



PAPER

OPEN ACCESS

RECEIVED
14 June 2021REVISED
29 July 2021ACCEPTED FOR PUBLICATION
9 August 2021PUBLISHED
2 September 2021

Original content from
this work may be used
under the terms of the
[Creative Commons
Attribution 4.0 licence](#).

Any further distribution
of this work must
maintain attribution to
the author(s) and the
title of the work, journal
citation and DOI.



Linear response in large deviations theory: a method to compute non-equilibrium distributions

Nahuel Freitas*, Gianmaria Falasco and Massimiliano Esposito

Complex Systems and Statistical Mechanics, Department of Physics and Materials Science, University of Luxembourg, L-1511
Luxembourg, Luxembourg

* Author to whom any correspondence should be addressed.

E-mail: nahuel.freitas@uni.lu**Keywords:** large deviations theory, linear response, stochastic thermodynamics, non-equilibrium steady states

Abstract

We consider thermodynamically consistent autonomous Markov jump processes displaying a macroscopic limit in which the logarithm of the probability distribution is proportional to a scale-independent rate function (i.e. a large deviations principle is satisfied). In order to provide an explicit expression for the probability distribution valid away from equilibrium, we propose a linear response theory performed at the level of the rate function. We show that the first order non-equilibrium contribution to the steady state rate function, $g(\mathbf{x})$, satisfies $\mathbf{u}(\mathbf{x}) \cdot \nabla g(\mathbf{x}) = -\beta \dot{W}(\mathbf{x})$ where the vector field $\mathbf{u}(\mathbf{x})$ defines the macroscopic deterministic dynamics, and the scalar field $\dot{W}(\mathbf{x})$ equals the rate at which work is performed on the system in a given state \mathbf{x} . This equation provides a practical way to determine $g(\mathbf{x})$, significantly outperforms standard linear response theory applied at the level of the probability distribution, and approximates the rate function surprisingly well in some far-from-equilibrium conditions. The method applies to a wealth of physical and chemical systems, that we exemplify by two analytically tractable models—an electrical circuit and an autocatalytic chemical reaction network—both undergoing a non-equilibrium transition from a monostable phase to a bistable phase. Our approach can be easily generalized to transient probabilities and non-autonomous dynamics. Moreover, its recursive application generates a virtual flow in the probability space which allows to determine the steady state rate function arbitrarily far from equilibrium.

1. Introduction

Equilibrium statistical mechanics provides the probability distribution over the states of a system in terms of their energy. Finding an explicit form that generalizes the Gibbs distribution to non-equilibrium stationary conditions has been the object of extensive research in the past and remains a central issue in non-equilibrium physics nowadays [1–11]. The efforts in this direction may be divided into two research lines. The first is based on expanding the non-equilibrium probability around the known equilibrium distribution. Close to equilibrium, the correction to the Gibbs distribution was first related to the dissipation of the system by McLennan [3]. Since the dissipation is an extensive quantity of the trajectory duration and is quadratic in the distance from equilibrium, more recent work [7] has been devoted to clarify which part of the dissipation enter the McLennan formula and the suitable ordering of the long time and near to equilibrium limits. Far from equilibrium, the stationary distribution has been obtained as formal series either based on all the cumulants of the (transient) entropy production or involving the excess dynamical activity induced by the non-equilibrium forces [8]. Even at linear order around equilibrium, the difficulty in obtaining explicit results lies in the calculations of conditional averages over the full set of possible trajectories [6, 12]. The second research line is based on Freidlin–Wentzell large deviations (LDs) theory [13], which states that the logarithm of the stationary distribution of systems with weak noise is proportional to a rate function, called the quasi-potential. Since explicit computations of the rate function are in general very difficult, expansions around a reference quasi-potential have been developed [14, 15].

However, mainly systems affected by Gaussian noise were considered [14, 15], and without leveraging the thermodynamic structure of the underlying dynamics.

In this paper we combine the previous two approaches by considering thermodynamically consistent stochastic dynamics displaying a macroscopic limit in which the probability distribution satisfies a LDs principle. Then, the steady state probability of a state \mathbf{x} can be approximated by $P_{ss}(\mathbf{x}) \propto \exp(-\Omega f(\mathbf{x}))$, for a sufficiently large scale parameter Ω , where $f(\mathbf{x})$ is the rate function or quasi-potential associated to the steady state. For a system in equilibrium at inverse temperature β , the usual Gibbs distribution indicates that $f(\mathbf{x})$ must be the free energy density $\phi(\mathbf{x})$ times β (assuming the free energy is extensive in Ω). Our first result is an explicit expression to compute the first order non-equilibrium correction $g(\mathbf{x})$ to the steady state rate function. We do so for systems subjected to Poissonian noise, i.e. microscopically described by a Markov jump process, thus complementing the LDs approach and providing a clear thermodynamic interpretation of the results. We show that $g(\mathbf{x})$ is fully determined by the deterministic dynamics, described by a drift vector field $\mathbf{u}(\mathbf{x})$, and the scalar field $\dot{W}(\mathbf{x})$ that indicates what is the rate at which work is performed on the system for a given state \mathbf{x} . This result offers a practical method to determine $g(\mathbf{x})$, which is applied to exactly solvable problems in electronics and chemistry, and is shown to be more accurate than usual linear response (LR) at the level of the probability distribution [16–21]. It can also be generalized to transient evolutions and non-autonomous settings. Second, we extend our formalism and derive a general LR formula for the quasi-potential around an arbitrary reference state. Finally, we exploit these results to devise a virtual dynamics of the quasi-potential in a generic parameter space that, upon integration, allows one to calculate non-equilibrium rate functions starting from the known system free energy. This method considerably simplifies the complexity of the original problem, which amounts to the solution of an infinite-order algebraic equation in the gradient of the quasi-potential.

2. Macroscopic limit and large deviations principle

We consider a Markovian jump process on a discrete space $\mathbf{n} \in \mathbb{N}^k$ with time-independent transition rates $\lambda_\rho(\mathbf{n})$, associated to jumps $\mathbf{n} \rightarrow \mathbf{n} + \Delta_\rho$. Namely, the time evolution of the probability distribution $P(\mathbf{n}, t)$ over states is given by the master equation

$$\partial_t P(\mathbf{n}, t) = \sum_\rho [\lambda_\rho(\mathbf{n} - \Delta_\rho) P(\mathbf{n} - \Delta_\rho, t) - \lambda_\rho(\mathbf{n}) P(\mathbf{n}, t)]. \quad (1)$$

We assume that a macroscopic limit exists: there is a scaling parameter Ω such that the transition rates scale as Ω and that the typical values of the density $\mathbf{x} = \mathbf{n}/\Omega$ are finite for $\Omega \rightarrow \infty$. Important examples satisfying the previous assumptions include chemical reaction networks [22–28] in the large volume limit, electronic circuits [29, 30] in the limit of large capacitances (see example in section 8.1), and some coarse-grained models of interacting many-body systems [31]. Then, in the limit $\Omega \rightarrow \infty$, the probability distribution $P(\mathbf{x}, t)$ satisfies a LDs principle [32, 33]:

$$P(\mathbf{x}, t) \asymp e^{-\Omega f(\mathbf{x}, t)}. \quad (2)$$

Indeed, plugging the previous ansatz in the master equation and keeping only the dominant terms in $\Omega \rightarrow \infty$, we obtain the following dynamical equation for the rate function $f(\mathbf{x}, t)$ (see appendix A for details):

$$\partial_t f(\mathbf{x}, t) = \sum_\rho \omega_\rho(\mathbf{x}) [1 - e^{\Delta_\rho \cdot \nabla f(\mathbf{x}, t)}], \quad (3)$$

where $\omega_\rho(\mathbf{x}) = \lim_{\Omega \rightarrow \infty} \lambda_\rho(\Omega \mathbf{x})/\Omega$ are the scaled rates and the gradient operator is with respect to \mathbf{x} . Thus, the steady state rate function $f_{ss}(\mathbf{x})$ must satisfy [34]:

$$0 = \sum_\rho \omega_\rho(\mathbf{x}) [1 - e^{\Delta_\rho \cdot \nabla f_{ss}(\mathbf{x})}]. \quad (4)$$

Equation (3) is a Hamilton–Jacobi equation for the action $f(\mathbf{x}, t)$, with the Hamiltonian $H(\mathbf{x}, \mathbf{p}) = \sum_\rho \omega_\rho(\mathbf{x}) [1 - e^{\Delta_\rho \cdot \mathbf{p}}]$ and generalized momenta defined as usual as $\mathbf{p} = \nabla f$. Therefore, the solutions of equation (3) can also be found by integrating the associated Hamiltonian equations [35–37]. Equation (4) represents the stationary version, corresponding to a Hamiltonian dynamics on the manifold of constant null energy $H(\mathbf{x}, \mathbf{p}) = 0$. The steady state rate function solving the latter equation is also known as a ‘quasi-potential’ or ‘non-equilibrium potential’. This is in part because of the analogy between equation (2) and the usual equilibrium Gibbs distribution, but more importantly because $f_{ss}(\mathbf{x})$ can be shown to be a Lyapunov function of the deterministic dynamics [38], as the potential energy is in overdamped equilibrium settings (see appendix B). Away from equilibrium, the time derivative of $f_{ss}(\mathbf{x})$

along deterministic trajectories was related to the adiabatic/non-adiabatic decomposition of the entropy production rate [39–41] in reference [34]. Also, the quasi-potential provides information about the lifetime of non-equilibrium metastable states, in full analogy to equilibrium reaction-rate theory [30, 37, 42, 43].

3. Thermodynamic structure

Thermodynamic consistency is enforced by assuming that for each transition with rate $\lambda_\rho(\mathbf{n})$ there is a corresponding reverse transition with rate $\lambda_{-\rho}(\mathbf{n})$ (thus, $\Delta_{-\rho} = -\Delta_\rho$), and that those rates are related by the local detailed balance (LDB) conditions

$$\log \frac{\lambda_\rho(\mathbf{n})}{\lambda_{-\rho}(\mathbf{n} + \Delta_\rho)} = -\beta [\Phi(\mathbf{n} + \Delta_\rho) - \Phi(\mathbf{n}) - W_\rho(\mathbf{n})], \quad (5)$$

where $\Phi(\mathbf{n})$ is the energy associated to a given state \mathbf{n} , and $W_\rho(\mathbf{n})$ is a non-conservative work contribution associated to the transition $\mathbf{n} \rightarrow \mathbf{n} + \Delta_\rho$. For simplicity we consider isothermal settings at inverse temperature β , but all of the following results hold also in more general situations where the system might interact with reservoirs at different temperatures. In that case, the LDB conditions take a form equivalent to equation (5), with the energy $\beta\Phi(\mathbf{n})$ being replaced by a generalized Massieu potential, and the work contributions can be expressed in terms of a minimal set of fundamental forces [44].

In the macroscopic limit we will assume that the energy of the system is an extensive quantity and therefore the limit $\phi(\mathbf{x}) = \lim_{\Omega \rightarrow \infty} \Phi(\mathbf{n} = \Omega\mathbf{x})/\Omega$ is well defined. For the work contributions we will abuse notation and write $W_\rho(\mathbf{x}) = \lim_{\Omega \rightarrow \infty} W_\rho(\mathbf{n} = \Omega\mathbf{x})$. Thus, we consider situations in which the work contributions associated to a transition do not scale with the system size, as is indeed the case in electronic circuits [29, 30] and chemical reaction networks [23–28]. According to the previous definitions, for large Ω the LDB conditions become:

$$\log \frac{\omega_\rho(\mathbf{x})}{\omega_{-\rho}(\mathbf{x})} = -\beta [\Delta_\rho \cdot \nabla \phi(\mathbf{x}) - W_\rho(\mathbf{x})]. \quad (6)$$

4. Deterministic dynamics

From equation (2) we see that as we increase the scale parameter Ω , the distribution $P(\mathbf{x}, t)$ becomes increasingly localized around the most probable states, i.e. the minima of the rate function $f(\mathbf{x}, t)$. If we let $\mathbf{x}(t)$ be one of the minima of $f(\cdot, t)$ for a given time t , then it is possible to show that it evolves according to the following closed deterministic dynamics [22, 30, 37]:

$$d_t \mathbf{x}(t) = \mathbf{u}(\mathbf{x}(t)), \quad (7)$$

where the deterministic drift $\mathbf{u}(\mathbf{x})$ is given by

$$\mathbf{u}(\mathbf{x}) = \sum_\rho \omega_\rho(\mathbf{x}) \Delta_\rho = \sum_{\rho > 0} I_\rho(\mathbf{x}) \Delta_\rho. \quad (8)$$

In the last equality we have exploited the fact that, due to the requirement of thermodynamic consistency, every transition has a reversed associated one, and expressed the drift $\mathbf{u}(\mathbf{x})$ in terms of the average forward currents $I_\rho(\mathbf{x}) = \omega_\rho(\mathbf{x}) - \omega_{-\rho}(\mathbf{x})$. Indeed, a second order truncation in the Kramers–Moyal expansion of the master equation in equation (1) leads to a Fokker–Planck equation with the vector field $\mathbf{u}(\mathbf{x})$ defined above as the drift. However, note that such expansion is not globally valid, but only locally around the minima of $f(\cdot, t)$ [45]. An alternative proof of the previous result, based solely on equation (3), is given in appendix B, where we also show that the steady state rate function satisfying equation (4) is a Lyapunov function for the deterministic dynamics [38].

5. Equilibrium state and linear response

In this and the following section we will consider the perturbation of equation (4) around the equilibrium steady state. For this, it is useful to define $\omega_\rho^{(0)}(\mathbf{x})$ to be the scaled rates in the absence of non-conservative sources of work (i.e. when $W_\rho(\mathbf{x}) = 0$). According to equation (6), they satisfy

$$\log \frac{\omega_\rho^{(0)}(\mathbf{x})}{\omega_{-\rho}^{(0)}(\mathbf{x})} = -\beta \Delta_\rho \cdot \nabla \phi(\mathbf{x}). \quad (9)$$

In that case, the dynamics of the system is said to be (unconditionally) detailed balanced, and for long times a thermal equilibrium state will be reached, which is exactly given by the Gibbs distribution $P_{\text{eq}}(\mathbf{n}) \propto \exp(-\beta\Phi(\mathbf{n}))$. Thus, by equation (2), the stationary rate function corresponding to this situation is $f_{\text{ss}}(\mathbf{x}) = \beta\phi(\mathbf{x})$. In a general situation with non-vanishing work contributions $W_\rho(\mathbf{x})$, we can always consider the decomposition

$$f_{\text{ss}}(\mathbf{x}) = \beta\phi(\mathbf{x}) + g(\mathbf{x}), \quad (10)$$

where $g(\mathbf{x})$ quantifies the deviations with respect to thermal equilibrium. We will show in what follows that, to first order in the quantities $\beta W_\rho(\mathbf{x})$, $g(\mathbf{x})$ must satisfy the equation

$$\mathbf{u}^{(0)}(\mathbf{x}) \cdot \nabla g(\mathbf{x}) = -\beta \dot{W}^{(0)}(\mathbf{x}), \quad (11)$$

where the vector field $\mathbf{u}^{(0)}(\mathbf{x})$ and the scalar field $\dot{W}^{(0)}(\mathbf{x})$ are given by

$$\mathbf{u}^{(0)}(\mathbf{x}) = \sum_{\rho>0} I_\rho^{(0)}(\mathbf{x}) \Delta_\rho \quad \dot{W}^{(0)}(\mathbf{x}) = \sum_{\rho>0} I_\rho^{(0)}(\mathbf{x}) W_\rho(\mathbf{x}), \quad (12)$$

in terms of the scaled currents

$$I_\rho^{(0)}(\mathbf{x}) = \omega_\rho^{(0)}(\mathbf{x}) - \omega_{-\rho}^{(0)}(\mathbf{x}). \quad (13)$$

Before proceeding with the proof of equation (11), we discuss the interpretation of the previous expressions. In the first place, $I_\rho^{(0)}(\mathbf{x})$ is just the average current along transition ρ according to the detailed balanced dynamics, given that the state of the system is \mathbf{x} . Then, the vector field $\mathbf{u}^{(0)}(\mathbf{x})$ is the net drift in state space for state \mathbf{x} , as discussed in the previous section. As a consequence, the field lines associated with $\mathbf{u}^{(0)}(\mathbf{x})$ are just the deterministic trajectories of the detailed balanced system. The quantity $I_\rho^{(0)}(\mathbf{x}) W_\rho(\mathbf{x})$ is the average scaled work rate associated to the transition ρ in the state \mathbf{x} (to lower order in $W_\rho(\mathbf{x})$), and $\dot{W}^{(0)}(\mathbf{x})$ is thus the total scaled work rate (the actual work rate is $\Omega \dot{W}^{(0)}(\mathbf{x})$).

Equation (11) displays an apparent ‘gauge invariance’, meaning that if $g(\mathbf{x})$ is any given solution, then $\tilde{g}(\mathbf{x}) = g(\mathbf{x}) + h(\mathbf{x})$ will also be a solution provided that $\mathbf{u}^{(0)}(\mathbf{x}) \cdot \nabla h(\mathbf{x}) = 0$. However, this indeterminacy is removed if one demands that the function $g(\mathbf{x})$ must be a single-valued function for all \mathbf{x} . This can be seen by considering a situation in which the deterministic dynamics has a unique fixed point \mathbf{x}^* (i.e. $\mathbf{u}(\mathbf{x}^*) = 0$). Then, for any initial condition \mathbf{x}_0 , the ensuing trajectory $\mathbf{x}(t|\mathbf{x}_0)$ satisfies $\lim_{t \rightarrow +\infty} \mathbf{x}(t|\mathbf{x}_0) = \mathbf{x}^*$. Also,

$$d_t g(\mathbf{x}(t|\mathbf{x}_0)) = d_t \mathbf{x}(t|\mathbf{x}_0) \cdot \nabla g(\mathbf{x}(t|\mathbf{x}_0)) = \mathbf{u}^{(0)}(\mathbf{x}(t|\mathbf{x}_0)) \cdot \nabla g(\mathbf{x}(t|\mathbf{x}_0)) = -\beta \dot{W}^{(0)}(\mathbf{x}(t|\mathbf{x}_0)), \quad (14)$$

which gives, once integrated over time,

$$g(\mathbf{x}_0) = g(\mathbf{x}^*) + \beta \int_0^{+\infty} dt \dot{W}^{(0)}(\mathbf{x}(t|\mathbf{x}_0)). \quad (15)$$

Thus, fixing the value $g(\mathbf{x}^*)$ at the fixed point, the previous equation allows to determine $g(\mathbf{x}_0)$ for any other point \mathbf{x}_0 . This can be generalized to situations with multistability at equilibrium, where the deterministic dynamics has multiple fixed points. In that case one can construct local solutions $g(\mathbf{x})$ for each basin of attraction in the same way as before, fixing an arbitrary value for $g(\mathbf{x})$ at each fixed point. These values must be later adjusted in order for the different local solutions to match continuously with each other at the separatrices between different basins of attraction (the existence of a continuous quasi-potential for diffusive systems with multiple coexisting attractors is discussed in [35]).

6. Proof

We now proceed with the proof of equation (11). Since we are assuming thermodynamic consistency, we can rewrite equation (4) as:

$$\begin{aligned} 0 &= \sum_{\rho>0} \{ \omega_\rho(\mathbf{x}) [1 - e^{\Delta_\rho \cdot \nabla f_{\text{ss}}(\mathbf{x})}] + \omega_{-\rho}(\mathbf{x}) [1 - e^{-\Delta_\rho \cdot \nabla f_{\text{ss}}(\mathbf{x})}] \} \\ &= \sum_{\rho>0} [\omega_\rho(\mathbf{x}) - \omega_{-\rho}(\mathbf{x}) e^{-\Delta_\rho \cdot \nabla f_{\text{ss}}(\mathbf{x})}] [1 - e^{\Delta_\rho \cdot \nabla f_{\text{ss}}(\mathbf{x})}]. \end{aligned} \quad (16)$$

Now we use the LDB conditions in equation (6) to write:

$$0 = \sum_{\rho>0} \omega_\rho(\mathbf{x}) [1 - e^{\beta [\Delta_\rho \cdot \nabla \phi(\mathbf{x}) - W_\rho(\mathbf{x})]} e^{-\Delta_\rho \cdot \nabla f_{\text{ss}}(\mathbf{x})}] [1 - e^{\Delta_\rho \cdot \nabla f_{\text{ss}}(\mathbf{x})}]. \quad (17)$$

We see from the last expression that when $W_\rho(\mathbf{x}) = 0$, the choice $f_{ss}(\mathbf{x}) = \beta\phi(\mathbf{x})$ is indeed a solution, and in fact it makes each of the terms in the sum vanish independently. Thus, considering $f_{ss}(\mathbf{x}) = \beta\phi(\mathbf{x}) + g(\mathbf{x})$, we have

$$0 = \sum_{\rho>0} \omega_\rho(\mathbf{x}) [1 - e^{-\beta W_\rho(\mathbf{x})} e^{-\Delta_\rho \cdot \nabla g(\mathbf{x})}] [1 - e^{\beta \Delta_\rho \cdot \nabla \phi(\mathbf{x})} e^{\Delta_\rho \cdot \nabla g(\mathbf{x})}]. \quad (18)$$

Close to equilibrium $W_\rho(\mathbf{x})$ is small by definition and $g(\mathbf{x})$ is assumed to be small and of the same order. It is therefore justified to expand equation (18) to first order in both of them. Doing that, we notice that the first factor in square brackets becomes linear in $W_\rho(\mathbf{x})$ and $g(\mathbf{x})$, and therefore, to first order, all the other factors can be evaluated at zero order,

$$\begin{aligned} 0 &= \sum_{\rho>0} \omega_\rho^{(0)}(\mathbf{x}) [\beta W_\rho(\mathbf{x}) + \Delta_\rho \cdot \nabla g(\mathbf{x})] [1 - e^{\beta \Delta_\rho \cdot \nabla \phi(\mathbf{x})}] \\ &= \sum_{\rho>0} I_\rho^{(0)}(\mathbf{x}) [\beta W_\rho(\mathbf{x}) + \Delta_\rho \cdot \nabla g(\mathbf{x})], \end{aligned} \quad (19)$$

where in the last line we used the detailed balance conditions in equation (9) to construct $I_\rho^{(0)}(\mathbf{x}) = \omega_\rho^{(0)}(\mathbf{x}) - \omega_{-\rho}^{(0)}(\mathbf{x})$. The result of equation (11) follows immediately.

7. Dynamical extensions

By considering a similar splitting of the time-dependent rate function $f(\mathbf{x}, t) = \beta\phi(\mathbf{x}, t) + g(\mathbf{x}, t)$, and repeating exactly the same derivation, this time starting from the dynamical equation of equation (3), we obtain the following dynamical equation for $g(\mathbf{x}, t)$:

$$\partial_t g(\mathbf{x}, t) = \beta \dot{W}^{(0)}(\mathbf{x}) + \mathbf{u}^{(0)}(\mathbf{x}) \cdot \nabla g(\mathbf{x}, t). \quad (20)$$

We can also consider a situation with time dependent rates. For a time-dependent scaled energy $\phi(\mathbf{x}, t)$, $\beta\phi(\mathbf{x}, t)$ is the rate function that defines an instantaneous equilibrium state. Considering now the splitting $f(\mathbf{x}, t) = \beta\phi(\mathbf{x}, t) + g(\mathbf{x}, t)$, we obtain

$$\partial_t g(\mathbf{x}, t) = \beta [\dot{W}^{(0)}(\mathbf{x}, t) - \partial_t \phi(\mathbf{x}, t)] + \mathbf{u}^{(0)}(\mathbf{x}, t) \cdot \nabla g(\mathbf{x}, t), \quad (21)$$

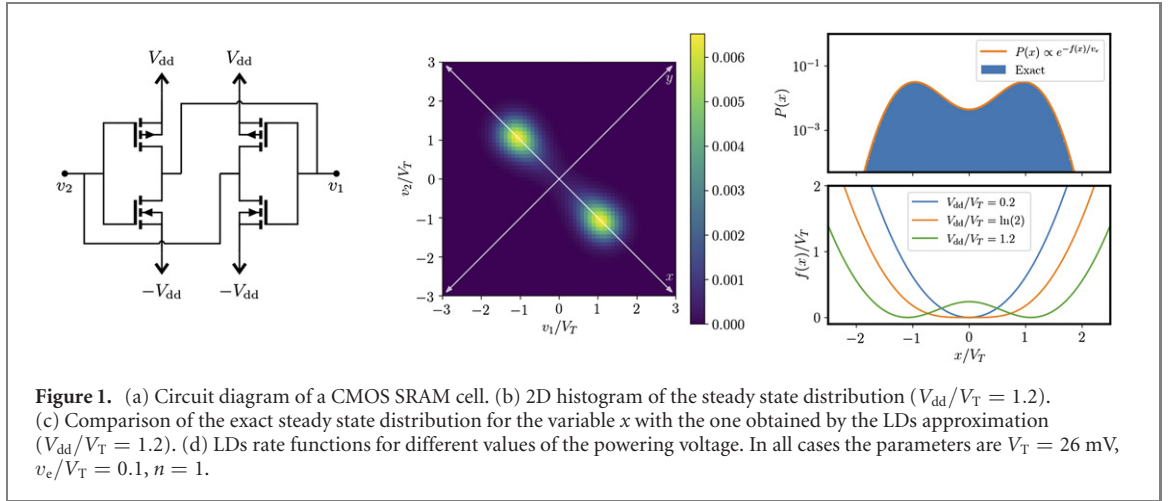
where the time-dependent work rate $\dot{W}^{(0)}(\mathbf{x}, t)$ and drift $\mathbf{u}^{(0)}(\mathbf{x}, t)$ are defined in exactly the same way as before, but in terms of the time-dependent currents $I_\rho^{(0)}(\mathbf{x}, t) = \omega_\rho^{(0)}(\mathbf{x}, t) - \omega_{-\rho}^{(0)}(\mathbf{x}, t)$. Since the previous equation is a linear partial differential equation for $g(\mathbf{x}, t)$, for periodically driven systems one may also apply Floquet theory.

8. Examples

8.1. CMOS SRAM cell

We consider the model of a low-power static random access memory (SRAM) cell developed in [29, 30]. The usual CMOS implementation of this kind of memories involves two inverters or NOT gates connected in a loop, each of which is composed of two MOS transistors, as shown in figure 1(a). The memory is powered by applying a voltage bias $2V_{dd}$, which takes the system out of thermal equilibrium. The stochastic model of this device has two degrees of freedom: the charges q_1 and q_2 at the output of each inverter, which correspond to voltages $v_{1/2} = q_{1/2}/C$, where C is a value of capacitance characterizing the device. Two Poisson rates are associated to each of the four transistors, corresponding to forward and backward conduction events. In each conduction event, or jump, the voltages $v_{1/2}$ can change by the elementary voltage $v_e = q_e/C$, where q_e is the positive electron charge. Also, all Poisson rates are proportional to I_0/q_e , where I_0 is a characteristic current. Under certain physical scalings of the transistors, as discussed in [30], both I_0 and C increase linearly with the size of the device, and therefore we can take C , or equivalently v_e^{-1} , as the scaling parameter Ω above. Thus, the macroscopic limit $\Omega \rightarrow \infty$ must in this case be understood as the limit $v_e/V_T \rightarrow 0$ and $v_e/V_{dd} \rightarrow 0$, where $V_T = (\beta q_e)^{-1}$ is the thermal voltage. A typical steady state distribution for this system (in its bistable phase, see below) is shown in figure 1(b). The rate function $f(v_1, v_2)$ corresponding to such steady state can be computed in terms of new variables $x = (v_1 - v_2)/2$ and $y = (v_1 + v_2)/2$. For example, the rate function of the reduced distribution for x is exactly given by [30]:

$$f(x) = \frac{x^2 + 2V_{dd}x}{V_T} + \frac{2nV_T}{n+2} [L(x, V_{dd}) - L(x, -V_{dd})], \quad (22)$$



where $L(x, V_{dd}) = \text{Li}_2(-\exp((V_{dd} + x(1 + 2/n))/V_T))$, and $\text{Li}_2(\cdot)$ is the polylogarithm function of second order, with n a parameter characterizing the transistors. In figure 1(c) we compare the distribution obtained from the previous rate function with exact numerical results. The agreement is essentially perfect even if only a few tens of electrons are involved (the scaling parameter is only $v_e = 0.1V_T$, and $V_{dd} = 1.15V_T$). Also, in figure 1(c) we show that there is a transition from a monostable phase into the bistable phase that allows the storage of a bit of information, occurring at the critical powering voltage $V_{dd}^* = \ln(2)V_T$ for $n = 1$. Expanding the previous expression to first order in V_{dd} , we obtain:

$$f(x) = \underbrace{\frac{x^2}{V_T}}_{\beta\phi(x)} - 4V_{dd} \underbrace{\frac{n}{n+2} \ln\left(2 \cosh\left(\frac{n+2}{2n} \frac{x}{V_T}\right)\right)}_{g(x)} + \mathcal{O}(V_{dd}^2), \quad (23)$$

where we have identified the equilibrium and non-equilibrium contributions as in equation (10). We will now recover this last result using the formalism developed here.

We begin by writing down the deterministic equations of motion, which read:

$$\begin{aligned} d_t v_1 &= I(v_1, v_2) - I(-v_1, -v_2) \\ d_t v_2 &= I(v_2, v_1) - I(-v_2, -v_1), \end{aligned} \quad (24)$$

where the scaled current $I(v_1, v_2)$ is defined in terms of the scaled rates $\omega_{\pm}(v_1, v_2)$ as $I(v_1, v_2) = \omega_+(v_1, v_2) - \omega_-(v_1, v_2)$. In turn, these rates are given by:

$$\begin{aligned} \omega_+(v_1, v_2) &= (I_0/C) e^{(V_{dd}-v_2-V_{th})/(nV_T)} \\ \omega_-(v_1, v_2) &= \omega_+(v_1, v_2) e^{-(V_{dd}-v_1)/(V_T)}. \end{aligned} \quad (25)$$

Also, because of the symmetry of the device, the work contribution of each forward elementary transition is $q_e V_{dd}$ [29], and therefore the total work rate is

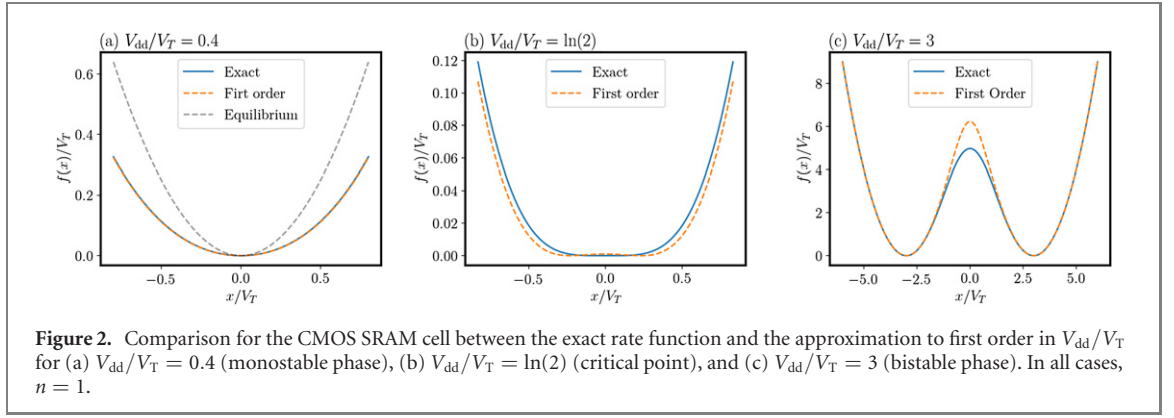
$$\dot{W}(v_1, v_2) = q_e V_{dd} [I(v_1, v_2) + I(-v_1, -v_2) + I(v_2, v_1) + I(-v_2, -v_1)]. \quad (26)$$

To proceed, we write the equations of motion in the previously defined variables x and y :

$$\begin{aligned} d_t x &= [I(x, y) - I(-x, y) - I(-x, -y) + I(x, -y)]/2 \\ d_t y &= [I(x, y) + I(-x, y) - I(-x, -y) - I(x, -y)]/2, \end{aligned} \quad (27)$$

where the change of variables in the function $I(\cdot)$ is implicit. We see from the previous equations that if we consider an initial condition where $y(0) = 0$, then the ensuing trajectory will satisfy $y(t) = 0$ at all times, since $d_t y|_{y=0} = 0$ for all x . This is true, in particular, for the equilibrium dynamics (that is, for $V_{dd} = 0$) which is the one we must consider in equation (11). Thus, over a trajectory with $y(0) = 0$, equation (11) reduces to:

$$d_t x^{(0)} d_x g(x) = -\beta \dot{W}^{(0)}(x), \quad (28)$$



where the superscript (0) means in this case that the currents should be evaluated at $V_{dd} = 0$. Then, we obtain:

$$\begin{aligned}
 g(x) &= -2 \frac{V_{dd}}{V_T} \int_0^x dx' \frac{I(x', 0) + I(-x', 0)}{I(x', 0) - I(-x', 0)} \Big|_{V_{dd}=0} \\
 &= -4V_{dd} \frac{n}{n+2} \ln \left(2 \cosh \left(\frac{n+2}{2n} \frac{x}{V_T} \right) \right), \quad (29)
 \end{aligned}$$

in full agreement with equation (23).

In figure 2 we compare the exact rate function of equation (22) with the first order approximation obtained with our formalism. We see in figure 2(a) that while the approximation is a first order expansion in $V_{dd}/V_T \ll 1$, it is essentially exact for values as high as $V_{dd}/V_T = 0.4$. Even more remarkably, the first order approximation is able to predict the transition to bistability, although as we see in figure 2(b) the critical point is underestimated (since at the exact critical point $V_{dd}^* = \ln(2)V_T$ the approximation already developed a bistability). The accuracy of the first order approximation can be understood by noting that the next order vanishes by symmetry, as well as all even powers of V_{dd} in the expansion of equation (22). Finally, although the approximation is not expected to work at all for $V_{dd}/V_T \gg 1$, we see in figure 2(c) that it correctly gets the most probable values (which correspond to the fixed points of the deterministic dynamics) as well as the curvature around them, while it moderately fails to describe the height of the barrier separating the two fixed points.

8.2. Schlögl model

We consider now the Schlögl model [46], a well known autocatalytic chemical model displaying bistability far from equilibrium. It consists in the following set of chemical reactions,



Here, species A and B will be considered to be chemostated reservoirs (i.e. we fix their concentration), and the only degree of freedom for the problem is the number of molecules of species X. The reactions are assumed to take place in a well-mixed container of volume V , which can be taken to be the macroscopic parameter Ω . Thus, in the macroscopic limit we deal with the concentration x of species X, and the scaled rates (ensuing the mass-action law) read:

$$\omega_1(x) = k_{+1}ax^2, \quad \omega_{-1}(x) = k_{-1}x^3, \quad \omega_2(x) = k_{+2}b, \quad \omega_{-2}(x) = k_{-2}x. \quad (31)$$

We can set $k_{+1}a = 1 = k_{+2}b$ choosing appropriate units for time and concentrations. Then, in this case the LDB conditions impose the relation $k_{-1} = k_{-2}e^{-\Delta\mu}$, where $\Delta\mu$ is the chemical potential difference between species A and B (measured in units such that $\beta = 1$ in the following). Thermal equilibrium is realized at $\Delta\mu = 0$, for which the deterministic concentration of x is $x^{(0)} = 1/k_{-1} = 1/k_{-2}$. The system can display bistability at $\Delta\mu \neq 0$ only for $k_{-2} \gtrsim 1.7$ (so we will set $k_{-2} = 2$ in the following plots). The exact rate function of the non-equilibrium steady state was obtained for example in [46] and reads:

$$f(x) = \int_0^x \log \left(\frac{yk_{-2}(1 + e^{-\Delta\mu}y^2)}{1 + y^2} \right) dy \quad (32)$$

$$= x \log \left(\frac{xk_{-2}(1 + e^{-\Delta\mu}x^2)}{1 + x^2} \right) + 2e^{\Delta\mu/2} \arctan \left(e^{-\frac{\Delta\mu}{2}} x \right) - x - 2 \arctan(x) \quad (33)$$

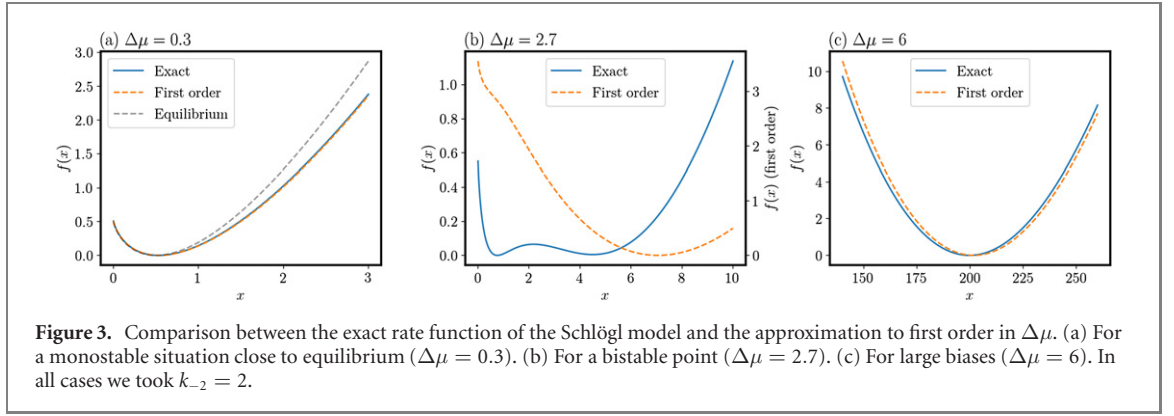


Figure 3. Comparison between the exact rate function of the Schlögl model and the approximation to first order in $\Delta\mu$. (a) For a monostable situation close to equilibrium ($\Delta\mu = 0.3$). (b) For a bistable point ($\Delta\mu = 2.7$). (c) For large biases ($\Delta\mu = 6$). In all cases we took $k_{-2} = 2$.

$$= \underbrace{x \log(k_{-2}x) - x}_{\beta\phi(x)} + \underbrace{\Delta\mu(\arctan(x) - x)}_{g(x)} + \mathcal{O}((\Delta\mu)^2). \quad (34)$$

We now recover the previous result by employing our formalism. The equilibrium currents are

$$I_1^{(0)}(x) = (1 - k_{-2}x)x^2, \quad I_2^{(0)}(x) = 1 - k_{-2}x. \quad (35)$$

The work per reactions are $W_1 = \Delta\mu$ and $W_2 = 0$, the total work rate is $\dot{W}^{(0)}(x) = \Delta\mu x^2(1 - k_{-2}x)$ and $u^{(0)}(x) = (1 - k_{-2}x)(1 + x^2)$. Then, the equation for the first order correction to the rate function is

$$(1 - k_{-2}x)(1 + x^2) dx g(x) = -x^2(1 - k_{-2}x)\Delta\mu, \quad (36)$$

from which we obtain $g(x) = -\Delta\mu \int_0^x y^2/(1 + y^2) dy = \Delta\mu[\arctan(x) - x]$, in full agreement with equation (34).

In figure 3(a) we compare the equilibrium rate function with the exact and first order approximation for $\Delta\mu = 0.3$, showing that the first order approximation is indeed accurate. However, in contrast to the previous example, the first order approximation fails to offer an even qualitative description for values of $\Delta\mu$ at which the exact rate function develops a bistability (in fact, the first order approximation never develops a bistability). This model becomes monostable again for larger values of $\Delta\mu$, and in that case the first order approximation correctly estimates the most probable value and the curvature around it, even if it is not expected to work in that regime, as was also observed in the previous example.

9. Comparison with usual linear response

The approach that we presented exploits the existence of a macroscopic limit with an LD principle and applies LR to the LD rate function. A natural question to ask is whether this offers any advantage with respect to usual LR theory [16–21], which directly gives the first order correction to the probability distribution (instead of the rate function associated to it), and does not require any macroscopic limit. We now show that our approach is more accurate than usual LR and valid further away from equilibrium.

Assuming that the first order correction $g(x)$ is already known, the LD approximation for the steady state distribution is:

$$\begin{aligned} P_{ss}(\mathbf{x}) &\simeq \frac{1}{Z} e^{-\Omega\beta f(\mathbf{x})} \\ &\simeq \frac{1}{Z} e^{-\Omega(\beta\phi(\mathbf{x}) + g(\mathbf{x}))} \\ &= \frac{1}{Z_{eq}} e^{-\Omega\beta\phi(\mathbf{x})} \frac{Z_{eq}}{Z} e^{-\Omega g(\mathbf{x})} \\ &\simeq P_{eq}(\mathbf{x}) \frac{Z_{eq}}{Z} e^{-\Omega g(\mathbf{x})}. \end{aligned} \quad (37)$$

In the first line of the previous equation the LD approximation is involved, while in the second line the LR approximation at the level of the rate function is involved. Also, $P_{eq}(\mathbf{x}) = \exp(-\Omega\beta\phi(\mathbf{x}))/Z_{eq}$ is the LD approximation of the equilibrium state. The partition function Z is

$$Z = \int d\mathbf{x}' e^{-\Omega\beta f(\mathbf{x}')} = \int d\mathbf{x}' e^{-\Omega\beta\phi(\mathbf{x}')} e^{-\Omega g(\mathbf{x}')}, \quad (38)$$

and $Z_{\text{eq}} = Z|_{g(\mathbf{x})=0}$. For the LR approximation to be sensible we must have $g(\mathbf{x}) \ll \beta\phi(\mathbf{x})$. However, if we consider the more stringent conditions $\Omega g(\mathbf{x}) \ll 1$, we can write

$$P_{\text{ss}}(\mathbf{x}) \simeq P_{\text{eq}}(\mathbf{x}) \left(1 + \Omega \int d\mathbf{x}' P_{\text{eq}}(\mathbf{x}') [g(\mathbf{x}') - g(\mathbf{x})] \right) + \mathcal{O}(\Omega^2 g^2), \quad (39)$$

and

$$Z \simeq Z_{\text{eq}} \left(1 - \Omega \int d\mathbf{x}' P_{\text{eq}}(\mathbf{x}') g(\mathbf{x}') \right) + \mathcal{O}(\Omega^2 g^2). \quad (40)$$

From the previous equation we see that the first order perturbation to the probabilities, $\delta P(\mathbf{x}) = P_{\text{ss}}(\mathbf{x}) - P_{\text{eq}}(\mathbf{x})$, is related to the first order perturbation to the LD rate function by

$$\delta P(\mathbf{x})/P_{\text{eq}}(\mathbf{x}) = -\Omega g(\mathbf{x}) + c, \quad (41)$$

where c is a constant ensuring $\int d\mathbf{x} \delta P(\mathbf{x}) = 0$. Thus, we see that $\delta P(\mathbf{x})$ and $g(\mathbf{x})$ contain the same information, as one can be computed from the other. However, in the macroscopic limit ($\Omega \gg 1$), the LR approximation at the level of the rate function is expected to be more accurate than the LR approximation at the level of the probabilities, and also valid for stronger perturbations. The reason is that for $\Omega \gg 1$ the conditions $g(\mathbf{x}) \ll \beta\phi(\mathbf{x})$ might be satisfied even if the conditions $\Omega g(\mathbf{x}) \ll 1$ are not.

We now illustrate the previous results in the example of section 8.1 (CMOS SRAM cell). For a given value of the powering voltage V_{dd} , we will compare three probability distributions: (i) the LD approximation with the exact rate function of equation (22): $P_{\text{ex}}(\mathbf{x}) \propto \exp(-f(\mathbf{x})/v_e)$ (note that, as shown in figure 1(c), this distribution can be considered exact already for $v_e/V_T \simeq 0.1$), (ii) the LD approximation combined with the LR approximation of the rate function: $P_g(\mathbf{x}) \propto \exp(-(x^2/V_T + g(\mathbf{x}))/v_e)$, where $g(\mathbf{x})$ is given by equation (29), and (iii) the usual LR approximation at the level of the probability distribution: $P_\delta(\mathbf{x}) = P_{\text{eq}}(\mathbf{x}) + \delta P(\mathbf{x}) = P_{\text{eq}}(\mathbf{x})(1 - g(\mathbf{x})/v_e + c)$ (see equation (41)). In figure 4(a) we see that for $V_{\text{dd}}/V_T = 0.2$, both $P_g(\mathbf{x})$ and $P_\delta(\mathbf{x})$ correctly describe the deviations from the equilibrium distribution. However, further away from equilibrium (in particular, at the critical point $V_{\text{dd}} = V_{\text{dd}}^* = V_T \ln(2)$), $P_g(\mathbf{x})$ remains an acceptable approximation while $P_\delta(\mathbf{x})$ does not, as shown in figure 4(b). Finally, in figure 4(c) we compare the Hellinger distances $H(P_g, P_{\text{ex}})$ and $H(P_\delta, P_{\text{ex}})$ between the exact distribution and the LR approximations at the level of the rate function or at the level of the probability distribution, respectively, as a function of the powering voltage V_{dd} (the Hellinger distance $0 \leq H(p, q) \leq 1$ between distributions $p(\mathbf{x})$ and $q(\mathbf{x})$ is defined by $H^2(p, q) = 1 - \sum_i \sqrt{p(x_i)q(x_i)}$). We see that $P_g(\mathbf{x})$ is always a better approximation with respect to $P_\delta(\mathbf{x})$, and that the gap between the two increases as we go deeper into the macroscopic limit (i.e. if we decrease the value of the elementary voltage v_e), as predicted above. We note that for sufficiently large V_{dd} the function $P_\delta(\mathbf{x})$ stops being positive definite (a common issue of the usual LR approximation at the level of the probability distribution) and therefore the Hellinger distance cannot be computed. Also, in this particular case the distance $H(P_g, P_{\text{ex}})$ reaches a maximum for large V_{dd} , after which it decreases again. This is related to the observation made in the previous section that the LR approximation at the level of the rate function correctly describes the most probable values and the curvature around them for large biases, even if *a priori* it is not expected to work in that regime.

10. Linear response of a non-equilibrium steady state

In this section we generalize our previous results by considering the problem of perturbing a general non-equilibrium steady state of an autonomous system. The rate function $f(\mathbf{x})$ of the steady state corresponding to some given work contributions $W_\rho(\mathbf{x})$ must satisfy equation (17):

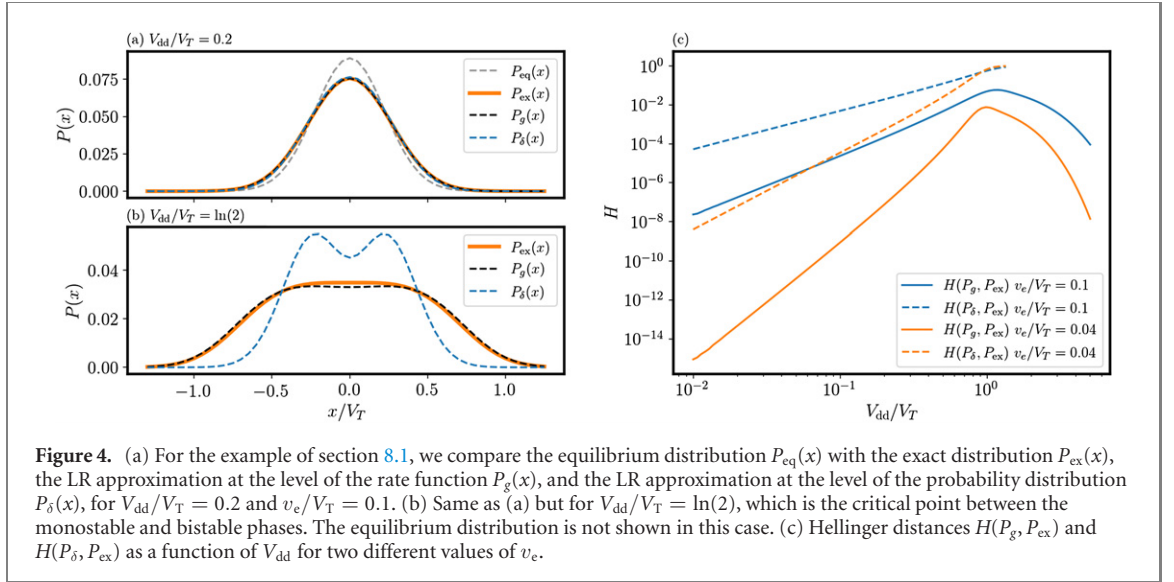
$$0 = \sum_{\rho>0} \omega_\rho(\mathbf{x}) \left[1 - e^{\beta[\Delta_\rho \cdot \nabla \phi(\mathbf{x}) - W_\rho(\mathbf{x})]} e^{-\Delta_\rho \cdot \nabla f(\mathbf{x})} \right] \left[1 - e^{\Delta_\rho \cdot \nabla f(\mathbf{x})} \right]. \quad (42)$$

Let us now consider the perturbation $W_\rho(\mathbf{x}) \rightarrow W'_\rho(\mathbf{x}) = W_\rho(\mathbf{x}) + \delta W_\rho(\mathbf{x})$, and the corresponding response $f(\mathbf{x}) \rightarrow f'(\mathbf{x}) = f(\mathbf{x}) + g(\mathbf{x})$. By expanding the previous equation to first order in $\delta W_\rho(\mathbf{x})$ and $g(\mathbf{x})$, we find

$$\mathcal{U}(\mathbf{x}) \cdot \nabla g(\mathbf{x}) = -\beta \dot{W}(\mathbf{x}). \quad (43)$$

The previous equation has the same structure as equation (11) for perturbations around equilibrium, but this time the vector field $\mathcal{U}(\mathbf{x})$ and the scalar field $\dot{W}(\mathbf{x})$ are given by

$$\mathcal{U}(\mathbf{x}) = \sum_{\rho>0} (\omega_{-\rho}(\mathbf{x}) e^{-\Delta_\rho \cdot \nabla f(\mathbf{x})} - \omega_\rho(\mathbf{x}) e^{\Delta_\rho \cdot \nabla f(\mathbf{x})}) \Delta_\rho, \quad (44)$$



and

$$\dot{W}(\mathbf{x}) = \sum_{\rho>0} \omega_{-\rho}(\mathbf{x}) (e^{-\Delta_\rho \cdot \nabla f(\mathbf{x})} - 1) \delta W_\rho(\mathbf{x}). \quad (45)$$

The previous results of section 5 are easily recovered by considering $f(\mathbf{x}) \rightarrow \beta\phi(\mathbf{x})$, $W_\rho(\mathbf{x}) \rightarrow 0$, and using the LDB conditions at equilibrium, since in that way we see that $\mathbf{U}(\mathbf{x})$ reduces to the equilibrium deterministic drift $\mathbf{u}(\mathbf{x})$, and $\dot{W}(\mathbf{x})$ to the work rate $\dot{W}(\mathbf{x})$. We note that $\mathbf{U}(\mathbf{x})$ does not correspond to the field of deterministic trajectories, and also that $\dot{W}(\mathbf{x})$ cannot be interpreted as the average work rate for a given state, as is the case for perturbations of the equilibrium distribution. However, it is interesting to note that the fixed points of the true unperturbed deterministic dynamics are shared by the dynamics corresponding to the alternative drift $\mathbf{U}(\mathbf{x})$, and also that $\dot{W}(\mathbf{x})$ vanishes at those fixed points. These properties are important to analyze the errors of the numerical method introduced in the next section, and are a consequence of the fact that the gradient of the exact rate function $f(\mathbf{x})$ vanishes at the deterministic fixed points (see appendix B).

11. Virtual evolution of non-equilibrium steady states

In equilibrium statistical mechanics the unnormalized thermal distribution $P(\mathbf{x}|\beta) = \exp(-\beta\Phi(\mathbf{x}))$ at a given inverse temperature β can be obtained by solving the evolution equation $d_\beta P(\mathbf{x}|\beta) = -\Phi(\mathbf{x})P(\mathbf{x}|\beta)$, with a known initial condition (typically, the infinite temperature distribution $P(\mathbf{x}|0) = 1$). Thus, the previous differential equation allows to obtain the distribution for an inverse temperature $\beta + d\beta$ given the knowledge of the distribution at inverse temperature β . Interestingly, equation (43) provides the basis for a similar procedure, in which a non-equilibrium steady state distribution is evolved starting, for example, from the equilibrium one. In order to illustrate this idea we will consider a system with a single state-independent work parameter λ (as the examples given in section 8). Then, the steady state rate function $f(\mathbf{x}|\lambda)$ has a parametric dependence on λ , and we can rewrite equation (43) as:

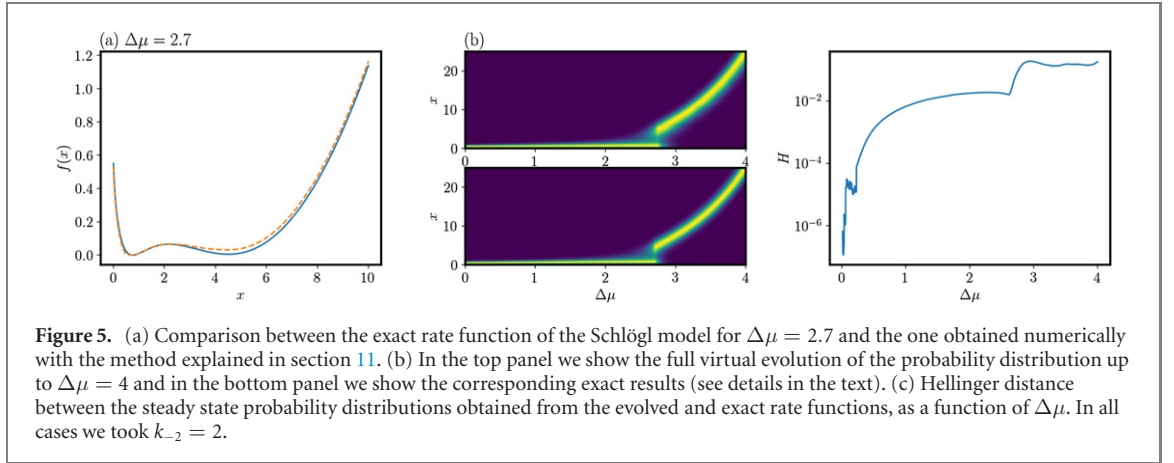
$$\mathbf{U}(\mathbf{x}|\lambda) \cdot \nabla[\partial_\lambda f(\mathbf{x}|\lambda)] = -\beta \dot{W}(\mathbf{x}|\lambda). \quad (46)$$

Solving this equation we can obtain $\partial_\lambda f(\mathbf{x}|\lambda)$ up to an irrelevant constant, and if $f(\mathbf{x}|\lambda)$ is known, we can approximate $f(\mathbf{x}|\lambda + d\lambda) \simeq f(\mathbf{x}|\lambda) + d\lambda \partial_\lambda f(\mathbf{x}|\lambda)$. Iterating this procedure it is in principle possible to obtain the rate function $f(\mathbf{x}|\lambda)$ for arbitrary λ starting from the equilibrium distribution $f(\mathbf{x}|0)$.

We now apply this method to the Schlögl model discussed in section 8.2. In that case the work parameter is $\lambda = \Delta\mu$ and, since the problem is one-dimensional, we can write equation (46) as:

$$\partial_{\Delta\mu} f(x|\Delta\mu) = -\beta \int_0^x \frac{\dot{W}(x'|\Delta\mu)}{\mathcal{U}(x'|\Delta\mu)} dx'. \quad (47)$$

We numerically solve the previous differential equation using a Runge–Kutta method of order 4, starting from the equilibrium rate function $f(x|0) = x \log(k_{-2}x) - x$ (see equation (34)). In figure 5(a) we compare the exact rate function given by equation (34) with the one obtained with our evolution method up to $\Delta\mu = 2.7$, a bistable point of the system. To obtain the evolved rate function shown in the plot we have



sampled the intervals $x \in [0, 10]$ and $\Delta\mu \in [0, 2.7]$ at 4×10^3 points each. In the top panel of figure 5(b) we show the results obtained by performing the evolution up to $\Delta\mu = 4$, this time sampling the interval $x \in [0, 25]$ at 10^4 points and the interval $\Delta\mu \in [0, 4]$ at 4×10^3 points. The color plot indicates the value of the probability distribution $P_{\text{ev}}(x|\Delta\mu) \propto e^{-\Omega f(x|\Delta\mu)}$ (with $\Omega = 10$), where $f(x|\Delta\mu)$ is the rate function obtained with our method at each point in the evolution. For visualization purposes, it was normalized so that the maximum of $P_{\text{ev}}(x|\Delta\mu)$ for a given $\Delta\mu$ is always 1. In the bottom panel of figure 5(b) we show the exact probability distribution $P_{\text{ex}}(x|\Delta\mu)$ computed in the same way as before but in terms of the exact rate function in equation (34). In figure 5(c) we show the Hellinger distance $H(P_{\text{ev}}, P_{\text{ex}})$ as a function of $\Delta\mu$. We see that the distance between the distributions computed from the evolved and exact rate functions increases rapidly at the beginning of the evolution, reaches a plateau, and increases again after the bistable region around $\Delta\mu \simeq 2.7$ is crossed.

An important comment regarding the numerical stability of the proposed method is in order. For equation (43) to have a smooth and well behaved solution (that is, such that $\nabla g(\mathbf{x})$ is finite for all \mathbf{x}), the field $\dot{\mathcal{W}}(\mathbf{x})$ must vanish wherever the field $\mathcal{U}(\mathbf{x})$ does. As already noted in the previous section, if \mathbf{x}^* is a fixed point of the deterministic dynamics, then from equations (44) and (45) we see that $\mathcal{U}(\mathbf{x}^*) = 0$ and $\dot{\mathcal{W}}(\mathbf{x}^*) = 0$. This follows from the LDB conditions and the fact that $\nabla f(\mathbf{x}^*) = 0$ (see appendix B). Thus, assuming that the deterministic fixed points are the only zeros of $\mathcal{U}(\mathbf{x})$, the previous minimal consistency condition is indeed satisfied by equations (44) and (45). Importantly, for this to hold, the unperturbed rate function $f(\mathbf{x})$ involved in the definitions of $\mathcal{U}(\mathbf{x})$ and $\dot{\mathcal{W}}(\mathbf{x})$ must correspond exactly to the parameters at which the rates $\omega_\rho(\mathbf{x})$ are evaluated. However, that correspondence is broken in the intermediate steps of our iterative procedure, since at each step (except at the first one when the equilibrium distribution is perturbed) the fields $\mathcal{U}(\mathbf{x})$ and $\dot{\mathcal{W}}(\mathbf{x})$ are constructed from an approximate rate function $f(\mathbf{x})$, which might not exactly satisfy $\nabla f(\mathbf{x}^*) = 0$. This issue might lead to errors and discontinuities for values of \mathbf{x} close to the deterministic fixed points at each step, which will propagate forward in the virtual evolution. That is in fact the main source of errors contributing to the deviations between the exact and evolved rate functions in figure 5(a). A more sophisticated numerical implementation of the proposed method might take as an input the deterministic fixed points $\{\mathbf{x}^*\}$ at each step, in order to ensure that they match the minima of the approximated rate function. This point should be considered for applications to more complex systems in many dimensions.

12. Conclusion

For thermodynamically consistent stochastic systems displaying a macroscopic limit in which the stationary distribution fulfills a LDs principle, we have shown how to compute the correction to the rate function, or quasi-potential, to first order in the forces taking the system out of thermal equilibrium. We have applied this result to two examples: a realistic model of an important kind of electronic memory and the Schlögl model of a bistable chemical reaction. We also generalized our methods to time-dependent settings. In the final part of our work we considered the LR of arbitrary non-equilibrium steady states. Based on that result, we developed a method to incrementally evolve the stationary state rate function in the space of thermodynamic forces, which in principle allows to obtain the rate function arbitrarily far from equilibrium, starting from the known equilibrium one. This might lead to new numerical schemes to study non-equilibrium fluctuations at the mesoscopic level, providing an alternative to the usual stochastic simulations based on the Gillespie algorithm.

Our work makes use of the thermodynamic consistency of the dynamics and of the existence of a macroscopic limit. The results might remain nonetheless an excellent approximation even significantly away from the strict macroscopic limit, as is the case in the examples discussed in section 8. Furthermore, we have shown that in the mesoscopic or macroscopic regime the LR approach at the level of the rate function is more accurate than traditional LR theory at the level of the probability distribution, even if both approaches are first order expansions in the thermodynamic forces. As shown in the example in section 9 (figure 4), what actually happens is that both approaches become more accurate as the scale of the system is increased, but with a larger gap between the accuracy of LR at the level of the rate function and at the level of the probability distribution. This is understood as follows. A fixed change $\Delta \mathbf{x}$ in the density $\mathbf{x} = \mathbf{n}/\Omega$ requires an increasing number of elementary transitions as the scale of the system is increased. Also, the work contribution during a transition does not scale with the system size, while the associated change in free energy does. Thus, it is natural to expect LR approximations to be more accurate in the macroscopic regime, since they are based on the assumption that work contributions are small with respect to other energy scales. However, while the first order non-equilibrium correction to the rate function is independent of the scale (by definition of the rate function), the first order correction to the probability distribution $P(\mathbf{x})$ is proportional to the scale parameter (see equation (41)), and that constrains the strength of non-equilibrium forces for which the correction remains small.

Acknowledgments

We acknowledge funding from the INTER project ‘TheCirco’ (INTER/FNRS/20/15074473) and CORE project ‘NTEC’ (C19/MS/13664907), funded by the Fonds National de la Recherche (FNR, Luxembourg), and from the European Research Council, project NanoThermo (ERC-2015-CoG Agreement No. 681456).

Data availability statement

All data that support the findings of this study are included within the article (and any supplementary files).

Appendix A. Dynamics of the large deviations rate function

In this section we provide details of the derivation of equation (3) in the main text, starting from the master equation in equation (1). In the first place we represent a given discrete state \mathbf{n} of the system by its ‘density’ $\mathbf{x} = \mathbf{n}/\Omega$. For example, the components of \mathbf{n} are the numbers of molecules in a chemical reaction network or the numbers of charges in an electronic circuit, while the components \mathbf{x} are, respectively, the concentrations or the voltages. Thus, a transition $\mathbf{n} \rightarrow \mathbf{n} + \Delta_\rho$ corresponds to a change Δ_ρ/Ω in the density \mathbf{x} . We consider the LDs ansatz $P(\mathbf{x}, t) = \exp(-\Omega f(\mathbf{x}, t))/Z(t)$ for the time-dependent probability distribution in terms of \mathbf{x} , where $Z(t) = \sum_{\mathbf{n}} \exp(-\Omega f(\mathbf{n}/\Omega, t))$ takes care of the normalization. Then, we have

$$P(\mathbf{x} + \Delta_\rho/\Omega, t) = P(\mathbf{x}, t) e^{-\Delta_\rho \cdot \nabla f(\mathbf{x}) + \mathcal{O}(\Omega^{-1})}, \quad (\text{A1})$$

and

$$\mathrm{d}_t P(\mathbf{x}, t) = -P(\mathbf{x}, t) [\Omega \mathrm{d}_t f(\mathbf{x}, t) + \mathrm{d}_t \log(Z(t))]. \quad (\text{A2})$$

Equation (3) is then obtained by inserting the previous two expressions in the master equation of equation (1), dividing both sides by $\Omega P(\mathbf{x}, t)$, and taking the limit $\Omega \rightarrow \infty$ in which $\Omega^{-1} \log(Z(t)) \rightarrow 0$.

Appendix B. Deterministic dynamics

We define $\mathbf{x}(t)$ to be a minimum of the rate function $f(\cdot, t)$ at time t . Then, we have that $\partial_{x_i} f(\mathbf{x}(t), t) = 0$ and that the symmetric matrix with elements $[\partial_{x_i, x_j}^2 f(\mathbf{x}(t), t)]_{ij}$ is positive semidefinite. For time $t + \mathrm{d}t$ we can write:

$$\begin{aligned} 0 &= \partial_{x_i} f(\mathbf{x}(t + \mathrm{d}t), t + \mathrm{d}t) \\ &= \partial_{x_i} f(\mathbf{x}(t) + \mathbf{u}(t) \mathrm{d}t + \mathcal{O}(\mathrm{d}t^2), t + \mathrm{d}t) \\ &= u_j(t) \partial_{x_j, x_i}^2 f(\mathbf{x}(t), t) \mathrm{d}t + \partial_t \partial_{x_i} f(\mathbf{x}(t), t) \mathrm{d}t + \mathcal{O}(\mathrm{d}t^2), \end{aligned} \quad (\text{B1})$$

where we have introduced the velocity $\mathbf{u}(t) = d_t \mathbf{x}(t)$, and repeated indices are summed. Now we employ the differential equation satisfied by $f(\mathbf{x}, t)$, equation (3) in the main text, to obtain:

$$\begin{aligned} \partial_t \partial_{x_i} f(\mathbf{x}, t) &= \partial_{x_i} \partial_t f(\mathbf{x}, t) \\ &= \partial_{x_i} \left[\sum_{\rho} \omega_{\rho}(\mathbf{x}) \left[1 - e^{(\Delta_{\rho})_j \partial_{x_j} f(\mathbf{x}, t)} \right] \right] \\ &= \sum_{\rho} \partial_{x_i} \omega_{\rho}(\mathbf{x}) \left[1 - e^{(\Delta_{\rho})_j \partial_{x_j} f(\mathbf{x}, t)} \right] + \sum_{\rho} \omega_{\rho}(\mathbf{x}) \left[-e^{(\Delta_{\rho})_j \partial_{x_j} f(\mathbf{x}, t)} (\Delta_{\rho})_j \partial_{x_i x_j}^2 f(\mathbf{x}, t) \right]. \end{aligned} \quad (\text{B2})$$

Noticing that the first sum in the last line of the previous equation vanishes when evaluated at $\mathbf{x}(t)$, and combining the result with equation (B1), up to first order in dt we obtain:

$$\left[\mathbf{u}(t) - \sum_{\rho} \omega_{\rho}(\mathbf{x}(t)) \Delta_{\rho} \right]_j \partial_{x_j x_i}^2 f(\mathbf{x}(t), t) = 0. \quad (\text{B3})$$

Thus, whenever the matrix $[\partial_{x_i x_j}^2 f(\mathbf{x}(t), t)]_{ij}$ can be assumed to be positive definite, we have $\mathbf{u}(t) - \sum_{\rho} \omega_{\rho}(\mathbf{x}(t)) \Delta_{\rho} = 0$, in agreement with equations (7) and (8) in the main text.

If the rate function has a single global minimum, then the instantaneous average value $\langle \mathbf{x}(t) \rangle = \Omega^{-1} \langle \mathbf{n}(t) \rangle = \Omega^{-1} \sum_{\mathbf{n}} \mathbf{n} P(\mathbf{n}, t)$ also evolves as $d_t \langle \mathbf{x}(t) \rangle = \mathbf{u}(\langle \mathbf{x}(t) \rangle)$ in the $\Omega \rightarrow \infty$ limit. In order to see this, we multiply equation (1) in the main text by \mathbf{n} and sum, obtaining:

$$\partial_t \sum_{\mathbf{n}} \mathbf{n} P(\mathbf{n}, t) = \sum_{\rho, \mathbf{n}} \mathbf{n} [\lambda_{\rho}(\mathbf{n} - \Delta_{\rho}) P(\mathbf{n} - \Delta_{\rho}, t) - \lambda_{\rho}(\mathbf{n}) P(\mathbf{n}, t)] \quad (\text{B4})$$

$$= \sum_{\rho, \mathbf{n}} [(\mathbf{n} + \Delta_{\rho}) \lambda_{\rho}(\mathbf{n}) P(\mathbf{n}, t) - \mathbf{n} \lambda_{\rho}(\mathbf{n}) P(\mathbf{n}, t)] \quad (\text{B5})$$

$$= \sum_{\rho, \mathbf{n}} \Delta_{\rho} \lambda_{\rho}(\mathbf{n}) P(\mathbf{n}, t), \quad (\text{B6})$$

where we shifted $\mathbf{n} \rightarrow \mathbf{n} + \Delta_{\rho}$ in the first sum on the rhs of (B4). Dividing both sides by Ω , passing to integral over \mathbf{x} and recognizing that P peaks around the single minimum of f as $\Omega \rightarrow \infty$ we obtain the desired result.

Finally, we mention that if $f_{ss}(\mathbf{x})$ is the steady state rate function, i.e. the solution to equation (4) in the main text, then it follows that [34, 38]

$$d_t f_{ss}(\mathbf{x}(t)) = \sum_{\rho} \omega_{\rho}(\mathbf{x}) \Delta_{\rho} \cdot \nabla f_{ss}(\mathbf{x}(t)) \leq \sum_{\rho} \omega_{\rho}(\mathbf{x}(t)) [e^{\Delta_{\rho} \cdot \nabla f_{ss}(\mathbf{x})} - 1] = 0, \quad (\text{B7})$$

where we used equation (7) and $x \leq e^x - 1$. Thus, the function $f_{ss}(\mathbf{x})$ decreases monotonically along deterministic trajectories and, since the rate function can always be considered to be bounded by below, it is a Lyapunov function of the deterministic dynamics. As a consequence, if \mathbf{x}^* is a fixed point of the deterministic dynamics, then it is also a minimum of $f_{ss}(\mathbf{x})$, and therefore we have $\nabla f_{ss}(\mathbf{x}^*) = 0$ (assuming $f_{ss}(\mathbf{x})$ is continuous). A thermodynamic interpretation of $d_t \phi(\mathbf{x}(t))$ is given in [34].

References

- [1] Lebowitz J L and Bergmann P G 1957 Irreversible gibbsian ensembles *Ann. Phys., NY* **1** 1–23
- [2] Lebowitz J L 1959 Stationary nonequilibrium gibbsian ensembles *Phys. Rev.* **114** 1192
- [3] McLennan J A Jr 1959 Statistical mechanics of the steady state *Phys. Rev.* **115** 1405
- [4] Zubarev D N 1974 *Nonequilibrium Statistical Thermodynamics* (New York: Consultants Bureau)
- [5] Zubarev D N 1994 Nonequilibrium statistical operator as a generalization of Gibbs distribution for nonequilibrium case *Condens. Matter Phys.* **4** 7
- [6] Komatsu T S and Nakagawa N 2008 Expression for the stationary distribution in nonequilibrium steady states *Phys. Rev. Lett.* **100** 030601
- [7] Maes C and Netočný K 2010 Rigorous meaning of McLennan ensembles *J. Math. Phys.* **51** 015219
- [8] Colangelo M, Maes C and Wynants B 2011 A meaningful expansion around detailed balance *J. Phys. A: Math. Theor.* **44** 095001
- [9] Dhar A, Saito K and Hänggi P 2012 Nonequilibrium density-matrix description of steady-state quantum transport *Phys. Rev. E* **85** 011126
- [10] Ness H 2013 Nonequilibrium density matrix for quantum transport: Hershfield approach as a McLennan–Zubarev form of the statistical operator *Phys. Rev. E* **88** 022121
- [11] Falasco G and Esposito M 2021 Local detailed balance across scales: from diffusions to jump processes and beyond *Phys. Rev. E* **103** 042114
- [12] Crooks G E 2000 Path-ensemble averages in systems driven far from equilibrium *Phys. Rev. E* **61** 2361

- [13] Venttsel' A D and Freidlin M I 1970 On small random perturbations of dynamical systems *Russ. Math. Surv.* **25** 1–55
- [14] Bouchet F, Gawędzki K and Nardini C 2016 Perturbative calculation of quasi-potential in non-equilibrium diffusions: a mean-field example *J. Stat. Phys.* **163** 1157–210
- [15] Tél T, Graham R and Hu G 1989 Nonequilibrium potentials and their power-series expansions *Phys. Rev. A* **40** 4065
- [16] Proesmans K and van den Broeck C 2015 Onsager coefficients in periodically driven systems *Phys. Rev. Lett.* **115** 090601
- [17] Proesmans K, Cleuren B and van den Broeck C 2016 Linear stochastic thermodynamics for periodically driven systems *J. Stat. Mech.* **023202**
- [18] Proesmans K and Fiore C E 2019 General linear thermodynamics for periodically driven systems with multiple reservoirs *Phys. Rev. E* **100** 022141
- [19] Kay B, Saito K and Seifert U 2015 Thermodynamics of micro- and nano-systems driven by periodic temperature variations *Phys. Rev. X* **5** 031019
- [20] Bauer M, Brandner K and Seifert U 2016 Optimal performance of periodically driven, stochastic heat engines under limited control *Phys. Rev. E* **93** 042112
- [21] Tomé T and de Oliveira M J 2015 Stochastic approach to equilibrium and nonequilibrium thermodynamics *Phys. Rev. E* **91** 042140
- [22] Kurtz T G 1972 The relationship between stochastic and deterministic models for chemical reactions *J. Chem. Phys.* **57** 2976–8
- [23] Schmiedl T and Seifert U 2007 Stochastic thermodynamics of chemical reaction networks *J. Chem. Phys.* **126** 044101
- [24] Poletini M and Esposito M 2014 Irreversible thermodynamics of open chemical networks: I. Emergent cycles and broken conservation laws *J. Chem. Phys.* **141** 024117
- [25] Rao R and Esposito M 2016 Nonequilibrium thermodynamics of chemical reaction networks: wisdom from stochastic thermodynamics *Phys. Rev. X* **6** 041064
- [26] Lazarescu A, Cossetto T, Falasco G and Esposito M 2019 Large deviations and dynamical phase transitions in stochastic chemical networks *J. Chem. Phys.* **151** 064117
- [27] Avanzini F, Falasco G and Esposito M 2020 Thermodynamics of non-elementary chemical reaction networks *New J. Phys.* **22** 093040
- [28] Avanzini F, Penocchio E, Falasco G and Esposito M 2021 Nonequilibrium thermodynamics of non-ideal chemical reaction networks *J. Chem. Phys.* **154** 094114
- [29] Freitas N, Delvenne J-C and Esposito M 2020 Stochastic thermodynamics of non-linear electronic circuits: a realistic framework for computing around kT (accepted) (arXiv:2008.10578[cond-mat.stat-mech])
- [30] Freitas N, Proesmans K and Esposito M 2021 Reliability and entropy production in non-equilibrium electronic memories (arXiv:2103.01184[cond-mat.stat-mech])
- [31] Herpich T, Cossetto T, Falasco G and Esposito M 2020 Stochastic thermodynamics of all-to-all interacting many-body systems *New J. Phys.* **22** 063005
- [32] Touchette H 2009 The large deviation approach to statistical mechanics *Phys. Rep.* **478** 1–69
- [33] Touchette H and Harris R J 2011 Large deviation approach to nonequilibrium systems (arXiv:1110.5216[cond-mat.stat-mech])
- [34] Ge H and Qian H 2016 Mesoscopic kinetic basis of macroscopic chemical thermodynamics: a mathematical theory *Phys. Rev. E* **94** 052150
- [35] Graham R and Tél T 1986 Nonequilibrium potential for coexisting attractors *Phys. Rev. A* **33** 1322
- [36] Kamenev A 2011 *Field Theory of Non-equilibrium Systems* (Cambridge: Cambridge University Press)
- [37] Cossetto T 2020 Problems in nonequilibrium fluctuations across scales: a path integral approach *PhD Thesis* University of Luxembourg
- [38] Gang H 1986 Lyapounov function and stationary probability distributions *Z. Phys. B* **65** 103–6
- [39] Esposito M, Harbola U and Mukamel S 2007 Entropy fluctuation theorems in driven open systems: application to electron counting statistics *Phys. Rev. E* **76** 031132
- [40] Esposito M and van den Broeck C 2010 Three faces of the second law: I. Master equation formulation *Phys. Rev. E* **82** 011143
- [41] Ge H and Qian H 2010 Physical origins of entropy production, free energy dissipation, and their mathematical representations *Phys. Rev. E* **81** 051133
- [42] Peter H, Talkner P and Borkovec M 1990 Reaction-rate theory: fifty years after Kramers *Rev. Mod. Phys.* **62** 251
- [43] Bouchet F and Reygner J 2016 Generalisation of the Eyring–Kramers transition rate formula to irreversible diffusion processes *Ann. Henri Poincaré* **17** 3499–532
- [44] Rao R and Esposito M 2018 Conservation laws shape dissipation *New J. Phys.* **20** 023007
- [45] Gardiner C W 1985 *Handbook of Stochastic Methods* vol 3 (Berlin: Springer)
- [46] Vellela M and Qian H 2009 Stochastic dynamics and non-equilibrium thermodynamics of a bistable chemical system: the Schlögl model revisited *J. R. Soc. Interface.* **6** 925–40