Steady-State Thermodynamics for Heat Conduction: Microscopic Derivation

Teruhisa S. Komatsu,[1] Naoko Nakagawa,[2] Shin-ichi Sasa,[1] and Hal Tasaki[3]

[1]Department of Pure and Applied Sciences, University of Tokyo, Komaba, Meguro-ku, Tokyo 153-8902, Japan
[2]College of Science, Ibaraki University, Mito, Ibaraki 310-8512, Japan
[3]Department of Physics, Gakushuin University, Mejiro, Toshima-ku, Tokyo 171-8588, Japan

(Received 2 November 2007; revised manuscript received 19 March 2008; published 13 June 2008)

Starting from microscopic mechanics, we derive thermodynamic relations for heat conducting nonequilibrium steady states. The extended Clausius relation enables one to experimentally determine nonequilibrium entropy to the second order in the heat current. The associated Shannon-like microscopic expression of the entropy is suggestive. When the heat current is fixed, the extended Gibbs relation provides a unified treatment of thermodynamic forces in the linear nonequilibrium regime.

DOI: 10.1103/PhysRevLett.100.230602 PACS numbers: 05.70.Ln, 05.40.−a, 05.60.Cd

Thermodynamics (TD) is a theoretical framework that describes universal quantitative laws obeyed by macroscopic systems in equilibrium. A core of TD is the Clausius relation $\Delta S = Q/T$, which relates the entropy with the heat transfer caused by a change in the system. Combined with energy conservation, the Clausius relation leads to the Gibbs relation $TdS = dU + \sum_i f_i d\nu_i$, where ν_i is a controllable parameter and f_i the corresponding generalized force. The Gibbs relation is particularly useful since it represents the forces as gradients of suitable thermodynamic potentials. It also played a key role when Gibbs constructed equilibrium statistical mechanics.

Here we wish to address the fundamental question whether TD can be extended to nonequilibrium steady states (NESS) which, like equilibrium states, lack macroscopic time dependence. We shall call the possible extension steady-state thermodynamics (SST). The possibility of SST is far from trivial since NESS exhibit many properties which are very different from equilibrium states. First of all a naive extension of the Clausius relation to NESS is never possible since the heat transfer Q generally diverges linearly in time. It is also a deep theoretical question whether the long-range correlation universally observed in NESS [1] is consistent with SST. In addition to such abstract interests, there are nonequilibrium phenomena which may be better understood using SST. An interesting example is the force exerted on a small rigid body placed in a heat conducting fluid [2]. This force may be understood as a thermodynamic force in SST (see [3] for related ideas).

NESS sufficiently close to equilibrium can be characterized by the linear response theory. But this theory, which requires an ensemble of trajectories in space-time, does not lead us directly to SST. It is also clear that the theory only gives the result up to the first order in the "degree of nonequilibrium."

Although an extension of TD to NESS (or, equivalently, a construction of SST) may sound as a formidably difficult task, there are at least two branches of study which are encouraging. One is the series of works which reveal deep implications on NESS of the microscopic time-reversal symmetry. It has been shown that the simple symmetry (1) leads to various nontrivial results including the Green-Kubo relation, Kawasaki's nonlinear response relation, and the fluctuation theorem [4,5]. Although none of these works directly treat extensions of TD, techniques for characterizing NESS and energy transfer may be useful. The other is a series of works in which the theoretical consistency of SST was examined from purely phenomenological points of view. These works provide us with some useful guidelines for constructing SST. In [6] (see also [7]) it was proposed that heat Q in the Clausius relation should be replaced by "excess heat," which is the intrinsic heat transfer caused by the change of the state. In [8] it was conjectured that one should fix the total heat current J to get a Gibbs relation in a heat conducting NESS.

In the present Letter, we shall report a unification of these two branches, which has led us to a microscopic construction of SST (see [8,9] for early attempts to construct SST). More precisely we start from microscopic mechanics, and derive a natural extension of the Clausius relation to heat conducting NESS. The extended Clausius relation enables one to experimentally determine nonequilibrium entropy to the second order in the heat current. This extends the construction of entropy for NESS by Ruelle [10], who treated a simpler system with isokinetic thermostat. We further determine the precise form (4) of the entropy. In systems with a fixed heat current, we derive an extension of the Gibbs relation, which enables one to treat thermodynamic forces in the linear nonequilibrium regime in a new unified manner.

Setup.—The theory can be developed in various nonequilibrium settings including driven or sheared fluid. For simplicity we here focus on heat conduction, and consider a system which is attached to two heat baths and has controllable parameters (such as the volume).

We assume that the system consists of N particles whose coordinates are collectively denoted as $\Gamma = (\mathbf{r}_1, \ldots, \mathbf{r}_N; \mathbf{p}_1, \ldots, \mathbf{p}_N)$. We write its time reversal as $\Gamma^* = (\mathbf{r}_1, \ldots, \mathbf{r}_N; -\mathbf{p}_1, \ldots, -\mathbf{p}_N)$. When discussing time evolution of Γ, we denote by $\Gamma(t)$ its value at time t, and by $\hat{\Gamma} = (\Gamma(t))_{t\in[0,\mathcal{T}]}$ its history (or path) over the time interval $[0, \mathcal{T}]$. Given a path $\hat{\Gamma}$, we denote its time reversal as $\hat{\Gamma}^\dagger = (\Gamma(\mathcal{T} - t)^*)_{t\in[0,\mathcal{T}]}$.

We take a Hamiltonian satisfying the time-reversal symmetry $H_\nu(\Gamma) = H_\nu(\Gamma^*)$, where ν is the set of controllable parameters. Time evolution of the system is determined by the Hamiltonian and coupling to the two external heat baths with inverse temperatures β_1 and β_2. To model the heat baths, one may use Langevin noise or an explicit construction using Hamiltonian mechanics as in [11]. Our results are valid in both (and other physically natural) settings. We shall characterize our system using the set of parameters $\alpha = (\beta_1, \beta_2; \nu)$.

An external agent performs an operation to the system by changing α according to a prefixed protocol. A protocol is specified by a function $\alpha(t) = (\beta_1(t), \beta_2(t); \nu(t))$ of $t \in [0, \mathcal{T}]$. We denote by $\hat{\alpha} = (\alpha(t))_{t\in[0,\mathcal{T}]}$ the whole protocol. Again $\hat{\alpha}^\dagger = (\alpha(\mathcal{T} - t))_{t\in[0,\mathcal{T}]}$ denotes the time reversal of $\hat{\alpha}$. By (α) we denote a protocol in which the parameters are kept constant at α.

Consider a time evolution with a protocol $\hat{\alpha}$, and denote the probability weight for a path $\hat{\Gamma}$ as $\mathcal{W}_{\hat{\alpha}}[\hat{\Gamma}]$. It is normalized as $\int_{\Gamma(0)=\Gamma_i} \mathcal{D}\hat{\Gamma}\, \mathcal{W}_{\hat{\alpha}}[\hat{\Gamma}] = 1$ for any initial state Γ_i, where $\mathcal{D}\hat{\Gamma}$ denotes the path integral over all the histories (with the specified initial condition).

Time-reversal symmetry.—Let $J_k(\hat{\Gamma}; t)$ be the heat current from the kth bath to the system at time t in the history $\hat{\Gamma}$. Then the entropy production rate at t is given by $\theta_{\hat{\alpha}}(\hat{\Gamma}; t) = -\sum_{k=1}^2 \beta_k(t) J_k(\hat{\Gamma}; t)$. By integration, we get the entropy production $\Theta_{\hat{\alpha}}(\hat{\Gamma}) = \int_0^{\mathcal{T}} dt\, \theta_{\hat{\alpha}}(\hat{\Gamma}; t)$.

Then it has been shown [5,11] that the present (and other physically realistic) time evolution satisfies

$$\mathcal{W}_{\hat{\alpha}}[\hat{\Gamma}] = \mathcal{W}_{\hat{\alpha}^\dagger}[\hat{\Gamma}^\dagger] e^{\Theta_{\hat{\alpha}}(\hat{\Gamma})}, \tag{1}$$

which is the basis of the present work.

Steady state and its representation.—We assume that the system settles to a unique NESS when it evolves for a sufficiently long time with fixed α. We take $\mathcal{T}$ much larger than the relaxation time. We treat NESS with a small heat current, where convection hardly takes place.

In the NESS characterized by α, the expectation value of the current $J_k(\hat{\Gamma}; t)$ takes a constant value, which we denote as $\bar{J}_k(\alpha)$. We define the excess heat current from the kth bath as $J^{\rm ex}_{k,\hat{\alpha}}(\hat{\Gamma}; t) = J_k(\hat{\Gamma}; t) - \bar{J}_k(\alpha(t))$. Then $\theta^{\rm ex}_{\hat{\alpha}}(\hat{\Gamma}; t) = -\sum_{k=1}^2 \beta_k(t) J^{\rm ex}_{k,\hat{\alpha}}(\hat{\Gamma}; t)$ and its integration $\Theta^{\rm ex}_{\hat{\alpha}}(\hat{\Gamma}) = \int_0^{\mathcal{T}} dt\, \theta^{\rm ex}_{\hat{\alpha}}(\hat{\Gamma}; t)$ are the excess entropy production rate and the excess entropy production, respectively.

We denote by $\rho^{\rm st}_\alpha(\Gamma)$ the probability distribution for the NESS characterized by α. By using (1), it was shown in [12] (see also [11]) that the distribution has a concise representation

$$\rho^{\rm st}_\alpha(\Gamma) = \exp\left[-S(\alpha) + \frac{\langle\Theta^{\rm ex}_{(\alpha)}\rangle^{(\alpha)}_{\rm st,\Gamma} - \langle\Theta^{\rm ex}_{(\alpha)}\rangle^{(\alpha)}_{\Gamma^*,\rm st}}{2} + \tilde{R}(\alpha, \Gamma)\right], \tag{2}$$

where $S(\alpha)$ is determined by normalization, and $\tilde{R}(\alpha, \Gamma) = O(\epsilon^3)$. Here the degree of nonequilibrium ϵ is a dimensionless quantity proportional to the typical heat current. Throughout the present Letter, $\langle\cdots\rangle^{\hat{\alpha}}_{\Gamma_i,\Gamma_f}$ stands for the expectation taken with respect to the path probability $\mathcal{W}_{\hat{\alpha}}[\hat{\Gamma}]$ with the initial and the final conditions Γ_i and Γ_f, respectively, where the subscript st denotes the steady state [13]. The representation (2) plays a fundamental role in our construction of SST.

Extended Clausius relation.—Our first result is a natural extension of the Clausius relation.

Let $\hat{\alpha}$ be an arbitrary quasistatic protocol in which the parameters change slowly and smoothly from $\alpha_i = \alpha(0)$ to $\alpha_f = \alpha(\mathcal{T})$. Then the extended Clausius relation is

$$S(\alpha_f) - S(\alpha_i) = -\langle\Theta^{\rm ex}_{\hat{\alpha}}\rangle^{\hat{\alpha}} + R(\hat{\alpha}), \tag{3}$$

where $R(\hat{\alpha})$ is a small error about which we discuss shortly. (Here, and in what follows, $\langle\cdots\rangle^{\hat{\alpha}}$ is shorthand for $\langle\cdots\rangle^{\hat{\alpha}}_{\rm st,st}$.) Eq. (3) is the core of our SST.

When $\beta_1 = \beta_2$, we can show $R(\hat{\alpha}) = 0$, and hence (3) becomes precisely the standard Clausius relation. Note that the heat current in the original relation has been replaced in the extended relation (3) by the excess heat current, following the phenomenological proposals in [6,7]. Although the excess entropy production $\langle\Theta^{\rm ex}_{\hat{\alpha}}\rangle^{\hat{\alpha}}$ appears to depend on paths (in the parameter space) defined by the protocol $\hat{\alpha}$, (3) shows, rather strikingly, that it can be written as the difference of the entropy $S(\alpha)$, which is a function of α. This is far from a mere consequence of definitions, and represents a deep fact that NESS possess a nontrivial thermodynamic structure.

For an infinitesimal protocol $\hat{\alpha}$ [14], we will show that $R(\hat{\alpha}) = O(\epsilon^2\Delta)$, where Δ is a dimensionless quantity which characterizes the change $\alpha_f - \alpha_i$ [15]. (We know from examples [17] that this error estimate is optimal.)

The error term $R(\hat{\alpha})$ for a general quasistatic protocol $\hat{\alpha}$ can be obtained by summing up the errors in infinitesimal steps. In general $O(\Delta)$ sums up to $O(1)$, thus giving $R(\hat{\alpha}) = O(\epsilon^2)$. There are, however, important cases where we can set $R(\hat{\alpha}) = O(\epsilon^3)$. In such a case, the extended Clausius relation (3) is correct to $O(\epsilon^2)$, and hence goes beyond the linear response theory. Take, for example, the initial state α_i as an equilibrium state with $\beta_1 = \beta_2$. If we fix β_1 and change only β_2, the error $O(\epsilon^2\Delta)$ sums up to $R(\hat{\alpha}) = O(\epsilon^3)$ (see also [10]).

Nonequilibrium entropy.—$S(\alpha)$ in (3) was introduced as the normalization factor in the representation (2). It is interesting that this quantity plays the role of entropy in an operational thermodynamic relation.

In [18], we shall show that this entropy has an interesting symmetrized Shannon-like expression

$$S(\alpha) = -\int d\Gamma\, \rho^{\rm st}_\alpha(\Gamma) \log\sqrt{\rho^{\rm st}_\alpha(\Gamma)\rho^{\rm st}_\alpha(\Gamma^*)}. \tag{4}$$

Note that the right-hand side becomes precisely the Shannon entropy if $\rho^{\rm st}_\alpha(\Gamma) = \rho^{\rm st}_\alpha(\Gamma^*)$. Since the equilibrium distribution has this symmetry, the entropy $S(\alpha)$ approaches the Shannon entropy in the equilibrium limit.

The expression (4) shows that $S(\alpha)$ reflects certain properties of the steady-state distribution $\rho_\alpha^{\rm st}(\Gamma)$. Of particular interest is the long-range correlation [1], which should manifest itself as an anomalous size dependence of $S(\alpha)$ in the second order in ϵ. As we have examined above, the extended Clausius relation (3) allows one to compare the nonequilibrium and the equilibrium entropies, and determine $S(\alpha)$ in NESS with the precision of $O(\epsilon^2)$. One can thus detect the long-range correlation experimentally by means of calorimetry.

Extended Gibbs relation.—Our second major result is an extension of the Gibbs relation.

Let β be an arbitrary reference. Using the energy conservation, and noting that $\sum_{k=1}^2 \bar{J}_k(\alpha) = 0$, we get

$$\Theta_{\hat\alpha}^{\rm ex}(\hat\Gamma) = \beta\{W_{\hat\alpha}(\hat\Gamma) + H_{\nu(0)}(\Gamma(0)) - H_{\nu(\mathcal{T})}(\Gamma(\mathcal{T}))\} + \Phi_{\hat\alpha}(\hat\Gamma), \tag{5}$$

where $W_{\hat\alpha}(\hat\Gamma)$ is the total work done by the external agent who changes the parameters ν (the temperatures of the baths are changed without doing any work), and we defined $\Phi_{\hat\alpha}(\hat\Gamma) = -\sum_{k=1}^2 \int_0^{\mathcal{T}} dt[\beta_k(t) - \beta]J_{k,\hat\alpha}^{\rm ex}(\hat\Gamma; t)$.

If the average $\langle\Phi_{\hat\alpha}\rangle^{\hat\alpha}$ happens to be negligible, then (5) and the extended Clausius relation (3) imply

$$dS = \frac{dU}{T} + \sum_i \frac{f_i}{T} d\nu_i + O(\epsilon^2\Delta), \tag{6}$$

where we wrote $\langle W_{\hat\alpha}\rangle^{\hat\alpha} = -\sum_i f_i(\alpha) d\nu_i$, with $\nu = (\nu_1, \nu_2, \ldots)$, and $f_i(\alpha)$ being the (generalized) force conjugate to ν_i. Remarkably, (6) is identical to the standard Gibbs relation. We stress that all the terms in (6) can be determined experimentally by measuring heat currents and mechanical forces.

There may be several ways to make $\langle\Phi_{\hat\alpha}\rangle^{\hat\alpha}$ negligible. A natural strategy that comes from the phenomenological proposal in [8] is to consider a system with a fixed heat current J. To be more precise, we consider a "source-drain system," in which the two baths have different characters. Bath 1, which has a lower temperature and is coupled efficiently to the system, is a "heat drain." It helps the system to get rid of extra energy and reach the NESS rapidly. Bath 2, which has a higher fixed temperature, is a "heat source." It supplies a constant heat current J to the system in average when the system is disturbed by an external operation [19]. This means that $\langle J_{2,\hat\alpha}^{\rm ex}(t)\rangle^{\hat\alpha}$ is negligible. We now choose the reference as $\beta = \beta_1(0)$ so that $\beta_1(t) - \beta = O(\Delta)$. Since $\langle J_{1,\hat\alpha}^{\rm ex}(t)\rangle^{\hat\alpha} = O(\Delta)$, we find that $\langle\Phi_{\hat\alpha}\rangle^{\hat\alpha}$ is $O(\Delta^2)$ and hence negligible.

In a source-drain system, it is natural to characterize the NESS by parameters (T, J, ν), where $T = 1/\beta_1$. (β_2 is determined from T, J, and ν.) If we restrict ourselves to the operations in which only the parameter ν of the Hamiltonian changes, (6) gives

$$f_i(T, J, \nu) = -\frac{\partial}{\partial\nu_i} F(T, J, \nu) + O(\epsilon^2), \tag{7}$$

where the nonequilibrium free energy is defined by the familiar relation $F = U - TS$. The relation (7) shows that any thermodynamic force (including that exerted on a body in a heat conducting fluid) in the linear nonequilibrium regime is indeed a conservative force with the potential $F(T, J, \nu)$. Although any physical quantity can be evaluated to $O(\epsilon)$ by using the linear response theory, (7) may provide a novel point of view for analyzing thermodynamic forces in the setting with a fixed current. For example, (7) implies the Maxwell relation $\partial f_i/\partial\nu_j = \partial f_j/\partial\nu_i + O(\epsilon^2)$, which may be confirmed experimentally in suitable settings.

Derivation of main equality (3).—We consider an infinitesimal protocol $\hat\alpha$ [14]. Noting that $\Theta_{\hat\alpha^\dagger}(\hat\Gamma^\dagger) = -\Theta_{\hat\alpha}(\hat\Gamma)$, (1) implies $\mathcal{W}_{\hat\alpha}[\hat\Gamma]e^{-\Theta_{\hat\alpha}^{\rm ex}(\hat\Gamma)/2} = \mathcal{W}_{\hat\alpha^\dagger}[\hat\Gamma^\dagger]e^{-\Theta_{\hat\alpha^\dagger}^{\rm ex}(\hat\Gamma^\dagger)/2}$. By integrating over all paths satisfying $\Gamma(0) = \Gamma_{\rm i}, \Gamma(\mathcal{T}) = \Gamma_{\rm f}$, we get

$$\rho_{\alpha_{\rm f}}^{\rm st}(\Gamma_{\rm f})\langle\exp[-\Theta_{\hat\alpha}^{\rm ex}/2]\rangle_{\Gamma_{\rm i},\Gamma_{\rm f}}^{\hat\alpha} = \rho_{\alpha_{\rm i}}^{\rm st}(\Gamma_{\rm i}^*)\langle\exp[-\Theta_{\hat\alpha^\dagger}^{\rm ex}/2]\rangle_{\Gamma_{\rm f}^*,\Gamma_{\rm i}^*}^{\hat\alpha^\dagger}, \tag{8}$$

which is our starting point. We later show that

$$\langle\exp[-\Theta_{\hat\alpha}^{\rm ex}/2]\rangle_{\Gamma_{\rm i},\Gamma_{\rm f}}^{\hat\alpha}/\langle\exp[-\Theta_{\hat\alpha^\dagger}^{\rm ex}/2]\rangle_{\Gamma_{\rm f}^*,\Gamma_{\rm i}^*}^{\hat\alpha^\dagger} = \exp\left[-\frac{\langle\Theta_{\hat\alpha}^{\rm ex}\rangle_{\Gamma_{\rm i},\Gamma_{\rm f}}^{\hat\alpha} - \langle\Theta_{\hat\alpha^\dagger}^{\rm ex}\rangle_{\Gamma_{\rm f}^*,\Gamma_{\rm i}^*}^{\hat\alpha^\dagger}}{2} + R'(\hat\alpha; \Gamma_{\rm i}, \Gamma_{\rm f})\right], \tag{9}$$

with $R'(\hat\alpha; \Gamma_{\rm i}, \Gamma_{\rm f}) = O(\epsilon^3) + O(\epsilon^2\Delta)$. We assume here that various quantities can be expanded both in ϵ and Δ. We regard Δ as infinitesimal and omit $O(\Delta^2)$.

Note that $\langle\Theta_{\hat\alpha}^{\rm ex}\rangle_{\Gamma_{\rm i},\Gamma_{\rm f}}^{\hat\alpha} = \int_0^{\mathcal{T}} dt\langle\theta_{\hat\alpha}^{\rm ex}(t)\rangle_{\Gamma_{\rm i},\Gamma_{\rm f}}^{\hat\alpha}$ holds, and $\langle\theta_{\hat\alpha}^{\rm ex}(t)\rangle_{\Gamma_{\rm i},\Gamma_{\rm f}}^{\hat\alpha}$ attains non-negligible values only when t is near 0, $\mathcal{T}$, or $\mathcal{T}/2$ (where the system is out of steady states either by the imposed conditions or the operation). We can therefore decompose the expectation value as

$$\langle\Theta_{\hat\alpha}^{\rm ex}\rangle_{\Gamma_{\rm i},\Gamma_{\rm f}}^{\hat\alpha} = \langle\Theta_{(\alpha_{\rm i})}^{\rm ex}\rangle_{\Gamma_{\rm i},{\rm st}}^{(\alpha_{\rm i})} + \langle\Theta_{(\alpha_{\rm f})}^{\rm ex}\rangle_{{\rm st},\Gamma_{\rm f}}^{(\alpha_{\rm f})} + \langle\Theta_{\hat\alpha}^{\rm ex}\rangle_{\rm st,st}^{\hat\alpha}. \tag{10}$$

By substituting (9) and (10) into the identity (8), and comparing the result with the representation (2), we get

$$S(\alpha_{\rm f}) - S(\alpha_{\rm i}) = \frac{1}{2}\{\langle\Theta_{\hat\alpha^\dagger}^{\rm ex}\rangle_{\rm st,st}^{\hat\alpha^\dagger} - \langle\Theta_{\hat\alpha}^{\rm ex}\rangle_{\rm st,st}^{\hat\alpha}\} + R(\hat\alpha), \tag{11}$$

where $R(\hat\alpha) = R'(\hat\alpha; \Gamma_{\rm i}, \Gamma_{\rm f}) - \tilde{R}(\alpha_{\rm i}, \Gamma_{\rm i}) + \tilde{R}(\alpha_{\rm f}, \Gamma_{\rm f})$. Since $R(\hat\alpha) = 0$ if $\Delta = 0$, we must have that $R(\hat\alpha) = O(\epsilon^2\Delta)$. Noting the symmetry $\langle\Theta_{\hat\alpha^\dagger}^{\rm ex}\rangle_{\rm st,st}^{\hat\alpha^\dagger} = -\langle\Theta_{\hat\alpha}^{\rm ex}\rangle_{\rm st,st}^{\hat\alpha}$ [18] we get (3) for an infinitesimal process.

Derivation of (9).—We regard (only in this derivation) time-independent $H_{\nu(0)}$ as the Hamiltonian of the system, and interpret the force from $H_{\nu(t)} - H_{\nu(0)}$ as an "external force." Then the energy balance implies $H_{\nu(0)}(\Gamma(\mathcal{T})) - H_{\nu(0)}(\Gamma(0)) = W(\hat\Gamma) + \sum_{k=1}^2 \int_0^{\mathcal{T}} dt J_k^{\rm ex}(\hat\Gamma; t)$ where $W(\hat\Gamma)$ is the total work done by the external force. By defining $\tilde{\Phi}_{\hat{\alpha}}(\hat{\Gamma}) = \Phi_{\hat{\alpha}}(\hat{\Gamma}) + \beta W(\hat{\Gamma})$, we have $\Theta^{\rm ex}_{\hat{\alpha}}(\hat{\Gamma}) = \tilde{\Phi}_{\hat{\alpha}}(\hat{\Gamma}) + \beta\{H_{\nu(0)}(\Gamma(0)) - H_{\nu(0)}(\Gamma(\mathcal{T}))\}$ as in (5).

To simplify notation, we drop $\hat{\alpha}$ or $\hat{\alpha}^\dagger$, and abbreviate the expectations $\langle\cdots\rangle^{\hat{\alpha}}_{\Gamma_{\rm i},\Gamma_{\rm f}}$ and $\langle\cdots\rangle^{\hat{\alpha}^\dagger}_{\Gamma^*_{\rm f},\Gamma^*_{\rm i}}$ as $\langle\cdots\rangle$ and $\langle\cdots\rangle^\dagger$, respectively. We make use of the cumulant expansion $\log\langle e^{-\Theta^{\rm ex}/2}\rangle = -\langle\Theta^{\rm ex}\rangle/2 + \langle\Theta^{\rm ex};\Theta^{\rm ex}\rangle/8 + \cdots$, where $\langle\Theta^{\rm ex};\Theta^{\rm ex}\rangle = \langle(\Theta^{\rm ex})^2\rangle - \langle\Theta^{\rm ex}\rangle^2$. Since $H_{\nu(0)}(\Gamma(0)) - H_{\nu(0)}(\Gamma(\mathcal{T}))$ is constant in the present average, we have $\langle\Theta^{\rm ex};\Theta^{\rm ex}\rangle = \langle\tilde{\Phi};\tilde{\Phi}\rangle$. Similar identities also hold for higher order cumulants (see, e.g., [11]).

Let us denote by K the left-hand side of (9). The cumulant expansion yields

$$\log K = -\frac{\langle\Theta^{\rm ex}\rangle - \langle\Theta^{\rm ex}\rangle^\dagger}{2} + \frac{\langle\tilde{\Phi};\tilde{\Phi}\rangle - \langle\tilde{\Phi};\tilde{\Phi}\rangle^\dagger}{8} + O(\tilde{\Phi}^3). \tag{12}$$

To evaluate the second term, we observe that

$$\langle\tilde{\Phi};\tilde{\Phi}\rangle - \langle\tilde{\Phi};\tilde{\Phi}\rangle^\dagger = \langle\tilde{\Phi};\tilde{\Phi}\rangle_{\rm eq} - \langle\tilde{\Phi};\tilde{\Phi}\rangle^\dagger_{\rm eq} + O(\tilde{\Phi}^3), \tag{13}$$

where $\langle\cdots\rangle_{\rm eq}$ and $\langle\cdots\rangle^\dagger_{\rm eq}$ are averages in the corresponding equilibrium dynamics with the static Hamiltonian H and a common β. But the time-reversal symmetry in equilibrium dynamics implies $\langle\tilde{\Phi};\tilde{\Phi}\rangle_{\rm eq} = \langle\tilde{\Phi};\tilde{\Phi}\rangle^\dagger_{\rm eq}$. Since $\Phi = O(\epsilon)$ [16] and $\beta W = O(\Delta)$, we have $\tilde{\Phi} = O(\epsilon) + O(\Delta)$. Thus (12) and (13) imply the desired (9).

Discussions.—We treated a general classical model of heat conduction, and derived natural nonequilibrium extensions of the Clausius and the Gibbs relations. The mere existence of a consistent operational thermodynamics (i.e., SST) may be of great importance, but the way the extension has been done may also be quite suggestive.

The extended Clausius relation (3) and the associated microscopic expression (4) of the entropy form a theoretical core of the present work. They may provide us with a clue to developing the statistical mechanics for NESS that work beyond the linear response regime.

It is also suggestive that we obtained the extended Gibbs relation (6) in a special setting with "source" and "drain," in which the heat current is fixed. There is a possibility that this special setting is necessary for uncovering universal statistical properties of heat conducting systems, as such properties may be hidden in other settings. In this connection, it is exciting to explore implications of the "nonequilibrium order parameter" defined as $\Psi(T, J, \nu) = \partial F(T, J, \nu)/\partial J$ [8].

We hope that the present results trigger further nontrivial developments in nonequilibrium physics.

We wish to thank Glenn Paquette for crucial discussions which made us realize a flaw in the earlier version of the present work. We also thank the anonymous referee for letting us know of [10]. This work was supported by Grants No. 18740240 (T. S. K.), No. 19540392 (N. N.), and No. 19540394 (S. S.) from the Ministry of Education, Science, Sports and Culture of Japan.

[1] J. R. Dorfman, T. R. Kirkpatrick, and J. V. Sengers, Annu. Rev. Phys. Chem. **45**, 213 (1994).
[2] S. Wiegand, J. Phys. Condens. Matter **16**, R357 (2004).
[3] H. Brenner, Phys. Rev. E **72**, 061201 (2005); S. Duhr and D. Braun, Proc. Natl. Acad. Sci. U.S.A. **103**, 19 678 (2006).
[4] D. J. Evans, E. G. D. Cohen, and G. P. Morriss, Phys. Rev. Lett. **71**, 2401 (1993); G. Gallavotti and E. G. D. Cohen, Phys. Rev. Lett. **74**, 2694 (1995); J. Kurchan, J. Phys. A **31**, 3719 (1998); C. Maes, J. Stat. Phys. **95**, 367 (1999); G. E. Crooks, Phys. Rev. E **61**, 2361 (2000).
[5] C. Jarzynski, J. Stat. Phys. **98**, 77 (2000); C. Maes and K. Netočný, J. Stat. Phys. **110**, 269 (2003).
[6] Y. Oono and M. Paniconi, Prog. Theor. Phys. Suppl. **130**, 29 (1998).
[7] R. Landauer, Phys. Rev. A **18**, 255 (1978).
[8] S. Sasa and H. Tasaki, J. Stat. Phys. **125**, 125 (2006).
[9] T. Hatano and S. Sasa, Phys. Rev. Lett. **86**, 3463 (2001).
[10] D. Ruelle, Proc. Natl. Acad. Sci. U.S.A. **100**, 3054 (2003).
[11] T. S. Komatsu, N. Nakagawa, S. Sasa, and H. Tasaki, arXiv:0805.3023.
[12] T. S. Komatsu and N. Nakagawa, Phys. Rev. Lett. **100**, 030601 (2008).
[13] To be precise, we define expectations with various initial and final conditions as follows. We omit $\hat{\alpha}$ for simplicity. $\langle f\rangle_{\Gamma_{\rm i},\Gamma_{\rm f}} = \{\rho^{\rm st}_{\alpha_{\rm f}}(\Gamma_{\rm f})\}^{-1}\int_{\Gamma(0)=\Gamma_{\rm i},\Gamma(\mathcal{T})=\Gamma_{\rm f}}\mathcal{D}\hat{\Gamma}\,\mathcal{W}[\hat{\Gamma}]f(\hat{\Gamma})$, $\langle f\rangle_{\Gamma_{\rm i},{\rm st}} = \int_{\Gamma(0)=\Gamma_{\rm i}}\mathcal{D}\hat{\Gamma}\,\mathcal{W}[\hat{\Gamma}]f(\hat{\Gamma})$, $\langle f\rangle_{{\rm st},\Gamma_{\rm f}} = \{\rho^{\rm st}_{\alpha_{\rm f}}(\Gamma_{\rm f})\}^{-1}\times\int_{\Gamma(\mathcal{T})=\Gamma_{\rm f}}\mathcal{D}\hat{\Gamma}\rho^{\rm st}_{\alpha_{\rm i}}(\Gamma_{\rm i})\mathcal{W}[\hat{\Gamma}]f(\hat{\Gamma})$, and $\langle f\rangle_{\rm st,st} = \int\mathcal{D}\hat{\Gamma}\rho^{\rm st}_{\alpha_{\rm i}}(\Gamma_{\rm i})\mathcal{W}[\hat{\Gamma}]f(\hat{\Gamma})$. Here the normalization factors are simplified by assuming that $\rho^{\rm st}_{\alpha_{\rm f}}(\Gamma_{\rm f}) = \int_{\Gamma(0)=\Gamma_{\rm i},\Gamma(\mathcal{T})=\Gamma_{\rm f}}\mathcal{D}\hat{\Gamma}\,\mathcal{W}[\hat{\Gamma}]$ for any $\Gamma_{\rm i}$.
[14] An infinitesimal protocol $\hat{\alpha}$ is defined by $\alpha(t) = \alpha_{\rm i}$ for $t \in [0, \mathcal{T}/2)$ and $\alpha(t) = \alpha_{\rm f}$ for $t \in [\mathcal{T}/2, \mathcal{T}]$, where $\alpha_{\rm f} - \alpha_{\rm i}$ is infinitesimal.
[15] Let $\Delta\beta$ be the typical change in the inverse temperature, and W be the typical work done according to the change of ν. Then $\Delta \sim E_0\Delta\beta + \beta W$, where $E_0 = J_{\max}\tau_{\rm r}$ (see [16]) is a typical energy scale.
[16] We assume that the typical current in the system is $J \sim \kappa\delta\beta$ when there is a difference $\delta\beta$ in the inverse temperatures of the baths. Since the average of Φ is vanishing, we examine its behavior when the system is disturbed into a nontypical state. Then one expects a large current of the order $J_{\max} \sim \kappa\beta$. Since this decays rapidly within the relaxation time $\tau_{\rm r}$, we see $\Phi \sim \delta\beta J_{\max}\tau_{\rm r} \sim \beta J_{\rm ss}\tau_{\rm r} \sim \epsilon$, where $J_{\rm ss} \sim \kappa\delta\beta$ is the current in the steady state.
[17] G. C. Paquette (private communication).
[18] T. S. Komatsu, N. Nakagawa, S. Sasa, and H. Tasaki, report (to be published).
[19] Although not all heat baths act as a "heat source," one can design various sources [without violating the basic assumption (1)] which keeps the averaged current almost constant when the parameters are changed. An example is a bath with a very high temperature coupled very weakly to the system [18].