MIT Open Access Articles

Limits of predictions in thermodynamic systems: a review

Limits of Predictions in Thermodynamic Systems: A Review

Robert Marsland III

Department of Physics, Boston University, 590 Comm. Ave., Boston, Massachusetts 02215, USA[*]

Jeremy England

Physics of Living Systems, Massachusetts Institute of Technology,
400 Technology Square, Cambridge, Massachusetts 02139, USA

(Dated: November 16, 2017)

The past twenty years have seen a resurgence of interest in nonequilibrium thermodynamics, thanks to advances in the theory of stochastic processes and in their thermodynamic interpretation. Fluctuation theorems provide fundamental constraints on the dynamics of systems arbitrarily far from thermal equilibrium. Thermodynamic uncertainty relations bound the dissipative cost of precision in a wide variety of processes. Concepts of excess work and excess heat provide the basis for a complete thermodynamics of nonequilibrium steady states, including generalized Clausius relations and thermodynamic potentials. But these general results carry their own limitations: fluctuation theorems involve exponential averages that can depend sensitively on unobservably rare trajectories; steady-state thermodynamics makes use of a dual dynamics that lacks any direct physical interpretation. This review aims to present these central results of contemporary nonequilibrium thermodynamics in such a way that the power of each claim for making physical predictions can be clearly assessed, using examples from current topics in soft matter and biophysics.

I. INTRODUCTION

Modern society is built on the transformation of heat into mechanical work. The task of perfecting the heat engine was made possible by theoretical insights into fundamental bounds on its efficiency. By calculating the maximum amount of work extractable from a given quantity of heat by an engine operating between two thermal reservoirs of fixed temperature, Sadi Carnot was able to show that state-of-the-art steam engines in 1824 could still be significantly improved [17]. By determining the kinds of processes that saturate the bound, he was able to offer practical suggestions for making these improvements.

Biotechnology and biologically inspired design have opened up new engineering challenges that push the boundaries of thermodynamics. Living systems use heat to accomplish a wide variety of tasks in addition to performing mechanical work, such as replicating complex structures with high fidelity [52, 53, 62], maintaining robust clocks [6, 16], and sensing the state of their environment [42, 44, 66]. Most of these processes are intrinsically nonequilibrium, in the sense that they cannot be represented even approximately as a sequence of states of thermal equilibrium. Since they never approach the limit of zero entropy change, such processes are left unconstrained by classical thermodynamic theory. Similar issues arise in contemporary soft matter physics, particularly with the advent of active matter systems that constantly transduce chemical energy into mechanical force on microscopic scales [47, 57, 65].

In the 1950's-1970's, the theory of Linear Irreversible Thermodynamics was developed to analyze chemical reactions and various transport problems near thermal equilibrium from a thermodynamic perspective (cf. [23] for a thorough introduction). Many hoped that these results could be developed into a general theory of nonequilibrium thermodynamics, capable of making universal claims analogous to the Carnot bound on engine efficiency [55, 61]. Ilya Prigogine's Principle of Minimum Entropy Production initially seemed like a promising starting point for constructing a more comprehensive theory, but counterexamples were soon identified that revealed its restricted range of validity (cf. [61, p. 100ff.],[43]).

The past twenty years have seen a resurgence of interest in nonequilibrium thermodynamics, based on advances in the theory of stochastic processes and in their thermodynamic analysis. Stochastic Thermodynamics has matured into a systematic theory of nonequilibrium processes, in which the analogies to thermal equilibrium that Prigogine and others were searching for can be mathematically defined [71, 76]. But these general results carry their own limits: far from equilibrium, they demand precise knowledge of the probability distributions of the relevant quantities, which are often experimentally inaccessible [37], or they invoke a "dual" dynamics that lacks any direct physical interpretation [68]. This review aims to present a selection of central results from contemporary nonequilibrium thermodynamics in such a way that the power of each claim for making physical predictions can be clearly assessed.

After setting up our notation and theoretical framework in Section II, we proceed in Section III to consider the immediate consequences of imposing a consistent thermodynamic interpretation on a stochastic model. These include fluctuation theorems, the role of Shannon entropy, and the relationship of entropy to transition rate ratios at a coarse-grained level. We illustrate these ideas and their potential applications by reviewing a recent

[*] marsland@bu.edu

effort to obtain general design principles for nonequilibrium self-assembly from thermodynamic considerations.

The practical utility of the results of Section III is limited by their dependence on information about exponentially rare fluctuations, which becomes important far from equilibrium. Recent advances in the study of such fluctuations in the field of Large Deviation Theory have significantly pushed back this limit. In Section IV, we discuss the assumptions and consequences of a particularly powerful result that relates these rare fluctuations to the small fluctuations that could actually be observed in an experiment [29, 59]. This theorem leads to a "thermodynamic uncertainty relation," which places a limit on the allowed dynamical precision of a nonequilibrium process based on its rate of entropy production [5].

If we now imagine perturbing a driven system – compressing an active material, for instance – many analogies to classical thermodynamics suggest themselves. In equilibrium, we could have computed the force resisting the perturbation by taking a derivative of free energy; we could have predicted typical values of observables by minimizing the free energy; and we could have constrained the minimum work required to change they state of the system with the difference between initial and final free energies. In Section V, we discuss the extent to which these three kinds of thermodynamic predictions can be generalized to transitions between nonequilibrium steady states.

We believe that this selection of topics will provide a helpful unifying perspective on some of the most important themes in contemporary nonequilibrium thermodynamics, particularly for readers interested in biophysical or soft-matter applications. But a number of current lines of inquiry have been left outside the scope of this review. Stochastic Thermodynamics has shed new light on the original thermodynamic problem of heat engine performance (cf. [12, 63]), consolidating the considerable progress made in Finite-Time Thermodynamics throughout the second half of the 20th century [1, 56]. This new perspective has also helped elucidate the nature of fluctuation-response relations, showing how the traditional formulations can be generalized to far-from-equilibrium scenarios [4, 48]. A comprehensive review of Stochastic Thermodynamics from 2012 summarizes the development of both of these themes up through that year [71].

II. MICROSCOPIC REVERSIBILITY CONNECTS ENTROPY TO DYNAMICS

In thermal equilibrium, the principle of detailed balance relates transition rates between system states $\mathbf{x}$ to the energies of those states $U(\mathbf{x})$, by requiring that all probability flux vanish in the Boltzmann distribution. We can express this requirement in terms of the probability that the system passes through a continuous sequence of states $\mathbf{x}_t$ during a window of duration τ, starting from state $\mathbf{x}_0$ at time $t = 0$. We denote the whole trajectory by $\mathbf{x}_0^\tau$, and the time-reversed version of this trajectory by $\hat{\mathbf{x}}_0^\tau$. At time t after initialization, a system following the time-reversed trajectory will be found in state $\mathbf{x}_{\tau-t}^*$, where $\mathbf{x}^*$ is the state obtained from $\mathbf{x}$ by reversing the signs of all momentum degrees of freedom that may be contained in that vector.

Detailed balance requires that the ratio of forward and reverse trajectory probabilities $p(\mathbf{x}_0^\tau|\mathbf{x}_0)$ and $p(\hat{\mathbf{x}}_0^\tau|\mathbf{x}_\tau^*)$ conditioned on their respective initial states must be equal to the ratio of Boltzmann weights:

$$\frac{p(\hat{\mathbf{x}}_0^\tau|\mathbf{x}_\tau^*)}{p(\mathbf{x}_0^\tau|\mathbf{x}_0)} = e^{\frac{1}{k_B T}[U(\mathbf{x}_\tau) - U(\mathbf{x}_0)]}. \tag{1}$$

The cornerstone of contemporary nonequilibrium statistical mechanics is a generalization of Equation (1) known as "local detailed balance" or "microscopic reversibility," which applies to driven systems arbitrarily far from thermal equilibrium. Since this principle is so fundamental, it is important to understand its physical basis and its range of validity before proceeding to more specific results.

A. Stochastic processes from classical mechanics

Consider an isolated chunk of classical matter with Hamiltonian H_{tot}, whose state is described by a set of $\mathcal{N}$ coordinates q_i with conjugate momenta p_i. The dynamics are given by Hamilton's equations

$$\dot{q}_i = \frac{\partial H_{\text{tot}}}{\partial p_i} \tag{2}$$
$$\dot{p}_i = -\frac{\partial H_{\text{tot}}}{\partial q_i}.$$

We will keep track of some subset of the $\mathcal{N}$ degrees of freedom, which we call the "system" and denote by $\mathbf{x}$, and refer to the rest of the degrees of freedom $\mathbf{y}$ as the "environment." We can now split the Hamiltonian into three parts: one that depends on the system degrees of freedom alone, one that depends on the environment alone, and one that couples the two sets together:

$$H_{\text{tot}}(\mathbf{x}, \mathbf{y}, t) = H_{\text{sys}}(\mathbf{x}, \lambda_t) + H_{\text{env}}(\mathbf{y}) + h_{\text{int}}(\mathbf{x}, \mathbf{y}) \tag{3}$$

We have allowed for an explicit time-dependence in the Hamiltonian via a control parameter λ (e.g., the position of the piston in the classic example of gas compression), which can be used to do work on the system according to a pre-defined protocol λ_0^τ and drive it out of equilibrium.

We now fix the initial condition of the system $\mathbf{x}_0$, and choose a random initial condition $\mathbf{y}_0$ for the environment from a uniform distribution over states with total energy $H_{\text{tot}} = U$. Since the dynamics of the whole setup are deterministic, the trajectory $\mathbf{x}_0^\tau$ is fully determined by this initial random choice. We can collect all the $\mathbf{y}_0$'s that produce trajectories within a small window around a

of λ is

$$W = \int_0^\tau \frac{\partial H_{\text{sys}}(\mathbf{x}_t, \lambda_t)}{\partial \lambda} \dot{\lambda}_t dt. \tag{5}$$

Now we can compute the probability of seeing the reverse trajectory $\hat{\mathbf{x}}_0^\tau$ if we reverse the momenta, choose the environment conditions from this new energy surface, and run the protocol λ_t in reverse:

$$p(\hat{\mathbf{x}}_0^\tau | \mathbf{x}_\tau^*, \hat{\lambda}_0^\tau) = \frac{\Omega^*(\hat{\mathbf{x}}_0^\tau)}{\Omega(U + W - H_{\text{sys}}(\mathbf{x}_\tau, \lambda_\tau))}. \tag{6}$$

As illustrated in Figure 1, the time-reversibility of the Hamiltonian dynamics (2) can be expressed as the statement

$$\Omega(\mathbf{x}_0^\tau) = \Omega^*(\hat{\mathbf{x}}_0^\tau). \tag{7}$$

This relation expresses the fact that every trajectory $(\mathbf{x}_0^\tau, \mathbf{y}_0^\tau)$ satisfying the equations of motion (2) has a reverse trajectory $(\hat{\mathbf{x}}_0^\tau, \hat{\mathbf{y}}_0^\tau)$ that also satisfies these equations. (It also assumes that phase space volumes $\Omega = \int \prod_i dp_i \, dq_i$ are measured with the Liouville measure $\prod_i dp_i \, dq_i$, which is conserved by the equations of motion.) Note that Equation (7) remains true even in the presence of magnetic fields, as long as the signs of these fields are reversed for the purposes of calculating $\Omega^*(\hat{\mathbf{x}}_0^\tau)$.

Combining Equations (4), (6) and (7) leads immediately to

$$\frac{p(\hat{\mathbf{x}}_0^\tau | \mathbf{x}_\tau^*, \hat{\lambda}_0^\tau)}{p(\mathbf{x}_0^\tau | \mathbf{x}_0, \lambda_0^\tau)} = \frac{\Omega(U - H_{\text{sys}}(\mathbf{x}_0, \lambda_0))}{\Omega(U + W - H_{\text{sys}}(\mathbf{x}_\tau, \lambda_\tau))}. \tag{8}$$

We can now use Boltzmann's microcanonical definition of entropy to express the right hand side in terms of the thermodynamic entropy $S_e = k_B \ln \Omega$ of the environment:

$$\frac{p(\hat{\mathbf{x}}_0^\tau | \mathbf{x}_\tau^*, \hat{\lambda}_0^\tau)}{p(\mathbf{x}_0^\tau | \mathbf{x}_0, \lambda_0^\tau)} = e^{-\Delta S_e / k_B}. \tag{9}$$

Equation (9) is known as the "microscopic reversibility relation," since it encodes the time-reversibility of the full microscopic dynamics in a coarse-grained stochastic dynamics [19, 27].

B. Entropy and Heat

The microscopic reversibility relation (9) has a remarkably wide range of validity, although the above derivation has been made somewhat more restrictive for the sake of clarity and simplicity. In particular, we have assumed that the environment is initialized in thermal equilibrium, so that the initial environmental states can be sampled from a uniform distribution over a constant-energy surface. The microcanonical definition of temperature $1/T = \partial S / \partial U$ then yields
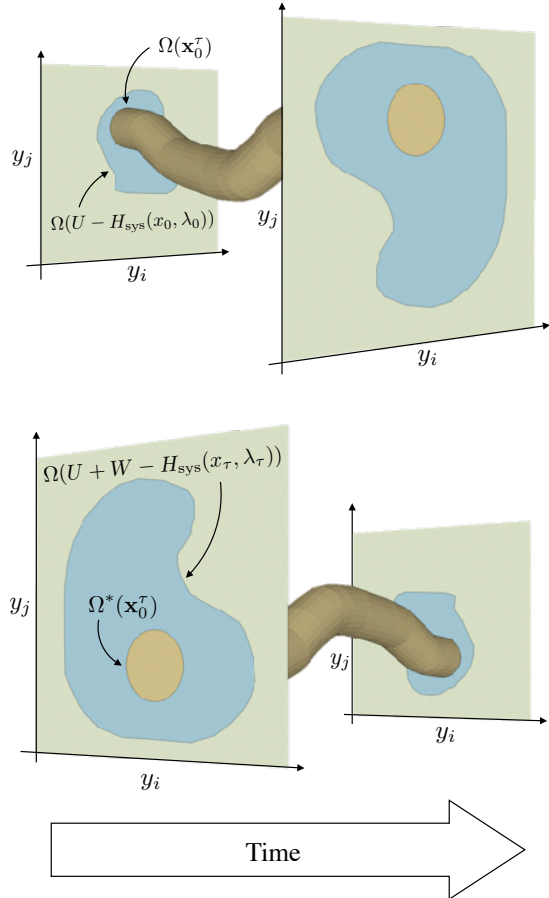
$$\Delta S_e = \frac{Q}{T} \tag{10}$$

FIG. 1. Color online. Time-reversal symmetry requires that the environmental phase space volumes giving rise to forward and reverse trajectories are identical. The planes represent the phase space of the environment, and the tubes represent the sets of environmental trajectories compatible with a given system trajectory $\mathbf{x}_0^\tau$ and its time-reverse $\hat{\mathbf{x}}_0^\tau$, respectively. The conservation of the Liouville measure implies that the beginning and end of the tube occupy the same phase space volume, and time-reversal symmetry requires that the initial conditions for the reverse tube are simply the momentum-reversed versions of the final conditions for the forward tube.

given $\mathbf{x}_0^\tau$, and denote the phase space volume occupied by this set of states as $\Omega(\mathbf{x}_0^\tau)$. The probability of observing a trajectory in the window is then given by the ratio of this volume to the total phase space volume $\Omega(U - H_{\text{sys}}(\mathbf{x}_0, \lambda_0))$ available to the environment at time $t = 0$:

$$p(\mathbf{x}_0^\tau | \mathbf{x}_0, \lambda_0^\tau) = \frac{\Omega(\mathbf{x}_0^\tau)}{\Omega(U - H_{\text{sys}}(\mathbf{x}_0, \lambda_0))}. \tag{4}$$

The energy of the environment changes over the course of the trajectory, as heat flows in and out of the system. At the end of the trajectory, it is equal to $E + W - H_{\text{sys}}(\mathbf{x}_\tau, \lambda_\tau)$, where the work W done by manipulation
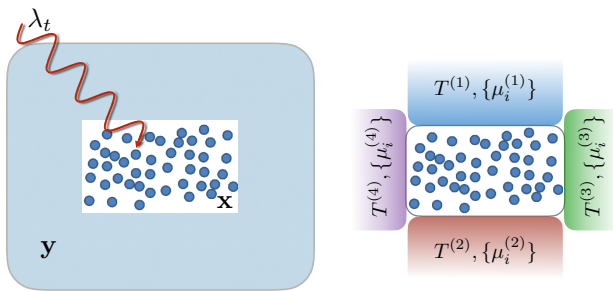
FIG. 2. Color online. Left: The $\mathcal{N}$ degrees of freedom in a piece of isolated matter are partitioned into two sets, a system with microstate $\mathbf{x}$ and an environment with microstate $\mathbf{y}$. The system is illustrated here as discrete particles, since we keep track of the full trajectories $\mathbf{x}_0^\tau$. The environment is a solid color, because we integrate it out to obtain an effective stochastic dynamics for $\mathbf{x}$. Work can be done on the system by externally imposed variations in a set of parameters λ_t, which affect only the system part of the Hamiltonian H_{sys}. Right: We consider situations where the environment can be modeled as a set of ideal thermal and chemical reservoirs with temperatures $T^{(\alpha)}$ and chemical potentials $\mu_i^{(\alpha)}$.

where $Q = W - \Delta H_{\text{sys}}$ is the amount of energy added to the environment over the course of the given trajectory.

But the derivation can be straightforwardly generalized to the case where different sectors of the environment are initially equilibrated separately with different temperatures $T^{(\alpha)}$ and are brought into contact with the system at time $t = 0$. The result also remains valid when particle exchange between system and environment is allowed, although this is more complicated to set up notationally (see [39] for derivation of a related expression allowing for particle exchange). If we include these possibilities, illustrated in Figure 2, we obtain a more general expression for the entropy change (cf. [60]):

$$\Delta S_e[\mathbf{x}_0^\tau] = \sum_\alpha \frac{1}{T^{(\alpha)}} \left(Q^{(\alpha)}[\mathbf{x}_0^\tau] - \sum_i \mu_i^{(\alpha)} \Delta n_i^{(\alpha)}[\mathbf{x}_0^\tau] \right). \tag{11}$$

where $Q^{(\alpha)}$ and $\Delta n_i^{(\alpha)}$ are the heat and number of particles of type i, respectively, delivered to sectors of the environment with temperatures $T^{(\alpha)}$ and chemical potentials $\mu_i^{(\alpha)}$.

In subsequent sections, it will be convenient to split ΔS_e into an equilibrium contribution $-\Delta H_{\text{sys}}/T$ and a nonequilibrium correction $\mathcal{W}$ that can be thought of as a generalized work:

$$T\Delta S_e = \mathcal{W} - \Delta H_{\text{sys}}. \tag{12}$$

The generalized work can be written explicitly as

$$\mathcal{W} = W + T\Delta S_e - \sum_\alpha Q^{(\alpha)}, \tag{13}$$

and reduces to the ordinary mechanical work W performed through the control parameters λ_t in the case

of an isothermal process with no particle exchange. This splitting highlights the fact that systems can be kept out of equilibrium by thermal or chemical "work" even if all the control parameters λ are held constant. In such cases, the system relaxes to a nonequilibrium steady state (in the limit of infinitely large reservoirs), which will be the main subject of Section V.

Since the Schrödinger Equation is time-reversible, a version of Equation (9) can also been obtained for quantum systems. In this case the system trajectory cannot be observed without interfering with the dynamics, however, and one has to consider the environment trajectory instead [21, 35]. The consequences of microscopic reversibility that we will discuss in Section III do not involve observation of fully-detailed system trajectories, and are thus easier to generalize to the quantum case [15].

In summary, the microscopic reversibility relation (9) depends on two basic assumptions: (a) the intrinsic dynamics of the system are time reversible, and (b) the environment is initialized in thermal equilibrium, or is partitioned into non-interacting sectors that are each separately initialized in thermal equilibrium.

C. Coarse-grained models

A nonequilibrium steady state can also be set up by other means, including stationary force fields with nonzero curl and externally imposed flow fields. Physically, all these forms of driving are ultimately reducible to the above framework. Experimental realizations of nonconservative force fields, for example, are actually created by rapidly varying some control parameters [11]. But it is convenient to model such a system by abstracting from the source of the flow field or effective force field, and these models can be numerically simulated without adding more information about the underlying oscillations or gradients.

Furthermore, many systems of interest in soft matter and biology consist of very small particles in aqueous solution. It is often a good approximation to let the momentum degrees of freedom and the microstate of the water molecules instantaneously relax to thermal equilibrium, taking the spatial location of the particles of interest to fully specify the system state $\mathbf{x}$. The dynamics of $\mathbf{x}$ are then described by an overdamped Langevin equation, which makes the particle's velocity proportional to the applied force and includes a random force due to the hidden degrees of freedom (cf. [28, 71]).

Chemically driven systems are usually represented with a similar sort of coarse-grained dynamics, with "fast" degrees of freedom rapidly reaching the Boltzmann distribution within a subset of microstates specified by "slow" variables. The slow variables in a chemical system refer to discrete metastable states, which could contain a collection of many similar protein conformations in the same free energy well, or specify the concentrations of

several molecular species in a well-mixed volume.

These coarse-grained models are no longer guaranteed to satisfy Equation (9), and it is necessary to directly verify that the equation does indeed hold under the natural definitions of heat and work in the model. This is straightforward for diffusion processes driven by non-conservative forces (cf. [68]) or by external flows [73], which satisfy (9) when the heat is obtained from the natural definition of work as force times distance and of energy as the integral of the conservative part of the force.

Chemical systems require more care. If the dynamics on the set of discrete states $\mathbf{x}$ is modeled as a Markovian stochastic process, and the assumption of separation of timescales holds, one obtains a modified form of (9). The heat Q in Equation (11) is replaced by the free energy difference $-\Delta F$ between the final and initial coarse-grained states. In general, this free energy is given by $F(\mathbf{x}) = -k_B T \ln Z_{\mathbf{x}}$ where $Z_{\mathbf{x}}$ is the sum of the Boltzmann weight $e^{-U/k_B T}$ over all the hidden degrees of freedom within the coarse-grained state $\mathbf{x}$ [71]. When $\mathbf{x}$ is a vector of chemical concentrations in a well-mixed volume, the free energy contains a part due to the intrinsic properties of the molecule, and part due to the sum over the possible spatial configurations of particles within the volume. In this case, the contribution to the free energy from each chemical species is equivalent to a chemical potential $\mu_i = \mu_{0,i} + k_B T \ln c_i$ [69].

In the limit of large system size, where the concentrations are naturally treated as continuous variables, chemical reaction dynamics are often modeled with a stochastic differential equation known as the Chemical Langevin Equation (CLE). Near equilibrium, the CLE still obeys Equation (9). But once the free energy change per reaction becomes of order $k_B T$, the ratio of reverse to forward trajectory probabilities no longer corresponds directly to any thermodynamic quantity [34].

III. CONSEQUENCES OF MICROSCOPIC REVERSIBILITY

The basic insights of equilibrium thermodynamics are founded on the Clausius relation

$$\Delta S \geq -\frac{Q}{T} \qquad (14)$$

which relates the change ΔS in the entropy of a system to the heat Q exhausted into an ideal thermal reservoir at temperature T. More generally, one can replace Q/T by ΔS_e, which could include contributions from particle exchange and multiple thermal reservoirs.

The entropy of equilibrium thermodynamics is only defined for equilibrium states. To obtain thermodynamic constraints on transitions between nonequilibrium states, it is necessary to define a generalized entropy that still obeys Equation (14). The Gibbs/Shannon entropy

$$S = -k_B \int d\mathbf{x}\, p(\mathbf{x}) \ln p(\mathbf{x}) \qquad (15)$$

is naturally generalizable to nonequilibrium states, since it remains well-defined even when $p(\mathbf{x})$ is not the Boltzmann distribution. J. Schnakenberg argued in the 1970's that Equation (14) should hold for this choice of the entropy in a broad class of Markovian stochastic processes on a finite set of states [69].

In this section, we review how a generalized version of Schnakenberg's result follows from microscopic reversibility (9), thus confirming that the Shannon entropy preserves the Clausius inequality away from thermal equilibrium. We then point out a further refinement that relates the average total entropy change to macroscopic irreversibility, placing a tighter constraint on processes that necessarily involve positive total entropy production. We end the section by discussing design principles for nonequilibrium self-assembly that can be derived from these results, and commenting on their range of applicability.

A. Fluctuation Theorem for Total Entropy Change

The full trajectories $\mathbf{x}_0^\tau$ that appear in Equation (9) are usually not directly observable. To obtain predictions from this symmetry, we need to integrate out the dependence on these microscopic details. As we will see, integrating Equation (9) over all trajectories leads directly to an integral fluctuation theorem (FT). This theorem involves all moments of the probability distribution for total entropy change. In this section, we will be concerned with the resulting inequality for the first moment, which is a form of the Clausius inequality. By retaining the information about fluctuations that is discarded in this derivation, one can obtain additional predictions, including the Fluctuation-Dissipation Theorem of classical linear response theory (cf. [20, Ch. 5]).

Multiplying $p(\mathbf{x}_0^\tau | \mathbf{x}_0, \lambda_0^\tau)$ by any normalized probability distribution $p_0(\mathbf{x}_0)$ generates a normalized distribution $\mathcal{P}(\mathbf{x}_0^\tau | \lambda_0^\tau)$ over the whole trajectory space. Likewise, multiplying $p(\hat{\mathbf{x}}_0^\tau | \mathbf{x}_\tau^*, \hat{\lambda}_0^\tau)$ by some $p_\tau(\mathbf{x}_\tau)$ generates another normalized distribution $\mathcal{P}(\hat{\mathbf{x}}_0^\tau | \hat{\lambda}_0^\tau)$. It follows immediately from Equation (9) and the normalization of these distributions that

$$\left\langle e^{-\frac{\Delta S_e}{k_B} + \ln \frac{p_\tau(\mathbf{x}_\tau)}{p_0(\mathbf{x}_0)}} \right\rangle = 1 \qquad (16)$$

where $\langle \mathcal{F}[\mathbf{x}_0^\tau] \rangle = \int \mathcal{D}[\mathbf{x}_0^\tau] \mathcal{F}[\mathbf{x}_0^\tau] \mathcal{P}(\mathbf{x}_0^\tau | \lambda_0^\tau)$ is the average value of the trajectory functional $\mathcal{F}[\mathbf{x}_0^\tau]$ over the forward trajectory distribution $\mathcal{P}(\mathbf{x}_0^\tau | \lambda_0^\tau)$.

If $p_\tau(\mathbf{x}_\tau)$ is chosen as the time-evolved version of $p_0(\mathbf{x}_0)$, then Equation (16) can be written in the more suggestive form

$$\left\langle e^{-\frac{1}{k_B}(\Delta S_e + \Delta s)} \right\rangle = 1 \qquad (17)$$

using the stochastic entropy

$$\Delta s = -k_B \ln p_\tau(\mathbf{x}_\tau) + k_B \ln p_0(\mathbf{x}_0). \qquad (18)$$

The average of this quantity over the forward trajectory ensemble gives the change in Shannon entropy of the system over the course of its forward evolution:

$$\langle \Delta s \rangle = -k_B \int d\mathbf{x}\, p_\tau(\mathbf{x}_\tau) \ln p_\tau(\mathbf{x}_\tau)$$
$$+ k_B \int d\mathbf{x}\, p_0(\mathbf{x}_0) \ln p_0(\mathbf{x}_0). \qquad (19)$$

Equation (17) is known as the integral fluctuation theorem for the total entropy change, and was first obtained in 2005 by Udo Seifert (who used it to justify the definition (18) of the stochastic entropy) [70]. This is a very powerful result, and almost everything we will discuss in the remainder of this review follows from it or from some closely related variant.

Using the fact that $e^x > 1 + x$, we recover the Clausius inequality

$$\langle \Delta s \rangle \geq -\langle \Delta S_e \rangle. \qquad (20)$$

This confirms Schnakenberg's thesis for any system obeying microscopic reversibility (9), including situations beyond the scope of his original paper where the dynamical rules change in time due to variations in a control parameter λ_t.

Equations (17) and (20) can be rewritten in terms of the thermodynamic work $\mathcal{W}$ defined in Equation (12), which will be particularly useful in Section V when discussing transitions between nonequilibrium steady states. The fluctuation theorem becomes

$$\left\langle e^{-\frac{1}{k_B T}(\mathcal{W} - \Delta U + T \Delta s)} \right\rangle = 1, \qquad (21)$$

This reduces to the Jarzynski Equality when the initial and final distributions are chosen as Boltzmann distributions, which turns $U - Ts$ into the equilibrium free energy F [36]. In general, this fluctuation theorem implies the isothermal form of the Clausius inequality:

$$\langle \mathcal{W} \rangle \geq \Delta \mathcal{F} \qquad (22)$$

where the nonequilibrium free energy is

$$\mathcal{F} = U - T \langle s \rangle. \qquad (23)$$

B. Statistical Irreversibility in Macroscopic Transitions

Our starting point in Equation (9) quantified the statistical irreversibility of a microscopic trajectory and related that quantity to the entropy change. By modifying a few steps in the above derivation of the integral fluctuation theorem, one can preserve more of this information during the integration over trajectories, bounding the entropy change with a positive number that quantifies the irreversibility of a macroscopic transition [25, 30, 33, 38].

One way of doing this is to categorize the microstates according to some empirical criterion [25]. This partitions the phase space into regions that we label

I, II, III, ... We can study the statistics of transitions between a pair of regions I and II by setting $p_0(\mathbf{x}_0) = 0$ everywhere outside of region I and $p_\tau(\mathbf{x}_\tau) = 0$ everywhere outside of region II. For simplicity, we will require the distributions within each region to satisfy $p(\mathbf{x}^*) = p(\mathbf{x})$.

The choice of $p_0(\mathbf{x}_0)$ is not compatible with our original derivation of the fluctuation theorem (17), which assumes that it is everywhere nonzero. But we can recover a sensible expression if we restrict our domain of integration to trajectories that begin in I and end in II [10, 75]. This yields:

$$\left\langle e^{-(\Delta S_e + \Delta s)/k_B} \right\rangle_{\mathrm{I} \to \mathrm{II}} = e^{\ln \frac{\pi(\mathrm{II} \to \mathrm{I})}{\pi(\mathrm{I} \to \mathrm{II})}} \qquad (24)$$

where $\pi(\mathrm{I} \to \mathrm{II})$ is the normalization of the trajectory distribution over the restricted domain

$$\pi(\mathrm{I} \to \mathrm{II}) = \int_{\mathrm{I} \to \mathrm{II}} \mathcal{D}[\mathbf{x}_0^\tau] p_0(\mathbf{x}_0) p(\mathbf{x}_0^\tau | \mathbf{x}_0, \lambda_0^\tau) \qquad (25)$$

and gives the total probability of arriving in II after time τ given that the system started in I.

The inequality $e^x \geq 1 + x$ then gives [25]:

$$\langle \Delta s \rangle_{\mathrm{I} \to \mathrm{II}} + \langle \Delta S_e \rangle_{\mathrm{I} \to \mathrm{II}} \geq -k_B \ln \frac{\pi(\mathrm{II} \to \mathrm{I})}{\pi(\mathrm{I} \to \mathrm{II})}. \qquad (26)$$

This is a tighter version of the Clausius inequality (20) for the case where $\pi(\mathrm{I} \to \mathrm{II}) > \pi(\mathrm{II} \to \mathrm{I})$. Generating internal order with a negative $\langle \Delta s \rangle$ requires dissipating *at least* the same amount of entropy into the environment, but this minimum can only be achieved when the system is just as likely to return back to I as it was to arrive in II from I.

In Equation (19) of the previous section, we noted that $\langle \Delta s \rangle$ is simply the change in the Shannon entropy of the distribution over states of the system. The restricted average appearing in Equation (26), however, has a more subtle interpretation. Using Bayes' Rule, we can express this quantity in terms of the conditional distribution $p_0(\mathbf{x}_0 | \to \mathrm{II})$ of finding the system in $\mathbf{x}_0$ at time 0 given that it ends up in II at time τ:

$$\langle \Delta s \rangle_{\mathrm{I} \to \mathrm{II}} = -k_B \int_{\mathrm{II}} d\mathbf{x}\, p_\tau(\mathbf{x}_\tau) \ln p_\tau(\mathbf{x}_\tau)$$
$$+ k_B \int_{\mathrm{I}} d\mathbf{x}\, p_0(\mathbf{x}_0 | \to \mathrm{II}) \ln p_0(\mathbf{x}_0). \quad (27)$$

The first term on the right hand side is the ordinary Shannon entropy of the final state, but the second term is a cross-entropy between the actual initial distribution and the new distribution conditioned on the final macrostate. The thermodynamic interpretation of $\langle \Delta s \rangle_{\mathrm{I} \to \mathrm{II}}$ as a change in internal entropy requires the additional assumption that the macrostate dynamics on the timescale of interest are effectively Markovian, so that the probability of a transition from I to II in time τ is independent of the exact initial microstate within I.

For a macroscopic system, even down to the level of a single biological cell, the relative probability of the reverse transition $\pi(\mathrm{II} \to \mathrm{I}) / \pi(\mathrm{I} \to \mathrm{II})$ quickly becomes

astronomically small. If I represents the state with a single bacterium on an agar plate, and II is the state with two bacteria, then the probability of the transition I to II in twenty minutes can be close to 1, but the probability of the reverse transition II → I has been estimated at $\exp(-10^{11})$ [25]. Even though this number is so small, Equation (26) does not allow us to set it to zero: that would make the right-hand side infinite, and require an infinite increase in entropy to maintain the inequality. The above estimate already sets the right-hand side to $10^{11}k_B$, which corresponds to significant heat production at 300 Kelvin. For a liter of exponentially growing bacteria at a reasonable density of 10^6/mL, half a Joule of heat per 30-minute division cycle must be dissipated even in the absence of a decrease in internal entropy.

The original form of the Clausius relation (20) requires only that the heat dissipation compensate the change in internal entropy, which has been estimated at $-10^{10}k_B$ [25]. This suggests that most of the heat generated by the bacterium during division goes into providing statistical irreversibility to the many intermediate steps of the process, ensuring that the key reactions reliably proceed in the correct direction.

C. Application to Self-Assembly

We now turn to a set of systems that can be explicitly modeled, and provide a framework for obtaining design principles from Equation (26). Consider a self-assembly process in which particles at chemical potential μ stick together to build up a structure. We parameterize the state of the assembly by the number of particles N it contains and a set of intensive parameters ω that characterize the internal structure and composition. If μ is large enough, the structure will spontaneously grow at a finite rate. One can now examine the thermodynamic constraints on achieving a given probability distribution over ω that differs from the Boltzmann distribution. Of particular interest are probability distributions that increase the yield of a desired structure, as occurs in kinetic proofreading situations. This problem has been studied by several authors in the context of a simple but surprisingly rich model of template-based replication [2, 3, 26, 67].

In this section, we use the tightened Clausius equality of Equation (26) to re-derive two recently obtained general thermodynamic bounds on the dissipative cost of a non-Boltzmann distribution over ω [54]. We first note that the environmental entropy change over a transition from N to $N + \Delta N$ is given by (11):

$$\langle \Delta S_e \rangle_{N \to N + \Delta N} = \frac{\mu \Delta N - \langle U \rangle_{N+\Delta N} + \langle U \rangle_N}{T} \qquad (28)$$

where $\langle U \rangle_N$ is the average energy at fixed N in the growing structure.

To express the internal entropy change, we write the nonequilibrium distribution over microstates at fixed N

in terms of an effective energy and free energy:

$$p_N(\omega) = e^{-\frac{1}{k_B T}[U^{\mathrm{eff}}(N,\omega) - F_N^{\mathrm{eff}}]} \qquad (29)$$

where $F_N^{\mathrm{eff}} = -k_B T \ln \sum_\omega e^{-\frac{U^{\mathrm{eff}}(N,\omega)}{k_B T}}$. Increasing the yield of a given structure ω is now equivalent to decreasing the relative $U^{\mathrm{eff}}(N,\omega)$ for that ω. A few lines of algebra confirm that

$$\langle s \rangle_N = -k_B \sum_\omega p_N(\omega) \ln p_N(\omega) = \frac{1}{T}\left(\langle U^{\mathrm{eff}} \rangle_N - F_N^{\mathrm{eff}} \right) \qquad (30)$$

in this parameterization.

In the large N limit, we expect that the intensive quantity ω should be an average of identically distributed independent random variables, and so the quantities $F_N^{\mathrm{eff}}, F_N, \langle U^{\mathrm{eff}} \rangle_N, \langle U \rangle_N$ appearing in the exponent of the probability distributions should all become proportional to N (cf. [74]). This implies that

$$\frac{F_{N+\Delta N}^{\mathrm{eff}} - F_N^{\mathrm{eff}}}{\Delta N} = \frac{F_N^{\mathrm{eff}}}{N} \qquad (31)$$

and likewise for the other quantities.

In this case, Equation (26) implies:

$$\delta\mu - \frac{(\langle U \rangle_N - F_N) - (\langle U^{\mathrm{eff}} \rangle_N - F_N^{\mathrm{eff}})}{N}$$
$$\geq \frac{k_B T}{\Delta N} \ln \frac{\pi(N \to N + \Delta N)}{\pi(N + \Delta N \to N)} \qquad (32)$$

where $\delta\mu \equiv \mu - F_N/N$ is the excess chemical potential beyond what is required to keep N stationary in the equilibrium ensemble. The left hand side has been rearranged so that the second term is the relative entropy between the actual distribution and the Boltzmann distribution (divided by N), which is always positive. Note that both averages are taken in the actual nonequilibrium distribution over microstates at fixed N.

As in the self-replication example, making predictions using this result requires computing the reversal probability $\pi(N + \Delta N \to N)$. The first thing we can say about this quantity is that it is less than $\pi(N \to N + \Delta N)$ if ΔN is positive and the structure is growing, which makes the right hand side of (32) greater than zero. The resulting inequality is one of the main conclusions of the paper by Nguyen and Vaikuntanthan [54], and extends an earlier result by the latter author on systems driven by externally varied parameters [75]. We have obtained it by a different route, which sheds light on the implications of the tightened Clausius inequality (26).

The top panel of Figure 3, generated with data from [54], shows the minimum allowed $\delta\mu$ as a function of the characteristic size ξ of spatial correlations in a 1-D filament assembled from two different monomer types. E^{eff} for a given ξ is found by inverting the expression for ξ as a function of binding energies at thermal equilibrium.
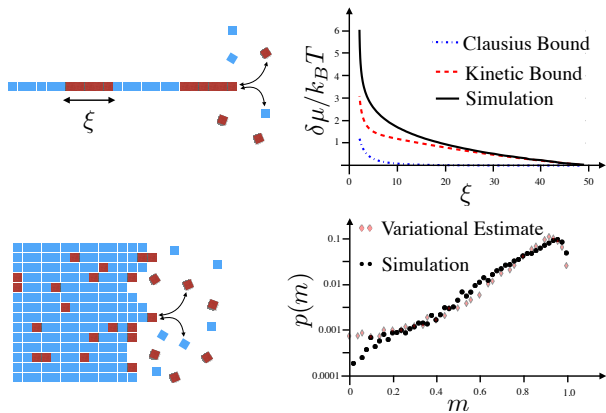
FIG. 3. Color online. Plots generated using simulation data originally reported in [54], with permission from the authors. Red and blue monomers are maintained at identical chemical potentials μ in a supersaturated solution, but like-colored pairs of monomers bind more strongly than unlike pairs. Top: Excess chemical potential required to reduce the typical size ξ of uniformly colored domains from its equilibrium value $\xi = 50$ in simulated assembly of a 1-dimensional filament. The black line is generated directly from simulation, while the blue line is obtained from Equation (32) by requiring that the left hand side be non-negative. The red line is the tighter bound of Equation (35), which includes kinetic information about the relative size of fluctuations in assembly speed. Bottom: Probability of observing a fraction m of blue monomers in a 2-dimensional filament that has been growing for a long time. The black dots are direct simulation data. The red diamonds are generated by sampling from an Ising distribution with coupling strength chosen to saturate the constraint (35).

Actual values of ξ obtained by growing the filament at finite $\delta\mu$ in simulation satisfy this weakened version of (32) that simply requires the left-hand side to be positive. But there is plenty of room to obtain a stronger bound that makes use of the right hand side.

The statistical irreversibility can be calculated explicitly in terms of a measurable drift velocity v and diffusion coefficient D if the dynamics of N are exactly described by a Gaussian drift-diffusion process. When inserted into the full expression for the irreversibility, this yields a simple ratio of v to D regardless of the exact value of ΔN:

$$\frac{1}{\Delta N} \ln \frac{\pi(N \to N + \Delta N)}{\pi(N + \Delta N \to N)} = \frac{v}{D}. \qquad (33)$$

Outside of this special scenario, we can still define a drift velocity and diffusion coefficient in the limit $\tau, \Delta N \to \infty$, where the central limit theorem guarantees that small fluctuations in ΔN become Gaussian. The statistics of these fluctuations about the mean value $\langle \Delta N \rangle = v\tau$ are fully determined by a diffusion coefficient $D = \frac{1}{2} \lim_{\tau \to \infty} \frac{\text{var}(\Delta N)}{\tau}$ where $\text{var}(\Delta N) = \langle (\Delta N)^2 \rangle - \langle \Delta N \rangle^2$ is the variance in ΔN over the observation time τ. But the return probability becomes exponentially small in this limit, and involves an extremely rare large fluc-

tuation that is not covered by the central limit theorem. As we will discuss at more length in Section IV, recent results in the theory of rare fluctuations in stochastic processes imply that the Gaussian extrapolation always underestimates the true irreversibility:

$$\frac{1}{\Delta N} \ln \frac{\pi(N \to N + \Delta N)}{\pi(N + \Delta N \to N)} \geq \frac{v}{D}. \qquad (34)$$

This allows us to use the observable ratio v/D on the right hand side of the tightened inequality of Equation (32), leading to the second main result in [54]:

$$\delta\mu - \frac{(\langle U \rangle_N - F_N) - (\langle U^{\text{eff}} \rangle_N - F_N^{\text{eff}})}{N} \geq \frac{v}{D}. \qquad (35)$$

Returning to the top panel of Figure 3, we see that this revised bound is significantly tighter. As in the self-replication example, the energetic cost of the process is mainly determined by the statistical irreversibility. As the forcing increases from zero, all the chemical work initially goes into breaking time-reversal symmetry and biasing the process towards positive growth. Only as $\delta\mu$ increases beyond $\sim 2k_BT$ does the internal entropy change begin to make a significant contribution.

Because the new bound can be saturated at finite distance from equilibrium, it can also be used to predict properties of the assembled structure without solving the dynamics. The bottom panel of Figure 3, also made with data from [54], displays a prediction of the distribution over the fraction $m = N_B/N$ of blue monomers in a 2-dimensional structure. This prediction was obtained by assuming that the nonequilibrium statistics of these fluctuations are still described by an equilibrium Ising model, but with a modified effective coupling energy. This effective coupling was chosen to make the distribution as different as possible from the actual equilibrium distribution while still satisfying the bound contained in Equation (35). The agreement between the prediction and the observed fluctuation spectrum is encouraging, and more work should be done to determine how far this procedure can be generalized. In principle, it could be a powerful way to obtain design principles for more complicated self-assembly scenarios that are challenging to simulate directly.

IV. DYNAMICAL PRECISION AND CONSTRAINTS ON RARE FLUCTUATIONS

The microscopic reversibility relation (9) relates entropy changes to statistical irreversibility. Results derived from this relation thus tend to be most useful when some information is available about the probabilities of trajectories that reverse the system's typical behavior. In the previous section we discussed the ideal case of perfectly Gaussian fluctuations, where the reversal probability is completely determined by drift and diffusion coefficients. But in general, computing the irreversibility requires evaluating the probabilities of exponentially

rare events, which do not fall under the central limit theorem, and may bear no relation to the small fluctuations actually observed in a finite-time experiment.

For this reason, the power of nonequilibrium results like those of Section III has been greatly augmented by advances in the mathematical theory of rare events in stochastic processes, known as Large Deviation Theory (see [74] for an accessible introduction to this field). In this section, we present the recently discovered bound underlying the constraint on compositional fluctuations of Equation (35) [29, 59]. This result also places constraints on dynamical design goals via a "thermodynamic uncertainty relation", which assigns an energetic cost to the suppression of fluctuations in the speed of a cyclic process [5]. We will end the section by pointing out a few applications in this area.

A. Bound on rare current fluctuations

We illustrate the bound with a simple model: the ring of N discrete states shown in the first panel of Figure 4. The system can hop from one state to either adjacent state in a Poisson process, whose rates for clockwise and counterclockwise hops w_+ and w_- are independent of the position along the ring. We can think of this as a highly simplified model of a biochemical clock, in which a protein cycles through a set of internal conformations or phosphorylation states [6].

Our goal is to constrain the statistical irreversibility of the typical dynamics based on small, observable fluctuations. Specifically, we can consider the probability $p_\tau(j)$ that the system completes $j\tau$ net cycles in the clockwise direction in time τ. In the limit of large τ, small fluctuations in j are described by a Gaussian distribution peaked at the typical value (cf. [77])

$$\bar{j} = \frac{1}{N}(w_+ - w_-) \tag{36}$$

with variance

$$\mathrm{var}(j) = \frac{1}{N\tau}(w_+ + w_-). \tag{37}$$

We can quantify the statistical irreversibility with an expression analogous to the right hand side of the tightened Second Law bound in Equation (32), by taking the logarithm of the ratio of the probabilities of observing the system's typical behavior and its time-reverse:

$$\sigma \equiv \lim_{\tau \to \infty} \frac{1}{\tau} \ln \frac{p_\tau(\bar{j})}{p_\tau(-\bar{j})}. \tag{38}$$

Depending on the physical interpretation of the model, σ may be equal to the rate of entropy production (in units of k_B). But for the moment, we are concerned only with the statistical properties of the model itself, abstracting from the thermodynamics, and have introduced this new symbol to reflect this distinction. In our simple ring
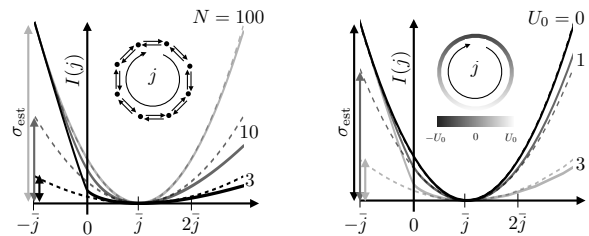

FIG. 4. Large deviation rate function $I(j) = -\lim \frac{1}{\tau} \ln p_\tau(j)$ for a ring of N discrete states, at fixed values of the typical current $\bar{j}$ and irreversibility σ. The solid lines are exact calculations, and dotted lines are Gaussian extrapolations $I_G(j)$ based on small fluctuations. Left: Uniform transition rates at three different N values with $\sigma, \bar{j}$ fixed. The extrapolated probability of reversing the mean current is significantly different from the true $I(-\bar{j})$ for $N = 3$, but improves as N increases, until the lines overlap at $N = 100$. These overlapping lines also represent the quadratic upper bound $I_B(j)$ for these $\sigma, \bar{j}$ values. Right: Three different sets of non-uniform transition rates at fixed $N, \sigma, \bar{j}$. The number of states is fixed at $N = 100$, and the log-ratio of rates $\ln w_+/w_-$ varies sinusoidally with amplitude U_0/N. This could be used as a model of driven diffusion through the sinusoidal potential energy landscape shaded on the ring in the inset, with amplitude U_0 in units of $k_B T$. As U_0 becomes significantly larger than 1, the Gaussian estimate again begins to deviate downward from the true value of the entropy production rate.

model, the exact value of σ can be found by examining the probabilities of the individual trajectories of $j\tau$ net positive steps, and noting that each one is related to the probability of its time-reversed version by the same factor $(w_+/w_-)^{Nj\tau}$. Inserting this into Equation (38) yields:

$$\sigma = (w_+ - w_-) \ln \frac{w_+}{w_-}. \tag{39}$$

If we instead estimate the irreversibility based on small fluctuations, replacing $p_\tau(j)$ with a Gaussian distribution of mean $\bar{j}$ and variance $\mathrm{var}(j)$, we find:

$$\sigma_{\mathrm{est}} = \frac{2}{\tau} \frac{\bar{j}^2}{\mathrm{var}(j)} = 2\frac{(w_+ - w_-)^2}{w_+ + w_-}. \tag{40}$$

One can easily verify that $\sigma_{\mathrm{est}} \leq \sigma$ in this case, with equality only when $(w_+ - w_-)/(w_+ + w_-) \ll 1$.

Barato and Seifert computed σ and σ_{est} in a diverse array of more complex models, and found that the true irreversibility was always bounded by this simple function of $\mathrm{var}(j)$ and $\bar{j}$ [5]. To prove this relation for a generic Markov process, Gingrich *et al.* made use of recently discovered properties of the large deviation rate function $I(j)$, which represents the scaled logarithm of $p_\tau(j)$ in the long time limit:

$$I(j) = -\lim_{\tau \to \infty} \frac{1}{\tau} \ln p_\tau(j). \tag{41}$$

This function is minimized at the most likely value of j, which is equal to the mean $\bar{j}$ in this limit. Since all the

probability concentrates at this point, we always have $I(\bar{j}) = 0$. (See [74] for a mathematically precise construction of $I(j)$, accounting for the fact that j on the right hand side must be an integer multiple of $1/\tau$, while j on the left can be any real number). For large enough values of τ, $I(j)$ plays a role analogous to free energy, and the distribution over j is given by

$$p_\tau(j) \propto e^{-\tau I(j)}. \tag{42}$$

The statistical irreversibility defined by Equation (38) can now be written as

$$\sigma = I(-\bar{j}). \tag{43}$$

And the estimate based on small fluctuations can be obtained from a quadratic expansion of $I(j)$ about $I(\bar{j})$:

$$I_G(j) \equiv \frac{1}{2} I''(\bar{j})(j - \bar{j})^2. \tag{44}$$

Inserting this expression into Equation (42) shows that the curvature of $I(j)$ controls the variance of the resulting Gaussian distribution via $\mathrm{var}(j) = 1/\tau I''(\bar{j})$. The estimated irreversibility from Equation (40) can now be written in a parallel form to Equation (43):

$$\sigma_{\mathrm{est}} = I_G(-\bar{j}). \tag{45}$$

We have plotted $I(j)$ and $I_G(j)$ in Figure 4 for various parameter values sharing the same σ and $\bar{j}$ in two different models. The first is the uniform ring model presented above, and the second is a ring with non-uniform rates.

Gingrich *et al.* showed that $I(j)$ is bounded from above by a different quadratic function which we will call $I_B(j)$ [29]. $I_B(j)$ is uniquely defined by requiring that $I_B(\bar{j}) = I'_B(\bar{j}) = 0$ and $I_B(-\bar{j}) = \sigma$, with $\bar{j}$ and σ taken from the original model. This bound was simultaneously discovered by Pietzonka, Barato and Seifert on the basis of numerical evidence [59]. In an infinitesimal neighborhood around $j = \bar{j}$, we have $I_G(j) = I(j)$, and so $I_B(j)$ is also an upper bound on the approximate rate function $I_G(j)$ in this local region. But since $I_G(j)$ and $I_B(j)$ are quadratic functions tangent to each other at their extremal point $j = \bar{j}$, it is impossible for them to intersect, and the local bound implies a global bound on $I_G(j)$:

$$I_G(j) \leq I_B(j). \tag{46}$$

Figure 4 illustrates this result for the current around a single ring of states. Both panels include one parameter set for which the true $I(j)$ is nearly quadratic, so that $I_B(j) \approx I_G(j) \approx I(j)$ ($N = 100$ for left panel, $U_0 = 0$ for left panel). $I_B(j)$ is the same for all parameter sets within each panel, since it is fully determined by σ and $\bar{j}$, and so this same curve can be used to compare $I_B(j)$ to $I(j)$ and $I_G(j)$ at different parameter values, revealing that $I_G(j)$ is always smaller than $I_B(j)$. Since $I_B(-\bar{j}) = \sigma$,

this implies that $\sigma_{\mathrm{est}} = I_G(-\bar{j})$ always underestimates the true irreversibility:

$$\sigma_{\mathrm{est}} \leq \sigma. \tag{47}$$

These bounds continue to apply when the allowed transitions between states are not arranged on a single ring, but connected in an arbitrary topology containing any number of interlocking cycles [29]. This makes the description of the dynamics more complicated, because the mean numbers of jumps per unit time for different transitions no longer need to be the equal. If $j_{xy}\tau$ is the net number of jumps from state x to state y observed in time τ, then we can define a coarse-grained current j_d by taking a linear combination of all these microscopic currents:

$$j_d \equiv \sum_{y>x} d_{xy} j_{xy} \tag{48}$$

where d_{xy} are arbitrary constant coefficients. The statistical irreversibility σ is now obtained as $\sigma = I(\{-\bar{j}_{xy}\})$ from the joint rate function $I(\{j_{xy}\})$ that depends on all the microscopic currents, and $I_B(j_d)$ is defined by $I_B(\bar{j}_d) = I'_B(\bar{j}_d) = 0$ and $I_B(-\bar{j}_d) = \sigma$. The Gaussian estimate is defined in the same way as in the single ring, using the mean and variance of j_d:

$$\sigma_{\mathrm{est}} \equiv I_G(-\bar{j}_d) = \frac{2}{\tau} \frac{\bar{j}_d^2}{\mathrm{var}(j_d)}. \tag{49}$$

The rate function bound in Equation (46) and the resulting bound on σ in Equation (47) are purely mathematical results, derived from the properties of Markov processes prior to any thermodynamic interpretation. They provides new information that supplements the thermodynamic reasoning of Section III, converting formal expressions involving astronomically rare events into meaningful bounds on measurable quantities.

B. Performance bounds in clocks, sensors and motors

In Section III C we saw how recent advances in nonequilibrium thermodynamics can tighten the Second Law bound on the cost of reducing the internal entropy of a thermodynamic system. The bound on rare current fluctuations contained in Equation (46) has spawned a new family of results with no parallel in thermal equilibrium, placing constraints on *dynamical* performance goals. In these applications, the Gaussian estimate of irreversibility σ_{est} defined by Equation (49) is itself the relevant figure of merit. Larger values of σ_{est} are associated with smaller relative fluctuations and more "precise" dynamics. If the Markov model under consideration adequately accounts for all the dissipative processes in the system, then microscopic reversibility (9) implies that the true statistical irreversibility σ is equal to the

steady-state rate of entropy change in the environment:

$$\sigma = \frac{1}{k_B} \langle \dot{S}_e \rangle. \tag{50}$$

Equation (47) relating σ and σ_{est} then specifies the minimum dissipative cost of achieving the desired precision, and is known as a "Thermodynamic Uncertainty Relation," since its constraint on maximum precision is reminiscent of the Heisenberg Uncertainty Principle of quantum mechanics. In this subsection, we illustrate the consequences of this relation with some recent results on the precision of clocks and the efficiency of molecular motors.

Molecular motors like kinesin or the F1-ATPase convert chemical energy into mechanical work on small length scales where thermal fluctuations are important. The efficiency η of a motor protein doing work against an applied force f is defined as the ratio between the output work rate fv and the mean input chemical work rate $\langle \dot{\mathcal{W}} \rangle$ in the steady state:

$$\eta = \frac{fv}{\langle \dot{\mathcal{W}} \rangle} \tag{51}$$

Here v is the average speed of the motor traveling (or rotating) opposite to the direction of the force. The chemical work that is not transduced into mechanical work is released into the environment as heat, so we find:

$$\langle \dot{S}_e \rangle = \frac{1}{T} \left(\langle \dot{\mathcal{W}} \rangle - fv \right) \tag{52}$$

$$= \frac{fv}{T} \left(\frac{1}{\eta} - 1 \right). \tag{53}$$

By inserting this into the bound of Equation (47) and writing σ_{est} in terms of the mean speed v and the diffusion coefficient D of the motor motion, Pietzonka *et al.* showed that the efficiency is constrained by [58]:

$$\eta \leq \frac{1}{1 + vk_BT/Df}. \tag{54}$$

Making the dynamics more deterministic by increasing the ratio v/D thus incurs an extra energy cost, reducing the motor's efficiency.

Fluctuations are also a relevant design consideration in the construction of a clock, whose purpose is to provide a stable reference time for other processes. Any stochastic process that includes at least one cycle can be thought of as a clock, whose readout is the net number of forward jumps $j\tau$ across one of the links in the cycle. The bound on σ_{est} in Equation (47) then gives the minimum thermodynamic cost of a given precision goal, as discussed in [6], where the relative error ϵ is defined by

$$\epsilon^2 \equiv \frac{\text{var}(j)}{\bar{j}^2} = \frac{2}{\sigma_{\text{est}}\tau}. \tag{55}$$

The minimum entropy change in the environment $\langle \dot{S}_e \rangle \tau = \sigma \tau$ corresponding to a 1% relative error, for

instance, is $20,000 k_B$. This bound is independent of τ, so the *rate* of entropy change $\langle \dot{S}_e \rangle$ depends on the timescale at which the precision is demanded. Since we are concerned with *relative* error, higher precision is easier to obtain for longer time intervals. For instance, a 1% error in the measurement of an hour using the stochastic clock only demands $20,000 k_B/\text{hour}$, but the relative error in the measurement of each minute within that hour will be larger by a factor of $\sqrt{60}$. Achieving 1% precision in the measurement of each minute would demand the much higher entropy production rate of $20,000 k_B/\text{minute}$. As illustrated in Figure 4, the maximum precision at fixed $\langle \dot{S}_e \rangle$ and $\bar{j}$ is obtained in the near-equilibrium regime of large N, where the entropy change per step $k_B \ln w_+/w_- = \sigma/N\bar{j}$ becomes small.

This precision/dissipation trade-off had been investigated empirically before the bound of Equation (47) was discovered, in specific kinetic models of biological and artificial chemical clocks [16]. The bound on σ_{est} puts these findings in a broader context, although it remains an open challenge to rigorously apply it to chemical concentration oscillations in an informative way. We can get an initial sense of the implications of the bound for these chemical clocks by replacing the full chemical reaction network with a phenomenological model like those of Figure 4, where the system simply hops along a one-dimensional ring of N states. In their original investigations, Cao *et al.* expressed the clock precision in terms of fluctuations in the cycle period, whose average over many consecutive cycles can be related to the time-averaged current j through a monitored link in our simple ring model as $\mathcal{T} = 1/j$ [16]. An ensemble of initially synchronized clocks will drift apart due to thermal fluctuations, and each member will complete its cycle in a different period of time. The variance of the distribution over completion times will increase with each successive cycle, and the central limit theorem guarantees that this growth eventually becomes linear in the number of completed cycles $N_c = \bar{j}\tau$. This allows us to define a "diffusion coefficient" D by

$$\text{var}(N_c\mathcal{T}) = DN_c\bar{\mathcal{T}} \tag{56}$$

where $\bar{\mathcal{T}} = 1/\bar{j}$ is the typical period. The dimensionless quantity $D/\bar{\mathcal{T}}$ measures the relative drift per cycle, and Cao *et al.* found an empirical trade-off between this measure of the drift and the entropy production per cycle in the four models they studied [16].

We can place a thermodynamic bound on $D/\bar{\mathcal{T}}$ by noting that the relationship between $\mathcal{T}$ and j for small fluctuations in the $\tau \to \infty$ limit is adequately described by the linear approximation

$$\mathcal{T} \approx \frac{1}{\bar{j}} \left(1 - \frac{j - \bar{j}}{\bar{j}} \right). \tag{57}$$

From this expansion, it follows that $\text{var}(j)/\bar{j}^2 = \text{var}(\mathcal{T})/\bar{\mathcal{T}}^2$, and so Equations (45), (50) and (47) imply

that

$$\frac{D}{\bar{\bar{\mathcal{T}}}} \geq \frac{2k_B}{\Delta S_c} \tag{58}$$

where $\Delta S_c \equiv \langle \dot{S}_e \rangle \tau / N_c$ is the steady-state entropy production per clock cycle. The findings of Cao *et al.* are consistent with this result, but they cannot saturate the inequality at large ΔS_c because the number of states per cycle N is a fixed feature of each model. As illustrated in Figure 4, when the entropy change per reaction becomes larger than k_B, the Gaussian extrapolation begins to significantly underestimate the true irreversibility, which is here equal to the entropy production. At fixed N, it can be shown that $D/\bar{\mathcal{T}} \geq 1/N$ even in the limit $\Delta S_c \to \infty$ [22]. This opens up an interesting question as to the factors that constrain the number of states per cycle. Larger values of N lead to improved precision at fixed ΔS_c, but other costs must eventually come into play, perhaps involving the absolute cycle speed or robustness against externally imposed perturbations. Including these additional factors in the analysis will allow for a tighter estimate of the maximum feasible efficiency.

V. THERMODYNAMIC POTENTIALS

In the previous two sections, we have seen how the Second Law can be tightened to provide informative constraints on far-from-equilibrium processes. We now examine how far the traditional machinery of thermodynamics can be generalized to new kinds of states. In particular, we would like to look at those nonequilibrium states that most resemble equilibrium, with macroscopic properties that are independent of initial conditions and do not change in time. These include familiar situations like steady heat conduction, steady shear flow, and steady chemical flux through a chemostatted reaction network, as well as stationary states of novel materials like the contractile actin cytoskeleton of a living cell or a solution of self-propelled Brownian particles.

In the absence of an externally maintained driving force, the steady state is characterized by a thermodynamic potential, such as the entropy S, the Gibbs free energy G or the Helmholtz free energy F. This thermodynamic potential plays three distinct roles. First of all, the force required to modify a control parameter (such as the volume of a cylinder of gas) is given by a partial derivative of this function. Second, the values of quantities that are left free to fluctuate (like the pressure of a gas at fixed volume) are found by extremizing the function. Third, the thermodynamic potential obeys a Clausius relation, which defines the minimum amount of work or heat that must be exchanged with the system in order to accomplish a transition from one steady state to another.

Much effort in nonequilibrium statistical mechanics over the past 75 years has been devoted to the search for generalizations of these potentials for nonequilibrium steady states (cf. [45, pp. 51-54]). It is now clear that there is no function that simultaneously plays all three roles in a generic nonequilibrium system. But considerable progress has been made in formalizing a thermodynamic approach to each of these aspects individually. In the regimes where these approaches provide concrete predictions, most of their content was already anticipated in classical linear response theory. The new results can be seen as a way of systematizing and integrating those older findings, facilitating their application to more complex systems and possibly extending their range of validity.

A common thread running through all these efforts is the role of "excess" heat and "excess" work. Since a nonequilibrium steady state requires a continual supply of work and is constantly dissipating heat into its environment, the work and heat associated with a quasi-static transition or with relaxation to the steady state become infinite. To make use of these key thermodynamic quantities in the definition of a generalized thermodynamic potential, one must subtract off the "housekeeping" heat flow of the steady state, preserving only the excess portion associated with the given fluctuation or transition.

In this section, we review the use of excess quantities in the calculation of statistical forces, the construction of variational principles, and the analysis of transitions between steady states.

A. Statistical Forces

Maes and coworkers have illustrated the role of excess work in near-equilibrium thermodynamics by investigating the thermodynamic forces f_λ conjugate to control parameters λ in isothermal steady states [7, 8]. If we consider the isothermal compression of a nonequilibrium material, for instance, such as an active matter suspension or a sheared fluid, the control parameter λ will be the volume V of the container, which we change by adjusting the position of one of the walls. The conjugate force f is the pressure P exerted by the material on that wall. In thermal equilibrium, the pressure is simply minus the derivative of the free energy: $P = -dF/dV$. The activity of the particles or the shear flow will modify this pressure, and we would like to know which aspects of this nonequilibrium activity determine the pressure change.

As a concrete example, consider the Brownian particle depicted in Figure 5, which diffuses in a one-dimensional ring with potential energy landscape $U(x) = U_0 \cos(x/2\pi - \phi_0)$ and periodic boundary conditions at $x = 0$ and $x = 1$. Our control parameter λ will be the phase ϕ_0 of the landscape. Because of the symmetry of the system, the free energy F is independent of ϕ_0, and so the conjugate force $f_\phi^{\text{eq}} = -dF/d\phi_0$ of equilibrium thermodynamics vanishes. Figure 5 shows what happens when a non-conservative driving force f_{nc} is turned on, which pushes the particle around the ring counterclockwise. The agent in control of ϕ_0 now feels a nonzero average force f_ϕ trying to push ϕ_0 in the direction of f_{nc}.
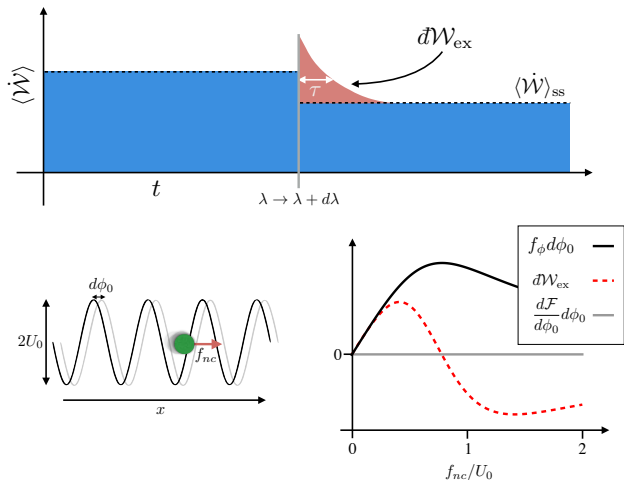
FIG. 5. Color online. Top: Definition of excess work. When the control parameter λ is suddenly changed by a small amount $d\lambda$, the average work rate $\langle\dot{\mathcal{W}}\rangle$ due to the steady driving forces is modified. The system relaxes to a new steady state on a time scale τ. The excess work $đ\mathcal{W}_{\mathrm{ex}}$ associated with this transition is the difference between the actual work done during this relaxation and the work that would have been done if the system had started in the steady state. Left: Schematic of driven diffusion with periodic boundary conditions and a sinusoidal potential energy landscape of amplitude U_0. The phase ϕ_0 of the periodic landscape can be used as a control parameter. Right: The work $-f_\phi d\phi_0$ required to quasi-statically change ϕ_0 is equal to $-đ\mathcal{W}_{\mathrm{ex}}$ for small driving force f_{nc}. This guarantees that their difference is equal to $d\mathcal{F} = 0$, as required by Equation (59). For larger driving, these two quantities diverge from each other, and Equation (59) is no longer satisfied.

Maes *et al.* pointed out that the first-order nonequilibrium correction to the statistical force is determined by the excess work associated with a change in the corresponding parameter. As illustrated in Figure 5, the excess work $đ\mathcal{W}_{\mathrm{ex}}$ is the transient part of the work done by the steady nonequilibrium driving forces after a sudden change of parameters from λ to $\lambda + d\lambda$. To first order in the strength of the driving force, microscopic reversibility (9) requires that infinitesimal isothermal transitions satisfy an extended Clausius relation [8, 41]:

$$d\mathcal{F} = -f_\lambda d\lambda + đ\mathcal{W}_{\mathrm{ex}} \tag{59}$$

where $d\mathcal{F}$ is the change in the nonequilibrium free energy defined by Equation (23): $\mathcal{F} = U - T\langle s\rangle$. To linear order in the nonequilibrium drive strength, $d\mathcal{F}$ remains equal to the change in equilibrium free energy dF [8]. The right hand side of the equation is simply the total remaining work after subtracting the constant background of the steady-state work rate. Equation (59) thus says that quasi-static transitions between near-equilibrium steady states follow the Clausius relation $\Delta F = \mathcal{W}$ after this

background subtraction.

As illustrated in Figure 5, however, the fact that $đ\mathcal{W}_{\mathrm{ex}}$ is not an exact differential makes some processes possible that would be forbidden in classical thermodynamics. When $f_{nc} > 0$, the force f_ϕ conjugate to ϕ_0 is positive all the way around the ring, and cannot be written as the gradient of any potential function. More generally, if multiple control parameters are allowed to vary, the integral of $f_\lambda \cdot d\lambda$ around a closed loop in control parameter space can be nonzero. This means that a cyclic process is possible that extracts work from the system, which would be a violation of the Second Law in classical thermodynamics. A nonequilibrium steady state relies on a constant supply of work even when the control parameters are not changing, and the right kind of change of parameters can extract some of this work.

A more surprising consequence of Equation (59) for thermodynamic reasoning is that the statistical force depends on the *kinetics* of the material. This stands in stark contrast to thermal equilibrium, where knowledge of microstate energies is sufficient to compute forces as derivatives of free energy. The source of this kinetic dependence is illustrated in the top panel of Figure 5, which shows how the excess work $đ\mathcal{W}_{\mathrm{ex}}$ done by the constant background driving forces depends on the time τ required to return to the steady state. This kinetic contribution enables boundary terms to significantly affect the pressure of macroscopic active materials, since the kinetics of relaxation after displacement of a boundary wall depend on the details of the interaction with that wall. Solon *et al.* have recently demonstrated this for a model of active Brownian particles, where the mean force exerted by the suspension against a given wall generically depends on the stiffness of that wall, in a way that is independent of system size [72]. This implies that the pressure generally fails to be fully determined by other intensive parameters like concentration and temperature, but remains sensitive to the properties of the container even in the thermodynamic limit.

As the strength of the driving force increases, the first-order approximation in Equation (59) eventually breaks down. Figure 5 shows how the excess work differential $đ\mathcal{W}_{\mathrm{ex}}$ remains well-defined for all values of the driving force, but need not bear any obvious relationship to f_λ or $d\mathcal{F}/d\lambda$. It may still be a useful quantity for characterizing transitions between nonequilibrium steady states, however. Its behavior as a function of f_{nc}, for example, is very rich near the critical force $f_{nc} = U_0$, and may be related to other phenomena like giant diffusion and increased effective temperature that have been observed in this regime [32, 64].

B. Variational Principles

Thermodynamic potentials also provide variational principles for predicting the properties of systems at thermal equilibrium. Returning to the example of isothermal

compression, we recall that the equilibrium probability $p_{eq}(P)$ of a given fluctuation away from the deterministic pressure P^* is determined by the free energy (cf. [24]):

$$p_{eq}(P) \propto e^{-F(P)/k_B T}. \qquad (60)$$

This relationship is essentially a coarse-grained version of the Boltzmann distribution. As the system size increases, the relative size of fluctuations decreases, and all the probability concentrates at the value $P = P^*$ that minimizes the free energy F.

In 1959, James McLennan showed that the correction to this distribution for near-equilibrium steady states is related in a simple way to the work done by the external driving forces [46, 51]. The coarse-grained version of his finding gives:

$$p_{ss}(P) \propto e^{-[F(P)-W_{ex}(P)]/k_B T} \qquad (61)$$

where $W_{ex}(P)$ is the extra work done on the way to a given fluctuation in P, over and above the steady-state work rate, as illustrated in Figure 6. In the linear-response limit considered by McLennan, the extra work on the way to a fluctuation is the same as the extra work during the relaxation of the fluctuation, and so the W_{ex} appearing here is essentially the same quantity discussed in the context of statistical forces above. Wherever Equation (61) is valid, steady-state values of observables can be computed by minimizing $F - W_{ex}$. By applying this variational principle, one can obtain the Green-Kubo linear response formula [39, 40, 46] and other standard results of linear response theory.

The microscopic reversibility relation (9) provides a foundation for deriving exact generalizations of Equation (61) for steady states arbitrarily far from equilibrium [9, 18, 19, 39, 40, 46]. These expressions involve quantities that are extremely challenging in general to measure or even estimate, and do not directly supply any new predictive power. But they provide a framework for systematically constructing perturbative expansions around the regime described by Equation (61), shedding new light on the physical factors that set the limits of linear response theory.

Following this line of reasoning, we have recently shown that the "near-equilibrium" variational principle contained in Equation (61) can accurately describe driven steady states far outside the realm of classical linear response theory [49, 50]. Our results complement those of Bunin *et al.*, who showed that linear-response type results can remain valid under strong driving as long as the third cumulant of the work distribution remains sufficiently small [13, 14]. Consider for example the sheared colloidal suspension illustrated in Figure 6. The microstate of the system is defined by the positions of a set of interacting Brownian particles diffusing in a viscous solvent of volume V. The spatial configuration of the particles equilibrates on a timescale τ_0 set by the diffusion coefficient, the number of particles, and the interparticle potential. A nonequilibrium steady state is
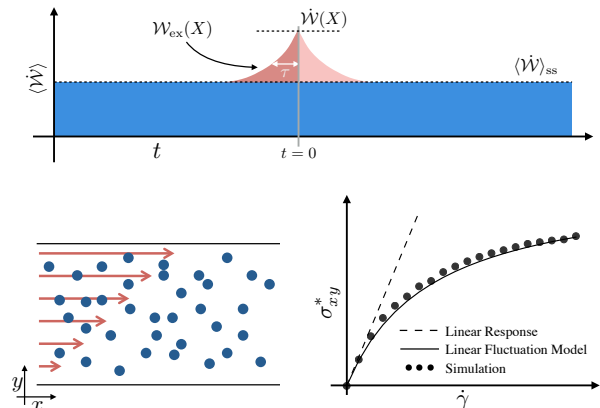

FIG. 6. Color online. Top: Excess work associated with steady-state fluctuations. The mean steady-state work rate $\langle \dot{\mathcal{W}} \rangle_{ss}$ is constant in time, but conditioning on a given fluctuation in an observable X at time $t = 0$ makes the average deviate from $\langle \dot{\mathcal{W}} \rangle_{ss}$. The excess work $\mathcal{W}_{ex}$ is the part of this deviation leading up to $t = 0$. Left: A nonequilibrium steady state can be set up in a colloidal suspension by imposing a fixed flow gradient $v_x = \dot{\gamma}y$ in the background solvent. Blue circles are Brownian particles interacting via a repulsive potential, and red arrows represent the imposed flow field. Right: The steady-state shear stress can be accurately predicted by minimizing $F - \mathcal{W}_{ex}$ using the simple phenomenological model of stress fluctuations contained in Equations (62) and (65).

created by moving the top of the container at a constant speed to set up a steady shear flow, where the solvent flow velocity in the x direction is proportional to the y position: $v_x = \dot{\gamma}y$. The constant of proportionality $\dot{\gamma}$ is called the shear rate, and is a measure of the nonequilibrium driving strength in this system. The instantaneous drag force per unit area σ_{xy} resisting this movement is a function of the particle configuration, since the mutual repulsion of the particles can either assist or oppose the flow gradient depending on their relative positions.

We showed that the variational principle of Equation (61) correctly predicts the typical value of σ_{xy} whenever its fluctuations are well described by a linear overdamped Langevin equation:

$$\frac{d}{dt}\sigma_{xy} = -\frac{1}{\tau}(\sigma_{xy} - \langle\sigma_{xy}\rangle_{ss}) + \sqrt{2D}\xi(t) \qquad (62)$$

where the Gaussian noise term is defined by $\langle \xi \rangle = 0$, $\langle \xi(t)\xi(t') \rangle = \delta(t - t')$, and τ, D are independent of σ_{xy}. The excess work in this model is easily computed in terms of σ_{xy}, $\dot{\gamma}$ and the relaxation time τ:

$$\mathcal{W}_{ex}(\sigma_{xy}) = V\tau\dot{\gamma}(\sigma_{xy} - \langle\sigma_{xy}\rangle_{ss}). \qquad (63)$$

Minimizing $F - \mathcal{W}_{ex}$ generates a prediction for the typical

drag force σ_{xy}^*:

$$\sigma_{xy}^* = \sigma_{xy}^0 + \frac{1}{k_B T}\dot{\gamma}\tau V \langle \sigma_{xy}^2 \rangle_{\text{eq}}, \qquad (64)$$

where σ_{xy}^0 is the drag force in the absence of inter-particle repulsion, and is independent of the particle configuration.

Near equilibrium, when the relaxation time τ is still equal to its equilibrium value τ_0, Equation (64) reproduces the Green-Kubo prediction of classical linear response theory. But at larger values of $\dot{\gamma}$, we expect the flow gradient to help the particle configuration relax more quickly. Eventually this convective relaxation becomes the dominant mechanism of particle rearrangement, and the relaxation timescale is roughly the time $1/\dot{\gamma}$ required for the flow gradient to push two adjacent particles past each other. We can capture both the $\dot{\gamma} \to 0$ and $\dot{\gamma} \to \infty$ limits by writing

$$\tau = \frac{\tau_0}{1 + k\dot{\gamma}\tau_0} \qquad (65)$$

where k is a constant fitting parameter. Figure 6 compares the σ_{xy}^* values directly obtained from a numerical simulation with the variational prediction of Equations (64) and (65). The simulation method and results are reported in more detail in [49, 50]. The equilibrium relaxation time τ_0 was obtained from the autocorrelation function of σ_{xy} at thermal equilibrium, and $k = 2$ was chosen to give the best fit to the rest of the curve. Also plotted is the linear response prediction using $\tau = \tau_0$, which only captures the initial slope of the σ_{xy} vs. $\dot{\gamma}$ curve.

We have generalized this procedure to encompass a broad class of systems and possible observables, as described in detail in [49]. The excess work $\mathcal{W}_{\text{ex}}$ provides the appropriate nonequilibrium correction to the equilibrium variational principle, as long as nonlinear corrections to the linearized fluctuation dynamics change $\mathcal{W}_{\text{ex}}$ by much less than $k_B T$ in typical fluctuations. But we also showed that the regime where Equation (61) applies is precisely the regime where the nonequilibrium driving force merely adds a linear tilt to the effective free energy. $\mathcal{W}_{\text{ex}}(X)$ is a linear function of the observable X, and if $F(X)$ is quadratic, then subtracting $\mathcal{W}_{\text{ex}}$ shifts the location of the global minimum without changing the shape of the function. This suggests that thermodynamically meaningful variational principles may not apply to pattern-formation scenarios or other cases where new phase transitions appear that did not exist in the equilibrium. In this sense, the extended Green-Kubo relation of Equation (64) is still a "near-equilibrium" result, since it only applies when the driving force incrementally modifies the equilibrium state, and fails when qualitatively new phenomena are generated.

C. Clausius Inequality

In Section V A, we saw that quasi-static processes near equilibrium satisfy the generalized form of the Clausius relation expressed in Equation (59). For transitions between equilibrium states, an even stronger statement can be made using the Clausius inequality obtained in Equation (22):

$$\mathcal{W} \geq \Delta\mathcal{F}, \qquad (66)$$

where $\mathcal{F} = U - T\langle s \rangle$ becomes equal to the equilibrium free energy F in the absence of driving forces. Our derivation of Equation (66) from the fluctuation theorem in Section III remains valid arbitrarily far from equilibrium, but the relation is not very informative for transitions between nonequilibrium steady states. Work is constantly being done even in the steady state to keep the system out of equilibrium, so $\mathcal{W} \to \infty$ in the quasi-static limit and the inequality is trivially satisfied.

A more interesting quantity is the finite "applied" work

$$\mathcal{W}_{\text{app}} = \int_0^\tau \left\langle \frac{\partial U}{\partial \lambda} \right\rangle_t \dot{\lambda}\, dt \qquad (67)$$

done by the external agent who manipulates the control parameters λ. The average is over the probability distribution at time t, obtained by evolving the initial distribution under the corresponding Fokker-Planck or Master equation. When the parameters are changed sufficiently slowly, the distribution is always very close to the steady state, and $\mathcal{W}_{\text{app}}$ is simply minus the integral of the steady-state force f_λ discussed in Section V A. A naïve observer ignorant of the internal forces or gradients sustaining the nonequilibrium steady states would believe that $\mathcal{W}_{\text{app}}$ is simply equal to the total work, and would expect it to satisfy Equation (66).

As we saw in Section V A, this expectation can already fail in the linear response regime. In the example of Figure 5, $d\mathcal{F} = 0$, but $d\mathcal{W}_{\text{app}} = -f_\phi d\phi_0$ is negative. To gain a deeper understanding of why the equilibrium predictions fail and how they should be corrected, it would be useful to have a modified version of the Clausius inequality for $\mathcal{W}_{\text{app}}$, in which equality can be achieved for quasi-static transitions. Hatano and Sasa showed in 2001 that this relation exists for systems described by overdamped Langevin dynamics [31], and their findings were later extended to Poisson processes [27] and fluctuating hydrodynamics [9]. It is now clear that a whole family of inequalities can be obtained using arguments very similar to our derivation of the original Clausius inequality in Section III.

The key insight behind these results lies in the fact that the trajectory probabilities $\mathcal{P}^*[\hat{\mathbf{x}}_0^\tau]$ of the reverse process are integrated out when moving from the microscopic reversibility relation (9) to the integral fluctuation theorem (16) [27, 71]. This means that integral fluctuation theorems for other quantities can be obtained by replacing the reverse trajectory probability with another properly

normalized distribution over trajectories $\mathcal{P}^{\dagger}[\hat{\mathbf{x}}_0^{\tau}]$. In this way, an integral fluctuation theorem $\langle e^{-\mathcal{A}} \rangle = 1$ can be obtained for any trajectory functional $\mathcal{A}[\mathbf{x}_0^{\tau}]$ that can be written in the form

$$\mathcal{A}[\mathbf{x}_0^{\tau}] = -\ln \frac{\mathcal{P}^{\dagger}[\hat{\mathbf{x}}_0^{\tau}]}{\mathcal{P}[\mathbf{x}_0^{\tau}]} \qquad (68)$$

for some choice of $\mathcal{P}^{\dagger}[\hat{\mathbf{x}}_0^{\tau}]$. As noted in Section III, the inequality $e^x \geq x + 1$ then implies that $\langle \mathcal{A} \rangle \geq 0$,

Hatano and Sasa identified a quantity that can be written in this form, reduces to the entropy change when the driving forces are turned off, and vanishes in a nonequilibrium steady state. This quantity is now known as the "non-adiabatic" entropy change ΔS_{na} [27]. It is closely related to two other quantities known in the literature as the "excess heat" [27, 31] and "excess work" [9]. These reduce to the heat and work, respectively, for systems whose steady state obeys detailed balance, and they exactly satisfy modified Clausius relations. But they do not correspond in general to any measurable physical property [9, 68]. For this reason, we reserve the symbol $\mathcal{W}_{\mathrm{ex}}$ in this review for the physically meaningful excess work found by subtracting the (measurable) average steady state work rate from the actual work done by the continually acting nonequilibrium driving forces. But the fluctuation theorem for ΔS_{na} still helps us understand how nonequilibrium steady states differ from equilibrium states. It turns out that ΔS_{na} can be written in terms of derivatives of the trajectory-level entropy $s = -k_B \ln p_{\mathrm{ss}}$. For systems described by overdamped Langevin dynamics (see Section II C), this allows us to rewrite $\langle \Delta S_{\mathrm{na}} \rangle \geq 0$ as

$$\mathcal{W}_{\mathrm{app}} \geq \Delta \mathcal{F} - \int_0^{\tau} \langle \nabla(U - Ts) \cdot \dot{\mathbf{x}} \rangle_t \, dt. \qquad (69)$$

This reduces to the ordinary Clausius inequality for undriven systems where $p_{\mathrm{ss}} \propto \exp(-U/k_B T)$, since then $T\nabla s = \nabla U$ and $\mathcal{F} = F$. But when the nonequilibrium steady-state distribution becomes different from the equilibrium distribution, new possibilities open up. It remains the case that equality is achieved only in the limit of quasi-static variation of λ. But unlike the work-free energy theorem of classical thermodynamics, this result contains path-dependent quantities on both sides of the inequality. Thus the quasi-static work no longer defines a state function, and there is no longer any guarantee that the slow protocol extracts the maximum amount of work for a given path in control-parameter space.

VI. DISCUSSION

The Second Law of Thermodynamics requires that the total entropy of the system plus environment can never decrease: so if the entropy of the system decreases, this must be compensated by the flow of heat into the environment. But in Section III we saw that the microscopic

reversibility relation also attaches an entropic cost to statistical irreversibility, leading to much higher levels of entropy production than would have been demanded by the Second Law alone. When the entropic requirements of statistical irreversibility are accounted for, the remaining dissipation can provide a remarkably tight bound on the allowed internal entropy. In the example of compositional fluctuations in two-dimensional self-assembly, the bound was tight enough to provide an accurate estimate of the distribution of fluctuations without any other knowledge of the system dynamics.

For many systems of interest – especially those with many degrees of freedom – the statistical irreversibility is difficult to compute or even estimate, since it involves the probability of extremely rare reverse trajectories. The self-assembly calculations of Section III relied on a recent result from Large Deviation Theory that relates the irreversibility to a measure of the size of typical fluctuations, which can be readily ascertained in an experiment or simulation. In Section IV, we described how this result guarantees that a naïve extrapolation of Gaussian fluctuations in a time-averaged probability current always underestimates the true irreversibility. We also discussed the new constraints it imposes on dynamical design goals, where the size of typical current fluctuations is itself the relevant figure of merit.

In a nonequilibrium steady state, thermodynamic potentials no longer provide shortcuts for calculating statistical forces, predicting typical values of fluctuating observables, or constraining the minimum amount of heat or work required to bring the system from one state to another. Efforts to identify nonequilibrium thermodynamic potentials have shed considerable new light on the reasons why these shortcuts fail, and on the special features of thermal equilibrium that make them work. We saw in Section V that empirically meaningful generalizations can be obtained close to equilibrium, where the thermodynamic quantities are modified by adding or subtracting the "excess" heat or work associated with a transition or fluctuation. This provides a powerful unified perspective in which to justify and organize the results of classical linear response theory, facilitating their application to new situations and extending their range of validity.

In summary, it appears that nonequilibrium thermodynamics is finally approaching full maturity. While some of the early enthusiasm for universal principles has been disappointed, rigorous investigation has revealed an even richer structure whose outlines are now becoming clear. This theoretical framework is now in a position to provide guidance for specific applications in biotechnology and materials science, and feedback from real design goals in these areas will be crucial for deepening and solidifying our understanding.

ACKNOWLEDGMENTS

---

[1] B. Andresen. Current trends in finite-time thermodynamics. *Angewandte Chemie International Edition*, 50(12):2690, 2011.

[2] D. Andrieux and P. Gaspard. Nonequilibrium generation of information in copolymerization processes. *Proc. Natl. Acad. Sci. U. S. A.*, 105:9516, 2008.

[3] D. Andrieux and P. Gaspard. Molecular information processing in nonequilibrium copolymerizations. *J. Chem. Phys.*, 130:14901, 2009.

[4] M. Baiesi and C. Maes. An update on the nonequilibrium linear response. *New J. Phys.*, 15:013004, 2013.

[5] A. C. Barato and U. Seifert. Thermodynamic uncertainty relation for biomolecular processes. *Phys. Rev. Lett.*, 114:158101, 2015.

[6] A. C. Barato and U. Seifert. Cost and precision of brownian clocks. *Phys. Rev. X*, 6:041053, 2016.

[7] U. Basu, C. Maes, and K. Netočný. How statistical forces depend on the thermodynamics and kinetics of driven media. *Phys. Rev. Lett.*, 114:250601, 2015.

[8] U. Basu, C. Maes, and K. Netočný. Statistical forces from close-to-equilibrium media. *New J. Phys.*, 17:115006, 2015.

[9] L. Bertini, A. De Sole, D. Gabrielli, G. Jona-Lasinio, and C. Landim. Macroscopic fluctuation theory. *Rev. Mod. Phys.*, 87:593, 2015.

[10] A. Bérut, A. Petrosyan, and S. Ciliberto. Information and thermodynamics: experimental verification of Landauer's Erasure Principle. *J. Stat. Mech. Theory Exp.*, 2015:P06015, 2015.

[11] V. Blickle, T. Speck, C. Lutz, U. Seifert, and C. Bechinger. Einstein relation generalized to nonequilibrium. *Phys. Rev. Lett.*, 98:210601, 2007.

[12] K. Brandner, K. Saito, and U. Seifert. Thermodynamics of micro- and nano-systems driven by periodic temperature variations. *Phys. Rev. X*, 5:031019, 2015.

[13] G. Bunin, L. D'Alessio, Y. Kafri, and A. Polkovnikov. Universal energy fluctuations in thermally isolated driven systems. *Nat. Phys.*, 7:913, 2011.

[14] G. Bunin and Y. Kafri. Dynamics of energy fluctuations in equilibrating and driven-dissipative systems. *J. Phys. A Math. Theor.*, 46:095002, 2013.

[15] M. Campisi, P. Hänggi, and P. Talkner. Colloquium : Quantum fluctuation relations: Foundations and applications. *Rev. Mod. Phys.*, 83:771, 2011.

[16] Y. Cao, H. Wang, Q. Ouyang, and Y. Tu. The free-energy cost of accurate biochemical oscillations. *Nat. Phys.*, 11:772, 2015.

[17] S. Carnot. *Reflexions on the Motive Power of Fire: A Critical Edition with the Surviving Scientific Manuscripts*. Manchester University Press, Manchester, 1986.

[18] M. Colangeli, C. Maes, and B. Wynants. A meaningful expansion around detailed balance. *J. Phys. A Math. Theor.*, 44:095001, 2011.

[19] G. E. Crooks. Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences. *Phys. Rev. E*, 60:2721, 1999.

[20] G. E. Crooks. *Excursions in Statistical Dynamics*. PhD thesis, University of California at Berkeley, 1999.

[21] G. E. Crooks. Quantum operation time reversal. *Phys. Rev. A*, 77:034101, 2008.

[22] A. David and S. Larry. The least variable phase type distribution is Erlang. *Commun. Stat. Stoch. Model.*, 3:467, 1987.

[23] S. R. de Groot. *Non-equilibrium Thermodynamics*. Interscience Publishers, New York, 1962.

[24] A. Einstein. The theory of the opalescence of homogeneous fluids and liquid mixtures near the critical state. *Ann. Phys.*, 33:1275, 1910.

[25] J. L. England. Statistical physics of self-replication. *J. Chem. Phys.*, 139:121923, 2013.

[26] M. Esposito, K. Lindenberg, and C. Van den Broeck. Extracting chemical energy by growing disorder: efficiency at maximum power. *J. Stat. Mech. Theory Exp.*, 2010:P01008, 2010.

[27] M. Esposito and C. Van Den Broeck. Three detailed fluctuation theorems. *Phys. Rev. Lett.*, 104:90601, 2010.

[28] C. W. Gardiner. *Stochastic methods: A Handbook for the Natural and Social Sciences*. Springer-Verlag, Berlin, 4th edition, 2009.

[29] T. R. Gingrich, J. M. Horowitz, N. Perunov, and J. L. England. Dissipation bounds all steady-state current fluctuations. *Phys. Rev. Lett.*, 116:120601, 2016.

[30] A. Gomez-Marin, J. M. R. Parrondo, and C. Van den Broeck. Lower bounds on dissipation upon coarse graining. *Phys. Rev. E*, 78:011107, 2008.

[31] T. Hatano and S.-i. Sasa. Steady-state thermodynamics of Langevin systems. *Phys. Rev. Lett.*, 86:3463, 2001.

[32] K. Hayashi and S.-i. Sasa. Effective temperature in nonequilibrium steady states of Langevin systems with a tilted periodic potential. *Phys. Rev. E*, 69:066119, 2004.

[33] J. Horowitz and C. Jarzynski. Illustrative example of the relationship between dissipation and relative entropy. *Phys. Rev. E*, 79:021106, 2009.

[34] J. M. Horowitz. Diffusion approximations to the chemical master equation only have a consistent stochastic thermodynamics at chemical equilibrium. *J. Chem. Phys.*, 143:044111, 2015.

[35] J. M. Horowitz and J. M. R. Parrondo. Entropy production along nonequilibrium quantum jump trajectories. *New J. Phys.*, 15(8):085028, 2013.

[36] C. Jarzynski. Nonequilibrium equality for free energy differences. *Phys. Rev. Lett.*, 78:2690, 1997.

[37] C. Jarzynski. Rare events and the convergence of exponentially averaged work values. *Phys. Rev. E*, 73:046105, 2006.

[38] R. Kawai, J. Parrondo, and C. van den Broeck. Dissipation: The phase-space perspective. *Phys. Rev. Lett.*,

98:80602, 2007.

[39] T. S. Komatsu and N. Nakagawa. Expression for the stationary distribution in nonequilibrium steady states. *Phys. Rev. Lett.*, 100:30601, 2008.

[40] T. S. Komatsu, N. Nakagawa, S.-i. Sasa, and H. Tasaki. Representation of nonequilibrium steady states in large mechanical systems. *J. Stat. Phys.*, 134:401, 2009.

[41] T. S. Komatsu, N. Nakagawa, S.-i. Sasa, and H. Tasaki. Exact equalities and thermodynamic relations for nonequilibrium steady states. *J. Stat. Phys.*, 159:1237, 2015.

[42] G. Lan, P. Sartori, S. Neumann, V. Sourjik, and Y. Tu. The energy-speed-accuracy trade-off in sensory adaptation. *Nat. Phys.*, 8:422, 2012.

[43] R. Landauer. Stability and entropy production in electrical circuits. *J. Stat. Phys.*, 13:1, 1975.

[44] A. H. Lang, C. K. Fisher, T. Mora, and P. Mehta. Thermodynamics of statistical inference by cells. *Phys. Rev. Lett.*, 113:148103, 2014.

[45] G. Lebon, D. Jou, and J. Casas-Vazquez. *Understanding Non-equilibrium Thermodynamics*. Springer-Verlag, Berlin, Heidelberg, 2008.

[46] C. Maes and K. Netočný. Rigorous meaning of McLennan ensembles. *J. Math. Phys.*, 51:015219, 2010.

[47] M. C. Marchetti, J. F. Joanny, S. Ramaswamy, T. B. Liverpool, J. Prost, M. Rao, and R. A. Simha. Hydrodynamics of soft active matter. *Rev. Mod. Phys.*, 85:1143, 2013.

[48] U. M. B. Marconi, A. Puglisi, L. Rondoni, and A. Vulpiani. Fluctuation-dissipation: Response theory in statistical physics. *Phys. Rep.*, 461:111, 2008.

[49] R. Marsland III. *The Edge of Thermodynamics: Driven Steady States in Physics and Biology*. PhD thesis, Massachusetts Institute of Technology, 2017.

[50] R. Marsland III and J. England. Far-from-equilibrium distribution from near-steady-state work fluctuations. *Phys. Rev. E*, 92:052120, 2015.

[51] J. McLennan. Statistical mechanics of the steady state. *Phys. Rev.*, 115:1405, 1959.

[52] A. Murugan, D. A. Huse, and S. Leibler. Speed, dissipation, and error in kinetic proofreading. *Proc. Natl. Acad. Sci. U. S. A.*, 109:12034, 2012.

[53] A. Murugan, D. A. Huse, and S. Leibler. Discriminatory proofreading regimes in nonequilibrium systems. *Phys. Rev. X*, 4:021016, 2014.

[54] M. Nguyen and S. Vaikuntanathan. Design principles for nonequilibrium self-assembly. *Proc. Natl. Acad. Sci. U. S. A.*, 113:14231, 2016.

[55] Y. Oono and M. Paniconi. Steady state thermodynamics. *Prog. Theor. Phys. Suppl.*, 130:29, 1998.

[56] H. Ouerdane, Y. Apertet, C. Goupil, and P. Lecoeur. Continuity and boundary conditions in thermodynamics: From carnot's efficiency to efficiencies at maximum power. *The European Physical Journal Special Topics*, 224:839, 2015.

[57] J. Palacci, S. Sacanna, A. P. Steinberg, D. J. Pine, and P. M. Chaikin. Living crystals of light-activated colloidal surfers. *Science*, 339:936, 2013.

[58] P. Pietzonka, A. C. Barato, and U. Seifert. Universal bound on the efficiency of molecular motors. *J. Stat. Mech. Theory Exp.*, 2016:124004, 2016.

[59] P. Pietzonka, A. C. Barato, and U. Seifert. Universal bounds on current fluctuations. *Phys. Rev. E*, 93:052145, 2016.

[60] M. Polettini, G. Bulnes-Cuetara, and M. Esposito. Conservation laws and symmetries in stochastic thermodynamics. *Phys. Rev. E*, 94:052117, 2016.

[61] I. Prigogine. *Thermodynamics of Irreversible Processes*. Charles C. Thomas, Springfield, 1955.

[62] R. Rao and L. Peliti. Thermodynamics of accuracy in kinetic proofreading: dissipation and efficiency trade-offs. *J. Stat. Mech. Theory Exp.*, 2015:P06001, 2015.

[63] O. Raz, Y. Subaşı, and R. Pugatch. Geometric heat engines featuring power that grows with efficiency. *Physical review letters*, 116:160601, 2016.

[64] P. Reimann, C. Van den Broeck, H. Linke, P. Hänggi, J. M. Rubi, and A. Pérez-Madrid. Giant acceleration of free diffusion by use of tilted periodic potentials. *Phys. Rev. Lett.*, 87:010602, 2001.

[65] T. Sanchez, D. T. N. Chen, S. J. DeCamp, M. Heymann, and Z. Dogic. Spontaneous motion in hierarchically assembled active matter. *Nature*, 491:431, 2012.

[66] P. Sartori, L. Granger, C. F. Lee, and J. M. Horowitz. Thermodynamic costs of information processing in sensory adaptation. *PLoS Comput. Biol.*, 10:e1003974, 2014.

[67] P. Sartori and S. Pigolotti. Kinetic versus energetic discrimination in biological copying. *Phys. Rev. Lett.*, 110:188101, 2013.

[68] S.-i. Sasa. Possible extended forms of thermodynamic entropy. *J. Stat. Mech. Theory Exp.*, 2014:P01004, 2014.

[69] J. Schnakenberg. Network theory of microscopic and macroscopic behavior of master equation systems. *Rev. Mod. Phys.*, 48:571, 1976.

[70] U. Seifert. Entropy production along a stochastic trajectory and an integral fluctuation theorem. *Phys. Rev. Lett.*, 95:040602, 2005.

[71] U. Seifert. Stochastic thermodynamics, fluctuation theorems and molecular machines. *Rep. Prog. Phys.*, 75:126001, 2012.

[72] A. P. Solon, Y. Fily, A. Baskaran, M. E. Cates, Y. Kafri, M. Kardar, and J. Tailleur. Pressure is not a state function for generic active fluids. *Nat. Phys.*, 11:673, 2015.

[73] T. Speck, J. Mehl, and U. Seifert. Role of external flow and frame invariance in stochastic thermodynamics. *Phys. Rev. Lett.*, 100:178302, 2008.

[74] H. Touchette. The large deviation approach to statistical mechanics. *Phys. Rep.*, 478:1, 2009.

[75] S. Vaikuntanathan and C. Jarzynski. Dissipation and lag in irreversible processes. *Europhysics Lett.*, 87:60005, 2009.

[76] C. Van den Broeck. The many faces of the second law. *J. Stat. Mech. Theory Exp.*, 2010:P10009, 2010.

[77] N. G. van Kampen. *Stochastic Processes in Physics and Chemistry*. Elsevier, Amsterdam, 3 edition, 2007.