Adaptive Resolution Simulation as a Grand Canonical Molecular Dynamics Scheme: Principles, Applications and Perspectives

Luigi Delle Site,[1, a)] Animesh Agarwal,[1, b)] Christoph Junghans,[2] and Han Wang[1, c)]

[1)] *Institute for Mathematics, Freie Universität Berlin, Germany*

[2)] *Theoretical Division (T-1), Los Alamos National Laboratory, Los Alamos, NM 87545, USA*[d)]

We describe the adaptive resolution multiscale method AdResS. The conceptual evolution as well as the improvements of its technical efficiency are described step by step, with an explicit reference to current limitations and open problems.

I. INTRODUCTION

Many problems in the field of condensed matter, intended in the broadest sense, are multiscale in nature. The concept of "multiscale" is here intended as the interplay between different scales, where such an interplay must have the fundamental role in the properties of a system. Unfortunately nowadays the term "multiscale" is often misused, and, above all in the community of molecular simulation, almost every problem is forced into this category. In reality, an exhaustive description of physical phenomena corresponding to a multiscale process requires, in principle, the simultaneous treatment of all the degrees of freedom (DOFs) of the system. This is obviously a prohibitive task not only because of the required computational resources but above all because of the large amount of data produced. In fact, we would have to deal with a huge amount of data that would mostly contain information not essential to the problem analyzed and will inevitably overshadow the underlying fundamental physics or chemistry of the system. In general, the desired goal is that of treating in a simulation only those DOFs which are strictly required by the specific problem and the choice of the relevant DOFs should be done in such a way that fundamental physical principles are not violated. This aim is at the basis of the modern research industry dedicated to the development of multiscale models in simulation, however, too often, the basic requirement of physical conceptual consistency is violated. From the practical point of view, in order to produce methods that allow us to study large scale phenomena, physical approximations are often unavoidable, however one must be really careful regarding this aspect. In fact it should be mandatory to address the following points: (a) why some methods work particularly well, despite conceptual conditions are not met; (b) a systematic determination of the error made because of the physical approximations done. Most of the multiscale methods in molecular simulation do not proceed along these directives, thus, despite their popularity, they do not gain credibility outside the field of molecular simulation.

In this perspective, this chapter will report about the development and application of, AdResS[1–3], a multiscale method based on the concept of adaptive molecular resolution in space. We will report its theoretical foundations, its merits, its limitations and its extension as a Grand Canonical Molecular Dynamics set up. Large space will be given to the discussion of why the method sometimes works and sometimes does not, to the necessity of providing, as much as possible, a mathematical/rigorous basis to the theoretical foundations and to the capability of defining criteria of control of errors caused by physical approximations. Our aim is to provide the reader with the tools for understanding the current level of conceptual consistency and the capability of controlling of approximations/errors; finally, several sections are dedicated to open problems where further development is still required.

II. THEORETICAL FOUNDATIONS

The AdResS approach falls in the general category of *concurrent coupling methods*[4], that is, those methods where scales cannot be separated and all the relevant DOFs must be treated simultaneously. In order to slowly introduce the idea we make here the similarity with the well established Quantum Mechanics/Molecular Mechanics (QM/MM) method[5]. QM/MM is based on the idea that a fixed subsystem is described with a quantum resolution while the remaining part of the system is treated at classical atomistic level, in this book there are several examples of such an approach. The similarity of QM/MM with AdResS, is that space is divided in different regions in each of which a different molecular

[a)]Electronic mail: luigi.dellesite@fu-berlin.de

[b)]Electronic mail: animesh@zedat.fu-berlin.de

[c)]Electronic mail: han.wang@fu-berlin.de

[d)]Electronic mail: junghans@lanl.org

resolution is employed according to the requirements of the problem. However, *per se*, the QM/MM approach leads to several conceptual problems, the relevant one for the discussion here is that the two regions are not open, thus the free exchange of molecules from one region to the other is forbidden and thus particles density fluctuations are automatically suppressed; these latter are crucial for the correct description of the thermodynamics of bio-systems. Of course one can take into account this aspect and quantify the possible error, however the predictive power of QM/MM is highly restricted when the interplay between thermodynamic fluctuations and electronic structure is a key aspect of a physical process. The idea of AdResS is instead that of allowing for the free exchange of particles, which, passing from one region to the other, change their resolution accordingly and preserve the thermodynamics of the system, that is the thermodynamic properties the system would have if it was treated as a whole at high resolution. The process of changing resolution and preserve the thermodynamic (and structural) properties requires the development of statistical mechanics concepts which do not follow the standard textbook guidelines and thus are often subject to the prejudice of the large portion of conservative researchers in the molecular simulation community. In the next sections we will describe, step by step, the concept we have developed, starting from an empirical intuitive idea up to the definition of a Grand Canonical-like set up. It must be underlined, that differently from QM/MM, at this stage, we couple only classical (atomistic and coarse-grained) systems, however, idea based on the AdResS scheme have been, in the meanwhile, employed to extend QM/MM to an open system set up. A brief discussion about this aspect can be found in the section dedicated to the extension of AdResS to quantum systems.

A. From AdResS to GC-AdResS

The starting point of the original AdResS idea is an intuitive empirical principle whose main components are:

- the simulation box is divided in three regions, one characterized by atomistic resolution, one at coarse-grained (e.g. spherical) resolution and, in between, a small transition region where molecule have a hybrid resolution as explained below.

- Molecules in the different regions are coupled through a spatial interpolation formula on the forces:

$$\mathbf{F}_{\alpha\beta} = w(X_\alpha)w(X_\beta)\mathbf{F}_{\alpha\beta}^{Atom} + [1 - w(X_\alpha)w(X_\beta)]\mathbf{F}_{\alpha\beta}^{CG} \tag{1}$$

  where α and β indicate two molecules, $\mathbf{F}^{Atom}$ is the force derived from the atomistic force field and $\mathbf{F}^{CG}$ from the corresponding coarse-grained potential, X is the x coordinate of the center of mass of the molecule and w is an interpolating function which smoothly goes from 0 to 1 (or vice versa) in the interface region, (Δ), where the lower resolution is slowly transformed (according to w) in the high resolution (or vice versa), as illustrated in Fig.1.
  According to Eq.1 two atomistic molecules interact via a force derived only from an atomistic potential, coarse-grained molecules, instead, interacts with all the others via a force derived only via a coarse-grained potential (coarse-grained molecules do not posses any atomistic degrees of freedom), while for the other cases molecules interact according to their coupled value of $w(X_\alpha)w(X_\beta)$, that is with some space-dependent hybrid combination of atomistic and coarse-grained potential . This implies that a molecule moving from the atomistic to the coarse-grained region, experience a slowly diminishing effect of the force on its atomistic degrees of freedom (rotations and vibrations) and experience an increasing effect (on the center of mass) of the coarse-grained potential. This implies that the molecule experiences the various interactions more and more as a sphere would do. The same process, in reverse order, (acquire degrees of freedom) happens to a coarse-grained molecule going towards the atomistic region. A detailed explanation of the technical implications of Eq.1 can be found in[6,8] . The (physical) empirical, idea behind Eq.1 is that a smooth interpolation on the forces will allow for a transition from an atomistic to a coarse-grained dynamics of a molecule when it goes from the atomistic to the coarse-grained region (and vice versa). Since the interpolation is smooth and the transition region is relatively small compared to the other regions, the perturbation of the dynamics in the atomistic or coarse-grained region, due to the change of resolution, is negligible and thus the behavior of molecules in the two regions is statistically very close to that of a full atomistic or full coarse-grained simulation (in each corresponding region).

- A thermostat is added to assure the overall thermodynamic equilibrium at a given temperature and to take care of the loss of energy due to the perturbation of the transition region.

This essential set up, though very simplified, turned out to be extremely accurate when implemented in an MD approach for toy systems[1,9,10]; later it was successfully employed to simulate physio-chemical systems[11–13]. The coarse-grained model was derived from full atomistic simulations in such a way that it reproduces the molecule-molecule radial distribution at the same temperature, density and pressure[14,15]. However in these studies it was
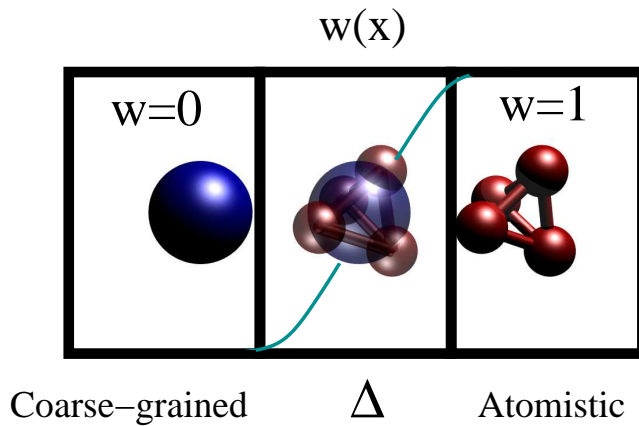
FIG. 1. Pictorial representation of the adaptive box and corresponding molecular representation. The region on the left, is the low resolution region (coarse-grained), the central part is the transition (hybrid) region Δ, where the switching function $w(x)$ is defined, and the region on the right, is the high resolution region (atomistic).

noted that in the transition region there was a small systematic discrepancy between the target particle number density and that obtained in the adaptive simulation. It is obvious that in this region the accuracy cannot be high but on the other hand this region is of no physical interest for the simulation. However the discrepancy mentioned above, implies that also in the atomistic and coarse-grained region there is a systematic error (though small); we will show in later sections that reproducing the density in the transition region is indeed crucial for the Grand-Canonical-like set up. The natural question arising at this point is why such a discrepancy occurs and whether it is possible to develop a scheme to control and improve the accuracy in the transition region. If such a scheme can be found then as a consequence we would be able to improve the accuracy in the region we are mostly interested, that is the high resolution (atomistic) region. The question posed above implies a deeper understanding of the thermodynamics and statistical mechanics of a system simulated with adaptive molecular resolution and the essence of the problem can be summarized as it follows:

- The adaptive approach is based on force interpolation (1) and such interpolated force cannot be derived from a potential (for a rigorous proof see[16,17]); this aspect will be discussed in a detailed way in one of the next sections.

- The lack of a global Hamiltonian implies that thermodynamic properties cannot be defined on the basis of standard statistical mechanics. The *laws* of standard thermodynamics should be reconsidered and properly introduced within the framework of the adaptive resolution system.

- Given the scenario reported above, it naturally comes the question of how, when using AdResS, one can assure *a priori* that thermodynamic and structural properties are properly described.

- However, *a posteriori* checks of thermodynamics and structural properties of simulated systems, show that the accuracy is actually very high. The only sizeable discrepancy is the previously reported enhanced value of the particle density in the transition region.

The approach employed to solve the question above consists of going beyond the traditional path of defining (at any cost) a Hamiltonian. This, in fact, would imply that we are likely to perform the calculation of relevant thermodynamic properties in a "forced/artificial" Canonical ensemble where artifacts may influence results in a sizeable manner. The question to pose is not about the strict definition of the statistical ensemble, rather the definition of the thermodynamic quantities of interest and how to properly calculate them in a molecular simulation characterized by the process of adaptive molecular representation. This is actually the key point, the adaptive resolution does not have any physical meaning, a system of N water molecules must have the overall properties of water independently from the molecular

representation. It will have a more detailed local description in the atomistic region, but its overall density, temperature and pressure must be the same everywhere (atomistic, coarse-grained and even transition region) and should be the same as those of a reference full atomistic system of N molecules. With this principle in mind, two key steps were done; the first regards a rigorous definition of temperature in the transition region.

1. Temperature in Δ

Temperature is well defined in the atomistic and coarse-grained region via the equipartition theorem, as the average kinetic energy divided by the total number of DOFs, $T_{Atom} = \frac{2\langle K_{Atom}\rangle}{n_{Atom}}$; $T_{CG} = \frac{2\langle K_{CG}\rangle}{n_{CG}}$. However, such a definition does not apply in the transition region where one would have: $T_\Delta = \frac{2\langle K_\Delta\rangle}{n_\Delta}$; here DOFs are partially removed/reintroduced as a function of the position thus one cannot properly define $\langle K_\Delta\rangle$, while $n_\Delta = n_{Atom}w(x) + n_{CG}(1 - w(x))$, with n the number of DOFs, molecules would have if they had full atomistic resolution. At this point, one must notice that as a matter of fact, the calculation of the average kinetic energy in an ideal canonical ensemble characterized by an adaptive molecular resolution, would imply that in the transition region each removed/reintroduced DOF does not fully contribute to the average kinetic energy. In fact in the integral of $\langle K\rangle$ the dimensionality of a transforming DOF goes between zero (a coarse-grained molecule does not have this DOF) and one (an atomistic molecule fully possesses such a DOF). This situation can be described in an exact way via the mathematical formalism of integration in fractional dimensions and implies only to consider an extra multiplicative term in the integral of $\langle K\rangle$[16,18]. The overall result shows that the temperature in each point, x in the Δ region can be calculated as: $T_\Delta = \frac{2\langle K_\Delta\rangle}{n_\Delta(x)}$; moreover it turned out that $\langle K_\Delta\rangle$ is proportional to $n_\Delta(x)$ and thus the ratio, $\frac{\langle K_\Delta\rangle}{n_\Delta(x)}$, leads to a constant value which is nothing else than the target temperature of the reference full atomistic system. Computational tests showed that indeed, with the definition of temperature in Δ, as derived above, the system was automatically kept at the expected thermal equilibrium during the simulation. This was the first essential step in the direction of having a deeper understanding and justification of the basic algorithm. Once the procedure for the calculation of the temperature was proven satisfactory, then we moved to next problem, that is address the question of why in Δ a small, but systematic, discrepancy for the particle density is observed in all simulations.

2. Balancing the chemical potential

Let us consider the typical AdResS set up and ask the question of what would preserve thermodynamic equilibrium, and thus uniform molecular density, if one interfaces three open systems with different molecular representations. If each molecular resolution was a different molecular species and, in first instance, one starts with the same density all over the space, then the different species would remain in equilibrium if the chemical potential is the same for all species. For an adaptive set up, molecules have only a different representation, they do not belong to different species and are free to move from one region to the other, however, for similarity to the case above, if the density must remain constant, then the chemical potential characterizing each molecular resolution (at the given temperature, density and pressure) should be the same as that of the reference full atomistic case. Let us suppose it is possible to calculate the chemical potential corresponding to each resolution, this would lead, over the simulation box, to a space dependent chemical potential $\mu(x)$, where $\mu(x) = \mu_{Atom}$ in the atomistic region, $\mu(x) = \mu_{CG}$ in the coarse-grained region and $\mu(x) = \mu_\Delta(x) = \mu(w(x))$ in the transition region. Next, taking the atomistic resolution as reference, it would be sufficient to balance the chemical potential in each region, that is, one can impose that: $\mu^{eff} = \mu_{Atom}$, (μ^{eff} is the effective chemical potential that characterizes the entire system) independently from the resolution and thus of the position, then we would have equilibrium and uniform density all over the simulation box. This target can be obviously accomplished if one imposes a force proportional to the gradient of the difference between the reference chemical potential and the space dependent chemical potential: $\mathbf{F}(x_\Delta) \approx \nabla[\mu_{Atom} - \mu(x)]$. Such a force should act on the center of mass of the molecules in Δ, and automatically vanishes in both atomistic and coarse-grained regions. While the calculation of μ_{Atom} and μ_{CG} can be done in simulation via thermodynamic integration or Widom, *insertion particle* approach[19,20] by considering a full atomistic system and a full coarse-grained system, the question arises on how to calculate $\mu(w(x))$ in Δ, in fact such techniques require the definition of a Hamiltonian. To this aim, a scheme was developed where Δ was ideally divided in several small regions each characterized by a constant value of w. Then each constant value of w was considered separately, that is, an auxiliary simulation was performed where all the molecules in the simulation box interact with an AdResS force with w fixed to a given value (not anymore space dependent). In such a set up it is possible to define a Hamiltonian (corresponding to the linear interpolation via the constant w of atomistic and coarse-grained potential) and thus calculate the chemical potential, $\mu(\hat{w})$ corresponding to the specific value of $w = \hat{w}$. Since each constant value of w characterizes a small region of Δ, $\mu(w)$, calculated in

the way described before, is an approximate, though numerically accurate, way to determine $\mu(x)$ in Δ. Next, one can take a force proportional to the gradient, $\mathbf{F}_{th} \approx \nabla[\mu_{Atom} - \mu(x)]$, called then *"thermodynamic force"*, apply it to the center of mass of the molecules and, as a consequence, a uniform molecular density should be obtained while preserving all other thermodynamic and structural properties (temperature, pressure, radial distribution functions, solvation structures, etc. etc.). This idea was proved to be successful not only for standard liquids, but also for liquid mixtures[21]. The key point is that there was no need to explicitly define a global (most probably artificial) Hamiltonian but yet the basic properties were accurately reproduced employing conclusions based on thermodynamics only. Despite this study provided a deeper understanding of how the process of adaptivity can be interpreted on the basis of standard thermodynamics, from the practical point of view, the calculation of the thermodynamic force required an excessive number of extra simulations. As a consequence this approach, conceptually satisfactory, was not optimal from the practical point of view, however it would stimulate further conceptual developments, as reported below.

3. Thermodynamics and Statistical mechanics of a generic open system

The approach employed above, requires that the atomistic model and the coarse-grained model, at given density and temperature, are characterized by the same pressure. A coarse-grained potential, derived only by the inversion of the molecule-molecule radial distribution function, would preserve the compressibility, but not the pressure. For this reason, usually a pressure correction term is added to the coarse-grained model at the price of loosing the correct compressibility[22]. However, starting from the basic idea of the previous section, that is obtaining a uniform density also in the transition region, one may think of a generalized principle that imposes a uniform density profile across the simulation box without correcting the pressure of the coarse-grained model. Said in these terms, this is the intuitive/practical idea, however one can formalize it on the basis of principle of thermodynamic and statistical mechanics. In fact, the adaptive set up consists of interfaced open systems; from the thermodynamic and statistical mechanics point of view, the key quantity of interest is the grand potential, $\Omega = PV$. The volume of each subsystem is a fixed quantity, as a consequence the equilibrium between the atomistic and the coarse-grained region is given by the equality of pressure. Thus in the interface region, Δ, conditions should be defined such that $p_{Atom} = p_{CG}$ is assured; this was done via a reformulation of the thermodynamic force in terms of balancing the pressure[2]:

$$p_{Atom} + \rho_0 \int_\Delta \mathbf{F}_{th}(\mathbf{r})d\mathbf{r} = p_{CG}, \tag{2}$$

where p_{Atom} is the target pressure of the atomistic system (region), p_{CG} is the pressure of the coarse-grained model, ρ_0 is the target molecular density of the atomistic system (region). In the MD set up, $\mathbf{F}_{th}$ is calculated via an iterative procedure as a function of the molecular density in Δ. The iterative formulas at the i-th step is

$$\mathbf{F}_{th}^{i+1}(\mathbf{r}) = \mathbf{F}_{th}^i(\mathbf{r}) - \frac{1}{\rho_0 \kappa_T^2} \nabla \rho^i(\mathbf{r}), \tag{3}$$

where κ_T the isothermal compressibility. The force is considered converged when the target density ρ_0 in Δ is obtained. In principle, this derivation implies a Grand-Canonical-like set up, in which each region acts as a reservoir for the other and the transition region as a filter that assures the correct entrance/exit of molecules from one region to the other. The approach as it is, cannot be defined Grand-Canonical because of the finiteness of the reservoir(s) and because one cannot calculate directly the chemical potential, μ; this can be only calculated in a separate way at high computational costs, as underlined in the previous section. Nevertheless this approach turned out to be satisfactory from both practical and conceptual point of view[2,23,24], and represented the basis for a further development toward a proper Grand-Canonical-like set up[8], this is reported in the next section.

4. Grand-Canonical-like formulation and its mathematical necessary conditions

The approach reported above leads, in a natural way, to the question of how much AdResS can be formulated in terms of Grand-Canonical ensemble and thus obtain a statistical mechanics framework that justifies, from the conceptual point of view, its computational robustness. For this purpose we must address the key question whether or not the probability distribution of the atomistic phase space variables $\boldsymbol{x}_{\text{Atom}}$ and the number of molecules in the atomistic region N_{Atom} are subject to the Grand-Canonical distribution. This implies the possibility of writing and numerically verifying the following formula:

$$p(\boldsymbol{x}_{\text{Atom}}, N_{\text{Atom}}) = \frac{1}{\mathcal{Z}_{\text{Atom}}} e^{\beta \mu_{\text{Atom}} N_{\text{Atom}} - \beta \mathcal{H}_{N_{\text{Atom}}}^{\text{Atom}}(\boldsymbol{x}_{\text{Atom}})} \tag{4}$$

Here $\mathcal{Z}_{\text{Atom}}$ is the grand-canonical partition function and the $\mathcal{H}^{\text{Atom}}$ is the atomistic Hamiltonian. In order to properly set the question, we have to consider the case of thermodynamic limit, where the atomistic region is large enough in order to have a statistical meaning, the transition region works as a relatively thin filter and the coarse-grained region acts as a large reservoir; this assumes $N_{\text{CG}} \gg 1$, $N_{\text{Atom}} \gg 1$, $N_{\text{CG}} \gg N_{\text{Atom}} \gg N_\Delta$. However, in practice, surprisingly, the method is characterized by high accuracy even in the "worst case scenario", in which the atomistic and CG regions can be even smaller than the hybrid region[3]. In this limiting process, as anticipated above, the size of the hybrid region is kept constant, therefore, it can be viewed as a thin filter that allows the change of resolution when a molecule freely goes from a high resolution region to a low resolution region and vice versa. In the initial condition set up of the adaptive simulation, different resolutions (regions) are at different thermodynamic state (different pressure), therefore, to assure thermodynamic equilibrium of the whole system (to be the same of the target full atomistic simulation), we can require that the hybrid region performs some works (denoted by ω_0 in this chapter) when a molecule goes through the hybrid region. In this section we discuss how such a work should be defined and what is its physical meaning in order to reproduce the target probability density in the atomistic region . Regarding the "accuracy" of AdResS, we compare the AdResS distribution with that of a full-atomistic simulation. It is clear that in the thermodynamic limit, a subsystem of a full atomistic system, that is of the same size as the atomistic region in AdResS, is a natural Grand Canonical ensemble. Therefore the full-atomistic reference system serves a good measure of the accuracy of AdResS as a Grand-Canonical MD set up.

A direct approach to an analysis of the distribution function (4) in an adaptive set up is rather complex, thus we will circumvent the problem through different steps the first of which consists in splitting the distribution as:

$$p(\boldsymbol{x}_{\text{Atom}}, N_{\text{Atom}}) = p(\boldsymbol{x}_{\text{Atom}}|N_{\text{Atom}})\, p(N_{\text{Atom}}) \tag{5}$$

As initial step we consider the first term of the R.H.S. of Eq. (5):

$$p(\boldsymbol{x}_{\text{Atom}}|N_{\text{Atom}}) = \sum_{N_\Delta} \int \mathrm{d}\boldsymbol{x}_\Delta\, p(\boldsymbol{x}_{\text{Atom}}|N_{\text{Atom}}; \boldsymbol{x}_\Delta, N_\Delta) \cdot p(\boldsymbol{x}_\Delta, N_\Delta|N_{\text{Atom}}) \tag{6}$$

The conditional distribution is a Boltzmann distribution

$$p(\boldsymbol{x}_{\text{Atom}}|N_{\text{Atom}}; \boldsymbol{x}_\Delta, N_\Delta) \propto \exp\{-\beta \mathcal{H}^{\text{Atom}}_{N_{\text{Atom}}}(\boldsymbol{x}_{\text{Atom}}; \boldsymbol{x}_\Delta, N_\Delta)\} \tag{7}$$

where

$$\mathcal{H}^{\text{Atom}}_{N_{\text{Atom}}}(\boldsymbol{x}_{\text{Atom}}; \boldsymbol{x}_\Delta, N_\Delta) = \sum_{i=1}^{N_{\text{Atom}}} \frac{1}{2}m_i \boldsymbol{v}_i^2 + \sum_{i<j}^{N_{\text{Atom}}} U^{\text{Atom}}(\boldsymbol{r}_i - \boldsymbol{r}_j) + \sum_{i=1}^{N_{\text{Atom}}} \sum_{j=N_{\text{Atom}}+1}^{N_{\text{Atom}}+N_\Delta} U^{\text{Atom}}(\boldsymbol{r}_i - \boldsymbol{r}_j). \tag{8}$$

In order to properly define an exact $\mathcal{H}^{\text{Atom}}_{N_{\text{Atom}}}$ without any artifact, we propose a technical modification of the transition region and of the corresponding definition of the switching function $w(x)$. In[8], the switching function was redefined in such a way that it allows the atomistic interactions to be separated from hybrid and coarse-grained interactions. This was done by extending the transition region into the original atomistic region for an amount corresponding to the cutoff of the molecule-molecule interaction (see also Fig. 2). As a consequence all the molecules of the (new) atomistic region interact with the rest of the system only in an atomistic way. The price to pay when extending the transition region is balanced by a significative conceptual advancement: now one can write an exact Hamiltonian for the atomistic region (see Eq.8), without any unphysical quantity involved, and consider the rest of the system as a bath/reservoir of particles. Eq. (7), is a valid approximation because the DOFs in the hybrid region, $\boldsymbol{x}_\Delta, N_\Delta$, can be viewed in terms of frozen configuration frames which, although cannot be treated as deterministic dynamical sequences within AdResS, can be treated in terms of statistical distribution of configurations (see the pictorial explanation of Fig.3). Therefore, according to Eq. (6), the necessary condition for having high accuracy for the distribution $p(\boldsymbol{x}_{\text{Atom}}|N_{\text{Atom}})$ implies high accuracy of the distribution $p(\boldsymbol{x}_\Delta, N_\Delta|N_{\text{Atom}})$. Ideally, it must be the same distribution of the corresponding subregion of the full-atomistic reference system. In the standard conditions, this cannot be true, however, one can search for (at least) necessary conditions which assure a satisfactory accuracy of $p(\boldsymbol{x}_\Delta, N_\Delta|N_{\text{Atom}})$. The condition, $\rho_\Delta = \rho_{\text{Atom}}$, assures a first order accuracy, while the second order is assured by the condition $g_\Delta(r) = g_{\text{Atom}}(r)$. In a similar way, higher orders of accuracy can be achieved by imposing that 3-body correlation, 4-body correlation functions and so on, are reproduced in Δ, however, in MD, the second order is usually already satisfactory. The condition $\rho_\Delta = \rho_{\text{Atom}}$ is fulfilled by the thermodynamic force introduced in the previous section, while the condition $g_\Delta(r) = g_{\text{Atom}}(r)$ can be fulfilled by a force which corrects the RDF in Δ to the target one.[8]. The new expression of the AdResS force acting on a generic molecule α becomes:

$$\mathbf{F}_\alpha = \sum_{\beta \neq \alpha} \left[w_\alpha w_\beta \mathbf{F}^{Atom}_{\alpha\beta} + (1 - w_\alpha w_\beta)\mathbf{F}^{CG}_{\alpha\beta} + w_\alpha w_\beta(1 - w_\alpha w_\beta)\mathbf{F}^{rdf}_{\alpha\beta} \right] + \mathbf{F}^{th}_\alpha \tag{9}$$
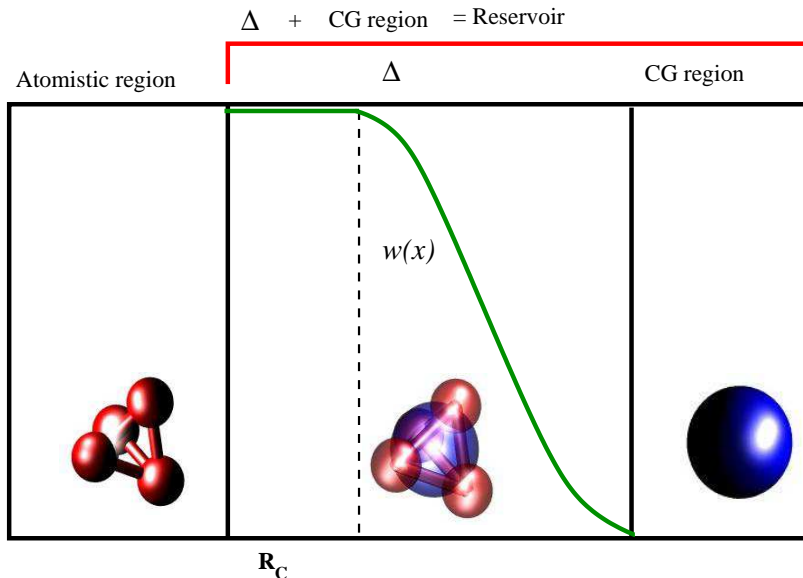
FIG. 2. Pictorial representation of the new adaptive box. $w(x)$ is extended into the atomistic region of an amount equal to R_c, the cut off distance of the atomistic interaction. This enables us to write an exact Hamiltonian for the atomistic region and identify the rest of the system as a generic reservoir.

where, $w_\alpha w_\beta (1 - w_\alpha w_\beta) \mathbf{F}^{rdf}_{\alpha\beta}$, is the corrective force for the radial distribution function, multiplied by $w_\alpha w_\beta (1 - w_\alpha w_\beta)$ so that it acts only in Δ. The force $\mathbf{F}^{rdf}_{\alpha\beta}$ is derived via an Iterative Boltzmann Inversion (IBI)[14,15], and in this case consists of:

$$U^{rdf}_{i+1} = U^{rdf}_i + k_B T \ln \left[\frac{g(r)_i}{g(r)_{Atom}} \right] \tag{10}$$

with $U^{rdf}(r)$ being the pairwise correction potential applied to a molecule in Δ, $g(r)_{Atom}$ is the target radial distribution function (full atomistic simulation) and $g(r)_i$ is the *average* radial distribution function at the $i - th$ step of iteration in Δ. With expression *average* here is meant that the $g(r)$ usually depends on the relative position of two molecules, $r = \mathbf{r}_{\alpha\beta}$, however the correction on the $g(r)$ in Δ in an adaptive scheme, is position dependent, that is some points in Δ need more/less correction that others. For this reason, in first approximation, we assume that the correction does not have such a dependence in Δ and we take the average over Δ so that the correction function is a function of $\mathbf{r}_{\alpha\beta}$ but uniform for all molecules in Δ. Next, we construct the prefactor, $w_\alpha w_\beta (1 - w_\alpha w_\beta)$, such that the spatial dependency of the correction in Δ is reproduced as much as possible, giving to the force the proper weight via $w(x)$. For example, in the region of Δ closer to the atomistic region, it is needed less correction than in the central region of Δ. However, since the application of $w_\alpha w_\beta (1 - w_\alpha w_\beta) \mathbf{F}^{rdf}_{\alpha\beta}$, may perturb the molecular number density in Δ, we perform another iteration for the thermodynamic force, and in turn iteratively correct $\mathbf{F}^{rdf}_{\alpha\beta}$ until the density and the radial distribution function in Δ converge both to the target ones. It is possible to derive also necessary conditions for the accuracy of the second term of the R.H.S. of Eq. (5), i.e. $p(N_{\mathrm{Atom}})$. The necessary condition in order to have a first order accuracy is: $\omega_0 = \mu_{\mathrm{CG}} - \mu_{\mathrm{Atom}}$, that is the work of in Δ should equalize the difference in chemical potential between the atomistic and the CG resolution. Instead, the necessary condition for the accuracy at second order is: $\kappa_{\mathrm{Atom}} = \kappa_{\mathrm{CG}}$, that is the equality of the compressibility in the two resolutions. Here the first order accuracy means that the deviation of $p(N_{\mathrm{Atom}})$ from the full-atomistic reference is controlled by $\mathrm{Const} \cdot N_{\mathrm{Atom}}/N_{\mathrm{CG}}$, while the second order accuracy means the deviation is controlled by $\mathrm{Const} \cdot (N_{\mathrm{Atom}}/N_{\mathrm{CG}})^2$. Under the thermodynamic limit, the deviation vanishes and the second order decays faster than the first order. It can be proved that when the thermodynamic force is imposed to the system, the first order necessary condition $\omega_0 = \mu_{\mathrm{CG}} - \mu_{\mathrm{Atom}}$ is satisfied [3] (see also next section for an extended discussion). The second order accuracy is achieved only when the coarse-grained model reproduces the atomistic compressibility.
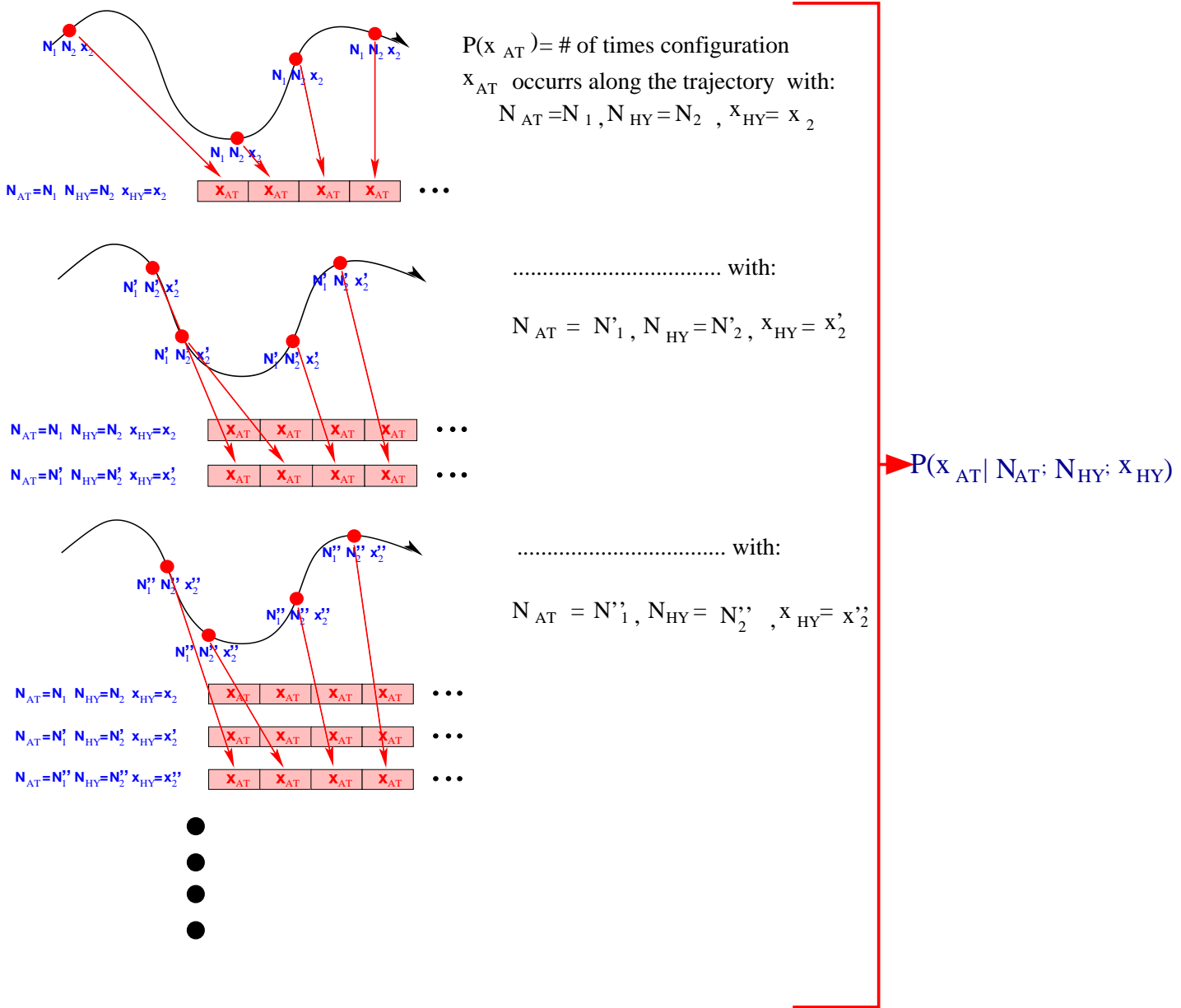
$P(x_{AT}) = $ # of times configuration x_{AT} occurrs along the trajectory with:

$$N_{AT} = N_1, N_{HY} = N_2, x_{HY} = x_2$$

.... with:

$$N_{AT} = N'_1, N_{HY} = N'_2, x_{HY} = x'_2$$

.... with:

$$N_{AT} = N''_1, N_{HY} = N''_2, x_{HY} = x''_2$$

$$P(x_{AT} | N_{AT}; N_{HY}; x_{HY})$$

FIG. 3. Pictorial representation of the "conceptual" process of how to derive $P(x_{Atom}|N_{Atom}; N_\Delta; x_\Delta)$ from a generic trajectory of AdResS. First, one ideally fixes N_{Atom} (here indicated by N_{AT}), N_Δ (here indicated by N_{HY}) and x_Δ (here indicated by x_{HY}), and counts the number of times a given configuration x_{atom} (here indicated by x_{AT}) occurs at under this condition. This is (ideally) repeated for all x_{Atom}. Let us suppose that the parameters chosen at this first step are: $N_{AT} = N_1; N_{HY} = N_2; x_{HY} = x_2$, then the process of sorting out of configurations under this condition leads to $P(x_{Atom}|N_1; N_2; x_2)$. At next step the parameters chosen are: $N_{AT} = N'_1; N_{HY} = N'_2; x_{HY} = x'_2$ and the sorting out process for all x_{AT} is repeated and one obtains: $P(x_{Atom}|N'_1; N'_2; x'_2)$. Next, a new sets of parameters are defined: $N_{AT} = N''_1; N_{HY} = N''_2; x_{HY} = x''_2$, and one obtains: $P(x_{Atom}|N''_1; N''_2; x''_2)$. When the process is iterated at infinity, the collection of $P(x_{Atom}|N^i_1; N^i_2; x^i_2)$ (with i generic step), corresponds to: $P(x_{Atom}|N_{Atom}; N_\Delta; x_\Delta)$. If the AdResS trajectory is, in principle, infinitely long, one can imagine that for each choice of the parameters $N_{AT}; N_{HY}; x_{HY}$ the statistical sorting out is equivalent to run a simulation with a Hamiltonian where $N_{AT}; N_{HY}; x_{HY}$ are fixed, and thus such probability has a Boltzmann form and as a consequence this holds also for $P(x_{Atom}|N_{Atom}; N_\Delta; x_\Delta)$.

B.   Essential quantity in the Grand Canonical ensemble: Automatic calculation of μ

An essential thermodynamic quantity in the Grand Canonical ensemble is the chemical potential μ. If the framework reported above is properly defined in a Grand Canonical ensemble, then we should be able to determine μ directly. In this section we outline the procedure by which μ can be calculated within AdResS.

In Ref.[3] it was shown that:

$$\mu_{CG} = \mu_{Atom} + \omega_{th} + \omega_Q \tag{11}$$

μ_{CG} is the chemical potential of the coarse-grained system, μ_{Atom} is the chemical potential of the atomistic system and $\omega_{th} = \int_\Delta \mathbf{F}_{th}(\mathbf{r})d\mathbf{r}$ is work performed by the thermodynamic force in Δ, finally, ω_Q is the heat given by the thermostat by which the inserted/removed degrees of freedom in Δ are slowly reintroduced/removed. ω_Q consists of two parts, one, ω_{extra}, which takes care of the dissipation of energy caused by the change of interactions in Δ; the other, ω_{DOF}, is related to the equilibration of the reinserted/removed degrees of freedom (e.g. rotational and vibrational). For the equipartition theorem, $\omega_{DOF} = \frac{1}{2}k_B T$ per degree of freedom. ω_{th} can be calculated by integrating the thermodynamic force over Δ. Instead, the calculation of ω_Q is not direct and in Ref.[3] we have proposed a procedure. The procedure consists of introducing an auxiliary Hamiltonian where the coarse-grained and atomistic potential are now interpolated, instead of forces as in standard AdResS. Next, we impose that the Hamiltonian system has the same thermodynamic equilibrium of the original system considered in a simulation of standard AdResS; this is done by introducing a thermodynamic force which, at the target temperature, assures that the density of particles across the system is the same as in the original system. In this way, in the Hamiltonian approach we obtain the same equilibrium of the original adaptive system and in addition a thermostat is not needed; this implies that the difference between the work of the original thermodynamic force and the work of the thermodynamic force calculated in the Hamiltonian approach gives ω_Q. We have also proven numerically, for liquid water, that $\omega_Q = \int_\Delta \langle w\nabla_x w(x)(V_{AT} - V_{CG})\rangle_\mathbf{r} d\mathbf{r}$, with V_{AT} and V_{CG} being the atomistic and coarse-grained potential. The result above means that ω_Q can be calculated in a straightforward way during the standard AdResS simulation. According to (11), if one knows μ_{CG}, then AdResS automatically gives μ_{Atom}. In MD usually the quantity of interest is not the total chemical potential, but the excess chemical potential μ_{AT}^{ex}, that is, the expression of (11) where the kinetic (ideal gas) part is removed. Regarding the kinetic part, the contribution coming from the center of mass for the coarse-grained and for the atomistic molecules is exactly the same and it is automatically removed in (11). The kinetic part of μ_{Atom}, due to, e.g., the rotational and vibrational degrees of freedom, ω_{DOF}, can be calculated by hand removing $\frac{1}{2}k_B T$ per degree of freedom. However, this calculation is actually not required in the current code, since the technical set up of AdResS considers the removed degrees of freedom as phantom variables but nevertheless equilibrate them in terms of thermal energy[6]. As a consequence the heat given by the thermostat for this part is the same in the atomistic and coarse-graining molecules and it is automatically removed in the difference. Finally, the calculation of μ_{CG}^{ex} is done with standard methods as Widom insertion particle[20] which, for simple coarse-grained models (e.g. spherical molecules) like those of the coarse-grained system, requires a small computational cost. The final expression writes:

$$\mu_{Atom}^{ex} = \mu_{CG}^{ex} - \int_\Delta F_{th}(\mathbf{r})d\mathbf{r} - \int_\Delta \langle w\nabla_x w(x)(V_{AT} - V_{CG})\rangle_\mathbf{r} d\mathbf{r} \tag{12}$$

This procedure has been successfully applied to various liquids and mixtures and shows the capability of AdResS to be a general tool for the calculation of thermodynamic properties of a system in a Grand Canonical fashion; for this reason we have renamed this formulation of AdResS as Grand Canonical AdResS (GC-AdResS), however in this paper we will use the name AdResS for any formulation of the method.

C.   What about a Hamiltonian approach?

The strongest criticism to AdResS is not about the principles on which is based, or its computational robustness and performance, but about the fact that the force of Eq.1 is not conservative, and thus the system does not have a global Hamiltonian. Other authors claim that the force of Eq.1 can be actually derived from a Hamiltonian where instead of the forces, the atomistic and coarse-grained potential are interpolated in the same technical fashion of AdResS[25–27]. However it has been shown both analytically[17] and numerically[28] that within this scheme, there is no possibility of deriving Eq.1 from a potential as that suggested in[25–27]. The lack of a global Hamiltonian leads to a deep cultural problem: the community of molecular simulation is used to treat systems (in equilibrium) with a well defined Hamiltonian and thus, unfortunately, is rather skeptical regarding procedures where the Hamiltonian is not used, despite they are based (as for AdResS) on well defined first principles of thermodynamics and statistical mechanics. In fact, after the appearance of AdResS in literature, a rush of attempts have been made for defining

and adaptive Hamiltonian approach. The one of Ensing *et al.* cited above, was essentially equivalent to AdResS but with an inexact Hamiltonian interpretation. A second attempt by the same group is based instead on a Lagrangian approach where the number of particles at high resolution is an explicit part of the equation associated to a Lagrangian multiplier[29]. Essentially this multiplier is equivalent to a chemical potential which regulates how many particles can be at high resolution, that is, in practice, something equivalent to the thermodynamic force of AdResS. It follows the question whether GC-AdResS can be written in terms of such a Lagrangian. At this stage we cannot give a final answer, however one must point out the following: the Lagrangian equation of[29] refers to the description of two open systems (atomistic and coarse-grained) and thus if the term related to the varying number of particles is properly defined, it should imply the existence of a solution for the Liouville equation with varying number of particles (in the atomistic region). Unfortunately, at the current state of the art, the question of the formulation of Liouville equation with varying number of particles is an open problem and actually an exact formulation has not been found yet[30]. It follows that the Lagrangian of[29], if not associated to a Liouville equation, may be artificial and would give correct results because of an *a priori* imposition of what one wants, rather than because of the natural result of a first-principles Lagrangian. Another attempt, to use the Lagrangian approach was developed by Heyden and Truhlar; the idea is based on dividing the space in three regions, as in AdResS, and the region at high resolution is usually centered around a fixed active site; a total Lagrangian is then written as a linear combination of Lagrangians, each corresponding to a possible combination of molecular resolutions and weighted according to the distance from the active center. This means that given a molecule A in the transition region at a distance r from the active center, interacting, for example, with molecules B, C, D, (also in the transition region) the corresponding Lagrangian is a linear combination of those Lagrangians with all possible combination of resolutions (all molecules have coarse-grained resolution, all have atomistic resolution, A is atomistic and the others are coarse-grained, A is coarse-grained and the others atomistic, A is atomistic, B is atomistic and C and D coarse-grained etc etc). Each of the terms of this combination is weighted by a function of the distance from the active site, the closer to the atomistic region, the higher the weight to the Lagrangians with more atomistic resolution and vice versa for the coarse-grained. Formally, the result is that one has a Hamiltonian approach and in average certainly takes into account all possible scenarios of transitions from atomistic to coarse-grained and vice versa. However, while formally elegant and robust, it is evident that practically is not feasible, in fact even for a small system of, let us suppose 50 molecules in the transition region, the amount of combinations simply explodes and cannot be calculated. In general, if a Hamiltonian approach exists, it is mandatory that satisfies the following necessary requirements:

- It should lead to the atomistic Hamiltonian in the atomistic region and the coarse-grained Hamiltonian in the coarse grained region

- It should automatically produce, at least in the atomistic region, a spatial probability distribution as that of the target full atomistic simulation.

Recently, Potestio *et al.* suggested a global Hamiltonian approach of AdResS[31]. As for the method of Refs.[25–27], this approach is based on the interpolation in space of the atomistic and coarse-grained potential. The procedure provides an elegant thermodynamic interpretation in terms of Kirkwood's (adiabatic) integration and suggests how to equilibrate the system. The problem is that similarly to the approach of[25–27], this method by construction, cannot satisfy the first requirement of the necessary conditions above, according to[17]. In fact, the extra force generated by the gradient of the weighting function, $w(x)$ remains the key problem. Such a force induces an unphysical flux of particles from one region to another, the balance of this flux is taken care by an additional field which must be added to the original Hamiltonian, such a field, in[31], is derived by an elegant thermodynamic procedure. However, in[17] it is shown that such a field (if properly defined) must be a solution of a first order partial differential equation. In order to describe a proper adaptive system, it is required that the field satisfies two boundary conditions; that is, it must be zero in the atomistic and in the coarse-grained region. However, the equation is of the first order and only one boundary condition can be used to fix the solution. For example, let us assume that we fix the boundary condition in the atomistic region; this implies, inevitably, that in the coarse-grained region the original potential is changed by an artificial, unphysical additional term. The consequence is that if we, ideally, reduce the atomistic region to zero, the original coarse-grained potential is not obtained (as instead should be). The additional term is a constant in space, but depends on the data of the specific system. In most of the cases, from the practical point of view, this is a non-relevant aspect, however, from the conceptual point of view, this makes the Hamiltonian, artificial. There is anyway a more relevant aspect; according to[3], as it stands now, the method of[31] can assure (*a priori*) that in the atomistic region (compared to a full atomistic target data), only the first order of the probability distribution, i.e., the molecular number density, can be obtained with high accuracy. Higher orders of the distribution, atom-atom (or even molecule-molecule) radial distribution functions or three-body correlations, do not come automatically. Other relevant limitations are the violation of Newton's' third law and the fact that technically the method cannot avoid the integration of the atomistic phantom DOF's in the coarse-grained region. In fact such DOF's are part of the

Hamiltonian, which means that in this method such DOF's are not phantom and that the coarse-grained zone is necessarily not defined by the coarse-grained representation only. As it will be discussed later, the integration of the phantom DOF's in standard AdResS can be avoided and this would lead to a better performance of the code in terms of computational resources required. Nevertheless, the approach of[31] represents the closest (somehow "first order") procedure for a truly adaptive Hamiltonian procedure with the major, advantage of the possibility of performing adaptive Monte Carlo simulations[32]. A similar problem can be found in an approach recently proposed by Wagoner and Pande[33]. The method id based on the coupling of a full atomistic region to a continuum via an intermediate region where particles are inserted and removed via Monte Carlo moves biased by nonequilibrium paths. Interestingly, the high resolution region can change shape and adapt according to the evolution of the system. However, despite an analytic expression of the probability distribution of the system is reported, the only check done is about its validity at the first order, that is, at the level of the particle number density. This is not sufficient to prove the robustness of the method, it must be proven that higher orders can be reproduced. Moreover, the direct coupling to the continuum, differently from the adaptive methods above, based on the treatment of large reservoir of coarse-grained particles, implicitly implies that particle number fluctuations are biased and thus the thermodynamics is artificial and can be properly described only if it is imposed externally.

The opinion of the authors is that a global Hamiltonian approach is not really required if it cannot be defined in a rigorous way; where rigorous means that it must satisfy all the basic requirements of an adaptive system. Alternative procedures based on fundamental principles of thermodynamics and statistical mechanics, if properly described and tested, and, as long as advantages and limitations are both reported, should be more welcome than artificial Hamiltonians. However, as stated before, it is our opinion that the cultural barrier in the community of molecular simulation is, at the current stage, too high to allow for such developments. In this perspective we have written this chapter; we describe the principles, gives our motivations and report the limitations and open problems of our method.

III. RECENT APPLICATIONS

A. How to run AdResS in your computer: Technical Description

Before starting any AdResS simulation, the coarse-grained potentials have to be specified first. This is done in VOTCA simulation package[15] by using Iterative Boltzmann inversion[14] or Force matching technique. In the latest version of AdResS (GC-AdResS) the derivation of the coarse-grained potential can be avoided by specifying any spherical potential (for simple molecular liquids). To calculate the thermodynamic force and thus, the excess chemical potential defined in Eq. 12, we use the AdResS version of GROMACS[34] together with VOTCA. In the settings.xml file in VOTCA, the method is specified in the inverse section:

```
<method>tf</method>
```

For each interaction, some additional options have to be specified in the settings.xml file. The options are given in units of Gromacs, which means nm, kJ/mol, K and ps. To specify in which region the thermodynamic force should be nonzero, the 'min' and 'max' properties are used. A smoothing function proportional to $cos^2(r)$ is used to make the force go smoothly to zero at the region specified by 'min' and 'max'. Additionally a tf section is needed for each interaction type. Below is the section of setting.xml file from a urea/water simulation.

```
<non-bonded>
  <!-- name of the interaction -->
  <name>SOL</name>
  <!-- thermodynmic force: apply in this range -->
  <min>.45</min>
  <max>3.30</max>
  <step>0.01</step>
  <!--spline:offset beyond the hyrbid zone to take into account for fitting the spline-->
  <inverse>
    <!-- target distribution -->
    <target>dens.SOL.xvg</target>
    <!-- name of the table for gromacs -->
    <gromacs>
      <table>tabletf_CMW.xvg</table>
```

```
        </gromacs>
        <tf>
          <spline_start>.4</spline_start>
          <spline_end>3.35</spline_end>
          <spline_step>0.42142857142857142857</spline_step>
          <molname>SOL</molname>
          <prefactor>0.04</prefactor>
      </tf>
      </inverse>
    </non-bonded>


    <non-bonded>
      <!-- name of the interaction -->
      <name>UREA</name>
      <!-- thermodynmic force: apply in this range -->
      <min>.45</min>
      <max>3.30</max>
      <step>0.01</step>
      <!--spline:offset beyond the hyrbid zone to take into account for fitting the spline-->
      <inverse>
        <!-- target distribution -->
        <target>dens.UREA.xvg</target>
        <gromacs>
          <table>tabletf_CMC.xvg</table>
        </gromacs>
        <tf>
          <spline_start>.4</spline_start>
          <spline_end>3.35</spline_end>
          <spline_step>0.42142857142857142857</spline_step>
          <molname>UREA</molname>
          <prefactor>0.04</prefactor>
      </tf>
      </inverse>
    </non-bonded>
```

Spline interpolation is used to smooth the force as the density highly fluctuates. To specify the spline interpolation range the 'spline_start' and 'spline_end' parameters are used. These can define a larger region than that between min and max as it is sometimes useful to extend the spline fit for numerical stability. The parameter 'spline_step' specifies the bin width of the fit grid. The 'molname' specifies the molecule used for calculating the density. The prefactor $\frac{1}{\rho_o^2 \kappa_T^{at}}$ appearing in Eq. 3 is specified in the 'prefactor' field. It is seen that, if a very large prefactor is taken, then the thermodynamic force as well as the density profile in the transition region strongly fluctuates, instead, with a small prefactor, the time taken to obtain a converged thermodynamic force is large, thus few attempts must be done to find a reasonable compromise. In the urea/water example, a prefactor of 0.04 for urea, and a prefactor of 0.08 for water was found to be a good choice. A target density file has to be specified for each interaction type, in most cases this will contain a flat density profile at the equilibrium density ρ_o. In addition to these fields, there are some other parameters that are specified in the inverse section. For each component, an initial guess of the thermodynamic force has to be provided in files 'tabletf_CMC.xvg' (CMC is the name of coarse-grained urea site) and 'tabletf_CMW.xvg' (CMW is the name of coarse-grained water site). An efficient initial strategy is to add the thermodynamic forces obtained from the simulation of pure components. In case, it is not available, then an initial guess with zero thermodynamic force is the most suitable option. There are some other parameters that are written in the inverse section of the settings.xml file.

```
    <!-- general options for inverse script -->
  <inverse>
    <!-- 300*0.00831451, 300K in gromacs units -->
    <kBT>2.49435</kBT>
    <!-- use gromacs as simulation program -->
    <program>gromacs</program>
```

```
    <!-- gromacs specific options -->
    <gromacs>
      <!-- skip so many ps before calculation -->
      <equi_time>200</equi_time>
      <!-- grid for table*.xvg !-->
      <table_bins>0.01</table_bins>
      <!-- extend the tables to this value -->
      <table_end>5.83452</table_end>
    </gromacs>
    <!-- these files are copied for each new run -->
    <filelist>spc.adress.itp urea.adress.itp grompp.mdp topol.top table.xvg
      index.ndx table_CMW_CMW.xvg table_CMW_CMC.xvg table_CMC_CMC.xvg</filelist>
    <!-- do so many iterations -->
    <iterations_max>20</iterations_max>
    <!-- ibi: inverse biltzmann imc: inverse monte carlo tf: thermody force -->
    <method>tf</method>
  </inverse>
</cg>
```

This section consists of parameters that are specific to gromacs mdrun. The files 'table_CMW_CMW.xvg', 'table_CMW_CMC.xvg' and 'table_CMC_CMC.xvg' consist of tabular coarse-grained potentials acting between water-water, water-urea and urea-urea coarse-grained molecules. The 'iterations_max' command specifies the number of iterations required for thermodynamic force calculation. The starting configuration has to be modified by adding a virtual coarse-grained site to all the molecules. For example, for the urea/water system, the starting configuration would be the following:

```
    1  SOL   OW    1   8.107  0.355  2.902 -0.0176  0.0989 -0.2993
    1  SOL   HW1   2   8.039  0.405  2.848 -0.2112  0.1880  0.0180
    1  SOL   HW2   3   8.131  0.270  2.855 -0.7787 -0.2104 -0.1498
    1  SOL   CMW   4   8.105  0.353  2.896 -0.0710  0.0866 -0.2732
     .
     .
     .
    2550 UREA  C110442  3.061  0.817  2.791  0.4370  0.0045  0.4733
    2550 UREA  O210443  3.019  0.936  2.780  0.0212 -0.1328  0.5610
    2550 UREA  N310444  3.001  0.706  2.741 -0.3024  0.4128  0.4390
    2550 UREA  H410445  3.039  0.614  2.736 -1.0670  0.1461 -2.0020
    2550 UREA  H510446  2.916  0.715  2.688  0.7484 -1.8617 -1.8763
    2550 UREA  N610447  3.182  0.798  2.847  0.7348 -0.0476 -0.1709
    2550 UREA  H710448  3.217  0.705  2.857  0.1912 -0.4974 -2.0263
    2550 UREA  H810449  3.244  0.874  2.865  1.5616 -0.0895 -2.5464
    2550 UREA  CMC10450 3.067  0.812  2.789  0.2179  0.0120  0.1647
```

The virtual sites (CMC and CMW) are created at the center of mass of the molecule. This requires the topology files (topol.top) to be modified by adding a 'virtual_sites' field.

```
    [ atoms ]
;   nr    type    resnr   residu   atom    cgnr   charge   mass
     1    CUrea      1     UREA      C1      1     0.142   11.999
     2    OUrea      1     UREA      O2      1    -0.390   15.999
     3    NUrea      1     UREA      N3      1    -0.542   13.999
     4     H         1     UREA      H4      1     0.333    1.008
     5     H         1     UREA      H5      1     0.333    1.008
     6    NUrea      1     UREA      N6      1    -0.542   13.999
     7     H         1     UREA      H7      1     0.333    1.008
     8     H         1     UREA      H8      1     0.333    1.008
     9    CMC        1     UREA      CMC     2     0.000
```

```
   [ virtual_sitesn ]
   ; Site funct atom indexes
      9     2      1 2 3 4 5 6 7 8

   [ atoms ]
;  nr   type   resnr residue   atom   cgnr      charge      mass
    1     OW       1   SOL       OW       1      -0.82   15.99940
    2      H       1   SOL      HW1       1       0.41    1.00800
    3      H       1   SOL      HW2       1       0.41    1.00800
    4    CMW       1   SOL      CMW       2      0.0000

   [ virtual_sites3 ]
   ; Site from funct a d
    4 1 2 3 1 0.05595E+00 0.05595E+00
```

Also the 'atomtypes' and 'nonbon_params' have to be specified separately for the coarse-grained sites.

```
   [ atomtypes ]
   ;name  mass         charge    ptype  C6                C12             ; sigma      epsilon
    CMW    0.00000      0.0000     V       0.000000E+00    1.000000E+00
    CMC    0.00000      0.0000     V       0.000000E+00    1.000000E+00
   [ nonbond_params ]
   ; i   j   func c6              c12
    CMW  CMW  1    0.000000E+00   1.000000E+00
    CMW  CMC  1    0.000000E+00   1.000000E+00
    CMC  CMC  1    0.000000E+00   1.000000E+00
```

In the GROMACS parameters file grompp.mdp, there are some additional parameters that are required with respect to a standard atomistic simulation. For performing an AdResS simulation, 'adress=yes' has to be specified in the grompp.mdp file.

```
   ; AdResS parameters
   adress = yes
   adress_type              = xsplit
   adress_const_wf          = 1
   adress_ex_width          = 0.5
   adress_hy_width          = 2.75
   adress_ex_forcecap       = 0
   adress_interface_correction = thermoforce
   adress_site              = com
   adress_reference_coords  = 6.0 1.44344 1.44258
   adress_tf_grp_names      = CMW CMC
   adress_cg_grp_names      = CMW CMC
   adress_do_hybridpairs    = no
```

There are two ways, in which one can construct the explicit and transition region in AdResS: a spherical region ('sphere' in 'adress_type') or a rectangular region ('xsplit' in 'adress_type'). 'adress_ex_width' and 'adress_ex_width' specify the size of explicit region and transition region respectively. If the thermodynamic force is used for the pressure correction, then it has to be specified in 'adress_interface_correction' and 'adress_reference_coords' specifies the coordinates of the center of explicit region.

B. Liquids and Mixture

In this section, as an example, we report a study of different liquids and mixtures. In particular we focus on the calculation of the excess chemical potential, μ^{ex}. Such a quantity is of particular interest for understanding the capability of the system to act as a solvent, a more detailed discussion can be found in Ref.[35]. Table I shows the values obtained from AdResS simulation, along with the values obtained by using thermodynamic integration approach and

Liquid component	Mole fraction of solute	GC-AdResS	TI full atomistic	Experiment
methane	–	-4.6 ± 0.1	-5.2 ± 0.1	–
ethane	–	-8.2 ± 0.3	-8.8 ± 0.1	–
propane	–	-8.5 ± 0.1	-9.5 ± 0.2	–
methanol	–	-20.1 ± 0.1	-20.6 ± 0.4	-20.5[36]
DMSO	–	-32.2 ± 0.3	-34.7 ± 0.7	-32.2[38]
methanol in methanol/water mixture	0.01	-18.1 ± 0.2	-19.7 ± 0.2	–
methane in methane/water mixture	0.006	9.1 ± 0.1	8.5 ± 0.2	–
urea in urea/water mixture	0.02	-56.1 ± 0.6	-58.2 ± 0.5	-57.8 ± 2.5[39]
ethane in ethane/water mixture	0.006	7.2 ± 0.2	7.4 ± 0.3	–
TBA in water/TBA mixture	0.001	-19.5 ± 0.3	-20.8 ± 0.6	-19.0[37]
DMSO in DMSO/water mixture	0.01	-31.4 ± 0.5	-33.2 ± 0.3	–
TBA in TBA/DMSO mixture	0.02	-24.8 ± 0.4	-24.0 ± 0.5	–

TABLE I. The excess chemical potential of different liquids and mixtures calculated from GC-AdResS and thermodynamic integration (TI) of full atomistic simulations. The units for μ^{ex} are kJ/mol.

the values available in literature[36,37]. Results show that indeed the accuracy of AdResS for μ^{ex} is rather satisfactory; moreover it offers the advantage that, since the calculation of μ^{ex} corresponds to the equilibration of the system (see sectionII B), every time an AdResS is performed the chemical potential is automatically obtained. Fig.4 shows the density profiles for a urea/water mixture in the transition region. The action of the thermodynamic force and thermostat in the transition region, that is the essential ingredients of μ^{ex}, ensure that the density is sufficiently close to the reference full atomistic density. Due to the accuracy of the density in the transition region, the density in the atomistic region reproduces in a very accurate way the target one (not shown), in addition, in Fig.5 it is shown that the various radial distribution functions in the atomistic region in the urea/water system calculated from AdresS are in very good agreement with the target ones. The essential thermodynamic (μ^{ex}) and structural ($\rho; g(r)$) quantities are properly reproduced.

IV. PERSPECTIVES

A. Extension to (some) quantum problems: Path Integral AdResS

The adaptive coupling of quantum and classical resolution is a rather delicate question, in fact the problem is that we are interfacing not only systems with a different number of degrees of freedom, but above all the equations governing their evolution are different: deterministic on the classical side (Newton's equation), probabilistic on the other (Schrödinger equation), and thus the idea of adaptivity necessarily implies some sort of conceptual discontinuity. For electrons this is a rather difficult problem in fact it is not clear how to switch on and off an electron, which by definition is a non localized particle. One must also keep in mind that the electronic properties in the quantum region of the adaptive box, must be the same as if the whole system was treated at quantum level. Practical cases where the adaptive idea is employed in the context of electronic structure calculations do not fulfill the requirement above, since the main properties of interest are not the electronic ones. In fact in this case the electronic structure is used only as an on-the-fly good approximation to refine the classical force field[26,40,41]. A rigorous solution has been found for a class of quantum systems, namely those where light atoms are present. In fact light atoms, e.g. hydrogens (protons), even at room temperature display non negligible effects of spatial delocalization; such a delocalization can be described through the formalism of the path integral of Feynman[42]. While in the classical description two atoms are represented by two spheres interacting with a given potential, in the path integral formalism two atoms are mapped onto "effective" classical polymer rings and the interaction distributed among pairs of beads with the same index (bead one or polymer one with bead one of polymer two etc etc) (see Fig.6). The fluctuation of the shape of the polymer ring in space describes the quantum delocalization of the atom and the statistical properties of such systems can be properly described via classical molecular dynamics (see e.g.[43]). For the adaptive approach this is an ideal situation; in fact interfacing a quantum and a classical description is equivalent to interface two regions with a different number of "effective" classical degrees of freedom, and thus the computational apparatus developed so far can be applied straightforwardly.

Indeed this was shown to be the case for the liquid of tetrahedral molecules, where in the quantum region each atom of the molecule is represented by a polymer ring with 30 beads while in the coarse-grained region the whole molecule is represented as a spherical coarse-grained object (see Fig.7 and Ref.[44]). Moreover a further test was made for a
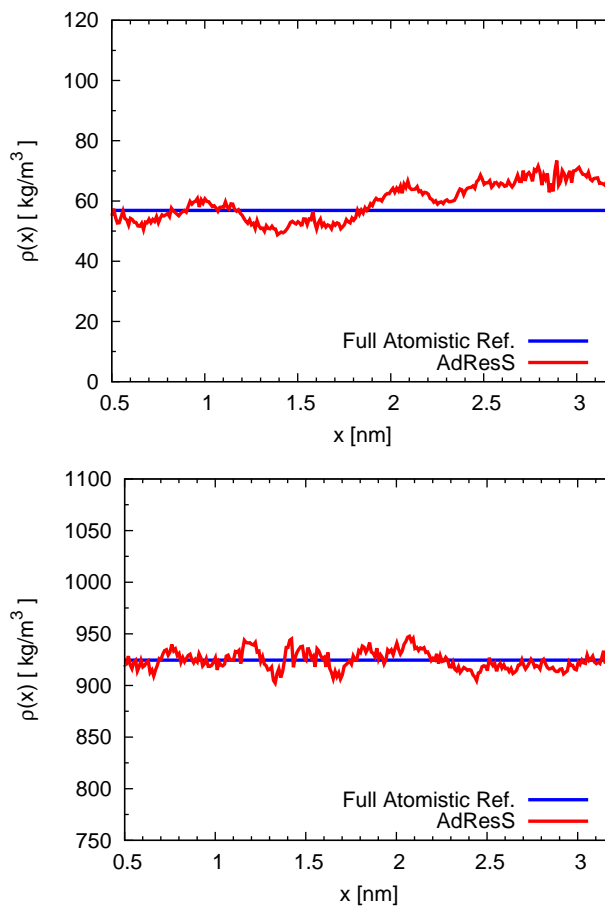
FIG. 4. Top: Molecular density profile in Δ for Urea/water mixture.

liquid of parahydrogen at extreme thermodynamic conditions where quantum effects play a major role and data from previous, full path integral, simulations were available as a reference. Also in this case the adaptive approach was able to reproduce the basic quantum properties of reference (see[45,46]). More recently we have tested the approach for liquid water at room conditions using a water model specifically designed for path integral simulations[47]. Fig.8 shows the various radial distribution functions calculated in the high resolution region of AdResS and compared with the equivalent quantities from a full path integral simulation: although one must consider such results at preliminary level, however they are definitively encouraging. Fritsch *et al.* have in the meanwhile developed a water model for path integral simulations which properly distinguishes between electronic and nuclear delocalization effects so that double counting of effects in the parameterization are avoided when the path integral approach is used; such model will be ideal for AdResS simulations[48] (for an extended discussion about path integral approach and AdResS see also[49]). What is the utility of such an approach for simulating biological molecules in solution?

Fig.9 pictorially shows the solvation of a biomolecule in water using AdResS with path integral. The flexibility of the hydrogen bond network may strongly depend on the spatial delocalization effects of the hydrogens which in turn may imply that the conformational properties of the molecule in solution can be very different from those obtained by standard atomistic simulation. The advantage of AdResS is that since it requires only a relatively small region at quantum level, the computational costs are not high and such simulations can be performed also by research groups which have got only standard (in house) computational resources; currently, full path integral simulation for such systems are possible only in presence of massive computational resources.

B.   On the dynamical properties of AdResS

An interesting question regarding AdResS is the accuracy of the method in determining dynamical properties in the atomistic region. In fact, due to the perturbation to the dynamics introduced by the transition region, trajectories
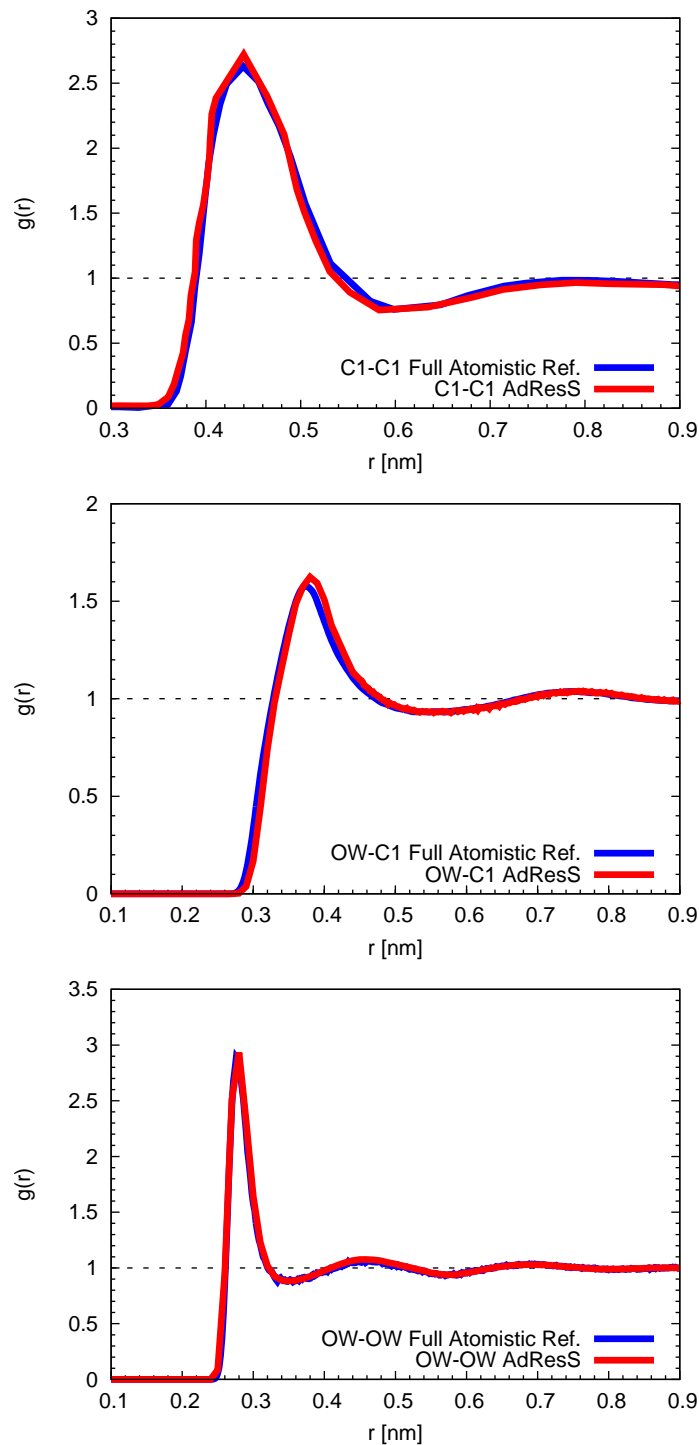
FIG. 5. Top: Urea-Urea radial distribution function (carbon-carbon); middle: the same plot for Urea-water (carbon-oxygen); bottom: for water-water (oxygen-oxygen).

within the atomistic region can be treated, strictly speaking, only as tools for statistical sampling of configurations, rather than for their intrinsic dynamical properties. However, one may expect that, at least for short time scales, dynamical processes should be correctly described. In order to test the validity of dynamics in the atomistic region, as a preliminary test, we have calculated the velocity auto-correlation function, $C(t) = \langle v(t) \cdot v(0) \rangle$ in an AdResS system (for more details, see Ref.[50]), and compared it with the the same quantity calculated in a full-atomistic system of reference. Technically, the full-atomistic system consists of 1728 SPC water molecules in a cubic box of length 3.724
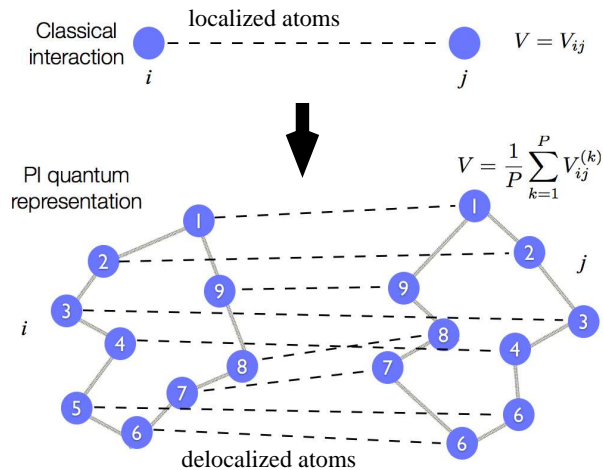
FIG. 6. Schematic representation of the mapping from a classical description of the interaction between atom i and atom j (through the potential V_{ij}) to the path integral/polymer ring description. P indicates the number of beads of the polymer, which in principle should tend to infinity. However for standard systems as been proved that 40 beads per atom is a satisfactory approximation.
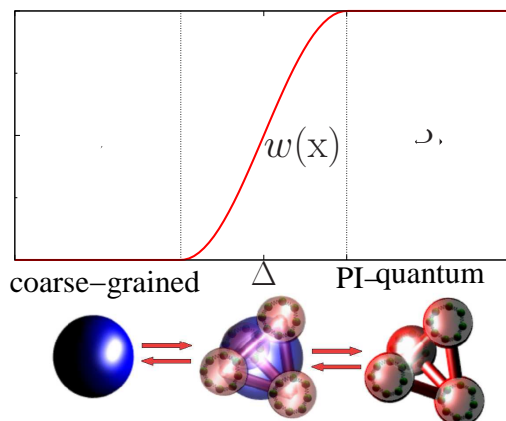

FIG. 7. Schematic representation of the adaptive resolution for the tetrahedral molecule which in our work is usually employed to test the adaptive principles. In the quantum region, (right) each atom is represented by a polymer ring of 30 beads, in the Δ region the molecule has hybrid coarse-grained/quantum resolution and in the coarse-grained region the molecule is represented by an effective spherical model obtained from a full path integral simulation.

nm. The time step is 2.0 fs and the neighbor list, which is built with a cut-off of 0.9 nm, is updated every 10 steps. We use reaction field with Coulomb cut-off 0.9 nm with dielectric constant of 80 in order to treat the electrostatics in the atomistic region. The simulation lasts for 2.0 ns, and the first 0.5 ns are removed. The AdResS simulation box is $7.529 \times 3.765 \times 3.765$ nm^3 containing 3456 water molecules. The system is split into atomistic, hybrid and coarse-grained resolution in x-direction. The size of the atomistic region is 1.0 nm and that of the hybrid region is 2.7 nm. The thermodynamics force is applied to the system to ensure a uniform density distribution in the system. All other running parameters in the AdResS simulation are the same as those in the full-atomistic simulation of reference. In the AdResS system, the velocity auto-correlation is calculated only when molecules are in the atomistic region. Results are reported in Fig. 10 where is shown a highly satisfactory consistency between the AdResS simulation and full-atomistic reference simulation. In this preliminary test we have considered only one system with a given size of the atomistic region, in future work we will perform a systematic study of $C(t)$ as a function of the size of the atomistic region[50]; this will allow us to conclude about the validity of a basic dynamical property as a function of the atomistic size and thus to suggest the limitation of AdResS in describing dynamics.
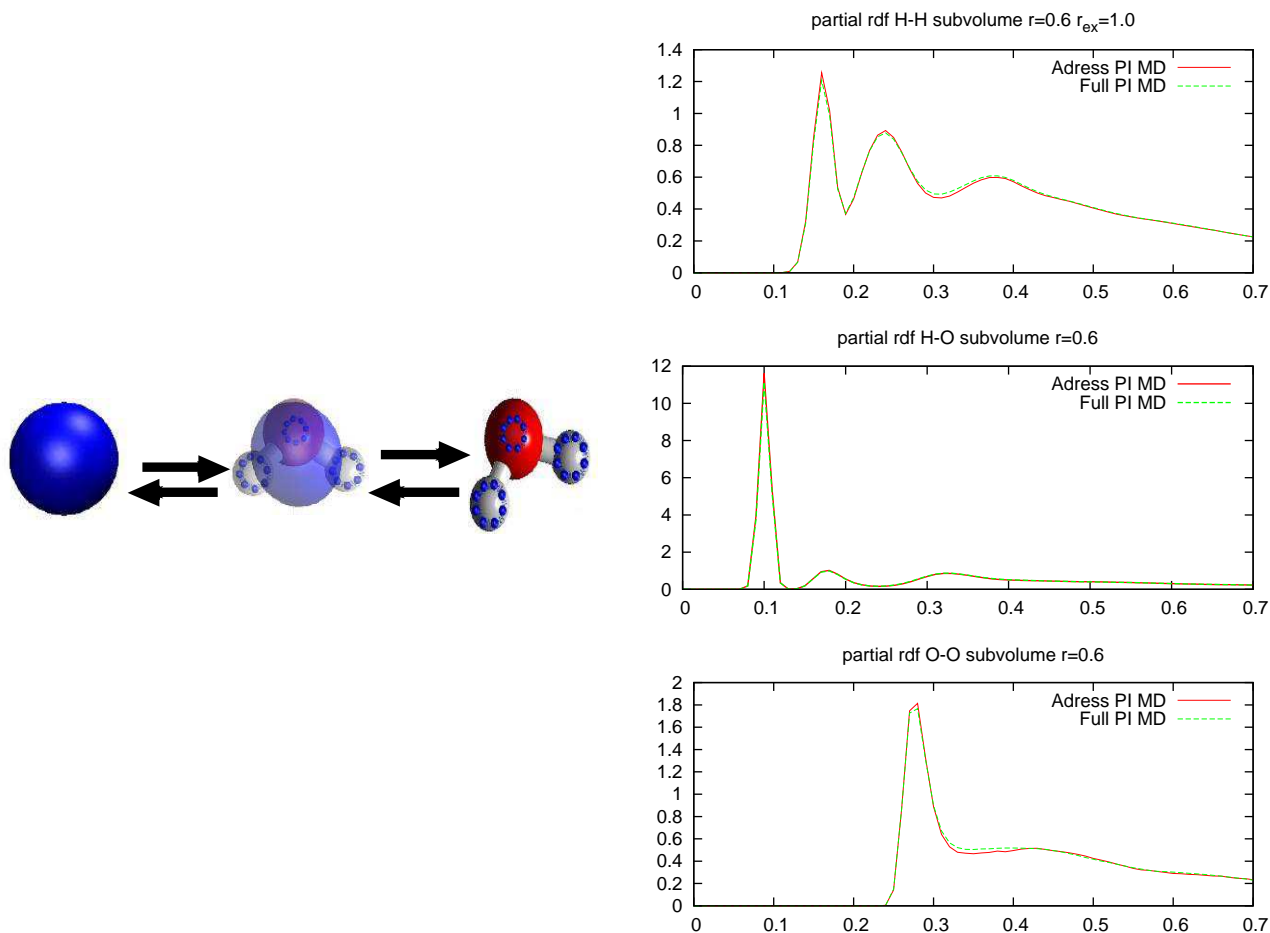
FIG. 8. Left: Schematic representation of the path integral adaptive resolution for a water molecule. In the quantum region each atom is represented by a polymer ring of 40 beads, in the transition region the molecule acquire hybrid coarse-grained/quantum resolution and in the coarse-grained region the molecule is represented by an effective spherical model obtained from the full path integral simulation. Right: radial distribution functions for hydrogen-hydrogen (top), oxygen-hydrogen (middle), oxygen-oxygen (bottom). The $g(r)$'s are calculated in the quantum region of AdResS and are compared with the corresponding function calculated in an equivalently large subregion of a full path integral simulation. Results show that AdResS satisfactorily reproduces the reference data.

C. Further Computational Optimization

AdResS has been implemented, so far, in different molecular dynamics codes, namely ESPResSo[6], ESPResSo++[7] and Gromacs[34]. All these implementations follow the "double representation trick"[6], which means the atomistic identities of the molecules are kept even in the coarse-grained zone. This approach leads to few computational advantages regarding the equilibration of degrees of freedom once the molecules enters into the hybrid zone. On the other hand, the force calculation, which is the computational most expensive part, is performed only on the physical relevant identities (degrees of freedom). The simplification of force calculation allows AdResS to provide a performance which, in terms of computational costs, is close to that of a pure coarse-grained simulation (for a small atomistic region). Moreover the atomistic as well as the coarse-grained zone of the simulation can be view as a standard MD simulation for which common performance tunings can be applied. For instance the Gromacs implementation[34] benefits from this fact and uses the highly optimized and vectorized force kernel in these two zones. Using the Gromacs version 4.5.1 a speedup of a factor 4 compared to full atomistic simulations has been reported for system of mixtures[23,24]. However in the successive version of GROMACS 4.6.1 the performance of atomistic simulations was highly improved while the corresponding implementation of AdResS is still not optimized. As a consequence, at the current state, AdResS can only assure a speed up between 2.0 and 2.5 for large systems (30000 molecules) compared to full atomistic simulations. In perspective, there is large room for improvements regarding the technical implementation in multiple
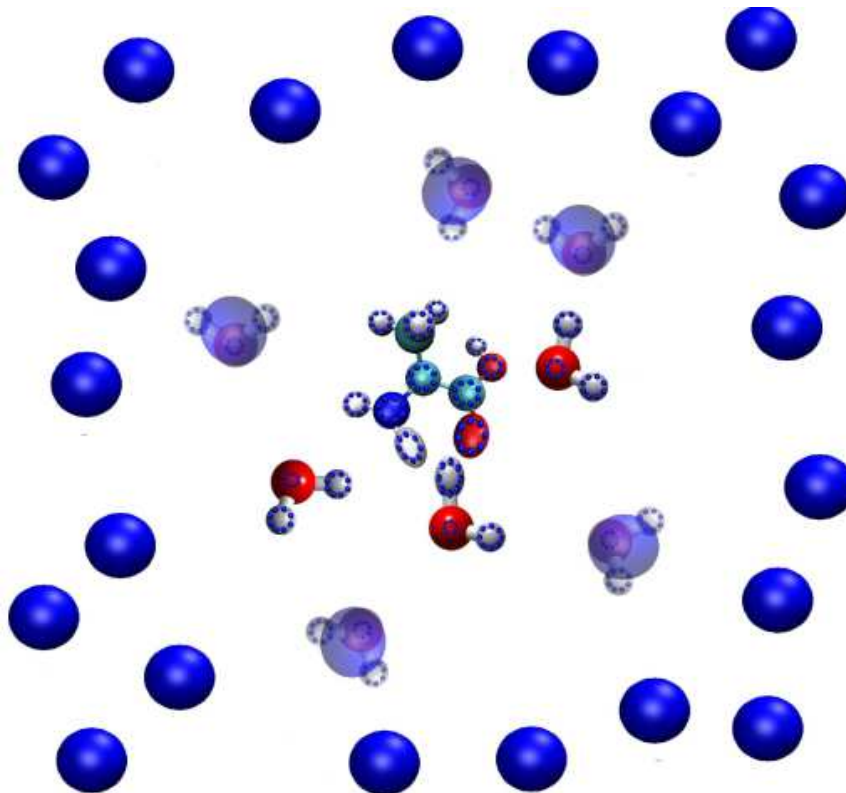
FIG. 9. Pictorial representation of the solvation of a small biomolecule in water within the AdResS path integral scheme. Around the solute in the first solvation shell, hydrogen and oxygens are treated at path integral level, thus the effects of spatial delocalization of the hydrogen is properly described; far way from the solute, water is represented by classical spherical particles. The transition from one representation to another occurs as in standard classical AdResS via a space dependent hybrid molecular representation.
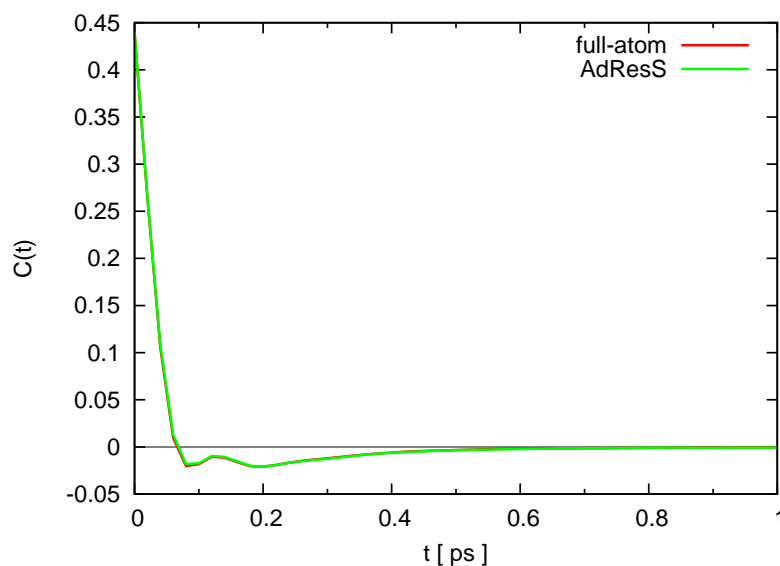

FIG. 10. The velocity auto-correlation function, $C(t) = \langle v(t) \cdot v(0) \rangle$ plotted against time. In green the data obtained with AdResS, in red the data obtained from a full atomistic simulation.

directions. First, the force calculations of the AdResS part could be vectorized, too. Second, as the coarse-grained

part is computationally less demanding than the rest of the simulation, the box needs to be divided into non-uniform domains among the different processors to decrease the load imbalance and the "double representation trick" can be avoided by removing the atomistic identity. Third, the time step in the coarse-grained region can be increased due to the increased softness of the coarse-grained interactions.

D. Possible extension to non equilibrium

A possible extension of AdResS is that of treating systems which are not in equilibrium. A possible path for such an extension is to merge AdResS with a dynamical approach to the nonequilibrium molecular dynamics (D-NEMD) as developed in Refs.[52,53]. In brief, such method considers the the response of systems in equilibrium to non-equilibrium perturbations following Onsager's regression hypothesis[54,55]. The hypothesis states that, up to a certain extent, external perturbations can be treated as (large) fluctuations within the equilibrium statistics of the system, and that the system's non-equilibrium response can be compared with fluctuations slowly returning back to equilibrium. An observable, $O(t)$, in a system subject to a perturbation which drives it out of equilibrium, can be calculated as:

$$O(t) = \int d\boldsymbol{x}\, \hat{O}(\boldsymbol{x}(t))\rho(\boldsymbol{x}, 0), \tag{13}$$

where $\boldsymbol{x}$ is the phase space variable of the system. $\hat{O}(\boldsymbol{x})$ is the microscopic observable, which is a only a function of the phase space position. $\rho(\boldsymbol{x}, 0)$ is the initial (equilibrium) probability density. The advantage is that one uses the equations and thermodynamic relations of equilibrium, that is, one uses equilibrium simulations to derive an ensemble of initial configuration along each of which a non equilibrium simulation is performed. $O(t)$ is then calculated averaging over the non equilibrium trajectories. The method has been developed for Canonical ensembles and strongly relies on the Liouville equation. In order to extend such an approach to (the atomistic region of) AdResS there are two major points that must be treated: (a) the method must be extended to the Grand Canonical case, which implicitly requires the extension of Liouville equation to open systems, thus with varying number of particles; (b) the thermostat used in AdResS for the removing/insertion of degrees of freedom should be such that it does not create artifacts by adsorbing the effects of the external perturbation. A proper solution of problem (a) requires a major conceptual step forward, instead point (b) has already been addressed in some recent work, by defining of a thermostat that acts only outside the region where the perturbation is applied[56]. Once problem (a) will be solved, then the application of AdResS to non equilibrium systems will be possible (for more details, see Ref.[51]).

CONCLUSION

We have reported the basic principles and the technical aspects of AdResS. Its conceptual evolution, from intuitive physical principles to a more rigorous formalization in terms of Grand Canonical-like set up, was extensively discussed. A list of open problems was given, in particular the fact that the conditions in the transition region, are only necessary but not sufficient. Moreover, the current computational gain, is not more than a factor two, compared with equivalent full atomistic simulations. Although such a gain is not trivial, expecially for large biological systems, however it can be highly improved upon technical optimization of the code; this aspect is probably the most relevant for applications.

ACKNOWLEDGMENTS

[1] 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).

[2] 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).

[3] 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).

[4] L.Delle Site What is a Multiscale Problem in Molecular Dynamics? *Entropy* **16** 23-40 (2014).

[5] M.C.Colombo, L.Guidoni, A.Laio, A.Magistrato, P.Maurer, S.Piana, U.Rohrig, K.Spiegel, M.Sulpizi, J.VandeVondele, M.Zumstein, U.Röthlisberger Hybrid QM/MM Car-Parrinello simulations of catalytic and enzymatic reactions. *CHIMIA* **56** 13 (2002).

[6] C.Junghans and S.Poblete A reference implementation of the adaptive resolution scheme in ESPResSo. *Comp.Phys.Comm.* **181** 1449 (2010).

[7] J.D.Halverson,T.Brandes, O.Lenz, A.Arnold, S.Bevc, V.Starchenko, K.Kremer, T.Stuehn, D.Reith ESPResSo++: A modern multiscale simulation package for soft matter systems *Comp.Phys.Comm.* **184** 1129 (2012).

[8] H.Wang, C.Schütte and L.Delle Site Adaptive Resolution Simulation (AdResS): A smooth thermodynamic and structural transition from atomistic to coarse grained resolution and vice versa in a Grand Canonical fashion. *J.Chem.Th.Comp.* **8** 2878 (2012).

[9] M.Praprotnik, L.Delle Site and K.Kremer Adaptive Resolution Scheme (AdResS) for Efficient Hybrid Atomistic/Mesoscale Molecular Dynamics Simulations of Dense Liquids. *Phys.Rev.E* **73** 066701 (2006).

[10] M.Praprotnik, L.Delle Site and K.Kremer A macromolecule in a solvent: Adaptive resolution molecular dynamics simulation. *J.Chem.Phys.* **126** 134902 (2007).

[11] M.Praprotnik, S.Matysiak, L.Delle Site, K.Kremer and C.Clementi Adaptive resolution simulation of liquid water *J.Phys.Cond.Matt.* **19** 292201 (2007).

[12] S.Matysiak, C.Clementi, M.Praprotnik, K.Kremer and L.Delle Site Modeling Diffusive Dynamics in Adaptive Resolution Simulation of Liquid Water. *J.Chem.Phys.* **128** 024503 (2008).

[13] B.P. Lambeth, C.Junghans, K.Kremer, C.Clementi, and L.Delle Site On the Locality of Hydrogen Bond Networks at Hydrophobic Interfaces. *J.Chem.Phys.* **133** 221101 (2010).

[14] D.Reith, M.Pütz and F.Müller-Plathe Deriving effective mesoscale potentials from atomistic simulations. *J.Comp.Chem.* **24** 1624-1636 (2003).

[15] V.Ruühle, C.Junghans, A.Lukyanov, K.Kremer and D.Andrienko Versatile object-oriented toolkit for coarse-graining applications. *J.Chem.Th.Comp.* **5** 3211-3223 (2009).

[16] M.Praprotnik, K.Kremer and L.Delle Site Fractional dimensions of phase space variables: A tool for varying the degrees of freedom of a system in a multiscale treatment. *J.Phys.A:Math.Th.* **40** F281-F288 (2007).

[17] L.Delle Site Some fundamental problems for an energy-conserving adaptive-resolution molecular dynamics scheme. *Phys.Rev.E* **76** 047701 (2007).

[18] M.Praprotnik, K.Kremer and L.Delle Site Adaptive molecular resolution via a continuous change of the phase space dimensionality. *Phys.Rev.E* **75** 017701 (2007).

[19] I. G. Tironi and W. F. van Gunsteren A molecular dynamics simulation study of chloroform *Mol.Phys.* **83** 381-403 (1994).

[20] B.Widom Some Topics in the Theory of Fluids *J.Chem.Phys.* **39** 2808 (1963).

[21] S.Poblete, M.Praprotnik, K.Kremer and L.Delle Site Coupling different levels of resolution in molecular simulations. *J.Chem.Phys.* **132** 114101 (2010).

[22] H.Wang, C.Junghans and K.Kremer Comparative atomistic and coarse-grained study of water: What do we lose by coarse-graining? *Eur.Phys.Jour.E* **28** 221-229 (2009).

[23] D.Mukherji, N.F.A.van der Vegt, K.Kremer and L.Delle Site Kirkwood-Buff analysis of liquid mixtures in an open boundary simulation. *J.Chem.Th.Comp.* **8** 375 (2012).

[24] D.Mukherji, N.F.A.van der Vegt, K.Kremer Preferential solvation of triglycine in aqueous urea: An open boundary simulation approach. *J.Chem.Th.Comp.* **8** 3536 (2012).

[25] B.Ensing, S.O. Nielsen, P.B. Moore, M.L. Klein, and M.Parrinello Energy conservation in adaptive hybrid atomistic/coarse-grain molecular dynamics. *J.Chem.Th.Comp.* **3** 1100-1105 (2007).

[26] R.E. Bulo, B.Ensing, J.Sikkema, and L.Visscher Toward a Practical Method for Adaptive QM/MM Simulations. *J.Chem.Th.Comp.* **5** 2212 - 2221 (2009).

[27] S.O. Nielsen, P.B. Moore, and B.Ensing Adaptive multiscale molecular dynamics of macromolecular fluids. *Phys.Rev.Lett.* **105** 237802 (2010).

[28] M.Praprotnik, S.Poblete, L.Delle Site and K.Kremer Comment on: Adaptive multiscale molecular dynamics of macromolecular fluids *Phys.Rev.Lett.* **107** 099801 (2011).

[29] B.Ensing, A.Laio, and S.O. Nielsen Hamiltonian adaptive hybrid atomistic/coarse-grain molecular dynamics. *Publication series of the John von Neumann Institute for Computing* **46** (2013).

[30] M.H.Peters An extended liouville equation for variable particle number systems. *arXiv:physics* **9809039** (1998).

[31] R.Potestio, S.Fritsch, P.Espanol, R.Delgado-Buscalioni, K.Kremer, R. Everaers and D.Donadio Hamiltonian adaptive resolution simulation for molecular liquids. *Phys.Rev.Lett.* **110**, 108301 (2013).

[32] R.Potestio, P.Espanol, R.Delgado-Buscalioni, R. Everaers, K.Kremer and D.Donadio Monte Carlo adaptive resolution simulation of multicomponent molecular liquids. *Phys.Rev.Lett.* **111**, 060601 (2013).

[33] J.A.Wagoner and V.S.Pande Finite domain simulations with adaptive boundaries: Accurate potentials and nonequilibrium movesets. *J.Chem.Phys.* **139** 234114 (2013).

[34] S. Fritsch, C. Junghans and K. Kremer, Structure formation of toluene around C60: Implementation of the Adaptive Resolution Scheme (AdResS) into GROMACS. *J.Chem.Th.Comp.* **8**, 398 (2012).

[35] A.Agarwal, H.Wang, C.Schütte and L.Delle Site Chemical potential of liquids and mixtures via Adaptive Resolution Simulation *J. Chem. Phys.* **141** 034102 (2014).

[36] D.P.Geerke and W.F. van Gunsteren Force Field Evaluation for Biomolecular Simulation: Free Enthalpies of Solvation of Polar and Apolar Compounds in Various Solvents. *ChemPhysChem* **7** 671 (2006).

[37] M.E.Lee and N.F.A. van der Vegt A new force field for atomistic simulations of aqueous tertiary butanol solutions. *J.Chem.Phys.* **122** 114509 (2005).

[38] J.R. Pliego Jr and J.M. Riveros Gibbs energy of solvation of organic ions in aqueous and dimethyl sulfoxide solutions *Phys. Chem. Chem. Phys.* **4** 1622 (2002).

[39] L.J. Smith and H.J.C. Berendsen and W.F. van Gunsteren Computer Simulation of Urea?Water Mixtures:? A Test of Force Field Parameters for Use in Biomolecular Simulation *J. Phys. Chem. B* **108** 1065 (2004).

[40] G.Csanyi, T.Albaret, M.C.Payne and A.De Vita "Learn on the fly": a hybrid classical and quantum-mechanical molecular dynamics simulation. *Phys.Rev.Lett.* **93**, 175503 (2004).

[41] K. Park, A. Götz, R.C. Walker and F. Paesani Application of adaptive QM/MM methods to molecular dynamics simulations of aqueous systems. *J.Chem.Th.Comp.* **8** 2868 (2012).

[42] R.P.Feynman and A.R.Hibbs Quantum Mechanics and Path Integrals. McGraw-Hill Inc.

[43] Path Integration via Molecular Dynamics. *Publication series of the John von Neumann Institute for Computing* **10** (2002).

[44] A.B.Poma and L.Delle Site Classical to Path-Integral Adaptive Resolution in Molecular Simulation: Towards a Smooth Quantum-Classical Coupling. *Phys.Rev.Lett.* **104** 250201 (2010).

[45] A.B.Poma and L.Delle Site Adaptive Resolution Simulation of Liquid Para-Hydrogen: Testing the robustness of the Quantum-Classical Adaptive Coupling. *Phys.Chem.Chem.Phys.* **13** 10510-10519 (2011).

[46] R.Potestio and L.Delle Site Quantum Locality and Equilibrium Properties in Low-temperature Parahydrogen: A Multiscale Simulation Study. *J.Chem.Phys.* **136** 054101 (2012).

[47] F. Paesani, W. Zhang, T.E. Cheatham, III, D.A. Case and G.A. Voth An accurate and simple quantum model for liquid water *J.Chem.Phys.* **125** 18457 (2006).

[48] S.Fritsch, R.Potestio, D.Donadio and K.Kremer Nuclear Quantum Effects in Water: A Multi-Scale Study. *J.Chem.Th.Comp.* **10** 816 (2014).

[49] S.Fritsch Scale Bridging Concepts in Molecular Simulation: Coarse-Graining and Thermodynamic Coupling. PhD Thesis in Physics, University of Mainz (2013).

[50] J.Zhu, H.Wang, A.Agarwal and L.Delle Site Equilibrium time correlation functions in open systems *arXiv:1412.0866* (2014).

[51] H. Wang and A.Agarwal Adaptive Resolution Simulation in Equilibrium and Beyond *arXiv:1412.2340* (2014).

[52] G.Ciccotti G.and Jacucci Direct computation of dynamical response by molecular dynamics: The mobility of a charged Lennard-Jones particle. *Phys.Rev.Lett.* **35** 789 (1975).

[53] G.Ciccotti, G. Jacucci and I.R.McDonald "Thought-experiments" by molecular dynamics. *J.Stat.Phys.* **21** 1-22 (1979).

[54] L. Onsager. Reciprocal relations in irreversible processes. i. *Phys. Rev.* **37** 405 (1931).

[55] L. Onsager. Reciprocal relations in irreversible processes. ii. Phys. Rev. **38** 2265 (1931).

[56] H.Wang, C.Schütte, G.Ciccotti and L.Delle Site Exploring the conformational dynamics of alanine dipeptide in solution subjected to an external electric field: A nonequilibrium molecular dynamics simulation. *J.Chem.Th.Comp.* **10** 1376 (2014).