Dynamics of systems with varying number of particles: From Liouville equations to general master equations for open systems

Mauricio J. del Razo[1,2]⋆ **and Luigi Delle Site**[1]†

1 Freie Universität Berlin, Department of Mathematics
and Computer Science, Berlin, Germany
2 Zuse Institute Berlin, Germany

⋆ m.delrazo@fu-berlin.de , † luigi.dellesite@fu-berlin.de

Abstract

A varying number of particles is one of the most relevant characteristics of systems of interest in nature and technology, ranging from the exchange of energy and matter with the surrounding environment to the change of particle number through internal dynamics such as reactions. The physico-mathematical modeling of these systems is extremely challenging, with the major difficulty being the time dependence of the number of degrees of freedom and the additional constraint that the increment or reduction of the number and species of particles must not violate basic physical laws. Theoretical models, in such a case, represent the key tool for the design of computational strategies for numerical studies that deliver trustful results. In this manuscript, we review complementary physico-mathematical approaches of varying number of particles inspired by rather different specific numerical goals. As a result of the analysis on the underlying common structure of these models, we propose a unifying master equation for general dynamical systems with varying number of particles. This equation embeds all the previous models and can potentially model a much larger range of complex systems, ranging from molecular to social agent-based dynamics.

1 Introduction

The conception of many-particle (or equivalently, many-body) systems has played a central role in modern theoretical and mathematical physics, from the many-electron systems of quantum mechanics to the tracers of fluid dynamics. Many-particle models are ubiquitous in condensed matter and provide the paradigm of reference for studying the properties of any existing substance. It is also an interesting approach per se in assessing the cascade of scales that characterize the physics of matter, since, in principle, it allows to systematically pass from the microscopic quantum mechanical resolution, up to the continuum hydrodynamics [1]. The models used in physics are in general based on an interaction potential, usually of a two-body nature, (e.g. electrostatic, Lennard-Jones) and the corresponding interlinked Newton's equations of motion (or Schrödinger equations in quantum mechanics). This means that particle systems can be mathematically described as dynamical systems moving along calculable trajectories, enabling powerful mathematical and physical machinery to model, simulate and analyze them.

Most of the machinery developed to model and simulate molecular or particle-based dynamical systems focuses on systems with a constant number of particles, this means they are closed but not necessarily isolated from heat exchange. However, this is not the case in many real-world applications, where we often have to focus on subsystems that exchange material with their surroundings; we refer to such systems as open systems. For instance, living cells constantly exchange molecules and energy with their environment; they consume chemical energy and dissipate heat. In terms of physical chemistry, every living system must be a nonequilibrium open system —a closed system has no life [2]. These nonequilibrium processes at the molecular scale often drive fundamental phenomena such as symmetry breaking, phase transitions and entropy production with profound impact on our understanding of living systems at meso- and macroscopic scales [3–5]. Moreover, the exchange of heat and matter is one of the main processes driving weather and climate phenomena, such as tipping events [6]. This manifests in the need for multiscale models [7], as well as reduced/coarse-grained models [8] and stochastic closure techniques [9]. Thus, to understand, model and simulate these processes, it is fundamental to develop mathematical and physical machinery to handle systems with varying particle number.

From a mathematical perspective, it is a complex problem. The main difficulty is the on-the-fly change in the number of degrees of freedom of the system. At a differential equation level, the number of equations would change as the system changes particle number. The solution is to lift the dynamics to the space of densities/distributions, where the dynamics become linear, albeit infinite-dimensional and thus complicated to analyze. This is analogous to switching representation from a nonlinear dynamical system to the linear Liouville equation, or in the case of stochastic dynamical systems to a Fokker-Planck equation. However, note some information is lost in the process, as we can no longer track individual particles. Following this reasoning, some theoretical machinery has been developed in previous works to handle systems with varying particle number. We will overview the main ones in this manuscript.

Other similar approaches often employ methods based on quantum field theory [10–13], which are not a requirement to neither write nor understand the equations. From a computational perspective, one possible solution would be to consider a large enough closed subsystem that includes a substantial part of the environment. In this context, numerical methods of particle simulation have enormously risen in the past decades, improving our capability of performing precise calculations of large many-particle systems (see e.g. [14, 15]). However, despite the growing computational power, many systems of major interest are not yet affordable with standard available models, so we need alternative solutions that combine novel mathematical developments with computational approaches.

In recent years, multiscale models and simulations have been developed to improve the efficiency of particle-based simulations. The basic idea is that one retains the degrees of freedom strictly required by the problem and simplifies the degrees of freedom that are not directly involved in the process of interest at coarser levels [16]. A prototype situation in most of the many-particle models is the necessity of coarse-graining the environment around the region of interest while retaining all the necessary details in the region itself; since the region of interest is subject to particle number fluctuations, the necessity of developing open many-particle systems rises naturally [17]. Moreover, once a mathematical model for the dynamics with varying number of particles is developed, it can organically be adapted and applied to systems where the particle number fluctuations are provided not only by the external environment but also by the change of composition of a system, when for example in a mixture different species interact forming a third one. Despite the high dimensionality of the resulting mathematical model, it can be used as a guiding framework to unify models at multiple scales and to derive physically consistent multiscale numerical schemes.

The concept of subsystem is closely related to the concept of subdynamics for which there is a vast literature available. For example, the model proposed by Prigogine and coworkers [18, 19], the model of Emch and Sewell [20], and the model of Robertson [21, 22], to cite but a few. The idea is based on the projection method of Zwanzig [23] for quantum systems (used also for classical systems). The evolution of the probability density of the system in the von Neumann or Liouville equation is projected on a subspace with a reduced number of variables and the rest of the system is formally coarse-grained. The models discussed in this paper share the idea of integrating/coarse-graining the degrees of freedom outside the subsystem, however they add the explicit treatment of the number of particles as a variable of the problem. This characteristic is not explicitly mentioned in the literature cited above, thus leaving the impression that the number of particles of the subsystem is assumed to be fixed. In a system with varying number of particles one should explicitly discuss the normalization of the probability density of the subsystem; this operation is substantially different from the case of a fixed number of particles. In addition, the equations proposed in [18–22], while mathematically rigorous and certainly elegant, are characterized by a rather complex structure difficult to implement in modern computational simulation techniques. Instead, for computational implementation, one needs models that are characterized by a certain level of mathematical rigor but at the same time are constructed to efficiently capture the relevant physics. The model of Bergman and Lebowitz [24, 25], discussed later on, can be considered the first historical attempt towards such an effort and became a source of inspiration for most of the progress reported here.

In this work, we first review two frameworks to handle classical systems with varying number of particles (motivated by different physical problems and numerical approaches). We introduce these frameworks and investigate the relations between them to show how a general master equation for systems varying number of particles emerges. The two main different points of view correspond to: (i) the approach based on Liouville-like equations of a subsystem embedded in a reservoir where the degrees of freedom are either explicitly integrated out,

reducing the effect of the environment on the boundary conditions of the subsystem of interest [26, 27], or implicitly integrated out into a coupling term [24]; and (ii) a master equation where the diffusion process based on the Fokker-Planck operator is coupled with a reaction process that can change the number of particles. In this case, the particle description of the environment is not considered and its effect is modeled empirically *a priori* [12, 28]. The first approach is inspired by molecular simulations where the microscopic single molecular trajectories with explicit chemical details are accessible and thus the probability distribution function of the phase space of system and environment can be explicitly sampled [29]. This means that a simulation of an open system with the simplified environment obtained analytically by formal integration of its related degrees of freedom should deliver the same result as an equivalent subsystem in a simulation of the entire system; as a consequence, the validity of the theoretical model of the open system can be directly tested numerically (see e.g. [30]). The second approach is inspired by problems occurring at a larger scale than the microscopic molecular scale, e.g. processes at the scale of biological living cells, where the number of molecules is large enough to render molecular models intractable but not large enough to consider macroscopic approaches that neglect inherent stochastic fluctuations [31]. From a physics perspective, in this approach, molecules are represented as particles undergoing diffusion and chemical reactions are coarse-grained into events that simply change the chemical composition of the system. Thus, the effect of the environment is modeled empirically through the diffusion constant, the reaction rates and the stochastic effects. This model does not only provide a probabilistic model for reaction-diffusion processes [12, 28] but also serves as a starting point to derive other models at different scales. This yields a unifying theory that serves as the backdrop to derive numerical schemes that are consistent across multiple scales [32, 33, 51].

Independently from the original inspiration, both models deal with varying number of particles and thus must retain a common general structure within a unique framework. In this work, we proceed with the formal analysis of the similarities and differences between the two approaches. Based on this comparison, a generalized structure emerges in the form of a unitary equation, from which the previous approaches are special cases. This generalized structure can be used for any dynamical system in an open setting, enabling a broader range of applications within physics and beyond. To finalize, we discuss mathematical perspectives and potential problems of interest in physical and chemical applications both from an analytical and numerical point of view. We further discuss possible applications of the general master equation beyond the fields that initially motivated the equation.

The paper is structured as follows. To motivate the reader, section 2 focuses on an intuitive overview of the final general master equation for systems with varying particle number, which emerges naturally in section 5. However, the equation itself is originally motivated by the specific approaches presented in sections 3 and 4. More specifically, in section 3, we explore two approaches to handle systems with Hamiltonian structure that are in contact with a material reservoir(s). In section 4, we show how to write the master equation to describe the probabilistic dynamics of reaction-diffusion processes, and we extend this to systems where the diffusion is governed by Langevin dynamics. In section 5, we show how the mathematical similarities between the approaches from sections 3 and 4 inspire formulating the general master equation and how it recovers the previous approaches as special cases. Finally, in section 6, we discuss the perspectives and future scope of this work in the context of physical modeling, numerical simulations and applications in other fields.

2 A general master equation for systems with varying number of particles

The general master equation presented in this section was inspired by the frameworks presented in sections 3 and 4. For educational purposes, we first carry out an intuitive derivation of the equation. We formulate the master equation for systems with varying number of particles starting from a general random dynamical system perspective. This together with section 5 constitute the main results of the paper.

We first consider a random dynamical system with fixed dimension n, e.g. a system with a fixed number of particles. The dynamics of the trajectories of the systems are given by the system of stochastic differential equations

$$\dot{\boldsymbol{x}} = F(\boldsymbol{x}), \tag{1}$$

where $\boldsymbol{x}$ could represent many things, depending on the application at hand. In physics, it could represent the position of particles or perhaps the positions and velocities of particles. In other fields like social systems, it could represent the position of an agent in opinion space [34] or the population of individuals of a certain species [35]. The function F incorporates the deterministic drift as well as noise components. One can equivalently write the dynamics of the probability distribution of the system in phase space given by the corresponding Fokker-Planck equation

$$\frac{\partial}{\partial t} f_n = \mathcal{A}_n f_n, \tag{2}$$

with $\mathcal{A}_n$ the infinitesimal generator of the Fokker-Planck equation for the corresponding n particle system, and $f_n := f_n(t, \boldsymbol{x})$ the probability density of the system being at $\boldsymbol{x}$ at time t, which naturally integrates to one when integrating over the whole phase space. Reasonable boundary conditions for f_n are usually reflective in a bounded domain or vanishing at infinity for a non-bounded domain.

To write a similar description for a system with a variable number of particles, one cannot write it in the form of eq. (1) since the dimension of the system changes as time evolves. However, we can write something similar to the master equation from eq. (2). Consider the family of distributions $f = (f_0, f_1, \ldots, f_n, \ldots)$, where $f_n := f_n(t, \boldsymbol{x}^n)$ is the probability density of having n particles (or whatever else we are modeling) at positions or states $\boldsymbol{x}^n = (x_1, \ldots, x_n)$ at time t. The phase space now has a much more complex structure with continuous and discrete degrees of freedom, so the normalization condition is

$$\sum_{n=0}^{\infty} \int f_n(t, \boldsymbol{x}^n) d\boldsymbol{x}^n = 1, \tag{3}$$

where the integral is over the whole available space. The master equation for the system with varying particle number can then be written as a family of Fokker-Planck equations, one for each n-particle level. These are then coupled by operators $\mathcal{Q}_{nm}$ that transfer probability from the m-particle state to the n one modeling processes that change the particle number (e.g. interactions with a material reservoir or reactions). These operators form a coupling matrix, which incorporates into the master equation

$$\frac{\partial}{\partial t} \begin{pmatrix} f_0 \\ f_1 \\ \vdots \\ f_n \\ \vdots \end{pmatrix} = \begin{pmatrix} \mathcal{A}_0 f_0 \\ \mathcal{A}_1 f_1 \\ \vdots \\ \mathcal{A}_n f_n \\ \vdots \end{pmatrix} + \begin{pmatrix} \mathcal{Q}_{00} & \mathcal{Q}_{01} & \cdots & \mathcal{Q}_{0n} & \cdots \\ \mathcal{Q}_{10} & \mathcal{Q}_{11} & \cdots & \mathcal{Q}_{1n} & \cdots \\ \vdots & \vdots & & \vdots & \cdots \\ \mathcal{Q}_{n0} & \mathcal{Q}_{n1} & \cdots & \mathcal{Q}_{nn} & \cdots \\ \vdots & \vdots & & \vdots & \ddots \end{pmatrix} \begin{pmatrix} f_0 \\ f_1 \\ \vdots \\ f_n \\ \vdots \end{pmatrix}. \tag{4}$$

Each row of the equation corresponds to the dynamics of the n-particle distribution. If the coupling matrix is removed, one obtains a set of uncoupled Fokker-Planck equations, each for a different particle number. If the space dependence is removed, one recovers a continuous-time Markov chain, where the operators $\mathcal{Q}_{nm}$ become transition rates. This master equation for systems with varying particle number can be written more compactly by focusing on the nth component

$$\left(\frac{\partial}{\partial t} - \mathcal{A}_n\right) f_n = \mathcal{Q}_n f \,, \tag{5}$$

with $\mathcal{Q}_n f = \sum_{m=0}^{\infty} \mathcal{Q}_{nm} f_m$. Although these infinite sums might seem problematic at first, in practical cases most $\mathcal{Q}_{nm}$ are zero and the sums remain finite, as we will see later. Moreover, in analogy with stochastic matrices in continuous-time Markov chains, we expect each matrix column to sum to zero to enforce the conservation of probability. As in this case, we have a matrix of operators and space dependence, the probability conserving condition translates into the following constraint imposed into the coupling operators

$$\sum_{n=0}^{\infty} \int (\mathcal{Q}_{nm}\eta_m)(\boldsymbol{x}^n) d\boldsymbol{x}^n = 0 \,, \tag{6}$$

for η_m any given test arbitrary probability density for m particles and where the integrals run over the whole available space. Each of these integrals represents the probability flux associated with a transition. By forcing their sum to be zero, we enforce that the net probability flux leaving one state(s) is the same as the one entering another state(s). Thus, enforcing probability conservation. As a final note, there is a connection between this description and hybrid switching diffusions [36–39]. However, in the particle context, hybrid switching diffusions can only account for a change in the state of a particle but cannot incorporate or remove particles.

In the next sections we will review the derivation of master equations of this form written for specific contexts, which have inspired the generalization above that applies to a larger class of systems with varying number of particles [10, 12, 25–28]. In sections 3 and 4, we will explore specific physical systems that involve varying particle number. Then, in section 5, we will show how these specific systems inspired the general master equation (eq. (5)) and how they are recovered as particular cases.

3 Liouville-like equations for classical open systems

The two approaches reported in this section were developed (the first one) or used (the second one) to conceptually frame a numerical method for the molecular simulation of open systems that exchange energy and matter with a large reservoir. The specific numerical code embedded in this theoretical framework is the Adaptive Resolution Simulation (AdResS) [29, 40, 41]; however, any molecular simulation approach that is characterized by the system-reservoir exchange of energy and particles could be framed as well in such models. The models presented here have been used as inspiration for designing and rationalizing the system-reservoir coupling [30, 42, 43]. Systems of molecular simulation are in general characterized by an explicit particle-particle Hamiltonian and by its corresponding phase-space probability density. This latter is not known explicitly, however it is statistically sampled either through a single long trajectory or through a collection of short trajectories, each with an initial condition uncorrelated to the initial condition of the others. As a consequence the statistical calculation of physical quantities is done by sampling and averaging the physical quantity of interest along
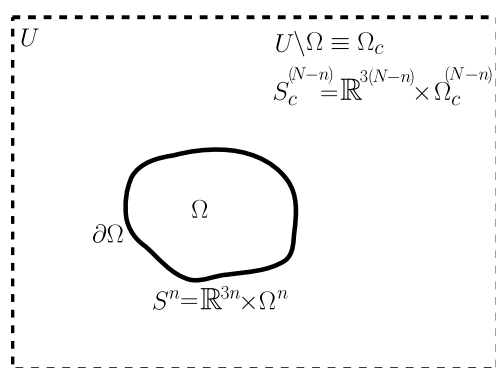
Figure 1: Graphical representation of the open system and associated formalism. The open system Ω, with its boundary surface $\partial\Omega$, is defined as a subsystem of U with n particles. The reservoir is defined as a large system, $U\backslash\Omega = \Omega_c$, with $N-n$ particles, since U contains N particles. The n dimensional domain of the phase-space of particles in Ω is defined as $S^n = \mathbb{R}^{3n} \times \Omega^n$ while the $N-n$ dimensional domain of the phase-space of particles in Ω_c is defined as $S_c^{N-n} = \mathbb{R}^{3(N-n)} \times \Omega^{(n-N)}$. This figure is adapted from Fig.1 of [27].

such trajectories [14]. The natural complete framework for the treatment of these systems is the Liouville equation, thus it seemed natural to manipulate the Liouville equation of the total system to obtain an equivalent equation for a subsystem where the surrounding (rest of the total system) has been explicitly integrated out in its particle degrees of freedom or, alternatively, the particle degrees of freedom have been empirically removed by modeling the reservoir as a generic thermodynamic bath.

3.1 From a large system of N particles to an open subsystem of n particles

The model of open system based on the Liouville equation of the total system of [26] is here described in its essential features. Let us consider a large dynamical system of N particles in equilibrium (called here Universe, U) and define an open subsystem, Ω, containing n particles (with the corresponding reservoir, $U\backslash\Omega = \Omega_c$, of $N-n$ particles, with $N >> N-n$), as illustrated pictorially in fig. 1. Starting from the Liouville equations for the probability in phase space of the Universe and integrating out all the degrees of freedom of $U\backslash\Omega = \Omega_c$ one would wish to derive a Liouville-like equation for the particles in Ω taking into account that particles can freely move from the subsystem to the reservoir and vice versa. The Hamiltonian of the Universe is defined as:

$$H_N = \sum_{i=1}^{N} \frac{\vec{p}_i^2}{2M} + V_{\text{tot}}(\mathbf{q}^N), \tag{7}$$

with $\vec{p}_i$ the momentum of the i-th particle, M is the mass of an individual particle; the potential of particle-particle interaction is defined as: $V_{\text{tot}}(\mathbf{q}^N) = \sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N} \frac{1}{2}V(\vec{q}_j - \vec{q}_i)$, with i and j labeling two different molecules with corresponding positions $\vec{q}_i$ and $\vec{q}_j$. The corresponding probability density in phase-space is given by:

$$F_N : \mathbb{R}^+ \times (U \times \mathbb{R}^3)^N \to \mathbb{R}(t, \mathbf{X}^N) \mapsto F_N(t, \mathbf{X}^N), \tag{8}$$

where

$$\int_{S^N} F_N \, d\mathbf{X}^N = 1, \tag{9}$$

with $\mathbf{X}^N \in S^N = (U \times \mathbb{R}^3)^N$ the space of the position and momentum variables of the N particles.

The Liouville equation for U is then given by:

$$\frac{\partial F_N}{\partial t} = \Lambda_N F_N\,, \tag{10}$$

with Λ_N the Liouville operator for an N particle system

$$\Lambda_N F_N = \{F_N, H_N\} = \sum_{i=1}^{N} \left[\nabla_{\vec{q}_i} \cdot (\vec{v}_i F_N) + \nabla_{\vec{p}_i} \cdot (\vec{F}_i F_N)\right]\,, \tag{11}$$

$\{*,*\}$ are the canonical Poisson brackets, $\vec{v}_i = \vec{p}_i/M_i$ is the velocity of the ith particle and the total force, $\vec{F}_i$, acting on particle i is given by

$$\vec{F}_i = -\nabla_{q_i} V_{\text{tot}}(\boldsymbol{q}^N) = -\sum_{j=1;j\neq i}^{N} \nabla_{\vec{q}_i} V(\vec{q}_i - \vec{q}_j)\,. \tag{12}$$

The Hamiltonian of the subsystem Ω is accordingly defined as:

$$H_n = \sum_{i=1}^{n} \frac{\vec{p}_i^2}{2M} + \sum_{i=1}^{n}\sum_{j\neq i}^{n} \frac{1}{2} V(\vec{q}_j - \vec{q}_i) \qquad (\vec{q}_i, \vec{q}_j \in \Omega)\,, \tag{13}$$

while the corresponding phase-space probability density is:

$$f_n : \mathbb{R}^+ \times S^n \to \mathbb{R}; \ (t, \boldsymbol{X}^n) \mapsto f_n(t, \boldsymbol{X}^n), \qquad \text{for} \qquad (n = 0, \ldots, N),$$

$$f_n(t, \boldsymbol{X}^n) = \binom{N}{n} \int_{(S_c)^{N-n}} F_N(t, \boldsymbol{X}^n, \boldsymbol{\Xi}_n^N)\, d\boldsymbol{\Xi}_n^N\,, \tag{14}$$

$$\boldsymbol{\Xi}_n^N \equiv [\Xi_{n+1}, \ldots, \Xi_{N-n}], \qquad \text{where} \qquad \Xi_i = (\vec{q}_i, \vec{p}_i) \in S_c^{N-n}\,.$$

The collection of n-particle functions, $(f_n)_{n=0}^N$ defines the probability density with the normalization condition derived from the normalization of F_N for the universe (e.g. (9)). As a consequence one has that:

$$\sum_{n=0}^{N} \int_{\Omega^n} \int_{(\mathbb{R}^3)^n} f_n(t, (\boldsymbol{q}, \boldsymbol{p}))\, d\boldsymbol{p}\, d\boldsymbol{q} = 1\,. \tag{15}$$

The integration of eq. (10) w.r.t. the variables of S_c^{N-n} implies a straightforward calculation procedure which does not carry any relevant conceptual aspects and thus it is not reported here; the corresponding details of can be found in [26, 43], here the final results are reported:

$$\frac{\partial f_n}{\partial t} + \Lambda_n f_n = \Psi_n + \Phi_n^{n+1}\,, \tag{16}$$

with

$$\Psi_n = -\sum_{i=1}^{n} \nabla_{\vec{p}_i} \cdot \left(\vec{F}_{\text{av}}(\vec{q}_i) f_n(t, \boldsymbol{X}^{i-1}, X_i, \boldsymbol{X}_i^{n-i})\right), \tag{17}$$

and

$$\vec{F}_{\text{av}}(\vec{q}_i) = -\int_{S^c} \nabla_{\vec{q}_i} V(\vec{q}_i - \vec{q}_j) f_2^{\circ}(X_j|X_i)\, dX_j\,. \tag{18}$$

The latter corresponds to the mean field force that the outer particles exert onto the ith particle in Ω. The quanity $f_2^\circ(X_{\text{out}}|X_{\text{in}})$ is defined as the conditional distribution of an outer particle at the given state of an inner particle; one can assume that it is known or a model can be proposed. Furthermore,

$$\Phi_n^{n+1} = (n+1) \int_{\partial\Omega} \int_{(\vec{p}_i \cdot \vec{n}) > 0} (\vec{p}_i \cdot \vec{n}) \left(f_{n+1}(t, X^n, (\vec{q}_i, \vec{p}_i)) - f_n(t, X^n) f_1^\circ(\vec{q}_i, -\vec{p}_i) \right) d^3 p_i d\sigma_i . \quad (19)$$

In eq. (19) $f_1^\circ(\vec{q}_i, -\vec{p}_i)$ corresponds to the (modeler's) assumption of the reservoir's one particle distribution calculated at the interface boundary $\partial\Omega$. It must be also added that the addition of an external thermostat in the reservoir does not change the equation and allows us to extend the model to situations of non-equilibrium like thermal gradient [27]. In general, such a theoretical framework rationalizes the AdResS molecular dynamics protocol, but in principle, as said above, it can be applied to any other molecular simulation method that aims at considering a varying number of molecules.

3.2 The Bergman-Lebowitz equation of open systems

Bergmann and Lebowitz, and Lebowitz and Shimony in two seminal papers [24, 25], have proposed an equation based on the extension of the Liouville equation that models open systems embedded in a reservoir of particles and energy. The essence of the model is to consider an impulsive, Markovian interaction between the reservoir and the system; the reservoir is considered stationary and not influenced by the changes occurring in the system, thus the thermodynamic state point of the reservoir is fixed. The interaction between the system and the reservoir is modeled as a discontinuous transition of the system from a state with N particles (X^n) to one with M particles (Y^m), where X^n corresponds to the canonical variables (position and momenta). The change of the state of the system state is described by a time-independent Markovian transition kernel, $K_{nm}(X^n, Y^m)$. The kernel, $K_{nm}(X^n, Y^m)$, corresponds to the probability per unit time that the system at Y^m makes a transition to X^n as a result of the interaction with the reservoir. The probability density, $f_n(X^n, t)$, in some point X^n of the phase space is then regulated by an extension of the Liouville equation:

$$\frac{\partial f_n(X^n, t)}{\partial t} = \Lambda_n f_n + \sum_{m=0}^{\infty} \int dY^m [K_{nm}(X^n, Y^m) f_m(Y^m, t) - K_{mn}(Y^m, X^n) f_n(X^n, t)] . \quad (20)$$

The first term of eq. (20) corresponds precisely to the Liouville equation for the n particle system (eq. (11)), while the second term establishes a coupling with the reservoir(s). This last term is composed of negative or positive contributions corresponding either to the outflux/loss of probability from the current n state or to the influx/gain of probability into the current state, respectively.

If the kernel satisfies the condition of flux balance:

$$\sum_{m=0}^{\infty} \int [e^{-\beta H(X^m) + \beta\mu m} K_{nm}(X^n, Y^m) - K_{mn}(Y^m, X^n) e^{-\beta H(X^n) + \beta\mu n}] dX^m = 0 , \quad (21)$$

it follows that the stationary Grand Ensemble is the Grand Canonical ensemble; here $\beta = \frac{1}{k_B T}$ where k_b is the Boltzmann constant and T the temperature, while μ is the chemical potential of the system. The key difference between the approach of section 3.1 and the approach of Bergman and Lebowitz lies in the assumption about the reservoir and the corresponding term of system-reservoir exchange in the equation. The model of section 3.1 does not require any assumption about the reservoir, but directly integrates its particle's degrees of freedom.

As a consequence, the system-reservoir coupling term is explicitly written in terms of particle quantities without any stochastic assumption. In the model shown here, the reservoir is modeled *a priori* without any explicit link to its particle resolution and the system-reservoir coupling term is modeled, as a consequence, with a probabilistic process. The model of Bergman and Lebowitz has been very important for the numerical implementation of the open system approach AdResS in an intermediate step. Such a model, differently from the model of section 3.1, could not provide an explicit numerical receipt for the system-reservoir coupling, nevertheless it allowed for a physical qualitative interpretation of the coupling conditions of the AdResS code [42, 44].

4 Master equations for reaction-diffusion processes

Molecular dynamics are limited to the study of complex biochemical phenomena at the scale of living cells. While the most sophisticated molecular simulations can achieve simulations of a few macromolecules in the scale of micro- or milliseconds when being optimistic, biochemical processes at these scales often involve thousands or millions of macromolecules and occur over timescales of seconds. Moreover, the chemical events at the molecular scale happen at a much faster scales than those relevant to life processes, and thus their detailed molecular kinetics do not play a key role in the dynamics. It is thus appropriate to consider particle-based reaction-diffusion models (fig. 2), where molecules are represented as particles undergoing random motion due to thermal fluctuations of the solvent (diffusion), and chemical reactions due to instantaneous reaction events, which often occur after chancy encounters between two or more molecules. The relevant features of the molecular scale are captured in the diffusion coefficients and the reaction rate functions.

Particle-based reaction-diffusion models are often the standard model to describe biochemical processes at the scale of living cells. At these scales, the so-called chemical diffusion master equation (CDME) provides a probabilistic model in terms of the number of particles of each of the chemical species involved, as well as their spatial configuration [10, 12, 28]. One of the original objectives of developing the CDME was to serve as an underlying ground model upon which one can construct particle-based reaction-diffusion simulations consistently. Moreover, one can in principle recover most other reaction-diffusion models as limits of the CDME, e.g. in the well-mixed limit, one recovers the well-known chemical master equation (CME) [31, 45–47]; in the thermodynamic limit, one recovers reaction-diffusion PDEs; and when doing spatial discretizations, one recovers the reaction-diffusion master equation [48] or the spatiotemporal master equation [49]. Thus, it serves as a unifying framework from which one can develop numerical schemes that are consistent across scales. For instance, some mathematicians are interested in branching and annihilating Brownian motion [50] and its connection to the Kolmogorov-Petrovsky-Piskunov-Fisher (KPPF) equation. Following an applied mathematics/mathematical physics approach, one can frame these processes in terms of the CDME and study the limiting cases to recover the KPPF reaction-diffusion PDE, as done recently for other reaction systems [51]. This could yield alternative insights to the research community in that field, as well as methodologies for multiscale simulations.

One of the main virtues of the CDME is its capability to handle systems with varying number of particles while maintaining spatial resolution. This is an inherent characteristic of the CDME since reaction events often change the number of particles in the system. In this section, we will overview the CDME for a simple example, and we will show how it extends when the diffusion term is replaced by Langevin dynamics.
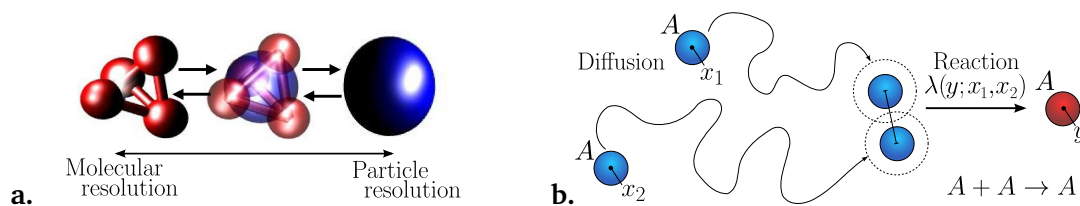
Figure 2: Illustration of particle-based models. **a.** Change of resolution from molecular to particle resolution, where each molecule is considered as one bead. **b.** A particle-based reaction-diffusion process for $A+A \to A$, where $\lambda(y, x_1, x_2)$ is the position dependent rate function. Figure (a) is adapted from Fig.1(a) of [52].

4.1 Chemical diffusion master equation: coupling diffusion with reaction processes

As a starting point, we follow [28] and consider a system with a varying number of particles, all corresponding to one chemical species, enclosed in a finite domain $\mathbb{X}$ (with reflecting boundaries in the boundaries of $\mathbb{X}$). The configuration of the system is given by the number of particles and their positions, so its probability distribution is given as an ordered family of probability density functions:

$$f = (f_0, f_1, f_2, \ldots, f_n, \ldots),$$ (22)

where $f_n := f_n(t, \boldsymbol{q}^n)$ is the probability density of finding n particles at the positions $\boldsymbol{q}^n = (q_1, \ldots, q_n) \in \mathbb{X}^n$ at time t, or simply $\boldsymbol{q}$ if clear from context. The phase space of these distributions is depicted in fig. 3. As the particles are statistically indistinguishable from each other, the densities must be symmetric with respect to permutations of labels, for instance, $f_3(t, x, y, z) = f_3(t, x, z, y) = f_3(t, y, x, z)$ and so on. The probability distribution should be normalized, thus

$$\sum_{n=0}^{\infty} \int_{\mathbb{X}^n} f_n(t, \boldsymbol{q}) d\boldsymbol{q} = 1.$$ (23)

For a system with M reactions, the nth component of the CDME has the general form

$$\frac{\partial f_n}{\partial t} = \mathcal{D}_n f_n + \sum_{r=1}^{M} \mathcal{R}_n^{(r)} f.$$ (24)

The generator of the CDME is decomposed into the reaction and diffusion components: $\mathcal{D}$ is the diffusion operator and $\mathcal{R}^{(r)}$ corresponds to the reaction operator for the rth reaction. Each reaction operator can be split into two contributions

$$\mathcal{R}_n^{(r)} = \mathcal{G}_n^{(r)} - \mathcal{L}_n^{(r)},$$ (25)

representing the gain or loss of probability respectively due to the rth reaction in a given configuration.

To exemplify the formulation of the CDME for a reaction involving only one chemical species, we focus on the following reaction (fig. 2b.)

$$A + A \to A.$$ (26)
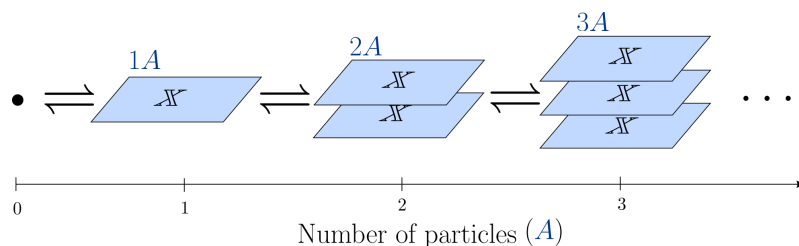
Figure 3: Phase space for a general reaction-diffusion process involving one chemical species. The space $\mathbb{X}$ represents the phase space of one particle A (e.g. $\mathbb{R}^3$), and the distributions described by the chemical diffusion master equation reside in this phase space. Figure adapted from Fig.1(a) of [51].

The rate at which a reaction event occurs is given by $\lambda(y; \boldsymbol{x}) > 0$, and it depends on the positions $\boldsymbol{x} = (x_1, x_2) \in \mathbb{X}^2$ of the reactants and the position $y \in \mathbb{X}$ of the products. Note that the rate function must be symmetric with respect to the positions of the reactants (as well of the products if more than one), i.e. $\lambda(y; x_1, x_2) = \lambda(y; x_2, x_1)$. The nth component of the CDME then is given by

$$\frac{\partial f_n}{\partial t} = \mathcal{D}_n f_n + \mathcal{G}_n f_{n+1} - \mathcal{L}_n f_n, \tag{27}$$

where $\mathcal{D}_n$, $\mathcal{G}_n$, $\mathcal{L}_n$ refer to the corresponding diffusion, gain and loss operators, respectively. Note as we are only considering one reaction, we dropped the upper index. Reactions at the n-particle state transfer probability to the $(n-1)$-particle state, so they are represented by the loss term. Reactions at the $(n+1)$-particle state transfer probability to the n-particle state state, so they are represented by the gain.

For non-interacting particles, the diffusion operator $\mathcal{D}_n$ is the infinitesimal generator of the n-particle Fokker-Planck equation.

$$\mathcal{D}_n f_n = \sum_{i=1}^{n} \nabla_i \cdot (A_i f_n) + \sum_{i,j=1}^{n} \nabla_i \cdot \left(D_{ij} \nabla_j f_n \right), \tag{28}$$

where $A_i = A_i(\boldsymbol{q}, t)$ is the drift, $D_{ij} = D_{ij}(\boldsymbol{q})$ are the 3×3 diffusion matrices and ∇_i denotes differentiation with respect to ith component of the position $\boldsymbol{q}$ [12]. If the drift is consequence of an interaction potential $U(\boldsymbol{q})$ then

$$A_i = -\sum_{j=1}^{n} D_{ij} \nabla_j U. \tag{29}$$

In the absence of drift and assuming isotropic diffusion, $\mathcal{D}_n$ would simply be the Laplacian, $\mathcal{D}_n = \sum_{i=1}^{n} D \nabla_i^2$ with a scalar diffusion constant D. Considering this structure and assuming there are no reactions, the CDME would correspond to an infinite family of uncoupled Fokker–Planck equations for the particle positions, where each member of the family corresponds to a different number of particles.

The loss operator acting on the n-particle density will output the total rate of probability loss of f_n due to all possible combinations of reactants. It is given in terms of the loss per reaction $L_{i,j}$, which acts on 2 particles at a time, with (i, j) denoting the indexes of the particles that it acts on. The loss per reaction quantifies how much probability is lost to the current state due to one reaction, it is thus the integral over the density and the rate function λ over all the possible positions of the products:

$$\left(L_{i,j} f_n \right)(\boldsymbol{q}) = f_n(\boldsymbol{q}) \int_{\mathbb{X}} \lambda(y; \boldsymbol{q}_{i,j}) dy, \tag{30}$$

with $\boldsymbol{q} \in \mathbb{X}^n$ and where $\boldsymbol{q}_{i,j} = (q_i, q_j)$ represent the ith and jth components of $\boldsymbol{q}$. The total loss is then the sum of the loss per reaction over all possible reactions,

$$\mathcal{L}_n = \sum_{1 \leq i < j \leq n} L_{i,j} \,. \tag{31}$$

The form of the ordered sum guarantees that we count all the possible ways of picking up k particles without double counting. Similarly, the gain operator acting on the n-particle density will output the total rate of probability gain of f_n. It can be expressed in terms of the gain per reaction resulting from 2 reacting particles producing one product with index k. The gain per reaction, termed G_k, quantifies how much probability is gained by the current state due to one reaction, it is thus the integral over the density and the rate function λ over all the possible positions of the reactants:

$$(G_k \rho_{n+1})(\boldsymbol{q}) = \int_{\mathbb{X}^2} \lambda(q_k; \boldsymbol{z}) \rho_{n+1}(\boldsymbol{q}_{\backslash\{k\}}, \boldsymbol{z}) d\boldsymbol{z} \,, \tag{32}$$

with $\boldsymbol{q} \in \mathbb{X}^n$ and where the subscript $\backslash\{k\}$ means that the entry with index k is excluded from the tuple $\boldsymbol{q}$ of particle positions. The total gain is then the sum of the gain per reaction over all possible reactions,

$$\mathcal{G}_n = \frac{n(n+1)}{2} \frac{1}{n} \sum_{k=1}^{n} G_k \tag{33a}$$

$$= \frac{(n+1)}{2} \sum_{k=1}^{n} G_k \,, \tag{33b}$$

where in the first line, the first fraction represents all the possible ways of picking two particles of the same species from state $n + 1$, and the second fraction and sum represent all the ways of adding a particle into the n state (while preserving symmetry). Gathering all the terms, the CDME then has the form

$$\frac{\partial f_n(\boldsymbol{q})}{\partial t} = \sum_{v=1}^{n} D_v f_n + \frac{(n+1)}{2} \sum_{k=1}^{n} \int_{\mathbb{X}^2} \lambda(q_k; \boldsymbol{z}) f_{n+1}(\boldsymbol{q}_{\backslash\{k\}}, \boldsymbol{z}) d\boldsymbol{z} - \sum_{1 \leq i < j \leq n} f_n(\boldsymbol{q}) \int_{\mathbb{X}} \lambda(y; \boldsymbol{q}_{i,j}) dy \,,$$

where the notation omitted the time dependence of f_n for simplicity. The extension to multiple species is reported in the following subsection. The CDME for general one-species and multiple-species reactions were formulated in detail in [12, 28].

Multiple species extension

To exemplify the CDME for a reaction involving multiple species, we follow [28] again and consider the reaction

$$A + B \rightarrow C \,, \tag{34}$$

with rate function $\lambda(y; x_A, x_B)$, where x_A and x_B are the locations of one pair of reactants and y is the location of the product. The stochastic dynamics of the system are described in terms of the distributions $f_{a,b,c}(\boldsymbol{q}^a, \boldsymbol{q}^b, \boldsymbol{q}^c)$, where a, b, c indicate the numbers of A, B, and C particles, respectively, and $\boldsymbol{q}^a$ indicates the positions of the A particles, $\boldsymbol{q}^b$ of the B particles, and $\boldsymbol{q}^c$ of the C particles. The normalization condition eq. (23) generalizes to multiple species

$$\sum_{a,b,c=0}^{\infty} \int_{\mathbb{X}^a \times \mathbb{X}^b \times \mathbb{X}^c} f_{a,b,c}(\boldsymbol{q}^a, \boldsymbol{q}^b, \boldsymbol{q}^c) d\boldsymbol{q}^a d\boldsymbol{q}^b d\boldsymbol{q}^c = 1 \,. \tag{35}$$

Applying analogous reasoning as before, we can derive the CDME for the bimolecular reaction [28]. Here we just state the final equation

$$\frac{\partial f_{a,b,c}}{\partial t} = \sum_{\mu=1}^{a} D_{\mu}^{A} f_{a,b,c} + \sum_{\nu=1}^{b} D_{\nu}^{B} f_{a,b,c} + \sum_{\xi=1}^{c} D_{\xi}^{C} f_{a,b,c}$$
$$+ \frac{(a+1)(b+1)}{c} \sum_{\xi=1}^{c} \int_{\mathbb{X}^2} \lambda\left(q_{\xi}^{c}; z, z'\right) f_{a+1,b+1,c-1}\left((\boldsymbol{q}^a, z), (\boldsymbol{q}^b, z'), \boldsymbol{q}_{\backslash\{\xi\}}^{c}\right) dz\, dz' \quad (36)$$
$$- f_{a,b,c} \sum_{\mu=1}^{a} \sum_{\nu=1}^{b} \int_{\mathbb{X}} \lambda\left(y; q_{\mu}^{a}, q_{\nu}^{b}\right) dy,$$

where to simplify notation, the dependence of $\rho_{a,b,c}$ on time t has been omitted, as well as its dependence on the positions $\left(x^{(a)}, x^{(b)}, x^{(c)}\right)$ if clear from context. The first line corresponds to the diffusion of each chemical species, the second line to the gain term and the last one to the loss. The works [12, 28] developed a comprehensive and formal mathematical framework to formulate the CDME for any given reaction system.

4.2 Langevin dynamics with varying number of particles

There is a close mathematical link between section 3 and section 4.1; the resulting equations from both approaches describe the dynamics of densities in the phases space of systems with varying number of particles. In section 4.1, we introduced the diffusion operator, as well as the densities, only dependent on the position of the particles. However, one can in principle define these operators on more general dynamics. For instance, the state of the system can also depend on the velocities of the particles, as well as on the interactions between particles. In section 3 we had already incorporated velocities, but in a classical context, where there was no stochastic component and the changes in particle number were due to being in contact with a reservoir. A natural question arises: what is the physical/mathematical connection between these approaches?

In this section, we focus on Langevin dynamics with varying copy numbers as a middle ground to hint at the similarities between the classical open systems approach from section 3 and the stochastic reaction-diffusion approach from section 4.1. This will hint at a general approach to model the dynamics of systems with varying number of particles, enabling richer models of open systems.

To start, consider the Langevin dynamics of a system with n point particles in a spatial domain $\mathbb{X}$ with positions and velocities $\boldsymbol{q} \in \mathbb{X}^n$ and $\boldsymbol{v} \in \mathbb{V}^n$, where $\mathbb{V}$ is the space of one-particle velocities (in general $\mathbb{R}^3$). The particle's masses are m, and they are under an interaction potential $U(\boldsymbol{q})$ with a configuration-dependent friction tensor η to model velocity-dependent hydrodynamic interactions (assuming approximation of pair-wise additivity [53]), then Langevin dynamics are given by

$$d\boldsymbol{q}(t) = \boldsymbol{v}(t)dt \qquad m \cdot d\boldsymbol{v}(t) = -\eta \cdot \boldsymbol{v}(t)dt - \nabla_q U(\boldsymbol{q})dt + \sqrt{2k_B T}\,\eta^{1/2} d\boldsymbol{w}(t), \quad (37)$$

where $\boldsymbol{w}(t)$ corresponds to a $3n$ dimensional vector of independent Wiener processes or standard Brownian motion; k_B is the Boltzmann constant, and T is the temperature. The corresponding Fokker-Planck equation for these dynamics —also named Klein-Kramers equation [54]— determines the dynamics of the probability density $f_n(t, \boldsymbol{q}, \boldsymbol{v})$ in phase space, and

it is given by [53]

$$\frac{\partial f_n}{\partial t} = \mathcal{K}_n f_n, \qquad \text{with} \tag{38}$$

$$\mathcal{K}_n = \sum_{i=1}^{n} \left(-v_i \cdot \nabla_{q_i} f_n + \frac{1}{m} \nabla_{q_i} U \cdot \nabla_{v_i} f_n \right) + \sum_{i,j=1}^{n} \frac{1}{m} \nabla_{v_i} \cdot \eta_{ij} \left(v_j f_n + \frac{k_B T}{m} \nabla_{v_j} f_n \right), \tag{39}$$

where $\mathcal{K}_n$ simply denotes the infinitesimal generator of this Fokker-Planck equation for an n particle system with positions and velocities—in three dimensions, the equation would be a Fokker-Planck equation in $6n$ dimensions. We would like to now incorporate processes that change the number of particles. This can be framed quite generally following section 4.1 in terms of reactions. Reaction events can be thought of more generally as instantaneous events that follow a Poisson process in terms of rate functions λ. Note these rates must not be constant; they can depend on time, on the state of the system or on external variables. The master equation for this process will have the same form as the CDME in eq. (24), but now instead of the diffusion operator, we have the Klein-Kramers one

$$\frac{\partial f_n}{\partial t} = \mathcal{K}_n f_n + \sum_{r=1}^{M} \mathcal{R}_n^{(r)} f. \tag{40}$$

Here M denotes the number of processes changing the number of particles, and we assume that $f = (f_0, f_1, \ldots, f_n, \ldots)$ now also depends on the velocities. Note we again assume the densities are symmetric with respect to particle indexing permutations and that the normalization from eq. (23) requires additional integration over the velocity variables

$$\sum_{n=0}^{\infty} \int_{\mathbb{X}^n \times \mathbb{V}^n} f_n(\boldsymbol{q}, \boldsymbol{v}) \, d\boldsymbol{q} \, d\boldsymbol{v} = 1. \tag{41}$$

In addition, the rate functions within the $\mathcal{R}$ operators could also depend explicitly on the velocities of particles, allowing for richer particle-base models and generalizing the depth of field of the CDME. The approach presented here based on Langevin dynamics is an ideal starting ground to couple open particle-based models with fluids and to incorporate hydrodynamic and electrostatic interactions. As a final remark, note that if we take the overdamped limit of the Langevin dynamics, we in principle recover the CDME from section 4.1.

5 Previous models as special cases

In this section, we first show what are the mathematical connections between the approaches from sections 3.1 and 3.2 and the other two approaches in sections 4.1 and 4.2. Based on the insight gained in section 4.2, we compare to the previous results and show how this inspired the general master equation for classical systems with varying number of particles from section 2.

In both previous sections 3 and 4, we introduced equations to describe the dynamics of densities in the phase space of systems with varying number of particles. The first insight is that mathematically the n-particle Liouville equation is a special case of the n particle Fokker-Planck equation (Klein-Kramers eq.). To show this, we start with the classical approach from section 3.1. The resulting equation for the distribution dynamics in the phase space of the open system is given by a family of Liouville n-particle equations coupled by the terms $\Psi_n + \Phi_n^{n+1}$ (eq. (16))

$$\frac{\partial f_n}{\partial t} + \sum_{i=1}^{n} \left(\nabla_{\vec{q}_i} \cdot (\vec{v}_i f_n) + \nabla_{\vec{p}_i} \cdot (\vec{F}_i f_n) \right) = \Psi_n + \Phi_n^{n+1}. \tag{42}$$

The Bergman-Lebowitz model from section 3.2 shares the same mathematical structure but with different coupling terms due to different conceptual starting points of view: section 3.1 employs an explicit particle reservoir whose coupling with the open system has been achieved by explicit integration of its degrees of freedom, while section 3.2 employs a reservoir modeled a priori with a stochastic term for the exchange with the open system. In practice, the Bergman-Lebowitz model conceptually lies in between the other two approaches shown here. We can compare this equation with the nth component of the master equation eq. (40) from section 4.2

$$\frac{\partial f_n}{\partial t} - \mathcal{K}_n f_n = \sum_{r=1}^{M} \mathcal{R}_n^{(r)} f \,, \tag{43}$$

with

$$\mathcal{K}_n f_n = \sum_{i=1}^{n} \left(-v_i \cdot \nabla_{q_i} f_n + \frac{1}{m} \nabla_{q_i} U \cdot \nabla_{v_i} f_n \right) + \sum_{i,j=1}^{n} \frac{1}{m} \nabla_{v_i} \cdot \eta_{ij} \left(v_j f_n + \frac{k_B T}{m} \nabla_{v_j} f_n \right) . \tag{44}$$

We can rewrite the first term of the diffusion operator in exactly the same form as eq. (42)

$$\mathcal{K}_n f_n = - \sum_{i=1}^{n} \left(\nabla_{q_i} (v_i \cdot f_n) + \frac{1}{m} \nabla_{v_i} \cdot (F_i f_n) \right) + \sum_{i,j=1}^{n} \frac{1}{m} \nabla_{v_i} \cdot \eta_{ij} \left(v_j f_n + \frac{k_B T}{m} \nabla_{v_j} f_n \right), \tag{45}$$

where F_i is the net force acting on particle i due to potential-based interactions and $\nabla_{p_i} = \nabla v_i / m$. If the noise term of the Langevin equation is removed, the second term of $\mathcal{K}_n f_n$ vanishes recovering exactly the Liouville equation. Thus, the Liouville equation can be mathematically understood as a special case of the Fokker-Planck equation for Langevin dynamics in the deterministic limit. Note the emphasis on "mathematically" since from a physical perspective Langevin dynamics are often understood as a coarse-grained representation of classical molecular dynamics.

The second insight is that these equations have an analogous mathematical structure. One part is essentially a transport term for the n-particle density given by a Fokker-Planck equation, which simplifies to a Liouville equation in deterministic cases. The other part is a coupling term that models the change in the number of particles across the family of densities $f = (f_0, f_1, \dots, f_n, \dots)$. We can write both equations in the form of eq. (5):

$$\underbrace{\left(\frac{\partial}{\partial t} - \mathcal{A}_n \right) f_n}_{\substack{n\text{-particle} \\ \text{transport term}}} = \underbrace{\mathcal{Q}_n f}_{\substack{\text{coupling} \\ \text{term}}} . \tag{46}$$

We can recover all models investigated in the previous sections as follows:

- Section 3.1, Liouville-like equation for an open subsystem:

$$\mathcal{A}_n f_n \to \Lambda_n f_n \,, \qquad \text{and} \qquad \mathcal{Q}_n f \to \Psi_n + \Phi_n^{n+1} \,. \tag{47}$$

  In this case, the dynamics of the system are deterministic with Hamiltonian structure, so we only need the Liouville operator for the transport part. All the physics for the coupling between the subsystem with n particles and the larger system (universe) with N particles are condensed in the coupling term. As an additional verification of physical consistency, it was further shown that in equilibrium and under the hypothesis of short-range interactions, this approach recovers automatically the standard stationary grand canonical distribution [26].

- Section 3.2, Bergman-Lebowitz equation of open systems:

$$\mathcal{A}_n f_n \to \Lambda_n f_n, \qquad \text{and}$$

$$\mathcal{Q}_n f \to \sum_{m=0}^{\infty} \int dY^m [K_{nm}(X^n, Y^m) f_m(Y^m, t) - K_{mn}(Y^m, X^n) f_n(X^n, t)]. \tag{48}$$

Once again the dynamics of the transport process are deterministic with Hamiltonian structure, so we can use again the Liouville operator. However, the interaction with the reservoir is modeled stochastically using the Markovian transition kernel K_{mn}. This can be understood under the same light as the reaction in section 4.1. We can even separate the terms into a general loss and a gain term $\mathcal{Q}_n = \mathcal{G}_n - \mathcal{L}_n$ where

$$\mathcal{L}_n f = \sum_{m=0}^{\infty} \int dY^m [K_{mn}(Y^m, X^n) f_n(X^n, t)], \tag{49}$$

$$\mathcal{G}_n f = \sum_{m=0}^{\infty} \int dY^m [K_{nm}(X^n, Y^m) f_m(Y^m, t)]. \tag{50}$$

In this sense, this approach is a middle ground between the approach in section 3.1 and the one in section 4.1.

- Section 4.1, Chemical diffusion master equation for reaction diffusion:

$$\mathcal{A}_n f_n \to \mathcal{D}_n f_n, \qquad \text{and} \qquad \mathcal{Q}_n f \to \sum_{r=1}^{M} \mathcal{R}_n^{(r)} f. \tag{51}$$

In this case, the transport dynamics are stochastic, but they correspond to Brownian dynamics (overdamped Langevin dynamics to be precise), so there is no velocity dependence. Thus the transport part is governed by the Fokker-Planck generator for the standard Brownian diffusion of n-particles: $\mathcal{D}_n$. The reaction operators $\mathcal{R}_n^{(r)}$ transport probability between configurations with different number of particles. These can be separated into the gain and loss parts: $\mathcal{R}_n^{(r)} = \mathcal{G}_n^{(r)} - \mathcal{L}_n^{(r)}$. The total gain and loss due to all the reactions is

$$\mathcal{G}_n = \sum_{r=1}^{M} \mathcal{G}_n^{(r)}, \qquad \mathcal{L}_n = \sum_{r=1}^{M} \mathcal{L}_n^{(r)}. \tag{52}$$

Following [28], we can use the local rate functions for each reaction with these relations and write a global Markovian transition kernel in the form of eqs. (49) and (50), which establishes a direct connection with the Bergman-Lebowitz approach. The main difference is that, in the Bergman-Lebowitz approach, this kernel can depend on velocities while here it would only depend on positions.

- Section 4.2, Langevin dynamics with varying particle number:

$$\mathcal{A}_n f_n \to \mathcal{K}_n f_n, \qquad \text{and} \qquad \mathcal{Q}_n f \to \sum_{r=1}^{M} \mathcal{R}_n^{(r)} f. \tag{53}$$

From a mathematical point of view, this case covers the most general transport dynamics from all the examples presented; all the other cases are special cases of this one. Here, the transport dynamics are stochastic and track positions and velocities. Thus, the transport operator is given by the generator of the Klein-Kramers equation; the Fokker-Planck

equation for Langevin dynamics. In the overdamped limit, we recover the diffusion operator $\mathcal{D}_n$ from section 4.1. Alternatively, if we remove the noise term, we recover the Liouville operator $\mathcal{H}_n$ from sections 3.1 and 3.2. The coupling term works exactly as before, with the slight difference that now the reaction rate functions can also depend on the velocity. Following the same reasoning as before, we could once again write this in the form of eqs. (49) and (50), but unlike the overdamped Langevin case, we would also have a dependence on velocities.

All the cases are special cases of eq. (46). However, its mathematical structure is even more general and does not need to be constrained to Langevin dynamics, Hamiltonian structure or a specific form of the coupling terms. The Liouville/Fokker-Planck term can be written for very general deterministic or stochastic dynamical systems, and the coupling terms can model reactions, interactions with external reservoirs or any other process that changes the number of particles. The only constraint is that every term in the equation must conserve the total probability. We know the Fokker-Planck equation part is probability preserving, so we only need to make sure that the couplings $\mathcal{Q}_n f$ also preserve the total probability (eqs. (15), (23) and (41)) following eq. (6). This was shown in [24,26] for classical molecular dynamics and in [12] for reaction-diffusion dynamics. For other specific applications, one must construct the corresponding transport and coupling operators.

6 Perspectives: Physical, numerical and beyond

The general formulation of the master equation for systems with varying particle number presented in this work provides a novel and synergistic link between different fields and how they handle open settings. Through its general formulation, it unveils new perspectives and opens up the application scope to a diverse range of fields within and beyond physics. There is ample potential for future applications, specifically in the design of multiscale theory and simulation of complex systems. In this section, we discuss new perspectives inspired by this work, as well as current and future applications.

From a physics perspective, understanding the dynamics of systems with varying number of particles requires imposing fundamental physical constraints to eq. (46). In most cases, these systems will have —partly— a Hamiltonian structure, which is represented by the Liouville part of the equation (section 3). Thus, in the spirit of the Bergman-Lebowitz approach, it is illustrative to present an alternative separation of eq. (46) into the following components:

$$\underbrace{\frac{\partial f_n}{\partial t} - \Lambda_n f_n}_{\substack{n\text{-particle}\\ \text{Liouville}}} = \underbrace{\mathcal{T}_n f_n}_{\substack{\text{heat}\\ \text{exchange}}} + \underbrace{\mathcal{Q}_n f}_{\substack{\text{material}\\ \text{exchange}}} \,, \tag{54}$$

where the infinitesimal generator in eq. (46) for the corresponding Fokker-Planck equation is $\mathcal{A}_n = \Lambda_n + \mathcal{T}_n$. The terms represent the Liouvillian dynamics (Hamiltonian); the contribution due to stochastic fluctuations, often a thermostat modeling heat exchange with a reservoir(s); and the coupling terms capturing the material/particle exchange with a reservoir(s). If we assume Langevin dynamics, the thermostat will have the form of section 4.2 and $\mathcal{A}_n \to \mathcal{K}_n$; however, eq. (54) is not limited to Langevin thermostats. The material exchange can model reactions, interactions with a reservoir or any other process that changes the number of particles. Analyzing and manipulating these components could offer insights into the intricate interplay between system dynamics and its surroundings, shedding light on emergent behaviors, equilibrium states and non-equilibrium phenomena. As an example in a simpler context, it has been shown that, in well-mixed biochemical systems, reservoir interaction drives phenomena fundamental for life processes such as symmetry breaking, entropy production and

phase transitions [3, 4, 31]. The form of eq. (54) hints at the possibility of performing a two-level coarse-graining of a Hamiltonian system. First, coarse-graining the fast-scales, e.g. the solvent dynamics into a thermostat. Then, a second coarse-graining following section 3.1 to model the material exchange of the subsystem with the environment. Although these two effects produced by different types of reservoir interaction were also captured in a seminal work [24] with equations in the form of eq. (20), the approach discussed in this work is more general and conceptually different.

In the context of coarse-graining techniques, the Mori-Zwanzig formalism stands out as a powerful tool for capturing the relevant/slow dynamics of complex systems while accounting for memory effects and non-Markovian behavior. Although it is not trivial how to apply it to systems with varying particle number, it sheds light on what kind of dynamics we should expect in the coarse-grained variables. When applying the Mori-Zwanzig formalism to a Hamiltonian system, it not only yields a noise term representing the thermostat but also a term modeling non-Markovian memory effects. What are the memory effects emerging from the coarse-graining due to particle exchange with the reservoir? Can these memory effects be incorporated into the master equation eq. (54)? Can we incorporate memory effects into reaction-diffusion processes and what is their relevance? These are open questions motivated by this work that are tremendously relevant to the physics community.

The quantum mechanical perspective could also be framed in similar terms, particularly through analysis of frameworks like the Lindblad equation [55], a generalization of the von Neumann's equation that describes the time evolution of density matrices/operators for open quantum systems subject to dissipative processes. Von Neumann's equation is the quantum analog to Liouville's equation, where the classical Poisson bracket is substituted by the commutator and the density by the density matrix/operator. The Lindblad equation corresponds to von Neumann's equation with an additional term to model the interactions with the environment. This term is constructed using the so-called jump operators that can be understood as modeling probability "jumping" from one state to another due to interactions with the environment, analogously to the $\mathcal{Q}_n f$ term in eq. (54). It is natural to suggest that one may be able to construct a framework, following the ideas of this manuscript, similar but alternative to the Lindblad master equation, to describe the exchange of particles between arbitrary quantum systems and the environment. Recently an approach similar to that of [26] and applied to the von Neumann equation has been proposed by one of the authors [56], while the other author reviewed field theory inspired representations of reaction-diffusion systems [13], which perhaps aids in extending these ideas to a quantum setting. Nevertheless, a more precise connection to quantum systems and situations out of equilibrium requires a more detailed analysis beyond the current scope and is left for future work.

From a numerical perspective, multiscale simulations play a pivotal role in understanding complex open systems. By having a master equation for the underlying continuous particle-based process, like the one presented in this work, one has a ground model from which coarser models emerge from the bottom up through discretizations, mean-field limits or moment expansions. In particular, systems in contact with material reservoirs are essential for real-world applications, such as molecular dynamics, biochemical kinetics and weather modeling. Reservoirs in these systems can exhibit dynamic behavior and undergo changes in properties over time. Modeling such reservoirs accurately requires sophisticated simulation techniques that account for multiscale phenomena and ensure physical consistency across scales. The formal derivation of the master equation and the coupling term with accurate statistical properties provides a guiding protocol for designing physically well-founded coupling schemes. Some examples of numerical schemes that handle the coupling with dynamic reservoirs accurately are in models related to this work [16, 29, 32, 33, 51], including thermodynamic aspects [33, 57]. Moreover, these master equations can be used to derive meso/macroscopic models by calcu-

lating expectations, higher moments [32,58] or even by applying probabilistic limits, such as the law of large numbers, the central limit theorem or large deviation principles [33,59,60]. Through these methodologies, a physical and mathematical consistency is established across scales, often in the form of relations between the parameters of the models used at different scales. Then these relations can be used to develop numerical schemes that are physically consistent across multiple scales [39,61–63].

In the context of data-driven modeling and simulation, a large range of methods have been recently constructed based on the Koopman operator [64–68]. One can intuitively construct the Koopman operator of system with fixed number of particles n as follows: (i) the infinitesimal generator of the process is given by $\mathcal{A}_n$ from eq. (46), this could be the Liouville, the Fokker-Planck generator or something more complex ($\mathcal{Q}_n$ is not relevant for now as the particle number is fixed); (ii) the solution of the equation is given in terms of the exponential of the infinitesimal generator $\mathcal{A}_n$, which for a chosen time-step defines a propagator. This propagator can be understood as the Ruelle or Perron-Frobenius operator, which propagates probability densities forward in time; (iii) the adjoint of this operator is the Koopman operator and instead of propagating probability densities, it propagates observables. This last property is why it is suitable for data-driven methods. To the best of our knowledge, we are not aware of the construction or application of the Koopman operator for particle-based systems with varying copy numbers. This work delivers the infinitesimal generator for general dynamical systems with varying particle number, $\mathcal{A}_n + \mathcal{Q}_n$ from eq. (46). This is the first cornerstone to derive the corresponding Koopman operator. Obtaining such an operator would enable applications in a diverse range of fields.

Many of these methods and techniques based on the general master equation can be applied to fields beyond physics. As the equation describes a general random dynamical system, it can be applied to complex systems with varying particle number beyond reaction-diffusion and molecular systems. This will potentially have significant applications in the development of numerical schemes and multiscale modeling of the spread of diseases [69–71], innovations [72], opinion dynamics [34,73], and power, transportation and communication systems [74,75] among others.

To finalize, we list some relevant examples of how the master equations presented in this work aid in the numerical simulation of complex molecular systems. In particular in the treatment of multiscale molecular simulations that go beyond equilibrium, for example, systems in a temperature gradient [42]. Such a situation can be realized by embedding the open system in two distinct reservoirs, each at a different thermodynamic condition. In such a case the prescription for the simulator is reduced to the boundary conditions of the equation of section 3 at the interface between the system and each reservoir. In general, the embedding of the open system in multiple reservoirs at different thermodynamic conditions allows the design of numerical algorithms, where the reservoir can be modeled as a thermodynamically fluctuating (in time) region via the fluctuating hydrodynamics method [30]. The instantaneous thermodynamic condition of the reservoir corresponds to boundary conditions for the open systems which then it is simulated and whose averaged molecular properties are given, in the next step, as input to the reservoir; the procedure is then repeated thus producing the dynamics of the open system. Without the information on the boundary conditions of the equations of section 3, such a numerical scheme could not be implemented with such a physical consistency.

In the realm of reaction-diffusion processes, the chemical diffusion master equation can be expanded in terms of classical creation and annihilation operators acting on the basis of the underlying space (a Fock space) [12,28]. Based on this formulation, it is straightforward to develop Galerikin discretizations of the master equation [12]. This immediately yields the so-called reaction-diffusion master equations [49,76], where the space is discretized in voxels, and the diffusion is modeled as jumps between neighboring voxels. However, unlike previous

standard derivations [76], the scaling of the rates for nonlinear reactions (involving two or more particles) is automatically adjusted to the size of the lattice grid chosen, which enforces a consistent convergence to the micro and macroscopic scale [12, 32, 77]. It further provides a relation between the microscopic parameters and the partial differential equation model at the macroscopic scale [32], enabling particle-based simulations that are consistent with the macroscopic model and thus facilitating the implementation of reservoirs as macroscopic models that are coupled to a particle-based model [32].

The methodologies emerging from the theoretical constructs presented in this work do not only have the potential to provide physical insight, numerical schemes and solutions to the master equations of a large range of complex systems with varying particle number; they also have unifying capabilities and potential to bring insights into the underlying physics through the analysis of convergence properties, stability, and approximation errors. Exploring the application of these methods in diverse complex systems will allow researchers from several fields to uncover novel phenomena, unify models across scales and design efficient multiscale computational strategies for studying systems with varying number of particles.

Acknowledgments

This work was partly supported by the Deutsche Forschungsgemeinschaft (DFG) Collaborative Research Center 1114 "Scaling Cascades in Complex Systems", project No.235221301, Project C01 (L.D.S.) "Adaptive coupling of scales in molecular dynamics and beyond to fluid dynamics". M.J.R. was supported by DFG grant no. RA 3601/1-1.

References

[1] A. Leonidovich Kuzemsky, *Statistical mechanics and the physics of many-particle model systems*, World Scientific, Singapore, ISBN 9789813145627 (2016), doi:10.1142/10169.

[2] H. Qian, *Phosphorylation energy hypothesis: Open chemical systems and their biological functions*, Annu. Rev. Phys. Chem. **58**, 113 (2007), doi:10.1146/annurev.physchem.58.032806.104550.

[3] H. Qian, S. Kjelstrup, A. B. Kolomeisky and D. Bedeaux, *Entropy production in mesoscopic stochastic thermodynamics: Nonequilibrium kinetic cycles driven by chemical potentials, temperatures, and mechanical forces*, J. Phys.: Condens. Matter **28**, 153004 (2016), doi:10.1088/0953-8984/28/15/153004.

[4] H. Qian, P. Ao, Y. Tu and J. Wang, *A framework towards understanding mesoscopic phenomena: Emergent unpredictability, symmetry breaking and dynamics across scales*, Chem. Phys. Lett. **665**, 153 (2016), doi:10.1016/j.cplett.2016.10.059.

[5] R. R. Netz, *Approach to equilibrium and nonequilibrium stationary distributions of interacting many-particle systems that are coupled to different heat baths*, Phys. Rev. E **101**, 022120 (2020), doi:10.1103/PhysRevE.101.022120.

[6] N. Wunderling et al., *Climate tipping point interactions and cascades: A review*, Earth Syst. Dynam. **15**, 41 (2024), doi:10.5194/esd-15-41-2024.

[7] U. Achatz, Y.-H. Kim and G. S. Voelker, *Multi-scale dynamics of the interaction between waves and mean flows: From nonlinear WKB theory to gravity-wave parameterizations in weather and climate models*, J. Math. Phys. **64**, 111101 (2023), doi:10.1063/5.0165180.

[8] A. J. Majda, C. Franzke and D. Crommelin, *Normal forms for reduced stochastic climate models*, Proc. Natl. Acad. Sci. **106**, 3649 (2009), doi:10.1073/pnas.0900173106.

[9] J. Berner et al., *Stochastic parameterization: Toward a new view of weather and climate models*, Bull. Am. Meteorol. Soc. **98**, 565 (2017), doi:10.1175/BAMS-D-15-00268.1.

[10] M. Doi, *Second quantization representation for classical many-particle system*, J. Phys. A: Math. Gen. **9**, 1465 (1976), doi:10.1088/0305-4470/9/9/008.

[11] P. Grassberger and M. Scheunert, *Fock-space methods for identical classical objects*, Fortschr. Phys. **28**, 547 (1980), doi:10.1002/prop.19800281004.

[12] M. J. del Razo, D. Frömberg, A. V. Straube, C. Schütte, F. Höfling and S. Winkelmann, *A probabilistic framework for particle-based reaction-diffusion dynamics using classical Fock space representations*, Lett. Math. Phys. **112**, 49 (2022), doi:10.1007/s11005-022-01539-w.

[13] M. J. del Razo, T. Lamma and W. Merbis, *Field theories and quantum methods for stochastic reaction-diffusion systems*, (arXiv preprint) doi:10.48550/arXiv.2409.13377.

[14] D. Frenkel and B. Smit, *Understanding molecular simulation*, Academic Press, Cambridge, USA, ISBN 9780122673511 (2002), doi:10.1016/B978-0-12-267351-1.X5000-7.

[15] D. P. Landau and K. Binder, *A guide to Monte Carlo simulations in statistical physics*, Cambridge University Press, Cambridge, UK, ISBN 9781107074026 (2014), doi:10.1017/CBO9781139696463.

[16] M. Praprotnik, L. Delle Site and K. Kremer, *Multiscale simulation of soft matter: From scale bridging to adaptive resolution*, Annu. Rev. Phys. Chem. **59**, 545 (2008), doi:10.1146/annurev.physchem.59.032607.093707.

[17] R. Cortes-Huerto, M. Praprotnik, K. Kremer and L. Delle Site, *From adaptive resolution to molecular dynamics of open systems*, Eur. Phys. J. B **94**, 189 (2021), doi:10.1140/epjb/s10051-021-00193-w.

[18] I. Prigogine, C. George and F. Henin, *Dynamical and statistical descriptions of N-body systems*, Physica **45**, 418 (1969), doi:10.1016/0031-8914(69)90270-5.

[19] C. George, *Subdynamics and correlations*, Physica **65**, 277 (1973), doi:10.1016/0031-8914(73)90346-7.

[20] G. G. Emch and G. L. Sewell, *Nonequilibrium statistical mechanics of open systems*, J. Math. Phys. **9**, 946 (1968), doi:10.1063/1.1664664.

[21] B. Robertson, *Equations of motion in nonequilibrium statistical mechanics*, Phys. Rev. **144**, 151 (1966), doi:10.1103/PhysRev.144.151.

[22] J. Seke, *Equations of motion in nonequilibrium statistical mechanics of open systems*, Phys. Rev. A **21**, 2156 (1980), doi:10.1103/PhysRevA.21.2156.

[23] R. Zwanzig, *On the identity of three generalized master equations*, Physica **30**, 1109 (1964), doi:10.1016/0031-8914(64)90102-8.

[24] J. L. Lebowitz and P. G. Bergmann, *Irreversible Gibbsian ensembles*, Ann. Phys. **1**, 1 (1957), doi:10.1016/0003-4916(57)90002-7.

[25] J. L. Lebowitz and A. Shimony, *Statistical mechanics of open systems*, Phys. Rev. **128**, 1945 (1962), doi:10.1103/PhysRev.128.1945.

[26] L. Delle Site and R. Klein, *Liouville-type equations for the n-particle distribution functions of an open system*, J. Math. Phys. **61**, 083102 (2020), doi:10.1063/5.0008262.

[27] R. Klein and L. Delle Site, *Derivation of Liouville-like equations for the n-state probability density of an open system with thermalized particle reservoirs and its link to molecular simulation*, J. Phys. A: Math. Theor. **55**, 155002 (2022), doi:10.1088/1751-8121/ac578f.

[28] M. J. del Razo, S. Winkelmann, R. Klein and F. Höfling, *Chemical diffusion master equation: Formulations of reaction-diffusion processes on the molecular level*, J. Math. Phys. **64**, 013304 (2023), doi:10.1063/5.0129620.

[29] L. Delle Site, C. Krekeler, J. Whittaker, A. Agarwal, R. Klein and F. Höfling, *Molecular dynamics of open systems: Construction of a mean-field particle reservoir*, Adv. Theory Simul. **2**, 1900014 (2019), doi:10.1002/adts.201900014.

[30] A. Gholami, R. Klein and L. Delle Site, *Simulation of a particle domain in a continuum, fluctuating hydrodynamics reservoir*, Phys. Rev. Lett. **129**, 230603 (2022), doi:10.1103/PhysRevLett.129.230603.

[31] H. Qian and L. M. Bishop, *The chemical master equation approach to nonequilibrium steady-state of open biochemical systems: Linear single-molecule enzyme kinetics and nonlinear biochemical reaction networks*, Int. J. Mol. Sci. **11**, 3472 (2010), doi:10.3390/ijms11093472.

[32] M. Kostré, C. Schütte, F. Noé and M. J. del Razo, *Coupling particle-based reaction-diffusion simulations with reservoirs mediated by reaction-diffusion PDEs*, Multiscale Model. Simul. **19**, 1659 (2021), doi:10.1137/20M1352739.

[33] M. J. del Razo, H. Qian and F. Noé, *Grand canonical diffusion-influenced reactions: A stochastic theory with applications to multiscale reaction-diffusion simulations*, J. Chem. Phys. **149**, 044102 (2018), doi:10.1063/1.5037060.

[34] L. Helfmann, N. Djurdjevac Conrad, P. Lorenz-Spreen and C. Schütte, *Modelling opinion dynamics under the impact of influencer and media strategies*, Sci. Rep. **13**, 19375 (2023), doi:10.1038/s41598-023-46187-9.

[35] M. Kot, *Elements of mathematical ecology*, Cambridge University Press, Cambridge, UK, ISBN 9780521001502 (2001), doi:10.1017/CBO9780511608520.

[36] X. Mao and C. Yuan, *Stochastic differential equations with Markovian switching*, Imperial College Press, London, UK, ISBN 9781860947018 (2006), doi:10.1142/p473.

[37] G. G. Yin and C. Zhu, *Hybrid switching diffusions*, Springer, New York, USA, ISBN 9781441911063 (2010), doi:10.1007/978-1-4419-1105-6.

[38] P. C. Bressloff, *Stochastic switching in biology: From genotype to phenotype*, J. Phys. A: Math. Theor. **50**, 133001 (2017), doi:10.1088/1751-8121/aa5db4.

[39] M. J. del Razo, M. Dibak, C. Schütte and F. Noé, *Multiscale molecular kinetics by coupling Markov state models and reaction-diffusion dynamics*, J. Chem. Phys. **155**, 124109 (2021), doi:10.1063/5.0060314.

[40] S. Fritsch, S. Poblete, C. Junghans, G. Ciccotti, L. Delle Site and K. Kremer, *Adaptive resolution molecular dynamics simulation through coupling to an internal particle reservoir*, Phys. Rev. Lett. **108**, 170602 (2012), doi:10.1103/PhysRevLett.108.170602.

[41] H. Wang, C. Hartmann, C. Schütte and L. Delle Site, *Grand-canonical-like molecular-dynamics simulations by using an adaptive-resolution technique*, Phys. Rev. X **3**, 011018 (2013), doi:10.1103/PhysRevX.3.011018.

[42] R. Ebrahimi Viand, F. Höfling, R. Klein and L. Delle Site, *Theory and simulation of open systems out of equilibrium*, J. Chem. Phys. **153**, 101102 (2020), doi:10.1063/5.0014065.

[43] R. Klein, R. Ebrahimi Viand, F. Höfling and L. Delle Site, *Nonequilibrium induced by reservoirs: Physico-mathematical models and numerical tests*, Adv. Theory Simul. **4**, 210071 (2021), doi:10.1002/adts.202100071.

[44] A. Agarwal, J. Zhu, C. Hartmann, H. Wang and L. Delle Site, *Molecular dynamics in a grand ensemble: Bergmann-Lebowitz model and adaptive resolution simulation*, New J. Phys. **17**, 083042 (2015), doi:10.1088/1367-2630/17/8/083042.

[45] C. W. Gardiner and S. Chaturvedi, *The poisson representation. I. A new technique for chemical master equations*, J. Stat. Phys. **17**, 429 (1977), doi:10.1007/BF01014349.

[46] D. T. Gillespie, *A rigorous derivation of the chemical master equation*, Phys. A: Stat. Mech. Appl. **188**, 404 (1992), doi:10.1016/0378-4371(92)90283-V.

[47] D. Schnoerr, G. Sanguinetti and R. Grima, *Approximation and inference methods for stochastic biochemical kinetics – A tutorial review*, J. Phys. A: Math. Theor. **50**, 093001 (2017), doi:10.1088/1751-8121/aa54d9.

[48] S. A. Isaacson and D. Isaacson, *Reaction-diffusion master equation, diffusion-limited reactions, and singular potentials*, Phys. Rev. E **80**, 066106 (2009), doi:10.1103/PhysRevE.80.066106.

[49] S. Winkelmann and C. Schütte, *The spatiotemporal master equation: Approximation of reaction-diffusion dynamics via Markov state modeling*, J. Chem. Phys. **145**, 214107 (2016), doi:10.1063/1.4971163.

[50] L.-P. Arguin, A. Bovier and N. Kistler, *The extremal process of branching Brownian motion*, Probab. Theory Relat. Fields **157**, 535 (2012), doi:10.1007/s00440-012-0464-x.

[51] M. J. del Razo, *Open reaction-diffusion systems: Bridging probabilistic theory across scales*, (arXiv preprint) doi:10.48550/arXiv.2404.07119.

[52] M. Praprotnik, L. Delle Site and K. Kremer, *Adaptive resolution molecular-dynamics simulation: Changing the degrees of freedom on the fly*, J. Chem. Phys. **123**, 224106 (2005), doi:10.1063/1.2132286.

[53] D. L. Ermak and J. A. McCammon, *Brownian dynamics with hydrodynamic interactions*, J. Chem. Phys. **69**, 1352 (1978), doi:10.1063/1.436761.

[54] N. G. Van Kampen, *Stochastic processes in physics and chemistry*, North Holland, Amsterdam, Netherlands (2007), doi:10.1016/B978-0-444-52965-7.X5000-4.

[55] D. Manzano, *A short introduction to the Lindblad master equation*, AIP Adv. **10**, 025106 (2020), doi:10.1063/1.5115323.

[56] L. Delle Site and A. Djurdjevac, *An effective Hamiltonian for the simulation of open quantum molecular systems*, J. Phys. A: Math. Theor. **57**, 255002 (2024), doi:10.1088/1751-8121/ad5088.

[57] A. Gholami, F. Höfling, R. Klein and L. Delle Site, *Thermodynamic relations at the coupling boundary in adaptive resolution simulations for open systems*, Adv. Theory Simul. **4**, 2000303 (2021), doi:10.1002/adts.202000303.

[58] S. A. Isaacson, J. Ma and K. Spiliopoulos, *Mean field limits of particle-based stochastic reaction-diffusion models*, SIAM J. Math. Anal. **54**, 453 (2022), doi:10.1137/20M1365600.

[59] D. F. Anderson and T. G. Kurtz, *Stochastic analysis of biochemical systems*, Springer, Cham, Swtizerland, ISBN 9783319168944 (2015), doi:10.1007/978-3-319-16895-1.

[60] S. A. Isaacson, J. Ma and K. Spiliopoulos, *How reaction-diffusion PDEs approximate the large-population limit of stochastic particle models*, SIAM J. Appl. Math. **81**, 2622 (2021), doi:10.1137/20M1365429.

[61] M. J. del Razo and H. Qian, *A discrete stochastic formulation for reversible bimolecular reactions via diffusion encounter*, Commun. Math. Sci. **14**, 1741 (2016), doi:10.4310/CMS.2016.v14.n6.a13.

[62] M. Dibak, M. J. del Razo, D. De Sancho, C. Schütte and F. Noé, *MSM/RD: Coupling Markov state models of molecular kinetics with reaction-diffusion simulations*, J. Chem. Phys. **148**, 214107 (2018), doi:10.1063/1.5020294.

[63] T. Hempel, M. J. del Razo, C. T. Lee, B. C. Taylor, R. E. Amaro and F. Noé, *Independent Markov decomposition: Toward modeling kinetics of biomolecular complexes*, Proc. Natl. Acad. Sci. **118**, e2105230118 (2021), doi:10.1073/pnas.2105230118.

[64] S. L. Brunton, M. Budišić, E. Kaiser and J. N. Kutz, *Modern Koopman theory for dynamical systems*, SIAM Rev. **64**, 229 (2022), doi:10.1137/21M1401243.

[65] C. Schütte and S. Klus, *Towards tensor-based methods for the numerical approximation of the Perron-Frobenius and Koopman operator*, J. Comput. Dynam. **3**, 139 (2016), doi:10.3934/jcd.2016007.

[66] S. Klus, F. Nüske, P. Koltai, H. Wu, I. Kevrekidis, C. Schütte and F. Noé, *Data-driven model reduction and transfer operator approximation*, J. Nonlinear Sci. **28**, 985 (2018), doi:10.1007/s00332-017-9437-7.

[67] M. Korda and I. Mezić, *Linear predictors for nonlinear dynamical systems: Koopman operator meets model predictive control*, Automatica **93**, 149 (2018), doi:10.1016/j.automatica.2018.03.046.

[68] M. O. Williams, I. G. Kevrekidis and C. W. Rowley, *A data-driven approximation of the Koopman operator: Extending dynamic mode decomposition*, J. Nonlinear Sci. **25**, 1307 (2015), doi:10.1007/s00332-015-9258-5.

[69] T. Britton and E. Pardoux (eds.), *Stochastic epidemic models with inference*, Springer, Cham, Switzerland, ISBN 9783030308995 (2019), doi:10.1007/978-3-030-30900-8.

[70] N. Malysheva, J. Wang and M. von Kleist, *Stochastic simulation algorithm for effective spreading dynamics on time-evolving adaptive networX (SSATAN-X)*, Math. Model. Nat. Phenom. **17**, 35 (2022), doi:10.1051/mmnp/2022035.

[71] S. Winkelmann, J. Zonker, C. Schütte and N. Djurdjevac Conrad, *Mathematical modeling of spatio-temporal population dynamics and application to epidemic spreading*, Math. Biosci. **336**, 108619 (2021), doi:10.1016/j.mbs.2021.108619.

[72] N. Djurdjevac Conrad, L. Helfmann, J. Zonker, S. Winkelmann and C. Schütte, *Human mobility and innovation spreading in ancient times: A stochastic agent-based simulation approach*, Europhys. J. Data Sci. **7**, 24 (2018), doi:10.1140/epjds/s13688-018-0153-9.

[73] C. Castellano, S. Fortunato and V. Loreto, *Statistical physics of social dynamics*, Rev. Mod. Phys. **81**, 591 (2009), doi:10.1103/RevModPhys.81.591.

[74] M. Newman, *Networks*, Oxford University Press, Oxford, UK, ISBN9780198805090 (2018).

[75] D. Helbing, *Traffic and related self-driven many-particle systems*, Rev. Mod. Phys. **73**, 1067 (2001), doi:10.1103/RevModPhys.73.1067.

[76] S. Hellander, A. Hellander and L. Petzold, *Reaction-diffusion master equation in the microscopic limit*, Phys. Rev. E **85**, 042901 (2012), doi:10.1103/PhysRevE.85.042901.

[77] S. A. Isaacson, *A convergent reaction-diffusion master equation*, J. Chem. Phys. **139**, 054101 (2013), doi:10.1063/1.4816377.