

# UC Davis

## UC Davis Previously Published Works

### Title

Toward Hamiltonian Adaptive QM/MM: Accurate Solvent Structures Using Many-Body Potentials

### Permalink

<https://escholarship.org/uc/item/2xv1h9t9>

### Journal

Journal of Chemical Theory and Computation, 12(8)

### ISSN

1549-9618

### Authors

Boereboom, Jelle M  
Potestio, Raffaello  
Donadio, Davide  
[et al.](#)

### Publication Date

2016-08-09

### DOI

10.1021/acs.jctc.6b00205

Peer reviewed

# Toward Hamiltonian Adaptive QM/MM: Accurate Solvent Structures using Many-body Potentials

Jelle M. Boereboom,<sup>†</sup> Raffaello Potestio,<sup>‡</sup> Davide Donadio,<sup>¶,‡,§,||</sup> and Rosa E.  
Bulo<sup>\*,†</sup>

<sup>†</sup>*Inorganic Chemistry and Catalysis group, Debye Institute for Nanomaterials Science,  
Utrecht University, Universiteitsweg 99, 3584 CG Utrecht, The Netherlands*

<sup>‡</sup>*Max Planck Institute for Polymer Research, Ackermannweg 10, 55128 Mainz, Germany*

<sup>¶</sup>*Department of Chemistry, University of California Davis, One Shields Avenue, Davis,  
CA, 95616*

<sup>§</sup>*Donostia International Physics Center, Paseo Manuel de Lardizabal, 4, E-20018  
Donostia-San Sebastian, Spain*

<sup>||</sup>*IKERBASQUE, Basque Foundation for Science, E-48011 Bilbao, Spain*

E-mail: R.E.Bulo@uu.nl

## Abstract

Adaptive quantum mechanical (QM) / molecular mechanical (MM) methods enable efficient molecular simulations of chemistry in solution by describing reactive subregions with an accurate many-body potential energy expression (QM) while the rest of the system is described in a more approximate manner (MM). As solvent molecules diffuse in and out of the reactive region, they are gradually included into (and excluded from) the many-body QM potential. **It would be desirable to model such system**

using an adaptive Hamiltonian, but so far it has resulted in distorted structures at the boundary between the two regions. Here, we propose a Hamiltonian scheme to describe adaptively solvent diffusion across a multi-scale boundary separating configurational potentials that cannot be expressed by a multi-body expansion. The adaptive expressions are entirely general, and complimentary to all standard (non-adaptive) QM/MM embedding schemes available. We demonstrate the validity of our approach on a system described by two different MM potentials (MM/MM'), in which long-range interactions are treated by many-body Ewald summation. Our Hamiltonian approach provides both energy conservation and the correct solvent structure everywhere in the system, thus enabling microcanonical adaptive QM/MM simulations that can be used to obtain vibrational spectra and dynamical properties.

## 1 Introduction

Molecular dynamics (MD) simulations of chemical processes in a complex environment can be significantly accelerated with a dual-resolution approach, modeling the region of interest (active or A-region, Figure 1) at high resolution (HR), while the environment (E-region, Figure 1) is modeled at lower resolution (LR).<sup>1-3</sup> Conventional dual-resolution approaches<sup>4,5</sup> define the active region as a preselected set of atoms. This strategy works well if the molecular system is rigid, but in a solute-solvent system HR solvent molecules readily diffuse away from the active region, to be replaced by LR solvent molecules. Adaptive resolution methods address this issue by dynamically assigning molecules HR or LR character based on their proximity to the active site.<sup>6-13</sup> Generally, this procedure involves a transition region that connects smoothly the active and environment regions (T-region, Figure 1). The solvent molecules in the T-region have partial HR and partial LR character, and the description of each solvent molecule  $s$  smoothly changes from HR to LR (or vice versa) as it moves across the region. The HR character fraction changes with the distance  $r_s$  from a predefined HR center (Figure 1, blue water molecule).

Adaptive resolution methods fall into two categories. The first category combines complex many-body potential energy expressions, of which the combination of quantum mechanics (QM) and molecular mechanics (MM) is the most common example (QM/MM). Other examples include QM/QM, but also MM/MM, since some MM descriptions employ many-body expressions like Ewald summations to describe interactions across periodic boundaries.<sup>20</sup> The second category combines potential energy expressions that can be reduced to a simple sum over particle pairs, such as most MM and coarse grained (CG) particle descriptions (MM/CG).<sup>14-17</sup> While the two problems are similar, MM/CG developments cannot always straightforwardly be extended to QM/MM, because the use of a QM potential in the HR region involves further layers of complexity, stemming from its many-body configurational nature. **(DD: among the other problems that we do not consider here, we could also mention that the number of electrons may not be conserved and that we remove electronic degrees of freedom)**

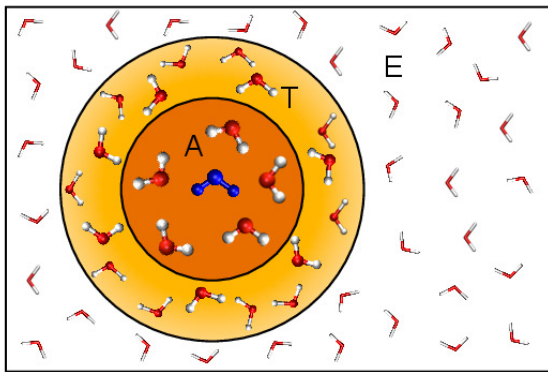


Figure 1: (color online). Schematic representation of an adaptive dual-resolution description of water in water, partitioned into an A-region [dark grey (orange)], a T-region [light grey (yellow)], and an E-region [white], around a central dark grey (blue) water molecule. Ball&Stick water molecules are HR, and LR molecules are depicted with thick lines. The HR character of a solvent molecule  $s$  is determined by its distance  $r_s$  from the HR center.

Currently, there are only few examples of adaptive QM/MM simulations based on a Hamiltonian formalism (energy conserving).<sup>6,8</sup> In all practical situations, the proposed methods failed to provide reliable solvent structures, while alternative non-Hamiltonian schemes

performed much better in this respect.<sup>8,18</sup> Similar findings also directed MM/CG developments<sup>19</sup>, until very recently, when a Hamiltonian approach was introduced (H-AdResS) that conserves both energy and solvent structure.<sup>17</sup> This approach relies on the MM/CG energy expression being written as a sum over pair-wise interactions. In this paper we reformulate the energy and force expressions for the state-of-the-art Hamiltonian adaptive QM/MM approaches (**DD: references here?**) in terms that match the simpler MM/CG expressions. From such expressions we derive a novel Hamiltonian scheme that connects different many-body potentials (e.g. QM/MM, QM/QM, or many-body MM/MM) and provides accurate geometries whilst conserving the total energy. The novel approach is an extension of current adaptive QM/MM formulations, and is very general in that it can be combined with many of the available flavors.<sup>6,8,9</sup>

This paper is organized as follows. In Section 2 we introduce the theory behind adaptive dual-resolution simulations. We first compare the latest adaptive MM/CG and QM/MM expressions, and we discuss the H-AdRes MM/CG correction to the Hamiltonian. We then rewrite the adaptive QM/MM (many-body) expressions so that we can define the criteria for a 'per particle' correction analogue to MM/CG. Finally, we discuss a simple correction that has been applied in previous works,<sup>7,8,22</sup> and introduce our novel HadQMMM correction. (**DD: don't like HadQMMM, as I think that it is misleading. How about calling it Hamiltonian adaptive many-body correction HAMBC?**) In Section 3, we illustrate a test case that we use to prove our adaptive simulation scheme, and we provide the computational details. In Section 4, we present the results of two possible 'per particle' corrections to the Hamiltonian, and show that the HadQMMM correction results in accurate solvent structures, whilst conserving the total energy.

## 2 Theory

The Hamiltonian of any system consists of a global kinetic energy function  $T(\mathbf{p})$  (of momentum vector  $\mathbf{p}$ ), and a potential energy function  $V(\mathbf{r})$  (of coordinate vector  $\mathbf{r}$ ) describing all interactions in the system. In this section we first compare the potential energy expressions  $V(\mathbf{r})$  currently used in adaptive MM/CG and QM/MM simulations. Then we explain why these uncorrected expressions produce inaccurate forces on the particles, and we describe the 'per particle' H-AdRes correction to the MM/CG Hamiltonian. We then deduce the criteria for an analogue 'per-particle' correction to the many-body QM/MM Hamiltonian, and discuss a simple correction that has been previously applied. In the final subsection we derive our novel HadQMMM correction from the criteria presented before.

### 2.1 Adaptive Dual-Resolution Potential Energy Expressions

The H-AdResS approach combines an MM and a CG potential ( $V^{\text{MM}}(\mathbf{r})$  and  $V^{\text{CG}}(\mathbf{r})$ ) into a global  $V^{\text{MM/CG}}(\mathbf{r})$ . Since  $V^{\text{MM}}(\mathbf{r})$  and  $V^{\text{CG}}(\mathbf{r})$  can both be expressed as a sum of pair-wise interaction terms, the combined potential energy can also be written as a sum over particle pairs,

$$V^{\text{MM/CG}}(\mathbf{r}) = \sum_{\alpha < \beta} \left( \frac{\lambda_\alpha + \lambda_\beta}{2} V_{\alpha\beta}^{\text{MM}} + \left(1 - \frac{\lambda_\alpha + \lambda_\beta}{2}\right) V_{\alpha\beta}^{\text{CG}} \right), \quad (1)$$

with  $V_{\alpha\beta} = V(\mathbf{r}_\alpha - \mathbf{r}_\beta)$  an interaction potential for the particle pair  $\alpha\beta$ , and  $\lambda_{\alpha/\beta} = \lambda(\mathbf{r}_{\alpha/\beta})$  a simple continuous function of the distance  $r_{\alpha/\beta}$  of particle  $\alpha/\beta$  to the center of the A-region. The function  $\lambda_{\alpha/\beta}$  constitutes the contribution (or weight) of the MM potential  $V_{\alpha\beta}^{\text{MM}}$ , and expresses the MM (HR) character of the particle.<sup>17</sup> It has a value 1 if particle  $\alpha/\beta$  is in the A-region, a value 0 in the E-region, and a fractional number between 1 and 0 in the T-region.

In contrast, a many-body interaction potential cannot be decomposed into pair contributions, and the dual-resolution potential must be expressed in global terms (**DD: what do you mean by "global"?**). Conventional QM/MM methods partition the molecular system

into a set of fully QM solvent molecules and a set of fully MM solvent molecules. Labeling a specific choice of QM/MM partitioning  $p$ , with the set of QM molecules named  $\mathcal{S}_p$  and the complementary set of MM molecules  $\mathcal{S}'_p$ , the QM/MM potential energy becomes,

$$E_p(\mathbf{r}) = V_{\mathcal{S}_p}^{\text{QM}} + V_{\mathcal{S}'_p}^{\text{MM}} + V_{\text{int}}(\mathbf{r}). \quad (2)$$

Here  $V_{\mathcal{S}_p}^{\text{QM}} = V^{\text{QM}}(\mathbf{r}_{\mathcal{S}_p})$  is the QM potential energy for the subsystem of molecules belonging to the set  $\mathcal{S}_p$ , and  $V_{\text{int}}(\mathbf{r})$  is an interaction energy between the two types of molecules, which can be defined in several different ways.<sup>21</sup> Mechanical embedding and electrostatic embedding are the most common choices for this interaction energy, although many other options are available. The expression in Eq. 2 is completely general, covering all types of QM/MM embedding.

An adaptive QM/MM simulation must account for the fact that all solvent molecules in the T-region have different partial QM characters. This can be achieved by including contributions from different partitions into the global potential energy expression.<sup>6</sup>

$$V^{\text{QM/MM}}(\mathbf{r}) = \sum_{p \in \mathcal{P}} \sigma_p(\mathbf{r}) E_p(\mathbf{r}). \quad (3)$$

Here  $\mathcal{P}$  is the set of all possible partitions  $p$  ( $2^n$  partitions in case of  $n$  solvent molecules). The function  $\sigma_p(\mathbf{r})$  denotes the contribution (or weight) of partition  $p$ , and the sum over all weights equals 1 ( $\sum_{p \in \mathcal{P}} \sigma_p(\mathbf{r}) = 1$ ). Each weight  $\sigma_p(\mathbf{r})$  is a function of the coordinates of all atoms. The fractional QM character  $\omega_s(\mathbf{r})$  of a solvent molecule with label  $s$  is then the sum of weights of the contributing partitions that describe this solvent molecule QM. It is the QM/MM analogue of  $\lambda_\alpha$  in MM/CG, and can be written as,

$$\omega_s(\mathbf{r}) = \sum_{p \in \mathcal{P}} \delta_s(\mathcal{S}_p) \sigma_p(\mathbf{r}), \quad (4)$$

with  $\delta_s(\mathcal{S}_p)$  the Dirac measure (**DD: this is not a Dirac distribution, it is a step**

**function - Heaviside - or if you want a Kronecker  $\delta$** ), which is 1 if  $s \in \mathcal{S}_p$  and 0 if  $s \in \mathcal{S}'_p$ . The concept of **combinatorial** partitions is schematically visualized in Figure 2 for the simple case of only two solvent molecules in the T-region. Partitions  $P_1 - P_4$  describe all solvent molecules in the A-region as QM and all molecules in the E-region as MM, and should therefore provide all important contributions to  $V^{\text{QM/MM}}(\mathbf{r})$ . Partitions  $P_5$  and  $P_6$  describe a molecule in the E-region as QM, or a molecule in the A-region as MM, and should therefore not contribute. The general form of the energy expression in Eq. (3) is used in many of the current adaptive QM/MM methods.<sup>6,8,9</sup> The expressions only differ in the functional form of the weights  $\sigma_p(\mathbf{r})$ .

The adaptive QM/MM methods that use the general expression in Eq. (3) are Permuted Adaptive Partitioning (PAP),<sup>6</sup> Sorted Adaptive Partitioning (SAP),<sup>6</sup> Difference-based Adaptive Solvation (DAS),<sup>8</sup> and Size-Consistent Multi-Partitioning (SCMP).<sup>9</sup> The PAP method defines non-zero weight functions for all QM/MM permutations of T-region molecules ( $P1 - P4$ , Figure 2), which comes down to exponential scaling of the computational cost with the number of molecules in the T-region ( $M$ ). The DAS method reduces the computational cost by assigning zero weight to a (much) larger number of partitions. In principle only contributions from 'ordered' partitions (P1-P3 Figure 2) are non-zero, but to ensure continuity of the forces extra partitions are included if two solvent molecules are at similar distance from the QM center. As a result, DAS scales approximately linear with  $M$  provided that the solvent structure in the T-region does not contain regions of extremely high density. Like DAS, the SAP weight-functions are only non-zero for 'ordered' partitions. The SAP computational costs scale linear with  $M$  in all cases, but this is achieved at the cost of the simplicity of the weight-functions. The SAP potential energy surface is very steep in places, and simulations require a small time-step for proper integration of the equations of motion. Finally, the SCMP weight functions are constructed in such a way that the number of contributing partitions is conserved throughout the simulation, and all contributing partitions always have the same number of QM (HR) molecules. As a result, an SCMP



simulation on a parallel platform can exhibit nearly perfect linear scaling behavior.

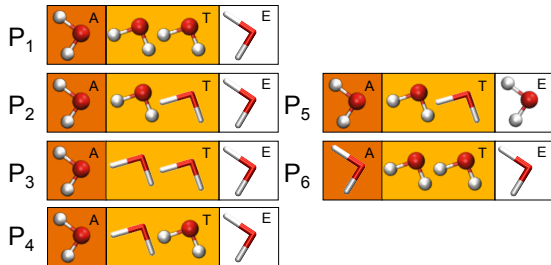


Figure 2: Six possible QM/MM partitions, with QM molecules depicted as Ball&Stick and MM molecules as thick lines. Partitions  $P_1 - P_4$  may all contribute to the adaptive QM/MM energy expression, while by construction  $P_5$  and  $P_6$  do not.

## 2.2 Transition forces

The forces derived from the adaptive QM/MM potential energy in Eq. 3 have the form,

$$\begin{aligned}
 F_i^{\text{Q/M}}(\mathbf{r}) &= - \sum_{p \in \mathcal{P}} \sigma_p(\mathbf{r}) \frac{\partial E_p(\mathbf{r})}{\partial x_i} - \sum_{p \in \mathcal{P}} \frac{\partial \sigma_p(\mathbf{r})}{\partial x_i} E_p(\mathbf{r}) \\
 &= F_i^{\text{ad}}(\mathbf{r}) + F_i^{\text{tr}}(\mathbf{r}),
 \end{aligned}
 \tag{5}$$

where  $x_i$  is a component of the vector  $\mathbf{r}$ . This negative gradient of the potential energy is a sum of two terms: a so called adaptive force  $F_i^{\text{ad}}(\mathbf{r})$  and a transition force  $F_i^{\text{tr}}(\mathbf{r})$ . The former term is a linear combination of force terms derived from a conventional QM/MM potential energy (Eq. (2)), but the second term (the transition force) contains the gradients of the weights, and causes anomalies in the structure of the solvent.<sup>8,18</sup>

A pragmatic solution to the problem of the distorted structures is to discard the Hamiltonian formalism, and simply neglect the offending term (e.g. use only  $F_i^{\text{ad}}(\mathbf{r})$  in Eq. (5) for propagation). It has been demonstrated repeatedly that, while such non-Hamiltonian simulations do not conserve energy (the applied forces do not integrate to a consistent potential energy<sup>19</sup>), they do provide reliable structures.<sup>8,18,22,23</sup> Nonetheless, giving up a Hamiltonian formalism has fundamental and practical consequences.<sup>17</sup> Without a well-defined energy a

partition function cannot be defined, and it is difficult to interpret the meaning of average values obtained from a simulation. In practice, the MD simulations require local thermostats with strong coupling to prevent heating, and the reliability of dynamical quantities (diffusion coefficients, time correlation functions) is compromised.<sup>22</sup>

A more rigorous solution is to adjust the Hamiltonian with an extra term in the potential energy expression of Eq. (3). The gradient of this term should then effectively cancel  $F_i^{tr}(\mathbf{r})$ . The H-AdResS approach pursues the MM/CG equivalent of this strategy. The MM/CG forces derived from Eq. (1) have the form,

$$\begin{aligned}
 F_{\alpha i}^{\text{MM/CG}}(\mathbf{r}) &= - \sum_{\beta} \left( \frac{\lambda_{\alpha} + \lambda_{\beta}}{2} \frac{\partial V_{\alpha\beta}^{\text{MM}}}{\partial x_{i\alpha}} + \left(1 - \frac{\lambda_{\alpha} + \lambda_{\beta}}{2}\right) \frac{\partial V_{\alpha\beta}^{\text{CG}}}{\partial x_{i\alpha}} \right) \\
 &\quad - \frac{1}{2} \frac{\partial \lambda_{\alpha}}{\partial x_{i\alpha}} \sum_{\beta} [V_{\alpha\beta}^{\text{MM}} - V_{\alpha\beta}^{\text{CG}}],
 \end{aligned}
 \tag{6}$$

where the last term, containing the gradient of the MM/CG weight function  $\lambda_{\alpha}$ , is the equivalent of  $F_i^{tr}(\mathbf{r})$  in Eq. (5), and has similar catastrophic effects on the structure of the system. Unlike the QM/MM force expression in Eq. (5), the simpler MM/CG force expression in Eq. (6) clearly shows that the transition force scales linearly with the difference between the two types of interaction energy (MM and CG) of particle  $\alpha$  with the rest of the system. **In a thermodynamic perspective this energy difference determines the free energy released or absorbed when particle  $\alpha$  is converted from CG to MM.<sup>17</sup> At fixed conformation of the system this energy difference is the vertical difference between two potential energy surfaces at a specific position  $\mathbf{r}$ .** For each particle  $\alpha$  the H-AdResS approach applies a correction term to the potential energy expression in Eq. (1). This term is constructed such that its gradient effectively cancels the vertical energy release for particle  $\alpha$ .

### 2.3 'Per Molecule' Correction to a Many-Body Hamiltonian

In this work we show that we can arrive at a similar 'per particle' correction for the QM/MM potential energy expression in Eq. (3). The first step is to rewrite Eq. (5) into a form

comparable to Eq. (6). We define an empty set of QM molecules  $\mathcal{S}_0$  associated to the QM/MM potential energy  $E_0(\mathbf{r})$  that describes the system fully MM. In Eq. (3) the sum of the weights  $\sigma_p(\mathbf{r})$  equals 1 ( $\sum_{p \in \mathcal{P}} \sigma_p(\mathbf{r}) = 1$ ), so that  $\sigma_0(\mathbf{r}) = 1 - \sum_{\substack{p \in \mathcal{P} \\ p \neq 0}} \sigma_p(\mathbf{r})$ . We can therefore rewrite Eq. (3) as,

$$V^{\text{QM/MM}}(\mathbf{r}) = \sum_{\substack{p \in \mathcal{P} \\ p \neq 0}} \sigma_p(\mathbf{r}) E_p(\mathbf{r}) + \left(1 - \sum_{\substack{p \in \mathcal{P} \\ p \neq 0}} \sigma_p(\mathbf{r})\right) E_0(\mathbf{r}), \quad (7)$$

and the transition force defined in Eq. (5) as,

$$F_i^{\text{tr}}(\mathbf{r}) = - \sum_{\substack{p \in \mathcal{P} \\ p \neq 0}} \frac{\partial \sigma_p(\mathbf{r})}{\partial x_i} (E_p(\mathbf{r}) - E_0(\mathbf{r})). \quad (8)$$

Eq. (8) shows that the transition force  $F_i^{\text{tr}}(\mathbf{r})$  is linearly dependent on the difference between the partition energies  $E_p(\mathbf{r})$  and the fully MM potential energy  $E_0(\mathbf{r})$ . This can be seen as the energy release when all the QM solvent molecules in a set  $\mathcal{S}_p$  are converted to MM. Analogously to the MM/CG expressions we can simply divide this term into 'per molecule' contributions, representing the energy released when the specific solvent molecule is converted from QM to MM,

$$\Delta E_p(\mathbf{r}) = E_p(\mathbf{r}) - E_0(\mathbf{r}) = \sum_{s \in \mathcal{S}_p} \Delta \epsilon_s^p(\mathbf{r}). \quad (9)$$

Note that  $\Delta E_p(\mathbf{r})$  is different for each global geometry  $\mathbf{r}$  and for each partition  $p$ , and that Eq. (9) assumes nothing about the definition of  $\Delta \epsilon_s^p(\mathbf{r})$ . In fact, there is no unique way to subdivide  $\Delta E_p(\mathbf{r})$  into molecular contributions, since they depend on the order in which the molecules are converted to MM.

The intuitive solution is to assume equal contributions from all molecules  $s$  in Eq. (9). Earlier attempts at a Hamiltonian expression went one step further,<sup>6,8</sup> and also assumed similar contributions for each partition  $p$  and geometry  $\mathbf{r}$ , attempting to cancel each  $\Delta \epsilon_s^p(\mathbf{r})$

with the same constant  $C$ . An appealingly simple choice for  $C$  is the difference in QM and MM energies for a single (gas phase) molecule at the geometries  $\mathbf{r}_s^{\text{QM}}$  and  $\mathbf{r}_s^{\text{MM}}$  optimized with the respective methods ( $C = V^{\text{QM}}(\mathbf{r}_s^{\text{QM}}) - V^{\text{MM}}(\mathbf{r}_s^{\text{MM}})$ ). A corrected potential energy expression of the form,

$$\tilde{V}^{\text{QM/MM}}(\mathbf{r}) = V^{\text{QM/MM}}(\mathbf{r}) - \sum_{p \in \mathcal{P}} \sigma_p(\mathbf{r}) \sum_{s \in \mathcal{S}_p} C, \quad (10)$$

has the desired corrected gradient,

$$\tilde{F}_i^{\text{QM/MM}}(\mathbf{r}) = F_i^{\text{ad}}(\mathbf{r}) - \sum_{p \in \mathcal{P}} \frac{\partial \sigma_p(\mathbf{r})}{\partial x_i} \left( E_p(\mathbf{r}) - \sum_{s \in \mathcal{S}_p} C \right) \quad (11)$$

$$= F_i^{\text{ad}}(\mathbf{r}) - \sum_{\substack{p \in \mathcal{P} \\ p \neq 0}} \frac{\partial \sigma_p(\mathbf{r})}{\partial x_i} \left( E_p(\mathbf{r}) - \sum_{s \in \mathcal{S}_p} C - E_0(\mathbf{r}) \right). \quad (12)$$

The sum in Eq. (12) equals the sum in Eq. (11) because for partition  $p = 0$  no correction term applies as  $\mathcal{S}_0$  is an empty set. If  $C$  is indeed a good representation of  $\Delta \epsilon_s^p(\mathbf{r})$  for each geometry  $\mathbf{r}$ , QM/MM partition  $p$ , and solvent molecule  $s$ , then each  $E_p(\mathbf{r})$  is corrected to be similar to  $E_0(\mathbf{r})$ , and the transition force (Eq. (8)) is approximately canceled. However, it has previously been shown that such a simple correction results in equilibrium solvent structures with an extreme depletion of solvent molecules across the T-region.<sup>7,8,22</sup>

The depletion in the transition region can be explained using insights presented in the H-AdResS paper.<sup>17</sup> The authors suggested that the energy release upon conversion of a particle  $\alpha$  from MM to CG ( $\sum_{\beta} [V_{\alpha\beta}^{\text{MM}} - V_{\alpha\beta}^{\text{CG}}]$ ) depends strongly on the fractional MM (HR) character of  $\alpha$  ( $\lambda_{\alpha}$ ) as reflected in its geometry. In our QM/MM simulations the analogue of  $\lambda_{\alpha}$  is  $\omega_s(\mathbf{r})$ . For a molecule in the A-region with  $\omega_s(\mathbf{r}) = 1$  and energy  $E_x$ , the geometry will resemble the equilibrium QM geometry. Therefore the MM potential energy of the molecule in this configuration will be higher than  $E_x$  and the energy increase upon converting molecule  $s$  from QM to MM will be negative ( $\langle \Delta \epsilon_s^p(\mathbf{r}) \rangle_{\omega_s=1} < 0$ : Figure 3, black versus red line). The reverse argument holds for the energy of a molecule in the E-region, for which  $\langle \Delta \epsilon_s^p(\mathbf{r}) \rangle_{\omega_s=0} > 0$

(Figure 3, black versus blue line). This already explains why a simple correction using a constant  $C$  cannot be sufficient. In the T-region, the molecule neither has the equilibrium QM geometry, nor the equilibrium MM geometry. Due to these non-equilibrium geometries both the QM and the MM energies will have a value higher than  $E_x$ . Any combination of a QM and an MM description will therefore raise the energy above  $E_x$ , which is reflected in the higher energies of particles in the T-region (Figure 3), leading to a local depletion of the water density in this region.

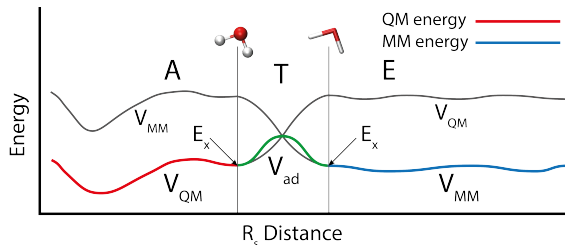


Figure 3: Schematic representation of QM, MM, and adaptive QM/MM potential energies after correction according to Eq. (10).

## 2.4 The HadQMMM Correction

The above argumentation implies a trend in the error of the water densities that depends on the distance of the water molecules to the QM center. This distance is directly related to the QM-character of the water molecules (1 at short distances, 0 at large distances). Such a trend complies with insights put forward in Ref.<sup>17</sup> stating that the energy release upon conversion of a particle  $\alpha$  from MM to CG ( $\sum_{\beta} [V_{\alpha\beta}^{\text{MM}} - V_{\alpha\beta}^{\text{CG}}]$ , Eq. (6)) depends strongly on its fractional MM character as reflected in its geometry. Similar to Ref.<sup>17</sup> the final step towards a working Hamiltonian approach is to approximate  $\sum_{s \in \mathcal{S}_p} \Delta \epsilon_s^p(\mathbf{r})$  with an expression that depends on the QM character  $\omega_s(\mathbf{r})$  of each molecule  $s$ . The chosen 'per molecule' correction term is an (ensemble) average over all coordinates  $\mathbf{r}$ , partitions  $p$ , and solvent molecules  $s$  at a specified value for  $\omega_s$ :  $\langle \Delta \epsilon_s^p(\mathbf{r}) \rangle_{\omega_s}$ . In the following, we will show that correcting the 'per molecule' energy release  $\Delta \epsilon_s^p(\mathbf{r})$  with this average is equivalent to correcting the potential

energy expression  $V^{\text{QM/MM}}(\mathbf{r})$  with a function  $H^c(\omega_s)$  for each molecule  $s$ , as long as  $\frac{dH^c(\omega_s)}{d\omega_s}$  equals  $\langle \Delta\epsilon_s^p(\mathbf{r}) \rangle_{\omega_s}$ . For a system of  $n$  solvent molecules,

$$\widehat{V}^{\text{QM/MM}}(\mathbf{r}) = V^{\text{QM/MM}}(\mathbf{r}) - \sum_{s=1}^n H^c(\omega_s). \quad (13)$$

The forces derived from this expression have the form,

$$\begin{aligned} \widehat{F}_i^{\text{QM/MM}}(\mathbf{r}) &= F_i^{\text{ad}}(\mathbf{r}) - \sum_{p \in \mathcal{P}} \frac{\partial \sigma_p(\mathbf{r})}{\partial x_i} E_p(\mathbf{r}) \\ &+ \sum_{s=1}^n \frac{dH^c(\omega_s)}{d\omega_s} \frac{\partial \omega_s(\mathbf{r})}{\partial x_i}. \end{aligned} \quad (14)$$

Since  $\omega_s(\mathbf{r})$  is a sum over  $\sigma_p(\mathbf{r})$  values for those partitions that describe molecule  $s$  QM (Eq. (4)), we can express its gradient to  $x_i$  as follows,

$$\sum_{s=1}^n \frac{\partial \omega_s(\mathbf{r})}{\partial x_i} = \sum_{s=1}^n \sum_{p \in \mathcal{P}} \delta_s(\mathcal{S}_p) \frac{\partial \sigma_p(\mathbf{r})}{\partial x_i} = \sum_{p \in \mathcal{P}} \frac{\partial \sigma_p(\mathbf{r})}{\partial x_i} \sum_{s \in \mathcal{S}_p} 1 \quad (15)$$

Inserting Eq. (15) into Eq. (14) we obtain,

$$\begin{aligned} \widehat{F}_i^{\text{QM/MM}}(\mathbf{r}) &= F_i^{\text{ad}}(\mathbf{r}) - \sum_{p \in \mathcal{P}} \frac{\partial \sigma_p(\mathbf{r})}{\partial x_i} \left( E_p(\mathbf{r}) - \sum_{s \in \mathcal{S}_p} \frac{dH^c(\omega_s)}{d\omega_s} \right) \\ &= F_i^{\text{ad}}(\mathbf{r}) - \sum_{\substack{p \in \mathcal{P} \\ p \neq 0}} \frac{\partial \sigma_p(\mathbf{r})}{\partial x_i} \left( E_p(\mathbf{r}) - \sum_{s \in \mathcal{S}_p} \frac{dH^c(\omega_s)}{d\omega_s} - E_0(\mathbf{r}) \right). \end{aligned} \quad (16)$$

If  $\frac{dH^c(\omega_s)}{d\omega_s}$  is a good estimate for the  $\Delta\epsilon_s^p(\mathbf{r})$  value of a solvent molecule  $s$  with QM character  $\omega_s(\mathbf{r})$  at any geometry  $\mathbf{r}$  accessed during the simulation, then  $E_p(\mathbf{r})$  is corrected towards the MM potential energy  $E_0(\mathbf{r})$ , and the transition force  $F_i^{\text{tr}}(\mathbf{r})$  in Eq. (8) is approximately canceled.

Since we know that the NVT ensemble of structures is well-reproduced by a non-Hamiltonian simulation using the force expression  $F_i^{\text{ad}}(\mathbf{r})$  in Eq. (5),<sup>23</sup> the ensemble average  $\frac{dH^c(\omega_s)}{d\omega_s} =$

$\langle \Delta \epsilon_s^p(\mathbf{r}) \rangle_{\omega_s}$  can be extracted from such a non-Hamiltonian simulation. The energy differences ( $\Delta \epsilon_s^p(\mathbf{r})$ ) that are obtained from the non-Hamiltonian simulation are the differences in energy between two QM/MM partitions that differ only in the description of one solvent molecule. For example, in Figure 2 the value of  $\Delta \epsilon_3^{P1}(\mathbf{r})$  for solvent molecule 3 and partition  $P1$  is defined as the difference between  $E_{P1}(\mathbf{r})$  and  $E_{P2}(\mathbf{r})$ . The correction term in the energy  $H^c(\omega_s)$  (Eq. (15)) can be obtained by thermodynamic integration<sup>33</sup> of this average over  $\omega_s$ ,

$$H^c(\omega_s) = \int_0^{\omega_s} \langle \Delta \epsilon_s^p(\mathbf{r}) \rangle_{\omega'_s} d\omega'_s. \quad (17)$$

The above integral is by definition zero for solvent molecules that are purely MM ( $\omega_s = 0$ ).

### 3 Computational Details

We demonstrate the performance of our approach on a test system of water in water, using the adaptive QM/MM weight-functions  $\sigma_p(\mathbf{r})$  as formulated in the SAP method.<sup>6</sup> The SAP method has been selected in this work for a practical reason. The number of contributing partitions to the SAP expression always equals  $M + 1$ , while the DAS method,<sup>8</sup> due to the construction of the weight functions, includes more partitions when the solvent structures exhibit extreme distortions in the T-region (e.g. Hamiltonian simulations with a simple correction). Please note that for simulations with homogeneous solvent structures the SAP and DAS methods compute an equal number of partition terms. All simulations were performed with FlexMD, a python library that serves as a wrapper around several molecular program packages, each providing the required QM or MM energies and forces. FlexMD itself is distributed with the ADF program package,<sup>27,28</sup> and uses the atomistic simulation environment (ASE)<sup>29</sup> for MD propagation. Our model system is a 30.025 Å periodic simulation box containing 915 water molecules. The active region (A-region) is a 5.5 Å sphere around a central water molecule, while the transition region (T-region) is a 0.9 Å thick layer around the A-region. The two selected potentials for our test simulations are both MM

potentials, and we chose the SPC-Fw force-field to describe the central 'QM' region,<sup>25</sup> while the environment region is described with the TIP3P-fs force-field.<sup>25,26</sup> SCP-Fw and TIP3P-fs energies and forces are both computed with the NAMD program,<sup>30</sup> and the many-body character of both descriptions is introduced by the long-range electrostatics, for which Particle Mesh Ewald (PME) is used.<sup>24</sup> For the QM/MM interaction  $V_{int}(\mathbf{r})$  Eq. (2) we used simple mechanical embedding.<sup>31,32</sup> **(DD: what do you mean here? Aren't they interacting via electrostatics as well?)**

All simulations were first equilibrated for 10 ps at a water density of 1.01 g/mL in the canonical ensemble (NVT), using a Langevin thermostat with a friction of 82.7 ps<sup>-1</sup>. The equilibrated structures were used as starting geometries for 5 ps NVE simulations **(DD: still so short?)** (0.1 fs time-step) using the Hamiltonian expressions  $\tilde{V}^{QM/MM}(\mathbf{r})$  Eq. (10) and  $\hat{V}^{QM/MM}(\mathbf{r})$  Eq. (13). Additionally, five non-Hamiltonian simulations, using the forces  $F_i^{ad}(\mathbf{r})$  Eq. (5) for propagation, of 10 ps each were performed. The results from these five simulations were combined to obtain the average structures of the solution, and used as reference for all other results. These simulations also use a Langevin thermostat with a friction of 82.7 ps<sup>-1</sup>. Each simulation has the same starting structure but varying starting velocities, which were randomly generated. Finally, five 10 ps Hamiltonian NVT simulations were performed with the simple correction term (Eq. (10)), using random starting velocities and a 0.5 fs time-step. The same procedure was repeated with the HadQMMM correction of Eq. (13). For all thermostatted simulations a time-step of 0.5 fs was used, and it was verified that this yielded the same equilibrated structures as the NVE simulations with a smaller time-step.

## 4 Results

As demonstrated previously,<sup>8,18,22,23</sup> the non-Hamiltonian simulations using only  $F_i^{ad}(\mathbf{r})$  to propagate the trajectory result in a radial distribution ( $g(r)$ ) of water-oxygens around the



central oxygen atom  $O^*$  that is very similar to the O-O radial distribution of SCP-Fw and TIP3P-fs.<sup>25,26</sup> The first sharp peak is located at 2.8 Å from the central oxygen atom, and the second and third much shallower peaks lie at 4.4 Å and 6.7 Å respectively (Figure 4, red line). We use this result as a reference for the performance of the Hamiltonian simulations.

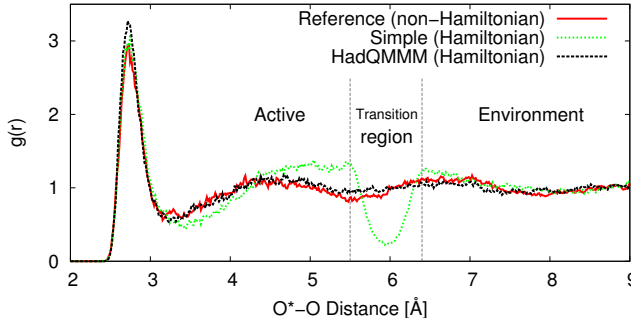


Figure 4: (color online). Radial distribution of oxygen around the central ‘QM’ water oxygen atom  $O^*$  for three different simulations: Non-Hamiltonian [dark grey (red)], Hamiltonian using a simple constant correction [light grey (green) dotted], and HadQMMM [black dotted].

The corrected Hamiltonian approach using  $\tilde{V}^{\text{QM/MM}}(\mathbf{r})$  (Eq. (10)) results in an  $O^*$ -O radial distribution (Figure 4, green line) that deviates strongly from the reference. The main difference is a depletion of the water density in the T-region, which is balanced by an increased density at the edges of the A- and E-regions. The water densities on either side of the T-region are similar, indicating that there is no significant chemical potential difference between a ‘QM’ water molecule in the A-region and an MM water molecule in the E-region. The correct density balance between the A- and E-regions is the result of the correction  $C$ , which proves to be a good approximation for the chemical potential difference between a QM and an MM water molecule at their respective equilibrium geometries.

Figure 4 also depicts the  $O^*$ -O  $g(r)$  resulting from a simulation using the potential energy expression  $\hat{V}^{\text{QM/MM}}(\mathbf{r})$  of Eq. (13) (HadQMMM). The approach clearly compensates for the transition forces  $F_i^{\text{tr}}(\mathbf{r})$  such that the correct solvent structure is obtained everywhere in the solution. In the Supplemental Material the tabulated function  $H^c(\omega_s)$  is given in Table SI.

The merit of our novel Hamiltonian approach is further confirmed with micro-canonical

(NVE) simulations that reveal no significant drift in total energy (no more than  $0.005 \text{ kcal mol}^{-1} \text{ ps}^{-1}$ ), which can be seen in Figure 5. This is in stark contrast to the total energy drift in the non-Hamiltonian simulations of  $10 \text{ kcal mol}^{-1} \text{ ps}^{-1}$ .

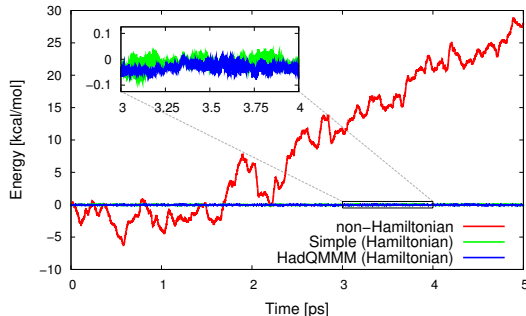


Figure 5: (color online). Total energy of the non-Hamiltonian [dark grey (red)], Hamiltonian with simple constant correction [light grey (green)], and HadQMMM [black (blue)] simulations against simulation time.

## 5 Conclusions

In summary, we propose a new Hamiltonian adaptive dual-resolution approach that combines two many-body potentials, and is able to correctly describe the structure of a molecular solution while simultaneously conserving the total energy. The forces in a straightforward Hamiltonian simulation (**DD: unclear, what is a straightforward Hamiltonian simulation**) consist of two terms,  $F_i^{ad}(\mathbf{r})$  and  $F_i^{tr}(\mathbf{r})$ . The former term by itself provides the correct equilibrium solvent structures, but the latter term causes artifacts. A crucial step in our derivation is the separation of the transition force  $F_i^{tr}(\mathbf{r})$  into characteristic contributions from each solvent molecule. We then introduce a 'per molecule' correction term that is a function of the QM character of the molecule and that cancels the undesirable transition force. When we extract the correction term from a non-Hamiltonian adaptive QM/MM simulation, the resulting Hamiltonian can be used in simulations that conserve energy and preserve the solvent structure throughout the molecular system. The new approach (HadQMMM) thus enables microcanonical simulations that provide meaningful fluctuations

and response functions, and can be used to obtain vibrational spectra and other dynamical properties.

## Acknowledgement

The research was supported by the Netherlands Organization for Scientific Research (NWO) (Vidi 723.012.104). The authors thank T. Jiang for fruitful discussions. RP and DD acknowledge financial support under the project SFB-TRR146 of the Deutsche Forschung Gemeinschaft.

## References

- (1) J. L. Gao and X. F. Xia, *J. Am. Chem. Soc.* **115**, 96679675 (1993).
- (2) J. L. Gao, *Acc. Chem. Res.* **29**, 298305 (1996).
- (3) K. Park, G. A. W., R. C. Walker, and F. Paesani, *J. Chem. Theory Comput.* **8**, 2868 (2012).
- (4) A. Warshel and M. Levitt, *J. Mol. Biol.* **103**, 227 (1976).
- (5) M. J. Field, P. A. Bash, and M. Karplus, *J. Comput. Chem.* **11**, 700 (1990).
- (6) A. Heyden, H. Lin, and D. G. Truhlar, *J. Phys. Chem. B* **111**, 2231 (2007).
- (7) S. Pezeshki and H. Lin, *J. Comput. Chem.* **35**, 1778 (2014).
- (8) R. E. Buló, B. Ensing, J. Sikkema, and L. Visscher, *J. Chem. Theory Comput.* **5**, 2212 (2009).
- (9) H. C. Watanabe, T. Kubař, and M. Elstner, *J. Chem. Theory Comput.* **10**, 4242 (2014).

- (10) T. Kerdcharoen, K. R. Liedl, and B. M. Rode, *Chem. Phys.* **211**, 313 (1996).
- (11) T. Kerdcharoen and K. Morokuma, *Chem. Phys. Lett.* **355**, 257 (2002).
- (12) N. Bernstein, C. Varnai, I. Solt, S. A. Winfield, M. C. Payne, I. Simon, M. Fuxreiter, and G. Csanyi, *Phys. Chem. Chem. Phys.* **14**, 646 (2012).
- (13) M. Bockmann, M. L. Doltsinis, and D. Marx, *J. Chem. Theory Comput.* **11**, 2429 (2015).
- (14) M. Praprotnik, L. Delle Site, and K. K. Kremer, *Annu. Rev. Phys. Chem.* **59**, 545 (2008).
- (15) A. Heyden and D. G. Truhlar, *J. Chem. Theory Comput.* **4**, 217 (2008).
- (16) J. H. Park and A. Heyden, *Mol. Simul.* **35**, 962 (2009).
- (17) R. Potestio, S. Fritsch, P. Español, R. Delgado-Buscalioni, K. Kremer, R. Everaers, and D. Donadio, *Phys. Rev. Lett.* **110**, 108301 (2013).
- (18) S. Pezeshki, C. Davis, A. Heyden, and H. Lin, *J. Chem. Theory Comput.* **10**, 4765 (2014).
- (19) L. Delle Site, *Phys. Rev. E* **76**, 047701 (2007).
- (20) E. P. P., *Annalen de Physik* **369**, 253287 (1921).
- (21) S. Pezeshki and H. Lin, *Mol. Simul.* **41**, 168 (2015).
- (22) R. E. Bulo, C. Michel, P. Fleurat-Lessard, and P. Sautet, *J. Chem. Theory Comput.* **9**, 5567 (2013).
- (23) K. Kreis, D. Donadio, K. Kremer, and R. Potestio, *EPL* **108**, 30007 (2014).
- (24) T. Darden, D. York, and L. Pedersen, *J. Chem. Phys.* **98**, 10089 (1993).

- (25) Y. Wu, H. L. Tepper, and G. A. Voth, *J. Chem. Phys.* **124**, 024503 (2006).
- (26) U. W. Schmitt and G. A. Voth, *J. Chem. Phys.* **111**, 9361 (1999).
- (27) ADF2010, SCM, Theoretical Chemistry, Vrije Universiteit, Amsterdam, The Netherlands, <http://www.scm.com>.
- (28) G. te Velde, F. M. Bickelhaupt, E. J. Baerends, C. Fonseca Guerra, S. J. A. van Gisbergen, J. G. Snijders, and T. Ziegler, *J. Comput. Chem.* **22**, 931 (2001).
- (29) S. Bahn and K. W. Jacobsen, *Comput. Sci. Eng.* **4**, 56 (2002).
- (30) J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kal, and K. Schulten, *J. Comput. Chem.* **26**, 1781 (2005).
- (31) F. Maseras and K. Morokuma, *J. Comput. Chem.* **16**, 1170 (1995).
- (32) D. Bakowies and W. Thiel, *J. Phys Chem.* **100**, 10580 (1996).
- (33) J. Kirkwood, *J. Chem. Phys.* **3**, 300 (1935).