Irreversible Thermodynamics

2 authors:

David Michael Rogers
University of South Florida
54 PUBLICATIONS   **638** CITATIONS

SEE PROFILE

Susan B Rempe
Sandia National Laboratories
183 PUBLICATIONS   **3,561** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project Simulation and analysis algorithms View project

Project Quasi-Chemical Theory View project

A First and Second Law for Nonequilibrium Thermodynamics: Maximum Entropy Derivation of the Fluctuation-Dissipation Theorem and Entropy Production Functionals

David M. Rogers[*] and Susan B. Rempe[†]

Center for Biological and Materials Sciences, MS 0895,

Sandia National Laboratories, Albuquerque, New Mexico 87185, USA

Abstract

This paper essentially appears in David M. Rogers and Susan B. Rempe. "Irreversible Thermodynamics." J. Phys.: Conf. Ser. 402:012014, 2012 – which deserves the citation for this work.

We derive a physically motivated theory for nonequilibrium systems from a maximum entropy approach similar in spirit to the equilibrium theory given by Gibbs. Requiring Hamilton's principle of stationary action to be satisfied on average during a trajectory, we derive constraints on the *transition* probability distribution that lead to a path probability of the Onsager-Machlup form. Additional maximum transition entropy constraints, derived from energy and momentum conservation laws, then introduce heat exchange and external driving forces into the system via a process reminiscent to the LagrangedAlembert principle. The Lagrange multipliers for these constraints express the instantaneous temperature and pressure of thermostatic external driving systems. The result is a fully time-dependent, non-local description of a nonequilibrium ensemble coupled to reservoirs at arbitrary thermostatic or mechanical states. Detailed accounting of the energy exchange and the change in information entropy of the central system then provides a description of the entropy production that is not dependent on the specification or existence of a steady-state or on any definition of thermostatic variables for the central system. We find that the maximum entropy route provides a clear derivation of the path free energy functional, path-integral, Langevin, Brownian, and Fokker-Planck statements of nonequilibrium processes. We show that the equilibrium notions of the Gibbs-Maxwell relations and free energy perturbation techniques carry over to fluctuation-dissipation theorems and nonequilibrium ensemble reweighting techniques as should be expected. We also show that the thermodynamic entropy production extends the generalized fluctuation theorem through the addition of an instantaneous information entropy term for the end-points, leading to a concise statement of the second law of thermodynamics.

---

* dmroge@sandia.gov

† slrempe@sandia.gov

I. INTRODUCTION

The Green-Kubo fluctuation theorems[1] relate equilibrium time-correlation functions with the time-response of system observables, a, to an external driving force. They are important for their ability to calculate the transport coefficients, L, appearing in Onsager's phenomenological relation,[2]

$$\langle \dot{a}_i(t) \rangle \approx \sum_{ij} L_{ij} X_j(t), \tag{1}$$

which identified X_j with the "entropic" driving forces, $\partial S / \partial a_j \approx \beta(t) - \beta_{\mathrm{eq}}$.[3]

The first law of thermostatics relates quasi-static changes in energies to their conjugate thermodynamic forces. By analogy, we should expect a first law of nonequilibrium thermodynamics that relates thermodynamic forces not completely balanced by the system to fluxes, for which the above relations form a rough draft. Combined with the usual second law prescription of increasing entropy, Onsager's relations establish the direction in which the system will relax toward equilibrium and an estimate of the entropy increase by this process. However, the linear transport equations have been derived by analogy with the equilibrium theory, and their interpretation must be made with respect to the entropy of a quasi-steady process. A more satisfactory development would therefore determine the range of conditions for which Eq. 1 holds, as well as provide a foundation for studying processes with arbitrary driving forces and defined without reference to any equilibrium state. If, in addition, the theory was able to resemble the well-known equilibrium statistical mechanics, it would offer a wealth of immediate insight into new applications for which now standard nonequilibrium methods may prove cumbersome and error prone. Such a resemblance must be reached by defining path functionals analogous to the energy and entropy of equilibrium states, and would therefore constitute a true statistical mechanics for thermodynamics (as opposed to thermostatics). Our first question in this investigation will be how such a microscopic first law of thermodynamics can be formulated.

Although much work has been devoted to these problems, these questions have been addressed from a large number of different viewpoints in the last sixty years and remain open.[1, 4–13] Further, there appears to be little consensus on a unifying, general, set of relations from which all nonequilibrium results may be derived.[14] Recent work has centered on a second question of deriving a microscopic second law of thermodynamics through proving the existence of fluctuation theorems[15] and exploring their consequences for nonequilibrium

systems.

A fundamental fluctuation theorem result is an expression for the "lost work" over and above the equilibrium free energy for stochastic processes that convert one thermostatic state to another[16] via mechanical driving. This lost work can be interpreted as an entropy increase. It is simple to show in the case of an isolated deterministic system[17] evolving from x_0 through the set of states $\{x\}_0^S$ to x_S at time S,

$$e^{\beta_0(W(\{x\}_0^S) - \Delta F)} = \mathcal{P}_0(x_0) / \mathcal{P}_S(x_S(x_0)).$$

Here the work, W, is the gain in system energy $\Delta U = U_S(x_S) - U_0(x_0)$, and both distributions, $\mathcal{P}(x)$, are required to be at the same temperature, k_B/β_0, to use the equilibrium relation $\beta_0 \Delta F = -\ln Z_S/Z_0$, with Z the thermostatic partition function.[18] Further work has provided examples of many fluctuation theorems.[14, 19] Each can be used to define a measure of irreversibility since for any two path probabilities, A and B, we can define a likelihood ratio,

$$e^{l(\{x\}_0^S)} \equiv \mathcal{P}_B(\{x\}_0^S) / \mathcal{P}_A(\{x\}_0^S) \tag{2}$$
$$\Rightarrow \langle l \rangle_B = \mathcal{D}[B|A] \geq 0,$$

where $\mathcal{D}$ is the Kullback-Leibler divergence (and necessarily positive). Although special significance is often attributed to time reversal (defined by replacing the time evolution operator from i to $i+1$ with a reversal of odd functions of time at step $i+1$, unaltered time evolution from x_{i+1}^* to x_i^*, and another reversal of odd functions at i),[12] this operation is not possible if one-way steps are present, so that only one of $i \to i+1$ or $i+1^* \to i^*$ has probability zero. Furthermore, the extension of this formalism to systems lacking momentum, such as discrete processes described by a transition probability matrix, is also unclear although suggestions involving a (possibly non-unique) stationary distribution have been offered[15]. Until these issues are resolved, there does not exist a definitive path functional able to answer our first question of finding an analogous first law obeyed by nonequilibrium systems. Could the thermostatic energy turn into total work, and the entropy functional something like dissipated energy?

Early works were principally focused on this first question. After a detailed picture of how time-correlations control the rate of relaxation to equilibrium[20, 21], a general set of relations were described using projector-operator theory[6, 7, 22, 23] from which the rate of

macroscopic relaxation may be derived. The principal content of the theory was to define a coarsening, or projection operator that removes information about the unmodeled degrees of freedom. The exact, non-Markovian kinetic equation for the probability distribution in the space of remaining variables then shows the relaxation process in the form of a time convolution of the time-correlation functions and the thermodynamic forces driving them to equilibrium.[22, 23] That is, Eq. 1 should be replaced with

$$\langle \dot{a}_i(t) \rangle \approx -\int_0^t \sum_{ij} L_{ij}(t-\tau)X_j(t-\tau)d\tau, \qquad (3)$$

where $L_{ij}(s) = \langle \dot{a}_i(t)\dot{a}_j(s)\rangle$. Despite conjectured relationships to a nonequilibrium entropy function,[4, 6, 24] its relation to the thermostatic entropy has not yet been fully justified in terms of the maximum entropy formalism. The work of Zubarev[9] and others[25] on defining Boltzmann-Gibbs distributions approximating transient distributions made some advances in applying the quasi-static perturbation theory further out of equilibrium. However, transient dynamics has not been clarified to the point where it is possible to derive Eq. 3 from a second derivative as sketched by Jaynes[8] – analogous to the Gibbs relations for equilibrium which give rise to obviously related quantities such as heat capacity, compressibility and coefficient of thermal expansion.

At this point, the situation shared a peculiar similarity to the circumstances surrounding Gibbs' classical text introducing the principle of maximum entropy.[26] The method for deriving fluctuation dissipation theorems (analogous to the first law) was to define a system evolving according to an exact Lagrangian, make a random phase approximation to yield an ensemble of exactly evolving trajectories, and then derive a corresponding "physical" distribution on trajectory space (analogous to phase-space, Table I). Because of the prevailing attitude regarding mechanics as the only possible method for solving thermostatic problems, Gibbs' use of maximum entropy methods was seen as a non-physical trick[27] to derive properties of molecular equilibrium. Again, introducing maximum entropy led to an expansive generalization of the fluctuation dissipation theorem by Jaynes[8, 28] – who, it should be noted, has recognized and written about this conflict of Gibbs[29].

Jaynes introduced maximum entropy following the program of Gibbs and using the subjective interpretation of probability used by Laplace and Jeffreys.[28, 30] Making the substitution from a state, containing a variable at a single time, to an entire trajectory immediately identifies a nonequilibrium analog of the Gibbs ensemble, complete with an information en-

tropy and free energy functional on trajectory space. Pursuing the analogy further, Jaynes showed that the first derivatives of the path free energy yielded averages of path functionals, and the second derivatives (minus the space and time-correlation functions) give their "first-order" perturbation with respect to changing the thermodynamic forces, λ[8, 29, 31]. Therefore we can shorten Eq. 3 to

$$\langle a \rangle \approx \langle a \rangle_0 - \sum_k \langle (a - \langle a \rangle_0)(b_k - \langle b_k \rangle_0) \rangle_0 (\lambda_k - \lambda_k^0) \tag{4}$$

for *any* path functional a we wish to average and *any* set of "controlled" quantities b_k that define our trajectory ensemble. For example, if b_k included time and space-dependent particle fluxes (i.e. k indexes both time and space) so that $\sum_k \to \iint dtdx$, then Eq. 4 easily generalizes Eq. 1 to fluctuation relations defined directly from the set of constrained path functionals, $\langle b_k \rangle$. Notice that the perturbation expansion above is not defined by reference to an equilibrium state, but instead with respect to a reference probability distribution on path space. Elegant derivations of fluctuation theorems have been given many times in works employing path integral methods[32], and can be related to the Onsager-Machlup action[4, 33]. Further work[12, 34, 35] has also shown how Jaynes' path entropy may be connected to fluctuation and entropy production theorems. The connection of this approach to discrete master equations has been explored in Refs. [36, 37].

Unfortunately, the argument leading to Eq. 4 contains a fatal flaw. To see this, first note that the average of a quantity at time t is dependent on the "control" parameters throughout the whole trajectory. This is because logical inference does not contain a preferential time direction. Knowledge that the system has a given property at time $t + \tau$ constrains the state at time t, as it should, but a force λ_k acting at a future time should not be able to influence the past trajectory. Although this could be alleviated by requiring only casual information to enter into the determination of the state at time t, this approach leads to a probability distribution valid only for x_t and not throughout a trajectory. A better approach is to maximize the entropy of the transition probability distribution, an approach argued for on average in the long-time limit by Ref. [38].

Limiting the scope of the maximum trajectory entropy procedure in this way automatically corrects a related re-normalization problem. Suppose phase space were to branch at a future time $t + \tau$. In this case, a uniform measure on path space would assign points at time t a different weight depending on future events. A simple example is the Monty

Hall problem with a prize assigned to the first door without loss of generality.[39] The contestant's choice of one of three doors plus Monty opening another (not concealing a prize) constitutes four possible paths, and maximum path entropy would give each path an equal weight – an intuitive, but incorrect, solution. The correct probability assignment is a uniform distribution for each *transition*, leading to ending weights of $1/3 \times 1/2 = 1/6$ for the two paths following the selection of the first door[?]. In more abstract terms, the marginal probability at each time should not depend on the future – a concept expressed mathematically by defining a progressively measurable function with respect to the natural filtration of stochastic processes[40].

In practice, this flaw can be avoided by considering only processes where assigning equal *a priori* weight to all paths is equivalent to assigning equal weight to all transitions. By Liouville's theorem, this is obviously true for deterministic processes. More generally, the approaches are equivalent when the number of possible transitions does not depend on the starting point. The path entropy approach of Jaynes is therefore valid in the absence of factors re-normalizing for starting-point dependent differences in the number of possible paths ($Z[\lambda_i, x_i] = Z[\lambda_i]$ in Eq. 8).

A further question remains on the application of fluctuation-dissipation theorems to both Langevin and Brownian (overdamped Langevin) processes. What is the appropriate order for coarse-grained equations of motion? Writing down a Langevin equation assumes Newtonian, second-order dynamics, whereas the first-order Brownian motion can also be derived via the same approach. Although the Green-Kubo relations were supposed to have solved these problems, ambiguity remains in this approach at a fundamental level because a strict derivation using the method of mechanics does not offer direct insight into the choice of macroscopic variables used for systematic coarse-graining. In fluid mechanics, either can be applied, and the choice between full Newtonian motion models or simplified advection-diffusion models is predictably based on the scales of length, relaxation time, and applied force involved. Several authors have made substantial contributions[41–44] in connecting fluid mechanics and nonequilibrium dynamics.

Although seemingly incompatible, there are advantages to both the mechanics-based and the statistical derivations. For example, considering the choice of equations of motion from the statistical perspective of Equation 4 at once allows us to see the consequences of each choice. If we choose a first-order equation of motion, then we substitute the velocity at a

point for a and expand it about a streaming velocity $\langle a \rangle_0$ in the particle fluxes, $b_j(x,t)$. If second order, then a becomes a change in momentum at a point and we expand about the average force, $\langle a \rangle_0$, in the stress tensor of the surrounding fluid, $b(x,t)$. We believe that further corrections sometimes employed in fluid mechanics may also be derived via extending the process that led to Eq. 4. If both the nonequilibrium first and second laws could be addressed from the same perspective, it seems that an expansive generalization of nonequilibrium statistical mechanics could be achieved.

In order to combine these two viewpoints, we introduce the following maximum entropy argument. Suppose the information about the time change of a set of dynamic variables, x, consists in a set of prescribed averages, $\{\langle f_k(x_{i+1}, x_i)|x_i\rangle\}_{k=1}^m$. Using only this information, we seek to construct a probability distribution for the dynamic variables at point $i+1$, given a known x at time i. According to the standard maximum information entropy ($\mathcal{H}$) machinery,[30] we set the variation of the transition entropy to zero under constraints for $\langle 1|x_i\rangle$, $\langle f_k(x_{i+1}, x_i)|x_i\rangle$ using Lagrange multipliers $Z[\lambda, x_i], \lambda_{k,i}[x_i]$

$$
\begin{aligned}
\delta \mathcal{H}_{i+1|i} = \delta \sum_{\{x_{i+1}\}} p(x_{i+1}|x_i) \Big[&- \ln \frac{p(x_{i+1}|x_i)}{p^0(x_{i+1}|x_i)} \\
&- (\ln Z[x_i] - 1) - \sum_{k=1}^m \lambda_{k,i} f_k(x_{i+1}, x_i) \Big]
\end{aligned}
\tag{5}
$$

to find

$$
p(x_{i+1}|x_i) = p^0(x_{i+1}|x_i) e^{-\eta_i[x]} / Z[\lambda_i, x_i]
\tag{6}
$$

$$
\eta_i[x] \equiv \sum_{k=1}^m \lambda_{k,i} f_k[x_{i+1}, x_i]
\tag{7}
$$

$$
Z[\lambda, x_i] = \sum_{\{x_{i+1}\}} p^0(x_{i+1}|x_i) e^{-\eta_i[x]}
\tag{8}
$$

$$
\mathcal{H}_{i+1|i} = \ln Z[\lambda, x_i] + \langle \eta_i | x_i \rangle
\tag{9}
$$

$$
d\mathcal{H}_{i+1|i} = \sum_{k=1}^m \lambda_k d\langle f_k | x_i \rangle.
\tag{10}
$$

These are the expressions relating to the statistical state at time $i+1$ given information on the transition probability distribution. The functional notation for quantities such as $\eta_i[x]$ has been used to indicate that, in general, these may be considered as functionals depending on the trajectory over all times at or before $i+1$. In the continuous limit, these quantities

8

exist between times i and $i+1$, and should be viewed in the Stratonovich definition. The appendix shows example calculations of the partition function for the Wiener process.

The essential difference between Eq. 6 and the maximum path entropy prescription is the role of the normalization constants. In the latter, the path probability is

$$\mathcal{P}\left(\{x_i\}_0^S\right) = \frac{e^{-\sum_{i=0}^{S-1} \eta_i[x_{i+1};x_i] - \beta U(x_0)}}{Z} \mathcal{P}_0\left(\{x_i\}_0^S\right),$$

with normalization

$$Z = \int \cdots \int e^{-\sum_{i=0}^{S-1} \eta_i[x_{i+1};x_i] - \beta U(x_0)} \mathcal{P}_0\left(\{x_i\}_0^S\right) \, d\{x_i\}_1^S.$$

This is replaced by individually normalized transitions,

$$\mathcal{P}\left(\{x_i\}_0^S\right) = \prod_{i=0}^{S-1} \frac{e^{-\eta_i[x_{i+1};x_i]}}{Z_i[x_i]} \frac{e^{-\beta U(x_0)}}{Z_0} \mathcal{P}_0\left(\{x_i\}_0^S\right),$$

in the former. The present conditional expansion implies a dual characterization of a stochastic process as a single ensemble of paths and a set of telescoping ensembles, each with well-defined, non-anticipating energy changes through $\Delta F_k = \sum_{I=0}^{J-1} \langle f_k[x_{i+1}, x_i] \rangle$.

Equilibrium	Nonequilibrium
Phase Space	Trajectory Space
Free Energy	Path Free Energy
Entropy	Caliber
Average Value Constraint	Average Flux Constraint
Equilibrium Average	Path Average
Conditional Free Energy (PMF)	Conditional Path log-Probability
Thermodynamic forces (δPMF)	Changes in Path Flux (work)
Heat Capacity	Thermal Conductivity
(none)	Irreversibility and Entropy Production

TABLE I. Correspondence between single-time and time-dependent path maximum entropy formulations of statistical mechanics.

Using this form of maximum transition entropy provides several distinct advantages for the derivation of nonequilibrium relations. Not least is the correspondence to the canonical,

maximum entropy form of equilibrium thermostatics pointed out in Table I. We begin by deriving the generalized Langevin and Brownian dynamics from a consideration of the action deviation as a constrained quantity. In Sec. III, we connect the Langevin equations derived from our stochastic action deviation principle to energy exchange and thermodynamic entropy production via interaction with external reservoirs.[5] We then show that this method extends the fluctuation theorem by including the instantaneous information entropy to give an exact form and an informational interpretation to the second law of thermodynamics. The coefficient of thermal expansion, isothermal compressibility, and heat capacity can be derived using second derivatives of the equilibrium free energy. In an analogous way, we show how Green-Kubo transport theory can be derived from second derivatives of Legendre transforms of the maximum transition entropy functional. These relations can describe the response of both steady- and nonsteady-states arbitrarily far from equilibrium. Applications of this result enable calculation of derivatives of the current-voltage curve at constant current or constant voltage and give general conditions under which Onsager reciprocity will hold.[28] A detailed numerical example is developed for the role of information in an optically trapped bead pulling experiment.[14]

II. DYNAMIC CONSTRAINTS

Starting from a mechanics problem specified by Lagrangian, $L(x, \dot{x})$, the usual mechanical prescription is to require stationary action:

$$\mathcal{A} \equiv \int_0^S L(x, \dot{x}) dt$$

$$\delta \mathcal{A} = \int_0^S \frac{\partial L}{\partial x} \delta x + \frac{\partial L}{\partial \dot{x}} \delta \dot{x} \; dt$$

$$= \frac{\partial L}{\partial \dot{x}} \delta x \Big|_0^S + \int_0^S \left(\frac{\partial L}{\partial x} - \frac{d}{dt} \frac{\partial L}{\partial \dot{x}} \right) \delta x(t) \; dt$$

$$\Rightarrow \frac{\delta \mathcal{A}}{\delta x(t)} = \frac{\partial L}{\partial x} - \frac{d}{dt} \frac{\partial L}{\partial \dot{x}} = F - \dot{p}. \tag{11}$$

In this report, we will freely substitute force, $F \equiv \frac{\partial L}{\partial x}$, and momentum, $\dot{p} \equiv \frac{d}{dt} \frac{\partial L}{\partial \dot{x}}$. The requirement for stationary action then reads $\frac{\delta \mathcal{A}}{\delta x(t)} \overset{\text{set}}{=} 0 \Rightarrow F = \dot{p}$.

When $L = \dot{x}^T M \dot{x}/2 - U(x)$, the above procedure directly gives Newtonian mechanics and has the advantage of being generally valid under coordinate transformations, $y = y(x), x =$

$x(y)$. However Eq. 11 gives second-order equations of motion, requiring $\dot{x} = dx(t)/dt$ by definition, and complicating discussions of numerical integration. A first-order form can be derived from an alternate Lagrangian,

$$L(\begin{bmatrix} q \\ v \end{bmatrix}, \begin{bmatrix} \dot{q} \\ \dot{v} \end{bmatrix}) = \left(\dot{q} - \tfrac{1}{2}v\right)^T Mv - U(q), \tag{12}$$

by treating $q(t)$ and $v(t)$ as separate dynamic variables. Using $x(t) = \{q(t), v(t)\}$ in (11), we find

$$\frac{\delta\mathcal{A}}{\delta q} = -\frac{\partial U}{\partial q} - M\dot{v} \tag{13}$$

$$\frac{\delta\mathcal{A}}{\delta v} = M(\dot{q} - v). \tag{14}$$

Setting these two equal to zero gives a result entirely equivalent to Newtonian mechanics, but in which we may consider a Verlet-type integration process of updating v with fixed q, and then updating q with v fixed. It has been found that Eq. 11 forms a solid basis for forming generalizations of physical laws[45] as well as deriving variational integrators with good conservation properties.[46]

A stochastic generalization may be to permit small deviations by constraining $f = \frac{\delta\mathcal{A}}{\delta x_i(t)} \frac{\delta\mathcal{A}}{\delta x_j(t-\Delta t)}\epsilon$ in Eq. 6 using Lagrange multipliers $G_{\Delta t}, \Delta t \in [0, \infty)$. The resulting matrices of constraint values, G, can restrict deviations in the action in a history-dependent and non-local way. Each deviation in the action can be thought of as arising from elastic collisions with un-modeled molecules from the surrounding 'bath' environment or as an unknown Lagrangian applied to the system between times $t - \epsilon$ and t[47] (with ϵ a small time increment). Without other constraints, this squared-deviation constraint implies interaction with a completely chaotic (infinite temperature) bath and does not conserve, for example, energy and momentum except in the deterministic limit ($\mathbf{G} \to diag(\infty)$).

When two systems are coupled, the combined system should obey a set of conservation laws. As formalized by Noether's theorem,[48] such conservation laws can be derived for a single system directly from the action formulation by considering continuous transformations of the trajectory $x(t) \to x(t) + q(t,a)$ in a region around $a = 0$, where $q(t,0) = 0$. Because the action is stationary with respect to small perturbations in $x(t)$, there exists a vanishing

11

quantity,

$$\frac{dI(t)}{dt} \equiv - \left.\frac{\partial \mathcal{A}}{\partial a}\right|_{a=0}$$

$$= - \left.\frac{\partial q(t,a)^T}{\partial a}\right|_{a=0} \frac{\delta \mathcal{A}}{\delta x(t)}, \tag{15}$$

(with T reserved for denoting transposition) and it is possible to define an invariant (using $\left.\frac{\partial q(t,a)}{\partial a}\right|_{a=0} \equiv y(t)$),

$$I(t) - I(\infty) = \int_{-\infty}^{t} -y^T(t')\frac{\delta \mathcal{A}}{\delta x(t')}dt'.$$

If the action functional is invariant to this transformation ($\mathcal{A}[x(t)] \to \mathcal{A}[x(t) + q(t,a)]$), then the corresponding invariant is a conventional conserved quantity. The substitution $x(t) \to x(t) + a$ ($y = \mathbb{1}$, the ones-vector) generates the momentum, $x(t) \to x(t + a)$ ($y = \dot{x}$) generates the energy, and so on.

It is instructive to consider a case where the action is not invariant to the transformation. For example, the form of the action is changed using the substitution for momentum in the simple harmonic oscillator ($L = m\dot{x}^2/2 - kx^2/2 \neq m\dot{x}^2/2 - k(x + a)^2/2$). However, if we assume the existence of a generalized momentum, $\hat{p}$, that is nonetheless an invariant of some 'complete' system, we can define (using Eq. 15)

$$\frac{d\hat{p}(t)}{dt} = \dot{p} - F = m\ddot{x} + kx$$

as a momentum *exchange*. If the net momentum of the observed system changes by more than its internal force, it implies that an external system has lost exactly this amount of momentum. In the presence of noise, this quantity may not be conserved ($d\hat{p}$ is stochastic). Thus one effect of the noise is to re-distribute $\hat{p}$ over the system and the bath.

In general, we can define a set of constraints on $\{\langle dI_j(t)/dt\rangle\}_{j=0}^{m}$ using an $N \times m + 1$ matrix Y, whose columns correspond to the constraint directions $\left.\frac{\partial q_j(t,a)}{\partial a}\right|_{a=0}$ and whose leftmost column is reserved for the energy $Y_{*,0} = \dot{x}$. In this case, the vector of exchanges for a given trajectory over a given time interval, ϵ, is

$$dI(t) \equiv -Y(t)^T \frac{\delta \mathcal{A}}{\delta x(t)} \epsilon. \tag{16}$$

In this equation, the presence of ϵ is used to denote the Stratonovich integral implicitly (see Appendix). The selection of constraints must be made based on the allowed interactions of the system with its reservoir.

12

These considerations have shown a simple method for including the influence of an external system on the dynamics. If we assume the existence of some "total" invariant between the system and the bath, then it makes sense to enforce a stochastic constraint on the average change, $\langle dI(t) \rangle$. We should note our fixed sign convention, where I is always taken to be a quantity belonging to the central system under consideration. Before moving on to discuss the obvious connection of these changes to the thermodynamic work, we shall first consider the significance of the Lagrange multipliers in this formalism.

Collecting constraints on $\frac{\delta\mathcal{A}}{\delta x(t)} \frac{\delta\mathcal{A}}{\delta x(t')} \epsilon$, dI and carrying out the maximum entropy procedure specified in Eq. 5 for determining $\frac{\delta\mathcal{A}}{\delta x(t)}$ for each time, t, given a history $x(t')$ for $t' \leq t - \epsilon, \epsilon \rightarrow 0^+$, we arrive at a transition probability

$$
\begin{aligned}
\mathcal{P}\left(\tfrac{\delta\mathcal{A}}{\delta x(t)}\epsilon \Big| x(t')_{t'=0}^{t-\epsilon}\right) &= \exp\left[-\epsilon \tfrac{\delta\mathcal{A}}{\delta x(t)}^T G \tfrac{\delta\mathcal{A}}{\delta x(t)} - \tfrac{\delta\mathcal{A}}{\delta x(t)}^T \epsilon \int_0^{t-\epsilon} G_{t-\tau} \tfrac{\delta\mathcal{A}}{\delta x(\tau)} d\tau + \epsilon \tfrac{\delta\mathcal{A}}{\delta x(t)}^T Y(t)\beta/2 \right] \\
&\quad \times Z[\beta, x(t')_{t'=0}^{t-\epsilon}]^{-1} \\
&= \exp\left[-(\tfrac{\delta\mathcal{A}}{\delta x(t)}\epsilon - \mu(t))^T \tfrac{G}{\epsilon}(\tfrac{\delta\mathcal{A}}{\delta x(t)}\epsilon - \mu(t)) + \mu(t)^T \tfrac{G}{\epsilon}\mu(t)\right] \\
&\quad \times Z[\beta, x(t')_{t'=0}^{t-\epsilon}]^{-1} \\
C &\equiv (2G)^{-1} \\
\mu(t) &\equiv C\left(Y(t)\beta\epsilon/2 - \epsilon \int_0^{t-\epsilon} G_{t-\tau} \tfrac{\delta\mathcal{A}}{\delta x(\tau)} d\tau\right).
\end{aligned} \tag{17}
$$

In this equation, $Y\beta$ is a vector with the dimension of the system coordinates, since $\beta = [\beta_0, \ldots, \beta_m]^T$. The action deviation, $\frac{\delta\mathcal{A}}{\delta x(t)}\epsilon$, then follows a normal distribution with mean $\mu(t)$ and single-time variance/covariance matrix $C\epsilon$. The history integral in Eq. 17 could alternatively be written in terms of a time-dependent covariance function. In this work, however, we will not be concerned with the calculation of history-dependent partition functionals for which this transformation becomes useful. Note that an external force, $CY\beta/2$, shifts the average of $\frac{\delta\mathcal{A}}{\delta x(t)}$ from zero to μ, reminiscent to the Lagrange-D'Alembert principle.

Enforcing a constraint for the Hamiltonian energy change $dH = -\frac{\delta\mathcal{A}}{\delta x(t)}^T \dot{x}\epsilon$ using the Lagrange multiplier $\beta_0/2$, along with a possibly empty set of additional constraints ($\tilde{Y}$) in the form of Eq. 16, leads directly to a Generalized Langevin equation by substitution of the mean and variance from Eq. 17 into

$$\tfrac{\delta\mathcal{A}}{\delta x(t)}\epsilon = \mu(t) + (C\epsilon)^{1/2} z(t),$$

yielding

$$\dot{p}\epsilon = F(x(t))\epsilon - C\left(\dot{x}\beta_0\epsilon/2 + \tilde{Y}\tilde{\beta}\epsilon/2 + \epsilon\int_0^{t-\epsilon}G_{t-\tau}(\dot{p}(\tau) - F(\tau))d\tau\right) + C^{1/2}d\mathcal{W}(t).$$

(18)

Here the centered (Stratonovich) Wiener process increment has been substituted for the standard normal random variate, $z(t)$, at time t using $d\mathcal{W}(t) = \epsilon^{1/2}z(t)$. It is well-known from the Fokker-Planck equation[49],

$$\frac{\partial \rho(x,p)}{\partial t} = -(M^{-1}p)^T\nabla_x\rho - \nabla_p^T\left[F\rho - C\dot{x}\beta_0\rho/2 - C\nabla_p\rho/2\right],$$

that the solution to this equation is the canonical distribution with temperature $\beta_0^{-1} = k_BT$, where k_B is the Boltzmann constant. See also Ref. 50 for the formulation in the presence of memory terms. This solution is independent of G, suggesting that a natural parametrization for the Langevin equation is in terms of the temperature and $C^{1/2}$, which is related to the thermal conductivity or rate of temperature equilibration (see Eq. 19). Generally, if an invariant, I, can be expressed as a function of x and p, then a similar result should hold for other common equilibrium thermodynamic ensembles as well, such as the N, P, T ensemble where $\langle dV \rangle$ constitutes an additional dynamic variable and constraint.[51] This establishes the physical interpretation of the Lagrange multipliers as the thermostatic variables of the bath that dictate the eventual equilibrium of the system. Note that increasing β tends to decrease $\langle dI \rangle$, for example increasing pressure will drive the volume downward.

The present work shares some conceptual similarity to the second entropy of Attard [52, 53]. In this report, however, we have been able to derive our results in a mathematically rigorous way directly from two extremum principles: a maximum entropy expression for the transition probability (Eq. 6), and constraints derived from an action functional (Eq. 11). This allows for trivial generalizations to systems coupled with arbitrary reservoirs. In addition, there is a clear physical motivation for the transition entropy and the full nonequilibrium entropy production that allows us to find the work done on the system by each constraint, as will be shown in the next section.

The constrained dynamical equations are also sufficient for defining nonequilibrium analogues of intensive thermodynamic variables such as the temperature. This can be done by adding a hypothetical constraint, $\langle dI_j \rangle$, defined for some set of atoms or region of space in the system. Analogous to the operation of a thermometer – where zero energy is exchanged

14

on imposing stochastic and damping terms – to define temperature, we then require that no work is done on average, $\langle dI_0 \rangle = 0$. Integrating using the Stratonovich rules developed in the appendix, we find

$$\langle dI_0 \rangle / \epsilon = \frac{1}{2}[\text{Tr}(M^{-1}C) - \beta_0 \langle \dot{x}_i^T C \dot{x}_i \rangle] \tag{19}$$

The result is the intuitive kinetic temperature. It is especially simple if we choose $C = 2M\gamma/\beta_0$ as is common for the Langevin equation. In that case, the ensemble average kinetic energy at each instant determines the temperature. For Boltzmann-distributed $\dot{x}$, regardless of the choice of C, the average heat flow is zero when β_0 determines the temperature. If different types of particles can be coupled to separate thermostats, such as in plasmas, then it becomes physically meaningful to speak of separate ionic and electronic temperatures.

To end this section, we show that it is possible to derive a Brownian limit from our action functional approach using the alternative Lagrangian, Eqns. 12-14. First, it can be shown that changes in the Hamiltonian are recovered by applying Eq. 16.

$$-\dot{q}^T \frac{\delta \mathcal{A}}{\delta q(t)} - \dot{v}^T \frac{\delta \mathcal{A}}{\delta v(t)} = \dot{q}^T(\frac{\partial U}{\partial q} + M\dot{v}) - \dot{v}^T M(\dot{q} - v)$$

$$= \frac{d}{dt}[U(q) + v^T M v / 2]$$

Next, applying Eq. 17 and assuming $\frac{\delta \mathcal{A}}{\delta q(t)} = 0$ exactly, we find a combined equation,

$$M\dot{v}\epsilon = F(q(t))\epsilon,$$

$$\dot{q}\epsilon = v\epsilon + C_q \left(\dot{p}\beta_0\epsilon/2 - \int_0^{t-\epsilon} G_{q,t-\tau}(\dot{q}(\tau) - v(\tau))d\tau \right) + C_q^{1/2} d\mathcal{W}(t). \tag{20}$$

The second equation has the form of a Generalized Brownian motion equation, but includes terms related to the process $v(t)$. In particular, if the process $v(t)$ becomes unknown, then the best guess, $M\dot{v} = F(x)$, and a streaming velocity $v = v_0$ generate an appealing equation for Brownian motion. More rigorously, if the process $v(t)$ is assumed to be unknown, updates $\dot{q}\epsilon$ should be made based on a stochastic realization of $v(t)$, whose average will generate the streaming velocity $\langle v(t) \rangle = v_0$.

To the best of the authors' knowledge, this is a novel derivation of the Brownian limit that does not require an explicit limiting process of an infinitely massive particle, or infinite momentum jumps between position changes. Instead, these two assumptions are implicitly present in assuming that $v, M\dot{v}$ are known during each position update. The memory term derived here is similar to the form postulated in Ref. 54, which led to a quantitative treatment

15

of memory effects. Here we can see it to be a natural consequence of placing constraints on the squared deviation of the action and the energy change at each time-step.

III.   IRREVERSIBLE THERMODYNAMICS

Having firmly established the connection between the constrained dynamical equations and the equilibrium distribution above, we next construct a high-level view of any process that employs a series of transition probabilities to effect a change in the state of the system. This construction will lead naturally to a view of the process in terms of a thermodynamic path transforming one type of energy into another with a concomitant irreversible entropy production.

To begin, we exactly define a system state, A, as any information known about a system that is sufficient to construct a probability distribution for its variables, $\mathcal{P}(x|A)$. The machinery of statistical mechanics can then be used to propagate this information to system states at other time points and under alternate possible processes.

The work of Joule and Thomson showed that there exists a series of mechanical operations that can be performed to effect a transition between any two thermostatic states;[3] however, this transition can only take place in the direction of increasing entropy. The entropy increase comes about because of experimental inability to control the detailed motions of all particles, and is therefore zero in the case of completely controllable mechanical work. Thus, it is important to define a mechanical, adiabatic process in which all work is completely controlled by letting $C \to 0$ with constant external force experienced by the system $F_{\text{ext},j}^{\text{sys}} \equiv -\frac{C}{2}y_j\beta_j \to \frac{\partial x(t)}{\partial I_j(t)}F_{\text{ext},j}(t)$. At a finite temperature, there is some amount of uncertainty about the exact state of I on short time scales. This leads to a discrepancy between the force exerted by the external system and its "long time scale" counterpart experienced by the system. We should therefore view the mechanical, adiabatic process formally as a zero-temperature continuous-time limit. In general, only a subset of work values can be controlled. Before proceeding it will be necessary to solidify the concept of controllable work.

If this work is to be delivered by an external thermostatic system, for example an adiabatically coupled piston, then the first law of thermostatics gives $F_{\text{ext},j} = -dU_{\text{ext}}/dI_j = dU_{\text{ext}}/dI_{\text{res},j} = -\beta_j/\beta_0$ if the force can be assumed constant over a sufficiently short time-step. Mechanically, this force corresponds to the force on a wall exerted by a spring placed

externally to it. The total force on the wall is, of course,

$$F_{\text{tot},j} = -\frac{dU_{\text{int}} + dU_{\text{ext}}}{dI_j} \equiv F_{\text{int},j} + F_{\text{ext},j}.$$

This implies that if a two spring system were disconnected after a change $d\tilde{I} \equiv [dI_1, \ldots, dI_m]^T$, their internal potential energies would have changed by an amount

$$dU_{\text{int}} = -d\tilde{I}^T F_{\text{int}}$$
$$dU_{\text{ext}} = -d\tilde{I}^T F_{\text{ext}} = d\tilde{I}^T \tilde{\beta}/\beta_0 \qquad (21)$$
$$\tilde{\beta} \equiv [\beta_1, \ldots, \beta_m]^T.$$

The sum of these two energy changes is not necessarily zero due to the possibility of momentum change. Using the known energy change of the system, it should then be possible to solve for the change in kinetic energy of the constraint. For the system, the *total* energy change is given by

$$dE = -\frac{\delta\mathcal{A}}{\delta x}^T \dot{x}\epsilon = dI_0$$
$$= dW + dQ$$
$$dW_j \equiv -\frac{\delta\mathcal{A}}{\delta x}^T y_j \dot{a}_j \epsilon$$
$$dW = d\tilde{I}^T \dot{a} \qquad (22)$$
$$dQ = -\left(\dot{x} - \tilde{Y}\dot{a}\right)^T \frac{\delta\mathcal{A}}{\delta x}\epsilon, \qquad (23)$$

where we have used Eq. 15 for $\frac{\partial\mathcal{A}}{\partial a}$, and multiplied by da/dt to define work values and a corresponding "non-mechanical" energy transfer, dQ. In the adiabatic limit, all energy transferred to the system by external forces should be reflected by known mechanical changes (related to $\{I\}$), which is precisely what $\dot{a}$ allows us to do. Note also that all quantities refer to the central system unless specifically denoted 'ext.' Thus dW means the work done on the system.

In the mechanical limit, $-\frac{\delta\mathcal{A}}{\delta x(t)}$ is just $\dot{p} - F_{\text{int}} = F_{\text{ext}}^{\text{sys}}$, the external force experienced by the system in the absence of the thermostatting random noise, and $Y_{*,0} = \dot{x}$. Substituting this quantity from the Langevin equation (18), an adiabatic, mechanical system must satisfy

$$dQ = -(\dot{x} - \tilde{Y}\dot{a})^T \frac{\delta\mathcal{A}}{\delta x}\epsilon$$
$$= -\frac{1}{2}(\dot{x} - \tilde{Y}\dot{a})^T C\tilde{Y}\tilde{\beta}\epsilon = 0. \qquad (24)$$

17

The last section gave some physical insight into the quantities $\dot{a}$. A mathematical consideration of the previous equation shows that $\tilde{Y}\dot{a} \equiv \sum_{j=1}^{m} y_j \dot{a}_j$ ($\tilde{Y}$ being identical to Y with the first column removed) can be understood as a projection, removing components of $\dot{x}$ parallel to $\delta\mathcal{A}/\delta x$. If we therefore define (writing the Moore-Penrose pseudoinverse of A as A^+)

$$\dot{a} \equiv (C^{-1/2}\tilde{Y})^+ C^{-1/2}\dot{x}, \tag{25}$$

then $dQ = 0$ identically. For an example application, $y = \mathbb{1}$ (ones-vector) is associated with the system momentum, and Eq. 25 generates the average velocity $\dot{a} = \mathbb{1}^T\dot{x}/N$ when $C = cI$. Eq. 25 is invariant to multiplication of C by a constant, and so persists in the deterministic limit. The work done on the system (Eq. 22) is

$$dW = -\frac{\delta\mathcal{A}}{\delta x}^T \tilde{Y}(C^{-1/2}\tilde{Y})^+ C^{-1/2}\dot{x}\epsilon.$$

Note the information-theoretic quality of the work defined by the equation above. If separate reservoirs existed that were able to influence the motion of each particle in the system independently, then $\tilde{Y}$ would become an identity matrix, and $d\mathcal{W}$ would equal dE identically. In the presence of noise, the work taken from the thermostatic reservoirs does not necessarily equal the work done on the system. In the $y = \mathbb{1}$ example above, the momentum work done on the system is

$$\frac{1}{N}(\sum_a \dot{p}_a - F_a)(\sum_a \dot{x}_a)\epsilon. \tag{26}$$

Here "$\sum_a$" denotes a sum over N (one-dimensional) particles, with obvious extension to multiple dimensions. Because this interaction controls only the total system momentum, the work is computed using the average velocity change. As shown in the appendix, the expectation of this stochastic integral is $\langle dW_p\rangle/\epsilon = \frac{\mathbb{1}^T\dot{x}}{N}\mathbb{1}^T C(\lambda'\mathbb{1} - \beta_0\dot{x})/2$, while $dU_{\text{ext},p}/\epsilon = \frac{\lambda'}{\beta_0}\mathbb{1}^T C(\lambda'\mathbb{1} - \beta_0\dot{x})/2$. Similarly, an electric field can couple only to the net dipole moment of a system. This implies the transformation of applied energy to heat if it cannot be manipulated to perfectly match fluctuations in the driven variable.

The kinetic energy change of I, ascribed to the reservoir, can be determined in the mechanical limit from

$$0 = dU_{\text{ext}} + dK - dW$$

$$\Rightarrow dK = -dI^T(\tilde{\beta}/\beta_0 - \dot{a}), \tag{27}$$

using Eqns. 21 and 22[**?**]. The equations above thus completely describe any exchange of mechanical energy between deterministic systems exerting known forces. If the two-spring system considered above were disconnected at time S, with an outside observer absorbing the kinetic energy of the wall (Eq. 27), the energy change of the reservoir would reduce to the usual thermostatic potential change $\int_0^S dU_{\text{ext}}(t)$.

A well known consequence of Liouville's theorem is that the entropy change is zero in a completely deterministic process.[25, 55] Using adiabatic processes, then, it is possible to propagate a starting state, A, to any state with constant entropy. If, however, phase space volume were not preserved, it would amount to discarding information on the state of the system at a given time (for example, by integrating the probability over short time intervals). Then the amount of work that can be recovered from the system will become less than the amount input. In an extreme case, all information about the system may have been lost, flushing the corresponding information content to zero. Starting from this unknown state, p^0, the probability of a given frequency distribution, p, is approached by $P(p) \propto e^{\mathcal{H}[p]}$, with $\mathcal{H}[p]$ the familiar information entropy functional,

$$\mathcal{H}[p] = -\int p_i \ln p_i / p_i^0. \tag{28}$$

Further, if exchanges of conserved quantities during some process $A \to B$ are known, then the set $I(B)$ are also known from A. This information can be usefully employed to increase the amount of work that can be recovered, showing the entropy as a measure of "lost work." Consequently, any adiabatic process should be described by not only the mechanical work values, but also the change in information entropy due to information loss.

Next, consider allowing heat exchange in addition to controllable work. Assuming an infinitely large reservoir (or a short enough time-step), added heat will cause a negligible change in reservoir temperature. Because we have assumed the work done on each reservoir (Eq. 21) can be reversibly stored, these are not associated with an entropy change. We therefore introduce the physical entropy change in the reservoir as due only to exchange of heat, or "non-work" energy, $dS_{\text{ext}} = \beta_0 dQ_{\text{ext}}$, with $\beta_0^{-1} = k_B T_{\text{ext}}$. In this article, different notations are used for the information entropy, $\mathcal{H}$, the physical entropy, S, the caliber, σ, and the caliber-like functional, σ^*. All of these quantities are defined to be unitless and have some relation to the information entropy of Eq. 28. The energy rejected to the reservoir as heat, dQ_{ext}, originates from heat removed from the system plus a kinetic energy, dK,

assumed to be recoverable only as heat.

$$dQ_{\text{ext}} = -dQ + dK = -dE - dU_{\text{ext}}$$

$$= \frac{\delta \mathcal{A}}{\delta x(t)}^T \dot{x}\epsilon - d\tilde{I}^T \tilde{\beta}/\beta_0 = -dI^T \beta/\beta_0 \tag{29}$$

This heat can alternately be understood as the total energy dumped to the environment minus the energy removed "reversibly" (dU_{ext}). This interpretation shows that if some of the changes in the environment were re-classified as irreversibly stored, so that the information $\beta_j dI_j$ becomes lost, then this is equivalent to adding that energy to the total dQ_{ext}. The total energy rejected to the environment is recoverable (making a given step mechanical as considered above) if and only if $dQ_{\text{ext}} = 0$, implying $dE = dI_0 = -d\tilde{I}^T \beta/\beta_0$. These considerations again highlight the subjective nature inherent in the definition of irreversibility.

Connecting back to the usual thermostatics, U_{ext} plays the role of an energy for the reservoir.

$$dU_{\text{ext}} = dE_{\text{ext}} - \beta_0^{-1} dS_{\text{ext}}$$

$$= d\tilde{I}^T \tilde{\beta}/\beta_0$$

To use this relation, we must provide an experimental justification for the ability to use or store the energy terms appearing in the last equation.

For any transformation, $A \to B$, the total entropy change deriving from information loss and interactions with the environment is given by the change in information entropy plus the heat exchange term, the total entropy change is the sum of the increments,

$$\Delta S_{\text{inf,tot}} = \mathcal{H}[P(x|B)] - \mathcal{H}[P(x|A)] + \beta_0 Q_{\text{ext}}. \tag{30}$$

It should be noted that this formula is still not complete if there is a change in the phase space between A and B, for example if particles are added/removed, or if the state space is uniformly dilated. In this case, we have extra information on the region of phase space occupied after a transition. In general, if the state at time i is known to be x_i, then the size of the region of configuration space accessible at time $i+1$ is $Z_i(\beta, x_i)$ (Eq. 8). This reduces the entropy at $i+1$ to $S_{i+1} - \ln Z_i[\beta, x_i]$. Writing this down for each transition,

$$\Delta S_{\text{tot}} = \left\langle \sum_{i=0}^{S-1} dS_i \right\rangle$$

$$dS_i \equiv -\ln \frac{\mathcal{P}_{i+1}(x_{i+1}) Z_i[\beta, x_i]}{\mathcal{P}_i(x_i) Z_i[-\beta, x_{i+1}]} - dI^T \beta. \tag{31}$$

This equation has been symmetrized by including the corresponding entropy decrease for i given that x_i was inferred from x_{i+1} using the forward step probability but with reversed forces, $-\beta$. Support for this form is given in the next section, where it is also shown that $\Delta S_{\text{tot}} \geq 0$ using the Gibbs inequality.

Nevertheless, Eq. 30 can already be applied to Langevin and Brownian motion with uniform diffusion constants. As another example, applying Eq. 30 to a process taking any point to the equilibrium distribution shows that the expected entropy change for this process is the usual system entropy difference plus $\beta_0 Q_{\text{ext}} = -\beta_0(\langle E|B\rangle - \langle E|A\rangle) - \Delta\tilde{I}^T\tilde{\beta} = (I(A) - I(B))^T\beta$. Because the end-point entropies are fixed, alternate processes for transforming $A \to B$ are restricted to varying $\Delta S_{\text{ext}} = \int \beta_0 dQ_{\text{ext}}$. We may give a direct microscopic interpretation for the Clausius form of the second law for such processes.

$$\Delta S_{\text{tot}} = \Delta\mathcal{H} + \int \beta_0 dQ_{\text{ext}} \geq 0 \tag{32}$$

Here $\Delta\mathcal{H}$ is a function of the end-points, and β_0 and dQ_{ext} are fully variable along the path. Choosing a path from a specified distribution, A, to a maximum entropy distribution, $I(A)$, and then to the ending maximum entropy distribution $I(B)$, we may employ a quasistatic, "reversible," path between the two maximum entropy distributions so that $\min \int \beta_0 dQ_{\text{ext}}(A \to I(B)) = \min \int \beta_0 dQ_{\text{ext}}(A \to I(A))$. The heat evolved in this best-case process has its origin in the re-classification of information that occurs during the coarsening of A, in accordance with conclusions on Maxwell's demon.[56]

The equilibrium theory is therefore contained in the present development in the form of slow process that quickly relaxes to a canonical distribution at each moment. This perspective shows the intimate connection between coarse-graining that assumes infinitely fast relaxation of the reservoir and the traditional theory of quasi-static processes. The complete theory also permits an investigation of both relaxation processes and entropy production in time and history-dependent processes moved arbitrarily far from equilibrium by coupling to simple thermostatic reservoir systems.

IV.   PATH AVERAGES AND FLUCTUATIONS

Analogous to the equilibrium theory, we should expect that a cumulant expansion of an appropriate path free energy will yield path averages and fluctuations. A path free

energy functional up to time i can be defined from the partition functions of the transition distribution (Eq. 8) as

$$\mathcal{F}[\lambda_i] \equiv -\big\langle \ln Z[\lambda_i, \{x\}_0^i] \big\rangle. \tag{33}$$

Expanding

$$\frac{\partial \mathcal{F}}{\partial \lambda_{k,i}} = \frac{\partial}{\partial \lambda_{k,i}} \int -\ln Z[\lambda_i] \mathcal{P}\left(\{x\}_0^{i+1}\right) \, d\{x\}_0^{i+1}$$
$$= \big\langle f_k(x_{i+1}, \{x\}_0^i) \big\rangle,$$

we see that the derivatives of $\mathcal{F}$ indeed give the path averages, $\langle f \rangle$.

This result is valid for any chosen set of λ and corresponding functions, f_k. It is therefore possible to formally use the expansion to compute arbitrary path expectations for both thermal and mechanical driving forces. If they do not affect the dynamics ($\lambda = 0$), this formula is still useful in computing averages by analogy to the equilibrium theory using a first law of time-dependent nonequilibrium thermodynamics,

$$d\mathcal{F}[\lambda_i] = \sum_k \big\langle f_k(x_{i+1}, \{x\}_0^i) | \lambda \big\rangle d\lambda_{k,i}. \tag{34}$$

We can similarly compute second derivatives to give transient Gibbs-Maxwell relations,

$$\frac{\partial^2 \mathcal{F}[\lambda_i]}{\partial \lambda_i \partial \lambda_{j \leq i}} = -\big\langle (f_i - \langle f_i | \{x\}_0^j \rangle)(f_j - \langle f_j | \{x\}_0^j \rangle) \big\rangle. \tag{35}$$

Using the Legendre transformations ($\mathcal{F} - \lambda_h \frac{\partial \mathcal{F}}{\partial \lambda_h}$), these fluctuations can be transformed to ensembles with constrained averages (acceleration, particle flux, etc.), $\langle f \rangle$, rather than thermodynamic forces, λ.[28] These transformations form a basis for connecting stochastic, Langevin dynamics simulations (e.g. Eq. 18) to constant kinetic energy, solute flux, and other constant flux ensembles studied extensively in nonequilibrium molecular dynamics simulations.[44]

In addition, the coupling between past and present currents appears as the starting point for the Mori projector-operator method[7] as discussed previously by Jaynes.[8, 29] It should be noted that the time-asymmetry inherent in our definition of $\mathcal{F}_i$ implies that thermodynamic forces applied in the future, $\lambda_{t>i}$, do not affect currents in the past. Note that a sum of $\mathcal{F}_j$ over all times j has the same properties as Eq. 34 with respect to differentiation by λ_i when $d\lambda_{t>i} = 0$, but with time-symmetry restored. A careful consideration of the time-symmetric form leads to a path functional that can carry out inference on previous current

fluctuations given future deviations. Future consideration of these topics will greatly extend the range of nonequilibrium thermodynamic theory.

The linear relations (Eq. 35) should not be expected to hold for large deviations in the nonequilibrium forces. We can progress beyond this limitation by analogy to the transition from thermodynamic integration to free energy perturbation in equilibrium free energy calculations. As discussed in the introduction, any two processes on the same path space, $\{\Gamma\}$, can be compared using a likelihood ratio as in Eq. 2. This ratio allows for sampling methods that re-weight averages on path space with likelihood ratios as weights.

It has been claimed that entropy production can be gauged by the ratio of forward to reverse path probabilities.[12, 57] For the process derived in Sec. II, we can define a reverse process by inverting the sign of the generalized forces, β_i, and normalizing the distribution separately for each x_{i+1}.[28] This reversal corresponds to an attempt at guessing whether energy has been added or subtracted during a step $i \rightarrow i + 1$, with the action deviation constraint, G, unchanged. Thus, we can define a ratio

$$e^{dS_i} \equiv \frac{\mathcal{P}\left(x_{i+1}|x_i, \beta\right) \mathcal{P}\left(x_i\right)}{\mathcal{P}\left(x_i|x_{i+1}, -\beta\right) \mathcal{P}\left(x_{i+1}\right)} \tag{36}$$
$$= \frac{\mathcal{P}\left(x_i\right) p_0(x_{i+1}|x_i) Z[-\beta, x_{i+1}]}{\mathcal{P}\left(x_{i+1}\right) p_0(x_i|x_{i+1}) Z[\beta, x_i]} e^{-dI^T \beta}.$$

Since $\frac{p_0(x_{i+1}|x_i)}{p_0(x_i|x_{i+1})} = \frac{p_0(x_{i+1})}{p_0(x_i)}$ by Bayes' theorem, this result exactly matches Eq. 31 arrived at through thermodynamic reasoning.

Moreover, several steps can be concatenated to give

$$e^{\sum_{i=0}^{S-1} dS_i} = \frac{\mathcal{P}\left(\Gamma|x_0, \beta\right) \mathcal{P}\left(x_0\right) / p_0(x_0)}{\mathcal{P}\left(\Gamma|x_S, -\beta\right) \mathcal{P}\left(x_S\right) / p_0(x_S)}. \tag{37}$$

This equation can therefore be viewed as a statistical basis for the second law of thermodynamics. Note that it is the presence of the probability for the end-states in this formula that leads to the instantaneous information entropy required for precisely defining dS_i, Eq. 31. Our formula is physically motivated by observing that the entropy increase is attributed to a combination of environmental entropy changes, dQ_{ext}, and information-like entropy changes,

$$- \ln \frac{\mathcal{P}\left(x_{i+1}\right) Z[\beta, x_i] / p_0(x_{i+1})}{\mathcal{P}\left(x_i\right) Z[-\beta, x_{i+1}] / p_0(x_i)}.$$

This term implies an extra contribution to the total entropy change beyond Eq. 30. To understand this contribution, consider a simple one-state system, to be transformed into a two-state system through the (unbiased) transition probability $\mathcal{P}\left(x_{i+1}|x_i\right) = p_0(x_{i+1}) = 1/2$. The

presence of the normalization in Eq. 31 leads to an additional factor of $-\ln Z[x_i]/p_0(x_{i+1}) = -\ln 2$ in the entropy. This term is canceled by the entropy of the resulting state, $\mathcal{P}(x_{i+1}) = 1/2$, so that the total entropy change for this process is zero. If instead we perform a bit-set operation by going in the opposite direction, the entropy will increase if $\mathcal{P}(x_{i+1})$ is any non-uniform state – corresponding to our loss of information on discarding the bit. A term of this form is exactly what we should have expected when writing down Eq. 30. Applying this equation to a situation where a particle is added to the system, we find that the re-normalization will physically compensate for the expansion of phase space, which leads to a difference in entropy measures. Mathematically, this implies that the "default" measure $p_0(x_{i+1})$ is replaced by the re-normalized measure $p_0(x_{i+1})Z_{\text{tr}}^*(x_{i+1})/Z_{\text{tr}}(x_i)$ if we have information about the previous state when calculating the entropy at state $i+1$.

Equation 37 is connected to the likelihood definition (2). To see this, define B as a "forward" process, starting from $\mathcal{P}(x_0)$ and employing $\lambda_B = \{G, \beta_{i,k}/2\}$, and A as a "reverse" process, employing $\lambda_A = \{G, -\beta_{i,k}/2\}$. Because

$$e^{l_{A\to B}} = \frac{\mathcal{P}(\Gamma|B)}{\mathcal{P}(\Gamma|A)} = \frac{\mathcal{P}(\Gamma|x_0, B)\,\mathcal{P}(x_0|B)}{\mathcal{P}(\Gamma|x_S, A)\,\mathcal{P}(x_S|A)}$$

and the end-point distributions $\mathcal{P}(x_0)$, $\mathcal{P}(x_S)$ have been pre-determined, we can write Eq. 31 as

$$\Delta S_{\text{tot}} = \langle l_{A\to B}\rangle_B, \tag{38}$$

proving

$$\Delta S_{\text{tot}} \geq 0.$$

This completes the connection between the physical entropy production defined in Sec. III and fluctuation theorems in the form of Eq. 2.

To complete the discussion, we connect the entropy production to Jaynes' caliber function,

$$\sigma_\Gamma \equiv \mathcal{H}_0 + \left\langle \sum dI^T \beta/2 \right\rangle_B + \left\langle \sum_{i=0}^{S-1} \ln \mathcal{Z}_i[\beta] \right\rangle_B. \tag{39}$$

To do this, we decompose Eq. 36 into the form of Eq. 39 as

$$\Delta S_{\text{tot}} = \sigma_\Gamma^* - \sigma_\Gamma \tag{40}$$

by defining

$$\sigma_\Gamma^* \equiv -\left\langle \sum dI^T \beta/2 \right\rangle_B + \left\langle \sum_{i=0}^{S-1} \ln \mathcal{Z}_i[-\beta] \right\rangle_B + \mathcal{H}_S. \qquad (41)$$

Although this definition is similar to the caliber, it is not an information entropy since the averages are taken with respect to the forward probability distribution, $\mathcal{P}\left(\Gamma|x_0,\beta\right)\mathcal{P}\left(x_0\right)$. The choice of the forward direction corresponds to the direction in which information propagates [28, 47]. It determines the target distribution for taking the divergence (Eq. 38).

V. ATOMIC FORCE MICROSCOPY EXAMPLE

To give concrete examples of Eqns. 35 and Eq. 31, we must choose a set of coarse variables of interest ($\langle f \rangle$) and follow their time-evolution, $\langle f(t) \rangle$. Most mesoscopic models contain hydrodynamic equations of motion for the solution density. A rigorous route[58] to their derivation is by forming suitable integrals of the probability distribution function appearing inside the Liouville equation. Much of the early literature on nonequilibrium problems is focused on this derivation. In the maximum transition entropy context, the resulting equations describe the propagation of a state of knowledge forward in time using an exact equation of motion. The exact internal and external forces on the system are required using this route, and the projector-operator formalism is used to add the uncertainty introduced by mixing processes occurring below the size and time scale of the density function.

In the stochastic formalism developed here, the Liouville equation has been replaced with the Klein-Kramers and Smoluchowski equations. In addition to convection, these equations also describe diffusion of probability that occurs because of loss of information during mixing processes. Summing the Fokker-Planck equation for Eq. 20 over particles of each species type, α, and integrating over coordinates other than that of a single, distinguished molecule (using the relations from Ref. 58) gives the one-particle evolution equation,

$$\frac{\partial \rho_\alpha(r,t)}{\partial t} = -\nabla_r \left(D(\beta F \rho(r,t) - \nabla_r \rho(r,t)) \right). \qquad (42)$$

In this process, the force and diffusion coefficients have become averages over the probability distributions of the molecules of each type. Here, D has been substituted for $C_x/2$.

We present a numerical calculation of one-dimensional, optically trapped bead or atomic-force microscope pulling experiment. Assuming very fast relaxation of the momentum of the

pulling coordinate over a potential energy surface, U, it is appropriate to use Eq. 20 with the average velocity set to zero. In the case of an optical trap with center $x_0(t)$ and force $F(x,t) = -\kappa(x - x_0(t))$, we can employ a change of coordinates to $y(t) = x(t) - x_0(t)$ (with $\dot{x}_0(t) \equiv v(t)$) so that during each time-step the position and trap center are updated to give

$$\dot{y}\epsilon = -v(t)\epsilon - \frac{\kappa}{\gamma}y\epsilon + C_q^{1/2}d\mathcal{W}, \ \gamma^{-1} \equiv \beta C_q/2. \tag{43}$$

The constraint on $\dot{x}^2\epsilon/C_q$ determines the rate at which the bead is allowed to dissipate energy into the solution while the energy constraint, β, acts as a driving force for net energy exchange.

In the Brownian case, the system cannot distinguish between internal and external forces, so that Eq. 22 shows energy decrease as work from all applied (assumed internal) forces is dissipated into the surroundings as heat at each time-step. Reversing our sign convention to treat the harmonic trap as external, the work done on the bath through the system is

$$\begin{aligned} \langle dW_F|y_i\rangle &= \langle F\dot{x}\epsilon|y_i\rangle = -\langle\kappa\bar{y}\dot{x}\epsilon|y_i\rangle \\ &= \frac{\kappa}{\gamma}(\kappa y^2 - 1/\beta)\epsilon. \end{aligned} \tag{44}$$

If a potential were present for the particle, at each step energy

$$\langle dW_U|y_i\rangle = \left\langle -\frac{\partial U}{\partial x}\dot{x}\epsilon|y_i\right\rangle$$

would be converted into heat from the bath system's internal potential, U.[?] We are not including a potential term in our analysis.

The entropy increase of Eq. 31 (Eq. 36) is formally a path average, and its evaluation requires specifying an initial state and a driving protocol. As in Ref. 14, we may choose to follow several velocity programs starting from a steady-state at constant pulling force,

$$\rho_{\text{ss}}(y) \propto e^{-\frac{\beta\kappa}{2}(y+\gamma v/\kappa)^2}. \tag{45}$$

The 'housekeeping heat' dissipated by the bath's removal of the bead's momentum at each time-step leads to a steady-state dissipation of $\langle dW_F\rangle/dt = \gamma v^2$. After a sufficiently long time, the instantaneous information entropy will reach its steady-state value, $\mathcal{H}[\beta, \kappa] = (1 - \ln\frac{\beta\kappa}{2\pi})/2$. Note that the entropy of the steady-state distribution is well-defined because it is invariant to the change of coordinates $x \to y$. For constant pulling force and temperature,

this expression says that over long time-periods the integral of the information entropy change will be zero, and the total entropy increase will be completely due to dissipated work.

During intermediate time-periods, however, the entropy increase will be a combination of changes in the position distribution plus the dissipated work.

$$dS_i = d\mathcal{H}_i - \beta \langle dE \rangle_i = \left\langle -\ln \frac{\mathcal{P}_{i+1}(y_{i+1})}{\mathcal{P}_i(y_i)} + \beta dW_F(y_{i+1}; y_i) \right\rangle_i \qquad (46)$$

Because the dynamics is Markovian, the average dissipated work can be easily calculated from the distribution at each time-step and the total of each type of work will be a sum of one-step stochastic integrals (Eq. 44). Assuming the distribution is Gaussian at the starting time and Fourier-transforming Eq. 45 gives a Gaussian distribution at all future times with parameters as derived in Ref. [59]. The information entropy and work follow

$$\frac{d\mathcal{H}}{dt} = D\left(\frac{1}{w} - \beta\kappa\right), \qquad\qquad \frac{dW_F}{dt} = \frac{\kappa}{\gamma}\left(\kappa(\mu^2 + w) - \frac{1}{\beta}\right).$$

Using $x + x^{-1} \geq 2$, we can easily verify that the second law holds for the average,

$$\langle dS_i \rangle = D\beta^2 \kappa^2 \mu^2 + D\beta\kappa \left[\beta\kappa w + (\beta\kappa w)^{-1} - 2\right] \geq 0. \qquad (47)$$

We have plotted (Fig. 1) the rate of entropy production vs. time for two hypothetical driving protocols using numerical integration. The units in the figure are the same as in Ref. 14, and their third protocol (followed by its time-reversal starting at 0.28 s) has been used for the upper two sets of panels. Because the variance of the distribution only responds to changes in diffusion constant, temperature or driving force, we have varied κ in the second set of calculations. The information entropy rate gain goes to zero and the heat production becomes constant at the onset of the eventual steady-state. As shown by the heat production during compression from $\kappa = 3$ to 5 pN/μm, excess heat production is required whenever the information entropy decreases. When the distribution expands in the second half of the protocol, heat production decreases as the mean begins to lag behind the trap center. This is counter-balanced by an increase in information entropy, leading to net dissipation.

We note that a large amount of additional complexity can be added to this model by adding information about the variables here treated as 'external' to the description of the dynamics. If local variations in the fluid velocity or temperature were included, then the dynamics would have to specify the equations of motion for these fields. The final entropy
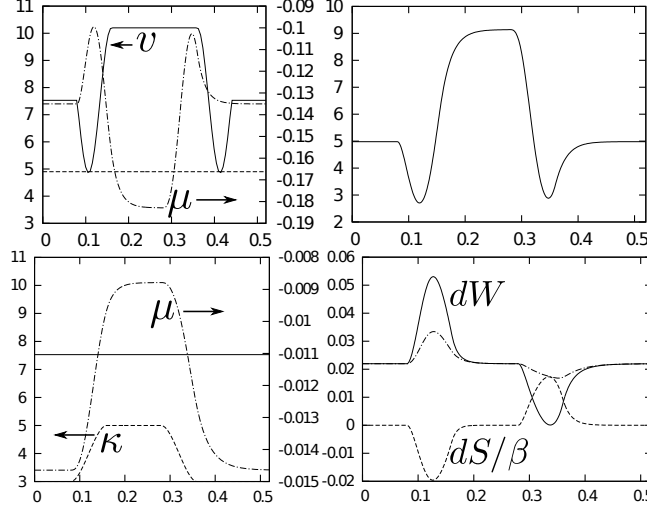
FIG. 1. Calculated entropy production during a transition between steady-states. The left set of panels show the imposed velocity, v (solid line), and force constant, κ (dashed line), on the left scale; and the response of the mean, μ (dot-dashed line, right scale) vs time (s). The right set of panels show the decomposition of the entropy increments (Eq. 46, pN-μm) into heat (solid) and information gain/loss (dashed). Whenever the information entropy decreases, an equal or larger amount of heat is produced so that the total (Eq. 47, dot-dashed) is always positive. For the upper two sets of panels, the force constant was held constant at $\kappa = 4.9$ pN/μm so no change in information entropy occurs. For the lower two, the distribution is compressed, then broadened by changing κ between 3 and 5 using a cubic interpolation lasting 80 ms.

increase may then be more or less than this result because these degrees of freedom may be responsible for additional heat production, but more information on the fluid state has been included, leading to decreased information loss.

VI. CONCLUSIONS

In this article, we have given a generalization of the theory for driven, irreversible processes. This theory does not require assuming a Boltzmann-Gibbs form for the starting distribution, nor does it rely on any particular definition of a stationary distribution. A set of transition probabilities obtained from applying maximum entropy methods defines the evolution equation for the system of interest. A special simplification is the case of Langevin and Brownian motion, which can be recovered as limits of a constrained action integral ap-

proach. Non-anticipating stochastic trajectories for classical particle and field motion can be cast in this form. The action functional interpretation gives a physical method for defining conserved quantities. The energy cost associated with transfers of these quantities from an external environment or experimental apparatus can easily be specialized by an appropriate choice of constraints and computed.

Deterministic dynamics are recovered from the Langevin equation when the deviation of the action functional is strongly constrained to zero (Sec. II). In this limit, the external forces, which appeared as statistical in the stochastic approach, become the mechanical Lagrange-D'Alembertian forces. Because both limits appear in this derivation, the fluctuation-dissipation theorems derived as Gibbs relations from Eq. 34 are applicable in the case of both thermal and mechanical driving forces. These equations are completely general because they apply not only arbitrarily far from equilibrium, but also during transient processes that do not possess a steady-state.

A particularly useful aspect of this approach is that it directly connects multiple length and time scales. The formulation of the equations has been in terms of particle motion, but coarse-grained relations are easy to define as appropriate ensemble averages over these motions. Examples of such averages include centers of mass for polymer units or average density and velocity fields. The coarse equations of motion will then lead to polymer coarse-graining models[60] or non-local hydrodynamic models.[1, 58, 61] For the time-evolution of average quantities, we expect the thermodynamic limit argument[47] to apply when the number of averaged degrees of freedom is large, so that the path realized by the system under a given set of constraints will fall arbitrarily close to the maximum entropy solution an overwhelming majority of the time. The present work is therefore a suitable foundation for the theory and analysis of nonequilibrium molecular dynamics.

Our method is free of projection operators and Fourier transforms that can often obscure the physical picture of driven dynamical systems. As a consequence, applications to simplified, standard examples such as circuit theory are easily accomplished. In addition, we find the physical basis of entropy production has become much more apparent. The Joule heating of a resistor, for example, can be seen from Eq. 26 as fundamentally arising from the difference between the velocity added to each ion individually *vs.* the usable energy in the average ion velocity. Because the energy added to the system in driving the ions is not expressible in terms of the average velocity alone, spreading in the distribution of ion

velocities becomes heat. The same remarks follow for driven convective transport, where a spreading in the distribution of forward fluid momentum leads to increases in the local temperature (Eq. 19).

Connections of this theory to the formal structure of maximum entropy thermodynamics and Bayesian inference have been elaborated upon in Ref. 28. These connections allow the definition of thermodynamic cycles expressing differences between driving protocols using the same free energy techniques commonly employed in the equilibrium theory. Some examples have already appeared in the literature for path re-weighting[62, 63]. It is expected that expression in terms of thermodynamic cycles will greatly simplify the derivation and interpretation of these studies.

We have identified a new generalization of the second law for irreversible processes by including the instantaneous information entropy. A traditional analysis shows that the total entropy increase (Eq. 31) is dependent on the details of system dynamics and exchange of conserved quantities with an external system. Connecting this with the fluctuation theorem (Eq. 36) gives a microscopic form for the second law of thermodynamics. The physical device of tracking work performed on individual particles as well as external reversible work sources allows us to track the flow of each type of work (and heat) through the system. Because these changes come directly from the forces on each degree of freedom, this analysis does not depend on arbitrary decompositions of energy functions or definitions of steady-states.

From an informational perspective, entropy increase comes about from discarding information and/or from the information loss associated with coupling to external reservoirs. This is distinguished from the entropy production functional of local equilibrium theory in that the entropy functionals developed here include long-range correlations and are not required to be extensive.[64] It is a simple matter to define more complicated baths, for example affecting only the average temperature in a given area for imposing thermal gradients. It should be noted that the analysis in Sec. III showed that increasing the number of controllable variables decreases the number of degrees of freedom associated with heat production.

Many questions remain open for attack by this method. Feynman's path action formulation of quantum mechanics[45] uses the action functional as a rotation operator, whereas the present relations are of the diffusion-type. Is there an appropriate limiting process whereby the latter may be obtained and the multipliers G identified with quantum fluctuations?[65–

67] Ref. [68] gives a Lagrangian related to the Clebsch parametrization for the field equations of Eulerian fluid mechanics, where the stream function ϕ plays the role of v in Eq. 20. Using the approach of this paper, energy constrained variations in $\partial \mathcal{A}/\partial v$ give the linear flux equations of local equilibrium hydrodynamics, variations in the mass conservation equation, $\partial \mathcal{A}/\partial \phi$ give a stochastic particle birth/death process, and variations in the energy equation, $\partial \mathcal{A}/\partial \rho$ give a pure strain term for momentum diffusion with a properly defined non-local and time-dependent viscosity. Addition of electric fields to this picture should result in derivations for the dielectric response in nanoscale fluid mechanics[69, 70].

The present report has not been specialized to cases where near-equilibrium approximations may be applied, and these should also be investigated for the potential of generating new relations in that regime. For such work, the nonequilibrium statistical operator approach[9] should prove helpful. Molecular insertion and deletion operations such as those considered from a maximum caliber approach in Ref. 36, 37 are an important next step for identifying a local formula for the chemical potential as was done here for the temperature. This step is non-trivial, as particle addition is a jump process, and cannot use the normal distribution form of the present report. The drift constraint, G, matters less for this analysis, but particle number-transition probabilities are not easily identifiable with chemical potentials (or differences between them) except in the case where all particle numbers are accessible at each addition/deletion step at a given position.

The simple form of Eq. 35 should aid in connecting constant flux integration schemes such as the Gaussian thermostat to constant force methods such as dual-reservoir nonequilibrium molecular dynamics calculations. A more complicated starting point for this task, the derivation of the adiabatic susceptibility, was done in chapter 4 of Ref. [71] by means of a quasi-static perturbation of a canonical distribution from a Hamiltonian. The general pattern derived in Sec. IV gives a welcome alternative.

ACKNOWLEDGMENT

Security Administration under contract DE-AC04-94AL85000.

[1] R. Kubo, Rep. Prog. Phys. **29**, 255 (1966).

[2] L. Onsager, Phys. Rev. **37**, 405 (1931).

[3] H. B. Callen, *Thermodynamics and an Introduction to Thermostatistics* (Wiley, New York, 1985), ISBN 0471862568, 2nd ed.

[4] L. Onsager and S. Machlup, Phys. Rev. **91**, 1505 (1953).

[5] P. G. Bergmann and J. L. Lebowitz, Phys. Rev. **99**, 578 (1955).

[6] R. Zwanzig, Ann. Rev. Phys. Chem. **16**, 67 (1965).

[7] H. Mori, Prog. Theor. Phys. **33**, 423 (1965).

[8] E. T. Jaynes, Ann. Rev. Phys. Chem. **31**, 579 (1980).

[9] D. N. Zubarev, J. Math. Sci. **16**, 1509 (1981), ISSN 1072-3374, URL `http://dx.doi.org/10.1007/BF01091712`.

[10] D. Jou, J. Casas-Vázquez, and G. Lebon, Rep. Prog. Phys. **51**, 1105 (1988).

[11] D. Jou, J. Casas-Vázquez, and G. Lebon, Rep. Prog. Phys. **62**, 1035 (1999).

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

[13] S. Kjelstrup and D. Bedeaux, *Non-Equilibrium Thermodynamics Of Heterogeneous Systems (Series on Advances in Statistical Mechanics)* (World Scientific, 2008), ISBN 9812779132.

[14] E. H. Trepagnier, C. Jarzynski, F. Ritort, G. E. Crooks, C. J. Bustamante, and J. Liphardt, Proc. Nat. Acad. Sci. USA **101**, 15038 (2004).

[15] G. E. Crooks, Phys. Rev. E **61**, 2361 (2000).

[16] C. Jarzynski, Phys. Rev. Lett. **78**, 2690 (1997).

[17] C. Jarzynski, J. Stat. Phys. **98**, 77 (2000).

[18] B. Palmieri and D. Ronis, Phys. Rev. E **75**, 011133 (2007).

[19] V. Y. Chernyak, M. Chertkov, and C. Jarzynski, J. Stat. Mech. p. P08001 (2006).

[20] J. G. Kirkwood, J. Chem. Physics **14**, 180 (1946), URL `http://link.aip.org/link/?JCP/14/180/1`.

[21] H. B. Callen and T. A. Welton, Phys. Rev. **83**, 34 (1951).

[22] R. Zwanzig, Phys. Rev. **124**, 983 (1961).

[23] R. Zwanzig, K. S. J. Nordholm, and W. C. Mitchell, Phys. Rev. A **5**, 2680 (1972), correction to Zwanzig, 1961.

[24] L. Onsager, Phys. Rev. **37**, 405 (1931).

[25] B. Robertson, in *Physics and Probability: essays in honor of Edwin T. Jaynes*, edited by W. T. Grandy (Cambridge Univ. Press, 1993), pp. 251–260.

[26] J. W. Gibbs, *Elementary principles in statistical mechanics* (C. Scribner's sons, 1902).

[27] E. Schrödinger, *Statistical thermodynamics* (Cambridge, 1967), ISBN 0486661016, 9780486661018.

[28] D. M. Rogers, T. L. Beck, and S. B. Rempe, J. Stat. Phys. **145**, 385 (2011).

[29] E. T. Jaynes, in *The Maximum Entropy Formalism*, edited by R. D. Levine and M. Tribus (M.I.T Press, Cambridge, 1979), p. 498, ISBN 0262120801,9780262120807.

[30] E. T. Jaynes, *Probability Theory: the Logic of Science* (Cambridge, Cambridge, 2003).

[31] E. T. Jaynes, in *Complex Systems - Operational Approaches*, edited by H. Haken (1985), p. 254.

[32] L. Y. Chen and P. L. Nash, arxiv preprint pp. 1–15 (2004).

[33] K. L. C. Hunt and J. Ross, J. Chem. Phys. **75**, 976 (1981), URL `http://link.aip.org/link/?JCP/75/976/1`.

[34] J. L. Lebowitz and H. Spohn, J. Stat. Phys. **95**, 333 (1999), ISSN 0022-4715, URL `http://dx.doi.org/10.1023/A:1004589714161`.

[35] R. Dewar, J. Phys. A: Math. Gen. **36**, 631 (2003).

[36] G. Stock, K. Ghosh, and K. A. Dill, J. Chem. Phys. **128**, 194102 (2008), ISSN 00219606, URL `http://dx.doi.org/doi/10.1063/1.2918345`.

[37] S. Press, K. Ghosh, and K. A. Dill, J. Phys. Chem. B **115**, 6202 (2011).

[38] A. A. Filyukov and V. Y. Karpov, J. Eng. Phys. Thermophys. **13**, 416 (1967), ISSN 1062-0125, URL `http://dx.doi.org/10.1007/BF00828961`.

[39] J. Gill, *Bayesian Methods* (Chapman & Hall/CRC, 2002), ISBN 1584882883,978-1584882886.

[40] A. Bobrowski, *Functional Analysis for Probability and Stochastic Processes: An Introduction* (Cambridge University Press, 2005), ISBN 0521831660.

[41] J. P. Boon and S. Yip, *Molecular Hydrodynamics* (Courier Dover Publications, 1991), ISBN 0486669491.

[42] R. Zwanzig, *Nonequilibrium Statistical Mechanics* (Oxford Univ. Press, 2001), ISBN 0195140184.

[43] G. Karniadakis, A. Beşkök, and N. R. Aluru, *Microflows and Nanoflows: fundamentals and simulation* (Springer, 2005), ISBN 0387221972.

[44] D. J. Evans and G. Morriss, *Statistical Mechanics of Nonequilibrium Liquids* (Cambridge Univ. Press, 2008), ISBN 0521857910, 2nd ed.

[45] F. Dyson, *Advanced Quantum Mechanics* (World Scientific, 2007), ISBN 9812706224, transcribed from 1951 Lecture Notes by David Derbes.

[46] J. E. Marsden and M. West, Acta Numerica **10**, 357 (2001).

[47] E. T. Jaynes, Phys. Rev. **108**, 171 (1957).

[48] E. Noether, Transp. Theory Stat. Phys. **1**, 186 (1971), English translation by M. A. Tavel.

[49] H. Risken, *The Fokker-Planck Equation: methods of solution and applications* (Springer, 1996), ISBN 354061530X, 2nd ed.

[50] T. D. Frank, Phys. Lett. A **360**, 552 (2007), ISSN 0375-9601, URL http://www.sciencedirect.com/science/article/B6TVM-4KTMSKY-2/2/44359866dfdd8fb20c7b27237223ebb0.

[51] S. E. Feller, Y. Zhang, R. W. Pastor, and B. R. Brooks, J. Chem. Phys. **103**, 4613 (1995), URL http://link.aip.org/link/?JCP/103/4613/1.

[52] P. Attard, J. Chem. Phys. **124**, 024109 (pages 10) (2006), URL http://link.aip.org/link/?JCP/124/024109/1.

[53] P. Attard, J. Chem. Phys. **130**, 194113 (pages 10) (2009), URL http://link.aip.org/link/?JCP/130/194113/1.

[54] P. Attard and A. Gray-Weale, J. Chem. Phys. **128**, 114509 (pages 16) (2008), URL http://link.aip.org/link/?JCP/128/114509/1.

[55] B. Robertson and W. C. Mitchell, J. Math. Phys. **12**, 563 (1971).

[56] H. S. Leff and A. F. Rex, *Maxwell's demon 2: entropy, classical and quantum information, computing* (IOP Publishing, London, 2003), ISBN 0750307595.

[57] G. E. Crooks, Phys. Rev. E **60**, 2721 (1999).

[58] R. Lehoucq and A. Von Lilienfeld-Toal, J. Elasticity **100**, 5 (2010), ISSN 0374-3535, URL http://dx.doi.org/10.1007/s10659-010-9246-9.

[59] O. Mazonka and C. Jarzynski, arXiv:cond-mat/9912121v1 (1999), report number LAUR-99-6303.

[60] R. L. C. Akkermans and W. J. Briels, J. Chem. Phys. **113**, 6409 (2000), URL `http://link.aip.org/link/?JCP/113/6409/1`.

[61] W. T. Grandy, *Foundations of Statistical Mechanics* (Kluwer, Boston, 1987), ISBN 902772489X, 9789027724892.

[62] D. M. Zuckerman and T. B. Woolf, J. Chem. Phys. **111**, 9475 (1999), URL `http://link.aip.org/link/?JCP/111/9475/1`.

[63] D. D. L. Minh and J. D. Chodera, The Journal of Chemical Physics **131**, 134110 (2009), URL `http://dx.doi.org/10.1063/1.3242285`.

[64] E. T. Jaynes, in *Maximum Entropy and Bayesian Methods*, edited by C. R. Smith, G. J. Erickson, and P. O. Neudorfer (Kluwer, Dordrecht, 1992), pp. 1–22, ISBN 978-0792320319.

[65] N. Christopeit, K. Helmes, M. Kohlmann, and K. Yasue, in *Stochastic Differential Systems* (Springer Berlin/Heidelberg, 1986), vol. 78 of *Lecture Notes in Control and Information Sciences*, pp. 352–361, URL `http://dx.doi.org/10.1007/BFb0041177`.

[66] I. M. Davies, J. Phys. A: Math. Gen. **22**, 3199 (1989), ISSN 0305-4470, URL `http://stacks.iop.org/0305-4470/22/i=16/a=010?key=crossref.34898115d47453fff1b3eb2605525560`.

[67] H. Dammak, Y. Chalopin, M. Laroche, M. Hayoun, and J.-J. Greffet, Phys. Rev. Lett. **103**, 190601 (2009).

[68] S. D. Mobbs, Proc. Royal Soc. London. A. Math. Phys. Sci. **381**, 457 (1982), URL `http://rspa.royalsocietypublishing.org/content/381/1781/457.abstract`.

[69] J. Hubbard and L. Onsager, The Journal of Chemical Physics **67**, 4850 (1977), URL `http://link.aip.org/link/?JCP/67/4850/1`.

[70] V. Ballenegger, R. Blaak, and J.-P. Hansen, in *Computer Simulations in Condensed Matter Systems: From Materials to Chemical Biology Volume 2*, edited by M. Ferrario, G. Ciccotti, and K. Binder (Springer Berlin / Heidelberg, 2006), vol. 704 of *Lecture Notes in Physics*, pp. 45–63, URL `http://dx.doi.org/10.1007/3-540-35284-8_3`.

[71] R. Kubo, M. Toda, and N. Hashitzume, *Statistical Physics II: Nonequilibrium Statistical Mechanics* (Springer, 2003), ISBN 354053833X, 2$^{\text{nd}}$ ed. Translated by M. Toda, R. Kubo, N. Saito, and N. Hashitsume.

[72] C. W. Gardiner, *Handbook of Stochastic Methods: for Physics, Chemistry and the Natural Sciences* (Springer, 2004), ISBN 3540208828, 3rd ed.

Appendix A: Analytical Calculation of Stochastic Integrals

Despite the wealth of literature on the Langevin and Wiener processes, the procedure for calculating expectations of time-integrals given in standard references such as Gardner[72] and Risken[49] remains complicated. Because Stratonovich integrals appear prominently in the present paper, often usurping the role of thermodynamic potentials, we present here two alternative methods. Both rely on replacing expressions to be evaluated at the midpoint of a time-step with the first-order expansion, $f(\bar{x}) \approx f(x) + \frac{\partial f(x)}{\partial x}\dot{x}\epsilon/2$.

Using Eq. 16 with $Y_0 = \bar{x}$ to find the energy change, we expand the average velocity about the mid-point,

$$
\begin{aligned}
\bar{x} = \frac{\dot{x}_{i+1} + \dot{x}_i}{2} &= \dot{x}_i + \epsilon/2 M^{-1}\dot{p} \\
&= \dot{x}_i + \tfrac{1}{2}M^{-1}\left(F\epsilon - \tfrac{\epsilon}{2}C\tilde{Y}\tilde{\beta} - \tfrac{\epsilon}{2}C\bar{x}\beta_0 + C^{1/2}d\mathcal{W}\right) \\
&= (I + \tfrac{\epsilon\beta_0}{4}M^{-1}C)^{-1}\left[\dot{x}_i + \tfrac{\epsilon}{2}M^{-1}(F - C\tilde{Y}\tilde{\beta}/2) + \tfrac{1}{2}M^{-1}C^{1/2}d\mathcal{W}\right] \\
&= (I - \tfrac{\epsilon\beta_0}{4}M^{-1}C)\dot{x}_i + \tfrac{\epsilon}{2}M^{-1}(F - C\tilde{Y}\tilde{\beta}/2) + \tfrac{1}{2}M^{-1}C^{1/2}d\mathcal{W} + O(\epsilon^{3/2}) \\
&= \dot{x}_i + \tfrac{\epsilon}{2}M^{-1}\dot{p}' + O(\epsilon^{3/2}),
\end{aligned}
$$

where $\dot{p}'$ is computed using only quantities at the time-step i. Multiplying this with $-\frac{\delta\mathcal{A}}{\delta x(t)}$ from Eq. 18, we get

$$
dI_0 = \bar{x}(\dot{p} - F)\epsilon = \tfrac{1}{2}d\mathcal{W}^T C^{1/2}M^{-1}C^{1/2}d\mathcal{W} + d\mathcal{W}^T C^{1/2}\dot{x}_i - \tfrac{\epsilon}{2}\beta^T Y'^T C\dot{x}_i + O(\epsilon^{3/2}).
$$

Since $M^{-1/2}C^{1/2}d\mathcal{W}$ is normally distributed with mean zero and variance-covariance matrix $M^{-1/2}CM^{-1/2}\epsilon$, dI_0 has a noncentral χ^2 distribution with expectation

$$
\tfrac{\epsilon}{2}\left[\operatorname{Tr}(M^{-1}C) - \beta^T Y'^T C\dot{x}_i\right].
$$

For a single constraint, $Y' = \dot{x}$, we find a definition of the kinetic temperature, Eq. 19. If, in addition, we include a constant pulling force, $\beta_1 = -\lambda$, we find

$$
\langle dI_0 \rangle = \tfrac{1}{2}\left[\operatorname{Tr}(M^{-1}C) - \beta_0(\lambda/\beta_0 \mathbb{1} - \dot{x}_i)^T C(\lambda/\beta_0 \mathbb{1} - \dot{x}_i) + \lambda(\lambda/\beta_0 \mathbb{1} - \dot{x}_i)^T C\mathbb{1}\right]. \tag{A1}
$$

What emerges is a kinetic temperature with respect to the terminal velocity, λ/β_0, as well as a heating term.

This method can also be used to prove Eq. 44, starting from the expansion

$$\bar{y} = y_i + \epsilon\dot{y}/2$$
$$= y_i - \tfrac{1}{2}\left(v(t)\epsilon + \tfrac{\kappa}{\gamma}\bar{y}\epsilon - C^{1/2}d\mathcal{W}\right)$$
$$= (1 + \tfrac{\kappa\epsilon}{2\gamma})^{-1}(y_i - \tfrac{v\epsilon}{2} + \tfrac{1}{2}C^{1/2}d\mathcal{W})$$
$$= (1 - \tfrac{\kappa\epsilon}{2\gamma})y_i - \tfrac{v\epsilon}{2} + \tfrac{1}{2}C^{1/2}d\mathcal{W} + O(\epsilon^{3/2}).$$

As discussed in the text, these integrals should also result from differentiating a partition function (Eq. 33). We present an extended derivation of the main results of this approach here. Both Langevin and Brownian equations can be derived as appropriate limits of the constraints

$$\eta(t)/\epsilon = \begin{bmatrix} \frac{\partial\mathcal{A}}{\partial q} \\ \frac{\partial\mathcal{A}}{\partial v} \end{bmatrix}^T \left(\begin{bmatrix} G_p & \\ & G_q \end{bmatrix}\begin{bmatrix} \frac{\partial\mathcal{A}}{\partial q} \\ \frac{\partial\mathcal{A}}{\partial v} \end{bmatrix} - \begin{bmatrix} \dot{q}\beta/2 + g(\bar{q})\lambda/2 \\ \dot{v}\beta/2 + h(\bar{q})\lambda/2 \end{bmatrix}\right),$$

where $\lambda g(q)$ and $\lambda h(q)$ introduce external forces. Next, we make the one-half step substitutions,

$$\begin{bmatrix} \frac{\partial\mathcal{A}}{\partial q} \\ \frac{\partial\mathcal{A}}{\partial v} \end{bmatrix} \rightarrow \begin{bmatrix} M(f(q) + F(q)\dot{q}\epsilon/2 - \dot{v}) \\ M(\dot{q} - v - \dot{v}\epsilon/2) \end{bmatrix}, \quad g(\bar{q}) \rightarrow g(q) + G(q)\dot{q}\epsilon/2,$$

using the force-per-mass, f, the appropriate derivative matrices $\{F, G, H\}_{IJ} = \partial\{f, g, h\}_I/\partial q_J$, and defining $J_q \equiv MG_qM \equiv C_q^{-1}/2$, $J_p \equiv MG_pM \equiv MC^{-1}M/2$. Factoring η gives a normal distribution for $[\dot{v}, \dot{q}]^T\epsilon$ with penalty matrix (inverse of the variance-covariance matrix, keeping terms below $O(\epsilon)$),

$$P = \begin{bmatrix} J_p + \epsilon M\beta/4 & \epsilon(MG\lambda/4 - F^T J_p - J_q)/2 \\ \epsilon(G^T M\lambda/4 - J_p F - J_q)/2 & J_q - \epsilon(MH\lambda + MF\beta)/4 \end{bmatrix}/\epsilon,$$

and mean (to first order in ϵ),

$$\left\langle \begin{bmatrix} \dot{v}\epsilon \\ \dot{q}\epsilon \end{bmatrix} \right\rangle = \begin{bmatrix} f - M^{-1}C(\tfrac{\beta}{2}v + \tfrac{\lambda}{2}g) \\ v + C_q M(\tfrac{\beta}{2}f + \tfrac{\lambda}{2}h) \end{bmatrix}\epsilon. \tag{A2}$$

These expressions are in accord with Eqns. 18 and 20.

The residual terms contribute to form the transition free energy functional (again to order ϵ),

$$\mathcal{F} = \tfrac{1}{2}\Big[\ln\tfrac{|2\epsilon P|}{(2\pi)^d} - (\beta v + \lambda g)^T C(\beta v + \lambda g)\epsilon/4 \tag{A3}$$

$$- (\beta M f + \lambda M h)^T C_q(\beta M f + \lambda M h)\epsilon/4\Big]. \tag{A4}$$

Note that the Fokker-Planck equation can be used to prove that the Boltzmann distribution is stationary under either the Langevin ($C_q \to 0$) or Brownian ($C \to 0$) limits, but not both. For the Langevin limit, it can be checked that the derivative of this equation with respect to $\beta/2$ gives Eq. A1. For the Brownian limit, we find again Eq. 44. These rely on the following expansion for the derivative of the log-determinant term,

$$\frac{\partial \ln|\epsilon P|}{\partial \alpha} = \mathrm{Tr}\left((\epsilon P)^{-1}\frac{\partial(\epsilon P)}{\partial \alpha}\right)$$

$$\epsilon P = P_0 + \epsilon P_1 + O(\epsilon^2), \ P_0 \equiv \begin{bmatrix} J_p & \\ & J_q \end{bmatrix}$$

$$(\epsilon P)^{-1} = P_0^{-1} - \epsilon P_0^{-1} P_1 P_0^{-1} + O(\epsilon^2).$$

Since $\frac{\partial(\epsilon P)}{\partial \alpha}$ should contain a prefactor of ϵ, the second term is usually unimportant, so that

$$\frac{\partial \ln|\epsilon P|}{\partial \alpha} = \mathrm{Tr}\left(2\begin{bmatrix} M^{-1}CM^{-1} & \\ & C_q \end{bmatrix}\frac{\partial(\epsilon P)}{\partial \alpha}\right).$$

Using a derivative, it is simple to find the constrained-momentum averages of Eq. 26

$$\langle dI_p \rangle = \left\langle -\tfrac{\delta \mathcal{A}}{\delta x(t)}\mathbb{1}\right\rangle = \frac{\partial \mathcal{F}}{\partial \lambda/2} \tag{A5}$$

This is just the derivative of Eq. A3 with $C_q = 0, g = \mathbb{1}$. The associated work, $dW_p = -\epsilon\tfrac{\delta \mathcal{A}}{\delta x(t)}\mathbb{1}\mathbb{1}^T \dot{x}/N$, is more complicated. It can be found by introducing another constraint, $b = \mathbb{1}\mathbb{1}^T \dot{x}/N$, with multiplier $\gamma/2$. Then its average is

$$\langle dW_p \rangle = \left\langle -\tfrac{\delta \mathcal{A}}{\delta x(t)}(\mathbb{1}\mathbb{1}^T \dot{x}\epsilon/N)\right\rangle \tag{A6}$$

$$= \frac{\partial}{\partial \gamma/2}\left[-\tfrac{\epsilon}{8}(\beta_0 v + \lambda\mathbb{1} + \gamma b)^T C(\beta_0 v + \lambda\mathbb{1} + \gamma b)\right]\Big|_{\gamma=0} \tag{A7}$$

$$= -\tfrac{\epsilon}{2}b^T C(\beta_0 v + \lambda\mathbb{1}) \tag{A8}$$

The relations below Eq. 26 re-define $\lambda \to -\lambda'$ from the above to increase rather than decrease the momentum in the direction $\mathbb{1}$, so that if the momentum encounters zero resistance, $\langle v \rangle = \lambda'\mathbb{1}/\beta_0$.

Appendix B: Jarzynski's Equality

In the special case where a stationary distribution is known during each time-propagation step, a set of useful equalities can be derived simply from the re-weighting equation,

$$\langle a(x(T))|\text{ss}, T \rangle = \left\langle a(x(T)) \prod_{t=1}^{T} \frac{\mathcal{P}_{\text{ss}}\left(x(t), t\right)}{\mathcal{P}_{\text{ss}}\left(x(t), t-1\right)} | x_0 \sim \text{ss}, 0 \right\rangle. \tag{B1}$$

The first average represents a single-time average over the steady-state distribution used to propagate the system during the transition $T \to T+1$. The second average is a path average over stochastic trajectories beginning in the steady-state at time zero. For the particular choice $a = 1$, the normalization constants cancel between successive $\mathcal{P}_{\text{ss}}\left(x(t), t\right)$, leaving only the ratio between the end-point partition functions to be expressed as a path average.

The proof[16] is by recursion from the property of the stationary distribution under time-propagation,

$$\mathcal{P}_{\text{ss}}\left(x(t+1), t\right) = \int \mathcal{P}_{\text{tr}}\left(x(t+1)|x(t), t\right) \mathcal{P}_{\text{ss}}\left(x(t), t\right) \; dx(t).$$

These relations will work for any stationary distribution of the transition probability. Equality of the starting and ending temperatures is not required. Combinations of equations 2 and B1 can give a very large number of useful fluctuation theorems.

Appendix C: Reviewer Comments

This paper essentially appears in David M. Rogers and Susan B. Rempe. "Irreversible Thermodynamics." J. Phys.: Conf. Ser. 402:012014, 2012 – which deserves the citation for this work. It was originally submitted to a different Journal. The initial feeback was borderline reject based, I believe, on the paper's complicated presentation. These are the responses included in the re-submission (which also failed). Some credit goes to the reviewers for identifying areas of complication. Since feedback generated during that review process may be of general interest, it has been included here. The reviewer comments and responses are included, except for translating some math into LaTex and removing the title of the original journal.

Monday, December 12, 2011 Editor, [journal]

Dear Editor;

We resubmit for review our revised manuscript entitled "A first and second law for nonequilibrium thermodynamics: Maximum entropy derivation of the fluctuation-dissipation theorem and entropy production functionals."

We appreciate the comments provided by the reviewers and have responded to each in detail below. We take the issue of manuscript length seriously and have revised our manuscript to shorten it, to highlight the new results, and to add clarity where requested. We believe that the manuscript is now in suitable form for publication.

Sincerely, David M. Rogers

1. Report of the First Referee

I have looked through the Rogers/Rempe paper and believe it is not suitable for [this journal].The paper presents a relatively slow-moving set of claims for results in nonequilibrium statistical mechanics using a maximum-entropy basis.

My objections to the paper are threefold:

1) It is too long for the content.

This was also noted by the second reviewer and has been corrected in the present revision.

2) The content does not contain sufficient new results (things that can be used to understand results from experiment of simulation in a novel manner).

As noted by the second reviewer, our report shows the application of ideas stemming from Jaynes to problems that postdate his work. The energy analysis is novel compared to the standard local equilibrium treatment because it does not assume a linear dissipation law or a local energy or entropy at any point. This was highlighted in the application to the optically trapped bead, where an exact expression for the entropy increase was obtained in terms of information entropy plus dissipated work. Other treatments of this problem have left the question of a universal entropy measure for arbitrary pulling processes (i.e. not between steady-states) unanswered. The paper gives methods for doing constant force NEMD, comparing constant force and constant flux NEMD, and computing local, time-dependent temperatures and system entropy production rates. These should greatly benefit the simulation community.

3) The maximum-entropy approach is itself flawed–the fact that many coexisting solutions of Rayleigh-Benard flow fit exactly the same boundary conditions

(different numbers of convective rolls with different orientations) but have different entropies is a simple counterexample.

Although we did not specifically treat continuum hydrodynamics in this report, we can make some comments to inform the reviewer on this point. First, there has historically been a debate on the extremum principle for deriving nonequilibrium relations (minimum vs. maximum). The application of these ideas to the Rayleigh-Benard problem gained wide interest in the work of Paltridge on heat conduction in atmospheric climate models. Ozawa, Ohmura, Lorenz, and Pujol (Rev. Geophys., 41(4):1018 2003) review this body of work and conclude with Paltridge and Jaynes (1980), that "long-term mean states ... correspond to a certain extent to a unique state in which the rate of entropy production due to thermal and viscous dissipation is at a maximum."

The caveat they noted concerns the linear transport regime, in which Prigogine showed that the linear transport laws lead to a minimization of the entropy production. Jaynes (1980) quite convincingly showed that this minimum principle reduces to energy and species conservation laws in all cases in which it is applicable. Ozawa gave a short argument that this principle is valid only when the assumption of no advective transport holds, and it has not been successfully extended to the nonlinear case.

Our report does not make the maximum entropy production claim, but instead a weaker claim of maximum entropy of the transition probability distribution. We showed quite clearly how this leads to the equation for Brownian motion, upon which the linear transport theory rests. It also synthesizes the Langevin equation describing advection of momentum. The steady-states of the system correspond to the stationary points of the associated Fokker-Planck equations and do not necessarily correspond to states of minimum or maximum thermodynamic entropy production. No preference between alternative solutions (except by reason of their dependence on initial conditions) can be generated by the arguments in our report, and it is unclear in what sense the reviewer is using the word entropy. The difference between the thermodynamic entropy production and transition entropy have been discussed in the work of Attard.

4) I suggest that the manuscript not be published in [this journal] as it does not contain sufficient useful novel results for its target audience.

The opinion of this reviewer appears to be biased against maximum entropy methods to the extent that new results using this method are deemed too lengthy to justify the

content. We contend that our report is relatively short for introducing an extremum principle that can be used to derive a large class of material in this field. Other works have run into similar difficulties expressing nonequilibrium ideas in a short space. See, for example, the treatments of the nonequilibrium statistical operator method by Zubarev or extended irreversible thermodynamics by Jou or the string of papers by Attard. The second reviewer noted interest in the novel results of Jaynes' caliber and had useful suggestions for shortening the manuscript which we have acted upon in the present revision.

2.   Report of the Second Referee

General comments:

The authors use a path ensemble method(Jaynes' caliber) to derive a path probability ensemble. The constraints are formulated in terms of path transition probabilities expressed using the system's classical action. The result is an elaboration of Jaynes' theory which the authors use to derive equations of motion(Langevin equation) as well as results of recent interest in non-eq. stat mech(fluctuation theorems) which postdate Jaynes' work. Jaynes rather sketched derivations of already known relations at his time such as those of Onsager. This manuscript also draws analogies to information theory and the second law.

The main conclusions of the paper are the path probability(Eq.(18)) and ensuing Langevin equation(Eq.(19) through Eq.(21)). While scientifically sound, the paper gives the impression of a first draft where every possible tangent is explored without discrimination. Many of the topics discussed can simply be referenced as they are not original to the current manuscript. More care should be given to those ideas which are original to this work. A serious re-focusing of the manuscript is required upon a second iteration.

A significant amount of work has gone into this paper to present a consistent re-interpretation of nonequilibrium thermodynamic theory. The format is necessary for removing extraneous assumptions, such as a dependence on the Boltzmann distribution and expressions only valid as linear deviations from it, that often enter the conventional theory. The topics explored are confined to the underpinnings of the theory and represent only a

42

small fraction of the important questions and applications which it can serve to expand upon. Some discussion of these further topics has been added to the conclusions. We also comment on the originality of our findings in our response to point 13.

On a different note, there are many citations(when there are too many, they stop being useful because the reader does not now which one to turn to). Having said this, the authors should cite others who have used caliber to derive equations of motion: Stock, Ghosh and Dill in J. Chem. Phys.(since these have also derived Langevin equations using caliber), Presse, Ghosh and Dill in J Physical Chemistry B.(these have derived mass action laws using caliber). The authors should also include a reference to the classic work of Zubarev on path ensembles and that of Filyukov and Karpov on the same topic (These last two methods are related to Jaynes' caliber).

We believe our selection of references is adequately annotated in the text, with the intention of providing appropriate background material on statements which require further qualification. The literature on nonequilibrium is already large and complex, and it is partly for this reason that we are attempting to show how alternative viewpoints relate to our own. In this respect, we are grateful to the reviewer for providing recent references on the maximum caliber approach, and have included them in the revised version.

The manuscript is focused around showing the consequence of the maximum transition entropy principle. Its importance stems from the fact that it can be used to derive such a large portion of the existing theory, and therefore requires some length to motivate and derive these relations. The definition of work values and total entropy as a sum of system information plus heat exchange (revised Eqns. 31 and 32) is unique to this paper. Previous treatments have rather worked backward from the transition probability ratio without a solid connection to work values. This has lead to improper emphasis on fluctuation theorems (of which there are infinitely many) and an associated uncertainty as to how to proceed even in the case of steady-state to steady-state transitions. Note that the stationary states of a given transition probability may not be unique, as in the case of Reyleigh-Bernard flow noted by the first reviewer. This is also a difficulty of the nonequilibrium statistical operator method of Zubarev, which relies on a Boltzmann-Gibbs probability form absent from our report. Starting instead from a prescription for the transition probability, we see all these special cases falling out, and it would be a disservice to the readers to remove these connections

from the manuscript.

Specific inquiries:

1) Beginning of the intro: can the authors be more specific regarding what they mean by first law of non-equilibrium thermo? Is this a new concept? The traditional laws of thermo are well described in the classic works of Tisza and Landsberg who, to the best of my knowledge, never describe first law of non-eq thermo. In addition, Onsager and Kubo never described their relations as being a first law either.

Yes, this is a new concept which comes by applying the equilibrium analysis to the partition function of the transition probability distribution. The text has been changed to indicate this.

2) Page 4, definition of Kullback-Leibler: is there a sign missing?

There is no sign missing, as the divergence is defined to be non-negative, whereas the (relative) entropy attains its maximum at zero.

3) Eqns.(4)-(11): These expressions need to be cleaned up. The definition of the information measure is the first line of Eq.(4). Where is the second line of Eq.(4) coming from? Next, the partition function Z depends on different variables in Eq.(4) and (7). What is going on? I am guessing there are two definitions of Z that are not being made explicit and that somehow this results in the second term of Eq.(4) which I don't understand. Eq.(10) and (11) define a ψ. How is this related to the current discussion?

Eqns. now numbered 5-10 are the standard set of relations obtained by applying Jaynes' derivation method (ref. 30) to the transition probability distribution, given a fixed starting point x_i. Both Z-s are equivalent. Some discussion has been added.

4) Bottom of page 8: Define A in the exponential.

A more specific comparison between the path probability expressions (three unnumbered equations following Eq. 10) has been given that removes possible ambiguity.

5) On page 9: In discussing the Langevin derivation from caliber, the authors should cite Stock, Ghosh and Dill in J. Chem. Phys. as well as Presse, Ghosh and Dill in J Physical Chemistry B at this point.

There are many Langevin equations, and the discrete form arrived at in those refs is more applicably cited when discussing particle insertion/deletion later in the report. We have also

added a reference to this work in the section describing other research using the maximum caliber approach.

6) Above Eq.(14): do you mean $\{x, \dot{x}\} = \{q, v\}$ rather than $x = \{q, v\}$? If not, I don't understand how terms like $\ddot{v}$ arise.

$x = \{q, v\}$, meaning both $q(t)$ and $v(t)$ are treated as free parameters in the variation of $A[x]$, and the Lagrangian is $L(\{q, v\}, \{\dot{q}, \dot{v}\})$. There is no $\ddot{v}$ term, the result of the variation gives first-order equations $\dot{q} = v$ and $M\dot{v} = F$.

7) Top of page 11:define G.

G is defined in the discussion following (revised #) Eq. 14 as the Lagrange multiplier constraining dA/dx.

8) Middle of page 11: what does $G \to I\infty$ mean?

added: G is a matrix, I is the identity (fixed avoid overloading I)

9) Bottom of page 12 (Important question that needs to be addressed): Why do you select only these particular constraints amongst many possible other options? Is is related to Noether's thm?

The selection of constraints is problem-specific, and is discussed in the paragraphs beginning: "Note the information-theoretic quality..." and "Connecting back to the usual thermostatics..." We have re-emphasized our references these statements in the conclusion section.

10) First line of Eq.(18): what constraint does the integral term (the second term in the exponential) come from? I understand where the first and third terms come from. Not the second.

This was discussed in the definition of G (following Eq. 14) as constraining the correlation between $dA/dx(t)$ and $dA/dx(t')$. In this formulation, history-dependence enters through these constraints rather than through a projection-operator.

11) Eq.(19): this expression is important and the authors should put in one or two steps to derive it. Right now, I can see how they are probably using the path probability at the top of the page and integrating over μ but there needs to be more detail.

The derivation comes from substituting the quantities from Eq. 17 into the sampling formula for the normal distribution: $dA/dx(t) = \mu + C^{1/2}z$ The explanation in the text has been revised.

12) second line of Eq.(34): shouldn't the probability appearing in the numerator of the logarithm be the prior(thus making the argument of the logarithm 1)? Otherwise, definitions at the bottom of page 21 need to be clarified.

According to the reviewer's suggestion, we have removed an explicit discussion of the caliber from our report. To address the reviewer, since the report does not assume a Boltzmann distribution for any state, the initial state is treated as a general, known, initial distribution $P(x_0)$. The caliber could be defined either including the entropy of this initial state or not by a simple re-definition of the reference, prior, probability $P_0(x_0)$. Usually, we are thinking of the prior probabilities as uniform distributions as in the usual Cartesian interpretation of stat. mech. In any case, (old eq. number) 34 and 35 were consistent using this definition. Our page numbering is slightly offset from the reviewer's, and so we are not sure exactly what definition is being referenced. Eqn.s now numbered 36 and 37 are also correct, since the reference transition probabilities can be converted to reference stationary probabilities using Bayes' theorem noted just below Eq. 36.

13) Most of the content of pages 14-30 are not original to this manusript (they have appeared elsewhere). Simple citations would suffice.

Although we would argue along with Stock, Presse, Ghosh, and Dill that the existence of a path free energy functional greatly clarifies the status of the Green-Kubo and fluctuation theorems as arising from information theory, we agree with the reviewer that this material is scattered throughout other works written from a large number of different perspectives. We have gone through these sections and attempted to find text that can be replaced by appropriate references. Unfortunately, the equation statements required for defining these issues comprise most of the repeated material. A major problem with, e.g. Gallavotti's 1996 PRL paper connecting the FT and GK is the large number of implicit assumptions and limits taken before the result can be derived, leaving us without a simple way to show when violations should be expected. Our text generalizes these relations, fixes the definition of the path free energy functional for application to maximum *transition* entropy paths, and as adds new ideas. By doing this, we have expounded an approach that is free of implicitly defined operators and Fourier transforms that obscure the physical picture of driven dynamical systems, higher-order moments of the fluctuation theorem are apparent, and the physical basis of entropy production has become much more apparent. In particular, we would like to point out that our application to the optically-trapped bead shows, in

addition to the heat flux considered in other works, the importance of the instantaneous information entropy (Eqns. 31-32 the main, novel result of Sec. 3) in completing a general statement for the second law.

As another example, Stock, Ghosh, and Dill (2008) note in connection with the second derivatives of the caliber, "Perhaps expressions such as Eq. (2.13) will be useful for dynamics in the same way that Maxwell's relations are for equilibria." In fact, the definition of the path free energy functional in Eq. 33 necessary for deriving this connection in Section 4 has never been considered before. The derivation of ensemble transformation formulas which have been omitted from the present revision occupy two pages of Callen's book, and arrive at a weaker statement involving matrix minors. The derivation of the adiabatic susceptibility (Eq. 4.1.19) of Kubo, Toda, and Saito's second volume on nonequ. is by means of a quasi-static perturbation of a canonical distribution from a Hamiltonian, and the student may never guess that the transformation from constant temperature to constant heat flux is an application of this general pattern. Again for the equation just before Eq. 38, the physical content of the fluctuation relations as probability ratios for paths with different constraints is often obscured in the literature by the consideration of time-reversal, the assumption of a time-independent temperature, and a reference to a steady-state. This has lead to the recognition of multiple fluctuation theorems with different physical interpretations – see, e.g. Chernyak, Chertkov, and Jarzynski, J. Stat. Mech. P08001 (2006) and Esposito and Broeck, Phys. Rev. E 82:11143 (2010). The second half of section 4 shows how such relations may be derived in a few lines and how they may be applied to the novel entropy measure of Eq. 31. This important section concludes by showing that the caliber and the thermodynamic entropy production are not the same, arguing against a simple, universal minimum/maximum (thermostatic) entropy production views of nonequ. thermo.

Nevertheless, we have removed some of the content from these sections in accordance with the wishes of the reviewer.