Unifying theories for nonequilibrium statistical mechanics

David M Rogers
University of South Florida, Tampa, FL, United States of America
E-mail: davidrogers@usf.edu

Received 4 February 2019
Accepted for publication 6 June 2019
Published 19 August 2019

Online at stacks.iop.org/JSTAT/2019/084010
https://doi.org/10.1088/1742-5468/ab3193

Abstract. The question of deriving general force/flux relationships that apply out of the linear response regime is a central topic of theories for nonequilibrium statistical mechanics. This work applies an information theory perspective to compute approximate force/flux relations and compares the result with traditional alternatives. If it can be said that there is a consensus on the form of response theories in driven, nonequilibrium transient dynamics, then that consensus is consistent with maximizing the entropy of a distribution over transition space. This agreement requires the problem of force/flux relationships to be described entirely in terms of such transition distributions, rather than steady-state properties (such as near-equilibrium works) or distributions over trajectory space (such as maximum caliber). Within the transition space paradigm, it is actually simpler to work in the fully nonlinear regime without relying on any assumptions about the steady-state or long-time properties. Our results are compared to extensive numerical simulations of two very different systems. The first is the periodic Lorentz gas under constant external force, extended with angular velocity and physically realistic inelastic scattering. There, we compare predicted and simulated distributions of the cumulative horizontal displacement after falling through 10 rows of fixed scatterers. The second is an $\beta$-Fermi–Pasta–Ulam–Tsingou (FPUT) chain, extended with a Langevin thermostat that couples to just two of its harmonic modes. This system tests whether the known long-time correlations in the FPUT system alter the properties of cumulative heat conduction away from the maximum entropy prediction. Although both systems are simulated for short, fixed times under physically realistic dynamical models, the maximum entropy structure of the time-integrated flows are clearly evident. The result encourages further development of empirical laws for nonequilibrium statistical mechanics by
employing analogies with standard maximum entropy techniques—even in cases where large deviation principles cannot be rigorously proven from an underlying model.

**Keywords:** driven diffusive systems, fluctuation theorems, large deviations in non-equilibrium systems, numerical simulations

---

1. **Introduction**

There is a growing consensus [1–4] that theories describing the full force/flux curve are connected by simple, general principles. However, the most compelling, simple examples are based on proving large deviation laws for sums of random numbers [5], empirical distributions [6] or Markov chains [7]. In this mathematical context, it can be difficult to make creative applications to simple physical systems, like a rotating dipole or fluid flow through a channel. Our goal in this work is to present an alternative point of view on the large deviation theory by justifying maximum entropy as a default model that applies approximately to the great majority of statistical dynamical laws. Maximum transition entropy provides a canonical form for nonequilibrium statistical mechanics that predicts key properties of force/flux relationships. We show these properties are recovered in two non-obvious (and non-Gaussian) examples: the intermediate-time
streamwise motion of a Lorentz gas under constant external force and the heat transferred between harmonic modes in FPUT chains. In the short-time limit, it instead predicts a Langevin equation of motion (that does have Gaussian statistics).

There are multiple theories of nonequilibrium statistical mechanics that have developed into essentially complete programs for studying open systems—defined here as mechanical systems driven by stochastic forces. Perhaps the earliest among these is thermodynamics itself, originally developed to describe the energy flows in engines driven by nonequilibrium flows of work and heat. The first and second laws are founded on the laws of conservation of energy, volume, mass, and charge, and therefore apply to all macroscopic nonequilibrium situations. Moreover, the equilibrium relations provide a default model for reservoirs (external to the system of interest) that store and deliver these quantities from the laboratory into an arbitrary dynamical system in any state [8]. In the thermodynamic limit, the equilibrium theory of statistical mechanics predicts the general form for probabilities of conserved quantities from information about the environmental reservoirs [9].

It is the goal of nonequilibrium statistical mechanics to provide the general form for rates of movement of conserved quantities within and between systems. Such equations of motion are the nonequilibrium analogues for the equilibrium equations of state. Also known as force/flux relationships, these equations of motion should give probabilities for the kinetics of processes given information about the state of the system and environment. For a general theory, we set the goal that the predictions of the theory should apply equally well to all types of open systems without relying on special assumptions of time-scale separation or preferred treatment of the atomic length scale.

The peculiar approach that will be taken in this work is to tackle the subjective problem of ascribing probabilities to the motion of a physical system that is interacting with a noisy environment. Since the environment will only be described in a statistical sense, the resulting probabilities may be greatly in error if there are conserved quantities in the dynamics that are not accounted for in the model. This is exactly the old problem with assuming the ergodic hypothesis when applying equilibrium statistical mechanics [10–12]. We paraphrase Jaynes in claiming a dual use for the results so obtained. Where the results are accurate, it provides us a canonical form for nonequilibrium thermodynamics. Where they disagree with experiment (either observations from physical or more accurate theoretical models), the disagreement shows evidence that the maximum entropy procedure did not account for relevant, reproducible information. Failure of the ‘canonical’ model implies the existence of conserved quantities in the dynamical laws governing the system/environment pair. Similarly, failure of the ‘canonical nonequilibrium’ model presented in this work prompts us to search for additional conserved quantities in the dynamics, eventually leading to new discoveries.

The sections that follow lay out the ‘canonical’ form (section 2), and then carry out two novel and non-trivial example applications introduced in section 3. We use the canonical form in two distinct modes. In ‘input mode’ it provides a microscopic, stochastic equation of motion for each system (section 3). In ‘output mode’ (in section 4), it is applied at a coarse-grained level to analyze their long-time dynamics. It is important to distinguish between these two modes—the first shows the utility of a canonical model for giving open boundary conditions to atomistic systems, and the second tests aggregate properties of the microscopic simulation for adherence to a
Unifying theories for nonequilibrium statistical mechanics

A general model should be applicable in both situations, since a coarse-grained model is also an open system with an implicit environment. Section 5 shows how this ‘canonical’ idea can be traced to the original Mori-Zwanzig and Green-Kubo theories, and provides some comparison to exact fluctuation theorems.

2. Maximum transition entropy

This section sets out our ‘canonical form’ for a transition probability distribution. It can be applied over a small or large discrete time steps—making it suitable for modeling approximate equations of motion for both microscopic and coarse-grained dynamical systems.

We adopt a setup similar to [13] for deterministic and stochastic discrete-time processes. The problem is cast as one of starting from a set of known coordinates for the subsystem of interest (e.g. $x_0$ at time $t_0$) and moving forward by some transition, labelled $g$, to a new set of coordinates, $x_1 = M(g; x_0)$, at time $t_1$. The map $M$ is a single-valued function. Small timesteps ($t_1 - t_0 = \Delta t$) are used later for input mode, and large timesteps ($t_1 - t_0 = T$) are used in output mode. The dynamics are made stochastic by introducing a probability space of possible transitions, $g \in G$, with an unknown underlying measure, $d\mu$. In general, the measure, $\mu$ is a function of the system’s past trajectory. We take it as fixed and time-invariant so that its associated dynamics are Markov. For continuous distributions, $d\mu(g)/dg =: P^0(g)$ is the probability distribution of $g$. Time is discretized, $t_1 \leq t_2 \leq \ldots$, according to any useful convention (equal time slices, first collision time, etc), and each transition (labeled $g_j$ for the transition taking $t_j \to t_{j+1}$) is associated with some (usually bounded) flow, $J(g_j, x_j)$. Each flow must measure the total amount of a conserved quantity ($J$) that was exchanged between the system and its environment between times $t_j$ and $t_{j+1}$, given that the starting coordinate was $x_j$. Here, conserved means that the flows would all be exactly zero if the system were not interacting with an external environment.

Within this setup, we can phrase a maximum entropy problem as follows: find the probability measure, $d\nu$, characterizing time interval $t_j \to t_{j+1}$, that maximizes the relative entropy,

$$S[d\nu|d\mu] = - \int_G d\nu \log \frac{d\nu}{d\mu}$$

under the constraint,

$$\langle J(g, x_j) | x_j \rangle := \int J(g, x_j) d\nu.$$  \hfill (1)

The solution is just the usual canonical distribution,

$$d\nu(g, x_j) = d\mu(g, x_j) e^{\lambda J(g, x_j)} Z_{\text{kin}}(\lambda, x_j),$$  \hfill (2)
(so that $P(g|\lambda) = \frac{d\nu}{dg}$) with Lagrange multiplier determined by the derivative of

$$Z_{\text{kin}}(\lambda, x_j) = \int_G e^{\lambda J(g, x_j)} d\mu.$$  

(4)

The distribution, $d\nu$, and its associated free energy $-\ln Z_{\text{kin}}(\lambda, x_j)$ and entropy (equation (1)), characterize the single temporal update occurring between time $t_j$ and $t_{j+1}$. Some related examples of this idea are present in [14, 15].

The motivation for using maximum entropy here is a subjective uncertainty about the underlying stochastic process [12, 16]. The ending result is the tilted exponential [17–19] and large deviation functions announced and studied by several authors [4, 7, 20]. However, we do not rely on complete knowledge of the underlying ‘default’ measure, $d\mu$. In the results below (as in experimental tests), $d\mu$ is treated as an empirical observable. Note that for a deterministic dynamical system, $d\mu$ would be a delta function.

Angle brackets are used throughout this work to denote averaging over an ensemble of trajectories $\Gamma_t = (x_0, x_1, \ldots, x_t)$. The trajectory’s starting conditions are given after a vertical line wherever there is a chance for ambiguity.

2.1. Input mode

The idea above can be used to derive a stochastic dynamical equation of motion. The required inputs are a Lagrangian and a description of the system-environment coupling. In the case of classical mechanical trajectories, $x(t)$, the uniform measure for $d\mu$ and a certain choice of $g$ leads to generalized Langevin equations. We sketch this application here, since we have used it to derive numerical integrators for both of the mechanical systems studied. In this context, it also dovetails nicely with the discrete Lagrangian method for deriving time-reversible symplectic integration methods.

Our choice is to define transition events, $g$, to be equal to deviations from an Euler–Lagrange equation of motion,

$$g = \frac{\delta A[x]}{\delta x},$$

(5)

where $A$ is a classical action functional. Then equation (5) uniquely determines the map, $\mathcal{M}$. It turns out that this ansatz has a plausible origin in quantum decoherence [21, 22]. Maximum entropy constraints are placed on

$$\langle D|\Gamma_i \rangle \equiv \langle g^2 dt|\Gamma_i \rangle, \quad \text{and} \quad \langle dE|\Gamma_i \rangle \equiv -\langle g dx|\Gamma_i \rangle.$$  

(6)

This leads to the well-known Langevin equation. The Lagrange multiplier conjugate to $D$ determines the rate at which equilibrium is approached, but does not affect the coordinate distribution in the stationary state. We define that multiplier to be $1/2\sigma^2$ in equation (7). In order to recover the canonical stationary distribution, the Lagrange multiplier conjugate to $dE$ must be $\beta/2$ (where $\beta^{-1} \equiv k_B T$ is the thermal temperature). The result for $\frac{\delta A[x]}{\delta x} = F(x) - \dot{p}$ (force minus momentum change) is the Langevin equation [8],

$$P(g = F - \dot{p}|x, p) \propto \exp \left( -\frac{g^2 dt}{2\sigma^2} + \beta dx \cdot g/2 \right).$$  

(7)

https://doi.org/10.1088/1742-5468/ab3193
This ansatz shows the Langevin equation to be the ‘canonical’ distribution for each step of dynamics carried out under equation (3).

2.2. Output mode

The formalism of section 2 can also be used to interpret dynamical processes where incomplete information is available. This type of problem is often seen in experimental tests of large deviation laws. Such laws can usually be discovered in cases where the underlying distribution is unknown, but a conservation law exists that demands a conjugate relationship between applied forces and corresponding flows. In this work, we do a similar experimental test by analyzing statistics of (scalar) integrated flows over time intervals spanning hundreds to thousands of integration time steps.

This conjugate relationship between forces and flows is clear in equation (3). The simplest consequence if equation (3) holds true is that the ‘kinetic’ partition function is the first integral of the force/flow curve,

\[ \langle J|\lambda, \Gamma \rangle = \frac{\partial \log Z_{\text{kin}}(\lambda, \Gamma)}{\partial \lambda}, \]  

while the fluctuation-dissipation theorem parallels the role of fluctuations in equilibrium,

\[ \langle \delta J^2|\lambda, \Gamma \rangle = \frac{\partial^2 \log Z_{\text{kin}}(\lambda, \Gamma)}{\partial \lambda^2}. \]  

These averages are conditional on the process history, \( \Gamma \), by the dependence of \( G, d\mu \) on \( \Gamma \) [20]. Indirectly showed their use for deriving Onsager reciprocity. Here, Onsager reciprocity applies to \( Z_{\text{kin}} \) in direct mathematical parallel to the equilibrium Maxwell-Gibbs relations based on the partition function. The full, causal, analogue of the Green–Kubo relations was demonstrated in [16].

New relations between transition probabilities can be shown directly from the ratio of equation (3) at two different applied forces,

\[ \log \frac{P(J|\lambda')}{P(J|\lambda)} = (\lambda' - \lambda)J - \log \frac{Z_{\text{kin}}(\lambda')}{Z_{\text{kin}}(\lambda)}. \]  

The relation obviously holds for the generalized Langevin equation (equation (7)), but also holds in all nonlinear cases where ‘canonical’ nonequilibrium statistical mechanics applies. However, because of its origin in maximum entropy rather than exact dynamics, it is better to be named a (forward) fluctuation relation then a fluctuation theorem proper. It should be qualified as ‘forward’ because it does not rely on time-reversal symmetry, but instead relies on conservation laws (associated by Noether’s theorem to continuous symmetries).

For the coarse-grained ‘experimental’ observables studied in later sections, the distribution over trajectories is taken to be an unknown that we wish to compare to equation (10). It is likely to hold only approximately because of the nonlinear nature of our systems. This contrasts with ‘input mode’—used for single time-steps of the microscopic dynamics.

In output mode, the correspondence between \( \lambda \) and the applied force, \( (E \text{ or } \Delta \beta \text{ in later examples}) \) is no longer direct. Instead, MaxTrans only predicts a canonical form
for transition probability distributions. In the same spirit as the Boltzmann/Gibbs distribution, \( \lambda \) and \( J \) are a conjugate pair, and their relation to a physical external field, \( E \), can usually be described by some function, \( \lambda(E) \). This relationship between generalized forces, \( \lambda \), and an applied physical force, is identifiable by any of four equivalent methods:

i. checking the ratio of equation (10) as a function of \( J \) for two different physical forces,

ii. matching mean and variance of the flux to the expansion,

\[
\langle J | \lambda' \rangle = \langle J | \lambda \rangle + (\lambda - \lambda) \langle \delta J^2 | \lambda \rangle + O(\Delta \lambda^2),
\]

iii. Green–Kubo style integration of the conjugate flux starting from \( \lambda = E \) at short time-scales (compare equation (7) to equation (44)) [8, 23], or

iv. differentiating \( \lambda(J) = d\sigma(\langle J \rangle)/d\langle J \rangle \), where

\[
\sigma(\langle J \rangle) = -\int dJ \ P(J|\lambda) \log \frac{P(J|\lambda)}{P(J|0)}. \tag{11}
\]

3. Stochastic modification of deterministic dynamical systems

This section describes the application of MaxTrans in input mode. We apply the constraints of equation (6) to derive two novel stochastic numerical integration schemes. Because the systems are simple enough, we are then able to provide analytical expressions for the resulting steady-state probabilities.

3.1. Inelastic periodic Lorentz gas

The periodic Lorentz gas describes a system of fixed scattering centers that cause rigid-body collisions of a single, spherical gas particle. The deflections of the studied gas particle cause it to undergo a random walk, mimicking an ideal gas. We simulated free flight of a single particle under constant external field \( \vec{E} = E \hat{x} + g \hat{y} \in \mathbb{R}^2 \) as a series of parabolic segments interrupted by discrete collisions. Scattering centers were placed on a regular 2D hexagonal lattice with side length \( L \). Numerically, collisions were detected by solving the quartic equation required to find the time of intersection of parabolic trajectories with one circular scatterer at the origin. By monitoring collisions with the unit cell boundaries and translating appropriately, only one particle-scatterer interaction needed testing during each computational update cycle.

On each collision, the particle’s location is unchanged, and an impulsive force is chosen at random following equation (7). An extra maximum-entropy constraint is added to enforce reflection of the particle’s velocity. The geometry in figure 1 is used in the following and defines the decomposition of the particle’s center of mass velocity into normal and tangential components and shows its angular velocity. We assume the
The particle is a uniform circular disk of radius $r$ with mass $M$ and moment of inertia $Mr^2/2$. The angular velocity is not considered in most treatments of the Lorentz gas, but must be included for a consistent set of energy equations. It is also needed to compute the tangential velocity, $v_t$, at the contact point.

Straightforward application of equation (7) (replacing $\beta$ with an unknown multiplier, $\tilde{\beta}$) would lead to

$$
\mathrm{d}p_n = -\left(\lambda + \frac{\tilde{\beta}}{2}\right)\sigma^2 v_n \mathrm{d}t + \sigma \mathrm{d}W_n
$$

(12)

$$
\mathrm{d}p_t = -\left(\frac{\tilde{\beta}\sigma^2}{2}\right)v_t \mathrm{d}t + \sigma \mathrm{d}W_t.
$$

(13)

Here, $\mathrm{d}p_n, \mathrm{d}p_t$ represent the normal and tangential forces added to the particle during the period of contact and $\mathrm{d}W_n, \mathrm{d}W_t$ are independent Wiener processes. To reach the impulsive force limit, we insist that an ‘inelasticity parameter’ $\gamma \equiv \frac{\tilde{\beta}\sigma^2}{2} \mathrm{d}t/2M$ remains finite in the limit $\mathrm{d}t \to 0$ so that $\sigma \mathrm{d}W = R \sqrt{2M\gamma/\tilde{\beta}}$, with $R$ a sample from the standard normal distribution. The impulses (now labeled $I_n = \mathrm{d}p_n$ and $I_t = \mathrm{d}p_t$) are then drawn from two standard normal distributions $(R_n, R_t)$,

$$
\frac{I_n}{M} = -\frac{\lambda\sigma^2}{M} v_n \mathrm{d}t - \gamma v_n + R_n \sqrt{2\gamma/M\tilde{\beta}}
$$

(14)

$$
\frac{I_t}{M} = -\gamma v_t + R_t \sqrt{2\gamma/M\tilde{\beta}}.
$$

(15)

This work used $\gamma = 0.01$. We also set $\lambda\sigma^2 \mathrm{d}t/M = 2$ to accomplish perfect reflection when $\gamma = 0$.

Adding this impulsive force to a rigid body results in the following stochastic map, $v \mapsto v'$, for updating all velocity components,

$$
-v'_n = (1 - \gamma)v_n + R_n \sqrt{2\gamma/\tilde{\beta}M}
$$

(16)

$$
v'_t - v_t = -\gamma(v_t + r\omega) + R_t \sqrt{2\gamma/\tilde{\beta}M}
$$

(17)

$$
r(\omega' - \omega)/2 = v'_t - v_t.
$$

(18)

The update happens at each collision time, $t$, corresponding to intersection of a parabolic ‘free’ trajectory with a circular scatterer’s boundary. The particle’s radius, $r$, need not be specified separately, since the equation of motion depends only on the product, $r\omega$.

Based on equation (7), we expect Lagrange multipliers, $1/2\sigma^2$ and $\tilde{\beta}$, to be associated to the thermal noise scale and the equilibrium temperature, respectively. To show this explicitly, it can be verified that the Boltzmann distribution,

$$
P(v_n, v_t, r\omega) \propto e^{-\frac{2M}{\beta}[\frac{(r\omega)^2}{2} + v_n^2 + v_t^2]},
$$

(19)

is a steady-state of the map (equation (16)) when $\vec{E} = \vec{0}$ and,

$$
\beta = (1 - 3\gamma/2)\tilde{\beta}.
$$

(20)
Damping increases the effective temperature. The difference between $\beta$ and $\tilde{\beta}$ is decreased slightly if the impulse (equation (14)) occurs at the center of a timestep, as for the Stratonovich stochastic calculus [24]. Our numerical simulations used the value of $\tilde{\beta}$ required by equation (20) to achieve a fixed inverse temperature, $(\beta M)^{-1} = 5.292 \cdot 10^{-19} \tau^2/L^2$.

Although highly unlikely because of the extremely large value of $\beta M$ used here, it is technically possible that the random increment to $v_n$ causes $v_n'$ to remain inward. Our simulation is therefore set to sample the appropriate truncated Gaussian for $R_n$ by generating random trials until one is found that leaves $v_n'$ pointing outward (away from the scatterer). Random noise is required by the fluctuation-dissipation theorem. With friction but no random noise, numerical simulations showed a few trajectories that settled into a stable limit cycle, stuck bouncing back and forth between the same two scatterers. The angular momentum did not play a role in the limit cycle, since it went quickly to zero. Our simulations included the small random noise, eliminating such occurrences.

The system simulated is illustrated by four randomly chosen trajectories shown in the top half of figure 2(a). The bottom half of figure 2(a) shows the complete $x_{final}$ histogram collected at row 10 for a small positive value of $E$. Further details are in section 4.1.

This type of model (under constant field) has been applied to study electron motion through insulators [25]. The zero-forcing case with elastic scattering was studied analytically by Sinai [10, 26], who showed that the trajectory of the particle over long times converges to a Brownian random walk, and that the expected direction of motion remains constant over time. The evolution of the probability density can be shown to converge to a Boltzmann transport equation [27, 28], and even has intuitive diffusive properties under a small, constant external force [29]. A review of approaches to the Lorentz gas was given by Spohn [30]. In the real-world case, the parabolic trajectories followed by the particle make exact analysis difficult. An analysis using a constant kinetic energy thermostat showed strong chaotic properties, including fractal scaling of the probability distributions for particle-scatterer impact [31]. With elastic collisions,
the kinetic energy of the particle must increase linearly as the particle falls. In that case, it has recently been shown that the particle velocity grows with time as $t^{1/3}$, and that (analogous to the Gambler’s ruin problem) for large enough starting velocity the particle will return to its initial height with probability 1 [32]. Our setup differs from earlier studies because of the presence of constant external force, inelastic collisions, and angular velocity.

### 3.2. Mode-coupled FPUT chains

To examine the time-course of energy redistribution between harmonic modes of a crystal lattice, Fermi, Pasta, Ulam and Tsingou (FPUT) simulated $N = 32$ points moving in 1D with unit masses and coordinates [33], $x_j, j = 1, \ldots, N$. This work uses periodic boundaries, so $x_0 = x_N$. The potential energy function is,

$$U(\vec{x}) = \sum_{j=0}^{N-1} V(x_{j+1} - x_j)$$

$$= \frac{1}{2} \sum_{j=0}^{N-1} (x_{j+1} - x_j)^2 + U_4(x)$$  \hspace{1cm} (21)

$$V(r) = r^2/2 + \epsilon r^4/4. \hspace{1cm} (22)$$

They discovered that for small anharmonic terms, energy did not seem to exchange, but only to oscillate regularly between harmonic modes.

Lack of fast energy exchange presents a physical problem for nonequilibrium models, since it means that simple diffusive dynamics is not appropriate. Rapid equilibration in these types of models does seem to occur if the initial energy is above some threshold value. Chirikov and Izrailev explained the existence of this threshold based on overlap of time-domain power spectral densities between spatial modes [34].

Recent, much longer, simulations and theory have shown that systems with small $\epsilon$ or small initial energy do, in fact, equilibrate but on a very long time-scale on the order of $\epsilon^{-1}$ [35]. Here, $\epsilon' = \epsilon(U(x^{(0)}) - U_4(x^{(0)}))/N$ is the nonlinearity parameter in Lvov and Onorato [35]. Those authors showed that the equilibration rate has a smooth cross-over at $\epsilon' \approx 0.12$. Above this value, ‘fast, Chirikov’ equilibration occurs on a time-scale of $\epsilon^{-1}$. All the simulations described here were thermal states of a pure harmonic system ($\epsilon = 0$) near $\beta = 1$ and $\epsilon = 0.1$, so $\langle \epsilon' \rangle = \epsilon/\beta = 0.1$.

To simulate this system numerically, we began by deriving a symplectic, volume-preserving dynamical integration scheme based on the Lagrangian,

$$L(x, \dot{x}) = \sum_j \frac{\dot{x}_j^2}{2} - U(x). \hspace{1cm} (24)$$

Following the procedure of Marsden [36], we make the substitutions, $\dot{x} \rightarrow (x^{(t)} - x^{(t-1)})/\Delta t$, $x \rightarrow x^{(t)}$, to construct a discrete action functional,
Figure 2. Left panels (a) and (c) show horizontal flow through the scatterer array—defined as the x-coordinate at first reaching \( y = 10 \). Trajectories are parabolic between collisions due to the constant external force \( (E + g) \). Right panels (b) and (d) illustrate the integrated heat flux from harmonic mode 7 to mode 3 in the Fermi–Pasta–Ulam–Tsingou lattice. The flow is nonzero because of the quartic term in the Hamiltonian. A steady-state exists because both modes are individually coupled to Langevin thermostats (temperatures \( \beta_{3}^{-1} \) and \( \beta_{7}^{-1} \)). Top panels (a) and (b) show example trajectories, along with a histogram of total flux at a single applied force. Bottom panels (c) and (d) summarize all simulations by plotting the average flux as a function of applied force. The insets of (c) and (d) show that histograms of the flux at large values of the force are non-Gaussian.

\[
A[x] = \sum_{t=1}^{T} \Delta t \, L_d(x^{(t)}, x^{(t-1)})
\]

\[
L_d(x', x) = \sum_{j} \frac{(x'_j - x_j)^2}{2\Delta t^2} - U(x').
\]
Requiring stationary action with respect to varying $x^i(t)$ would yield the Verlet (leapfrog) integration scheme common in molecular dynamics [37].

To illustrate the versatility of the Lagrangian approach, we show instead how to derive the stochastic integration method in Fourier space. First, define the coordinate transform:

$$X_k = \sum_{j=1}^{N} u^{-jk} x_j, \quad u \equiv e^{2\pi i/N}, \quad \omega_k^2 \equiv |u^k - 1|^2.$$  \hspace{1cm} (27)

Note that $X_{-k} = X_k^*$ and $X_{k+N} = X_k$ maintains the same number of degrees of freedom. Next, re-write the discrete action (equation (26)) in these new coordinates (defining $P_k \equiv (X_k' - X_k)/\Delta t$ for simplicity),

$$L_d(X', X) = \frac{1}{2N} \sum_{k=0}^{N-1} \left(|P_k|^2 - \omega_k^2 |X_k'|^2\right) - U_4(X').$$  \hspace{1cm} (28)

Here, $U_4$ represents the quartic potential terms as broken up in equation (22). According to MaxTrans, the exponent of the transition probability should be (using backward-differencing for $dX = P\Delta t$),

$$\Delta t \sum_{k=0}^{N-1} \frac{N}{2\sigma^2_k} \left| \frac{\delta A}{\delta X_k(t)} \right|^2 + \frac{\beta_k}{2} P_k(t) \frac{\delta A}{\delta X_k(t)}.$$  \hspace{1cm} (29)

Note how equation (6) behaves under a change in coordinate systems. The energy constraint is coordinate-free. The variance, $\sigma^2$, in the minimum deviation constraint undergoes a covariant transformation to $(\partial x/\partial X)^* \sigma^2 \partial x/\partial X$.

The thermostatted equations of motion for $X_k$ can be read off from the Gaussian mean and variance found by factoring equation (29)

$$- \left( \frac{\delta A}{\delta X_k(t)} \right)^* = - \frac{\beta_k \sigma^2_k}{2N} P_k + \frac{\sigma_k}{\sqrt{N\Delta t}} Z_k\hspace{1cm} (30)$$

Here $Z_k = Z^*_{-k}$ is a complex standard normal random variable ($\langle |Z_k|^2 \rangle = 2$).

The steady-state of equation (30) at $\epsilon = 0$ can be determined analytically to be a multivariate Gaussian with variances:

$$\langle X_k^2 \rangle = \frac{N}{\beta_k \omega_k^2} \left( 1 + \Delta t \frac{16 - \beta_k^2 \sigma_k^4}{4 \beta_k^2 \sigma_k^2} \right) + O(\Delta t^2)$$  \hspace{1cm} (31)

$$\langle P_k^2 \rangle = \frac{N}{\beta_k} \left( 1 + \frac{4\Delta t}{\beta_k \sigma_k^2} \right) + O(\Delta t^2)$$  \hspace{1cm} (32)

$$\langle X_k P_k \rangle = - \frac{N \Delta t}{2 \beta_k} \left( 1 + \frac{4\Delta t}{\beta_k \sigma_k^2} \right) + O(\Delta t^3).$$  \hspace{1cm} (33)
4. Results

This section defines coarse-grained variables and analyzes simulation results that test equation (10) and the resulting properties listed in section 2.2. This reflects a typical experimental situation, where the fine-grained dynamical details are hidden or unobserved.

4.1. Inelastic periodic Lorentz gas

Stochastic numerical simulations of the Lorentz gas were carried out as described in section 3.1. As a coarse-level observable, we created a histogram of $x_{\text{final}}$, the $x$-coordinate when $y$ first reaches 10 rows down. A histogram collected for that particular $E$ value is shown in the bottom half of figure 2(a). Two additional histograms are inset into figure 2(c) showing non-Gaussian behavior.

All results presented here were collected from the hexagonal lattice shown. The particle-to-scatter contact distance is $0.4L$. The time scale, $\tau$, was set so that the graviation constant is $9.8 \text{ ms}^{-2} = \frac{1}{10^2} L^2/\tau^2$. 102 400 trajectories were simulated with uniform random starting locations on the line $y = 0, x \in (-0.1L, 0.1L)$ and velocities chosen from a Gaussian distribution with variance $(\beta M)^{-1} = 5.292 \cdot 10^{-19} \tau^2/L^2$. This is consistent with a physical scatterer diameter of 6.35 cm and mass of 5 g.

Figure 3(a) plots the ratios, $P(x_{\text{final}}|E')/P(x_{\text{final}}|E)$ for successive values $(E, E')$ of the applied field strength. Despite the transient initial conditions, the relatively short simulation times, and the non-Gaussian nature of the histogram (figure 2(c)), the MaxTrans postulate, equation (3), appears to hold with $d\mu = P(x_{\text{final}}|E = 0)dx$. The mapping between the applied field, $E$, and the parameter, $\lambda$, is nonlinear.

Figure 4(a) compares properties $i$ and $ii$ (listed at the end of section 2) by plotting $d\langle x_{\text{final}}|\lambda \rangle/d\lambda$ (lines) from the force-flux curves of figure 2 against the fluctuations $\langle \delta x^2 \rangle_\lambda$ (shown as points). The two smooth lines on each plot come from the derivatives obtained from 3rd and 7th order numerical differencing schemes. Differences between these lines are a measure of the uncertainty in collected force versus flux data. The correspondence, $\lambda = \lambda(E)$ was made by the fits to the slopes in figure 3. Linear response predicts that $\langle x_{\text{final}} \rangle$ will be proportional to $\lambda$, $\lambda$ proportional to the force, and hence $\langle \delta x^2_{\text{final}} \rangle$ will be constant. Such a constant region appears near $E = 0$ in our results. For larger values of the field, the fluctuations become large, but match the derivatives of the $x_{\text{final}}(\lambda)$ curve. That curve was computed by integrating $d\lambda(x_{\text{final}})$, where $\lambda(x_{\text{final}}') = \lambda(x_{\text{final}})$ was fitted as shown in figure 3.

4.2. Mode-coupled FPUT chains

Stochastic numerical simulations of FPUT chains were carried out as described in section 3.2. In our simulations, we coupled modes $k = \pm 3$ and $k = \pm 7$ to a stochastic environment, but let the others retain their deterministic dynamics (equivalent to setting $\sigma_k^2 = 0$). This models a molecular crystal where the external environment can be strongly coupled (optically or acoustically) to a few harmonic modes. That coupling carries with it both random noise and deterministic dissipation. The specific parameters used were $\beta_3 = 1 + \Delta \beta$, $\beta_7 = 1 - \Delta \beta$, and $\sigma_3 = \sigma_7 = 0.1$ and $\Delta t = 0.01$. 

https://doi.org/10.1088/1742-5468/ab3193
The coarse variable chosen for this system is energy transfer between harmonic modes, and requires some explanation. Energy exchange between modes can be monitored by examining various decompositions of the energy change (using equations (30) in (6)),

$$\frac{\Delta E}{\Delta t} = \sum_k P_k^* \left( \frac{P' - P}{\Delta t} - F^{(2)} - F^{(4)} \right)_k.$$  

(34)

Writing the momentum update implied by equation (30) as $\frac{P' - P}{\Delta t} = F^{(2)} + F^{(4)} + F^{(\text{lang})}$, and noting that the first two forces conserve energy, we get

$$\frac{\Delta E}{\Delta t} = \sum_k P_k^* F_k^{(\text{lang})} = \sum_k \frac{\Delta E_k}{\Delta t} - P_k^* F_k^{(4)}.$$  

(35)

Here the harmonic energy in mode $k$ is defined as,

$$E_k = (|P_k|^2 + \omega_k^2 |X_k|^2) / 2N.$$  

(36)

From this point of view, the time-derivative of $E_k$ represents the flux of energy from both the anharmonic system and the Langevin thermostat combined.

To filter out noise coming from the Langevin thermostat, we define the ‘heat flux’ into mode $k$ to be just the quartic term, $\Delta Q_k/\Delta t = P_k^* F_k^{(4)}$. It was computed numerically from equation (35) as the difference between the harmonic oscillator’s energy change and the energy added from the Langevin thermostat. Comparing to equation (36) shows $E_k(t) - E_k(0) = Q_k(t) + Q_k^{(\text{lang})}(t)$. By conservation, $E_k(t)$, should reach a steady-state so $Q_k$ and $Q_k^{(\text{lang})}$ will mirror one another (and $Q_3 + Q_7$ will fluctuate about a fixed constant for large $t$). Each coarse variable, $Q_k$, thus reflects net heat flow from anharmonic energy exchange. For individual modes that are coupled to a ‘hot’ reservoir, we will accordingly observe heat flow out of that mode into anharmonic degrees of freedom. No heat flow between modes is possible when $\epsilon = 0$. This was verified numerically to test our implementation.
Figure 2(b) (left side) shows example trajectories of integrated energy flow versus time from mode \( k = 7 \) to \( k = 3 \) in the FPUT system at \( \epsilon = 0.1 \). Distributions of \( Q = Q_3(T) - Q_7(T) \) presented here were calculated at time \( T = 1638.4 \). Its distribution for that particular value of \( \Delta \beta \) is shown on the right side of figure 2(b). All trajectories include an initial transient of approximately 100 time units because the initial conditions were chosen from the canonical distribution for the harmonic system \( (\epsilon = 0) \) at uniform temperature \( \beta = 1 \).

Figure 2(d) also shows that the flow is a nonlinear function of the applied force \( (\Delta \beta \) for the FPUT lattice). This is especially apparent at large values of the forcing, where the histograms over flow \( (Q, \text{figure 2(b), right and figure 2(d), inset}) \) are markedly non-Gaussian. While fluctuations in the flow were almost constant near the origin for the Lorentz gas, linear response does not hold as well in the FPUT system. Nevertheless, our numerical data show the applicability of equation (10) well out of the linear response regime. Figure 3 illustrates property \( i \) and figure 4 tests property \( ii \). The fluctuations of the current still correspond to the slope of the flux-force curve away from \( \lambda = 0 \).

5. Discussion and comparison

Often, applied literature provides specialized fluctuation-dissipation or fluctuation theorems that give little hint as to how they may be generalized or extended. In fact, the original derivations allow quite a bit of flexibility in defining what forces and flows can enter, and can be put into a form very much resembling our basic results (equations (8)–(10)). We now discuss these alternative viewpoints by re-stating those earlier works in terms of time-derivatives (flows) rather than absolute positions.
5.1. Projection-operator and fluctuation-dissipation theorems

The projection operator-operator theory gives a rigorous, general equation of motion for the probability distribution of coarse coordinates like the particle position or the energy in each mode \([38, 39]\). The theory clearly indicates where closure relations are required. This section shows how the simplest closure relations with Gaussian noise can be derived by analogy to Gaussian processes. The result provides time-dependent Green–Kubo relations applicable at nonzero driving force. They are linear because they model only the slope of the flow versus force curve \([16]\).

An accessible derivation of the projection operator theory was given by Nordholm and Zwanzig \([40]\) with the result,

\[
\frac{\partial}{\partial t} \mathbb{P} f(t, x) = -\mathbb{P}i\mathcal{L} \mathbb{P} f(t, x) \\
+ \int_0^t ds \mathbb{P} \mathcal{L} e^{-i(1 - \mathbb{P}) \mathcal{L}} (1 - \mathbb{P}) i \mathcal{L} \mathbb{P} f(t - s, x) \\
- \mathbb{P} i \mathcal{L} e^{-i(1 - \mathbb{P}) \mathcal{L}} (1 - \mathbb{P}) f(0, x).
\]

(37)

The operator, \(\mathbb{P}\), projects the phase-space probability density, \(f(t, x)\), onto a subspace of the full phase space, \(\{x \in \Omega\}\). The Liouville operator is defined in terms of the Poisson brackets, \(i\mathcal{L} J = \{J, H\} = dJ/dt\). There is no difficulty interpreting this subspace as an arbitrary manifold lying inside \(\Omega\). For any point, \(x\), we can define the projected point on the manifold as \(\phi(x)\) so that

\[
\mathbb{P} f(x) = \int_\Omega dx' \delta(x - \phi(x')) f(x').
\]

(38)

The projected equation of motion (equation (37)) implicitly defines the probability distribution of transition events, \(g\), corresponding to movements on the projected manifold, \(\phi(x) \rightarrow \phi(x')\). It is trivial to re-cast it in this way, since \(dt \frac{\partial}{\partial t} \mathbb{P} f(t, x) = \int dx' \mathbb{P} (x| x', dt) f(t, x') - f(t, x)\) as the timestep, \(dt \rightarrow 0\).

The push-forward operation, \(i\mathcal{L} \mathbb{P} f(t, x)\), in both of the first two terms refers explicitly to this transition. The three parts of the equation of motion on the manifold (equation (37)) have the interpretation of \(i\) the deterministic transition (\(\mathbb{P}i\mathcal{L}\mathbb{P}\)) for points on the manifold, \(ii\) the memory function describing the predictable, but delayed effect due to earlier transitions, \((1 - \mathbb{P})i\mathcal{L} \mathbb{P} f(t - s, x)\), and \(iii\) the ‘random’ noise part due to initial conditions not on the manifold.

Because the real system is not constrained to travel on the manifold, the construction makes it clear that closure relations are required for describing parts \((ii)\) and \((iii)\). Specific choices for those closures form the starting points for mode coupling theory \([40]\) and nonlinear fluctuating hydrodynamics \([41]\). The latter has also been applied to dynamics in large-FPUT chains \([42]\).

These three parts directly translate to the terms in equation (14) or equation (30). Those can be traced to the following parts of the MaxTrans formalism. The deterministic transition, \(i\) and \(ii\), comes from the default measure, \(d\mu\). That default measure is a function of the coordinate history, and should be centered around the most likely next step. The random noise in \(d\mu\) provides \((iii)\) directly. Adding Lagrange multipliers to tilt
\( d\mu \) provides external forcing and memory terms together. Some additional ideas on this comparison can be found in [43].

In practice, most applications of the theory have used linear closure relations, which give rise to linear transport equations [44]. Many nonlinear closures are most easily understood by comparison to the linear theory [45, 46]. It has been pointed out [47] that the principle results of the linear theory are identical to linear regression.

The linear regression case has been treated in a very general way in the Gaussian process literature [48]. The critical assumptions are that the random noise obeys Gaussian statistics and that the coefficients of the memory function depend only on time, not on the process history. The equations below relate to the two systems considered here by replacing \( g \) with the horizontal motion of the disk, \( dx \), or the heat transfer, \( dQ \), over a small amount of time. The regression equations can be summarized by the assumption,

\[
P(\{g(t_i)\}_{0}^{n}) = GP[m(t_i), k(t_i, t_j)]
\]

which implies the following generating process,

\[
g(t_n) = m(t_n) + \sum_{j=k=1}^{n} k(t_n, t_{n-j})k_{n-j,n-k}^{-1} \times (g(t_{n-k}) - m(t_{n-k})) + \sigma_n R_n
\]

\[
\sigma_n^2 = k(t_n, t_n) - \sum_{j,k=1}^{n} k(t_n, t_{n-j})k_{n-j,n-k}^{-1}k(t_n, t_{n-k}).
\]

Here, \( GP \) denotes a Gaussian process, which is a multivariate normal distribution with mean \( m(t) \) and variance-covariance matrix \( k(t, t') \). Equation (40) states the applicable fluctuation-dissipation theorem—namely that the probability of \( g \) at the next step has a Gaussian distribution with a mean that is linear in the random increments, \( g(t_k) - m(t_k) \), and a variance, \( \sigma_n^2 \), that is reduced by knowledge of the process history. The variable \( R_n \) is an independent sample from the standard normal distribution.

This closure is demonstrated by noting the terms in equation (40) correspond 1:1 with those of equation (37). Brownian motion theory is recovered when \( g \) is taken to be the momentum. Then \( m(t) \) is the drift velocity and \( k(t, t) = k_B T/m \) near equilibrium. If, instead, \( g_n \) were taken to be the momentum update, \( p_n - p_{n-1} \), then \( m \) must be the deterministic part of the force, and equation (40) describes Langevin dynamics.

In the Langevin case, a little algebra shows that applying a single external force at time \( t_0 \) will not only directly shift \( g_0 \rightarrow g_0 + F_{0}^{ext} \), but will also accumulate a net effect at later times, \( t_k \) of,

\[
\delta g_{k+0} = [k(t_k, t_J)]^{T} k_{J,I}^{-1} \begin{bmatrix} \delta g_0 \\ \delta g_{1+0} \\ \vdots \\ \delta g_{k-1+0} \end{bmatrix}.
\]

The recursion is solved by

\[
\delta g_{k+0} = \frac{k(t_k, t_0)}{k(t_0, t_0)} F_{0}^{ext}.
\]
In comparison with our main result (equation (8)), this externally forced process could have been derived extremely easily by adding an exponential bias to the basic Gaussian process (equation (39)),

\[
P(\bar{g}|F^{\text{ext}}) \propto \exp \left\{ - (\bar{g} - \bar{m})^T k^{-1} (\bar{g} - \bar{m})/2 + (\bar{g} - \bar{m})^T (F^{\text{ext}}/\text{diag}(k)) \right\}\tag{44}
\]

This is the revised Onsager–Machlup action functional approach [19, 49]. It is also the result of adding a mean-value constraint of the form \(\delta A/\delta x(t) \cdot F^{\text{ext}}\) to the MaxTrans postulate in equation (7).

The linear transport theory can thus be cast into the language of Gaussian processes. The technical content of the celebrated fluctuation-dissipation theorem in this case is a statement of how dissipation of an external force, \(\sum_{j<n} k(t_n, t_{n-j})\), is related to fluctuations of the current, \(k(t_n, t_j) = \langle \delta g_n \delta g_j \rangle\).

We can see that this line of attack applies to time-dependent processes, but Gaussian processes do not make it clear how to extend the theory into the nonlinear regime. The major contribution of section 2 was to replace the fitting ansatz of equation (39) with a single-step fitting ansatz at time \(t_n\). This uses only the transition probabilities, \(P(g_n|\{g(t_i)\}_{0}^{n-1})\), instead of equation (39). The mean and variance, \(m_n\) and \(k_{nj}\) can be arbitrary nonlinear functions of the history, \(\{g\}_{0}^{n-1}\)—just as allowed by a careful read of the projection operator theory. Equation (10) continues to work in an obvious and transparent way, however, it is much harder to arrive analytically at the steady-state properties.

5.2. Fluctuation theorems and chaotic hypothesis

Fluctuation theorems address the probabilities of transitions even more directly. Specifically, they transform symmetries of the dynamical equations into symmetries of integrated quantities such as work and heat. They have a history stretching back to Callen and Welton [50], who proved a fluctuation theorem showing the odds of heat, \(Q\), flowing from cold to hot versus the reverse process go exponentially with the entropy change of ideal, equilibrium environmental reservoirs (proportional to \(\exp(Q\Delta\beta)\)).

Though the literature on fluctuation theorems is large, I provide here only a few results. The first fluctuation theorems about atomistic trajectories were developed by several groups [51, 52], who proved theorems of the form,

\[
\frac{1}{t} \log \frac{P(g = \sigma(t, x)/t|x_0 \to x_n)}{P(g = -\sigma(t, x)/t|x_n \to x_0)} \asymp g.\tag{45}
\]

The symbol, \(\asymp\), means asymptotic convergence with large time, \(t\). The conditioning on coordinates, \(x_0\) or \(x_n\), indicates whether the trajectory is initiated from a starting or ending point. For the transient fluctuation theorem, the microscopic entropy production is identified with the time-integral over a trajectory of length \(t\) starting from \(x(0) = x\),

\[
\sigma(t, x) = \int_0^t dt \ D(x(t)).\tag{46}
\]
where $D(x)$ is the ratio of phase space volume between the last and next time-step, $|i\mathcal{L}(dx)|/|dx|$. Because of the dependence on the starting/ending point, there are differences in the relations and proofs depending on whether the starting states are fixed or chosen at random from an SRB measure (steady-state), whose existence and uniqueness requires additional assumptions [20].

In the case where a dynamical system can be modeled as a finite-state Markov process, a new version of the fluctuation theorem (equation (45)) can be shown [53, 54], where $\sigma$ ($W$ of [53]) is the log-ratio of forward to reverse transition probabilities over $t = n$ steps of the Markov process,

$$\sigma(n, x) = \sum_{i=0}^{n-1} \log \frac{P(x_{i+1}|x_i)}{P(x_i|x_{i+1})}. \quad (47)$$

Although they apply in different cases, the two fluctuation relations essentially express the same measure of irreversibility, since the probability of a transition scales inversely with the starting volume, $P(x_{i+1}|x_i) \propto 1/|dx_i|^1$.

Since the log-ratio of transition probabilities are often related to work and entropy production, the fluctuation theorems can make quantitative statements about energy exchange during transitions of a dynamical system. Although equation (45) appears to be a special case of equation (10), equation (45) has been proven to hold under more general conditions. Its relation to symmetry provides it with a unique status in that it is closer to a dynamical law than a statistical one.

It is also possible to use $\sigma$ for describing simultaneous transition probabilities of several coarse variables [51, 55]. This interpretation often refers to $\sigma$ as a dissipation function, or, in reference to the Onsager–Machlup theory, as an action function. Our alternative derivation of $\sigma$ as a generalized entropy connected to equation (1) provides a more canonical explanation of this connection between $\sigma$ and force/flux relations.

Postulates like equation (3) have been hinted at before in connection with Onsager–Machlup theory [56], but their generality and use for deriving fluctuation theorems has not been widely appreciated.

A similar theory of Jaynes, named maximum caliber (MaxCal) [47], applied maximum entropy to the set of all trajectories. Unfortunately, that theory does not maintain causality [8], since forces applied in the future influence the entire history of the process. This shortcoming has not caused identifiable problems where the theory has been applied [57–59], since MaxCal and MaxTrans make the same probability assignments when there is time translation invariance of the forces. In those situations, non-causality only causes MaxCal to break down when applying the fluctuation theorems. The situation can be summed up by noting that path counting gives non-causal weights to trajectories, whereas the transition probabilities are always causal. This defect was noted in later revisions of MaxCal that bring it into line with MaxTrans [60].

Both Jaynes [61] and Haken [62, 63] investigated maximizing the entropy of the transition distribution as a restriction of MaxCal. Those early works chose steady-state averages as constraints, rather than the instantaneous flows as done here. Since the Lagrange multipliers had to be identified through the Fokker–Planck equation, that

---

1 To make this statement rigorous requires comparing the number of transitions that can be made out of state $x_i$ versus those out of state $x_{i+1}$.
choice hid the connection to the equilibrium, Boltzmann/Gibbs distribution. Had they made the latter choice, they would have immediately discovered our kinetic partition function, $Z_{\text{kin}}$, along with its attendant force-flux relationships and FDTs.

The canonical form of equation (7) codifies flux/force relations in a coordinate-independent way using a maximum entropy structure. Because of this, it provides facile derivations for both time-dependent Green–Kubo response theories and the fluctuation theorems when the postulate of equation (3) holds. Its shortcoming is that it does not directly predict the relationship between the forces and flows, e.g. $\beta$ and $\langle dE \rangle$. This, however, is exactly the well-known problem of determining the equations of state in equilibrium statistical mechanics.

6. Conclusions

We have constructed a probability distribution over values of flows that occur during short- and intermediate-time transitions using maximum entropy. The nonlinear, maximum entropy structure applies when there is an unknown, and hence subjectively random, environment. For classical mechanical systems, an ansatz was given to derive a Langevin equation from an action functional. Each step of the Langevin equation obeys the fluctuation relation (equation (3)) exactly by construction. Integrating over steps then leads to traditional Green–Kubo expressions. Because the atomic scale is not special, it is also possible to jump over steps by guessing that equation (3) holds for coarse-grained properties over long time-scales. In that case, it is a statement of a maximum entropy postulate that must be experimentally verified wherever it is used.

Our experience shows that conjugate force/flux pairs are connected to conservation laws governing interactions between the coarse-grained system and its environment. Numerical simulations of driven diffusion in 2D gas models and heat diffusion between modes of a crystal supported this view. We expect the postulate to be more easily satisfied as the number of transitions grows—even for transient, driven processes. The success of the linear response methods and of the fluctuation theorems both rest on exploiting the existence of this structure within the transition distribution.

This theory has been directly related to complementary methods of attack. The Green–Kubo relations provide quick estimates for the conductivities, even in transient steady-states. This work showed the connection by focusing on the transition quantities in equation (40), which have the same structure as equations (8) and (9). Fluctuation theorems provide connections with distributions of work and notions of irreversibility. We have shown the forward fluctuation relation of equation (10) derives equation (45) when the flux is related to number of transitions. Transition distributions provide an entry point to these theories, but can also be applied in their own right as a maximum entropy structure. Although it derives mathematically identical results, it is logically distinct from large deviation theories because maximum entropy infers transition distributions even when the underlying microscopic noise process is unknown. It is also distinct from Gaussian processes or near-equilibrium theories that use the idea of local entropies because it focuses specifically on one-step transitions [47, 49, 64, 65]. All general approaches are structural in the sense that finding analytical expressions for
long-time force/flux relationships is a difficult task, even for the Lorentz gas [25, 29] and FPUT model [66–68]. Our line of reasoning avoids that problem, and instead parallels the reasoning used to derive equilibrium equations of state. Its most common criticism—that certain analytic properties of the microscopic distribution must be proven—directly parallels the integrability objection to the ergodic hypothesis.

Acknowledgments

I thank the anonymous reviewers for suggesting improvements to the presentation of this work. Financial support was provided by the USF Research Foundation and NSF MRI CHE-1531590.

References

[1] Bustamante C, Liphardt J and Ritort F 2005 The nonequilibrium thermodynamics of small systems Phys. Today 58 43–8
[2] Maes C, Netočný K and Wynants B 2008 Steady state statistics of driven diffusions Physica A 387 2675–89
[3] Jarzynski C 2015 Diverse phenomena, common themes Nat. Phys. 11 105–7
[4] Chetrite R and Touchette H 2015 Nonequilibrium Markov processes conditioned on large deviations Ann. Inst. Henri Poincaré 16 2005–57
[5] Varadhan S R S 2008 Large deviations Ann. Probab. 36 397–419
[6] Bonetto, Gallavotti G and Garrido P L 1997 Chaotic principle: an experimental test Physica D 105 226–52
[7] Touchette H 2009 The large deviation approach to statistical mechanics Phys. Rep. 478 1–69
[8] Rogers D M and Rempe S B 2012 Irreversible thermodynamics J. Phys.: Conf. Ser. 402 012014
[9] Jaynes E T 1957 Statistical mechanics Phys. Rev. 106 620–30
[10] Lebowitz J and Penrose O 1973 Modern ergodic theory Phys. Today 26 23–9
[11] Jaynes E T 1979 Where do we stand on maximum entropy? The Maximum Entropy Formalism ed R D Levine and M Tribus (Cambridge, MA: MIT Press) p 498
[12] Christian Maes 1999 The fluctuation theorem as a Gibbs property J. Stat. Phys. 95 367–92
[13] Pincus S M 1992 Approximating Markov chains Proc. Natl Acad. Sci. USA 89 4432–6
[14] Derrida B 2007 Non-equilibrium steady states: fluctuations and large deviations of the density and of the current J. Stat. Mech. P07023
[15] Baiesi M, Maes C and Netočný K 2009 Computation of current cumulants for small nonequilibrium systems J. Stat. Phys. 135 57–75
[16] Rogers D M, Beck T L and Rempe S B 2011 An information theory approach to nonlinear, nonequilibrium thermodynamics J. Stat. Phys. 145 385–409
[17] Jiang D-Q, Qian M and Zhang F-X 2003 Entropy production fluctuations of finite Markov chains J. Math. Phys. 44 4176
[18] Rondoni L and Mejía-Monasterio C 2007 Fluctuations in nonequilibrium statistical mechanics: models, mathematical theory, physical mechanisms Nonlinearity 20 R1
[19] Bertini L, De Sole A, Gabrielli D, Jona-Lasinio G and Landim C 2015 Macroscopic fluctuation theory Rev. Mod. Phys. 87 593–636
[20] Gallavotti G 1996 Extension of Onsager’s reciprocity to large fields and the chaotic hypothesis Phys. Rev. Lett. 77 4334–4337
[21] Rogers D M 2017 An information theory model for dissipation in open quantum systems J. Phys.: Conf. Ser. 880 012039 (8th Int. Workshop DICE2016)
[22] Gao S (ed) 2018 How to Teach and Think About Spontaneous Wave Function Collapse Theories: Not Like Before (Cambridge: Cambridge University Press) pp 3–11
[23] Evans D J, Searles D J and Williams S R 2016 Fundamentals of Classical Statistical Thermodynamics (New York: Wiley)
[24] Stratonovich R L 1992 Nonlinear Nonequilibrium Thermodynamics I: Linear and Nonlinear Fluctuation-Dissipation Theorems vol 57 (Berlin: Springer)
[25] Chepelianskii A D and Shepelyansky D L 2001 Phys. Rev. Lett. 87 034101
Unifying theories for nonequilibrium statistical mechanics

[63] Haken H 1993 *Application of the Maximum Entropy Principle to Nonlinear Systems Far from Equilibrium* (Cambridge: Cambridge University Press) pp 239–50

[64] Prigogine I 1965 Steady states and entropy production *Physica* **31** 719–24

[65] Spohn H and Lebowitz J L 1978 Irreversible thermodynamics for quantum systems weakly coupled to thermal reservoirs *Advances in Chemical Physics* vol 38, ed S A Rice (New York: Wiley) pp 109–42

[66] Eckmann J P, Pillet C A and Rey-Bellet L 1999 Non-equilibrium statistical mechanics of anharmonic chains coupled to two heat baths at different temperatures *Commun. Math. Phys.* **201** 657–97

[67] Spohn H 2014 Nonlinear fluctuating hydrodynamics for anharmonic chains *J. Stat. Phys.* **154** 1191–227

[68] Onorato M, Vozella L, Proment D and Lvov Y V 2015 Route to thermalization in the $\alpha$-Fermi–Pasta–Ulam system *Proc. Natl Acad. Sci. USA* **112** 4208–13