

## Detailed balance has a counterpart in non-equilibrium steady states

This article has been downloaded from IOPscience. Please scroll down to see the full text article.

2005 J. Phys. A: Math. Gen. 38 293

(<http://iopscience.iop.org/0305-4470/38/2/001>)

View [the table of contents for this issue](#), or go to the [journal homepage](#) for more

Download details:

IP Address: 171.66.16.66

The article was downloaded on 02/06/2010 at 20:12

Please note that [terms and conditions apply](#).

# Detailed balance has a counterpart in non-equilibrium steady states

**R M L Evans**

University of Leeds, LS2 9JT, UK

Received 30 August 2004, in final form 27 October 2004

Published 15 December 2004

Online at [stacks.iop.org/JPhysA/38/293](http://stacks.iop.org/JPhysA/38/293)

## Abstract

When modelling driven steady states of matter, it is common practice either to choose transition rates arbitrarily, or to assume that the principle of detailed balance remains valid away from equilibrium. Neither of those practices is theoretically well founded. Hypothesizing ergodicity constrains the transition rates in driven steady states to respect relations analogous to, but different from, the equilibrium principle of detailed balance. The constraints arise from demanding that the design of any model system contains no information extraneous to the microscopic laws of motion and the macroscopic observables. This prevents over-description of the non-equilibrium reservoir, and implies that not all stochastic equations of motion are equally valid. The resulting recipe for transition rates has many features in common with equilibrium statistical mechanics.

PACS numbers: 05.20.-y, 05.70.Ln, 83.50.Ax

## 1. Introduction

I address the question: what is the appropriate stochastic equation of motion to use when modelling a driven steady state (including chaotic and fluctuating steady states) such as that of a fluid under continuous shear flow? At equilibrium, the solution is well understood. To generate configurations consistent with the equilibrium ensemble, one may use any equation of motion that respects the principle of detailed balance, which is a constraint on ratios of forward and reverse transition rates. That condition ensures that every thermally-driven flux is balanced by an equal and opposite flux. For non-equilibrium systems in continuously driven steady states, no such guidance is hitherto available in choosing an equation of motion consistent with the *mechanically* (externally) driven fluxes, so arbitrary choices are often made. The aim of this work is to eliminate arbitrariness, and determine what transition rates are implied by the macroscopic state of the non-equilibrium system, i.e. its mean energy and flux, combined with our knowledge of the microscopic laws of physics. The objective is to use *only* the

information that is available, without unwittingly introducing any arbitrariness, deriving from personal prejudices. The method for keeping the amount of information constant throughout the calculation is Jaynes' information-theoretic method of maximum entropy inference [1, 2] (MaxEnt), which is often misunderstood in the context of non-equilibrium thermodynamics, despite recent notable achievements [4]. It has been successfully used to derive fluctuation theorems [5] and linear transport theory [3], and to explain self-organized criticality [4].

Jaynes gives a nice explanation of maximum entropy inference in his original paper on the subject [1], where he uses the method to re-derive equilibrium statistical mechanics without the need for many microscopic details that had previously been considered necessary. The application of the method to equilibrium systems is uncontroversial. However, the history of non-equilibrium information theory can be confusing because it has been used in so many different ways, some of them exact, some only approximate. In fact, information theory itself is not a physical theory, but a mathematical method, providing a logical structure. Some physical input is required if such a method is to make physical predictions. If one throws away too much relevant information about some non-equilibrium system before applying MaxEnt, it will still provide answers, but they will be inaccurate. For instance, using the method to minimize the information content of the momentum distribution in a non-equilibrium gas, although efficacious, is not an exact method, as was recently shown [6]. In fact, there is no justification for discarding all information content except for some averaged features. Indeed, particles possess their individual velocities for a reason: they have each come from somewhere, and are going somewhere, and their journeys will affect the trajectories of other particles. These facts are relevant to the physics of a non-equilibrium system, and lead to temporal correlations.

At the other extreme, if one retains all the details of a system's phase-space trajectory, allowing no stochastic input (e.g. from a reservoir), then MaxEnt becomes a null procedure, since it is asked to choose the most likely distribution from a choice of only one physical scenario—a delta function distribution of trajectories. Such a null procedure may be regarded as an extreme case where MaxEnt can correctly 'predict' any and all physics. There is thus no reason in principle why MaxEnt should be expected to fail in non-equilibrium situations, if we ask it the right questions.

The choice of the prior set of options that is presented to MaxEnt is of crucial importance. It should be a set of physical paths through phase space, that each obeys Newton's laws, so that all physics (the Navier–Stokes equation, long-range correlations, etc) is respected *a priori*. MaxEnt then tells us which of these trajectories is most likely to be chosen, under the influence of a non-equilibrium reservoir that is coupled to the system but uncorrelated with it<sup>1</sup>. This is the application of information theory that should be understood here and in [7, 8]. I derive its implications for transition rates.

Alternatively, as is often done in theoretical modelling, one can settle for a physically imprecise prior set of dynamical rules—such as Brownian particles, or a discrete state space, or discrete time steps—so that things become easy to solve. Then applying the methods below will, accordingly, yield only approximate physics, but at least one will know exactly what information went into the simplified model. Such an application of the present theory would yield transition rates that are somewhat arbitrary, due to the arbitrariness of the prior rates that are chosen. However, it will provide strictly *the least* arbitrary model. Such a model will be derived in section 4 for a stochastically hopping particle that demonstrates some

<sup>1</sup> The 'reservoir' may in fact stand for the rest of the ensemble of systems, as in the equilibrium derivation of the Gibbs ensemble.

features of the method. It is important to realize that the approximations introduced in that section are only for expediency in that particular model. The general derivation of the method for obtaining transition rates from prior dynamical rules combined with non-equilibrium macroscopic observables, presented in sections 2 and 3, remains exact.

The conditions derived here for macroscopically driven steady states are analogous to the equilibrium principle of detailed balance. Like detailed balance, the conditions are not sufficient to completely determine the microscopic transition rates, but are necessary to be satisfied by any equation of motion that generates an unbiased ergodic driven steady-state ensemble. The derivation of detailed balance relies on two assumptions: time-reversal symmetry of the microscopic laws of motion, and the ergodic hypothesis which implies that a heat reservoir can be characterized by the Boltzmann distribution with temperature as the only parameter. Similarly, the non-equilibrium conditions assume the same microscopic laws that govern equilibrium motions (therefore implicitly requiring microscopic time-reversal symmetry, broken only by imposition of the macroscopic flux), and rely also on a hypothesis of ergodicity implying that the driven reservoir is fully characterized by its macroscopic observables (mean energy and flux). Many quiescent systems (those without fluxes) are at thermodynamic equilibrium, but exceptions include glasses [9], granular media [10, 11] and certain cellular automata [12], in which the ergodic hypothesis and/or microscopic reversibility fails. Boltzmann's law and the principle of detailed balance apply only to that class of quiescent systems that are, by definition, at equilibrium. That class of systems has of course proved to be large, significant and interesting. Similarly, not every non-equilibrium steady state should be expected to respect the conditions presented here; exceptions include traffic flow and fluids of molecular motors, in which the constituents violate time-reversal symmetry. The ergodic hypothesis may also fail in some systems, implying that hidden information that is not apparent in the macroscopic observables is nonetheless significant. However, it is anticipated that the ergodicity criteria are respected by the transition rates of many macroscopically driven systems, defining a special and important class.

The method outlined in section 2 was presented in a recent letter [7]. It is explained here in more detail, and the analysis extended to an alternative non-equilibrium ensemble in section 3. The method is demonstrated in section 4.1 where rates are derived for the stochastic transitions of a particle hopping in a non-trivial energy landscape, subject to a driving force. Applications to other models are also discussed in section 4.

## 2. The method

### 2.1. Information entropy

Using Jaynes' interpretation of Gibbs' entropy [1], it is possible to make a 'maximum entropy inference' [1, 4] to assess the probability that a system, subject to random influences, (whether at equilibrium or not) takes a particular trajectory  $\Gamma_0$  through its phase space, thus allowing us to assess the *reproducible part* [1, 3] of the system's motion. The recipe for the probability  $p(\Gamma_0)$  of trajectory  $\Gamma_0$  is to maximize the Shannon entropy, or information entropy,

$$S_I \equiv - \sum_{\Gamma} p(\Gamma) \ln p(\Gamma) \quad (1)$$

subject to constraints that some averaged properties of the trajectories conform with our knowledge of the macroscopic features such as mean energy, volume, flux etc.

In principle, this formalism gives us a full solution of the statistics of any ensemble, be it at equilibrium or not. In the absence of any macroscopic fluxes (i.e. at equilibrium), the prescription reduces to a maximization of the Gibbs entropy with respect to a distribution of

instantaneous states rather than trajectories, yielding Boltzmann's law. In the non-equilibrium case, MaxEnt gives us the probability of an entire trajectory  $\Gamma_0$ . It would be more useful to have a formula for the probability of a short segment of the trajectory, a single transition from a state  $a$  to a subsequent state  $b$ . Such a transition probability is what we require for designing a stochastic model or simulation. This would allow us to generate trajectories belonging to the non-equilibrium ensemble. Let us now derive that formula. We begin with some trivial calculations to establish notation.

### 2.2. Prior probability

At any instant  $t$ , the entire state of a system is represented classically by its phase-space position vector  $x(t)$ . This is a high-dimensional vector specifying the positions and momenta of all the particles constituting the system. As time progresses from the beginning  $t = 0$  to the end  $t = \tau_0$  of the experiment or simulation,  $x(t)$  traces out a trajectory  $\Gamma_0$  through phase space. It will prove useful to label each probability distribution function with a subscript indicating the duration of the trajectories to which it applies thus:  $p_{\tau_0}(\Gamma_0)$ . For a deterministic system with definite initial conditions, only one trajectory is possible, so the probability distribution is a delta function. In the presence of randomness, such as coupling to a reservoir of systems with similar properties, the distribution is finite for all trajectories that respect some prior dynamical rules, such as conservation of momentum for all internal degrees of freedom not directly coupled to the reservoir.

In the *absence* of any posterior constraints other than normalization,

$$\sum_{\Gamma} p_{\tau_0}(\Gamma) = 1, \quad (2)$$

all trajectories of a given duration  $\tau_0$  have equal *a priori* probability. That is not an independent postulate, but is embodied in the maximum entropy principle of information theory [1, 3], since the entropy-maximizing distribution is given by

$$\frac{\partial}{\partial p_{\tau_0}(\Gamma_0)} \sum_{\Gamma} \{-p_{\tau_0}(\Gamma) \ln p_{\tau_0}(\Gamma) + \lambda p_{\tau_0}(\Gamma)\} = 0 \quad (3)$$

with a Lagrange multiplier  $\lambda$  chosen for consistency with equation (2). Equation (3) is solved by  $p_{\tau_0}(\Gamma_0) = \text{constant}$ , indicating that the unconstrained ('prior') set of trajectories of a given duration have equal probability.

### 2.3. Equilibrium ensemble

We now impose a posterior constraint, and calculate the statistical properties of that sub-set of trajectories that respect the constraint. Let us not necessarily conserve the energy  $E$  of the system at each instant (since we allow energy exchange with a reservoir), but rather demand that its time-average over the whole trajectory  $\Gamma_0$  is fixed at  $E_0$ . We shall use a bar to indicate time averages, so that

$$\overline{E_{\Gamma_0}} \equiv \frac{1}{\tau_0} \int_0^{\tau_0} E_{\Gamma_0}(t) dt = E_0. \quad (4)$$

Let us divide the trajectory  $\Gamma_0$  into shorter segments  $\Gamma$ , each of duration  $\tau$ . Then the constraint on the time-averaged energy may be written as

$$\tau \sum_{\Gamma} \overline{E_{\Gamma}} = \tau_0 E_0. \quad (5)$$

Assuming ergodicity, time-averages are equivalent to ensemble-averages in the limit  $\tau_0/\tau \rightarrow \infty$ . So this constraint, for a time-average on  $\Gamma_0$ , defines the equilibrium canonical ensemble for  $\Gamma$ . In other words, the conditional probability  $p_\tau(\Gamma|\overline{E_{\Gamma_0}} = E_0)$  of encountering a particular trajectory segment  $\Gamma$  of duration  $\tau$ , given that the whole trajectory has a time-averaged energy  $E_0$ , is found by maximizing the information entropy for  $\Gamma$  subject to equation (5). The maximization involves Lagrange multipliers  $\beta$  for this energy constraint, and  $Z^{-1}$  for the normalization constraint, and yields

$$p_\tau(\Gamma|E_0) = Z^{-1} \exp(-\beta \overline{E_\Gamma})$$

where the condition  $\overline{E_{\Gamma_0}} = E_0$  is represented for brevity by  $E_0$ . As expected, this is Boltzmann's law, and we interpret the Lagrange multipliers as the temperature parameter  $\beta = 1/k_B T$  and partition function  $Z = \sum_\Gamma \exp(-\beta \overline{E_\Gamma})$ .

#### 2.4. Transition rates

A transition rate, for any transition between states  $a \rightarrow b$  say, can be written as a conditional probability. If we consider a trajectory segment  $\Gamma'$  of duration  $\Delta t$ , representing the transition  $a \rightarrow b$ , then the transition rate at some time, which we may define without loss of generality to be  $t = 0$ , is

$$\omega_{\Gamma'}^{\text{prior}} = p_{\Delta t}(\Gamma'|x(0) = a)/\Delta t. \quad (6)$$

This is the probability (per unit time) of encountering the trajectory  $\Gamma' \equiv \{x(0) = a, x(\Delta t) = b\}$ , given that we begin at  $a$ . Equation (6) gives the *prior* rate of the particular transition. The rate in the equilibrium ensemble is given by a probability subject to *two* conditions:

$$\omega_{\Gamma'}^{\text{eq}} = p_{\Delta t}(\Gamma'|a, E_0)/\Delta t \quad (7)$$

where the condition  $x(0) = a$  is represented for brevity by  $a$ . To re-cap, equation (7) defines the probability of encountering trajectory segment  $\Gamma'$  (a transition  $a \rightarrow b$ ) given that we are in state  $a$ , and that the entire trajectory  $\Gamma_0$  of duration  $\tau_0$  will turn out to have a mean energy  $E_0$ .

#### 2.5. Driven ensemble

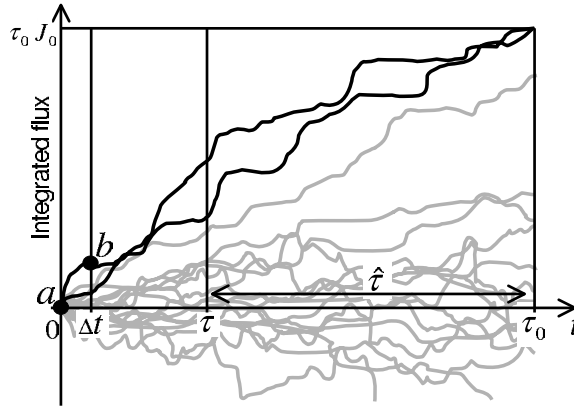
We have looked so far at the prior phase-space trajectories, and those for systems in the equilibrium ensemble. Our goal is to determine the transition rates appropriate to a non-equilibrium ensemble, for which there is an imposed flux  $J$ . Again, we should not over-constrain the dynamics. Let us allow the flux to fluctuate, and demand only that the dynamics will result in some finite value  $J_0$  of the flux time-averaged over the whole trajectory:

$$\overline{J_{\Gamma_0}} \equiv \frac{1}{\tau_0} \int_0^{\tau_0} J_{\Gamma_0}(t) dt = J_0. \quad (8)$$

We ask: what is the probability, in time  $\Delta t$ , of encountering the transition  $\Gamma' = \{a \rightarrow b\}$ , given that we begin in state  $a$ , and that the dynamics will eventually conspire to produce a mean flux  $J_0$  and energy  $E_0$ ? Again we relate this conditional probability to a transition rate:

$$\omega_{\Gamma'}^{\text{dr}} = p_{\Delta t}(\Gamma'|a, J_0, E_0)/\Delta t. \quad (9)$$

Figure 1 depicts some of the trajectories that have been discussed. Time  $t$  is shown on the horizontal axis, and all trajectories have a total duration  $\tau_0$ . The vertical axis represents the phase-space coordinates though, of course, this is a reduced representation of the vastly high-dimensional phase space, since it has been projected onto a single axis. For definiteness, let us say that this axis represents integrated flux, i.e. the flux that the system has accrued since



**Figure 1.** A sample of phase-space trajectories of duration  $\tau_0$ . One phase-space coordinate (cumulative flux) is shown as a function of time  $t$ . At equilibrium, most trajectories lie near the horizontal axis. The sub-set of trajectories shown in black belongs to the microcanonical-flux ensemble. Their time-averaged flux is  $J_0$ . The probability of a transition, in microscopic time  $\Delta t$ , between the phase-space points  $a$  and  $b$  shown, is enhanced for this sub-set, relative to its equilibrium likelihood.

$t = 0$ . We must imagine that all the other coordinates required to fully describe the state of the system are on axes perpendicular to the page.

A sample of trajectories representing the equilibrium distribution is shown (in grey and black). These trajectories are concentrated close to the time axis (zero flux). If another axis measuring instantaneous energy  $E(t)$  were constructed perpendicular to the page, then the density of trajectories would be exponentially distributed along that axis, by Boltzmann's law. Equation (7) gives the frequency of observing a particular trajectory segment shown in figure 1 (the single transition  $a \rightarrow b$ ) of microscopic duration  $\Delta t$ , given that we are currently (at  $t = 0$ ) in state  $a$ , and that the whole trajectory belongs to this equilibrium set. Equation (9) asks for the frequency with which that trajectory segment  $\{a \rightarrow b\}$  occurs in the sub-set of trajectories shown in black in figure 1, for which a given integrated flux will be accumulated by time  $t = \tau_0$ . This sub-set of trajectories is the driven ensemble.

Note that we shall not require  $\Delta t$  to vanish. The discussion will cover discrete-time processes for which the microscopic time step is  $\Delta t \equiv 1$ , as well as continuous-time dynamics for which  $\Delta t \rightarrow 0$ .

## 2.6. Bayesian evaluation

To mathematically manipulate conditional probabilities, we appeal to Bayes' theorem. It states that the joint probability of two outcomes  $X$  and  $Y$  both occurring, given a third fact  $Z$ , may be written in two equivalent ways:

$$p(X|Z)p(Y|X, Z) = p(Y|Z)p(X|Y, Z) \quad (10)$$

where  $p$  is simply used to mean 'probability' for any event (appropriately normalized), as opposed to a particular distribution function. We can now assign the following meanings:  $X$  is the fact that the transition  $a \rightarrow b$  takes place within  $\Delta t$ , represented by the trajectory  $\Gamma'$ ;  $Y$  says that the flux has a mean value  $\overline{J_{\Gamma_0}} = J_0$  averaged over the entire duration  $\tau_0$ ; and  $Z$  is the combined statement that the initial state at  $t = 0$  is  $a$  and that the trajectory's time-averaged energy will be  $\overline{E_{\Gamma_0}} = E_0$ . Thus, equation (10) expresses the probability of the transition taking



place within  $\Delta t$  and the flux averaged over  $\tau_0$  being  $J_0$ , for the given initial state and average energy. It is re-written thus:

$$p_{\tau_0}(J_0|a, E_0)p_{\Delta t}(\Gamma'|a, J_0, E_0) = p_{\Delta t}(\Gamma'|a, E_0)p_{\tau_0}(J_0|a, \Gamma', E_0). \quad (11)$$

Note that it is redundant to specify the two conditions  $(a, \Gamma')$ , since the trajectory segment  $\Gamma'$  is the transition  $a \rightarrow b$  which includes the initial state  $a$ . Substituting from equations (7) and (9) yields a theorem for transition rates in the driven steady-state ensemble,

$$\omega_{a \rightarrow b}^{\text{dr}} = \omega_{a \rightarrow b}^{\text{eq}} \lim_{\tau_0 \rightarrow \infty} \frac{p_{\tau_0}^{\text{eq}}(J_0|a \rightarrow b)}{p_{\tau_0}^{\text{eq}}(J_0|a)}. \quad (12)$$

Note that all quantities on the RHS of equation (12) are defined *at equilibrium*, not on the driven ensemble. This is indicated by the superscript ‘eq’, which is equivalent to the condition fixing  $\overline{E}_{\Gamma_0}$ , the time-averaged energy. Equation (12) tells us that the transition rate in the driven ensemble is given by the transition rate in the equilibrium ensemble, multiplied by an enhancement or attenuation factor. We shall see below that the theorem makes intuitive sense.

Given that the dynamics must be consistent with the macroscopically observable mean energy and flux, and with the same microscopic laws of motion that hold sway in an equilibrium system, MaxEnt yields an unbiased description of the dynamics, and thereby constrains the system the least. Equation (12) specifies explicitly the dynamical rules implied by MaxEnt. How do we know that equation (12) constrains the dynamics the least? It does, because all the quantities on the RHS are defined for the maximum-entropy ensemble at equilibrium, i.e. without the extra constraint on the flux. Given that we start with an unbiased set (the equilibrium ensemble), Bayes theorem gives us the least biased set subject to the extra posterior constraint.

Let us examine the enhancement factor in equation (12) in detail. It is a ratio of conditional probabilities for encountering a flux  $J_0$  in the equilibrium ensemble. Of course, we do not expect a system at equilibrium to exhibit any net flux, averaged over its whole trajectory. The chance of such a flux arising spontaneously at equilibrium is vanishingly small (as  $\tau_0 \rightarrow \infty$ ). So the RHS of equation (12) is the ratio of two vanishingly small terms. However unlikely it may be for an equilibrium system to spontaneously exhibit the desired macroscopic flux, we ask, how much would that probability be enhanced as a result of the putative transition  $a \rightarrow b$ ? If the dynamics of the transition itself contributes some flux to the trajectory, it is favoured by the enhancement factor. The factor also favours transitions to configurations that give a greater than average probability of subsequently obtaining the desired flux, for the given starting point. If the new state  $b$  is more likely to initiate high-flux trajectories, then the transition rate is boosted over and above the equilibrium rate.

We shall examine the implications of equation (12) in some examples, but firstly let us interpret the meaning of its derivation. Imagine that a lazy physicist wishes to collect data from a driven steady state, such as continuous shear flow of a complex fluid. Our physicist has a computer program that simulates the fluid at equilibrium (with free or frictionless boundaries, say), and is too lazy to write a new program that simulates shear. Instead, (s)he runs the equilibrium simulation, in the hope that it will spontaneously exhibit shear flow. It does not. So the dilettante updates the program’s random number generator and runs it again. Having tenacity and little imagination, the physicist repeats this process countless times until, one day, the fluid fluctuates into a state of sustained shear flow. The delighted simulator records this fluke, but continues the project for many more years until a large number of such accidents have been observed, exhibiting the same shear rate. Finally, the researcher discards an enormous set of simulated trajectories, and publishes only that subset which happened to perform the desired shear. On analysing this subset of trajectories, one might expect to observe the equilibrium



transition rates that were coded into the algorithm. But this is a biased data set, subject to an *a posteriori* constraint of shear flux  $J_0$ . So this sub-set of the equilibrium ensemble exhibits exactly the transition rates specified by equation (12). Although the programmer has published a biased account of the equilibrium simulations, there was no unwarranted or subjective bias other than the flux constraint, hence the project was a success in producing the physics of shear flow.

Note that, despite extracting a sub-ensemble from the equilibrium ensemble, the lazy physicist has *not* produced a near-equilibrium approximation. The ‘sub-ensemble dynamics’ of equation (12) has features qualitatively different from the equilibrium dynamics.

In section 4, I shall use some examples to demonstrate the correctness of the physics generated by sub-ensemble dynamics (equation (12)). Before doing so, in section 3, I develop a useful variant of equation (12), analogous to an alternative thermodynamic ensemble.

### 3. Alternative dynamic ensembles

#### 3.1. Microcanonical-flux ensemble

Equation (12) gives the frequency of observing a particular trajectory segment (e.g. a single transition  $a \rightarrow b$ ) of microscopic duration  $\Delta t$ , in the driven ensemble which is a sub-set of all trajectories, shown in black in figure 1. These trajectories lie in the extreme tails of the equilibrium distribution. Note that they have common end points, since we have specified the exact net flux that must flow during the duration of the experiment. Any nearby trajectories, that do not have *exactly* the specified flux, do not contribute to the quantities appearing in equation (12). Even very nearby trajectories are completely discarded by equation (12). This can be seen by re-writing the probability of the specified flux  $J_0$  as a sum over trajectories  $\Gamma_0$  with fluxes  $\overline{J}_{\Gamma_0}$ , so that equation (12) becomes

$$\begin{aligned} \frac{\omega_{a \rightarrow b}^{\text{dr}}}{\omega_{a \rightarrow b}^{\text{eq}}} &= \frac{\int dJ p_{\tau_0}^{\text{eq}}(J|a \rightarrow b) \delta(J - J_0)}{\int dJ p_{\tau_0}^{\text{eq}}(J|a) \delta(J - J_0)} \\ &= \frac{\sum_{\Gamma_0} p_{\tau_0}^{\text{eq}}(\Gamma_0|a \rightarrow b) \delta(\overline{J}_{\Gamma_0} - J_0)}{\sum_{\Gamma_0} p_{\tau_0}^{\text{eq}}(\Gamma_0|a) \delta(\overline{J}_{\Gamma_0} - J_0)}. \end{aligned} \quad (13)$$

Here, the Dirac delta functions kill all trajectories with anything but the exact net flux  $J_0$ . This can be a disadvantage for practical applications of the formula. (The lazy physicist, discussed above, must discard data even from simulations that produce *nearly* the right flux.) An alternative expression is now derived, that samples trajectories with less stringent conditions on their eventual flux content.

#### 3.2. Canonical-flux ensemble

In equilibrium statistical mechanics, the constraint of energy conservation is relaxed by dividing the isolated microcanonical system into a relatively small sub-section, defining the canonical system, and the large remainder, known as the reservoir. Similarly, we shall relax the strict constraint on the time-averaged flux, by dividing the total trajectory of duration  $\tau_0$  into a part (see figure 1) of duration  $\tau$  (where  $\Delta t \ll \tau \ll \tau_0$ ), whose properties are examined in detail, and the large remaining part of duration  $\hat{\tau} \equiv \tau_0 - \tau$ , for which the system’s motion is uncorrelated with the earlier trajectory segment.

We may express the conditional probability  $p_{\tau_0}(J_0|a)$  of a net flux  $J_0$  in the full duration  $\tau_0$ , as an integral over all possible fluxes during the interval  $\tau$  thus:

$$p_{\tau_0}^{\text{eq}}(J_0|a) = \int_{-\infty}^{\infty} dJ p_{\tau}^{\text{eq}}(J|a) p_{\hat{\tau}}^{\text{eq}}(\hat{J}|a, J) \quad (14)$$

where  $p_{\hat{\tau}}(\hat{J}|a, J)$  is the probability of an appropriate flux  $\hat{J}$  during interval  $\hat{\tau}$  given that the system began in state  $a$  at  $t = 0$ , and then flowed with mean flux  $J$  for the duration  $\tau$ . The required flux  $\hat{J}$  is given by

$$\tau J + \hat{\tau} \hat{J} = \tau_0 J_0. \quad (15)$$

Given that  $\tau$  exceeds any correlation time  $\tau_{\text{cor}}$ , the probability  $p_{\hat{\tau}}(\hat{J}|a, J)$  becomes independent of  $a$ , because the system has forgotten its initial state by the time  $\tau$  at which the interval  $\hat{\tau}$  commences. In fact, at time  $\tau$ , the system is in a state drawn at random from the driven steady-state ensemble, since the integral in equation (14) is dominated by  $J \approx J_0$ . So we may make the replacement

$$p_{\hat{\tau}}(\hat{J}|a, J) \rightarrow \sum_c p^{\text{dr}}(c) p_{\hat{\tau}}^{\text{eq}}(\hat{J}|c) \quad (16)$$

where  $p^{\text{dr}}(c)$  is the steady-state distribution of instantaneous microstates in the driven ensemble. Not only is the above formula independent of the initial state  $a$ , it actually takes a universal (exponential) form as a function of  $J$ , as shown in appendix A using the theory of large deviations. This is because the extremely unlikely value of the flux,  $\hat{J}$ , is the result of many unlikely realisations of the flux during the many uncorrelated intervals that comprise the large duration  $\hat{\tau}$ . As a result, equation (12) can be re-cast, using equations (14), (15) and (16), and the derivation in appendix A, as

$$\frac{\omega_{a \rightarrow b}^{\text{dr}}}{\omega_{a \rightarrow b}^{\text{eq}}} = \lim_{\tau/\tau_{\text{cor}} \rightarrow \infty} \frac{\int dJ p_{\tau}^{\text{eq}}(J|a \rightarrow b) e^{\tau \nu J}}{\int dJ p_{\tau}^{\text{eq}}(J|a) e^{\tau \nu J}} \quad (17)$$

where the control parameter  $\nu$  is conjugate to the time-averaged flux, and is fixed by the relation

$$\frac{\partial Q}{\partial \nu} = J, \quad (18)$$

in terms of the function

$$Q(\nu) \equiv \lim_{\tau \rightarrow \infty} \frac{\ln \langle e^m \rangle_{\text{dr}}}{\tau}. \quad (19)$$

Here,  $\langle \dots \rangle_{\text{dr}}$  is an ensemble average with respect to the steady-state distribution of microstates  $p^{\text{dr}}(c)$ . We have defined

$$m_c(\nu, \tau) \equiv \ln \int_{-\infty}^{\infty} p_{\tau}^{\text{eq}}(J|c) e^{\tau \nu J} dJ \quad (20)$$

that is a property of an instantaneous state  $c$  of the system. Note that  $m_c(\nu, \tau)$  has non-trivial  $\tau$ -dependence, containing transients for  $\tau < \tau_{\text{cor}}$ , and becoming linear in  $\tau$  for  $\tau \gg \tau_{\text{cor}}$ , while  $Q(\nu)$  is independent of  $\tau$ .

The above equations have a structure that is familiar from equilibrium thermodynamics. Clearly, in equation (19),  $Q$  plays the role of a thermodynamic potential, and its derivative  $J$  is conjugate to the temperature-like parameter  $\nu$ .

The conditional probabilities in the integrands of equation (17) describe the likelihood of any particular flux during the interval  $\tau$ , given the initial state and/or transition. The exponential factor measures the change in the weight of the large remainder of the trajectory of duration  $\hat{\tau}$ , due to the initial part accepting a flux  $J$  rather than postponing it until after  $\tau$ .

Compare equations (13) and (17). The expressions become very similar under a change of notation  $\tau \leftrightarrow \tau_0$ . The difference in the new formulation (equation (17)) is that trajectories with the wrong flux are not eliminated by a delta function, but re-weighted by an exponential weight factor.

The two alternative formulations are exactly akin to alternative ensembles in equilibrium statistical mechanics. We can regard the duration of a trajectory as being analogous to the size of a system at equilibrium, and the flux as analogous to energy-density. Originally we demanded that the integrated flux was fixed exactly, just as energy is fixed in the micro-canonical ensemble, and we enquired, in equation (13), about how the instantaneous ('local') conditions are affected by correlations in the rest of the trajectory ('system'). The formulation of equation (17) is akin to using the canonical ensemble. Again, we enquire about conditions at an instant ('locality')  $\Delta t$ , within a trajectory segment ('system') of duration ('size')  $\tau$ . But now, the integrated flux ('energy') is not strictly conserved, but can be exchanged with the rest of the trajectory ('a reservoir') of duration ('size')  $\hat{\tau}$  much longer ('larger') than the initial trajectory segment ('system'). Since all important correlations are contained within the 'system', the nature of the interface between 'system' and 'reservoir' becomes unimportant, and the 'reservoir' is characterized by a single parameter,  $\nu$ . Let us refer to this as the 'canonical-flux' ensemble. So long as the trajectory duration ('system size') is much greater than any correlation time ('length'), the properties at instant ('locality')  $\Delta t$  are unaffected by whether integrated flux ('energy') is exactly conserved, and the ensembles are equivalent in the infinite-time ('thermodynamic') limit.

It is possible to derive equation (17) (via equations (7) and (9)) by direct maximization of the information entropy of a set of trajectories, at fixed ensemble-averaged flux and energy. In that case, as with the above derivation, great care is required to compare the relevant time scales with correlation times, to avoid unwittingly averaging over the correlations present in  $p_\tau^{\text{eq}}(J|a \rightarrow b)$ . That would produce a mean-field expression, in which the rate of each transition is simply boosted exponentially according to its immediate flux contribution. Such a scheme is popular in simple models, but should not be mistaken for the exact theorem derived above.

### 3.3. Factors affecting transition rates

The expression for transition rates, equation (17), appears to depend on the arbitrary quantity  $\tau$ . It can be re-written in an a much clearer form that is explicitly independent of  $\tau$ , as we now show.

Although  $m_c(\nu, \tau) \rightarrow \infty$  as  $\tau \rightarrow \infty$ , the difference  $m_b - m_a$ , for two states  $a$  and  $b$ , has a finite asymptote, embodying the different transient influences that the two states have on the system, before it returns to a statistically steady state. So, let us define a function that contains that transient information, but is independent of the arbitrary quantity  $\tau$ , thus:

$$q_a(\nu) \equiv \lim_{\tau \rightarrow \infty} \{m_a(\nu, \tau) - \tau Q(\nu)\} \quad (21)$$

$$= \ln \lim_{\tau \rightarrow \infty} \frac{e^{m_a(\nu, \tau)}}{\langle e^{m(\nu, \tau)} \rangle_{\text{dr}}} \quad (22)$$

$$= \ln \lim_{\tau \rightarrow \infty} \frac{\int p_\tau^{\text{eq}}(J|a) e^{\tau \nu J} dJ}{\int \langle p_\tau^{\text{eq}}(J) \rangle_{\text{dr}} e^{\tau \nu J} dJ} \quad (23)$$

so that  $q_a - q_b = m_a - m_b$  in the long-time limit.

We require one further piece of notation. The dynamics is described by a set of transitions  $a \rightarrow b$  carrying integrated flux  $J_{ab} \Delta t$ . For continuous dynamics,  $J_{ab} \Delta t \rightarrow 0$  as  $\Delta t \rightarrow 0$ , but for discrete transitions,  $J_{ab} \Delta t$  remains finite whether or not time steps are made vanishingly small. As above, the following discussion applies to either case.

In terms of these physically meaningful quantities, transition rates in the driven ensemble are given by

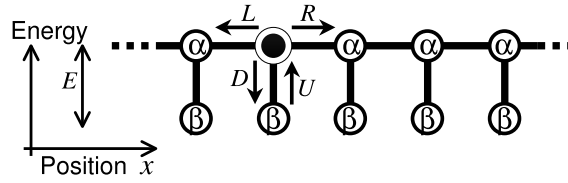
$$\omega_{a \rightarrow b}^{\text{dr}} = \omega_{a \rightarrow b}^{\text{eq}} \exp[v J_{ab} \Delta t + q_b(v) - q_a(v) - Q(v) \Delta t]. \quad (24)$$

The derivation of equation (24) from equation (17) is given in appendix B. It is now clear, in equation (24), that three distinct factors determine the rate of a transition  $a \rightarrow b$  in the driven steady-state ensemble. (1) The rate is proportional to the rate at equilibrium. So, all else being equal, energetically expensive transitions are slow, while down-hill transitions take precedence. (2) The rate is exponentially enhanced for transitions that contribute a favourable flux. (3) The dependence on  $q_b - q_a$  is overlooked by mean-field models. It says that a transition's likelihood depends also on the state in which it leaves the system. Its rate is enhanced if it puts the system into a state that is more likely to exhibit flux in the future. The effect of this factor on the driven steady-state distribution of microstates is to increase (relative to the Boltzmann distribution) the weight of states that are more-than-averagely willing to accept a future flux. We shall see an example of this effect in the following section. In the case of a shear flux, this means that low-viscosity states are favoured, as is often observed.

#### 4. Applications

We have a recipe for constructing a model of any given driven system, which is guaranteed to yield the desired flux, and to respect all the physical laws that are obeyed by the equilibrium version of the model, and that is guaranteed to have no artefacts from statistical bias. If we choose to provide this machinery with an equilibrium model that obeys all of Newton's laws—i.e., a fluid whose internal interactions conserve momentum, angular momentum, and energy, while stochastic forces from the reservoir couple only to particles at the boundary—then it will yield dynamical rules that also respect Newton's laws for the boundary-driven fluid. In other words, the method has the capacity to produce exact physics if provided with an exactly physical prior. It provides a description of the reservoir, by characterizing the stochastic part of the equations of motion. The way in which that reservoir couples to the system is up to the user to decide. In the above example, it is coupled only at the boundary, but we may instead consider a driven Brownian system, for which the heat bath is more strongly and uniformly coupled, dominating all momentum variables. Another alternative is to apply the method to a model whose prior (equilibrium) physical properties are simplified for the sake of clarity and analytical expediency. In that case, of course, the result of the recipe will be approximate and unreliable, but still the least arbitrary choice of transition rates for the given degree of simplification.

The micro-canonical flux ensemble introduced in sections 2 and 3.1 was first presented in [7], where it was used analytically to construct a continuum model of driven diffusion, and heuristically to discuss the features of a lattice model of dimers under shear. The latter model had a much more complex energy landscape including jammed states. Another analytically solvable model was constructed in [8], using the micro-canonical flux method. It was another one-dimensional driven diffusion model, but this time with a discrete state space and discrete time step. In the following section, we shall analytically construct a model of a driven system with a non-trivial energy landscape that demonstrates some features of more complex systems, such as sheared complex fluids, with states that are locally trapped so that they cannot easily be



**Figure 2.** Comb-shaped state-space for the hopping model. Circles represent the possible states of the system, of types  $\alpha$  and  $\beta$ . The single particle, shown as a filled circle, currently occupies an  $\alpha$  state. It may escape into other  $\alpha$  states to the left or right with rates  $L$  and  $R$  respectively, or downwards in energy, with rate  $D$ , to state  $\beta$ . Each  $\beta$  state can be exited with rate  $U$  upwards to the  $\alpha$  state at the same location only.

driven. The model reduces to simple one-dimensional driven diffusion in a certain limit, and has a discrete state space but continuous time, to complement the earlier published models. We shall use the canonical-flux ensemble of sections 3.2 and 3.3, to demonstrate the utility of this method.

#### 4.1. The model

Consider a particle that can hop stochastically among a set of discrete states that have the connectivity shown in figure 2. The particle will be driven by its non-equilibrium heat bath so that it has, on average, a drift velocity  $u$  from left to right. At each location  $x$ , it may occupy one of two states: state  $\alpha$ , from which it may escape to the left or right with rates  $L$  and  $R$  to other  $\alpha$  states at different locations  $x$ , or downwards with rate  $D$  into a lower energy trapped state  $\beta$ ; once in a  $\beta$  state, the particle cannot exhibit any flux, i.e., cannot move left or right, but can only wait for a random excitation at rate  $U$  back up to the  $\alpha$  state at the same location.

Note that, if we set  $D = 0$ , the model reduces to a continuous-time discrete-space linear hopping model, like the versions that were previously studied with both space and time continuous [7] or discrete [8].

The equilibrium version of this model (with no mean drift) is very straightforward. Detailed balance requires that  $U^{\text{eq}} = D^{\text{eq}} \exp(-E)$ , where  $E$  is the energy difference between states  $\alpha$  and  $\beta$  measured in units of  $k_B T$ , and that  $R^{\text{eq}} = L^{\text{eq}} \equiv \omega_0$  where we may measure all rates in units of  $\omega_0$  so that  $\omega_0 \equiv 1$  without loss of generality. The occupancy of  $\alpha$  states is given by Boltzmann as  $1/(1 + \exp(E))$ , and the only remaining parameter that we are free to choose is  $D^{\text{eq}} \equiv \rho$ , which specifies the ratio of vertical to horizontal mobilities.

When the model is not at equilibrium, but is driven at drift velocity  $u$ , the naïve expectation would be either that we are free to choose all four rates  $U, D, L, R$ , since non-equilibrium models traditionally have no rules, or that detailed balance still governs the ratio  $U/D$ . However, as discussed above, neither of these statements is true. There is, in fact, a least-arbitrary set of rates that corresponds to driving by an uncorrelated heat bath that is characterized only by its temperature and velocity. We shall now calculate that set of rates, using the canonical-flux ensemble.

The rates are given by equation (24). Defining our unit of length to be one inter-site spacing, the integrated flux of a transition to the right (left) is  $J_R \Delta t = 1$ , ( $J_L \Delta t = -1$ ), while transitions between  $\alpha$  and  $\beta$  states carry no flux as they leave the particle's displacement unchanged. Since time is continuous, the time-step is infinitesimal,  $\Delta t \rightarrow 0$ , so that the last term in the exponential of equation (24) vanishes, and it prescribes the following rates in the driven ensemble:

$$R = e^\nu \quad (25a)$$

$$L = e^{-\nu} \quad (25b)$$

$$D = \rho e^{q_\beta - q_\alpha} \quad (25c)$$

$$U = U^{\text{eq}} e^{q_\alpha - q_\beta} = \rho e^{-E + q_\alpha - q_\beta}. \quad (25d)$$

To complete the calculation of the rates, we require only  $q_\alpha - q_\beta$ , the difference in the willingness of the two states to admit a flux. This could be evaluated by ‘brute force’ using equation (23), if we first calculate the Green function for the equilibrium model, i.e. the probability that the particle travels a given distance in a given time, given the initial state  $\alpha$  or  $\beta$ . However, that calculation can be avoided, using the derivation in appendix C to show that, for this ‘comb’ model,

$$q_\alpha(\nu) - q_\beta(\nu) = \ln(1 + Q(\nu)/U^{\text{eq}}). \quad (26)$$

This is purely a result of the facts that state  $\beta$  can only be quit via state  $\alpha$ , and that escape times are distributed exponentially.

We can now construct a differential equation for  $Q(\nu)$ , as follows. Due to the model’s translational symmetry, the steady-state occupancy of  $\alpha$  states is just

$$f_\alpha = \frac{U}{U + D} \quad (27)$$

and, since displacements are allowed only from  $\alpha$  states, the mean drift velocity is

$$u = (R - L)f_\alpha = 2f_\alpha \sinh \nu. \quad (28)$$

Now, using equation (18), we obtain an ordinary differential equation,

$$\frac{dQ}{d\nu} = \frac{2(\rho e^{-E} + Q)^2 \sinh \nu}{\rho^2 e^{-E} + (\rho e^{-E} + Q)^2}$$

that can be integrated for  $\cosh \nu$  as a function of  $Q$ . The constant of integration is fixed by  $Q(0) = 0$  which follows from normalization of the probability distribution in the definition of  $Q$  (equations (19) and (20)). Finally, we obtain the required ‘potential’,

$$Q(\nu) = \cosh(\nu) - 1 - \frac{\rho}{2}(1 + e^{-E}) + \sqrt{\left[\cosh(\nu) - 1 - \frac{\rho}{2}(1 - e^{-E})\right]^2 + \rho^2 e^{-E}}$$

which, with equations (25), (26), (27) and (28), leads to four constraints on the four transition rates in the driven system, from which the abstract quantities  $\nu$  and  $Q$  have been eliminated:

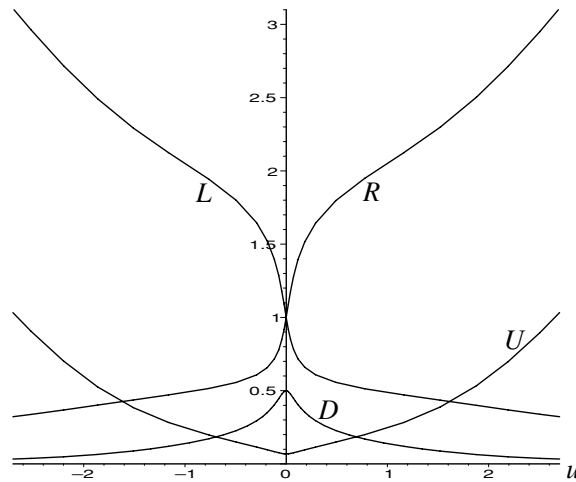
$$R - L = (1 + D/U)u \quad (29a)$$

$$RL = 1 \quad (29b)$$

$$UD = \rho^2 e^{-E} \quad (29c)$$

$$R + L + D - U = 2 + (1 - e^{-E})\rho \quad (29d)$$

One of these four equations is obvious; the others are not. Equation (29a) is simply a re-statement of equation (28), and gives the drift velocity  $u$  that results from any choice of the four transition rates. So, if we applied the usual ad hoc construction of non-equilibrium stochastic models, we would pluck four rates out of the air, use equation (29a) to find the resulting drift velocity, and have no other constraints. The other three constraints have arisen from our demand that the design of the model incorporates the prior dynamics, the large-scale flux, and no other design features.



**Figure 3.** Rates  $R$ ,  $L$ ,  $U$  and  $D$  for the driven comb model, as functions of the drift velocity  $u$ , using the parameters  $E = 2$ ,  $\rho = 0.5$ .

Note that equations (29b) and (29c) express relations between forward and reverse transition rates that are generic to any continuous-time model with instantaneous transitions, that *the product of the forward and reverse rates of a transition is equal in the driven and equilibrium ensembles*. This follows directly from equation (24) with  $J_{ab}\Delta t$  finite as  $\Delta t \rightarrow 0$ .

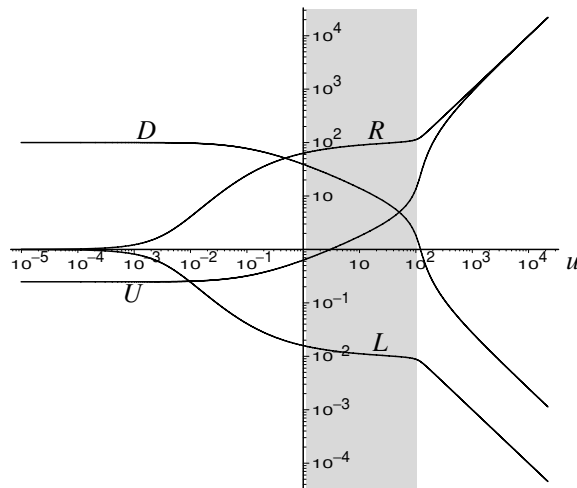
The four transition rates defined by equations (29) have exactly the same number of free parameters as in an equilibrium model: for a given energy gap  $E$  and flux  $u$ , the four rates are defined up to one parameter,  $\rho$ , that specifies the prior ratio of vertical to horizontal mobilities, as is the case in the equilibrium version of the model that was required to respect detailed balance. While the equilibrium occupancy, given by Boltzmann's law, is independent of  $\rho$ , the occupancy in the driven ensemble (equation (27)) does depend on this kinetic parameter.

#### 4.2. Properties of the model

The transition rates prescribed by sub-ensemble dynamics are plotted as functions of velocity  $u$  in figure 3 for an energy gap  $E = 2$  and mobility ratio  $\rho = 0.5$ . Due to the symmetries of the comb structure, the rates of transitions up and down ( $U$ ,  $D$ ) between  $\alpha$  and  $\beta$  states are even functions of  $u$ . At  $u = 0$ , the rates take their equilibrium values,  $R = L \equiv 1$  and  $D = U \exp E$ . On increasing velocity, hops to the right ( $R$ ) become more frequent, while hops to the left ( $L$ ) are suppressed, as expected. Also the particle becomes less likely to fall down ( $D$ ) into a trapped  $\beta$  state, and is increasingly dragged out of traps ( $U$ ) by the driving force.

The rates are re-plotted on log–log axes (for positive  $u$ ) in figure 4, using parameter values  $E = 6$ ,  $\rho = 100$ , that were chosen to provide a separation of time scales, emphasizing the features of the graphs. Three regimes of drift velocity  $u$  become apparent. On the left of the figure (low  $u$ ) is the near-equilibrium regime, where the rates  $D$ ,  $U$ , of transitions that do not carry a flux, remain approximately constant, respecting detailed balance. This fulfils the naive expectation, often applied to non-equilibrium models, that detailed balance continues to describe the physics of activated processes. Meanwhile, the rate of hops to the right,  $R$ , is enhanced and to the left,  $L$ , is suppressed, so that the sparsely populated  $\alpha$  states exhibit the required drift velocity.





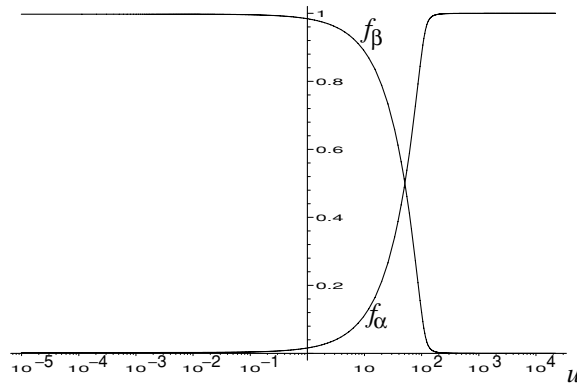
**Figure 4.** The transition rates of the driven comb model, shown on log–log axes as functions of the positive drift velocity  $u$ . Parameter values are  $E = 6$ ,  $\rho = 100$ . Three regimes of behaviour are visible. The regime at intermediate values of  $u$  is shaded.

The second regime of the driving velocity,  $\omega_0 < u < \rho$  is shaded grey in figure 4. In this regime, the flux constraint can no longer be satisfied by the small population of thermally-activated  $\alpha$  states. The states become mechanically activated, with particles in the immobile  $\beta$  state promoted into the mobile state by the driving force. As  $u$  increases through the shaded part of the figure, the unequal hopping rates to the right and left remain approximately constant, while the rate of activation  $U$  increases and rate of trapping  $D$  decreases, so that the drifting  $\alpha$  population increases. This is also apparent in figure 5, which shows the occupancies of the two states as a function of velocity, for this same set of parameters.

Once the mobile states are fully populated, and the trapped states have negligible occupancy, the bias on hops to the right can no longer remain constant while satisfying an increasing flux constraint. Hence, a third regime exists at the highest values of  $u$  (figure 4), where rates  $R$  and  $U$  both become proportional to the flux  $u$ , while the flux-impeding transitions have rates  $L$  and  $D$  inversely proportional to  $u$ .

#### 4.3. Comparison with shear flow

The model we have studied here, with its simple comb-shaped state space, has some features that are generic to driven systems. We saw, in section 3.3, that the rate of a transition in a driven ensemble depends on three factors: its rate at equilibrium, the amount of flux that it contributes, and the difference in the willingness of the initial and final states to allow the required flux in the future. Transitions that contribute a nonzero amount of flux were called ‘type A’ in [7], while transitions between states with different promise for future flux were labelled ‘type B’. In previous articles [7, 8], the rates prescribed by sub-ensemble dynamics were calculated explicitly only for simple models that exhibited only type A transitions due to the simplicity of their state spaces. The comb model, on the other hand, has both type A ( $\alpha \rightarrow \alpha$ ) and type B ( $\alpha \rightleftharpoons \beta$ ) transitions. Another example of such a model, that was previously discussed only heuristically [7, 8], is a set of dimers (particles that occupy two adjacent lattice sites) that perform random walks on a two-dimensional triangular lattice, while the lattice itself is driven into shear flow by sporadically cleaving and re-positioning



**Figure 5.** Occupancies of  $\alpha$  and  $\beta$  states for the same parameters as in figure 4, shown as functions of the drift velocity, on semi-logarithmic axes. Three zones of behaviour are again visible: Boltzmann-distributed states at low  $u$ ; re-population of states at intermediate  $u$ ; the mobile  $\alpha$  state fully occupied at high  $u$ .

its horizontal layers. Certain arrangements of the dimers (analogous to  $\alpha$  states of the comb model) allow these quanta of shear, while other states (analogous to  $\beta$  states) are prevented from shearing, due to the unbreakable dimers straddling two layers of the lattice, thus jamming the system.

Although any such many-particle system has a very complex state-space, its crucial features are reproduced in the comb model. When the comb model is stuck in a  $\beta$  state, the driving force (that derives from the statistics of the driven ensemble) pushes it into a more mobile state in order to flow. Likewise, when the dimer model is in a state that will not admit a flux, it must first re-arrange its particles. The driving force achieves this by imposing a shear stress on the particles, causing them to re-orient mechanically (as opposed to thermally, by Brownian motion). The sub-ensemble rules prescribe (for a given prior dynamics) the rate of that mechanically imposed re-alignment, thereby specifying the constraints that must be met by a physically acceptable constitutive relation for the flowing system.

## 5. Summary and outlook

There are certain constraints that must be satisfied by any candidate for a statistical mechanical theory of driven steady states: it must satisfy the known laws of motion, and it must give rise to the required macroscopic observables (flux, energy etc). In this paper, we have assumed; indeed, demanded; that those are the *only* constraints, and derived the transition rates implied by that assumption. Comparison with experimental observations will determine *a posteriori* whether a particular system belongs to the ergodic class that is well described by these unbiased rates, just as empirical comparison determines whether or not a static system is at thermodynamic equilibrium. If one is privy to prior information indicating that the driven system's motion is biased in some way that is not apparent in its macroscopic flux and conserved quantities, then the dynamical rules set out here should be disregarded. To violate the rules *a priori* without such a justification is to bias the model with arbitrary information derived from prejudice rather than from physics. Such arbitrariness is not condoned for equilibrium models, and the same should be the case for macroscopically driven steady states.

For example, consider how we design a stochastic model of an equilibrium system. The system is defined by some set of available states, and we must choose the rates of transitions

between those states. Canonical equilibrium is defined by a fixed volume, particle number, and mean energy of the system. We might choose any arbitrary set of rates, and then measure or calculate the mean energy that results when the system arrives at a steady state. Certainly, that procedure would give rise to a well-defined mean energy, volume and number, but that is not sufficient for us to say that the system is at equilibrium and that the transition rates are acceptable. There are constraints arising from the principle of detailed balance, which ensure that  $(E, N, V)$  are the *only* parameters characterizing the macroscopic state of the ensemble, beyond the definition of the system in terms of its accessible states and reversibility. We have found the generalization of those constraints to non-equilibrium steady states.

The prior is central to the formalism, and is often misinterpreted in non-equilibrium applications of information theory. In the present context, it is used to mean the complete set of *physically valid* paths that a system might take in response to the stochastic forces arising from a particular coupling to a non-equilibrium reservoir. If the reservoir can exchange energy with the system, then conservation of energy can be violated in the prior. If the coupling is only to particles at the system's boundary, then energy, momentum and angular momentum must be conserved by all internal interactions, so the prior does not include scenarios for which those laws are violated.

This has not been the usual definition of the prior, in previous attempts at non-equilibrium applications of information theory. It is often assumed that our knowledge of microscopic dynamics can be discarded, and MaxEnt will correctly reconstruct that missing information. Such optimism cannot be justified. For instance, MaxEnt has been used to choose between phase-space paths that are characterized by their actions [14], discarding our knowledge of Hamilton's principle of least action. The result is an exponential distribution in which the paths of least action are the most likely, but that result is incorrect. Paths on which the action is extremized are not just *likely*; they are the *only* paths of a classical system, and therefore the only paths that should appear in the prior if an exact calculation is wanted.

The central results of this paper are the formulae for transition rates in a driven ensemble. These are formulated in two alternative ways. In the 'microcanonical-flux ensemble', the flux is constrained to an exact value when time-averaged over the duration (tending to infinity) of each system's passage through phase-space, resulting in equation (12). The 'canonical-flux ensemble', in which only the ensemble-averaged flux is constrained, leads to equation (24) for the transition rates, which is exactly equivalent to the microcanonical-flux prescription. The canonical-flux equation (24) makes explicit the three factors influencing a transition rate. As at equilibrium, energetics are important, making a system reluctant to take up-hill steps in its energy landscape. Secondly, an exponential factor, that one might have guessed, favours transitions that impart the desired flux. The third factor prescribed by equation (24) is more subtle. It describes the importance of correlations, and depends on a well-defined quantity ascribed to each microstate, which quantifies its promise for future flux. A transition is favoured if it takes the system to a state of higher promise that is more amenable to future flux-carrying transitions.

The sub-ensemble scheme has previously been demonstrated to produce the standard equations of motion for diffusion with drift, both for continuous [7] and discrete [8] random walks. In the present paper, the dynamical rules were evaluated for a more complex model. We have seen that, for thermally activated processes that are governed by detailed balance at equilibrium, the sub-ensemble rules describe *mechanical* activation by the driving force, although detailed balance is recovered in the low-flux regime.

In the context of shear flow, mechanical activation corresponds to stress-induced rearrangement. The fact that this statistical formalism describes the effects of non-equilibrium

stresses in a natural way, makes it a promising approach for the study of shear-banding, jamming, and other shear-induced transitions of complex fluids.

At the risk of repetition, we have a recipe for constructing a model of any given driven system, which is guaranteed to yield the desired flux, and to respect all the physical laws that are obeyed by the equilibrium version of the model. It is also guaranteed to have no artefacts from statistical bias. This machinery has the capacity to produce exact physics if provided with an exactly physical prior. Otherwise, it will yield the least arbitrary model for the given degree of approximation.

### Acknowledgments

Many thanks go to Alistair Bruce, Michael Cates, Richard Blythe, Peter Olmsted, Tom McLeish, Alexei Likhtman, Suzanne Fielding and Hal Tasaki for informative discussions. RMLE is grateful to the Royal Society for support.

### Appendix A. The canonical-flux potentials

As stated in equation (16), the distribution of flux  $\hat{J}$  during interval  $\hat{\tau}$ , that appears in equation (14), is uncorrelated with the initial state  $a$ , and can therefore be written

$$p_{\hat{\tau}}(\hat{J}|a, J) \rightarrow f_{\hat{\tau}}(\hat{J}) \equiv \sum_c p^{\text{dr}}(c) p_{\hat{\tau}}^{\text{eq}}(\hat{J}|c) \quad (\text{A.1})$$

in terms of the instantaneous steady-state distribution of states  $p^{\text{dr}}(c)$ . The distribution  $f_{\hat{\tau}}(\hat{J})$  can be evaluated if we sub-divide the interval  $\hat{\tau}$  into  $n$  sub-intervals of duration  $\tau$ , where  $n \gg 1$  since  $\hat{\tau} \sim \tau_0 \gg \tau$ . The system begins each of these sub-intervals in a state drawn randomly and independently from the steady-state distribution  $p^{\text{dr}}(c)$ . These initial states are uncorrelated because  $\tau \gg \tau_{\text{cor}}$ . The overall flux in the interval  $\hat{\tau}$  is the mean of the fluxes in these  $n$  independent sub-intervals, so that

$$f_{\hat{\tau}}(\hat{J}) = \int_{-\infty}^{\infty} dJ_1 \cdots dJ_n f_{\tau}(J_1) \cdots f_{\tau}(J_n) \delta\left(\hat{J} - \frac{1}{n} \sum_i^n J_i\right). \quad (\text{A.2})$$

This limit distribution gives the likelihood (under equilibrium dynamics, with a non-equilibrium initial state) that the  $n$  independent flux measurements have an improbably-large mean value  $\hat{J}$ . Cam er's theorem of large deviations [13] states that the weight in the tail of the distribution of the mean of  $n$  independent identically distributed random variables behaves as

$$\lim_{n \rightarrow \infty} \frac{1}{n} \ln \int_j^{\infty} f_{n\tau}(J') dJ' = -I(\hat{J}, \tau). \quad (\text{A.3})$$

That is, the weight in the tail decays exponentially with  $n$ , at a rate  $I$  given [13] by

$$I(J, \tau) = \sup_{\theta} \left[ \theta J - \ln \int_{-\infty}^{\infty} f_{\tau}(J') e^{\theta J'} dJ' \right]. \quad (\text{A.4})$$

Dividing both sides of equation (A.3) by the constant  $\tau$  gives

$$\lim_{\hat{\tau}/\tau \rightarrow \infty} \frac{1}{\hat{\tau}} \ln \int_j^{\infty} f_{\hat{\tau}}(J') dJ' = \frac{-I(\hat{J}, \tau)}{\tau}. \quad (\text{A.5})$$

Since the LHS of equation (A.5) is independent of the arbitrary choice of  $\tau$ , we can infer that  $I \propto \tau$ . Let us define the function

$$H(J) \equiv I(J, \tau)/\tau \quad (\text{A.6})$$

that is independent of the arbitrary quantity  $\tau$ . Writing the exponential decay law explicitly, with an unknown prefactor  $A(\hat{J})$  that varies only slowly with  $\hat{\tau}$ ,

$$\int_j^\infty f_{\hat{\tau}}(J') dJ' \rightarrow A(\hat{J}) \exp[-\hat{\tau} H(\hat{J})] \quad \text{as } \tau/\hat{\tau} \rightarrow 0$$

and differentiating with respect to  $\hat{J}$  gives

$$f_{\hat{\tau}}(\hat{J}) \rightarrow [A(\hat{J})\hat{\tau}H'(\hat{J}) - A'(\hat{J})] \exp[-\hat{\tau}H(\hat{J})]. \quad (\text{A.7})$$

Now, substituting for  $\hat{J}$  from equation (15) and Taylor-expanding  $H(\hat{J})$  to first order in  $\tau/\hat{\tau}$  allows us to take the limit  $\hat{\tau}/\tau \rightarrow \infty$  when substituting equations (14), (A.1) and (A.7) into (12), yielding

$$\frac{\omega_{a \rightarrow b}^{\text{dr}}}{\omega_{a \rightarrow b}^{\text{eq}}} = \lim_{\tau/\tau_{\text{cor}} \rightarrow \infty} \frac{\int_{-\infty}^{\infty} dJ p_{\tau}^{\text{eq}}(J|a \rightarrow b) e^{\tau H'(J_0)J}}{\int_{-\infty}^{\infty} dJ p_{\tau}^{\text{eq}}(J|a) e^{\tau H'(J_0)J}}. \quad (\text{A.8})$$

The supremum in equation (A.4) can be evaluated by defining the functions in equations (19) and (20). From equations (A.1), (A.4) and (A.6), we have

$$H(J) = \nu J - Q(\nu) \quad (\text{A.9})$$

with the parameter  $\nu(J)$  [equal to  $\theta/\tau$  in equation (A.4)] given by equation (18). Thus the parameter  $H'(J_0)$  in equation (A.8) can be evaluated by differentiating equation (A.9) and substituting from equation (18), to give

$$\frac{\partial H}{\partial J} = \nu \quad (\text{A.10})$$

resulting in equation (17). Note that  $Q$  is a Legendre transform of  $H$ , and that  $\nu$  and  $J$  in equations (A.10) and (18) are conjugate variables.

## Appendix B. Re-formulation of the canonical-flux expression for transition rates

Let us make a change of variable in equation (17), and replace the integration over *average* flux  $J$  by one over *total* (integrated) flux  $K \equiv \tau J$ . Then, using  $p_{\tau}^{\text{eq}}(K|\cdot\cdot\cdot) dK$  now to represent the normalized probability of finding a total flux  $K$  on an equilibrium trajectory of length  $\tau$ , we can write

$$\frac{\omega_{a \rightarrow b}^{\text{dr}}}{\omega_{a \rightarrow b}^{\text{eq}}} = \lim_{\tau/\tau_{\text{cor}} \rightarrow \infty} \frac{\int dK p_{\tau}^{\text{eq}}(K|a \rightarrow b) e^{\nu K}}{\int dK p_{\tau}^{\text{eq}}(K|a) e^{\nu K}}. \quad (\text{B.1})$$

Now, the expression  $p_{\tau}^{\text{eq}}(K|a \rightarrow b)$  is the probability of accumulating an integrated flux  $K$  during interval  $\tau$ , given that the initial part  $\Delta t$  of that interval is taken up with a transition from state  $a$  to  $b$ . Since that transition carries an integrated flux  $K_{ab} \equiv J_{ab} \Delta t$ , we can replace the expression by the probability of accumulating the remaining flux  $K - K_{ab}$  in the remaining time, starting from state  $b$ , i.e.

$$p_{\tau}^{\text{eq}}(K|a \rightarrow b) = p_{\tau - \Delta t}^{\text{eq}}(K - K_{ab}|b).$$

Hence, after a change of variable, equation (B.1) gives

$$\ln \frac{\omega_{a \rightarrow b}^{\text{dr}}}{\omega_{a \rightarrow b}^{\text{eq}}} = \nu K_{ab} + \lim_{\tau \rightarrow \infty} [m_b(\nu, \tau - \Delta t) - m_a(\nu, \tau)] \quad (\text{B.2})$$

$$= \nu K_{ab} + \lim_{\tau \rightarrow \infty} [m_b(\nu, \tau) - m_a(\nu, \tau)] - \zeta_b(\nu, \Delta t) \quad (\text{B.3})$$

where

$$\zeta_b(v, \Delta t) \equiv \lim_{\tau \rightarrow \infty} [m_b(v, \tau) - m_b(v, \tau - \Delta t)] \quad (\text{B.4})$$

and, by substituting  $\tau \rightarrow \tau + \Delta t$  into equation (B.2), we find  $\zeta_a = \zeta_b \forall a, b$ , i.e.,  $\zeta_b$  is state-independent. Given that the limit in equation (B.4) exists, we can write

$$\zeta_b(v, \Delta t) = \Delta t \lim_{\tau \rightarrow \infty} \left( \frac{\partial m_b}{\partial \tau} \right)_v \quad (\text{B.5})$$

even for finite  $\Delta t$ , since  $m_b$  asymptotes to a linear function of  $\tau$ . The state-independence of  $\zeta_b$  can now be used to factor out the time-derivate of  $m$  from the ensemble average when differentiating equation (19) with respect to  $\tau$ , yielding

$$\zeta_b(v, \Delta t) = Q(v) \Delta t \quad \forall b. \quad (\text{B.6})$$

Finally, substituting equation (B.6) into (B.3) gives a very simple expression for the ratio of transition rates,

$$\ln \frac{\omega_{a \rightarrow b}^{\text{dr}}}{\omega_{a \rightarrow b}^{\text{eq}}} = v K_{ab} - Q(v) \Delta t + \lim_{\tau \rightarrow \infty} [m_b(v, \tau) - m_a(v, \tau)] \quad (\text{B.7})$$

from which equation (24) follows.

### Appendix C. Calculation for continuous-time hopping on a comb

For the discrete states of the comb model of section 4.1, with the integrated flux  $J \Delta t$  quantized into discrete values of the displacement  $x$ , equation (20) becomes

$$m_\beta(v, \tau) = \ln \sum_{x=-\infty}^{\infty} G_\beta(x, \tau) e^{vx}$$

where  $G_\beta(x, \tau)$  is the equilibrium Green function for  $\beta$  states. That is the probability of attaining a displacement  $x$  in time  $\tau$  given that the particle initially occupies a  $\beta$  state. An equivalent expression holds for  $m_\alpha$ .

In the continuous-time model, a particle occupying state  $\beta$  at time 0 will escape to the corresponding  $\alpha$  state at a time  $t$  that is drawn stochastically from the exponential probability distribution  $p_{\beta\alpha}(t) = U^{\text{eq}} \exp(-U^{\text{eq}}t)$ . Once excited to the  $\alpha$  state, the particle is governed by the corresponding Green function  $G_\alpha(x, \tau)$ , so that the Green function for a particle occupying state  $\beta$  is given by

$$\begin{aligned} G_\beta(x, \tau) &= \int_0^\tau p_{\beta\alpha}(\tau - t) G_\alpha(x, t) dt \\ &= U^{\text{eq}} e^{-U^{\text{eq}}\tau} \int_0^\tau e^{U^{\text{eq}}t} G_\alpha(x, t) dt \end{aligned}$$

from which it follows that

$$\exp(U^{\text{eq}}\tau + m_\beta(v, \tau)) = U^{\text{eq}} \int_0^\tau \exp(U^{\text{eq}}t + m_\alpha(v, t)) dt.$$

Differentiating with respect to  $\tau$  yields

$$U^{\text{eq}} + \frac{\partial m_\beta(v, \tau)}{\partial \tau} = U^{\text{eq}} \exp(m_\alpha(v, \tau) - m_\beta(v, \tau)).$$

In the limit of large  $\tau$ , the time derivative of  $m_\beta$  is just  $Q$ , as given by equation (B.6), so that, with the definition of  $q_\alpha$  in equation (21), the required result, equation (26) follows.

## References

- [1] Jaynes E T 1957 *Phys. Rev. Lett.* **106** 620  
Jaynes E T 1957 *Phys. Rev. Lett.* **108** 171
- [2] Shannon C E 1948 *Bell Syst. Tech. J.* **27** 379  
Shannon C E 1948 *Bell Syst. Tech. J.* **27** 623
- [3] Jaynes E T 1979 *Maximum Entropy Formalism* ed R D Levine and M Tribus (Cambridge, MA: MIT Press)
- [4] Dewar R 2003 *J. Phys. A: Math. Gen.* **36** 631 and references therein
- [5] Maes C 1999 *J. Stat. Phys.* **95** 367  
Evans D J 2004 *Preprint cond-mat/0408195*
- [6] Hyeon-Deuk K and Hayakawa H 2003 *J. Phys. Soc. Japan* **72** 2437
- [7] Evans R M L 2004 *Phys. Rev. Lett.* **92** 150601
- [8] Evans R M L 2004 *Physica A* **340** 364
- [9] Götze W 1989 *Liquids Freezing and Glass Transition (Les Houches Session LI)* ed J-P Hansen, D Levesque and J Zinn-Justin (Amsterdam: Elsevier)  
Bouchaud J-P 1992 *J. Phys. I France* **2** 1705
- [10] Wittmer J P, Claudin P, Cates M E and Bouchaud J-P 1996 *Nature* **382** 336
- [11] Cates M E 2000 *Soft and Fragile Matter* ed M E Cates and M R Evans (Bristol: Institute of Physics Publishing)  
Silbert L E, Farr R S, Melrose J R and Ball R C 1999 *J. Chem. Phys.* **111** 4780  
Cates M E, Wittmer J P, Bouchaud J-P and Claudin P 1998 *Phys. Rev. Lett.* **81** 1841
- [12] Ernst M H and Bussemaker H J 1995 *J. Stat. Phys.* **81** 515  
Gardner M 1970 *Sci. Am.* **223** 120  
Bagnoli F, Rechtman R and Ruffo S 1991 *Physica A* **171** 249
- [13] Varadhan S R S 1984 *Large Deviations and Applications* (Philadelphia, PA: SIAM)
- [14] Wang Q A 2004 *Preprint cond-mat/0312329*
- [15] Evans D J, Cohen E G D and Morriss G P 1993 *Phys. Rev. Lett.* **71** 2401  
Evans D J, Cohen E G D and Morriss G P 1993 *Phys. Rev. Lett.* **71** 3616  
Evans D J and Morriss G P 1990 *Statistical Mechanics of Nonequilibrium Liquids* (London: Academic)