)$$

for an infinitesimal time duration between \tilde{t} and $\tilde{t} + d\tilde{t}$ with $\tilde{t} = t + dt$ and $\tilde{t} + d\tilde{t} = t$ ($d\tilde{t} = dt$). During this time interval, both dynamic equations should be identical, so $\tilde{\Pi}_{\tilde{t}} = \Pi_t$. This implies that the time variation of

the Hamiltonian or driving force for a finite time interval is reversed in the time-reverse process. For example, if one increases the external force in the time-forward process, then the external force should decrease in the time-reverse process, which seems natural.

To define the irreversibility for each trajectory, we choose a time-reverse trajectory as the exact reversal of a time-forward trajectory by setting $\tilde{q}_{\bar{t}} = \epsilon q_{t+dt}$, $\tilde{q}_{\bar{t}+d\bar{t}} = \epsilon q_t$, where ϵ is the parity operator as $\epsilon q = (x, -v)$, see Fig. 2. Furthermore, we require that the initial probability for the time-reverse process is exactly the same as the final probability for the time-forward process: $\tilde{p}_{\bar{t}}(\tilde{q}_{\bar{t}}) = p_{t+dt}(q_{t+dt})$. This implies that the time-reverse process starts right after the end of the time-forward process with the same ensemble except for inverting the velocity direction (note that $\tilde{v}_{\bar{t}} = -v_{t+dt}$).

We define the irreversibility of an infinitesimal dynamic trajectory \mathbf{q} as

$$d\text{Irr}(\mathbf{q}) \equiv \ln \frac{\mathcal{P}(\mathbf{q})}{\tilde{\mathcal{P}}(\tilde{\mathbf{q}})}, \quad (11)$$

where the time-reverse trajectory $\tilde{\mathbf{q}}$ and its initial ensemble are defined as above. It is trivial to show that the Jacobian of $\tilde{\mathbf{q}} = \pi(\mathbf{q})$ is unity. Thus, the integral FT is automatic by the probability theory as

$$\langle e^{-d\text{Irr}} \rangle = 1 \quad \text{and} \quad \langle d\text{Irr} \rangle \geq 0. \quad (12)$$

It is clear that the involution property holds for the conditional probability; $f^2 \circ \Pi_t(q_{t+dt}|q_t) = \Pi_t(q_{t+dt}|q_t)$. However, it holds for the initial PDF only in the steady state, guaranteeing $\tilde{p}_{\bar{t}+d\bar{t}}(\tilde{q}_{\bar{t}+d\bar{t}}) = p_t(q_t)$. As the time-dependent dynamics does not allow a steady state, the involutarity is valid in the steady state with a time-independent propagator Π . In this case, the detailed FT holds as

$$P(d\text{Irr}) = P(-d\text{Irr})e^{d\text{Itt}}, \quad (13)$$

where we drop ‘ \sim ’ in the PDF of the right hand side because the time-reverse process is identical to the time-forward process with the same initial ensemble (steady state) and the same propagator. This reflection symmetry with a specific exponential weight is generally called as the *Gallavotti-Cohen* (GC) symmetry of the PDF.

How is the irreversibility $d\text{Irr}$ related to the thermodynamic quantity such as the total entropy production? From Eqs. (9) and (10), one can see that $d\text{Irr}$ can be divided into two parts: The first part is the change of the system entropy $dS(\mathbf{q}) = -\ln p_{t+dt}(q_{t+dt}) + \ln p_t(q_t)$ defined in Eq. (1) and the second part is the logarithm of the conditional probability ratio. Here comes the key nontrivial derivation originally by Schnakenberg [26] and later by many others in general stochastic systems [4,6,27] as

$$\ln \frac{\Pi_t(q_{t+dt}|q_t)}{\Pi_t(\tilde{q}_{\bar{t}+d\bar{t}}|\tilde{q}_{\bar{t}})} = \frac{dQ(\mathbf{q})}{T} = dS_r(\mathbf{q}), \quad (14)$$

where $dQ(\mathbf{q})$ is stochastic heat dissipated into the reservoir with temperature T . Thus, the irreversibility is directly related to the EP as

$$d\text{Irr}(\mathbf{q}) = dS(\mathbf{q}) + dS_r(\mathbf{q}) = dS_{\text{tot}}(\mathbf{q}). \quad (15)$$

Therefore we finally get the integral and detailed FT's for the total EP as

$$\langle e^{-dS_{\text{total}}} \rangle = 1 \quad \text{and} \quad \frac{P(dS_{\text{total}})}{P(-dS_{\text{total}})} = e^{dS_{\text{total}}}. \quad (16)$$

It is trivial to extend the FT's to any finite time interval, due to the Markovianity of the process. We emphasize that the integral FT holds for any initial condition, as is so the thermodynamic second law $\langle dS_{\text{total}} \rangle \geq 0$, but the detailed FT (GC symmetry) only holds in the steady state.

One may guess that this symmetry could be recovered in the infinitely long time limit with any non-steady-state initial condition, because the system reaches its steady state and the initial condition information should be lost in this limit. However, it turns out that this guess is incorrect and the initial memory persists forever in the far tail of the PDF, where the extreme rare events dominate [28–32]. One may also ask a question whether the integral FT for heat may hold in the steady state as $\langle e^{-dQ/T} \rangle_{\text{ss}} = 1(?)$, because the average of the system EP vanishes ($\langle dS \rangle_{\text{ss}} = 0$) and also ask a question whether accumulated heat for a long time in the steady state satisfies the integral FT because accumulated heat should be extremely larger than the system entropy change. Both are not true, simply because dS is a fluctuating quantity which can be very large (unbounded) even with a very small probability. As the integral FT is simply the integral (or sum) of $e^{-dS_{\text{total}}}$ with the weight of the PDF, extremely large negative dS_{total} contributes to the integral significantly even though its probability is exponentially small. It is shown that the integral FT for heat can be satisfied only with the special initial condition of the uniform PDF [31,32].

We will not derive the Schnakenberg formula, Eq. (14), in this article, but only mention that this formula was rigorously derived for various general stochastic systems, *e.g.*, described by the Langevin equation and the Master equation. One can still make a rough argument inferring from the Arrhenius reaction (hopping) rate such as $e^{-dH/T}$ that the logarithmic ratio of two propagators is just a energy difference between two states q_t and q_{t+dt} divided by T , which should dissipate as heat into the reservoir.

It is noteworthy to point out that there are some examples where the Schnakenberg formula breaks down and generates an extra term besides dS_r . One well-studied case is when the system is influenced by velocity-dependent forces [10–13,33]. The Lorentz magnetic force is fundamental and velocity-dependent, but fortunately one can show the FT's for the total EP by taking the time-reverse process with the opposite direction of the

magnetic field B in Eq. (10) as $\tilde{\Pi}_t(B) = \Pi_t(-B)$. For general velocity-dependent forces, this remedy does not work and one can not avoid an extra EP term in the thermodynamic second law and the FT's [12, 34, 35]. These phenomenological velocity-dependent forces could be found in many realistic processes including active particle dynamics, cold damping models, information-assisted engines, and so on, where the standard thermodynamic second law is significantly modified [33,36–41].

2. Work and Free energy

It is usually difficult to measure the EP and heat experimentally. However, work done by the external force can be estimated rather easily. One can derive the so-called work-free-energy relation almost trivially, by only changing the initial ensemble in the previous discussion. This relation may be very useful and easily checked by experiments and simulations.

We take the same time-forward and time-reverse trajectories as before in Fig. 2 along with Eqs. (9) and (10), but with a special choice of the initial conditions as EQ Boltzmann distributions:

$$\begin{aligned} p_t(q_t) &= e^{(F_t - H_t(q_t))/T}, \\ \tilde{p}_{\tilde{t}}(\tilde{q}_{\tilde{t}}) &= e^{(F_{t+dt} - H_{t+dt}(q_{t+dt}))/T}, \end{aligned} \quad (17)$$

where the Hamiltonian H_t may vary with time as well as the free energy defined by its Hamiltonian for a given time. Note that the final PDF in the time-forward process $p_{t+dt}(q_{t+dt})$ can be arbitrary, depending on the dynamics and the time duration, which has nothing to do with the initial EQ PDF in the time-reverse process $\tilde{p}_{\tilde{t}}(\tilde{q}_{\tilde{t}})$. In this sense, the trajectory PDF $\tilde{\mathcal{P}}(\tilde{\mathbf{q}})$ is not the usual PDF of the time-reverse process.

However, the trajectory probability ratio leads to an interesting relation as

$$\ln \frac{\mathcal{P}(\mathbf{q})}{\tilde{\mathcal{P}}(\tilde{\mathbf{q}})} = \frac{-dF + dH(\mathbf{q}) + dQ(\mathbf{q})}{T} = \frac{dW(\mathbf{q}) - dF}{T}, \quad (18)$$

where $dF = F_{t+dt} - F_t$, $dH(\mathbf{q}) = H_{t+dt}(q_{t+dt}) - H_t(q_t)$, and stochastic work $dW(\mathbf{q}) = dH(\mathbf{q}) + dQ(\mathbf{q})$. The previous simple probability theory reads

$$\langle e^{-dW/T} \rangle = e^{-dF/T} \quad \text{and} \quad \frac{P(dW)}{\tilde{P}(-dW)} = e^{(dW - dF)/T}, \quad (19)$$

which are known as the *Jarzynsky identity* [3] and the *Crooks relation* [5], respectively. The corresponding thermodynamic second law is $\langle dW \rangle \geq dF$.

Note that the involution property is always satisfied in this case because each initial PDF is given by the

Boltzmann distribution with the Hamiltonian at its initial time. However, the time-reverse process is not identical to the time-forward process, due to different initial conditions $p_t \neq \tilde{p}_{\tilde{t}}$ as in Eq. (17). Furthermore, the time-reverse process for a finite time interval is different from the time-forward process, because the direction of the time variation in the external force or Hamiltonian is opposite. Therefore, we can not drop ‘ \sim ’ in the detailed FT in Eq. (19). We also note that work can be done by a non-conservative time-independent force. In this case, the Hamiltonian is independent of time, so $p_t = \tilde{p}_{\tilde{t}}$ and then we can drop ‘ \sim ’ along with $dF = 0$ in Eq. (19).

Finally, we emphasize that the work free energy FT's in Eq. (19) holds only when we start with EQ distributions. So, if one measures the cumulated work starting from EQ, the FT's are valid. However, both integral and detailed FT's break down for general NEQ processes starting an arbitrary initial PDF, so the average work can become smaller than the free energy difference.

3. Excess and House-keeping EP

As noticed from the simple probability theory, any fluctuating quantity written as a form of relative entropy in Eq. (4) satisfies the integral FT with unit Jacobian J . Even for $J \neq 1$, a modified integral FT holds as $\langle e^{-R'} \rangle = 1$ with $R' = R - \ln J$. Moreover, one can choose an arbitrary process represented by the trajectory PDF $\tilde{\mathcal{P}}(\tilde{\mathbf{q}})$ (not time-reverse process), where the conditional probabilities Π and $\tilde{\Pi}$ in Eqs. (9) and (10) are not related to each other. Thus, in principle, there are arbitrarily many quantities with valid FT's and so their averages never decrease in time [42].

What is important is whether these quantities could be identified by physical thermodynamic variables, which can be measured in average by experiments. Here, we mention a couple of examples of such cases. The first example is the so-called *excess EP* [7,43], dS_{ex} , which represents entropy (heat) production responsible for transitions between different steady states plus the system EP. This can be obtained by taking [25]

$$\tilde{\Pi}_{\tilde{t}}^*(\tilde{q}_{\tilde{t}+d\tilde{t}}|\tilde{q}_{\tilde{t}}) = \Pi_t(\tilde{q}_{\tilde{t}}|\tilde{q}_{\tilde{t}+d\tilde{t}}) \frac{p_{\text{ss}}^{(t)}(\tilde{q}_{\tilde{t}+d\tilde{t}})}{p_{\text{ss}}^{(t)}(\tilde{q}_{\tilde{t}})}, \quad (20)$$

where $p_{\text{ss}}^{(t)}$ is the *instantaneous* steady-state PDF defined as the expected steady-state PDF if the time-dependent Hamiltonian is kept unchanged at a given time t such that $\sum_q \Pi_t(q'|q) p_{\text{ss}}^{(t)}(q) = p_{\text{ss}}^{(t)}(q')$. The stochasticity of this *-dynamics is preserved by $\sum_{\tilde{q}_{\tilde{t}+d\tilde{t}}} \tilde{\Pi}_{\tilde{t}}^*(\tilde{q}_{\tilde{t}+d\tilde{t}}|\tilde{q}_{\tilde{t}}) = 1$ from Eq. (20), which makes $\tilde{\mathcal{P}}^*(\tilde{\mathbf{q}}) = \tilde{p}_{\tilde{t}}(\tilde{q}_{\tilde{t}}) \tilde{\Pi}_{\tilde{t}}^*(\tilde{q}_{\tilde{t}+d\tilde{t}}|\tilde{q}_{\tilde{t}})$ normalized.

We set $\tilde{q}_{\tilde{t}} = q_{t+dt}$ and $\tilde{q}_{\tilde{t}+d\tilde{t}} = q_t$ with $\tilde{t} = t + dt$, $\tilde{t} + d\tilde{t} = t$. Notice that the variable changes between q and \tilde{q} are different from the standard time-reverse

transformation (no parity operator). With the choice of $\tilde{p}_{\tilde{t}}(\tilde{q}_{\tilde{t}}) = p_{t+dt}(q_{t+dt})$, we define

$$dS_{\text{ex}}(\mathbf{q}) \equiv \ln \frac{\mathcal{P}(\mathbf{q})}{\tilde{\mathcal{P}}^*(\tilde{\mathbf{q}})} = dS(\mathbf{q}) + \frac{dQ_{\text{ex}}(\mathbf{q})}{T}, \quad (21)$$

with the excess heat $dQ_{\text{ex}}(\mathbf{q})$ as

$$dQ_{\text{ex}}(\mathbf{q}) = T \ln \frac{p_{\text{ss}}^{(t)}(q_{t+dt})}{p_{\text{ss}}^{(t)}(q_t)}. \quad (22)$$

With the time-independent propagator Π , the instantaneous steady state $p_{\text{ss}}^{(t)}$ does not depend on time t any more. If one starts with the steady state, then $dS_{\text{ex}}(\mathbf{q})$ is identically zero for any trajectory \mathbf{q} . Thus, the excess EP is responsible for evolution of the instantaneous steady states.

With the unit Jacobian again, we get the integral FT automatically as

$$\langle e^{-dS_{\text{ex}}} \rangle = 1 \quad \text{and} \quad \langle dS_{\text{ex}} \rangle \geq 1. \quad (23)$$

In the adiabatically slow limit, the system is almost always in the steady state, then $dS_{\text{ex}}(\mathbf{q}) = 0$ again for any \mathbf{q} . This is why dS_{ex} is also called the *non-adiabatic* EP [25]. It is interesting to note that the involution property holds for the conditional probability, but does not hold for the initial ensemble in general except for the adiabatic limit, where the detailed FT is meaningless due to $P(dS_{\text{ex}}) = \delta(dS_{\text{ex}})$.

Another interesting quantity is the so-called *house-keeping* EP [7,8,43], dS_{hk} , which usually represents entropy (heat) generated in order to maintain the NEQ steady state. This is simply defined as the remaining EP excluding dS_{ex} such as

$$\begin{aligned} dS_{\text{hk}}(\mathbf{q}) &\equiv dS_{\text{total}}(\mathbf{q}) - dS_{\text{ex}}(\mathbf{q}) \\ &= \ln \frac{\Pi_t(q_{t+dt}|q_t)p_{\text{ss}}^{(t)}(q_t)}{\Pi_t(\epsilon q_t|\epsilon q_{t+dt})p_{\text{ss}}^{(t)}(q_{t+dt})} = \frac{dQ_{\text{hk}}(\mathbf{q})}{T}, \end{aligned} \quad (24)$$

where the house-keeping heat $dQ_{\text{hk}}(\mathbf{q}) = dQ(\mathbf{q}) - dQ_{\text{ex}}(\mathbf{q})$. Note that $dS_{\text{hk}}(\mathbf{q})$ is independent of initial conditions and is also called the *adiabatic* EP [25], in contrast to dS_{ex} . We emphasize that $dS_{\text{hk}}(\mathbf{q}) = 0$ for any trajectory \mathbf{q} for systems without any non-conservative or time-dependent (NEQ) force, because their steady states should be EQ states and their dynamics satisfy two (independent) EQ properties such as the detailed balance; $\Pi(q'|q)p_{\text{ss}}(q) = \Pi(\epsilon q|\epsilon q')p_{\text{ss}}(\epsilon q')$ and the parity symmetry of the PDF; $p_{\text{ss}}(q) = p_{\text{ss}}(\epsilon q)$ [44].

One may ask whether dS_{hk} also satisfies the FT's. This turns out to be correct only in the overdamped dynamics [8], where all state variables have the even parity ($\epsilon = 1$). In this case, we find that the term in Eq. (24) is identical to the propagator of the time-forward stochastic *-dynamics as

$$\Pi_t(q_t|q_{t+dt})p_{\text{ss}}^{(t)}(q_{t+dt})/p_{\text{ss}}^{(t)}(q_t) = \Pi_t^*(q_{t+dt}|q_t), \quad (25)$$

which leads to $dS_{\text{hk}}(\mathbf{q}) = \ln[\mathcal{P}(\mathbf{q})/\mathcal{P}^*(\mathbf{q})]$ with normalized $\mathcal{P}^*(\mathbf{q}) = p_t(q_t)\Pi_t^*(q_{t+dt}|q_t)$. Thus, we get the FTs for the systems only with the even-parity state variables as

$$\langle e^{-dS_{\text{hk}}} \rangle = 1 \quad \text{and} \quad \frac{P(dS_{\text{hk}})}{P^*(-dS_{\text{hk}})} = e^{dS_{\text{hk}}}, \quad (26)$$

where the initial conditions are arbitrary for both integral and detailed FT's. With the odd-parity variables like velocities, both FT's for dS_{hk} break down [10,11], but one can divide dS_{hk} into two parts which are responsible for the breakage of the detailed balance and for the parity asymmetry of the steady-state PDF. The FT's hold for the former, but not for the latter [11,13].

For the NEQ dynamics driven by a time-independent non-conservative force (time-independent propagator Π), $dS_{\text{ex}}(\mathbf{q}) = 0$ for any \mathbf{q} starting from the steady state (note that $dQ_{\text{ex}}(\mathbf{q}) \neq 0$, even though $\langle dQ_{\text{ex}} \rangle_{\text{ss}} = 0$). Therefore, $dS_{\text{hk}}(\mathbf{q})$ is exactly equivalent to $dS_{\text{total}}(\mathbf{q})$, which should satisfy the FT's even with the odd-parity variables. The house-keeping heat can be written as $dQ_{\text{hk}}(\mathbf{q}) = dQ(\mathbf{q}) + TdS(\mathbf{q})$, implying that the stochastic house-keeping heat differs from the total stochastic heat even in the steady state, but their steady-state averages are the same: $\langle dQ_{\text{hk}} \rangle_{\text{ss}} = \langle dQ \rangle_{\text{ss}}$. The thermodynamic second law also guarantees $\langle dQ_{\text{hk}} \rangle_{\text{ss}} \geq 0$.

4. Information entropy

Finally, we briefly mention how to incorporate information into the FT [9]. This is almost trivial because the information is usually expressed by the Shannon entropy as in the stochastic thermodynamics. The simplest example involving information is a composite system (system + memory), where only the system is in contact with a thermal reservoir. The stochastic entropy of the composite system can be decomposed as

$$S_{\text{comp}}(q, m) = S_{\text{sys}}(q) + S_{\text{mem}}(m) - I(q, m), \quad (27)$$

where q and m denote state variables for the system and the memory, respectively, and I is the mutual information, representing the correlation between the system and the memory. Of course, $dS_{\text{total}} = dS_{\text{comp}} + dQ_{\text{sys}}/T$ satisfies the integral fluctuation theorem for any initial condition.

It is useful to take a special example such as the post-measurement process [9,45,46] in the so-called Maxwell's demon problem, where the memory state does not change in time during the feedback/relaxation process after the measurement. Furthermore, we start from the EQ initial distribution for the system such as $p_t(q_t, m_t) = p_t(q_t)p_t(m_t|q_t)$ with $p_t(q_t) = e^{(F_t - H_t(q_t))/T}$ with the system Hamiltonian H_t and an arbitrary conditional PDF of the memory $p_t(m_t|q_t)$, which is related to the mutual information as $p_t(m_t|q_t) = p_t(m_t)e^{I_t(q_t, m_t)}$ by Eq. (27).

Now, we take the similar methodology as in Subsec. III.2 with $\tilde{p}_{\tilde{t}}(\tilde{q}_{\tilde{t}}, \tilde{m}_{\tilde{t}}) = \tilde{p}_{\tilde{t}}(\tilde{q}_{\tilde{t}})\tilde{p}_{\tilde{t}}(\tilde{m}_{\tilde{t}}|\tilde{q}_{\tilde{t}})$ by setting $\tilde{q}_{\tilde{t}} = \epsilon q_{t+dt}$, $\tilde{q}_{\tilde{t}+d\tilde{t}} = \epsilon q_t$, $\tilde{m}_{\tilde{t}} = \epsilon m_t$, $\tilde{m}_{\tilde{t}+d\tilde{t}} = \epsilon m_t$, with $\tilde{t} = t + dt$, $\tilde{t} + d\tilde{t} = t$. We choose $\tilde{p}_{\tilde{t}}(\tilde{q}_{\tilde{t}}) = e^{(F_{t+dt} - H_{t+dt}(q_{t+dt}))/T}$ and $\tilde{p}_{\tilde{t}}(\tilde{m}_{\tilde{t}}|\tilde{q}_{\tilde{t}}) = \tilde{p}_{\tilde{t}}(\tilde{m}_{\tilde{t}})e^{\tilde{I}_{\tilde{t}}(\tilde{q}_{\tilde{t}}, \tilde{m}_{\tilde{t}})}$ with $\tilde{p}_{\tilde{t}}(\tilde{m}_{\tilde{t}}) = p_t(m_t)$ and $\tilde{I}_{\tilde{t}}(\tilde{q}_{\tilde{t}}, \tilde{m}_{\tilde{t}}) = I_{t+dt}(q_{t+dt}, m_t)$. Then, we get the integral FT only when the system starts from EQ as

$$\langle e^{-dW/T+dI} \rangle = e^{-dF/T}, \quad (28)$$

where $dW = dH + dQ_{\text{sys}}$ and $dI = I_{t+dt}(q_{t+dt}, m_t) - I_t(q_t, m_t)$. The corresponding second law is $\langle dW \rangle \geq T\langle dI \rangle + dF$. Without getting help from the free energy difference (in the case of $dF = 0$), one can utilize the information gained in the measurement process for the work extraction (negative $\langle dW \rangle$) during the feedback/relaxation process where the mutual information (correlation) decreases as $\langle dI \rangle < 0$. This is the key idea of an information engine.

IV. CONCLUDING REMARKS

In this article, we focus on the FT's and thermodynamic second laws for classical Markovian dynamics with a single heat reservoir. There have been some progresses in various generalizations such as multiple reservoirs [47,48], quantum systems [22], strong coupling to environments [19], non-Markovian processes [20], and non-thermal noises [21]. We believe that the thermodynamic second law for the total EP would be valid for any fundamental and natural phenomenological processes. Thanks to the stochastic thermodynamics, we can prove the fluctuation theorems easily for some systems, which shed a novel light on the directional mystery of the macroscopic time arrow. However, there are still many realistic systems waiting for being explored.

ACKNOWLEDGMENTS

We thank many colleagues who contributed to some works presented here and also for useful discussions. In particular, Chulan Kwon, Jae Dong Noh, Joonhyun Yeo, Hyun Keun Lee, Jaegon Um, Jae Sung Lee, and Haye Hinrichsen are deeply appreciated. This research was supported by the NRF Grant No. 2017R1D1A1B06035497.

REFERENCES

[1] D. J. Evans, E. G. D. Cohen and G. P. Morriss, Phys. Rev. Lett. **71**, 2401 (1993).

[2] G. Gallavotti and E. G. D. Cohen, Phys. Rev. Lett. **74**, 2694 (1995).
 [3] C. Jarzynski, Phys. Rev. Lett. **78**, 2690 (1997).
 [4] J. Kurchan, J. Phys. A **31**, 3719 (1998).
 [5] G. E. Crooks, J. Stat. Phys. **90**, 1481 (1998).
 [6] J. L. Lebowitz and H. Spohn, J. Stat. Phys. **95**, 333 (1999).
 [7] T. Hatano and S. I. Sasa, Phys. Rev. Lett. **86**, 3463 (2001).
 [8] T. Speck and U. Seifert, J. Phys. A **38**, L581 (2005).
 [9] T. Sagawa and M. Ueda, Phys. Rev. Lett. **100**, 080403 (2008); *ibid.* **102**, 250602 (2009); *ibid.* **104**, 090602 (2010).
 [10] R. E. Spinney and I. J. Ford, Phys. Rev. Lett. **108**, 170603 (2012); Phys. Rev. E **85**, 051113 (2012); *ibid.* **86**, 021127 (2012).
 [11] H. K. Lee, C. Kwon and H. Park, Phys. Rev. Lett. **110**, 050602 (2013).
 [12] C. Kwon, J. Yeo, H. K. Lee and H. Park, J. Korean Phys. Soc. **68**, 633 (2016).
 [13] J. Yeo, C. Kwon, H. K. Lee and H. Park, J. Stat. Mech. 093205 (2016).
 [14] J. Liphardt, S. Dumont, S. B. Smith, I. Tinoco Jr. and C. Bustamante, Science **296**, 1832 (2002).
 [15] D. Collin, F. Ritort, C. Jarzynski, S. B. Smith, I. Tinoco Jr. and C. Bustamante, Nature **437**, 231 (2005).
 [16] D. Andrieux, P. Gaspard, S. Ciliberto, N. Garnier, S. Joubaud and A. Petrosyan, Phys. Rev. Lett. **98**, 150601 (2007).
 [17] S. Ciliberto, A. Imparato, A. Naert and M. Tanase, Phys. Rev. Lett. **110**, 180601 (2013).
 [18] D. Y. Lee, C. Kwon and H. K. Pak, Phys. Rev. Lett. **114**, 060603 (2015).
 [19] U. Seifert, Phys. Rev. Lett. **116**, 020601 (2016).
 [20] T. Speck and U. Seifert, J. Stat. Mech. **2007**, L09002 (2007).
 [21] Y. Utsumi, D. S. Golubev, M. Marthaler, G. Schön and K. Kobayashi, Phys. Rev. B **86**, 075420 (2012).
 [22] J. Kurchan, arXiv:cond-mat/0007360; H. Tasaki, arXiv:cond-mat/0009244; M. Campisi, P. Talkner and P. Hänggi, Phys. Rev. Lett. **102**, 210401 (2009); M. Campisi, P. Hänggi and P. Talkner, Rev. Mod. Phys. **83**, 771 (2011).
 [23] K. Sekimoto, Prog. Theor. Phys. Suppl. **130**, 17 (1998).
 [24] U. Seifert, Phys. Rev. Lett. **95**, 040602 (2005).
 [25] M. Esposito and C. Van den Broeck, Phys. Rev. Lett. **104**, 090601 (2010).
 [26] J. Schnakenberg, Rev. Mod. Phys. **48**, 571 (1976).
 [27] H. Hinrichsen, C. Gogolin and P. Janotta, J. Phys.: Conf. Ser. **297**, 012011 (2011).
 [28] J. Farago, J. Stat. Phys. **107**, 781 (2002); Physica A **331**, 69 (2004).
 [29] R. van Zon and E. G. D. Cohen, Phys. Rev. Lett. **91**, 110601 (2003).
 [30] J. D. Noh and J-M. Park, Phys. Rev. Lett. **108**, 240603 (2012).
 [31] J. S. Lee, C. Kwon and H. Park, Phys. Rev. E **87**, 020104(R) (2013); J. Stat. Mech. **2013**, P11002 (2013).
 [32] K. Kim, C. Kwon and H. Park, Phys. Rev. E **90**, 032117 (2014).
 [33] H. K. Lee, S. Lahiri and H. Park, Phys. Rev. E **96**, 022134 (2017).
 [34] C. Ganguly and D. Chaudhuri, Phys. Rev. E **88**, 032102

- (2013); D. Chaudhuri, Phys. Rev. E **90**, 022131 (2014).
- [35] H-M. Chun and J. D. Noh, J. Stat. Mech. **2018**, 023208 (2018).
- [36] K. H. Kim and H. Qian, Phys. Rev. Lett. **93**, 120602 (2004); Phys. Rev. E **75**, 022102 (2007).
- [37] S. Ito and M. Sano, Phys. Rev. E **84**, 021123 (2011).
- [38] F. Schweitzer, *Brownian Agents and Active Particles* (Springer, Berlin, 2002); P. Romanczuk, M. Bär, W. Ebeling, B. Lindner and L. Schimansky-Geier, Eur. Phys. J. Special Topics **202**, 1 (2012).
- [39] M. C. Marchetti, J. F. Joanny, S. Ramaswamy, T. B. Liverpool, J. Prost, Madan Rao and R. Aditi Simha, Rev. Mod. Phys. **85**, 1143 (2013); É. Fodor, C. Nardini, M. E. Cates, J. Tailleur, P. Visco and F. van Wijland, Phys. Rev. Lett. **117**, 038103 (2016).
- [40] S. Ramaswamy, R. A. Simha and J. Toner, Europhys. Lett. **62**, 196 (2003).
- [41] H. Hirakawa, S. Hiramatsu and Y. Ogawa, Phys. Lett. **63A**, 199 (1977); J-M. Courty, A. Heidmann and M. Pinard, Eur. Phys. J. D **17**, 399 (2001); G. Jourdan, G. Torricelli, J. Chevrier and F. Comin, Nanotechnology **18**, 475502 (2007).
- [42] C. Pérez-Espigares, A. B. Kolton and J. Kurchan, Phys. Rev. E **85**, 031135 (2012).
- [43] Y. Oono and M. Paniconi, Prog. Theor. Phys. **130**, 29 (1998).
- [44] In Langevin dynamics, the detailed balance guarantees the parity symmetry, but it does not in Markovian jump dynamics [11–13].
- [45] C. Kwon, J. Um and H. Park, EPL **117**, 10011 (2017).
- [46] J. Um, H. Hinrichsen, C. Kwon and H. Park, New J. Phys. **17**, 085001 (2015).
- [47] Y. Murashita and M. Esposito, Phys. Rev. E **94**, 062148 (2016); M. Esposito and C. Van den Broeck, Phys. Rev. E **82**, 011143 (2010).
- [48] J. S. Lee and H. Park (unpublished).