Large deviations and dynamical phase transitions in stochastic chemical networks

Alexandre Lazarescu,[1] [ID] Tommaso Cossetto,[2] Gianmaria Falasco,[2] and Massimiliano Esposito[1,2,a) [ID]

AFFILIATIONS

[1] CPHT, CNRS, École Polytechnique, IP Paris, F-91128 Palaiseau, France
[2] Complex Systems and Statistical Mechanics, Physics and Material Science Research Unit, University of Luxembourg, L-1511 Luxembourg, Luxembourg

a) Electronic mail: alexandre.lazarescu@polytechnique.edu

ABSTRACT

Chemical reaction networks offer a natural nonlinear generalization of linear Markov jump processes on a finite state-space. In this paper, we analyze the dynamical large deviations of such models, starting from their microscopic version, the chemical master equation. By taking a large-volume limit, we show that those systems can be described by a path integral formalism over a Lagrangian functional of concentrations and chemical fluxes. This Lagrangian is dual to a Hamiltonian, whose trajectories correspond to the most likely evolution of the system given its boundary conditions. The same can be done for a system biased on time-averaged concentrations and currents, yielding a biased Hamiltonian whose trajectories are optimal paths conditioned on those observables. The appropriate boundary conditions turn out to be mixed so that, in the long time limit, those trajectories converge to well-defined attractors. We are then able to identify the largest value that the Hamiltonian takes over those attractors with the scaled cumulant generating function of our observables, providing a nonlinear equivalent to the well-known Donsker-Varadhan formula for jump processes. On that basis, we prove that chemical reaction networks that are deterministically multistable generically undergo first-order dynamical phase transitions in the vicinity of zero bias. We illustrate that fact through a simple bistable model called the Schlögl model, as well as multistable and unstable generalizations of it, and we make a few surprising observations regarding the stability of deterministic fixed points and the breaking of ergodicity in the large-volume limit.

Published under license by AIP Publishing. https://doi.org/10.1063/1.5111110

I. INTRODUCTION

The fundamental question of statistical physics is whether it is possible to describe a microscopic model, with many components and relatively simple rules, using only a small set of mesoscopic variables and to determine the emerging laws of the system at that scale. A good but rather trivial example of this can be found in the behavior of independent random walkers (particles) on a lattice: although the time-evolution of the complete probability distribution of configurations is relatively complicated, one can write a much simpler equation for the distribution of the local average density of particles, under an appropriate rescaling of time and space.[1] That equation takes the form of a Langevin equation, with a drift and a diffusion that are both linear in the density, and a noise whose variance is linear as well. The resulting process is called a linear stochastic diffusion and is said to be *hydrodynamic* in the sense that the behavior of the local density is autonomous (i.e., does not depend on higher moments of the density), by analogy with fluid mechanics.

A nonlinear generalization of this is found in the macroscopic fluctuation theory (MFT[2]), which deals with interacting particles close to equilibrium: in essence, if mesoscopic portions of the system get asymptotically close to equilibrium as the size of the system is increased, then the local density of particles obeys a Langevin equation similar to the one mentioned above but with nonlinear drift, diffusion, and noise variance.

The appropriate language to tackle such questions is that of large deviations,[3,4] for the same reason that an equilibrium system is characterized by its free energy. Indeed, the process equivalence that we invoke when stating such results is a logarithmic one:[5] the logarithm of the probability distribution of the mesoscopic time-dependent density profiles of our system, rescaled by the appropriate

parameter (volume or total number of particles), converges in the limit of that parameter going to infinity to the rescaled logarithm of the probability distribution of the corresponding Langevin equation. The argument of the exponential distribution is then generally called the *action*, and for a Markovian process, it takes the form of a time integral over a functional of states (position and density) and fluxes (velocity and currents) called the Lagrangian or the *level* 2.5 large deviation function.[6] The resulting probability density of trajectories is then a path-integral in time, amenable to all standard techniques from quantum path integrals.[7] In the case of diffusions, or of single particles subject to a Gaussian noise with small variance (known as the Freidlin-Wentzell theory of large deviations[8]), that Lagrangian is a quadratic function of the fluxes, and the general structure of those processes has been the focus of many studies in recent years.[9–12] Much effort has also been put toward obtaining them as continuous limits of discrete lattice gas models, through the aforementioned MFT[2] or equivalently the so-called *additivity principle*.[13,14]

From those path integrals, one can obtain in principle the probability of any less detailed observable, such as, for instance, any time average of densities or fluxes, by finding the trajectories that most likely realize specific values of those observables. This can be performed either by minimizing the action under a constraint or equivalently by introducing a Lagrange multiplier and computing the scaled cumulant generating function (SCGF) of the observable instead.[9] Obtaining the SCGF of a time-averaged observable, as well as the optimal trajectory extremizing the Lagrangian, gives important information on the way that the system can sustain unlikely fluctuations of said observables, potentially leading to strategies to enhance said fluctuations (e.g., to boost the efficiency of a machine[15,16]) or suppress them (e.g., to prevent catastrophic events[17]). This SCGF is in essence a free energy of the system with a virtual forcelike parameter conjugated to the current, similar to a virtual magnetic field being added to the Ising model to study its phase transitions. However, unlike for equilibrium statistical physics, adding this extra parameter to the path probabilities of the original process does not result in another process with different forces but rather in what is called a *biased process*,[18] which is not a true stochastic process (it does not conserve probability, for instance) except in very special cases.[19] Despite this conceptual difference from equilibrium free energies, dynamical cumulant generating functions are quite useful to compute large deviation functions in practice, as we shall demonstrate later.

The computation of SCGFs, both analytically and numerically, has been covered by many studies. It is particularly important in the context of transport models (or interacting lattice gases), where the fluctuations of the flow of carriers are central to the physics of the system, especially when far from equilibrium.[20] As regards analytical computations, the systems considered are usually one-dimensional, as they are more likely to be solvable, and include models close to equilibrium, as the symmetric simple exclusion process with periodic[21] and open boundary conditions[22,23] or on an infinite line,[24,25] the weakly asymmetric simple exclusion process,[26] the KMP model,[27] and the zero-range process with periodic[28] or open boundary conditions,[29] as well as those same models driven far from equilibrium such as the asymmetric simple exclusion process with periodic[30–33] and open boundary conditions[34–36] or on an

infinite line,[37,38] with some studies focusing on the probability of extremely unlikely currents.[39–41] Numerical methods have also been devised to compute those unlikely probabilities, which are not easily accessible through standard stochastic algorithms due to the exponentially large sample they would require. Those focus on much of the same models and include numerical minimization of the Lagrangian for diffusions,[42] density matrix renormalization group (DMRG) techniques,[43,44] and cloning algorithms which provide a dynamical equivalent of importance sampling and have proven very efficient.[45–49]

A particularly interesting feature of some such systems is so-called *dynamical phase transitions*, which are qualitative changes in the way a system typically produces and maintains atypical values of certain observables. Many such transitions are known in the context of interacting lattice gases or diffusions.[14,50–54] For instance, for many driven one-dimensional models with hard-core interactions, which exhibit a nonvanishing current of particles in their stationary state, it is known that positive and negative fluctuations of that current are qualitatively different:[41,55,56] in order to produce a current lower than usual, the system had to better produce localized blockages (leading to inhomogeneous product states), whereas to produce a current larger than usual, it needs to microscopically order the particles everywhere (leading to homogeneous correlated states,[57] sometimes called *hyperuniform*[58]). The probability cost of those two types of events is also qualitatively different, which can be seen either as a change in analyticity in the large deviation function of the current or equivalently in the SCGF. Other examples involve simple diffusions[59] and kinetically constrained models.[60]

The topic of macroscopic fluctuations is well-established and well-studied for systems which are rescaled to a continuous-space process, but they make no less sense in situations with an intrinsically discrete geometry. A simple Markov jump process on a finite state-space, for instance, can be interpreted as the evolution equation for the typical density distribution of a large number of independent random walkers, and the probability distribution of that density in the limit of many walkers can be described through the large deviation formalism.[61] The resulting process is linear in the density as for diffusions due to the independence of its components.

In this paper, we consider a specific class of interacting jump processes, which provide a natural nonlinear generalization of Markov jump processes, namely, *chemical reaction networks*. Their deterministic versions are the well-known chemical rate equation systems,[62] which are structurally equivalent to a jump process where the probabilities are replaced by monomials of concentrations (the so-called *mass-action* kinetics). Their behavior can however be much more complex than their linear counterpart: whereas an irreducible Markov jump process on a finite state-space has a single steady state in the stationary limit, a chemical system can exhibit multistability,[63] limit cycles,[64] or even strange attractors.[65] It should be noted that this nonlinearity does not come from interactions between the particles, as it would in interacting diffusions, for instance, Ref. 66, but is of an entropic nature due to the reactions requiring all reactants to be at the same place. Beside their usefulness as nonlinear versions of jump processes, chemical network models have self-evident applications in chemistry, with much recent interest specifically in biochemistry, where the stochastics

and energetics of mesoscopic random systems out of equilibrium have direct consequences on the way living cells function.[67–70] Many models of populations (e.g., of microbes) have also been studied with the same formalism.[71]

The study of stochastic chemical systems is by no means a recent topic, with roots dating back to the 1960s.[72] Early works, which focused on solving simple models of elementary reactions with few molecules,[73] contained the essential ingredients of what would later be called the chemical master equation,[74] which is now well-established as a physically reasonable microscopic model for well-mixed chemical systems.[75] In this formalism, chemical reaction networks are a linear Markov jump process on a composition space (usually $\mathbb{N}^N$, where N is the number of chemical species) and have been extensively studied as such.[76–79] An intriguing property of those systems is the apparent paradox between the unique microscopic fixed point of the dynamics and the potentially numerous attractors that should emerge in the macroscopic limit, which can be understood as a breaking of ergodicity[80] due to the emergence of different time scales in those systems,[81–83] with a structure very similar to that of equilibrium phase transitions[84,85] and with kinetic properties that can be accessed through a Hamiltonian formalism.[86]

The large deviations of these models, as well as other types of population dynamics, have been a growing subject of interest in the past decade from various communities,[85,87–94] though to our knowledge a full theory of their dynamical fluctuations and rare event probabilities is still missing.

In the present work, we describe the large deviation behavior of generic chemical reaction networks, starting from the chemical master equation[74] which describes their microscopic dynamics. This paper aims at being pedagogical and self-contained, and therefore contains many formulae and derivations that are not entirely necessary for practical computations of, say, large deviation functions of time-averaged currents. The reader mainly interested in such concrete matters is advised to first refer to Sec. VI, which provides a summary of the relevant formulae and results.

In Sec. II, we introduce those models, along with a few important concepts to analyze them, and we define the mesoscopic variables that will be relevant in the large-volume limit.

In Sec. III, we take the said large-volume limit and show that the dynamics of those systems can be described through a path integral with a Lagrangian that can be computed explicitly. That path integral is dominated by a trajectory solving an Euler-Lagrange equation or equivalently Hamilton's equations for an appropriate Hamiltonian. This will be done in several different ways, depending on whether one takes the time-derivative of the density or all the separate chemical currents as a flux variable in the Lagrangian and on whether one puts a bias on some observables through a Lagrange multiplier. A connection will also be made between our approach and the Doi-Peliti formalism.[95,96] This section is meant as a pedagogical overview of the Lagrangian/Hamiltonian formalism for large-volume stochastic processes and is detailed accordingly. As such, it contains many results that are likely not original, although we have been unable to find references to some of them in the literature (such as for the detailed formalism of Sec. III B). The focus is however put on the biased version of Sec. III C, which we do believe to be new in its most general form, although notable results already exist for biases specifically on densities and dynamical activity (see Refs. 85

and 97 and more recently Ref. 98). Readers familiar with these concepts can safely skip this section or simply skim it for notations. The proofs relating to these results are provided in the Appendix, and a practical summary for the computation of long time large deviation functions, focused on only the necessary formulae, is given in Sec. VI.

Section IV contains the main original results of this work: we apply said formalism to compute the scaled cumulant generating (SCGF) function of currents and densities of our models, which is the Legendre transform of the large deviation function of stationary currents and densities. We show that it can be expressed in terms of the maximal value of the biased Hamiltonian defined in Sec. II among those it takes on its critical manifolds, which generalizes the result by Donsker and Varadhan mentioned earlier.[99–102] We then show that systems whose deterministic equations have several attractors (i.e., that are multistable) generically undergo first-order dynamical phase transitions around small values of the dynamical biases. This is proven when those attractors are fixed points and conjectured for higher-dimensional attractors.

In Sec. V, we illustrate this last result by exhibiting a few of those dynamical phase transitions in variants of a simple bistable system called the Schlögl model.[63] We remark that, surprisingly, even unstable fixed points of the deterministic dynamics can become stable under bias and that the ergodicity of the microscopic process, broken by the large-volume limit, is usually restored. We also show that, in some cases where the deterministic chemical equations have stable or unstable regimes depending on their initial condition, a bias can turn the whole system unstable.

In Sec. VI, we give a hands-on summary of how one may compute large deviation functions of dynamical observables in a chemical reaction network, meant as a quick reference or a minimal guide to readers who may wish to skip the details of the derivations.

Finally, we conclude with a few interesting open questions on the topic.

## II. DEFINITION OF THE MICROSCOPIC PROCESS AND OBSERVABLES

We will be looking at the standard chemical master equation with mass-action kinetics.[62] Section II contains many definitions and notations that are specific to chemical reaction models, and that will be essential to understand some of the upcoming computations and results. For that reason, we provide a simple illustrative example in Sec. II D, where all those objects are made explicit. The nonspecialist reader is encouraged to refer to that example as they read to get a clearer picture of their meaning and properties.

A. Definition of the process

Our system contains chemical constituents (or *particles*), with N different species A_x, $x \in [[0, N[[$, and a microscopic state n of the system is given by the number of (indistinguishable) particles of each species, which we will write as n_x (we will treat n as a vector with components n_x). We assume that the system, of volume V, is well mixed and has no spatial dimension. The system evolves through *chemical reactions* where a certain set of particles (reactants) is destroyed and replaced by another set (products), with a certain rate at Poisson-distributed times. This defines a Markov

jump process on the state-space $\mathbb{N}^N$ (possibly on a subset of it if all states are not connected dynamically due to conservation laws; see Sec. IV C).

In order to define the process in a convenient way, we will not differentiate reactants and products (as reactions will often be reversible), and we will call all sets of particles occurring on either side of our reactions as *complexes*. A complex γ is defined by a set of non-negative integers ν_x^γ called *stoichiometric coefficients*, giving the number of particles of species A_x entering in complex γ, which we sort into the stoichiometric vectors $\nu^\gamma = \{\nu_x^\gamma\}$. The standard notation for the reaction destroying complex γ and creating complex γ' is then

$$\sum_x \nu_x^\gamma A_x \rightarrow \sum_x \nu_x^{\gamma'} A_x. \tag{1}$$
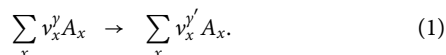
The rate at which this reaction is occurring from state n will be written as $W_{\gamma'\gamma}(n)$, taking the system from $\{n_x\}$ to $\{n_x - \nu_x^\gamma + \nu_x^{\gamma'}\}$. It will be convenient to define the difference between those states to be a generalized divergence,

$$\nabla_{\gamma'\gamma} = \nu^\gamma - \nu^{\gamma'}, \tag{2}$$

which is a matrix acting on functions of reactions (i.e., edges of the graph of connected complexes, i.e., ordered pairs of complexes), whose columns are differences of stoichiometric vectors, and which yields a variation of particle number. Its action on a test function $\lambda_{\gamma'\gamma}$ is given by

$$(\nabla\lambda)_x = \sum_{\gamma'\gamma} (\nu_x^\gamma - \nu_x^{\gamma'})\lambda_{\gamma'\gamma}. \tag{3}$$

Note that this operator ∇ can be factorized as $\nabla = \nu\mathfrak{d}$, where $\mathfrak{d}$ is the standard discrete divergence on the graph of complexes (also called the incidence matrix): $\mathfrak{d}_{\gamma'\gamma} = \delta_\gamma - \delta_{\gamma'}$, acting on reactions and yielding an antisymmetric function on the vertices of that graph (i.e., on complexes), and where ν is the matrix containing all the stoichiometric coefficients, acting on complexes and yielding particle numbers. We will only need this factorization in Sec. II C when considering conservation laws.

Moreover, the conjugate of this divergence defines a natural gradient from functions f of the species x to antisymmetric functions of the reactions,

$$(\nabla f)_{\gamma'\gamma} = \sum_x (\nu_x^{\gamma'} - \nu_x^\gamma)f_x. \tag{4}$$

We use the same notation ∇ as for the divergence, by analogy with the usual continuous case. However, this should not create any ambiguity, as the left and right spaces are different. We can then perform integrations by parts on functions of reactions: for two arbitrary test functions f_x and $\lambda_{\gamma'\gamma}$ dependent on species and reactions, respectively, we have

$$f\nabla \cdot \lambda = -\lambda \cdot \nabla f. \tag{5}$$

We can now write the chemical master equation with rates $W_{\gamma'\gamma}(n)$, for a probability distribution $P(n)$,

$$d_t P(n) = \sum_{\gamma',\gamma} W_{\gamma'\gamma}(n + \nabla_{\gamma',\gamma})P(n + \nabla_{\gamma',\gamma}) - W_{\gamma'\gamma}(n)P(n), \tag{6}$$

which we can also write algebraically in the bra-ket notation as $d_t|P\rangle = W|P\rangle$, with the Markov matrix W,

$$W = \sum_{\gamma',\gamma,n} W_{\gamma'\gamma}(n) |n - \nabla_{\gamma',\gamma}\rangle\langle n| - W_{\gamma'\gamma}(n) |n\rangle\langle n|. \tag{7}$$

A common and natural choice for $W_{\gamma'\gamma}(n)$ is the so-called *mass action* prescription, where the dependence in n simply comes from the combinatorial number of ways to choose the reactants, i.e., the product of $\binom{n_x}{\nu_x^\gamma}$ over γ. since $\nu_x^\gamma!$ is fixed, we may absorb it into the constant prefactor. Moreover, we will later need the prefactor to scale with the volume V with a specific exponent for V large, so for the sake of convenience we will write that term explicitly already. We get

$$W_{\gamma'\gamma}(n) = k_{\gamma'\gamma} \prod_x \frac{[n_x]!}{[n_x - \nu_x^\gamma]!} V^{1-\sum_x \nu_x^\gamma}, \tag{8}$$

where $k_{\gamma'\gamma}$ is called the *kinetic constant* of the reaction and is independent of n and of V.

Our aim will be to describe the fluctuations of such models around its typical behavior in the $V \rightarrow \infty$ limit. The order of V in each rate has been chosen so that the rates $W(n)$ be all linear in V in the large volume limit, which corresponds to a so-called *hyperbolic* scaling (also known as ballistic scaling). We will see in Sec. III B 2 that this scaling is compatible with the standard deterministic mass-action chemical equation systems.[62]

B. Time-additive dynamical observables

When taking the large volume limit, we will reduce the number of dynamical observables in our system: we will not be interested in the complete probability distribution of n and in the complete set of microscopic reactions dependent on the starting state n at every instant but only in the typical concentration of particles and total chemical currents integrated over an appropriate time window.

Let us therefore consider a single realization $n(\tau)$ of the microscopic process described above, and let us take a mesoscopic time step δt, which for now is simply a positive real constant. We can then define the empirical *concentration* ρ_x of species X as

$$\rho_x(t) \, V \, \delta t = \int_t^{t+\delta t} n_x(\tau)d\tau \tag{9}$$

and the empirical chemical currents $\lambda_{\gamma'\gamma}$ as

$$\lambda_{\gamma'\gamma}(t) \, V \, \delta t = w_{\gamma'\gamma} \equiv \#\big[\gamma \rightarrow \gamma'\big]_{t,t+\delta t}, \tag{10}$$

where the last term simply means the number of times reaction $\gamma \rightarrow \gamma'$ happens between t and $t + \delta t$, regardless of the state. Those observables are called *time-additive* because they obey Chasles relation in time on any history of the system. As a consequence, they are typically linear with respect to the observation time (hence the factors δt assumed in the left-hand sides).

Given those definitions, one can check that those quantities are related in the following way:

$$d_t\rho_x = \sum_{\gamma,\gamma'} (\nu_x^{\gamma'} - \nu_x^\gamma)\lambda_{\gamma'\gamma}, \quad \text{which we write as} \quad d_t\rho = -\nabla \cdot \lambda. \tag{11}$$

This is the continuity equation of our chemical system, and it is verified individually by each microscopic realization.

C. A caveat on conserved quantities

An interesting feature of chemical reaction networks is that they may have many nontrivial conserved quantities, i.e., specific linear combinations of concentrations that are left unchanged by the dynamics. For instance, if our system of chemical reactions involves benzene cycles with various functional groups, but none of the reactions actually destroys or creates a cycle, then the total number of benzene cycles is conserved (those unbreakable units are often called *moieties*). Other cases might not be so easily interpretable but can be characterized systematically by looking at algebraic properties of the process.[103,104] Those conservation laws are crucial features of chemical networks with interesting consequences on their behavior.[105]

Formally, looking at the continuity equation (11), we see that it involves the operator ∇ which might not be invertible, for instance, if it is not of full rank, which always happens if there are more species than reactions. For any vector $\mathfrak{c}_x$ annihilated by the gradient (i.e., in the left kernel of ∇), the quantity $\mathfrak{c}\rho$ is conserved,

$$d_t \mathfrak{c}\rho = -\mathfrak{c}\nabla \cdot \lambda = \lambda \cdot \nabla \mathfrak{c} = 0. \tag{12}$$

We call this equation a *conservation law*, and we will use the same term to designate the vector $\mathfrak{c}$ itself.
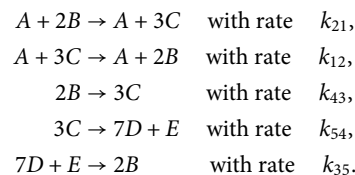
The effect of this is that the microscopic state-space of the system is split into many ergodic components, each corresponding to one set of values of all conserved quantities. This also means that the variable $\dot{\rho}$ is potentially of lower dimension than the number of species, which means that one has to be careful when taking Legendre transforms with respect to it, as we will very often do. In the present work, we will dismiss this issue by implicitly defining ρ on

a *reduced state space orthogonal to the conserved quantities*, i.e., one single ergodic component. In such a space, the matrix ∇ is invertible on the concentration side (i.e., to the right for the gradient and to the left for the divergence). We will however have to address the issue in Sec. IV C when considering initial conditions that span several ergodic components.

D. Example

In this section, we give a simple example in order to make all the previous definitions and notations more concrete, which can serve as a quick reference for the reader whenever in doubt.

Let us consider a system with $N = 5$ species $\{A, B, C, D, E\}$ and the following set of reactions:

$$
\begin{aligned}
A + 2B &\to A + 3C &\text{with rate}&\quad k_{21}, \\
A + 3C &\to A + 2B &\text{with rate}&\quad k_{12}, \\
2B &\to 3C &\text{with rate}&\quad k_{43}, \\
3C &\to 7D + E &\text{with rate}&\quad k_{54}, \\
7D + E &\to 2B &\text{with rate}&\quad k_{35}.
\end{aligned}
$$

The second reaction is the reverse of the first one, and the other three form a nonreversible cycle. We define the complexes $\gamma \in [[1, 5]]$ as the left-hand side of those reactions: $1 = (A + 2B)$, $2 = (A + 3C)$, $3 = (2B)$, $4 = (3C)$, $5 = (7D + E)$. The kinetic constants of the reactions are defined such that $k_{\gamma'\gamma}$ is the rate of destroying complex γ to create complex γ'.

The stoichiometry matrix v, adjacency matrix $\mathfrak{d}$, and chemical divergence ∇ are given by

$$
v = \begin{array}{c} \\ \\ \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 3 & 0 & 3 & 0 \\ 0 & 0 & 0 & 7 \\ 0 & 0 & 0 & 1 \end{bmatrix} \end{array}
\quad
\mathfrak{d} = \begin{bmatrix} +1 & -1 & 0 & 0 & 0 \\ -1 & +1 & 0 & 0 & 0 \\ 0 & 0 & +1 & 0 & -1 \\ 0 & 0 & -1 & +1 & 0 \\ 0 & 0 & 0 & -1 & +1 \end{bmatrix},
\quad
\nabla = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 2 & -2 & 2 & 0 & -2 \\ -3 & 3 & -3 & 3 & 0 \\ 0 & 0 & 0 & -7 & 7 \\ 0 & 0 & 0 & -1 & 1 \end{bmatrix} \tag{13}
$$

with $\nabla = v\mathfrak{d}$. Each column of v corresponds to the composition of one of the complexes, with rows indicating the number of particles of each species (labeled by the corresponding lower-case letter). Each column of $\mathfrak{d}$ indicates, for a given reaction, which is the reactant complex (+1) and the product complex (−1), and we have spit the matrix into blocks to emphasize the fact that it is block-diagonal due to the fact that the network of complexes has two connected components (sometimes called *linkage classes*). The columns of ∇ give, for each reaction, the net change in number of each species.

The mean-field chemical rate equations governing the evolution of the concentration in a large well-mixed system are then

$$d_t \rho_a = 0, \tag{14}$$

$$d_t \rho_b = -2\, k_{21}\, \rho_a \rho_b^2 + 2\, k_{12}\, \rho_a \rho_c^3 - 2\, k_{43}\, \rho_b^2 + 2\, k_{35}\, \rho_d^7 \rho_e, \tag{15}$$

$$d_t \rho_c = 3\, k_{21}\, \rho_a \rho_b^2 - 3\, k_{12}\, \rho_a \rho_c^3 + 3\, k_{43}\, \rho_b^2 - 3\, k_{54}\, \rho_c^3, \tag{16}$$

$$d_t \rho_d = 7\, k_{54}\, \rho_c^3 - 7\, k_{35}\, \rho_d^7 \rho_e, \tag{17}$$

$$d_t \rho_e = k_{54}\, \rho_c^3 - k_{35}\, \rho_d^7 \rho_e. \tag{18}$$

This can be written as $d_t \rho = -\nabla \cdot (k\rho^v)$, where $(k\rho^v)$ is the vector of average chemical currents for each reaction,

$$(k\rho^v) = \left[\; k_{21}\, \rho_a \rho_b^2 \;,\; k_{12}\, \rho_a \rho_c^3 \;,\; k_{43}\, \rho_b^2 \;,\; k_{54}\, \rho_c^3 \;,\; k_{35}\, \rho_d^7 \rho_e \;\right]. \tag{19}$$

This system has three conserved quantities, all of which can be easily found in this case by examining the reactions:

- The number of A is constant: the only reactions involving it (21 and 12) do not change that number. This is visible on the first row of ∇. The corresponding vector is $\mathfrak{c}^1 = [1, 0, 0, 0, 0]$. It is such that $\mathfrak{c}^1 \nu$ is in the left kernel of $\eth$, i.e., $\mathfrak{c}^1 \nu \eth = 0$ but $\mathfrak{c}^1 \nu \neq 0$.

- Given that D and E only appear together in complex 5, they always get created and destroyed together (with the appropriate multiplicities) so that $\rho_d - 7\rho_e$ is constant, which is clear from the rate equations. The corresponding vector is $\mathfrak{c}^2 = [0, 0, 0, 1, -7]$, which is in the left kernel of ν, i.e., $\mathfrak{c}^2 \nu = 0$.

- The last conserved quantity involves all the reactions: note that the first two can convert $2B$ into $3C$ and back, with A acting as a catalyst, and that the other three reactions also turn $2B$ into $3C$ but also $7D + E$. This means that the total number of groups of $2B$ or $3C$ or $7D + E$ is conserved. We can choose one representative from each group (one out of every pair of B, every trio of C, and every E), leading to the conserved quantity $\rho_b/2 + \rho_c/3 + \rho_e$, with a vector $\mathfrak{c}^3 = [0, 3, 2, 0, 6]$ renormalized to have integer entries. As in the first case, we have $\mathfrak{c}^3 \nu \eth = 0$ but $\mathfrak{c}^3 \nu \neq 0$.

Those vectors are not uniquely defined as they can be combined together: this constitutes a somewhat arbitrary basis of the left kernel of ∇. Note that those conservation laws are relevant for the study of statistical processes in general, not only in the context of chemical reactions, and that they are not always as obvious as here.[103]

E. Aside: Long-time large deviations

We will be interested in the distribution of additive observables in the limit of a large volume and small time step, but let us first examine a different case.

Let us consider for a moment a finite volume V and a single time step $\delta t = t$ which is large with respect to the relaxation time of the system. It is not guaranteed in general that the said relaxation time be well defined due to the possibility of the state-space being infinite, but we will assume that it is (which we can guarantee if the system has a certain type of conservation law but which is generally true if $P(n)$ is mostly contained in a compact at all times[62]). Let us also assume that our initial distribution has support in only one connected (ergodic) component of the state space.

In this context, let us examine the probability distributions $P_t(\lambda, \rho)$ of chemical currents and concentrations averaged over a single long time step. Given that this is a time-averaged additive observable in a Markov jump process on a finite state space (or a well-controlled infinite one), we know the following:[4] the aforementioned distribution follows a large deviation principle with a scale t, i.e., we can define a so-called *long-time large deviation function* $g(\lambda, \rho)$ such that

$$\lim_{t \to \infty}\left[-\frac{1}{t}\ln(P_t(\lambda, \rho))\right] = g(\lambda, \rho), \tag{20}$$

which we write as $P_t(\lambda, \rho) \asymp e^{-tg(\lambda, \rho)}$. The function g is, by construction, positive convex and vanishes only at the stationary value of the currents λ and concentrations ρ. Moreover, we can define the Legendre transform $E(\mathfrak{s}, h)$ of $g(\lambda, \rho)$, with a variable $\mathfrak{s}_{\gamma'\gamma}$ conjugate to $\lambda_{\gamma'\gamma}$ and h_x conjugate to ρ_x,

$$E(\mathfrak{s}, h) = \mathfrak{s} \cdot \lambda + h\rho - g(\lambda, \rho) \text{ with } \mathfrak{s}_{\gamma'\gamma} = \partial_{\lambda_{\gamma'\gamma}}g(\lambda, \rho) \text{ and } h_x = \partial_{\rho_x}g(\lambda, \rho), \tag{21}$$

which is also the scaled cumulant generating function (SCGF) of λ and ρ,

$$\left\langle e^{t(\mathfrak{s}\cdot\lambda + h\rho)}\right\rangle = e^{tE(\mathfrak{s}, h)}, \tag{22}$$

where $\langle\cdot\rangle$ represents an average over realizations, starting from any initial condition (which does not matter in the long-time limit). It is a classical result from the Donsker-Varadhan theory of large deviations[99–102] that the function $E(\mathfrak{s}, h)$ is also the *unique largest eigenvalue* of a deformed Markov matrix $W_{\mathfrak{s},h}$, where the nondiagonal entries carry an extra exponential weight $e^{\mathfrak{s}_{\gamma'\gamma}}$ and the diagonal entries have an extra linear term $\sum_x h_x \rho_x$,

$$W_{\mathfrak{s},h} = \sum_{\gamma',\gamma,n} e^{V\mathfrak{s}_{\gamma'\gamma}} W_{\gamma'\gamma}(n) |n - \nabla_{\gamma',\gamma}\rangle\langle n| - W_{\gamma'\gamma}(n) |n\rangle\langle n|$$
$$+ \sum_{x,n} h_x n_x |n\rangle\langle n|. \tag{23}$$

As demonstrated in the Appendix, this biased Markov matrix $W_{\mathfrak{s},h}$ is useful even if the observation time t is small: the role of its largest eigenvalue is lost, but it can still be used as a generator of the biased dynamics which produces the cumulant generating function $E(\mathfrak{s}, h)$ in the long-time limit [which in that case will be the same as in (22) up to a factor V].

Finally, note that in the case where the initial distribution spans several ergodic components α of the state-space, the term with the largest eigenvalues $E_\alpha(\mathfrak{s}, h)$ and a nonzero initial probability will exponentially dominate the others. This can lead to first-order phase transitions with respect to $\mathfrak{s}$ or h, as those eigenvalues are allowed to cross.

III. DYNAMICAL LARGE DEVIATION FORMALISM

In this section, we take the limits $V \to \infty$ and $\delta t \to 0$ on our microscopic process and describe the exponential rate functions (Lagrangians) of the probability distributions of the observables defined in Sec. II B, as well as their Legendre transforms (Hamiltonians), and obtain equations for the corresponding minimization problems. The proof of the large-volume limit is provided in the Appendix, where we see that it is also essential that, in the limit, $V\delta t \to \infty$. All of the following derivations also apply to population models with more complex rates (i.e., if $k_{\gamma'\gamma}$ depends on n), except for those of Sec. IV C which rely on mass-action kinetics.

We will look at three versions of the formalism: one where the flux-type variable of the Lagrangian is the time-derivative of the concentration (standard), one where we keep track of every individual chemical current (detailed), and one where we constrain the dynamics on a certain time-averaged value of the concentrations and currents (biased). The first one describes the probability of any history

of the concentration with minimal observables but cannot generally be written explicitly, whereas the second has more detail than necessary but is always explicit. They are both useful in constructing the third, which is what we need ultimately in order to have access to the scaled cumulant generating functions of chemical currents and concentrations.

In each subsection, we will derive the appropriate large volume large deviation rate function (Lagrangian) and the corresponding cumulant generating function for fluxes (Hamiltonian) and derive the equations of motion that solve the corresponding extremization problems (i.e., that describe the typical behavior of the system) along with the appropriate boundary conditions.

A. Standard formalism

We first describe the standard large deviation formalism for large volume Markov processes, sometimes called the WKB formalism[87,91,92] although in this case the low-noise property is proven rather than assumed.

Let us consider the integrated transition rate of our microscopic process, for a finite runtime t, between an initial state n_i and a final state n_f,

$$P_t[n_f|n_i] = \langle n_f|e^{tW}|n_i\rangle. \tag{24}$$

If we now observe the process over K mesoscopic time steps δt, corresponding to a total time $t = K\delta t$, the probability that it is in state n_k at time $k\delta t$ is given by

$$P_t[\{n_k\}] = \prod_{k=1}^{K} P_{\delta t}[n_k|n_{k-1}]. \tag{25}$$

Note that here, the index k corresponds to the time step rather than a chemical species, and every n_k is a full composition vector. The total transition rate between n_0 and n_K can then be decomposed over all possible paths,

$$P_t[n_K|n_0] = \sum_{k=1}^{K-1}\sum_{n_k=0}^{\infty} P_t[\{n_k\}] = \sum_{k=1}^{K-1}\sum_{n_k=0}^{\infty}\prod_{k=1}^{K} P_{\delta t}[n_k|n_{k-1}]. \tag{26}$$

1. Standard Lagrangian

For an appropriate scaling of time and volume such that $V\delta t \to \infty$, we expect each mesoscopic transition rate $P_{\delta t}[n_k|n_{k-1}]$ to have a large deviation form with a rate which we will call the *Lagrangian* of the process. More precisely, let us define

$$\mathscr{L}(\dot{\rho},\rho) = -\lim_{V\delta t\to\infty}\left[\frac{1}{V\delta t}\ln\left(P_{\delta t}[V\rho + V\delta t\dot{\rho}|V\rho]\right)\right]. \tag{27}$$

Equation (26) can then be rewritten, in the limit $V \to \infty$, $\delta t \to 0$, $V\delta t \to \infty$, $K \to \infty$, $K\delta t = $ cst, as a path integral,

$$P_t[\rho_t|\rho_0] \asymp \int \exp\left[-V\int_{\tau=0}^{t}\mathscr{L}(\dot{\rho}(\tau),\rho(\tau))d\tau\right]\mathscr{D}[\rho]. \tag{28}$$

Note that this scaling is significantly different from that used in Sec. II E, in that the concentrations ρ, and in particular, the initial and final states, are also scaled in the limit. This means that, even though the number of transitions occurring in a step δt is of order

$V\delta t \to \infty$, which is enough to guarantee the exponential scaling of the mesoscopic rates (cf. the Appendix), the relaxation time of the system becomes of order V, and the system does not have time to reach its stationary state.

The last ingredient missing from the picture is the boundary conditions. The previous expression is appropriate when starting and ending at fixed concentrations ρ_0 and ρ_t, but we may want instead to start from a certain initial distribution and trace over the final state with a certain cost function. Those boundary conditions have a crucial impact on the long-time behavior of our systems, as we will see in Sec. IV B. Let us therefore consider an initial distribution $P_0(\rho)$ and a final cost $\mathscr{O}_t(\rho)$ of the form

$$P_0(\rho) \asymp e^{-VU(\rho)} \quad \text{and} \quad \mathscr{O}_t(\rho) \asymp e^{-V\theta(\rho)}, \tag{29}$$

where U and θ are non-negative functions that vanish at one point at least (this can be guaranteed by normalizing them appropriately). Given that we are interested only in the exponential scaling of probabilities, we do not need to worry about the prefactors, but note that these expressions do include cases where P_0 and $\mathscr{O}_t$ do not scale exponentially with V: if the scaling is subexponential, then $U = 0$ or $\theta = 0$, and if the scaling is superexponential, they should be replaced by theta functions centered at their maxima.

Once those terms have been taken into account, the path integral representation of the process becomes

$$\langle\mathscr{O}_t\rangle_{P_0} \asymp \int \exp\left[-V\left(\int_{\tau=0}^{t}\mathscr{L}(\dot{\rho}(\tau),\rho(\tau))d\tau + U(\rho_0) + \theta(\rho_t)\right)\right]\mathscr{D}[\rho]. \tag{30}$$

2. Equations of motion

Considering that Eq. (30) is a sum over exponentials with a large exponent, the terms dominating the path integral correspond to the trajectories which minimize the rate of the exponential (or *action*). By taking a functional derivative of the action, we find the deterministic equations of motion describing the typical behavior of our system.

The differentiation of the action with respect to ρ_τ at every time τ gives

$$\int_{\tau=0}^{t}(\partial_\rho\mathscr{L}\delta\rho_\tau + \partial_{\dot\rho}\mathscr{L}\delta\dot\rho_\tau)d\tau + \partial_\rho U(\rho_0)\delta\rho_0 + \partial_\rho\theta(\rho_t)\delta\rho_t = 0. \tag{31}$$

An integration by parts of the second term yields

$$\int_{\tau=0}^{t}\left(\partial_\rho\mathscr{L} - \frac{d}{dt}\partial_{\dot\rho}\mathscr{L}\right)\delta\rho_\tau\,d\tau + [\partial_{\dot\rho}\mathscr{L}\delta\rho_\tau]_0^t$$
$$+ \partial_\rho U(\rho_0)\delta\rho_0 + \partial_\rho\theta(\rho_t)\delta\rho_t = 0. \tag{32}$$

Canceling the integrated term will yield the standard Euler-Lagrange equation,

$$\partial_\rho\mathscr{L} - \frac{d}{dt}\partial_{\dot\rho}\mathscr{L} = 0, \tag{33}$$

and the boundary terms fix the boundary conditions,

$$\partial_{\dot\rho}\mathscr{L}(\rho_0,\dot\rho_0) = \partial_\rho U(\rho_0) \quad \text{and} \quad \partial_{\dot\rho}\mathscr{L}(\rho_t,\dot\rho_t) = -\partial_\rho\theta(\rho_t). \tag{34}$$

3. Standard Hamiltonian and Hamilton's equations

Let us start once more from Eq. (26), and let us define a rescaled cumulant generating function (i.e., a log-Laplace transform) of $P_{\delta t}[n|n_i]$ with respect to $n - n_i$, with a parameter f_x conjugate to n_x, which we call the standard *Hamiltonian*,

$$\mathcal{H}(f, n_i) = \frac{1}{V\delta t} \ln\left(\sum_n e^{(n-n_i)f} \langle n|e^{\delta t W}|n_i\rangle\right). \tag{35}$$

Since the difference in state only depends on the integrated chemical currents through the continuity equation, this generating function can be rewritten as

$$\mathcal{H}(f, n_i) = \frac{1}{V\delta t} \ln\left(\sum_n \langle n|e^{\delta t W_{\nabla f,0}}|n_i\rangle\right), \tag{36}$$

with the biased generator $W_{\sigma,h}$ defined in Sec. II E taken at $h_x = 0$ and $\sigma_{\gamma'\gamma} = \nabla_{\gamma'\gamma} f = (\nu^{\gamma'} - \nu^{\gamma})f$.

In the large volume and small time step limit, with $n_i = V\rho$, this expression simply becomes

$$\mathcal{H}(f, \rho) = \sum_{\gamma,\gamma'} k_{\gamma'\gamma}\rho^{\nu^{\gamma}}\left(e^{(\nu^{\gamma'}-\nu^{\gamma})f} - 1\right), \tag{37}$$

where $\rho^{\nu^{\gamma}} = \prod \rho_x^{\nu_x^{\gamma}}$, as detailed in the Appendix. Moreover, the Laplace transform becomes a Legendre transform, which means that the Hamiltonian is related to the Lagrangian through

$$\mathcal{H}(f, \rho) = f\dot{\rho} - \mathcal{L}(\dot{\rho}, \rho) \quad \text{with} \quad f = \partial_{\dot{\rho}}\mathcal{L}, \tag{38}$$

hence its name. Note that $\mathcal{H}$ is explicit even though $\mathcal{L}$ is not in general.

The Euler-Lagrange equations can be recast in terms of the Hamiltonian. Let us consider Eq. (33), with $f = \partial_{\dot{\rho}}\mathcal{L}$. The Euler-Lagrange equation becomes

$$\partial_{\rho}\mathcal{L} = \frac{d}{dt}\partial_{\dot{\rho}}\mathcal{L} = \dot{f}. \tag{39}$$

Considering $\mathcal{H} = f\dot{\rho} - \mathcal{L}$ so that $\partial_{\rho}\mathcal{L} = -\partial_{\rho}\mathcal{H}$ and $\dot{\rho} = \partial_f \mathcal{H}$, the Euler-Lagrange equation simply becomes one of the standard Hamilton equations,

$$\dot{f} = -\partial_{\rho}\mathcal{H} \quad \text{with} \quad \dot{\rho} = \partial_f \mathcal{H}. \tag{40}$$

The boundary conditions simply translate to

$$f_0 = \partial_{\rho}U(\rho_0) \quad \text{and} \quad f_t = -\partial_{\rho}\theta(\rho_t), \tag{41}$$

which are implicit equations that define a manifold in phase space at each time. Let us also remark that the value of $\mathcal{H}$ is conserved by that dynamics,

$$\dot{\mathcal{H}} = \dot{\rho}\,\partial_{\rho}\mathcal{H} + \dot{f}\,\partial_f \mathcal{H} = 0. \tag{42}$$

Note that, in contrast with mechanical or quantum Hamiltonians, the "momentum" f does not correspond to a physical observable and cannot be interpreted as a velocity. Instead, it represents the *effective potential* of the force that the noise provides to the system.

4. Two-field picture

In the specific case of mass-action kinetics, we can transform the standard Hamiltonian $\mathcal{H}$ to a form closer to the famous Doi-Peliti action,[87,95,96] sometimes called Liouville functional,[106] which will be convenient for certain computations below.

Let us define two new variables (fields) $\psi_x = e^{-f_x}\rho_x$ and $\phi_x = e^{f_x}$ such that $\rho_x = \phi_x\psi_x$ and $f_x = \ln(\phi_x)$, which is to say that $\{\phi, \psi\} \to \{f, \rho\}$ is a Cole-Hopf transform. In terms of those variables, the Hamiltonian (37) becomes

$$\mathcal{H}(\phi_x, \psi_x) = \sum_{\gamma,\gamma'} k_{\gamma'\gamma}\psi^{\nu^{\gamma}}\left(\phi^{\nu^{\gamma'}} - \phi^{\nu^{\gamma}}\right) = \langle\phi^{\nu}|K|\psi^{\nu}\rangle, \tag{43}$$

where ψ^{ν} and ϕ^{ν} are vectors in the space of complexes and K is the Markov matrix containing the kinetic rates k. Moreover, using Eq. (40), the dynamics of those two variables is given by

$$\begin{aligned}
\dot{\phi}_x = \phi_x\dot{f}_x &= -\phi_x\sum_{\gamma,\gamma'}\frac{\nu_x^{\gamma}}{\rho_x}k_{\gamma'\gamma}\rho^{\nu^{\gamma}}\left(e^{(\nu^{\gamma'}-\nu^{\gamma})f} - 1\right) \\
&= -\sum_{\gamma,\gamma'}\frac{\nu_x^{\gamma}}{\psi_x}k_{\gamma'\gamma}\psi^{\nu^{\gamma}}\left(\phi^{\nu^{\gamma'}} - \phi^{\nu^{\gamma}}\right) \\
&= -\partial_{\psi_x}\mathcal{H},
\end{aligned} \tag{44}$$

$$\begin{aligned}
\dot{\psi}_x = \frac{\dot{\rho}_x}{\phi_x} - \psi_x\dot{f}_x &= \frac{1}{\phi_x}\sum_{\gamma,\gamma'}k_{\gamma'\gamma}\rho^{\nu^{\gamma}}\left(\nu_x^{\gamma'} - \nu_x^{\gamma}\right)e^{(\nu^{\gamma'}-\nu^{\gamma})f} \\
&\quad + \psi_x\sum_{\gamma,\gamma'}\frac{\nu_x^{\gamma}}{\rho_x}k_{\gamma'\gamma}\rho^{\nu^{\gamma}}\left(e^{(\nu^{\gamma'}-\nu^{\gamma})f} - 1\right) \\
&= \frac{1}{\phi_x}\sum_{\gamma,\gamma'}k_{\gamma'\gamma}\psi^{\nu^{\gamma}}\left(\nu_x^{\gamma'}\phi^{\nu^{\gamma'}} - \nu_x^{\gamma}\phi^{\nu^{\gamma}}\right) \\
&= \partial_{\phi_x}\mathcal{H},
\end{aligned} \tag{45}$$

which is to say that they obey the Hamilton equations with the appropriate Hamiltonian. The change of variables between $\{f, \rho\}$ and $\{\phi, \psi\}$ is therefore canonical. The corresponding boundary conditions become

$$\phi_0 = \exp\left[\psi_0^{-1}\partial_{\phi}U(\phi_0\psi_0)\right] \quad \text{and} \quad \phi_t = \exp\left[-\psi_t^{-1}\partial_{\phi}\theta(\phi_t\psi_t)\right]. \tag{46}$$

The usefulness of these so-called *two-field* variables $\{\phi, \psi\}$, as opposed to the *density-phase* variables $\{f, \rho\}$, is twofold. First, we may in certain cases use the algebraic structure of Eq. (43) to get information on the stationary states of the system;[107] for instance, if K is nonreducible, and under certain conditions on ν, we may use the Perron-Frobenius theorem on K to deduce that the process has a single stable fixed point. This is very specific to mass-action kinetics, as otherwise the Hamiltonian would not be a bilinear form of the complex weights $\{\psi^{\nu}, \phi^{\nu}\}$, although the change of variables would be canonical regardless. Second, this formalism turns out to be more appropriate for certain types of boundary conditions, as we will see in Sec. IV C.

Note that, unlike the standard Doi-Peliti formalism, all variables here are real-valued, although it makes no practical difference when dealing with the corresponding path integrals. Moreover, in the very special case where the chemical process is linear (i.e., each complex is made of a single element of a unique species), which

is essentially a Markov jump process, Eqs. (44) and (45) become decoupled, and the second one yields the master equation for that process.

B. Detailed formalism

In practice, the Lagrangian $\mathscr{L}(\dot{\rho}, \rho)$ is in general difficult to compute explicitly due to the potentially large number of trajectories resulting in the same $\dot{\rho}$. We will see in Sec. V that this can still be achieved for systems with a single chemical species,[108,109] but in general there is a more appropriate approach: rather than only keeping track of the variation of concentrations, we can differentiate the mesoscopic transition rates depending on the rate of each individual reaction. This corresponds to the so-called 2.5 *level* of large deviations,[6] where the highest possible mesoscopic detail is kept from microscopic trajectories. Several examples of the Lagrangians and Hamiltonians thus produced can be found in Ref. 110 in a different context.

Define $\hat{P}_{\delta t}[w_{\gamma'\gamma}|n_i]$ to be the probability, from n_i, to perform $w_{\gamma'\gamma}$ times reaction $\gamma \to \gamma'$ during a time δt. This rate is such that

$$P_{\delta t}[n_f|n_i] = \sum_{\gamma',\gamma} \sum_{w_{\gamma'\gamma}} \hat{P}_{\delta t}[w_{\gamma'\gamma}|n_i] \, \mathbb{I}(\nabla \cdot w = n_f - n_i), \quad (47)$$

where $\mathbb{I}$ is the indicator function (i.e., a Kronecker delta), used for typographic convenience.

1. Detailed Lagrangian

This new decomposition allows us to define a new large deviation function, which we will call the *detailed Lagrangian* of our process,

$$L(\lambda, \rho) = -\frac{1}{V\delta t} \ln\Big(\hat{P}_{\delta t}[V\delta t\lambda|V\rho]\Big). \quad (48)$$

Equation (26) can be rewritten as a new path integral, with the indicator function becoming a continuous delta function,

$$P_t[\rho_t|\rho_0] \asymp \int_{\rho_0} \exp\Big[-V\int_{\tau=0}^{t} L(\lambda(\tau), \rho(\tau))dt\Big] \delta(\dot{\rho} + \nabla \cdot \lambda) \, \mathscr{D}[\lambda], \quad (49)$$

and including the boundary conditions yields

$$\langle \mathscr{O}_t \rangle_{P_0} \asymp \int \exp\Big[-V\Big(\int_{\tau=0}^{t} L(\lambda(\tau), \rho(\tau))d\tau + U(\rho_0) + \theta(\rho_t)\Big)\Big] \times \delta(\dot{\rho} + \nabla \cdot \lambda) \, \mathscr{D}[\lambda]. \quad (50)$$

The advantage of this approach is that, unlike $\mathscr{L}$, the detailed Lagrangian L can always be computed explicitly, as shown in the Appendix. In the case of mass-action kinetics, we find

$$L(\lambda, \rho) = \sum_{\gamma,\gamma'} \lambda_{\gamma'\gamma} \ln(\lambda_{\gamma'\gamma}/k_{\gamma'\gamma}\rho^{\nu^{\gamma}}) - \lambda_{\gamma'\gamma} + k_{\gamma'\gamma}\rho^{\nu^{\gamma}}, \quad (51)$$

and we should note that the method can be as easily applied to different models and dynamics as long as the criteria detailed in the Appendix are met. This is the first important result of this paper. Notice the similarity with the long-time large deviation function of

currents and densities in a Markov jump process,[61] which is due to the fact that transition events have Poissonian distributions in both cases, as is demonstrated in the Appendix. In that respect, the large volume limit of chemical reaction networks is a *natural nonlinear version of Markov jump processes*, in the same way that interacting diffusions, as described by the macroscopic fluctuation theory,[2] are a natural nonlinear version of Fokker-Planck equations. It should be noted that the interactions at the origin of the nonlinearity are, in our case, of an entropic nature: the chemical components are not subject to any interaction potential, but the fact that reactions only occur under some geometrical constraints (namely, the presence of all reactants at the same place) produces that nonlinearity. It was brought to our attention that a similar result can be found in Ref. 111 in a case where the transitions are limited to the exchange of a single particle.

The standard Lagrangian (27) can then be obtained through the contraction formula

$$\mathscr{L}(\dot{\rho}, \rho) = \min_{\dot{\rho} = -\nabla \cdot \lambda} L(\lambda, \rho). \quad (52)$$

2. Equations of motion

Deriving the equations of motion for the detailed Lagrangian L requires more care, as we have to take the conservation relation into account. This is done by introducing a Lagrange multiplier $\xi_x(t)$ conjugate to the conserved quantity $(\dot{\rho} + \nabla \cdot \lambda)_x(t)$ so that ρ and λ become independent variables. The bulk minimization of the path integral yields three independent parts

$$\delta[L + \xi(\dot{\rho} + \nabla \cdot \lambda)] = [\partial_\rho L \, \delta\rho + \xi \, \delta\dot{\rho}] + [\partial_\lambda L \cdot \delta\lambda + \xi\nabla \cdot \delta\lambda] + [\dot{\rho} + \nabla \cdot \lambda]\delta\xi = 0. \quad (53)$$

Each part has to vanish on its own. We can do an integration by parts on time in the first term and on space in the second using Eq. (5), which gives two equations in addition to the conservation of matter from the third term,

$$\partial_\rho L = \dot{\xi} \quad \text{and} \quad \partial_\lambda L = \nabla \xi. \quad (54)$$

Combining the two yields, the detailed Euler-Lagrange equation

$$\nabla \partial_\rho L - \frac{d}{dt}\partial_\lambda L = 0. \quad (55)$$

Moreover, the boundary terms from the partial integration of $f \, \delta\dot{\rho}$, combined with the differentials of U_0 and θ_t, yield the boundary conditions

$$\xi_0 = \partial_\rho U(\rho_0) \quad \text{and} \quad \xi_t = -\partial_\rho \theta(\rho_t) \quad (56)$$

so that

$$\partial_\lambda L(\lambda_0, \rho_0) = \nabla \partial_\rho U(\rho_0) \quad \text{and} \quad \partial_\lambda L(\lambda_t, \rho_t) = -\nabla \partial_\rho \theta(\rho_t). \quad (57)$$

Note the similarity with the standard Hamiltonian boundary conditions (41). However, in this case, ξ is not the variable of the Hamiltonian, as we will see in Sec. III B.

Those equations of motion must be verified by the typical trajectories of the system, in particular, if we start from a certain concentration ρ_0 and simply trace over the final concentration by taking $\theta_t = 0$. Since we have an explicit expression for L, we could

inject it in (55), but that is not necessary: we know that L is positive by construction and cancels only at

$$\lambda_{\gamma'\gamma} = k_{\gamma'\gamma}\rho^{\nu^{\gamma'}}. \tag{58}$$

This, combined with the conservation equation, yields the standard system of coupled first order differential equations describing the evolution of a dilute well-mixed chemical solution with mass-action kinetics,

$$\dot{\rho}_x = \sum_{\gamma,\gamma'} (\nu_x^{\gamma'} - \nu_x^{\gamma}) k_{\gamma'\gamma}\rho^{\nu^{\gamma'}}. \tag{59}$$

Moreover, since this solution gives a value for $\dot{\rho}$ that only depends on ρ, this is also the solution of Eq. (33). This proves that, as expected (and even required for consistency), the most likely evolution of the chemical master equation in the large volume limit is given by the deterministic mass-action rate equations.

3. Detailed Hamiltonian and Hamilton's detailed equations

As with the Lagrangian, we can also define a *detailed Hamiltonian*, starting from Eq. (47) and introducing a variable $\sigma_{\gamma'\gamma}$ conjugate to $w_{\gamma'\gamma}$,

$$H(\sigma, n) = \frac{1}{V\delta t} \ln\left(\sum_{\gamma',\gamma} e^{\sigma_{\gamma'\gamma}w_{\gamma'\gamma}}\hat{P}_{\delta t}[w_{\gamma'\gamma}|n_i]\right), \tag{60}$$

which, as above, can be rewritten as

$$H(\sigma, n) = \frac{1}{V\delta t} \ln\left(\sum_{n} \langle n|e^{\delta t W_{\sigma,0}}|n_i\rangle\right), \tag{61}$$

which we recognize as a generalization of the standard Hamiltonian, where the *virtual forces* are not necessarily of the gradient form. We use here the letter σ instead of $\mathfrak{s}$ for the variable conjugate to λ, in order to avoid confusion later when we introduce both a bias and a conjugation on λ simultaneously. Even though those variables occupy the same place in the definition of the Hamiltonians, they represent different physical objects.

In the large volume and small time step limit, with $n_i = V\rho$ and $w = V\delta t\lambda$, this expression simply becomes

$$H(\sigma, \rho) = \sum_{\gamma,\gamma'} k_{\gamma'\gamma}\rho^{\nu^{\gamma'}}\left(e^{\sigma_{\gamma'\gamma}} - 1\right), \tag{62}$$

and we can check that H is indeed the Legendre transform of L,

$$H(\sigma, \rho) = \sigma \cdot \lambda - L(\lambda, \rho) \quad \text{with} \quad \sigma_{\gamma'\gamma} = \partial_{\lambda_{\gamma'\gamma}}L. \tag{63}$$

The equivalent of the contraction formula (52) for Hamiltonians is a simple reduction of variables,

$$\mathcal{H}(f, \rho) = H(\nabla f, \rho), \tag{64}$$

which is one reason to favor Hamiltonians over Lagrangians for explicit computations.

Note that this contraction is compatible with the continuity equation $\dot{\rho} = -\nabla \cdot \lambda$ as setting $\sigma = \nabla f$ leads to

$$\sigma \cdot \lambda = -f\nabla \cdot \lambda = f\dot{\rho}, \tag{65}$$

which is to say that the Legendre scalar $\sigma \cdot \lambda$ is conserved through its contraction to $f\dot{\rho}$.

Starting from Eq. (55), and defining $\sigma = \partial_\lambda L$, we find

$$\nabla \partial_\rho L = \frac{\mathrm{d}}{\mathrm{d}t}\partial_\lambda L = \dot{\sigma}. \tag{66}$$

Defining $H = \sigma \cdot \lambda - L$ so that $\partial_\rho L = -\partial_\rho H$ and $\dot{\rho} = -\nabla \cdot \partial_\sigma H$, we get the detailed Hamilton equations,

$$\dot{\sigma} = -\nabla \partial_\rho H \quad \text{with} \quad \dot{\rho} = -\nabla \cdot \partial_\sigma H. \tag{67}$$

The boundary conditions become

$$\sigma_0 = \nabla \partial_\rho U(\rho_0) \quad \text{and} \quad \sigma_t = -\nabla \partial_\rho \theta(\rho_t). \tag{68}$$

The value of H is also a constant of motion

$$\dot{H} = \dot{\rho}\,\partial_\rho H + \dot{\sigma}\,\partial_\sigma H = -(\nabla \cdot \partial_\sigma H)\partial_\rho H - (\nabla \partial_\rho H)\partial_\sigma H = 0 \tag{69}$$

through the integration by parts of one of the terms.

All of the equations of this section simply reduce to those for the standard Hamiltonian by setting $\sigma = \nabla f$, which is compatible with the boundary conditions. σ can be interpreted as the *random force* produced by the noise.

C. Biased formalism

The aim of this work is ultimately to describe the large deviations of our systems when conditioned on certain values of the currents λ and densities ρ. This can be formally achieved by constraining all the equations of Sec. III B 3 on the time-average of those observables. However, as we saw in Sec. II E, it can be preferable in practice to look at generating functions instead. The procedure of adding biases to the microscopic process to define those generating functions is quite standard,[5] but the outcome in terms of a biased process at the mesoscopic level (i.e., in terms of the hydrodynamic variables ρ and λ) and the properties of its dynamics are nontrivial. We will investigate some of those properties in Sec. IV, but first we should describe the formalism itself.

1. Detailed biased dynamics

As shown in the Appendix, the generating function of the cumulants of those observables can be obtained by replacing W by $W_{\mathfrak{s},h}$, as defined in Eq. (23), in every term of Eq. (26). This simply leads to an equation of the exact same form as (50) but with L replaced by a new *biased* Lagrangian

$$L_{\mathfrak{s},h}(\lambda, \rho) = L(\lambda, \rho) - \mathfrak{s} \cdot \lambda - h\rho, \tag{70}$$

i.e.,

$$\langle \mathcal{O}_t e^{tV(\mathfrak{s}\cdot\lambda + h\rho)} \rangle_{P_0}$$
$$\asymp \int \exp\left[-V\left(\int_{\tau=0}^{t} L_{\mathfrak{s},h}(\lambda(\tau), \rho(\tau))\mathrm{d}\tau + U(\rho_0) + \theta(\rho_t)\right)\right]$$
$$\times \delta(\dot{\rho} + \nabla \cdot \lambda)\,\mathscr{D}[\lambda]. \tag{71}$$

The detailed Euler-Lagrange equation (55) still applies directly to the biased Lagrangian $L_{\mathfrak{s},h}$, from the same calculations. Those equations can be recast in terms of the detailed Lagrangian L as

$$\partial_\rho L = \dot{\xi} + h \quad \text{and} \quad \partial_\lambda L = \nabla\xi + \mathfrak{s}. \tag{72}$$

Combining the two, for $\mathfrak{s}$ constant in time, yields the biased Euler-Lagrange equation,

$$\nabla\partial_\rho L - \frac{d}{dt}\partial_\lambda L = \nabla h. \tag{73}$$

Note that the dependence on $\mathfrak{s}$ is not apparent from this equation but can be seen in the boundary conditions,

$$\partial_\lambda L_{\mathfrak{s},h}(\lambda_0,\rho_0) = \nabla\partial_\rho U(\rho_0) \quad \text{and} \quad \partial_\lambda L_{\mathfrak{s},h}(\lambda_t,\rho_t) = -\nabla\partial_\rho\theta(\rho_t), \tag{74}$$

which become, in terms of L,

$$\partial_\lambda L(\lambda_0,\rho_0) = \mathfrak{s} + \nabla\partial_\rho U(\rho_0) \quad \text{and} \quad \partial_\lambda L(\lambda_t,\rho_t) = \mathfrak{s} - \nabla\partial_\rho\theta(\rho_t). \tag{75}$$

Defining the biased Hamiltonian is straightforward, as biasing the microscopic dynamics W is an associative operation: $(W_{\mathfrak{s},h})_{\sigma,0} = W_{\mathfrak{s}+\sigma,h}$. Through the computations shown in the Appendix, we find

$$H_{\mathfrak{s},h}(\sigma,\rho) = H(\sigma + \mathfrak{s},\rho) + h\rho, \tag{76}$$

which we can check to be the Legendre transform of $L_{\mathfrak{s},h}$,

$$H_{\mathfrak{s},h}(\sigma,\rho) = \sigma \cdot \lambda - L_{\mathfrak{s},h}(\lambda,\rho) \quad \text{with} \quad \sigma_{\gamma'\gamma} = \partial_{\lambda_{\gamma'\gamma}}L_{\mathfrak{s},h}. \tag{77}$$

The Hamilton equations for $H_{\mathfrak{s},h}$ are the same as those for the detailed Hamiltonian and can be recast as

$$\dot{\sigma} = -\nabla\partial_\rho H(\sigma + \mathfrak{s},\rho) - \nabla h \quad \text{with} \quad \dot{\rho} = -\nabla \cdot \partial_\sigma H(\sigma + \mathfrak{s},\rho) \tag{78}$$

with the same boundary conditions

$$\sigma_0 = \nabla\partial_\rho U(\rho_0) \quad \text{and} \quad \sigma_t = -\nabla\partial_\rho\theta(\rho_t). \tag{79}$$

The value of $H_{\mathfrak{s},h}$ is a constant of motion for the same reason as before. Note that, in the case where $h = 0$, the value of H is also constant.

2. Contraction of the currents and two-field picture

We saw earlier that the detailed Lagrangian has the advantage of being explicit, which the standard Lagrangian usually is not. It has however a serious disadvantage in practice: the set of currents λ is of much higher dimension than the set of velocities $\dot{\rho}$. This also applies to the conjugate quantities f and σ, although all Hamiltonians are in fact explicit, which makes $\mathscr{H}$ preferable to H. The latter is useful in defining the biased Hamiltonian $H_{\mathfrak{s},h}$, but for the same reason, it makes sense to define a contracted biased Hamiltonian $\mathscr{H}_{\mathfrak{s},h}(f,\rho)$ with fewer variables. Note that it is possible to define $\mathscr{H}_{\mathfrak{s},h}$ directly from the microscopic dynamics, without going through $H_{\mathfrak{s},h}$, as is done in Ref. 85 in the case of a bias on densities and dynamical activity (see, in particular, the Appendix).

Let us formally define a contracted biased Lagrangian $\mathscr{L}_{\mathfrak{s},h}$ by contraction of $L_{\mathfrak{s},h}$,

$$\mathscr{L}_{\mathfrak{s},h}(\dot{\rho},\rho) = \min_{\dot{\rho}=-\nabla\cdot\lambda} L_{\mathfrak{s},h}(\lambda,\rho). \tag{80}$$

This cannot be written explicitly in general, but this leads to the following contraction of the Hamiltonians

$$\mathscr{H}_{\mathfrak{s},h}(f,\rho) = H_{\mathfrak{s},h}(\nabla f,\rho) = H(\mathfrak{s} + \nabla f,\rho) + h\rho \tag{81}$$

so that

$$\mathscr{H}_{\mathfrak{s},h}(f,\rho) = \sum_{\gamma,\gamma'} k_{\gamma'\gamma}\rho^{\nu^\gamma}\left(e^{(\nu^{\gamma'}-\nu^\gamma)f+\mathfrak{s}_{\gamma'\gamma}} - 1\right) + \sum_x h_x\rho_x. \tag{82}$$

Note that this contraction is compatible with the previous Hamilton equations by setting $\sigma = \nabla f$ and removing ∇ (remember that, as stated in Sec. II C, we are implicitly working in a space where ∇ is invertible on the concentration side), which simply yields

$$\dot{f} = -\partial_\rho H(\nabla f + \mathfrak{s},\rho) - h = -\partial_\rho\mathscr{H}_{\mathfrak{s},h} \quad \text{with} \quad \dot{\rho} = \partial_f H(\nabla f + \mathfrak{s},\rho) = \partial_f\mathscr{H}_{\mathfrak{s},h} \tag{83}$$

with boundary conditions

$$f_0 = \partial_\rho U(\rho_0) \quad \text{and} \quad f_t = -\partial_\rho\theta(\rho_t). \tag{84}$$

As always, the value of this Hamiltonian is constant along the solutions of those equations.

As in Sec. III A 4, we can define $\psi_x = e^{-f_x}\rho_x$ and $\phi_x = e^{f_x}$ such that the Hamiltonian becomes

$$\mathscr{H}_{\mathfrak{s},h}(\phi_x,\psi_x) = \sum_{\gamma,\gamma'} k_{\gamma'\gamma}\psi^{\nu^\gamma}\left(e^{\mathfrak{s}_{\gamma'\gamma}}\phi^{\nu^{\gamma'}} - \phi^{\nu^\gamma}\right) + \sum_x h_x\phi_x\psi_x$$
$$= \langle\phi^\nu|K_{\mathfrak{s}}|\psi^\nu\rangle + \langle\phi|h|\psi\rangle, \tag{85}$$

where $K_{\mathfrak{s}}$ is the biased Markov matrix containing the kinetic rates k. Through the same calculations as before, we can check the Hamilton equations for those variables,

$$\dot{\phi}_x = -\partial_{\psi_x}\mathscr{H}_{\mathfrak{s},h} \quad \text{and} \quad \dot{\psi}_x = \partial_{\phi_x}\mathscr{H}_{\mathfrak{s},h}, \tag{86}$$

with boundary conditions

$$\phi_0 = \exp\left[\psi_0^{-1}\partial_\phi U(\phi_0\psi_0)\right] \quad \text{and} \quad \phi_t = \exp\left[-\psi_t^{-1}\partial_\phi\theta(\phi_t\psi_t)\right]. \tag{87}$$

Here again, for a linear process, Eqs. (86) decouple and can be identified as the Feynman-Kač equations for the time-dependent generating functions of cumulants of our observables.[7,112] The vectors ϕ and ψ solving the equations are then the left and right eigenvectors of the biased Markov matrix of the corresponding Markov jump process. The two-field picture is the standard formalism in that case: the equations on $\{f,\rho\}$ are still coupled and therefore more complicated.

IV. STATIONARY LARGE DEVIATIONS AND DYNAMICAL PHASE TRANSITIONS

Now that the formalism has been set up, we can examine what happens in the long-time limit in order to obtain the SCGF $E(\mathfrak{s},h)$, as defined in Sec. II E but with a term V factored out,

$$\left\langle e^{tV(\mathfrak{s}\cdot\lambda+h\rho)}\right\rangle_{P_0} \sim e^{tVE(\mathfrak{s},h)} \quad \text{for} \quad t \to \infty, \tag{88}$$

without having to solve the microscopic eigenvalue problem, which is potentially infinite-dimensional.

As we saw in Sec. III C, this expression is a special case of the path integral (71) with $\theta_t = 0$, which we may write in terms of $\mathscr{L}_{\mathfrak{s},h}$ instead, and is dominated by the term solving the corresponding Euler-Lagrange equations. The value of the dominating term is then given by

$$\left\langle e^{tV(\mathfrak{s}\cdot\lambda+h\rho)} \right\rangle_{P_0} \sim \exp\left[-V\left(\int_{\tau=0}^{t} \mathscr{L}_{\mathfrak{s},h}(\dot{\rho}(\tau),\rho(\tau))\mathrm{d}\tau + U(\rho_0) \right) \right] \tag{89}$$

along the solution. Rewriting this in terms of the Hamiltonian, we get

$$\left\langle e^{tV(\mathfrak{s}\cdot\lambda+h\rho)} \right\rangle_{P_0} \sim \exp\left[V\left(\int_{\tau=0}^{t} (\mathscr{H}_{\mathfrak{s},h}(f(\tau),\rho(\tau)) - f\dot{\rho})\mathrm{d}\tau - U(\rho_0) \right) \right] \tag{90}$$

along the solution of the Hamilton equations $\{\rho^\star(\tau), f^\star(\tau)\}$. If the solution is not unique, the one which gives the largest value of the argument of the exponential dominates.

We saw that the value of $\mathscr{H}_{\mathfrak{s},h}$ is conserved along this solution so that it comes out of the time integral. Moreover, the term $U(\rho_0)$ is not extensive in time so that it is negligible in the long time limit, although it does appear implicitly through $f^\star$ and $\rho^\star$ which depend on it. We are left with the *kinetic term*

$$\mathcal{K}(f^\star, \rho^\star) = \int_{\tau=0}^{t} f^\star \dot{\rho}^\star \, \mathrm{d}\tau, \tag{91}$$

which may or may not be extensive in time. Assuming that it is not (which we will show to be true if all attractors of the deterministic dynamics are *fixed points*, in Sec. IV B), we finally get

$$\left\langle e^{tV(\mathfrak{s}\cdot\lambda+h\rho)} \right\rangle_{P_0} \sim \exp\left[tV\mathscr{H}_{\mathfrak{s},h}(f^\star, \rho^\star) \right]. \tag{92}$$

This means that, for systems with only fixed points, the SCGF $E(\mathfrak{s}, h)$ is equal to the largest value of the biased Hamiltonian $\mathscr{H}_{\mathfrak{s},h}$ along solutions of the Hamilton equations. In cases where the kinetic term is extensive in time (e.g., along a limit cycle), it must be included in the maximization.

In the rest of this section, we will describe general features of those solutions including their attractors and the consequences of choosing different initial conditions. We will then show that *multistable* systems (i.e., that have more than one attractor) generically undergo a first-order dynamical phase transition when the bias crosses 0. This will be proven rigorously when all the attractors are fixed points and left as a conjecture for more complex cases.

A. Boundary conditions

The choice of boundary conditions can be a determining factor in how the system under consideration fluctuates. Since we are mostly interested in the SCGF defined above, we will restrict ourselves to a flat final condition, i.e.,

$$\theta(\rho_t) = 0 \quad \text{so that} \quad f_t = 0 \quad \text{or} \quad \phi_t = 1 \tag{93}$$

depending on which are the appropriate variables.

Let us remark that the case of a fixed final condition ρ_t is also of special interest, as it allows us to define the quasipotential of the process[10,113] and gives access to its stationary measure. Moreover, if both initial and final conditions are set to be densities of fixed points or other critical manifolds of the dynamics, the optimal path connecting them is called an *instanton* and is of particular importance to estimate transition times between stationary

states of multistable systems[114,115] or extinction times for metastable populations.[109,116]

For the initial condition, one possibility is to start from a fixed concentration ρ_i. In this case, the initial condition for the Hamiltonian trajectory is simply

$$\rho_0 = \rho_i \quad \text{or} \quad \phi_0 = \frac{\rho_i}{\psi_0} \tag{94}$$

and the density-phase variables $\{f, \rho\}$ are more appropriate, as they make the initial condition explicit.

Another popular choice is to start from a multi-Poisson distribution with average $\bar{\rho}$,

$$P_0(\rho) = \prod_x \frac{(V\bar{\rho}_x)^{V\rho_x}}{[V\rho_x]!} e^{-V\bar{\rho}_x} \quad \text{so that} \quad U(\rho) = \sum_x \rho_x \ln\left(\frac{\rho_x}{\bar{\rho}_x}\right) - \rho_x + \bar{\rho}_x, \tag{95}$$

which is to say that U is a relative entropy between ρ and a reference $\bar{\rho}$. In this case, the initial conditions become

$$(f_0)_x = \ln\left(\frac{\rho_x}{\bar{\rho}_x}\right) \quad \text{or} \quad (\psi_0)_x = \bar{\rho}_x \tag{96}$$

in which case the two-field variables $\{\phi, \psi\}$ become preferable.

B. Density-phase picture and global Hamiltonian attractors

Let us first examine the behavior of long-time trajectories in the density-phase picture,

$$\dot{f} = -\partial_\rho \mathscr{H}_{\mathfrak{s},h} \quad \text{and} \quad \dot{\rho} = \partial_f \mathscr{H}_{\mathfrak{s},h} \tag{97}$$
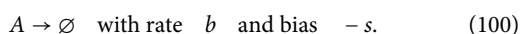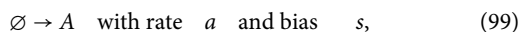
with the boundary conditions

$$\rho_0 = \rho_i \quad \text{and} \quad f_t = 0. \tag{98}$$

A first important remark that has to be made is that the dynamics is Hamiltonian, i.e., the volume of any subset of phase-space is conserved by the dynamics. If our boundary conditions were all at the initial or final time, that would mean that trajectories cannot converge, i.e., that no stable attractor exists, because that would require a contraction of phase-space during convergence. In other terms, all critical manifolds (fixed points, critical cycles, or stranger critical manifolds) of a Hamiltonian dynamics have either marginal or mixed stability (i.e., as many stable directions as unstable ones) and cannot be attractors if one starts from a complete initial condition (f_i, ρ_i).

However, in our case, we have by construction *mixed boundary conditions*: half of the boundary conditions are at the initial time and the other half at the final time. This allows trajectories to converge toward a critical manifold of the Hamiltonian as long as the initial condition crosses its stable manifold and the final condition crosses its unstable manifold. We will say that the critical manifold is *dynamically connected* to the boundary conditions, and we will use this term in general to mean that there exists a trajectory (heterocline) from one manifold to another. These types of problems are sometimes called *two-point Hamiltonian problems* in the context of optimal control,[117] although there usually both boundary conditions are for the density.

To illustrate this point, let us consider an extremely simple model with one biased reaction,

$$\varnothing \to A \quad \text{with rate} \quad a \quad \text{and bias} \quad s, \tag{99}$$

$$A \to \varnothing \quad \text{with rate} \quad b \quad \text{and bias} \quad -s. \tag{100}$$

The corresponding Hamiltonian is

$$\mathscr{H}(f,\rho) = a(e^{f+s} - 1) + b\rho(e^{-f-s} - 1) \tag{101}$$

with fixed point $\{f = -s, \rho = a/b\}$. The stable manifold of this fixed point is $f = -s$, which intersects with any initial condition ρ_i. The unstable manifold is $\rho = e^{f+s}a/b$, which intersects with the final condition $f = 0$.

In Fig. 1, on the left, we illustrate how the trajectory solving the Hamilton equations with those boundary conditions converges to the fixed point at long times, from blue to green to red. The initial condition is the vertical purple dashed line, the final condition is the horizontal orange dashed line, the fixed point is marked in bright green (at a density value $\rho^\star$), and its stable and unstable manifolds are marked in thick black lines. Note that the time that each trajectory takes is not a simple function of its length in phase-space, but the only way to produce a long trajectory is to get closer to the fixed point, where the velocities vanish. As a result, the bulk of the trajectory converges to the fixed point, with only finite-time boundary layers connecting it to both boundary conditions. This convergence is guaranteed purely by the boundary conditions, and the critical point can be rightly called an attractor. The plots below show the evolution of $\rho(t)$ for those same trajectories, with the gray one representing the asymptotic limit for infinite time (where the middle horizontal section should in fact be infinitely long). Each trajectory shows a plateau at a value that converges to $\rho^\star$, of increasing length, and two boundary layers whose shapes also converge and are bounded between ρ_i and the maximal value ρ_m that ρ takes on the unstable manifold when $f = 0$.

Similarly, we can illustrate the case of an *unstable* fixed point, which occurs when the initial condition *never* crosses the stable

manifold or the final condition *never* crosses the unstable manifold, which we can obtain by flipping the directions of all trajectories in the previous case (i.e., by inverting time). In that case, the long-time trajectories diverge to infinity in some direction. As we see in the figure below, on the right, if the unstable manifolds of the fixed point (red circle) are parallel to the final condition (horizontal orange dashed line), the trajectory for long times will diverge in that direction. In the figure below, we represent once again the evolution of $\rho(t)$. In this case, there is no plateau of increasing length, and ρ is not bounded but diverges upwards. Note however that this case is particularly unlikely to occur globally (i.e., without another stable attractor to catch the trajectory) and requires very specific symmetries of the Hamiltonian, as those manifolds have to be exactly parallel to the corresponding boundary condition. We will see an example of this in Sec. V C.

We may remark that, while close to a fixed point, the kinetic term of Eq. (90) vanishes because $\dot{\rho} \sim 0$, which justifies the assumption that it is negligible. Whether it remains true for more complex attractors remains to be determined.

Given these considerations, we may state the main result of this paper: the SCGF $E(\mathfrak{s}, h)$ is equal to the largest value that the biased Hamiltonian $\mathscr{H}_{\mathfrak{s},h}$ takes at *attractors* of the mixed-boundary condition problem, defined as the critical manifolds that are dynamically connected to the boundary conditions, if they exist. In case the kinetic term (91) does not vanish at all those attractors, its average value should be added to the value of the Hamiltonian. The consequences of this are left for future investigation.

C. Conserved quantities in the two-field picture

We mentioned in Sec. II C that chemical systems may have conservation laws: starting from a certain concentration ρ, the dynamics may only span a subset of phase-space (i.e., one of many ergodic components). This is not an issue in the density-phase picture, as the solution is simply to restrict the definition space of ρ to the ergodic component containing the initial condition. This is not so in the two-field picture.
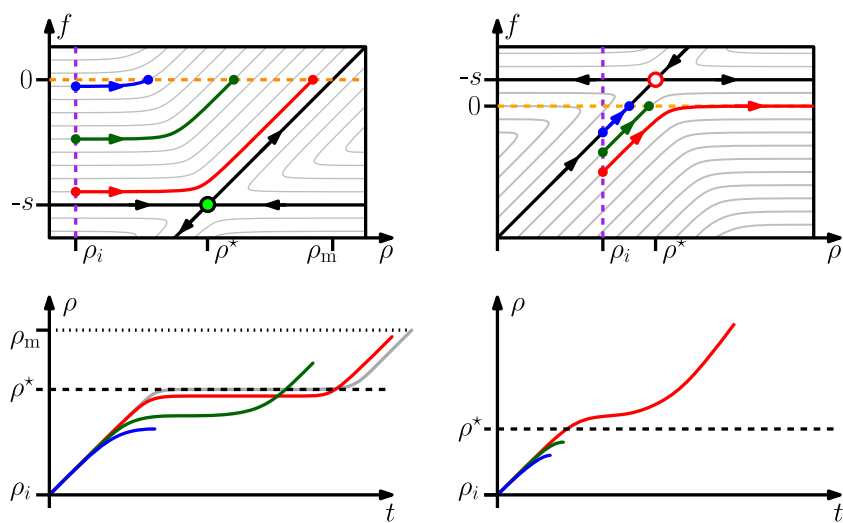

FIG. 1. Top: sketch of Hamiltonian trajectories close to a fixed point, for increasing runtimes (blue to green to red). Bottom: evolution of $\rho(t)$ for the same trajectories. Left: the fixed point is stable. Right: the fixed point is unstable.

Let us assume that the system we are considering has several conserved quantities $\mathscr{C}^\alpha = \mathfrak{c}^\alpha \rho$ such that

$$d_t \mathscr{C}^\alpha = d_t \mathfrak{c}^\alpha \rho = -\mathfrak{c}^\alpha \nabla \cdot \lambda = 0 \qquad (102)$$

in the Lagrange frame. Let us also place ourselves in the full space of concentrations ρ_x, in anticipation of the fact that we will start from an initial condition that spans all space.

In the Hamiltonian frame, this has several consequences. First of all, using Eq. (78), the previous equation expectedly becomes

$$\mathfrak{c}^\alpha \dot{\rho} = -\mathfrak{c}^\alpha \nabla \cdot \partial_\sigma H_{\mathfrak{s},h}(\sigma, \rho) = 0 \quad \text{with} \quad \sigma = \nabla f. \qquad (103)$$

Moreover, when going from (78) to (83), we can no longer simply invert ∇, but we have to instead consider an arbitrary integration constant which is a linear combination of the conservation laws $\mathfrak{c}^\alpha$,

$$\nabla \dot{f} = -\nabla \partial_\rho H_{\mathfrak{s},h}(\nabla f, \rho) \quad \Rightarrow \quad \dot{f} = -\partial_\rho H_{\mathfrak{s},h}(\nabla f, \rho) + A_\alpha \mathfrak{c}^\alpha, \qquad (104)$$

where the coefficients A_α are entirely unconstrained and can even depend on f, ρ, or t. Both of those statements boil down to the fact that the Hamiltonian

$$\mathscr{H}_{\mathfrak{s},h}(f, \rho) = H_{\mathfrak{s},h}(\nabla f, \rho) \qquad (105)$$

is invariant under $f \rightarrow f + \mathfrak{c}^\alpha$ for every α. Note that this is a gauge transformation and not a canonical change of variables: the equation of motion for $\dot{f}$ explicitly depends on A_α.

This underdetermination of the equations of motion is not actually an issue in the density-phase picture since we may simply use ∇f instead of f to determine the existence of attractors. Doing the same in the two-field picture would be particularly cumbersome since the variable f has been split between ϕ and ψ. We may however solve the issue in a different way in the case where we expect our dominant attractors (those with the highest value of the Hamiltonian) to be fixed points. Assume that we know the stationary values $\rho^\star$ and $\nabla f^\star$ of ρ and ∇f for a specific value of the conserved quantities $\{\mathscr{C}^\alpha\}$. We then have

$$\nabla \dot{f}^\star = -\nabla \partial_\rho H_{\mathfrak{s},h}(\nabla f^\star, \rho^\star) = 0, \qquad (106)$$

which means that $\partial_\rho H_{\mathfrak{s},h}(\nabla f, \rho)$ is a linear combination of the conservation laws $\mathfrak{c}^\alpha$. We can therefore choose A^α uniquely for each choice of $\mathscr{C}^\alpha$ such that

$$A_\alpha \mathfrak{c}^\alpha = \partial_\rho H_{\mathfrak{s},h}(\nabla f^\star, \rho^\star). \qquad (107)$$

This, in turn, implies that

$$\dot{f} = -\partial_\rho H_{\mathfrak{s},h}(\nabla f, \rho) + A_\alpha \mathfrak{c}^\alpha = 0 \quad \text{when} \quad \rho = \rho^\star \ , \ \nabla f = \nabla f^\star, \qquad (108)$$

which means that any compatible value of $\{f, \rho\}$ is a fixed point. If we now consider that we span all possible values of $\mathscr{C}^\alpha$, every corresponding A_α and resulting stationary value of $H_{\mathfrak{s},h}$ is accessible, and the dominant one is the largest one, as a function of A_α.

The final step is to translate those equations in the two-field picture. Since the only difference is in $\dot{f}$, considering the derivation in Sec. III A 4, we simply get

$$\dot{\psi}_x = \partial_{\phi_x} \mathscr{H}_{\mathfrak{s},h} - A_\alpha \mathfrak{c}_x^\alpha \psi_x = 0 \quad \text{and} \quad \dot{\phi}_x = -\frac{1}{\psi_x} \partial_{\phi_x} \mathscr{H}_{\mathfrak{s},h} + A_\alpha \mathfrak{c}_x^\alpha \phi_x = 0. \qquad (109)$$

In conclusion, we arrive at a second formulation of our main result, in the case where all attractors are fixed points: the SCGF $E(\mathfrak{s}, h)$ is equal to the largest value that the biased Hamiltonian $\mathscr{H}_{\mathfrak{s},h}$ takes at solutions of

$$\frac{1}{\psi_x} \partial_{\phi_x} \mathscr{H}_{\mathfrak{s},h} = \frac{1}{\phi_x} \partial_{\psi_x} \mathscr{H}_{\mathfrak{s},h} = -A_\alpha \mathfrak{c}^\alpha, \qquad (110)$$

where the factors A_α are unknowns of the equations.

Note that once more, in the special case of a linear chemical network, where every complex is a different species (or a different conformation of a single molecule) and where the only conserved quantity is the total mass (i.e., $\mathfrak{c} = 1$), those equations translate precisely as eigenvalue equations for $K_{\mathfrak{s},h}$, and we recover the well-known result by Donsker and Varadhan[99–102] mentioned in Sec. II E, of which this result is therefore a generalization.

In order to illustrate this rather formal result, let us look at a simple example. Consider the reversible reaction $B \leftrightarrow 2C$ with rates k_{21} for the forward reaction and k_{12} for the backward reaction. This system as one conservation law $\mathfrak{c} = [2, 1]$, i.e., $\mathscr{C}^0 = 2\rho_b + \rho_c$ is constant in time and is simple enough that we know it has a unique fixed point for any bias.

The biased Hamiltonian for that system is given by

$$\mathscr{H}_{\mathfrak{s},h}(f, \rho) = k_{21} \rho_b (e^{2f_c - f_b + s_{21}} - 1) + k_{12} \rho_c^2 (e^{f_b - 2f_c + s_{12}} - 1) + h_b \rho_b + h_c \rho_c. \qquad (111)$$

This Hamiltonian is invariant under any shift of f by a vector $[2X, X]$ for any real number X, which is straightforward to check.

We then define the two-field variables $\psi_x = \rho_x e^{-f_x}$ and $\phi_x = e^{f_x}$ and write the two-field version of the Hamiltonian,

$$\mathscr{H}_{\mathfrak{s},h}(\phi, \psi) = k_{21} \psi_b (\phi_c^2 e^{s_{21}} - \phi_b) + k_{12} \psi_c^2 (\phi_b e^{s_{12}} - \phi_c^2) + h_b \psi_b \phi_b + h_c \psi_c \phi_c. \qquad (112)$$

The fixed point solution to Hamilton's equations must then verify (110),

$$k_{21}(\phi_c^2/\phi_b e^{s_{21}} - 1) + h_b = 2X = k_{12} e^{s_{12}} \psi_c^2/\psi_b - k_{21} + h_b,$$

$$2k_{12} \psi_c(\phi_b/\phi_c e^{s_{12}} - \phi_c) + h_c = X = 2\phi_c(k_{21} e^{s_{21}} \psi_b/\psi_c - k_{12} \psi_c) + h_c. \qquad (113)$$

From the first equation, we find that

$$k_{21} e^{s_{21}} \phi_c^2/\phi_b = k_{12} e^{s_{12}} \psi_c^2/\psi_b = 2X - h_b + k_{21}, \qquad (114)$$

which we can plug into the second to get

$$\phi_c \psi_c = \frac{X - h_c}{2k_{12}} \left(\frac{k_{21}e^{s_{12}+s_{21}}}{2X - h_b + k_{21}} - 1 \right)^{-1} \quad \text{and subsequently} \quad \phi_b \psi_b = \frac{k_{12}k_{21}e^{s_{12}+s_{21}}}{(2X - h_b + k_{21})^2} \phi_c^2 \psi_c^2, \tag{115}$$

which, put together, allows us to compute the value of $\mathcal{H}_{\mathfrak{s},h}$ in terms of X,

$$\frac{(X - h_c)\big(2X(X - h_c)(2X - h_b + k_{21}) - (2X - h_b + k_{21} - k_{21}e^{s_{12}+s_{21}})(2X(X - h_c) + (X + h_c)(2X - h_b + k_{21}))\big)}{4k_{21}(2X - h_b + k_{21} - k_{21}e^{s_{12}+s_{21}})^2}.$$

At this stage, we can either fix the value of X by choosing a value for $\mathcal{C}^0 = 2\rho_b + \rho_c$, thus yielding the microcanonical value of $E(\mathfrak{s}, h)$ on a given sector of the conserved quantity, or maximize $\mathcal{H}_{\mathfrak{s},h}$ with respect to X and deduce the optimal value of $\mathcal{C}^0$ from it, which yields the canonical value of $E(\mathfrak{s}, h)$. In both cases, the average values of ρ and λ can be computed through the first derivatives of $E(\mathfrak{s}, h)$.

Note that all values of X are not necessarily accessible: some might correspond to negative values of ρ and are therefore unphysical. Moreover, for certain values of h and $\mathfrak{s}$, the optimal densities might end up on the boundaries of the system ($\rho_x = 0$ for some x) or at infinity. These are consequences of the Legendre transforms we have performed, and it is important to remember that the physical variables that we fix are ρ and λ, not h and $\mathfrak{s}$, which are merely computational tools.

D. Generic dynamical phase transitions for multistable systems

One of the interesting features of this last result is the following: if the dynamics has several fixed points with different values of the Hamiltonian, whenever the two largest values cross each other due to variation of the parameters $\mathfrak{s}$ and h, the system undergoes a first order dynamical phase transition. In this section, we will see that this scenario is almost guaranteed in systems where the deterministic (unbiased) dynamics has several fixed points.

We recall the expression of the contracted biased Hamiltonian,

$$\mathcal{H}_{\mathfrak{s},h}(f,\rho) = \sum_{\gamma,\gamma'} k_{\gamma'\gamma}\rho^{\nu^\gamma}\left(e^{(\nu^{\gamma'} - \nu^\gamma)f + \mathfrak{s}_{\gamma'\gamma}} - 1\right) + \sum_x h_x\rho_x \tag{116}$$

with dynamics

$$\dot{f} = -\partial_\rho \mathcal{H}_{\mathfrak{s},h} \quad \text{and} \quad \dot{\rho} = \partial_f \mathcal{H}_{\mathfrak{s},h} \tag{117}$$

and boundary conditions

$$\rho_0 = \rho_i \quad \text{and} \quad f_t = 0. \tag{118}$$

The unbiased version $\mathfrak{s} = h = 0$ takes a particularly simple form: it is straightforward to check that $\dot{f} = 0$ whenever $f = 0$, which implies that $f = 0$ along any solution due to the final condition. The equation on ρ then becomes

$$\dot{\rho}_x = \partial_f \mathcal{H}_{0,0}(0,\rho) = \sum_{\gamma,\gamma'} (\nu_x^{\gamma'} - \nu_x^\gamma)k_{\gamma'\gamma}\rho^{\nu^\gamma}, \tag{119}$$

which is the usual deterministic mass-action dynamics.

Consider the case where this equation admits several fixed points $\dot{\rho}_i^{(0)} = 0$. The value of the unbiased Hamiltonian is 0 along the whole $f = 0$ space, which means it is 0 at each of those fixed points, and since they are the only ones dynamically connected to the final condition, they are all potential convergence points, depending on the initial condition.

We now turn on the bias, with a global infinitesimal variable ε, and we look for the positions of the new fixed points to first nonzero order in ε,

$$\rho_i(\varepsilon) \sim \rho_i^{(0)} + \varepsilon\rho_i^{(1)} + \frac{\varepsilon^2}{2}\rho_i^{(2)} + \cdots, \tag{120}$$

$$f_i(\varepsilon) \sim 0 + \varepsilon f_i^{(1)} + \frac{\varepsilon^2}{2}f_i^{(2)} + \cdots, \tag{121}$$

$$\mathcal{H}_\varepsilon(f,\rho) \sim \mathcal{H}_{0,0}(f,\rho) + \varepsilon\mathcal{H}^{(1)}(f,\rho) + \frac{\varepsilon^2}{2}\mathcal{H}^{(2)}(f,\rho) + \cdots. \tag{122}$$

Equations (117) give us, to the first order (order zero trivially vanishes),

$$\dot{\rho}_i^{(1)} = \partial_f\mathcal{H}^{(1)}(0,\rho_i^{(0)}) + \partial_f^2\mathcal{H}_{0,0}(0,\rho_i^{(0)})f_i^{(1)}$$
$$+ \partial_\rho\partial_f\mathcal{H}_{0,0}(0,\rho_i^{(0)})\rho_i^{(1)} = 0, \tag{123}$$

$$\dot{f}_i^{(1)} = -\partial_\rho\mathcal{H}^{(1)}(0,\rho_i^{(0)}) - \partial_f\partial_\rho\mathcal{H}_{0,0}(0,\rho_i^{(0)})f_i^{(1)}$$
$$- \partial_\rho^2\mathcal{H}_{0,0}(0,\rho_i^{(0)})\rho_i^{(1)} = 0, \tag{124}$$

which is a linear equation on $\{f_i^{(1)}, \rho_i^{(1)}\}$ that always has a solution if the Wronskian $W_i = \partial^2\mathcal{H}_{0,0}(0,\rho_i^{(0)})$ (i.e., the matrix of second derivatives) is invertible, which is generally the case

$$\begin{bmatrix} f_i^{(1)} \\ \rho_i^{(1)} \end{bmatrix} = -W_i^{-1} \cdot \begin{bmatrix} \partial_f\mathcal{H}^{(1)}(0,\rho_i^{(0)}) \\ \partial_\rho\mathcal{H}^{(1)}(0,\rho_i^{(0)}) \end{bmatrix}. \tag{125}$$

The new value of the Hamiltonian is then

$$\mathcal{H}_\varepsilon(f_i,\rho_i) \sim \mathcal{H}_{0,0}(0,\rho_i^{(0)}) + \varepsilon\Big(\partial_f\mathcal{H}_{0,0}(0,\rho_i^{(0)})f_i^{(1)}$$
$$+ \partial_\rho\mathcal{H}_{0,0}(0,\rho_i^{(0)})\rho_i^{(1)} + \mathcal{H}^{(1)}(0,\rho_i^{(0)})\Big), \tag{126}$$

which is to say

$$\mathcal{H}_\varepsilon(f,\rho) \sim \varepsilon\mathcal{H}^{(1)}(0,\rho_i^{(0)}). \tag{127}$$

This expression does not depend on the specific values of $\{f_i^{(1)}, \rho_i^{(1)}\}$ but only on their existence and the value of the perturbation at

the original fixed point. This should come as no surprise to any-one familiar with perturbation theory for eigenvalues and eigen-states of matrices: the first order in perturbation of the eigenval-ues only depend on the value of the perturbation in the original eigenstates.

For generic values of biases, the value of that perturbation will be different at every fixed point, which means that they all become dynamically disconnected (since all trajectories conserve the value of the Hamiltonian). Moreover, the fixed point where that value is highest will generically depend on the direction of the perturbation, which means that the dominant attractor of our dynamics will jump from one fixed point to another when crossing $\varepsilon = 0$. This allows us to conclude that multistable chemical systems gener-ically undergo first-order dynamical phase transitions at zero bias. We will see examples of this in simple cases in Sec. V. In particular, we will exhibit examples where the dominant stable fixed points of the biased dynamics are *unstable* fixed points in the unbiased system, which get stabilized by fluctuations.

Note that this result requires $\mathscr{H}^{(1)}(0,\rho)$ to not be independent of ρ, which is the case for most nontrivial biases. Moreover, the argu-ment above can just as easily be made for more complex attractors, such as limit cycles. However, the possible contribution of the kinetic term (91) in the extremization problem makes it less clear that which attractor will dominate for small biases. We therefore leave this case as a conjecture for the time being.

V. A FEW EXAMPLES OF DYNAMICAL PHASE TRANSITIONS

In this section, we illustrate our results with various examples of dynamical phase transition that we may observe in multistable stochastic chemical reaction networks. We will consider the Schlögl model,[63] which is one of the simplest bistable chemical models, as well as variants of it with different numbers of fixed points, and we will see how those fixed points undergo phase transitions as we vary the bias. Two surprising observations will be made: fluctuations can restore broken ergodicity in the large volume limit and can stabilize unstable fixed points of the deterministic dynamics.

Even though all previous formulae apply to both density and current biases, we will only consider the latter, i.e., we take $h = 0$, for the sake of simplicity. An example of a dynamical phase transition involving a density bias can be found in Ref. 98.

A. Schlögl model

In the Schlögl model, two reversible chemical reactions can occur, with the following kinetic rates:

$$\varnothing \to A \quad \text{with rate} \quad k_0 \quad \text{and bias} \quad s_0,$$
$$A \to \varnothing \quad \text{with rate} \quad k_1 \quad \text{and bias} \quad s_1,$$
$$2A \to 3A \quad \text{with rate} \quad k_2 \quad \text{and bias} \quad s_2,$$
$$3A \to 2A \quad \text{with rate} \quad k_3 \quad \text{and bias} \quad s_3.$$
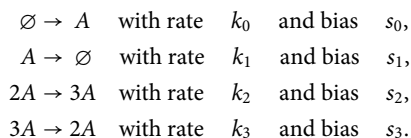
The biased Hamiltonian of the model is given by

$$\mathscr{H}_{s,0}(f,\rho) = k_0(e^{f+s_0} - 1) + k_1\rho(e^{-f-s_1} - 1)$$
$$+ k_2\rho^2(e^{f+s_2} - 1) + k_3\rho^3(e^{-f-s_3} - 1), \quad (128)$$

and the deterministic equation is

$$\dot{\rho} = \partial_f \mathscr{H}_{0,0}(f,\rho) = k_0 - k_1\rho + k_2\rho^2 - k_3\rho^3. \quad (129)$$

We start from a fixed initial condition ρ_0.

For an appropriate choice of rates (e.g., $k_0 = 0.7$, $k_1 = 2$, $k_2 = 0.9$, and $k_3 = 0.1$), this equation can have three fixed points, of which the middle one is unstable and the others are stable. In Fig. 2, we rep-resent the Hamiltonian trajectories for the unbiased system, where the stable fixed points are represented in green and the unstable one is represented in red. The sign of the Hamiltonian is indicated in each region, as a reference for the following deformations, in order to identify the critical point with the highest value of $\mathscr{H}$. Note that, in this context, the whole dynamics takes place on the $f = 0$ line due to the final condition and that the notion of stability of fixed points is therefore different from the fluctuating case. Also note that the stable fixed point that will be reached depends on the initial condi-tion, which is due to the breaking of ergodicity in the large-volume limit.[80,85]

We now turn on the bias, for example, by conditioning on the current of the reaction $\varnothing \leftrightarrow A$ alone: $s_0 = -s_1 = s$ and $s_2 = s_3 = 0$. The new trajectories are shown on the following figure, depending on the sign of s.

We represent in Fig. 3 the fixed point with the highest value of the Hamiltonian in green. As we see, it jumps from right to left when s crosses 0, and in both cases, this fixed point is unique regardless of the initial condition, as the other two are exponentially subdomi-nant. This can be easily understood when looking at the value of the perturbation at first order in s across the $f = 0$ line,

$$s\mathscr{H}^{(1)}(0,\rho) = k_0 s_0 + k_1\rho s_1 + k_2\rho^2 s_2 + k_3\rho^3 s_3 = s(k_0 - k_1\rho), \quad (130)$$

which is increasing if $s < 0$ and decreasing if $s > 0$, consistently with which fixed point is favored.
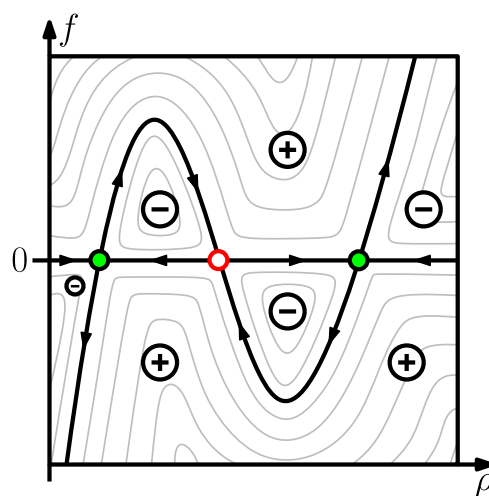
FIG. 2. Hamiltonian trajectories of the unbiased bistable Schlögl model. The critical manifolds $\mathscr{H} = 0$ are represented in black, and the arrows indicate the direction of the flow. The two stable fixed points are represented in green, and the unsta-ble one is represented in red. We indicate the sign of the Hamiltonian in each region.
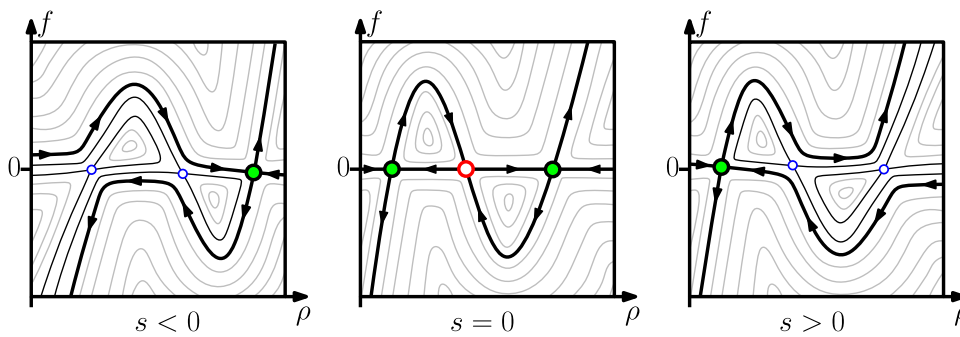
FIG. 3. Hamiltonian trajectories of the biased bistable Schlögl model. The critical manifolds are represented in black, and the arrows indicate the direction of the flow. The dominant fixed points are represented in green. A dynamical phase transition can be seen around zero bias, where the dominant fixed point jumps from right to left.

In Fig. 4, on the left, we plot the values of $\mathcal{H}$ at all three fixed points as a function of s. Those values are obtained by first expressing $f^\star(\rho)$ analytically using one of Hamilton's equations, plugging the result into the second and then computing $\rho^\star$ numerically.

The red curve corresponds to the lowest density, the blue one to the highest, and the green one to the central fixed point. As we see, for a positive s, the blue curve is the highest, i.e., the rightmost point is dominant, whereas for a slightly negative s, the red curve (leftmost point) is dominant. The SCGF $E(s)$ is obtained by keeping the maximum of all three curves and has a nonanalyticity at $s = 0$. In the right figure, we draw the resulting long-time large deviation function $g(\lambda)$ obtained by computing numerically the Legendre transform of $E(s)$. We see that it exhibits a horizontal plateau, approximately between $\lambda = 0.2$ (as visible on the inset) and $\lambda = 12$, which is an expected signature of a first-order phase transition at $s = 0$ (the slope of the plateau is consistent with the value of s). The sharp increase in slope for $\lambda < 0$ is due to the fluctuation relation and is smooth, as can be seen in the inset, so that there is no phase transition at $\lambda = 0$ even though the figure might suggest it.
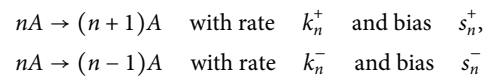
A surprising observation is that, on both sides of the transition, the ergodicity of the microscopic system, lost in the large-volume limit, is restored: any initial condition leads to the same fixed point as its stable manifold spans all density-space.

In light of Eq. (130), we may wonder if it is possible to favor the central fixed point instead. It is indeed the case as it is quite easy to engineer a bias which makes the perturbation maximal around that value of ρ. Consider, for instance, $s_0 = -4s$, $s_1 = 3s$, $s_2 = -s$, and $s_3 = 0$ for s small and positive. The resulting trajectories are sketched in Fig. 5.

The conclusion may seem quite counterintuitive: there is no fundamental difference between stable and unstable deterministic attractors when looking at biased trajectories, and fluctuations can stabilize unstable stationary states, just as an extra external driving can.[118]

B. Generalized Schlögl model

It is straightforward to generalize the previous case to one with any number of fixed points, by considering reactions of any order,

$$nA \rightarrow (n+1)A \quad \text{with rate} \quad k_n^+ \quad \text{and bias} \quad s_n^+,$$
$$nA \rightarrow (n-1)A \quad \text{with rate} \quad k_n^- \quad \text{and bias} \quad s_n^-$$

and choosing the rates such that the deterministic equation

$$\dot{\rho} = \sum_{n=n_{\min}}^{n_{\max}} (k_n^+ - k_n^-)\rho^n \tag{131}$$

has the desired number of fixed points. Any of the fixed points can then become dominant under the appropriate perturbation,

$$s\mathcal{H}^{(1)}(0,\rho) = \sum_{n=n_{\min}}^{n_{\max}} (k_n^+ s_n^+ + k_n^- s_n^-)\rho^n. \tag{132}$$

Note that as long as the lowest reaction is creative ($k_{n_{\min}}^- = 0$) and the highest one is destructive ($k_{n_{\max}}^+ = 0$), the first and last fixed points will be stable in the deterministic case. This also implies that the nonzero solution of $\mathcal{H}_{0,0}(f,\rho) = 0$ spans the whole f-space: we have
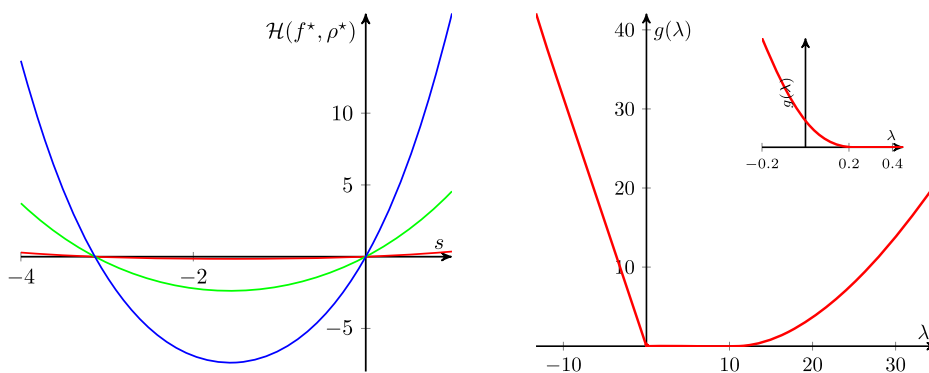
FIG. 4. Left: value of the Hamiltonian at each fixed point. Right: Legendre transform of the dominant value of $\mathcal{H}$.
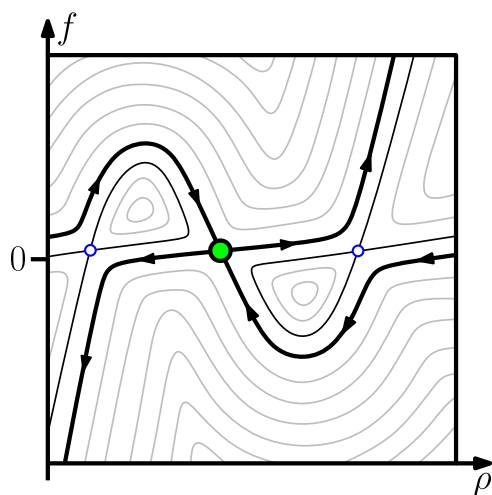
FIG. 5. Hamiltonian trajectories of the bistable Schlögl model with a bias favoring the unstable fixed point. Surprisingly, that fixed point becomes stable and dominant for the biased dynamics.

$$\mathscr{H}_{0,0}(f,\rho) = \sum_{n=n_{\min}}^{n_{\max}} (e^f - 1)k_n^+ \rho^n + (e^{-f} - 1)k_n^- \rho^n$$

$$= (e^f - 1) \sum_{n=n_{\min}}^{n_{\max}} \left(k_n^+ - e^{-f} k_n^- \right)\rho^n, \qquad (133)$$

which cancels at $f = 0$ or at

$$f^\star(\rho) = \ln\left(\frac{\sum_n k_n^- \rho^n}{\sum_n k_n^+ \rho^n} \right) \qquad (134)$$

such that

$$f^\star(\rho \to 0) \to -\infty \quad \text{and} \quad f^\star(\rho \to \infty) \to +\infty. \qquad (135)$$

This, in turn, implies that there will always be at least one fixed point dynamically connected to the boundary conditions. In the following example, we see what may happen when this is not the case.

C. Runaway Schlögl model

As a final example, we consider a variant of the Schlögl model whose deterministic dynamics is not always stable: the so-called
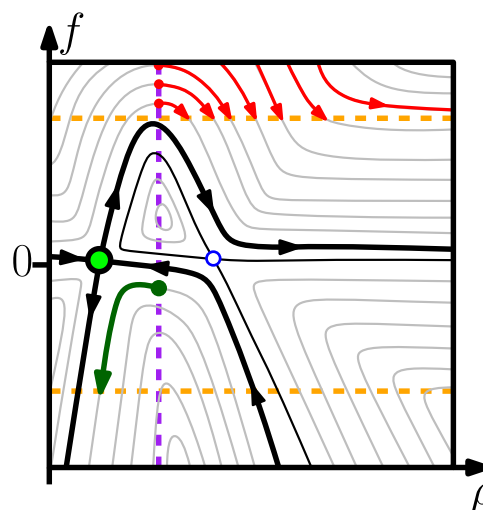
FIG. 7. The trajectories of the biased runaway Schlögl model depend heavily on the final condition for f. The green trajectory converges to the stable fixed point, whereas the red trajectories diverge.

runaway Schlögl model.[108] It is obtained by taking $k_3 = 0$ in the Schlögl model so that the stability condition mentioned above is not true anymore: the highest-order reaction is creative, and any initial condition larger than the largest fixed point will diverge to infinity.

Let us now consider a bias $s_0 = s + m$, $s_1 = -s - m$, and $s_2 = m$. This lets us effectively change the final condition through m and favor one or the other fixed point through s. In Fig. 6, we draw the Hamiltonian trajectories of the model for $m = 0$ and $s < 0$, $s = 0$, and $s > 0$.

Once again, we find a first-order dynamical phase transition between the two fixed points, one of them being originally unstable. Moreover, we see that in all cases, the unstable manifolds of the fixed points cannot reach beyond a certain value $f_{\max}$ of f. If we set m to be above that critical value, the fixed point becomes dynamically disconnected from the boundary conditions, and the system becomes unstable. This is shown in Fig. 7, where the initial condition is shown in purple (vertical dashed line) and the final conditions are shown in orange (horizontal dashed lines, for two different values of m). For a final
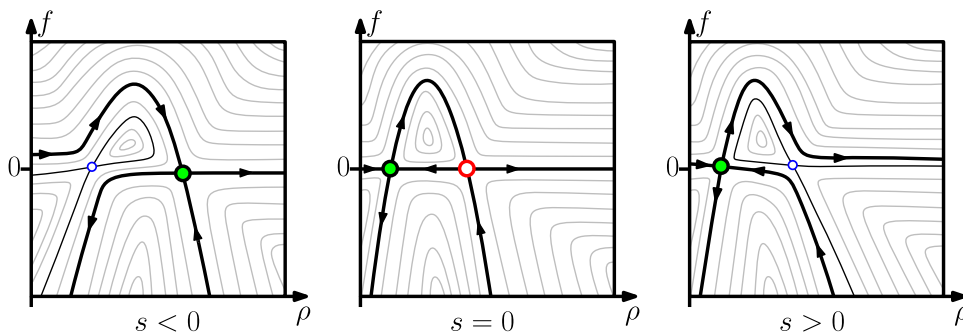
FIG. 6. Hamiltonian trajectories of the biased runaway Schlögl model. A first-order dynamical phase transition occurs between the two fixed points around zero bias.

condition below the dominant fixed point, there is a stable trajectory converging to it, shown in green. However, for a final condition above f_{max}, the only trajectories that have the right length are those shown in red and diverge as time grows. If the final condition is between those two regions, both types of trajectories exist, and it is unclear which dominates, as the red trajectories have a diverging value of the Hamiltonian but a nonzero kinetic term $f\dot{\rho}$.

This example gives us yet another type of dynamical phase transitions that can be found in some stochastic chemical networks, specifically ones that are deterministically unstable. A similar phase transition was observed in Ref. 119 in an Ornstein-Uhlenbeck process.

VI. PRACTICAL SUMMARY OF FORMULAE AND RESULTS

In this final section, we give a practical overview of our results to serve as a minimal reference for computing large deviation functions of time-averaged dynamical observables.

Let us consider a well-mixed chemical system in a volume V, with mass-action kinetics, species A_x with concentrations ρ_x, and elementary reactions between chemical complexes γ with stoichiometric coefficients v_x^γ. The kinetic rate of reaction $\gamma \to \gamma'$ is given by $k_{\gamma'\gamma}$ so that its complete mass-action rate is $k_{\gamma'\gamma}\rho^{v^\gamma}$. We define a fluctuating chemical current $\lambda_{\gamma'\gamma}$ of which this mass-action rate is the typical value.

We are interested in the probability of observing, over a long time span t, an atypical average value of a certain linear combination of chemical concentrations and currents a (which can be a vector), defined as

$$a = \sum_{\gamma,\gamma'} \mathfrak{s}_{\gamma'\gamma} \lambda_{\gamma'\gamma} + \sum_x h_x \rho_x. \tag{136}$$

This probability obeys a large deviation principle $P_t(a) = e^{-tVg(a)}$ with a large deviation function $g(a)$. We will need to define its Legendre transform $E(\mu) = \mu a - g(a)$ with a implicitly defined through $\mu = g'(a)$, which is called the scaled cumulant generating function of a.

In order to compute $g(a)$, we first need to define the biased Hamiltonian $\mathcal{H}_\mu$, with momenta variables f_x,

$$\mathcal{H}_\mu(f,\rho) = \sum_{\gamma,\gamma'} k_{\gamma'\gamma}\rho^{v^\gamma}\left(e^{(v^{\gamma'}-v^\gamma)f+\mu\mathfrak{s}_{\gamma'\gamma}} - 1\right) + \mu \sum_x h_x \rho_x. \tag{137}$$

Note that, to make all expressions simpler, we only have one global biasing variable $\mu \in \mathbb{R}$ instead of the full vector $\{\mathfrak{s},h\}$, which are here treated as constants.

This Hamiltonian is associated with Hamilton's equations

$$\dot{f} = -\partial_\rho \mathcal{H}_\mu \quad \text{with} \quad \dot{\rho} = \partial_f \mathcal{H}_\mu \tag{138}$$

with a final condition $f(t) = 0$. The initial condition depends on the distribution from which the initial state of the system is drawn, but we will assume that it is fixed: $\rho(0) = \rho_0$.

We now need to determine the critical manifolds of $\mathcal{H}_\mu$. Assuming they are all fixed points (we leave the more complex cases for future works), we need to find all solutions to $\partial_\rho \mathcal{H}_\mu = 0$ and

$\partial_f \mathcal{H}_\mu = 0$, which we will denote by $\{f_i^\star, \rho_i^\star\}$. In some particularly simple cases, these equations can be solved analytically, although in general we would have to rely on numerics. A few points are worth noting here:

- The dynamical system which we have to solve is Hamiltonian, which implies that the sum of all Lyapunov coefficients at any fixed point is 0, i.e., they are all of mixed or marginal stability. This means that, in terms of the dynamical equations (138), they are never attractors and cannot be reached by simply following those equations numerically.
- That being said, the fixed points of (138) are also the fixed points of the (unphysical) equations obtained by flipping the sign of any subset of those equations (of which there are twice the number K of chemical species). Except in pathological cases, it is possible, for each fixed point $(f_i^\star, \rho_i^\star)$, to find a set of signs $\{\theta_x, \varepsilon_x\} \in \{-1, 1\}^{2K}$ for which that point is an *attractor* of the ad-hoc equations $\dot{f} = -\theta\partial_\rho\mathcal{H}_\mu$ and $\dot{\rho} = \varepsilon\partial_f\mathcal{H}_\mu$. To give an example, consider the Hamiltonian $\mathcal{H}(f,\rho) = a(e^{f+s} - 1) + b\rho(e^{-f-s} - 1)$ used in Sec. IV B. Hamilton's equations

$$\dot{\rho} = a\, e^{f+s} - b\rho\, e^{-f-s} \quad \text{and} \quad \dot{f} = -b\rho(e^{-f-s} - 1) \tag{139}$$

  have a hyperbolic fixed point at $\rho = a/b$ and $f = -s$, but the equations

$$\dot{\rho} = a\, e^{f+s} - b\rho\, e^{-f-s} \quad \text{and} \quad \dot{f} = +b\rho(e^{-f-s} - 1), \tag{140}$$

  obtained by flipping the sign of $\dot{f}$, have a stable fixed point at the same place. This fixed point can then be attained by following those new equations numerically from appropriate initial conditions. In many cases, and in general for small values of μ, the appropriate signs will be $\theta_x = -1$ and $\varepsilon_x = 1$.
- Another way to find those fixed points is to consider that, for any fixed value of ρ, $\mathcal{H}_\mu(f,\rho)$ is a convex function of f, which has a single minimum $f^\star(\rho)$. This can be found analytically or numerically as a function of ρ and injected back into $\partial_\rho\mathcal{H}_\mu = 0$, making it easier to solve in its turn.

Once we have obtained the fixed points, the value of $E(\mu)$ is simply given by the maximum of $\mathcal{H}_\mu$ at those points, as long as the original chemical system is stable for all initial conditions (which is always the case if all reactions are reversible),

$$E(\mu) = \max_i\left[\mathcal{H}_\mu(f_i^\star, \rho_i^\star)\right]. \tag{141}$$

Note that the position of the corresponding fixed point will usually be continuous by parts: if the dominant fixed point is nondegenerate, it will be followed as μ varies, until it becomes degenerate with another one, at which point the second one may become dominant. This is essentially identical to ground-state eigenvalue crossing in quantum systems and is a generic phenomenology for first-order phase transitions.

Finally, the function $E(\mu)$ can then be turned into $g(a)$ through a numerical Legendre transform, yielding the desired probability distribution.

VII. CONCLUSION

We have proven that, in the limit of large volumes, a chemical master equation for a well-mixed chemical reaction network can be described through a large-deviation path integral over a hydrodynamic Lagrangian. We have derived that Lagrangian explicitly from the microscopic dynamics by keeping track of every individual chemical current. We have then derived the equations that the minimizer of the said path integral should verify, either through Euler-Lagrange equations or through the definition of an appropriate Hamiltonian and the corresponding Hamilton equations. We have shown that this formalism extends naturally to the case where the dynamics is biased with respect to time-additive observables and that the scaled cumulant generating function of those observables can be expressed in terms of the largest value that the biased Hamiltonian takes over all solutions at fixed time. Through a Cole-Hopf transform, we find a second set of canonical coordinates which, in the specific case of mass-action dynamics, leads to the Doi-Peliti Hamiltonian and is appropriate for dealing with conserved quantities.

The boundary conditions for those equations, which turn out to be split between the initial and final coordinates, are of crucial importance, as they allow the solutions to converge to global attractors even though the dynamics is Hamiltonian. The aforementioned cumulant generating function can then be identified as the maximal value the Hamiltonian takes across those attractors, provided that the kinetic term of the action is negligible. This generalizes a well-known result by Donsker and Varadhan relating the cumulant generating function of a Markov process to the largest eigenvalue of the corresponding biased Markov matrix.[99–102] Furthermore, in the case where the deterministic chemical reaction network has several fixed points, we have shown that, under bias, the system generically undergoes first-order dynamical phase transitions around zero bias, which we have illustrated through variants of the Schlögl model. Finally, we have shown an example of a dynamical phase transition where the system becomes unstable on one side of the transition.

Here are a few important takeaways from both our abstract results and illustrative examples:

- **tracking more is easier**: the detailed Lagrangian, which keeps track of all individual chemical currents, is explicit, whereas the standard Lagrangian is usually not;
- **slicing with precision is crucial**: taking a time step δt small enough to vanish but large enough that $V \, \delta t$ diverges is essential to obtain the correct Lagrangians;
- **boundary conditions make all the difference**: a Hamiltonian dynamics with a full initial condition cannot have attractors, but one with mixed boundary conditions often does;
- **stability is relative**: a fixed point which is unstable in the deterministic dynamics of a system may become stable under even an infinitesimal bias; and
- **what is broken can be fixed**: a system with microscopic ergodicity, broken due to some scaling limit, might well see it restored when constrained in certain ways.

There are many problems still open on the topic of large deviations of population models. Chief amongst them is the case of systems with attractors such as limit cycles—which are quite important in the context of biochemistry[120]—or even strange attractors, for which very little is known on the large-deviation scale, beyond some numerical results on entropy production.[121] In Ref. 116, the probability of extinction in a model with a cycle and a fixed point at zero population was estimated, by finding the instanton connecting the cycle to the fixed point through a Hamiltonian formalism. Perhaps similar methods can be applied for models biased on currents or densities, although even the shape of the biased attractors is yet to be determined.

Additionally, the issue of finding those attractors numerically is a particularly challenging one due to their systematically mixed stability. We have outlined a few ideas to remedy this in Sec. VI, but devising and testing actual algorithms is still an open problem as far as we know.

We have also avoided the case of systems sitting at the critical point of a stationary phase transition, where any bias might force the system into one phase or the other. The Schlögl model is simple enough to treat numerically at its critical point, but whether those critical points are special in terms of which dynamical phase transitions can occur under bias remains to be seen.

A crucial concept in chemical networks, which is only indirectly relevant to this work, is that of deficiency:[122] in essence, it is the number of conservation laws that is not visible immediately on the network of complexes. That number has important consequences on the types of stationary states that one may observe.[62] If it is 0, for instance, it is known that the system has exactly one stable fixed point and no other attractor.[107] It would be interesting to determine the consequences of deficiency in the context of dynamical large deviations.

We have mentioned the delicate case of dynamics which comes close to the boundaries of concentration space, where the number of individuals of one or more species comes back down to a microscopic scale. Those systems can be treated as so-called *hybrid models*, where part of the dynamics is approximated through a large deviation formalism and the rest is kept as a Markov jump process.[123,124] This would apply, for instance, to systems with a small number of enzymes which go through various conformations.[125] A few results exist as regards the large deviations of such systems,[126] but as far as we know, the probability distributions of time-averaged observables have not been tackled so far.

Finally, one may naturally wonder how the structure and phenomenology of those models would be affected if one adds spatial coordinates and lets the chemicals diffuse, as in reaction-diffusion models.[127] It is known that, for instance, in certain simple 1-species models with an active/inactive stationary transition (i.e., where, depending on a rate constant, the system has a fixed point at a nonzero/zero density), the position of the transition is strongly affected by the dimension of the diffusion space, to the point where special nonperturbative renormalization techniques[128] are required to account for it properly in $D \geq 3$.[129,130] Whether anything similar occurs for multistable models, and how this interacts with dynamical biases and phase transitions, shall be a very interesting question for a future work. We might also ask similar questions for systems that exhibit instabilities, leading to Turing patterns.[131,132]

APPENDIX: COMPUTATION OF THE CHEMICAL LAGRANGIANS AND HAMILTONIANS

In this appendix, we will compute explicitly the microscopic transition probabilities of our chemical networks, with or without bias, in the limit of a large volume and a short time step: $V \to \infty$, $\delta t \to 0$. This will be done by exploiting the fact that, in those limits, all separate reactions essentially commute.

Consider the microscopic chemical Markov matrix

$$W = \sum_{\gamma',\gamma,n} W_{\gamma'\gamma}(n) |n - \nabla_{\gamma',\gamma}\rangle\langle n| - W_{\gamma'\gamma}(n) |n\rangle\langle n|. \quad (A1)$$

The transition probability from n_i to n_f during time δt is

$$P_{\delta t}[n_f|n_i] = \langle n_f|e^{\delta t W}|n_i\rangle. \quad (A2)$$

Moreover, the generating function of exponential moments of the observables λ and ρ defined in Sec. II B between those states and during that time, with respective conjugate variables $\mathfrak{s}$ and h, is given by[4,36]

$$\left\langle e^{\delta t(\mathfrak{s}\cdot\lambda + h\rho)} \right\rangle_{n_i}^{n_f} = \left\langle n_f|e^{\delta t W_{\mathfrak{s},h}}|n_i\right\rangle, \quad (A3)$$

where the Markov matrix has been replaced by its biased version

$$W_{\mathfrak{s},h} = \sum_{\gamma',\gamma,n} e^{\mathfrak{s}_{\gamma'\gamma}} W_{\gamma'\gamma}(n) |n - \nabla_{\gamma',\gamma}\rangle\langle n| - W_{\gamma'\gamma}(n) |n\rangle\langle n|$$
$$+ \sum_{x,n} h_x n_x |n\rangle\langle n|. \quad (A4)$$

The effect of this biasing on the probability of each trajectory is to multiply it by $e^{\mathfrak{s}_{\gamma'\gamma}}$ for each occurrence of a reaction $\gamma \to \gamma'$ regardless of the starting composition n, thus generating the moments of λ, and by $e^{\int h n(\tau) d\tau}$ which generates moments of ρ.

We will now simplify this expression in the limit of a large volume V, in the case of mass-action kinetics,

$$W_{\gamma'\gamma}(n) = k_{\gamma'\gamma} \prod_x \frac{[n_x]!}{[n_x - v_x^\gamma]!} V^{1-\sum_x v_x^\gamma}, \quad (A5)$$

by showing that all individual reaction operators

$$A_{\mathfrak{s}}^{\gamma'\gamma} = \sum_n e^{\mathfrak{s}_{\gamma'\gamma}} W_{\gamma'\gamma}(n) |n - \nabla_{\gamma',\gamma}\rangle\langle n| \quad (A6)$$

as well as the corresponding escape rates

$$B^{\gamma'\gamma} = -\sum_n W_{\gamma'\gamma}(n) |n\rangle\langle n| \quad (A7)$$

and the concentration-counting operator

$$B_h = \sum_{x,n} h_x n_x |n\rangle\langle n| \quad (A8)$$

all commute on the large deviation scale for short times, as long as each particle number n is of order V.

Let us first consider $A_1 = A_{\mathfrak{s}}^{\gamma'\gamma}$, $A_2 = A_{\mathfrak{s}}^{\mu'\mu}$, and $B = B^{\gamma'\gamma}$ and compute their commutator,

$$A_2 A_1 - A_1 A_2 = \sum_n e^{\mathfrak{s}_{\gamma'\gamma} + \mathfrak{s}_{\mu'\mu}} \Big(W_{\mu'\mu}(n - \nabla_{\gamma',\gamma}) W_{\gamma'\gamma}(n)$$
$$- W_{\gamma'\gamma}(n - \nabla_{\mu',\mu}) W_{\mu'\mu}(n) \Big) |n - \nabla_{\gamma',\gamma} - \nabla_{\mu',\mu}\rangle\langle n|,$$

$$A_2 B - B A_2 = -\sum_n e^{\mathfrak{s}_{\mu'\mu}} \Big(W_{\mu'\mu}(n) W_{\gamma'\gamma}(n)$$
$$- W_{\gamma'\gamma}(n - \nabla_{\mu',\mu}) W_{\mu'\mu}(n) \Big) |n - \nabla_{\mu',\mu}\rangle\langle n|.$$

For n of order V, we can expand $W_{\gamma'\gamma}(n)$ in orders of V using Stirling's approximation, defining $\rho = n/V$,

$$W_{\gamma'\gamma}(n) \sim k_{\gamma'\gamma} V \prod_y \rho_y^{v_y^\gamma} \left(1 + \sum_x \frac{v_x^\gamma(1 - 2v_x^\gamma)}{2V\rho_x}\right),$$

$$W_{\gamma'\gamma}(n - \nabla_{\mu',\mu}) \sim k_{\gamma'\gamma} V \prod_y \rho_y^{v_y^\gamma} \left(1 + \sum_x \frac{v_x^\gamma(1 - 2v_x^\gamma)}{2V\rho_x} - v_x^\gamma \frac{v_x^{\mu'} - v_x^\mu}{V\rho_x}\right)$$

so that

$$A_2 A_1 - A_1 A_2 = V \sum_n e^{\mathfrak{s}_{\gamma'\gamma} + \mathfrak{s}_{\mu'\mu}} k_{\mu'\mu} k_{\gamma'\gamma} \rho^{v^\mu + v^\gamma} \left(\sum_x \frac{v_x^\gamma v_x^{\mu'} - v_x^\mu v_x^{\gamma'}}{\rho_x}\right)$$
$$\times |n - \nabla_{\gamma',\gamma} - \nabla_{\mu',\mu}\rangle\langle n|,$$

$$A_2 B - B A_2 = -V \sum_n e^{\mathfrak{s}_{\mu'\mu}} k_{\mu'\mu} k_{\gamma'\gamma} \rho^{v^\mu + v^\gamma} \left(\sum_x v_x^\gamma \frac{v_x^{\mu'} - v_x^\mu}{\rho_x}\right)$$
$$\times |n - \nabla_{\mu',\mu}\rangle\langle n|.$$

We therefore have that A_1, A_2, B, as well as their commutators, are of order V, as would be any higher order commutator. We can then apply the Baker-Campbell-Hausdorff formula to second order in δt to sums of those operators, such as $A_1 + A_2$,

$$e^{\delta t(A_1 + A_2)} \sim e^{\delta t A_1} e^{\delta t A_2} e^{-\delta t^2 [A_1, A_2]/2} \sim e^{\delta t A_2} e^{\delta t A_1} e^{\delta t^2 [A_1, A_2]/2}. \quad (A9)$$

In both expressions, the last term is negligible for $\delta t \to 0$, which concludes the proof. A similar computation leads to the commutation of $A_{\mathfrak{s}}^{\gamma'\gamma}$ with B_h.

Let us now estimate the value of a term of the form $\langle n_2|e^{\delta t(A_{\mathfrak{s}}^{\gamma'\gamma} + B^{\gamma'\gamma})}|n_1\rangle$ in the limits stated above, also considering that $V\delta t \to \infty$, which is to say that the number of reactions occurring in one time step is large. This is an important step to ensure that we are in a large deviation regime. Applying B to n_1 and expanding the other exponential, we find

$$\langle n_2 | e^{\delta t (A_{\mathfrak{s}}^{\gamma'\gamma} + B^{\gamma'\gamma})} | n_1 \rangle \asymp e^{-V\delta t k_{\gamma'\gamma} \rho_1^{\nu^y}} \sum_k \frac{\delta t^k}{k!} \langle n_2 | (A_{\mathfrak{s}}^{\gamma'\gamma})^k | n_1 \rangle$$

$$\asymp e^{-V\delta t k_{\gamma'\gamma} \rho_1^{\nu^y}} \frac{\left(V\delta t \, e^{\mathfrak{s}_{\gamma'\gamma}} k_{\gamma'\gamma} \rho_1^{\nu^y} \right)^k}{k!}$$

$$\times \mathbb{I}\big(n_2 - n_1 - k \nabla_{\gamma',\gamma} \big). \tag{A10}$$

Given that, for a large factor X, the function $X^k/k!$ has a maximum around $k = X$, we can rescale k by $X = V\delta t$ in the expression above, defining

$$k = V\delta t \, \lambda_{\gamma',\gamma}, \tag{A11}$$

where, as before, λ has the interpretation of a number of transitions per time step, rescaled by volume. One final use of the Stirling's formula then yields

$$\langle n_2 | e^{\delta t (A_{\mathfrak{s}}^{\gamma'\gamma} + B^{\gamma'\gamma})} | n_1 \rangle$$

$$\asymp \exp\left(-V\delta t \, \tilde{L}_{\mathfrak{s}}(\lambda_{\gamma',\gamma}, \rho_1) \right) \delta\big(\rho_2 - \rho_1 - \delta t \, \lambda_{\gamma',\gamma} \nabla_{\gamma',\gamma} \big), \tag{A12}$$

where we have replaced the discrete indicator function $\mathbb{I}$ with a continuous delta function, and

$$\tilde{L}_{\mathfrak{s}}(\lambda_{\gamma'\gamma}, \rho) = \lambda_{\gamma'\gamma} \ln(\lambda_{\gamma'\gamma}/k_{\gamma'\gamma} \rho^{\nu^y}) - \lambda_{\gamma'\gamma} + k_{\gamma'\gamma} \rho^{\nu^y} - \mathfrak{s}_{\gamma'\gamma} \lambda_{\gamma'\gamma}. \tag{A13}$$

This is the rate function of a single biased Poisson distribution, which is what we would expect for independent jumps processes. Similarly, we have

$$\langle n_2 | e^{\delta t (A_{\mathfrak{s}}^{\gamma'\gamma} + B^{\gamma'\gamma})} | n_1 \rangle \asymp e^{V\delta t \, h \rho_1} \, \delta\big(\rho_2 - \rho_1 \big). \tag{A14}$$

We can then consider the whole of $\langle n_f | e^{\delta t W_{\mathfrak{s},h}} | n_i \rangle$, by splitting the terms in $W_{\mathfrak{s},h}$ in any order we like and injecting the identity between every term. We are left with the product of all exponentials, with a global delta function that takes into account all the currents $\lambda_{\gamma'\gamma}$,

$$\langle n_f | e^{\delta t W_{\mathfrak{s},h}} | n_i \rangle \asymp \exp\left(-V\delta t \, L_{\mathfrak{s}}(\lambda, \rho_i) \right) \delta\big(\rho_f - \rho_i + \delta t \, \nabla \cdot \lambda \big) \tag{A15}$$

with

$$L_{\mathfrak{s}}(\lambda, \rho) = \sum_{\gamma',\gamma} \lambda_{\gamma'\gamma} \ln(\lambda_{\gamma'\gamma}/k_{\gamma'\gamma} \rho^{\nu^y}) - \lambda_{\gamma'\gamma} + k_{\gamma'\gamma} \rho^{\nu^y} - \mathfrak{s}_{\gamma'\gamma} \lambda_{\gamma'\gamma} - \sum_x h_x \rho_x. \tag{A16}$$

All that remains is to define $\rho_f - \rho_i = \dot\rho \, \delta t$ to recover formulae (51) for $\mathfrak{s} = 0$ and $h = 0$, and (70) otherwise.

To obtain an expression for the Hamiltonians instead, one may simply do a Legendre transform of the Lagrangians with respect to their respective flux variables. However, we may also start from their definition as a generating function,

$$H_{\mathfrak{s}}(\sigma, n_i) = \frac{1}{V\delta t} \ln\left(\sum_n \langle n | e^{\delta t W_{\sigma+\mathfrak{s},h}} | n_i \rangle \right). \tag{A17}$$

In this case, the sum over the endpoint n means we do not have to keep track of the number of particles exchanged. The same procedure as above, summed over the endpoint, yields

$$\sum_n \langle n | e^{\delta t (A_{\sigma+\mathfrak{s}}^{\gamma'\gamma} + B^{\gamma'\gamma})} | n_1 \rangle \asymp e^{-V\delta t k_{\gamma'\gamma} \rho_1^{\nu^y}} \sum_{k,n} \frac{\delta t^k}{k!} \langle n | (A_{\mathfrak{s}}^{\gamma'\gamma})^k | n_1 \rangle$$

$$\asymp e^{-V\delta t k_{\gamma'\gamma} \rho_1^{\nu^y}} \sum_k \frac{\left(V\delta t \, e^{\sigma_{\gamma'\gamma} + \mathfrak{s}_{\gamma'\gamma}} k_{\gamma'\gamma} \rho_1^{\nu^y} \right)^k}{k!}$$

$$\asymp \exp\left(V\delta t \, k_{\gamma'\gamma} \rho_1^{\nu^y} (e^{\sigma_{\gamma'\gamma} + \mathfrak{s}_{\gamma'\gamma}} - 1) \right).$$

Putting all terms together including B_h immediately yields

$$H_{\mathfrak{s}}(\sigma, n) = \sum_{\gamma',\gamma} k_{\gamma'\gamma} \rho_1^{\nu^y} \big(e^{\sigma_{\gamma'\gamma} + \mathfrak{s}_{\gamma'\gamma}} - 1 \big) + \sum_x h_x \rho_x, \tag{A18}$$

which is Eq. (76). Setting $\mathfrak{s} = 0$ and $h = 0$ yields Eq. (62). Finally, setting $\sigma = \nabla f$ yields Eq. (37).

One should note that these derivations very much rely on the fact that n stays of order V at all times so that a change in n_x of order ν_x^y is negligible in terms of ρ. For dynamics whose trajectories end up to any wall $n_x = 0$, one has to treat that species separately at the microscopic scale, resulting in a so-called *hybrid dynamics*.[126]

It should also be noted that those results can be easily generalized to dynamics other than mass-action, as long as one is able to show that all commutators between reaction operators are of order V (or of the same order as the operator themselves).

REFERENCES

[1] H. Spohn, *Large Scale Dynamics of Interacting Particles*, Texts and Monographs in Physics (Springer-Verlag, 1991).

[2] L. Bertini, A. De Sole, D. Gabrielli, G. Jona-Lasinio, and C. Landim, "Large deviation approach to non equilibrium processes in stochastic lattice gases," Bull. Braz. Math. Soc. **37**, 611–643 (2006); e-print arXiv:0602557 [math].

[3] F. Den Hollander, *Large Deviations* (American Mathematical Society, 2008).

[4] H. Touchette, "The large deviation approach to statistical mechanics," Phys. Rep. **478**, 1–69 (2009); e-print arXiv:0804.0327.

[5] R. Chetrite and H. Touchette, "Nonequilibrium Markov processes conditioned on large deviations," Ann. Henri Poincare **16**, 2005–2057 (2015); e-print arXiv:1405.5157.

[6] A. C. Barato and R. Chetrite, "A formal view on level 2.5 large deviations and fluctuation relations," J. Stat. Phys. **160**, 1154–1172 (2015); e-print arXiv:1408.5033.

[7] M. F. Weber and E. Frey, "Master equations and the theory of stochastic path integrals," Rep. Prog. Phys. **80**, 046601 (2017).

[8] M. I. Freidlin and A. D. Wentzell, *Random Perturbations of Dynamical Systems*, Grundlehren der Mathematischen Wissenschaften Vol. 260 (Springer Berlin Heidelberg, Berlin, Heidelberg, 2012).

[9] R. Chetrite and H. Touchette, "Variational and optimal control representations of conditioned and driven processes," J. Stat. Mech.: Theory Exp. **2015**, P12001; e-print arXiv:1506.05291.

[10] F. Bouchet, K. Gawędzki, C. Nardini, K. Gawedzki, and C. Nardini, "Perturbative calculation of quasi-potential in non-equilibrium diffusions: A mean-field example," J. Stat. Phys. **163**, 1157–1210 (2016).

[11] B. Derrida and T. Sadhu, "Large deviations conditioned on large deviations I: Markov chain and Langevin equation," J. Stat. Phys. (published online 2019); e-print arXiv:1807.06543.

[12] N. Tizón-Escamilla, V. Lecomte, and E. Bertin, "Effective driven dynamics for one-dimensional conditioned Langevin processes in the weak-noise limit," J. Stat. Mech.: Theory Exp. **2019**, 013201.

[13] T. Bodineau and B. Derrida, "Current fluctuations in nonequilibrium diffusive systems: An additivity principle," Phys. Rev. Lett. **92**, 180601-1 (2004); e-print arXiv:0406335 [cond-mat].

[14] T. Bodineau and B. Derrida, "Distribution of current in nonequilibrium diffusive systems and phase transitions," Phys. Rev. E **72**, 066110 (2005); e-print arXiv:0506540 [cond-mat].

[15] G. Verley, M. Esposito, T. Willaert, and C. Van den Broeck, "The unlikely Carnot efficiency," Nat. Commun. **5**, 4721 (2014).

[16] T. R. Gingrich, G. M. Rotskoff, S. Vaikuntanathan, and P. L. Geissler, "Efficiency and large deviations in time-asymmetric stochastic heat engines," New J. Phys. **16**, 102003 (2014).

[17] F. Ragone, J. Wouters, and F. Bouchet, "Computation of extreme heat waves in climate models using a large deviation algorithm," Proc. Natl. Acad. Sci. U. S. A. **115**, 24–29 (2018).

[18] R. L. Jack and P. Sollich, "Large deviations and ensembles of trajectories in stochastic models," Prog. Theor. Phys. Suppl. **184**, 304–317 (2009); e-print arXiv:0911.0211.

[19] G. Verley, "Nonequilibrium thermodynamic potentials for continuous-time Markov chains," Phys. Rev. E **93**, 012111 (2016).

[20] H. Touchette and R. J. Harris, "Large deviation approach to nonequilibrium systems," in *Nonequilibrium Statistical Physics of Small Systems: Fluctuation Relations and Beyond* (Wiley-VCH Verlag GmbH & Co., 2013); e-print arXiv:1110.5216.

[21] C. Appert-Rolland, B. Derrida, V. Lecomte, and F. Van Wijland, "Universal cumulants of the current in diffusive systems on a ring," Phys. Rev. E **78**, 021122 (2008); e-print arXiv:0804.2590.

[22] B. Derrida, B. Doucot, and P. E. Roche, "Current fluctuations in the one dimensional symmetric exclusion process with open boundaries," J. Stat. Phys. **115**, 717–748 (2003); e-print arXiv:0310453 [cond-mat].

[23] B. Derrida, "Non equilibrium steady states: Fluctuations and large deviations of the density and of the current," J. Stat. Mech.: Theory Exp. **2007**, P07023; e-print arXiv:0703762 [cond-mat].

[24] B. Derrida and A. Gerschenfeld, "Current fluctuations in one dimensional diffusive systems with a step initial density profile," J. Stat. Phys. **137**, 978–1000 (2009); e-print arXiv:0907.3294.

[25] P. L. Krapivsky, K. Mallick, and T. Sadhu, "Dynamical properties of single-file diffusion," J. Stat. Mech.: Theory Exp. **2015**, P09007; e-print arXiv:1505.01287.

[26] S. Prolhac and K. Mallick, "Cumulants of the current in the weakly asymmetric exclusion process," J. Phys. A: Math. Theor. **42**, 175001 (2009); e-print arXiv:0902.0570.

[27] V. Lecomte, A. Imparato, and F. Van Wijland, "Current fluctuations in systems with diffusive dynamics, in and out of equilibrium," Prog. Theor. Phys. Suppl. **184**, 276–289 (2009); e-print arXiv:0911.0564.

[28] P. Chleboun, S. Grosskinsky, and A. Pizzoferrato, "Lower current large deviations for zero-range processes on a ring," J. Stat. Phys. **167**, 64–89 (2017); e-print arXiv:1611.03729.

[29] R. J. Harris, A. Rákos, and G. M. Schütz, "Current fluctuations in the zero-range process with open boundaries," J. Stat. Mech.: Theory Exp. **2005**, P08003.

[30] B. Derrida and J. L. Lebowitz, "Exact large deviation function in the asymmetric exclusion process," Phys. Rev. Lett. **80**, 209 (1998); e-print arXiv:9809044 [cond-mat].

[31] S. Prolhac and K. Mallick, "Current fluctuations in the exclusion process and Bethe Ansatz," J. Phys. A: Math. Theor. **41**, 175002 (2008); e-print arXiv:0801.4659.

[32] S. Prolhac, "Tree structures for the current fluctuations in the exclusion process," J. Phys. A: Math. Theor. **43**, 105002 (2010); e-print arXiv:0911.3618.

[33] S. Prolhac, "Current fluctuations for totally asymmetric exclusion on the relaxation scale," J. Phys. A: Math. Theor. **48**, 06FT02 (2015); e-print arXiv:0101200 [cond-mat].

[34] T. Bodineau and B. Derrida, "Current large deviations for asymmetric exclusion processes with open boundaries," J. Stat. Phys. **123**, 277–300 (2006); e-print arXiv:0509179 [cond-mat].

[35] J. de Gier and F. H. L. Essler, "Large deviation function for the current in the open asymmetric simple exclusion process," Phys. Rev. Lett. **107**, 010602 (2011); e-print arXiv:1101.3235.

[36] A. Lazarescu, "The physicist's companion to current fluctuations: One-dimensional bulk-driven lattice gases," J. Phys. A: Math. Theor. **48**, 503001 (2015); e-print arXiv:1507.04179.

[37] M. Praehofer and H. Spohn, "Current fluctuations for the totally asymmetric simple exclusion process," In Out Equilib. **51**, 185–204 (2001); e-print arXiv:0101200 [cond-mat].

[38] G. B. Arous and I. Corwin, "Current fluctuations for TASEP: A proof of the Prähofer-Spohn conjecture," Ann. Probab. **39**, 104–138 (2011); e-print arXiv:0905.2993v2.

[39] Y. Baek, Y. Kafri, and V. Lecomte, "Extreme current fluctuations of boundary-driven systems in the large-N limit," J. Stat. Mech.: Theory Exp. **2016**, 053203.

[40] V. Popkov, G. M. Schütz, and D. Simon, "Asymmetric simple exclusion process on a ring conditioned on enhanced flux," J. Stat. Mech.: Theory Exp. **2010**, P10007; e-print arXiv:1007.4892.

[41] A. Lazarescu, "Generic dynamical phase transition in one-dimensional bulk-driven lattice gases with exclusion," J. Phys. A: Math. Theor. **50**, 254004 (2017).

[42] G. Bunin, Y. Kafri, and D. Podolsky, "Large deviations in boundary-driven systems: Numerical evaluation and effective large-scale behavior," Europhys. Lett. **99**, 20002 (2012); e-print arXiv:1202.0286v2.

[43] M. Gorissen and C. Vanderzande, "Current fluctuations in the weakly asymmetric exclusion process with open boundaries," Phys. Rev. E **86**, 051114 (2012); e-print arXiv:1201.6264.

[44] M. Gorissen, J. Hooyberghs, and C. Vanderzande, "Density-matrix renormalization-group study of current and activity fluctuations near nonequilibrium phase transitions," Phys. Rev. E **79**, 020101 (2009); e-print arXiv:0811.1718.

[45] C. Giardina, J. Kurchan, V. Lecomte, and J. Tailleur, "Simulating rare events in dynamical processes," J. Stat. Phys. **145**, 787–811 (2011); e-print arXiv:1106.4929.

[46] C. Giardinà, J. Kurchan, and L. Peliti, "Direct evaluation of large-deviation functions," Phys. Rev. Lett. **96**, 120603 (2006); e-print arXiv:0511248 [cond-mat].

[47] V. Lecomte and J. Tailleur, "A numerical approach to large deviations in continuous-time," J. Stat. Mech.: Theory Exp. **2007**, P03004; e-print arXiv:0612561 [cond-mat].

[48] J. Tailleur and V. Lecomte, "Simulation of large deviation functions using population dynamics," AIP Conf. Proc. **1091**, 212–219 (2009); e-print arXiv:0811.1041.

[49] T. Nemoto, V. Lecomte, S.-I. Sasa, and F. van Wijland, "Finite-size effects in a mean-field kinetically constrained model: Dynamical glassiness and quantum criticality," J. Stat. Mech.: Theory Exp. **2014**, P10001; e-print arXiv:1405.1658.

[50] T. Bodineau, B. Derrida, V. Lecomte, and F. van Wijland, "Long range correlations and phase transitions in non-equilibrium diffusive systems," J. Stat. Phys. **133**, 1013–1031 (2008).

[51] V. Lecomte, J. P. Garrahan, and F. van Wijland, "Inactive dynamical phase of a symmetric exclusion process on a ring," J. Phys. A: Math. Theor. **45**, 175001 (2012); e-print arXiv:1203.1600.

[52] G. Bunin, Y. Kafri, and D. Podolsky, "Cusp singularities in boundary-driven diffusive systems," J. Stat. Phys. **152**, 112–135 (2013); e-print arXiv:1301.1708v1.

[53] G. Bunin, Y. Kafri, and D. Podolsky, "Non-differentiable large-deviation functionals in boundary-driven diffusive systems," J. Stat. Mech.: Theory Exp. **2012**, L10001; e-print arXiv:1207.6117v1.

[54] C. Perez-Espigares, P. L. Garrido, and P. I. Hurtado, "Dynamical phase transition for current statistics in a simple driven diffusive system," Phys. Rev. E **87**, 032115 (2013).

[55] S. N. Majumdar and G. Schehr, "Top eigenvalue of a random matrix: Large deviations and third order phase transition," J. Stat. Mech.: Theory Exp. **2014**, P01012.

[56] Y. Baek, Y. Kafri, and V. Lecomte, "Dynamical symmetry breaking and phase transitions in driven diffusive systems," Phys. Rev. Lett. **118**, 030604 (2017); e-print arXiv:1609.06732.

[57] D. Karevski and G. M. Schütz, "Conformal invariance in driven diffusive systems at high currents," Phys. Rev. Lett. **118**, 030601 (2017); e-print arXiv:1606.04248.

[58] R. L. Jack, I. R. Thompson, and P. Sollich, "Hyperuniformity and phase separation in biased ensembles of trajectories for diffusive systems," Phys. Rev. Lett. **114**, 060601 (2015); e-print arXiv:1409.3986.

[59] P. Tsobgni Nyawo and H. Touchette, "A minimal model of dynamical phase transition," Europhys. Lett. **116**, 50009 (2016).

[60] J. P. Garrahan, R. L. Jack, V. Lecomte, E. Pitard, K. van Duijvendijk, and F. van Wijland, "First-order dynamical phase transition in models of glasses: An approach based on ensembles of histories," J. Phys. A: Math. Theor. **42**, 075007 (2009).

[61] C. Maes and K. Netočný, "Canonical structure of dynamical fluctuations in mesoscopic nonequilibrium steady states," Europhys. Lett. **82**, 30003 (2008).

[62] M. Feinberg, "Chemical reaction network structure and the stability of complex isothermal reactors—I. The deficiency zero and deficiency one theorems," Chem. Eng. Sci. **42**, 2229–2268 (1987).

[63] F. Schlögl, "Chemical reaction models for non-equilibrium phase transitions," Z. Phys. **253**, 147–161 (1972).

[64] D. Andrieux and P. Gaspard, "Fluctuation theorem and mesoscopic chemical clocks," J. Chem. Phys. **128**, 154506 (2008).

[65] L. F. Olsen, "An enzyme reaction with a strange attractor," Phys. Lett. A **94**, 454–457 (1983).

[66] M. H. Duong, M. A. Peletier, and J. Zimmer, "GENERIC formalism of a Vlasov–Fokker–Planck equation and connection to large-deviation principles," Nonlinearity **26**, 2951–2971 (2013).

[67] E. Levine and T. Hwa, "Stochastic fluctuations in metabolic pathways," Proc. Natl. Acad. Sci. U. S. A. **104**, 9224–9229 (2007).

[68] C. Ribrault, K. Sekimoto, and A. Triller, "From the stochasticity of molecular processes to the variability of synaptic transmission," Nat. Rev. Neurosci. **12**, 375–387 (2011).

[69] S. Huang, F. Li, J. X. Zhou, and H. Qian, "Processes on the emergent landscapes of biochemical reaction networks and heterogeneous cell population dynamics: Differentiation in living matters," J. R. Soc., Interface **14**, 20170097 (2017).

[70] A. Wachtel, R. Rao, and M. Esposito, "Thermodynamically consistent coarse graining of biocatalysts beyond Michaelis–Menten," New J. Phys. **20**, 042002 (2018).

[71] R. J. Allen and B. Waclaw, "Bacterial growth: A statistical physicist's guide," Rep. Prog. Phys. **82**, 016601 (2019).

[72] D. A. McQuarrie, "Stochastic approach to chemical kinetics," J. Appl. Probab. **4**, 413–478 (1967).

[73] D. A. McQuarrie, C. J. Jachimowski, and M. E. Russell, "Kinetics of small systems. II," J. Chem. Phys. **40**, 2914–2921 (1964).

[74] D. T. Gillespie, "A rigorous derivation of the chemical master equation," Physica A **188**, 404–425 (1992).

[75] I. Prigogine and G. Nicolis, "Self-organisation in nonequilibrium systems: Towards a dynamics of complexity," in *Bifurcation Analysis* (Springer Netherlands, Dordrecht, 1985), pp. 3–12.

[76] G. Nicolis and I. Prigogine, "Fluctuations in nonequilibrium systems," Proc. Natl. Acad. Sci. U. S. A. **68**, 2102–2107 (1971).

[77] J. Schnakenberg, "Network theory of microscopic and macroscopic behavior of master equation systems," Rev. Mod. Phys. **48**, 571–585 (1976).

[78] T. L. Hill, *Free Energy Transduction and Biochemical Cycle Kinetics* (Springer New York, New York, NY, 1989).

[79] J. Ross, *Thermodynamics and Fluctuations Far from Equilibrium*, Springer Series in Chemical Physics Vol. 90 (Springer Berlin Heidelberg, Berlin, Heidelberg, 2008).

[80] I. Oppenheim, K. E. Shuler, and G. H. Weiss, "Stochastic theory of nonlinear rate processes with multiple stationary states," Physica A **88**, 191–214 (1977).

[81] H. Ge and H. Qian, "Thermodynamic limit of a nonequilibrium steady state: Maxwell-type construction for a bistable biochemical system," Phys. Rev. Lett. **103**, 148103 (2009).

[82] M. Vellela and H. Qian, "Stochastic dynamics and non-equilibrium thermodynamics of a bistable chemical system: The Schlögl model revisited," J. R. Soc., Interface **6**, 925–940 (2009).

[83] H. Ge and H. Qian, "Non-equilibrium phase transition in mesoscopic biochemical systems: From stochastic to nonlinear dynamics and beyond," J. R. Soc., Interface **8**, 107–116 (2011).

[84] A. Nitzan, P. Ortoleva, J. Deutch, and J. Ross, "Fluctuations and transitions at chemical instabilities: The analogy to phase transitions," J. Chem. Phys. **61**, 1056–1074 (1974).

[85] H. Vroylandt and G. Verley, "Non-equivalence of dynamical ensembles and emergent non-ergodicity," J. Stat. Phys. **174**, 404–432 (2019).

[86] R. Kubo, K. Matsuo, and K. Kitahara, "Fluctuation and relaxation of macrovariables," J. Stat. Phys. **9**, 51–96 (1973).

[87] R. Dickman and R. Vidigal, "Path integrals and perturbation theory for stochastic processes," Braz. J. Phys. **33**, 73–93 (2003).

[88] E. Smith, "Large-deviation principles, stochastic effective actions, path entropies, and the structure and meaning of thermodynamic descriptions," Rep. Prog. Phys. **74**, 046601 (2011).

[89] A. Mielke, D. R. M. Renger, and M. A. Peletier, "A generalization of Onsager's reciprocity relations to gradient flows with nonlinear mobility," J. Non-Equilib. Thermodyn. **41**, 141–149 (2016).

[90] R. Rao and M. Esposito, "Nonequilibrium thermodynamics of chemical reaction networks: Wisdom from stochastic thermodynamics," Phys. Rev. X **6**, 041064 (2016).

[91] M. Assaf and B. Meerson, "WKB theory of large deviations in stochastic populations," J. Phys. A: Math. Theor. **50**, 263001 (2017).

[92] H. J. Park and A. Traulsen, "Extinction dynamics from metastable coexistences in an evolutionary game," Phys. Rev. E **96**, 042412 (2017).

[93] R. C. Kraaij and L. Mahé, "Well-posedness of Hamilton-Jacobi equations in population dynamics and applications to large deviations," e-print arXiv:1803.04525 (2018).

[94] A. Agazzi, A. Dembo, and J.-P. Eckmann, "Large deviations theory for Markov jump models of chemical reaction networks," Ann. Appl. Probab. **28**, 1821–1855 (2018).

[95] M. Doi, "Second quantization representation for classical many-particle system," J. Phys. A: Math. Gen. **9**, 1465–1477 (1976).

[96] L. Peliti, "Path integral approach to birth-death processes on a lattice," J. Phys. **46**, 1469–1483 (1985).

[97] H. Vroylandt, "Thermodynamics and fluctuations of small machines," Ph.D. thesis, Université Paris-Saclay, 2018.

[98] P. Zilber, N. R. Smith, and B. Meerson, "A giant disparity and a dynamical phase transition in large deviations of the time-averaged size of stochastic populations," Phys. Rev. E **99**, 052105 (2019).

[99] M. D. Donsker and S. R. S. Varadhan, "Asymptotic evaluation of certain Markov process expectations for large time, I," Commun. Pure Appl. Math. **28**, 1–47 (1975).

[100] M. D. Donsker and S. R. S. Varadhan, "Asymptotic evaluation of certain Markov process expectations for large time, II," Commun. Pure Appl. Math. **28**, 279–301 (1975).

[101] M. D. Donsker and S. R. S. Varadhan, "Asymptotic evaluation of certain Markov process expectations for large time—III," Commun. Pure Appl. Math. **29**, 389–461 (1976).

[102] M. D. Donsker and S. R. S. Varadhan, "Asymptotic evaluation of certain Markov process expectations for large time. IV," Commun. Pure Appl. Math. **36**, 183–212 (1983).

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

[104] R. Rao and M. Esposito, "Conservation laws and work fluctuation relations in chemical reaction networks," J. Chem. Phys. **149**, 245101 (2018).

[105] R. Rao and M. Esposito, "Conservation laws shape dissipation," New J. Phys. **20**, 023007 (2018).

[106] E. Smith and S. Krishnamurthy, "Path-reversal, Doi-Peliti generating functionals, and dualities between dynamics and inference for stochastic processes," e-print arXiv:1806.02001 (2018).

[107] D. F. Anderson, G. Craciun, and T. G. Kurtz, "Product-form stationary distributions for deficiency zero chemical reaction networks," Bull. Math. Biol. **72**, 1947–1970 (2010).

[108] V. Elgart and A. Kamenev, "Rare event statistics in reaction-diffusion systems," Phys. Rev. E **70**, 041106 (2004).

[109] M. Assaf and B. Meerson, "Extinction of metastable stochastic populations," Phys. Rev. E **81**, 021116 (2010).

[110] R. Kraaij, A. Lazarescu, C. Maes, and M. Peletier, "Deriving GENERIC from a generalized fluctuation symmetry," J. Stat. Phys. **170**, 492–508 (2018).

[111] C. Monthus, "Microscopic fluctuation theory (mFT) for interacting Poisson processes," J. Phys. A: Math. Theor. **52**, 135003 (2019); e-print arXiv:1811.04225.

[112] C. Gardiner, *Stochastic Methods: A Handbook for the Natural and Social Sciences*, 4th ed. (Springer, 2009).

[113] C. Bahadoran, "A quasi-potential for conservation laws with boundary conditions," e-print arXiv:1010.3624 (2010).

[114] J. Laurie and F. Bouchet, "Computation of rare transitions in the barotropic quasi-geostrophic equations," New J. Phys. **17**, 015009 (2015).

[115] L. S. Grigorio, F. Bouchet, R. M. Pereira, and L. Chevillard, "Instantons in a Lagrangian model of turbulence," J. Phys. A: Math. Theor. **50**, 055501 (2017).

[116] N. R. Smith and B. Meerson, "Extinction of oscillating populations," Phys. Rev. E **93**, 032109 (2016).

[117] V. M. Guibout and D. J. Scheeres, "Solving relative two-point boundary value problems: Spacecraft formulation flight transfers application," J. Guid., Control, Dyn. **27**, 693–704 (2004).

[118] U. Basu, P. de Buyl, C. Maes, and K. Netočný, "Driving-induced stability with long-range effects," Europhys. Lett. **115**, 30007 (2016).

[119] D. Nickelsen and H. Touchette, "Anomalous scaling of dynamical large deviations," Phys. Rev. Lett. **121**, 090602 (2018).

[120] B. Nguyen, U. Seifert, and A. C. Barato, "Phase transition in thermodynamically consistent biochemical oscillators," J. Chem. Phys. **149**, 045101 (2018).

[121] P. Gaspard, "Fluctuation theorem for nonequilibrium reactions," J. Chem. Phys. **120**, 8898–8905 (2004).

[122] M. Polettini, A. Wachtel, and M. Esposito, "Dissipation in noisy chemical networks: The role of deficiency," J. Chem. Phys. **143**, 184103 (2015).

[123] S. Winkelmann and C. Schütte, "Hybrid models for chemical reaction networks: Multiscale theory and application to gene regulatory systems," J. Chem. Phys. **147**, 114115 (2017).

[124] L. Andreis, W. König, and R. I. A. Patterson, "A large-deviations approach to gelation," e-print arXiv:1901.01876 (2019).

[125] T. Li and F. Lin, "Two-scale large deviations for chemical reaction kinetics through second quantization path integral," J. Phys. A: Math. Theor. **49**, 135204 (2016).

[126] P. C. Bressloff and J. M. Newby, "Path integrals and large deviations in stochastic hybrid systems," Phys. Rev. E **89**, 042701 (2014).

[127] G. Jona-Lasinio, C. Landim, and M. E. Vares, "Large deviations for a reaction diffusion model," Probab. Theory Relat. Fields **97**, 339–361 (1993).

[128] L. Canet, H. Chaté, and B. Delamotte, "General framework of the nonperturbative renormalization group for non-equilibrium steady states," J. Phys. A: Math. Theor. **44**, 495001 (2011).

[129] L. Canet, B. Delamotte, O. Deloubrière, and N. Wschebor, "Nonperturbative renormalization-group study of reaction-diffusion processes," Phys. Rev. Lett. **92**, 195703 (2004).

[130] D. Gredat, H. Chaté, B. Delamotte, and I. Dornic, "Finite-scale singularity in the renormalization group flow of a reaction-diffusion system," Phys. Rev. E **89**, 010102 (2014).

[131] G. Falasco, R. Rao, and M. Esposito, "Information thermodynamics of turing patterns," Phys. Rev. Lett. **121**, 108301 (2018).

[132] F. Avanzini, G. Falasco, and M. Esposito, "Thermodynamics of chemical waves," e-print arXiv:1904.08874 (2019).