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

Rupert Klein and Luigi Delle Site*

Freie Universität Berlin, Institute of Mathematics, Arnimallee 6, 14195 Berlin, Germany

E-mail: rupert.klein@math.fu-berlin.de and luigi.dellesite@fu-berlin.de

Abstract

A physico-mathematical model of open systems proposed in a previous paper (Delle Site and Klein 2020 *J. Math. Phys*. **61** 083102) can represent a guiding reference in designing an accurate simulation scheme for an open molecular system embedded in a reservoir of energy and particles. The derived equations and the corresponding boundary conditions are obtained without assuming the action of an external source of heat that assures thermodynamic consistency of the open system with respect to a state of reference. However, in numerical schemes the temperature in the reservoir must be controlled by an external heat bath otherwise thermodynamic consistency cannot be achieved. In this perspective, the question to address is whether the explicit addition of an external heat bath in the theoretical model modifies the equations of the open system and its boundary conditions. In this work we consider this aspect and explicitly describe the evolution of the reservoir employing the Bergmann–Lebowitz statistical model of thermostat. It is shown that the resulting equations for the open system itself are not affected by this change and an example of numerical application is reviewed where the current result shows its conceptual relevance. Finally, a list of pending mathematical and modelling problems is discussed the solution of which would strengthen the mathematical rigour of the model and offer new perspectives for the further development of a new multiscale simulation scheme.

*Author to whom any correspondence should be addressed.

(Some figures may appear in colour only in the online journal)

1. Introduction

Complex molecular systems of high relevance in science and technology are, in the majority, open systems embedded in a large thermodynamic environment which acts as a reservoir of energy and particles. In this work we define a system to be 'open' if it exchanges energy and matter with the exterior, whereas a closed system can exchange only energy but not matter. The necessity of extending the model of closed systems in physics, chemistry and biology, was suggested over the years by a strong activity of theoretical and mathematical physicists who developed several theoretical models of open subsystems embedded in a large thermodynamic reservoir of energy and particles [1–8]. Despite the availability of theoretical models, the corresponding implementation in numerical schemes for performing simulations of complex systems has not happened as expected. The mathematical abstraction and complexity of the models has represented a significant obstacle for their practical implementation in computer simulation schemes. However, applied physicists, and researchers in molecular simulation in particular, call for urgent progress in such a direction [9]. In fact, molecular dynamics (MD) [10, 11], considered nowadays a powerful tool for studying complex molecular systems [12], needs to overcome the restrictive principles of conservation of energy and matter to access a much larger spectrum of physical situations [13, 14]. Related progress of computational methods can only be driven by physico-mathematical models of open systems that are systematically based on first principles as much as possible but not restricted to ideal or abstract situations only. The success of such a project would assure that simulations of open systems are characterised by an increasing physical accuracy and reproducibility when treating realistic physical and chemical problems. In our previous work [15] a first attempt was made to employ a physico-mathematical model that rationalises the principles of an MD scheme whose main characteristic is the exchange of matter and energy between different subregions of the simulation box. Specifically, the adaptive resolution simulation (AdResS) scheme [16, 17] (see also section 7.1) was qualitatively mapped onto the Bergmann–Lebowitz (BL) model of open systems [1, 2] and this mapping helped identifying numerical procedures for the calculation of time correlation functions consistent with the physical principles of a system with variable number of particles [18]. The latest version of the AdResS scheme (see reference [19]) has in turn led to the development of a physico-mathematical model that derives boundary conditions at the interface between a molecular simulation region and its adjacent reservoirs from first principles [20], thereby clearly identifying the degrees of freedom available for the specification of the reservoirs' characteristics. Such systematically derived boundary conditions could, in perspective, guide the design of an evolved variant of the AdResS scheme that could properly account for situations of nonequilibrium [21] or reliably couple a region represented by molecular simulation with the fluctuating hydrodynamic regime of its large-scale environment [22] in a self-consistent multiscale simulation.

1.1. Aim of this work

In reference [20] (see also figure 1 for illustration) an open system Ω with n particles was defined as a subsystem of a large Universe, U, containing N particles. Next, from the
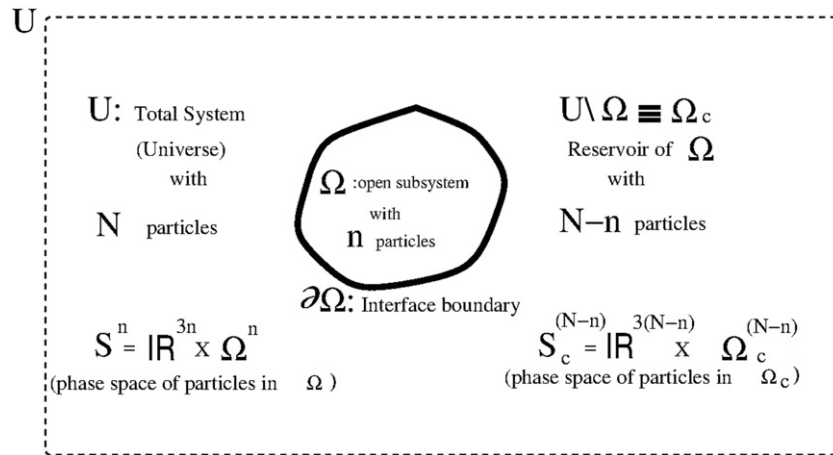
Figure 1. Pictorial illustration with corresponding formalism for the open system Ω, which is a subsystem with n particles of a large system U (Universe) containing N particles.

Liouville equation of the Universe the degrees of freedom in $U \backslash \Omega$ were integrated out, leading to a Liouville-type system of equations which defines a hierarchy of n-particle phase space distribution functions for the open system. However in numerical schemes an external thermostat is required to enforce thermodynamic consistency and thermal stability to the simulation, thus the current theoretical model would not match the conditions of a numerical scheme. For this reason, the aim of this work is to close the gap between the numerical implementation and the theoretical model by analyzing the effect of the introduction of an external thermostat acting in $U \backslash \Omega$ on the equations of f_n in Ω. The majority of heat baths/thermostats used in molecular simulation are, however, build upon empirical schemes aiming at numerical efficiency first. They all introduce an ad hoc alteration of the particle dynamics to impose a desired average temperature for particles in a given region (see e.g. reference [23]). The empirical character of numerical thermostats is such that it is difficult to cast them in the explicit mathematical form of a heat source acting on the system of interest which can be explicitly introduced in the Liouville equation. For this reason, for the mathematical derivation, we have chosen a popular model of heat bath suitable for the Liouville equation, that is the BL model [1, 2]. This model is general enough to express in a clear mathematical form the effective action of a numerical thermostat on a region of space and thus it represents an appropriate choice for the theoretical derivations. Thus, in this paper, we will adapt the BL thermostat to the hierarchical open systems model for the $U \backslash \Omega$ region and extend the idea to more reservoirs for treating situations with a thermal gradient. The paper is organised as follows: we first describe the essential characteristics of the physico-mathematical model of open system of references [20, 24]. Next the basic principles of the BL model for a heat source/bath are introduced and their inclusion into the equations of an open system of references [20, 24] are shown. The result of such an inclusion is the main formal result of this work: we show that the form of the equations is not altered by the addition of the thermal bath. We introduce the AdResS method and clarify the conceptual relevance of the current results for numerical applications by reviewing a simulation study where the thermostat plays a major role. Finally a list of mathematical open problems of the current model and of its potential evolution is discussed before a short summary with conclusions.

2. Liouville equations: from a large system of *N* particles to an open subsystem of *n* particles

This section recalls the basic information, required in the present context, of the model presented in reference [20]. Figure 1 pictorially represents the idea of partitioning a large system of N particles, named Universe, into a subsystem, Ω, that occupies a prescribed subdomain of U and is home to a time dependent number $n(t)$ of particles, and its (particle and energy) reservoir, $U \backslash \Omega = \Omega_c$ with $N - n(t)$ particles. The aim is to derive the Liouville equations for the probability in phase space of the particles in Ω by integrating out the degrees of freedom of particles in Ω_c from the Liouville equations of U. In this context, the Universe is characterised by its Hamiltonian:

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

where $\vec{p}_i$ is the momentum of the ith particle, M is the mass of an individual particle, while the interaction potential between particles i and j in positions $\vec{q}_i$ and $\vec{q}_j$ is given by: $V_{\text{tot}}(\mathbf{q}^N) = \sum_{i=1}^{N} \sum_{j=1, j \neq i}^{N} \frac{1}{2} V(\vec{q}_j - \vec{q}_i)$ and its probability density in phase-space

$$F_N : \mathbb{R}^+ \times (U \times \mathbb{R}^3)^N \to \mathbb{R} \quad (t, \mathbf{X}^N) \mapsto F_N(t, \mathbf{X}^N), \tag{2}$$

with

$$\int_{S^N} F_N \, \mathrm{d}\mathbf{X}^N = 1, \tag{3}$$

where $\mathbf{X}^N \in S^N = (U \times \mathbb{R}^3)^N$ subsumes the position and momentum variables of the N particles. With these basic ingredients the corresponding Liouville equation reads:

$$\frac{\partial F_N}{\partial t} = -\sum_{i=1}^{N} \left[\nabla_{\vec{q}_i} \cdot (\vec{v}_i F_N) + \nabla_{\vec{p}_i} \cdot \left(-\nabla_{q_i} V_{\text{tot}}(\mathbf{q}^N) F_N \right) \right], \tag{4}$$

where $\vec{v}_i = \vec{p}_i / M_i$ is the ith particle's velocity. The subsystem Ω, when it is (momentarily) occupied by n particles, is characterised by the Hamiltonian:

$$H_n = \sum_{i=1}^{n} \frac{\vec{p}_i^2}{2M} + \sum_{i=1}^{n} \sum_{j \neq i}^{n} \frac{1}{2} V(\vec{q}_j - \vec{q}_i) \quad (\vec{q}_i, \vec{q}_j \in \Omega), \tag{5}$$

and probability density:

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

$$f_n(t, \mathbf{X}^n) = \binom{N}{n} \int_{(S_c)^{N-n}} F_N(t, \mathbf{X}^n, \mathbf{\Xi}_n^N) \, \mathrm{d}\mathbf{\Xi}_n^N \tag{6}$$

$$\mathbf{\Xi}_n^N \equiv [\Xi_{n+1}, \dots, \Xi_N] \quad \text{where } \Xi_i = (\vec{q}_i, \vec{p}_i) \in S_c^{N-n},$$

(See the definitions of S^n and S_c^{N-n} in figure 1.) Importantly, for fixed time the tuple $(f_n)_{n=0}^{N}$ of n-particle functions is a probability density on the direct sum $\oplus_{n=0}^{N}(S^n \to \mathbb{R})$ of individually

un-normalised functions of n arguments that inherits its normalization condition from that for the Universe stated in (3), so that

$$\sum_{n=0}^{N} \int_{\Omega^n} \int_{(\mathbb{R}^3)^n} f_n(t,(\boldsymbol{q},\boldsymbol{p})) \, \mathrm{d}\boldsymbol{p} \, \mathrm{d}\boldsymbol{q} = 1. \tag{7}$$

As anticipated above, the aim is to integrate equation (4) w.r.t. the variables of S_c^{N-n} and to derive from the equation for $F_N(t, \boldsymbol{X}^N)$ a set of $N+1$ Liouville-type equations for the n-state probability density, $f_n(t, \boldsymbol{X}^n)$, in Ω. The details of the integration procedure can be found in references [20, 24], here we report the final result:

$$\frac{\partial f_n}{\partial t} + \sum_{i=1}^{n} \left(\nabla_{\vec{q}_i} \cdot (\vec{v}_i f_n) + \nabla_{\vec{p}_i} \cdot \left(\vec{F}_i f_n \right) \right) = \Psi_n + \Phi_n^{n+1}, \tag{8}$$

where

$$\vec{F}_i = -\sum_{j=1; j\neq i}^{N} \nabla_{\vec{q}_i} V(\vec{q}_i - \vec{q}_j), \qquad \Psi_n = -\sum_{i=1}^{n} \nabla_{\vec{p}_i} \cdot \left(\vec{F}_{\mathrm{av}}(\vec{q}_i) f_n(t, \boldsymbol{X}^{i-1}, X_i, \boldsymbol{X}_i^{n-i}) \right), \tag{9}$$

with

$$\vec{F}_{\mathrm{av}}(\vec{q}_i) = -\int_{S^c} \nabla_{\vec{q}_i} V(\vec{q}_i - \vec{q}_j) f_2^\circ(X_j | X_i) \mathrm{d}X_j, \tag{10}$$

representing the mean field force exerted by the outer particles onto the ith inner particle. Here $f_2^\circ(X_{\mathrm{out}} | X_{\mathrm{in}})$ is the conditional distribution for joint appearances of an outer particle given the state of an inner one. This quantity is either assumed to be known or can be modelled. Finally

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

$$\tag{11}$$

implements the knowledge that the Liouville equation for F^N is an advection equation in the state space of the Universe, so that the theory of characteristics must be used when monitoring which information can leave and enter the subdomain Ω (for details, see reference [20]). In equation (11) $f_1^\circ(\vec{q}_i, -\vec{p}_i)$ represents the modeler's assumption for the one particle distribution of the reservoir at the interface boundary $\partial\Omega$.

3. Multiple reservoirs

An interesting and instructive special case of the model presented in the previous section concerns the case of a subsystem Ω interfaced with two (or more) disjoint reservoirs, each kept at a different thermodynamic condition, that is, at different temperature and average particle density. Such a scenario, illustrated in figure 2, is relevant for cases of subsystems of a large complex system where the (natural or artificial) environment creates a thermal gradient acting on the subsystem of interest and induces thermophoresis [25, 26]. In this context, the derivation of Liouville-like equations for the probability density of the particles in Ω from the Liouville equation of U proceeds as in the previous section, except that now we can be more specific: the integration over the degrees of freedom of Ω_c now separates into the integration over the subdomains Ω_c^1 and Ω_c^2, which individually are assumed to feature spatially homogeneous statistics
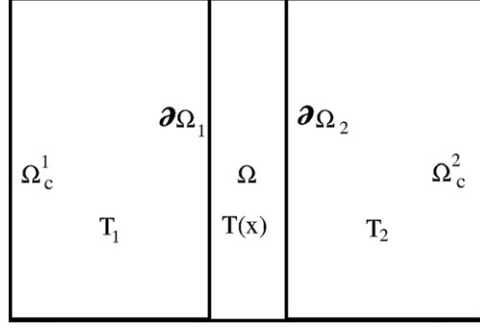
Figure 2. Pictorial illustration with corresponding formalism for the open system Ω interfaced with two disjoint reservoirs of particles at different temperatures. Here $U = \Omega_c^1 \cup \Omega \cup \Omega_c^2$, that is $\Omega_c = \Omega_c^1 \cup \Omega_c^2$, $\partial\Omega = \partial\Omega_1 \cup \partial\Omega_2$. T_1 and T_2 are the temperatures of particles in Ω_c^1 and Ω_c^2, kept constant by external thermostats/sources of heat acting over each thermostat separately. $T(x)$ is the temperature in Ω as a function of the position generated by the difference in temperature of the two reservoirs.

enforced by external heat baths. The detailed derivation is presented in reference [24], whereas here we report only the final results:

$$\frac{\partial f_n}{\partial t} + \sum_{i=1}^{n} \left(\nabla_{\vec{q}_i} \cdot (\vec{v}_i f_n) + \nabla_{\vec{p}_i} \cdot \left(\vec{F}_i f_n \right) \right) = \sum_{k=1,2} \left(\Psi_{n,k} + \Phi_{n,k}^{n+1} \right), \qquad (12)$$

where $k \in \{1, 2\}$ labels the reservoirs and

$$\Psi_{n,k} = -\sum_{i=1}^{n} \nabla_{\vec{p}_i} \cdot \left(\vec{F}_{\mathrm{av}}^k(\vec{q}_i) f_n(t, \mathbf{X}^{i-1}, X_i, \mathbf{X}_i^{n-i}) \right)$$

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

$$(13)$$

Of course this time the modeller needs separate assumptions for the statistics of both reservoirs at the interfaces $\partial\Omega_k$ (i.e., models for $f_1^{\circ,k}(\vec{q}_i, -\vec{p}_i)$, and $f_2^{\circ,k}(X_{\mathrm{out}}|X_{\mathrm{in}})$) for $k \in \{1, 2\}$ since the reservoirs are conditioned independently. The procedure can be straightforwardly extended to an arbitrary number of m disjoint reservoirs $1, 2, 3 \ldots m$, interfaced with Ω:

$$\frac{\partial f_n}{\partial t} + \sum_{i=1}^{n} \left(\nabla_{\vec{q}_i} \cdot (\vec{v}_i f_n) + \nabla_{\vec{p}_i} \cdot \left(\vec{F}_i f_n \right) \right) = \sum_{k=1}^{m} \Psi_{n,k} + \Phi_{n,k}^{n+1}. \qquad (14)$$

Such a formal derivation is rather general and does not imply a specific choice of the functions $f_1^{\circ,k}, f_2^{\circ,k}$, they can represent very general behaviour of the outside world, including discontinuous or continuous variation of their properties along the surface of Ω.

4. The essence of the BL model of thermal bath

The detailed description of the BL model can be found in the original papers of references [1, 2] here we report the key feature of the model required for the current discussion. The BL

model describes a close system of particles in contact with an ideal infinite reservoir at given temperature acting as a source/sink of energy/heat. It can be extended also to open systems with a source/sink of particles (as treated by us in reference [15]), however in the current context such an extension is not needed. For the case of a reservoir that acts as heat bath, the physical principle of the BL model states that the interaction between the system and the bath is impulsive (stochastic) and involves a discrete transition of the system in phase-space from a state, X, to a state X'. From the formal point of view, a transition from a state X to a state X' is governed by a contingent probability $K(X',X)\mathrm{d}X'\,\mathrm{d}t$ where the kernel $K(X',X)$ is a time-independent stochastic function. $K(X',X)$ expresses the probability per unit time that the system at X makes a transition to $X'N$ due to the interaction system–bath. The total interaction between the system and the bath writes: $\int \mathrm{d}X'[K(X,X')F(X',t) - K(X',X)F(X,t)]$. It follows that the general equation of time evolution of the probability density $F(X,t)$, is:

$$\frac{\partial F(X,t)}{\partial t} = -\{F(X,t),H(X)\} + \int \mathrm{d}X'[K(X,X')F(X',t) - K(X',X)F(X,t)]. \tag{15}$$

If direct microreversibility of the dynamics is assumed, then one has that:

$$K(X,X')F(X',t) - K(X',X)F(X,t) = 0; \quad \forall X,X',t. \tag{16}$$

However, for a stationary distribution, Bergmann and Lebowitz in references [1, 2] state that it is sufficient to satisfy a less restrictive condition, without assuming direct microreversibility:

$$\int \mathrm{d}X'[K(X,X')F(X',t) - K(X',X)F(X,t)] = 0. \tag{17}$$

The condition above is more appropriate for our case because direct microreversibility cannot be assumed in presence of a reservoir of particles where the microscopic degrees of freedom are either integrated out (theoretical model) or are fictitious (numerical scheme). In the presence or several, m, distinct baths the total action on the system corresponds to the sum of the individual action of each bath, i.e.:

$$\frac{\partial F(X,t)}{\partial t} = -\{F(X,t),H(X)\} + \sum_{i=1}^{m} \int \mathrm{d}X'[K_i(X,X')F(X',t) - K_i(X',X)F(X,t)]. \tag{18}$$

It must be reported that, in our knowledge, do not exist explicit expressions of $K(X,X')$ that can be used in numerical applications. This model aims at defining universal conditions that a reservoir must fulfil but it does not prescribe a straightforward receipt for applications, as we have discussed also in the previous sections. In next section we will adapt the scheme above to define the heat bath acting only on the particles in Ω_c but not (directly) acting on the particles in Ω.

5. BL-like thermostat for Ω_c

The thermostat acting on the region Ω_c continuously supplies/removes energy to/from Ω_c, i.e. the reservoir of Ω, but should not act directly on Ω. Below we will model the action of the thermostat in Ω_c through the stochastic kernel of BL that mimics the interaction of a system with an ideal infinite reservoir of energy. The Liouville equation for the Universe in presence

of a generic thermostat reads:

$$\frac{\partial F_N}{\partial t} + \sum_{i=1}^{N} \left[\nabla_{\vec{r}_i} \cdot (\vec{p}_i F_N) + \nabla_{\vec{p}_i} \cdot \left(-\nabla_{r_i} V_{\text{tot}}(\mathbf{q}^N) F_N \right) \right] = I(t, X^N), \tag{19}$$

where:

$$I(t, X^N) = \int_{S^N} \left\{ K(X^N, (X^N)') F_N(t, (X^N)') - K((X^N)', X^N) F_N(t, X^N) \right\} \mathrm{d}(X^N)', \tag{20}$$

for the case in which the model of thermostat given by BL is chosen.

In the following to avoid the heavy formalism of adding a parenthesis for labelling the prime, let us define: $(X^N)' \equiv Y^N; (X^n)' \equiv Y^n; (\Xi_n^N)' \equiv \Gamma_n^N$. If we now proceed with the integration w.r.t. the variables of S_c^{N-n} for equation (19), as done in the passage from equation (4) to equation (8), we obtain:

$$\frac{\partial f_n}{\partial t} + \sum_{i=1}^{n} \left(\nabla_{\vec{q}_i} \cdot (\vec{v}_i f_n) + \nabla_{\vec{p}_i} \cdot \left(\vec{F}_i f_n \right) \right) = \Psi_n + \Phi_n^{n+1} + \int_{S_c^{N-n}} I(t, X^N) \mathrm{d}\Xi_n^N. \tag{21}$$

We will show that $\int_{S_c^{N-n}} I(t, X^N) \mathrm{d}\Xi_n^N = 0; \forall t; \forall X^n \in S^n$ if we choose the form of thermostat of equation (20).

In fact, the key characteristics of the applied thermostat is that it must act only on the variables of the reservoir, that is on the variables of S_c, leaving the other variables untouched. This means that the transitions allowed in phase space due to the interaction with the thermostat are all those for which $X \equiv Y \, \forall \, X, Y \in S$, that is $(\vec{q}_i, \vec{p}_i) \equiv ((\vec{q})_i', (\vec{p})_i'), \forall \vec{q}_i, \vec{p}_i, (\vec{q})_i', (\vec{p})_i' \in S$

Let us rewrite $\int_{S_c^{N-n}} I(t, X^N) \mathrm{d}\Xi_n^N$ as:

$$\int_{S_c^{N-n}} I(t, X^N) \mathrm{d}\Xi_n^N = \binom{N}{n} \int_{S_c^{N-n}} \int_{S_c^{N-n}} \int_{S^n} \left\{ K(X^N, Y^N) F_N(t, Y^N) \right.$$
$$\left. - K(Y^N, X^N) F_N(t, X^N) \right\} \mathrm{d}Y^n \, \mathrm{d}\Gamma_n^N \, \mathrm{d}\Xi_n^N, \tag{22}$$

that is, we have partitioned the total integral $I(t, X^N)$ into the domain S^n and its complementary S_c^{N-n}, with $Y^n \in S^n$ and as before $\Gamma_n^N \in S_c^{N-n}$. The integrand in the inner part of the rhs of (22) can be rewritten, according to the above rules for the thermostat acting only on the variables of S_c, as:

$$K(X^N, Y^N) F_N(t, Y^N) - K(Y^N, X^N) F_N(t, X^n)$$
$$= K(X^n; \Xi_n^N, X^n; \Gamma_n^N) F_N(t, X^n, \Gamma_n^N) - K(X^n; \Gamma_n^N, X^n; \Xi_n^N) F_N(t, X^n, \Xi_n^N), \tag{23}$$

that is we have operated the partitioning of the variables of S^N in the variables of S^n and S_c^{N-n} inside the kernel and inside F_N, with the additional condition that the variables in S^n are left unchanged by the action of the thermostat, i.e. $X^n \equiv Y^n$. Finally, if we rewrite the integral (22) exchanging the integration order between $\mathrm{d}Y^n$ and $\mathrm{d}\Xi_n^N$, taking into account that $\mathrm{d}Y^n \equiv \mathrm{d}X^n$, one obtains:

$$\int_{S_c^{N-n}} I(t, X^N) \mathrm{d}\Xi_n^N = \binom{N}{n} \int_{S^n} \int_{S_c^{N-n}} \int_{S_c^{N-n}} \left\{ K(X^n; \Xi_n^N, X^n; \Gamma_n^N) \right.$$
$$\left. \times F_N(t, X^n, \Gamma_n^N) - K(X^n; \Gamma_n^N, X^n; \Xi_n^N) \times F_N(t, X^n, \Xi_n^N) \right\} \mathrm{d}\Gamma_n^N \, \mathrm{d}\Xi_n^N \, \mathrm{d}X^n. \tag{24}$$

The internal integrals are symmetric in Ξ_n^N and Γ_n^N and thus the double internal integration leads to:

$$
\int_{S_c^{N-n}} \int_{S_c^{N-n}} \left\{ K(X^n; \Xi_n^N, X^n; \Gamma_n^N) F_N(t, X^n, \Gamma_n^N) \right.
$$
$$
\left. - K(X^n; \Gamma_n^N, X^n; \Xi_n^N) F_N(t, X^n, \Xi_n^N) \right\} \, \mathrm{d}\Gamma_n^N \, \mathrm{d}\Xi_n^N = 0; \quad \forall t, \quad \forall X^n \in S^n. \tag{25}
$$

The result above tells us something relevant about the model of reference [20], that is, the effect of the thermalization of Ω_c on the particles of Ω remains to be expressed by the modeller's assumption regarding $f_1^\circ(\vec{q}_i, -\vec{p}_i)$ and $f_2^\circ(X_{\mathrm{out}}|X_{\mathrm{in}})$ at the interface boundary $\partial\Omega$. Then, if the particles in Ω_c are assumed to be thermalized at a given temperature T, $f_1^\circ(\vec{q}_i, -\vec{p}_i)$ and $f_2^\circ(X_{\mathrm{out}}|X_{\mathrm{in}})$ must be consistent with this choice. In particular, the probability that a particle in $\partial\Omega$ has a certain momentum must be consistent with a probability distribution of momenta at the wished temperature.

6. Differently thermostatted disjoint regions of Ω_c

Let us consider for simplicity the case of two reservoirs. One needs to show that the condition: $\int_{S_c^{N-n}} I(t, X^N) \mathrm{d}\Xi_n^N = 0; \forall t; \forall X^n \in S^n$ still holds. In this case the model implies the introduction of $K_1(X^N, Y^N)$ and $K_2(X^N, Y^N)$ which act separately in two disjoint regions interfaced with Ω. The introduction of separate kernel also implies that, analogously to the case of a single reservoir, the thermal state of the two different regions enters in the model trough the modeller choice at $\partial\Omega_1$ of $f_1^\circ(\vec{q}_i, -\vec{p}_i)$ and $f_2^{\circ,1}(X_{\mathrm{out}}|X_{\mathrm{in}})$ and of $f_2^\circ(\vec{q}_i, -\vec{p}_i)$, $f_2^{\circ,2}(X_{\mathrm{out}}|X_{\mathrm{in}})$ at $\partial\Omega_2$, i.e. their choice must be compatible with temperature T_1 and T_2 respectively. With the partitioning of space into the subsystem Ω and the two reservoir regions Ω_c^1 and Ω_c^2 one can write:

$$
\int_{S_c^{N-n}} I(t, X^N) \mathrm{d}\Xi_n^N = \int_{S_{c,1}^{N-n}} I_1(t, X^N) \mathrm{d}\Xi_n^N + \int_{S_{c,2}^{N-n}} I_2(t, X^N) \mathrm{d}\Xi_n^N, \tag{26}
$$

with $S_c \equiv S_{c,1} \cup S_{c,2}$ and $S_{c,1} \equiv (\Omega_c^1) \times \mathbb{R}^3$ and analogously for $S_{c,2}$, with the definition:

$$
I_1(t, X^N) = \int_{S_{c,1}^{N-n}} \int_{S^n} \left\{ K_1(X^N, Y^N) F_N(t, Y^N) - K_1(Y^N, X^N) F_N(t, X^N) \right\} \mathrm{d}Y^n \, \mathrm{d}\Gamma_n^N \, \mathrm{d}\Xi_n^N, \tag{27}
$$

analogously for $I_2(t, X^N)$. Here $K_1(\mathbf{X}^N, \mathbf{Y}^N), K_1(\mathbf{Y}^N, \mathbf{X}^N)$ acts on the variables, $(\vec{q}_i, \vec{p}_i), ((\vec{q}_i)', (\vec{p})') \in S_{c,1}$, and analogously, $K_2(\mathbf{X}^N, \mathbf{Y}^N), K_2(\mathbf{Y}^N, \mathbf{X}^N)$ acts on the variables $(\vec{q}_i, \vec{p}_i), ((\vec{q}_i)', (\vec{p})') \in S_{c,2}$. It follows:

$$
\int_{S_c^{N-n}} I(t, X^N) \mathrm{d}\Xi_n^N = \sum_{i=1,2} \int_{S_{c,i}^{N-n}} \int_{S_{c,i}^{N-n}} \int_{S^n} \left\{ K_i(X^N, Y^N) F_N(t, Y^N) \right.
$$
$$
\left. - K_i(Y^N, X^N) F_N(t, X^N) \right\} \mathrm{d}Y^n \, \mathrm{d}\Gamma_n^N \, \mathrm{d}\Xi_n^N, \tag{28}
$$

as for the case of a single reservoir, each term of the sum is equivalent the integral of equation (22). Thus each term separately leads to the expression equivalent to that of equation (25), that is to integrals that are symmetric in Ξ_n^N and Γ_n^N and thus to $\int_{S_{c,i}^{N-n}} I_i(t, X^N) \mathrm{d}\Xi_n^N = 0$.
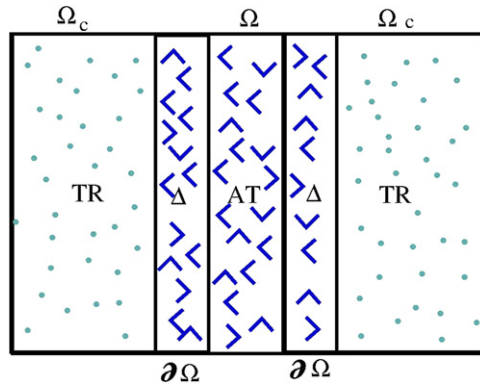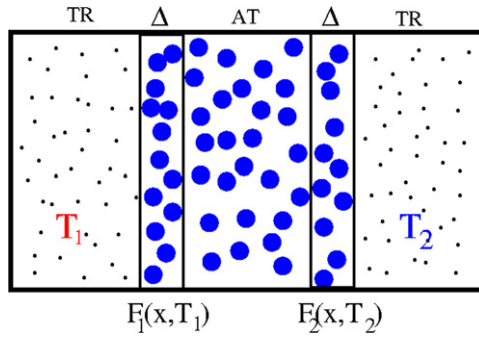
Figure 3. Pictorial illustration of the AdResS set up and the equivalence of its space partitioning with the space partitioning of the physico-mathematical model of reference [20] and section 7.1: $AT \equiv \Omega$; $\Delta \equiv \partial\Omega$; $TR \equiv \Omega_c$. In $AT \cup \Delta$ molecules are treated at atomistic resolution, while in TR the molecules become non-interacting point-particles (tracers). Molecules can freely move in space and across the different regions.

7. Application to molecular simulation

In this section we first introduce the AdResS technique and then we review an application regarding an open system in a thermal gradient. In the example the reservoirs play a key role and the results of the current paper assure conceptual/physical consistency to the numerical results.

7.1. AdResS: a molecular simulation technique that mimics an open system in contact with a reservoir of energy and particles

AdResS, is a scheme of MD that allows to treat molecules at different resolution according to their position in space. As a matter of fact the region of high resolution, that is the region of major interest, is equivalent to an open system embedded in a large reservoir of molecules at low resolution. Further progress has brought AdResS to be closer to the theoretical idea of an open subsystem of molecules embedded in a mean-field reservoir. In fact, in its current incarnation, the particles in the low resolution region have been reduced to the very essential representation, i.e. non interacting point-like particles (tracers) regulated by a mean-field [19] (see also figure 3). Specifically, the simulation box is partitioned in AT region, where molecules have atomistic resolution and where the observation of physical properties takes place; Δ regions, interfaced on both sides of the AT region, where molecules still have atomistic representation, and finally a particle reservoir region, TR, where molecules are represented by non-interacting point-particles. The Δ region, despite the atomistic resolution, is artificial and it is only needed to design proper boundary conditions for AT. In fact in Δ an external one-particle force, called thermodynamic force, imposes constraints on the molecular density so that the resulting density in Δ is compatible with the target molecular density of the AT region. The thermodynamic force has been derived by first principles of thermodynamic and statistical mechanics to assure that calculated physical quantities in the AT region are equivalent to the corresponding quantities in a simulation of reference where the whole box is at atomistic resolution [26, 28]. In addition an external thermostat fixes the temperature in $\Delta \cup TR$ at a wished value. The definition of the thermodynamic force and of the thermostat in Δ is, in an effective

Figure 4. Pictorial illustration of the AdResS set up for simulating a thermal gradient, $\Delta T = T_1 - T_2$, acting in the open (fully atomistic) region AT. At each $AT/\Delta + TR$ interface a different thermodynamic force, $F_i(x, T_i)$ with $i = 1, 2$, is calculated following the indications of the theoretical model of references [20, 24].

way, the numerical equivalent of the boundary conditions derived for a subsystem in reference [20]. A set up with a spherical AT region where Δ represents its external surface, while the rest of the box represent the TR region, is also used often in AdResS simulations; it represents a straightforward technical extension of the scheme of figure 3. It must be underlined that the tracer particles do not play any explicit role in the physical model and in the simulation results. They are required for technical reasons so that the numerical algorithm that dynamically feeds the molecular region with particles (or adsorbs particles, if in excess) is implemented in an efficient manner. The resulting numerical scheme has been shown to deliver highly efficient and accurate simulations of large systems with very complex molecular structures, e.g., hydrated biomembranes [29], to name one relevant example.

7.2. Open system in a thermal gradient

In references [21, 24] the AdResS approach was employed to study an open system in a thermal gradient, as illustrated in figure 4.

The boundary conditions at the two different interfaces between the AT and the (two distinct) Δ regions were derived following the results shown in section 6. Specifically, at each interface (separately) the one particle density is enforced to be as if the open system was in equilibrium with each single reservoir, as suggested by equation (12). In the terminology of AdResS this implies the definition of two different thermodynamic forces, one acting at the AT/Δ interface on the left and one acting at the AT/Δ interface on the right, as illustrated in figure 4. In addition, the imposition of a thermal gradient in the AT region requires also the use of distinct thermostats that keep the two disjoint reservoirs at the respective target temperature. With this set up the simulation results were highly satisfactory and an example is reported here in figure 5.

However, in the original theoretical model of references [20, 24] the description of a sharp thermal gradient, equivalent to that of the numerical study, is possible only under a restrictive approximation. In fact in absence of any external source that provides/removes heat, the different parts of $U \backslash \Omega$ connected via Ω would exchange particles and energy until eventually a state of equilibrium across the entire Universe U were reached. Thus one needs the restrictive physical approximation that the reservoirs are large enough to approximately
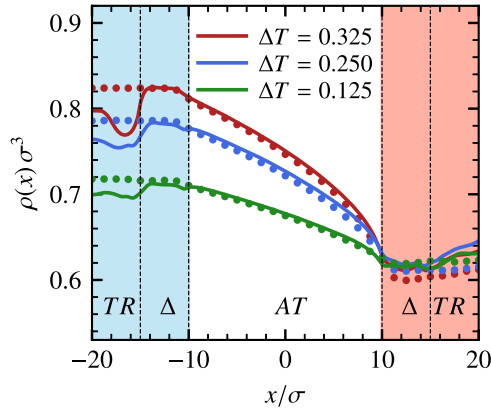
Figure 5. Particle density across the AdResS simulation box under different thermal gradients (continuous line) compared with a full atomistic simulation of reference (dotted line). The agreement in the *AT* region (i.e. region of interest) is very satisfactory and shows that the AdResS technique, with boundary conditions designed following the formal results of reference [20], is a reliable numerical approach for the simulation of an open system embedded in a reservoir of particle and energy, in and out of equilibrium. The quantities plotted are in Lennard-Jones units. [24] John Wiley & Sons. © 2021 The Authors. Advanced Theory and Simulations published by Wiley-VCH GmbH. CC BY-NC-ND 4.0.

maintain their fixed uniform temperatures over short time intervals of observation. Obviously there is an inconsistency between the theoretical model and the numerical implementation because, as a matter of fact, the reservoirs of the numerical model are finite (not much larger than the *AT* region) and there is no restriction on the time of observation in the simulation study. The results of the current paper solve this problem and thus legitimize the design of the boundary conditions in AdResS according to the equations of the theoretical model. In addition, it must be underlined that despite the fact that the BL kernel cannot be explicitly used in numerical applications, the results of AdResS, with boundary conditions taken from the theoretical model, are found to be highly satisfactory with several numerical thermostats. In particular the Andersen [30] thermostat, used in the application shown in this section, and the Langevin [31] thermostat, recently used in several AdResS applications for biophysical systems [32]. Such thermostats are very different in their practical realization, however they both fulfil the BL conditions in a practical, albeit different, manner.

8. Open mathematical problems and further modelling perspectives

In this section we list a series of mathematical and modelling pending problems related to the model of open systems. The solution of such questions would allow for a substantial step forward in the rigorous mathematical formalization of the model; this in turn will allow a further development of the numerical approach with the possibility of simulating a broader spectrum of physical and chemical systems.

8.1. Limit $N \to \infty$; $\Omega_c \to \infty$ at finite Ω

In our view the most relevant mathematical problem with important consequences for the applications, is the limit to $N \to \infty$; $\Omega_c \to \infty$ at fixed Ω, that is, an infinite reservoir for a finite open system. While it is not a major problem to properly define $F_N(\mathbf{X}_N, t)$ for a finite N, it may turn to be not trivial to define its corresponding limit for $N \to \infty$ in this context and derive, in such a limit, a corresponding hierarchy of equations for $f_n(\mathbf{X}_n, t)$. The problem has a similarity with the Boltzmann–Grad limit [33] for the Bogoliubov–Born–Green–Kirkwood–Yvon (BBGKY) hierarchy [34]. However, although the hierarchy of equations in references [20, 24] looks similar to the BBGKY hierarchy, there is a substantial difference: in the BBGKY approach for a system of N particles the Liouville equation is derived for a subsets of n particles with probability distribution function defined as the $N - n$-marginalised probability distribution function of the total system, but without the partitioning of space in Ω and Ω_c. Thus the hierarchy of equations in references [20, 24] is different in form and meaning from the BBGKY hierarchy; the modelling hypothesis and mathematical analysis of the $\partial\Omega$ boundary is crucial for the model of references [20, 24] and for its corresponding numerical implementation while such an aspect is not present in the BBGKY approach. Nevertheless the limit $N \to \infty$ implies similar challenges and thus existing results [35] for the Boltzmann–Grad limit may be used for the current problem. The limit $N \to \infty$ would be crucial for deriving equations that allow for the coupling of Ω to a hydrodynamic regime and thus suggest conditions at the interface $\partial\Omega$ so that the large scale information of Ω_c, with the macroscopic conditions and corresponding fluctuations, are passed onto Ω. The derived boundary conditions will then lead to a numerical approach for accurate multiscale simulations linking the molecular scale of Ω to the hydrodynamic scale of Ω_c.

8.2. Modelling beyond $f^\circ(\vec{q}_i, -\vec{p}_i)$ at $\partial\Omega$

The key result of references [20, 24] is the derivation of the terms: Ψ_n and Φ_n^{n+1}. These terms, in turn, lie on the modeller choice of a reasonable approximation for the particle probability distribution of the reservoir in $\partial\Omega$, that is $f^\circ(\vec{q}_i, -\vec{p}_i)$. The strong assumption is that the interaction potential is short-ranged so that corresponding particle–particle long-range correlations between particles in Ω and particles in Ω_c can be neglected. In molecular simulations one has electrostatic interactions and thus the short-range approximation is in principle not suitable. However simulations have often proven, *a posteriori*, that a strong charge screening effect may take place even for systems where the electrostatic interactions on the long range were thought to be crucial [36]. Nevertheless a more involved design of $f^\circ(\vec{q}_i, -\vec{p}_i)$ and $f_2^\circ(X_{\text{out}}|X_{\text{in}})$ where a rigorously derived effective range of interaction, due to e.g. screening effects, is considered would make the theoretical model closer to the reality of simulated systems. Furthermore, as underlined in reference [24], it is worth exploring the possibility of extending $f^\circ(\vec{q}_i, -\vec{p}_i)$ from a single-particle distribution, to two-,three-.... s-particle distribution at $\partial\Omega$ that depend explicitly on the particle number n of the open system. Such an extension would account for the effects of Ω on Ω_c because long-range interactions imply that particles in Ω do influence a sizeable number of particles in Ω_c. In turn, such an influence generates a response of the reservoir on the system, thus the mathematical analysis and corresponding modelling of such a process is very relevant for the correct description of the physics in Ω. The extension of $f^\circ(\vec{q}_i, -\vec{p}_i)$ to many-particle function will transform the terms Ψ_n and Φ_n^{n+1} in nonlinear functions of f_n and thus account for the description of non-linear coupling effects. It must be underlined that the particle-based exchange of information between the open system and the reservoir at $\partial\Omega$ is the actual novelty and conceptual advantage of the corresponding computational method compared to other numerical methods. In molecular simulation, usually, open systems with

varying number of particles are treated by introducing/removing particles from the simulation box through a Monte Carlo move regulated by an *a priori* fixed chemical potential [10, 13]. The particle-based properties of the reservoir and the corresponding passage of information between system and reservoir (and vice versa) at $\partial\Omega$, through a proper dynamical exchange of particles, are lost, thus providing only a partial description of the physical properties of an open system embedded in a large environment [14].

8.3. Extension to systems with quantum, electronic, degrees of freedom

The mathematical model discussed so far, describes only classical particles/molecules, thus the next natural step would be the inclusion of electronic degrees of freedom described at quantum mechanical level. The explicit consideration of electronic degrees of freedom not only requires different physical principles than the classical ones, but also, and above all, implies the use of more complex mathematical quantities. From the mathematical point of view, there exists a large amount of work which treats quantum/electronic degrees of freedom in open systems, however current available literature mostly focuses on simple systems. For example the Lindblad master equation [37, 38], that is von Neumann equation for the density matrix for open systems [18], is applied to ideal cases, but, so far, has not been incorporated in any numerical scheme for treating electronic degrees of freedom in complex molecular systems [39]. From the point of view of simulation, the difficulty of developing efficient simulation schemes with rigorous mathematical models has led to the use of numerical tools based on qualitative/empirical principles where often even basic physical accuracy comes into question. The class of current MD approaches that treats open systems of molecules with electronic degrees of freedom is called 'adaptive quantum mechanics/classical mechanics' (A-QM/MM) [40, 41]. The variation of number of molecules in an open system/subsystem is mimicked by performing a series of simulations of a closed system, each with a different number of molecules; static and dynamics properties of the subsystem are calculated by averaging over the closed-system simulations. Ayers and Miranda-Quintana have derived a rigorous result [42] which shows that, by applying the current A-QM/MM techniques, one samples an artificial ensemble not compatible with any physically well founded statistical ensemble, thus leading to artificial results (see also reference [43]). The consequence is that, as discussed in reference [40], results of A-QM/MM calculations are not reliable. The conclusion is that the extension of the model of references [20, 24] to electronic degrees of freedom is highly desirable, not only in the mathematical field, but, above all in the field of simulation.

9. Conclusions

We have revised the derivation made in references [20, 24] regarding the Liouville-like hierarchy of equations for the probability distribution in phase space of an open system with one or more particle reservoirs. The contribution of this work concerns the explicit introduction of thermal baths acting solely in the reservoir regions; to this aim the analytic form for the thermal bath was taken from a popular model by Bergmann and Lebowitz and adapted to the current model of open system. The results show that the explicit inclusion of the thermal baths does not change the form of the equations and that the thermal condition of the reservoirs is expressed through the modeller choice of the statistics of reservoir's particles at the boundary of the open system. Such a choice explicitly enters in the system–reservoir coupling term, $\Psi_n + \Phi_n^{n+1}$, in the Liouville-like equation for the probability density in phase space, $f_n(\mathbf{X}_n, t)$, of the open system Ω. This result is relevant to guide the modeller to make specific choices for the one-

and two-particle distribution of the reservoir at the interface boundary $\partial\Omega$, as required by the model. The explicit introduction of the thermal bath allows the theoretical model to resemble in a closer manner the numerical scheme employed in simulations and thus to explore the full power of the abstract model in designing optimal numerical schemes. Finally we list a series of mathematical and modelling pending problems whose solutions would push the mathematical model to a more rigorous framework and would allow the numerical simulation scheme to tackle a broader class of physical and chemical problems. Beyond the technical relevance of the current results, this work aim at showing a realistic case where the interplay between mathematical modelling and the development of a molecular simulation scheme led to relevant steps forward in both fields. In perspective, the hope is that such a synergy may be enforced further so that rigorous mathematics accompanied by impactful applications becomes the standard routine in molecular simulation.

Acknowledgments

Data availability statement

The data generated and/or analysed during the current study are not publicly available for legal/ethical reasons but are available from the corresponding author on reasonable request.

ORCID iDs

Rupert Klein ⓘ https://orcid.org/0000-0001-8032-3851
Luigi Delle Site ⓘ https://orcid.org/0000-0001-8115-8261

References

[1] Lebowitz J L and Bergmann P G 1957 Irreversible Gibbsian ensembles *Ann. Phys., NY* **1** 1
[2] Bergmann P G and Lebowitz J L 1955 New approach to nonequilibrium processes *Phys. Rev.* **99** 578
[3] Emch G G and Sewell G L 1968 Nonequilibrium statistical mechanics of open systems *J. Math. Phys.* **9** 946
[4] Robertson B and Mitchell W C 1971 Equations of motion in nonequilibrium statistical mechanics: III. Open systems *J. Math. Phys.* **12** 563
[5] Li K-H 1986 Physics of open systems *Phys. Rep.* **134** 1–85
[6] Quian H 2007 Phosphorylation energy hypothesis: open chemical systems and their biological functions *Annu. Rev. Phys. Chem.* **58** 113
[7] Rotter I 2009 A non-Hermitian Hamilton operator and the physics of open quantum systems *J. Phys. A: Math. Theor.* **42** 153001
[8] Mikhailov A and Ertl G 2017 *Chemical Complexity* (Berlin: Springer)
[9] Thoss M and Evers F 2018 Perspective: theory of quantum transport in molecular junctions *J. Chem. Phys.* **148** 030901
[10] Frenkel D and Smit B 2002 *Understanding Molecular Simulation* 2nd edn (New York: Academic)
[11] Tuckerman M E 2010 *Statistical Mechanics: Theory and Molecular Simulation* (Oxford: Oxford University Press)
[12] Hollingsworth S A and Dror R O 2018 Molecular dynamics simulation for all *Neuron* **99** 1129–43

[13] Delle Site L and Praprotnik M 2017 Molecular systems with open boundaries: theory and simulation *Phys. Rep.* **693** 1–56

[14] Ciccotti G and Delle Site L 2019 The physics of open systems for the simulation of complex molecular environments in soft matter *Soft Matter* **15** 2114

[15] Agarwal A, Zhu J, Hartmann C, Wang H and Delle Site L 2015 Molecular dynamics in a grand ensemble: Bergmann–Lebowitz model and adaptive resolution simulation *New J. Phys.* **17** 083042

[16] Praprotnik M, Delle Site L and Kremer K 2005 Adaptive resolution molecular-dynamics simulation: changing the degrees of freedom on the fly *J. Chem. Phys.* **123** 224106

[17] Praprotnik M, Delle Site L and Kremer K 2008 Multiscale simulation of soft matter: from scale bridging to adaptive resolution *Annu. Rev. Phys. Chem.* **59** 545–71

[18] Delle Site L 2016 Formulation of Liouville's theorem for grand ensemble molecular simulations *Phys. Rev.* E **93** 022130

[19] Delle Site L, Krekeler C, Whittaker J, Agarwal A, Klein R and Höfling F 2019 Molecular dynamics of open systems: construction of a mean-field particle reservoir *Adv. Theory Simul.* **2** 1900014

[20] Delle Site L and Klein R 2020 Liouville-type equations for the n-particle distribution functions of an open system *J. Math. Phys.* **61** 083102

[21] Ebrahimi Viand R, Höfling F, Klein R and Delle Site L 2020 Theory and simulation of open systems out of equilibrium *J. Chem. Phys.* **153** 101102

[22] Delle Site L, Praprotnik M, Bell J B and Klein R 2020 Particle-continuum coupling and its scaling regimes: theory and applications *Adv. Theory Simul.* **3** 1900232

[23] Hünenberger P H 2005 Thermostat algorithms for molecular dynamics simulations *Adv. Polym. Sci.* **173** 105–49

[24] Klein R, Ebrahimi Viand R, Höfling F and Delle Site L 2021 Nonequilibrium induced by reservoirs: physico-mathematical models and numerical tests *Adv. Theory Simul.* **4** 2100071

[25] Duhr S and Braun D 2006 Why molecules move along a temperature gradient *Proc. Natl Acad. Sci.* **103** 19678–82

[26] Aita L and Givli S 2014 A theoretical study of biological membrane response to temperature gradients at single-cell level *J.R.Soc. Interface* **11** 20131207

[27] Fritsch S, Poblete S, Junghans C, Ciccotti G, Delle Site L and Kremer K 2012 Adaptive resolution molecular dynamics simulation through coupling to an internal particle reservoir *Phys. Rev. Lett.* **108** 170602

[28] Gholami A, Höfling F, Klein R and Delle Site L 2021 Thermodynamic relations at the coupling boundary in adaptive resolution simulations for open systems *Adv. Theory Simul.* **4** 2000303

[29] Whittaker J and Delle Site L 2019 Investigation of the hydration shell of a membrane in an open system molecular dynamics simulation *Phys. Rev. Res.* **1** 033099

[30] Andersen H C 1980 Molecular dynamics simulations at constant pressure and/or temperature *J. Chem. Phys.* **72** 2384

[31] Hoover W G, Ladd A J C and Moran B 1982 High-strain-rate plastic flow studied via nonequilibrium molecular dynamics *Phys. Rev. Lett.* **48** 1818

[32] Delle Site L 2022 Investigation of water-mediated intermolecular interactions with the adaptive resolution simulation technique *J. Phys.: Condens. Matter.* **34** 115101

[33] Grad H 1949 On the kinetic theory of rarefied gases *Commun. Pure Appl. Math.* **2** 331–407

[34] Bogoliubov N 1946 Kinetic equations *J. Phys. USSR* **10** 265

[35] Langford O E 1975 Time evolution of large classical dynamical system *Lect. Notes Phys.* **38** 1–111

[36] Shadrack Jabes B and Delle Site L 2018 Nanoscale domains in ionic liquids: a statistical mechanics definition for molecular dynamics studies *J. Chem. Phys.* **149** 184502

[37] Kossakowski A 1972 On quantum statistical mechanics of non-Hamiltonian systems *Rep. Math. Phys.* **3** 247

[38] Lindblad G 1976 On the generators of quantum dynamical semigroups *Commun. Math. Phys.* **48** 119

[39] Manzano D 2020 A short introduction to the Lindblad master equation *AIP Adv.* **10** 025106

[40] Zheng M and Waller M P 2016 Adaptive quantum mechanics/molecular mechanics methods *WIREs Comput. Mol. Sci.* **6** 369–85

[41] Duster A, Wang C-H, Garza C, Miller D and Lin H 2017 Adaptive quantum/molecular mechanics: what have we learned, where are we, and where do we go from here? *WIREs Comput. Mol. Sci.* **7** e1310

[42] Miranda-Quintana R A and Ayers P W 2016 Interpolation of property-values between electron numbers is inconsistent with ensemble averaging *J. Chem. Phys.* **144** 244112

[43] Delle Site L 2018 Simulation of many-electron systems that exchange matter with the environment *Adv. Theory Simul.* **1** 1800056