Density-Functionalized QM/MM Delivers Chemical Accuracy For Solvated Systems

Xin Chen,[†] Jessica A. Martinez B.,[‡] Xuecheng Shao,[¶] Marc Riera,[§] Francesco Paesani,[§] Oliviero Andreussi,[∥] and Michele Pavanello[*,†]

E-mail: m.pavanello@rutgers.edu

Abstract

We present a reformulation of QM/MM as a fully quantum mechanical theory of interacting subsystems, all treated at the level of density functional theory (DFT). For the MM subsystem, which lacks orbitals, we assign an ad hoc electron density and apply orbital-free DFT functionals to describe its quantum properties. The interaction between the QM and MM subsystems is also treated using orbital-free density functionals, accounting for Coulomb interactions, exchange, correlation, and Pauli repulsion. Consistency across QM and MM subsystems is ensured by employing data-driven, many-body MM force fields that faithfully represent DFT functionals. Applications to water-solvated systems demonstrate that this approach achieves unprecedented, very rapid convergence to chemical accuracy as the size of the QM subsystem increases. We validate the method with several pilot studies, including water bulk, water clusters (prism hexamer and pentamers), solvated glucose, a palladium aqua ion, and a wet monolayer of MoS_2.

Significance

QM/MM is a powerful computational method used to model a critical, small portion of a complex molecular system—such as a protein's active site–using quantum mechanics, while treating the surrounding environment with classical force fields. While QM/MM has advanced our understanding of enzymes and biological systems, it often struggles with accuracy, even when the QM and MM regions are not covalently bonded. A notable challenge is the slow convergence of system-environment interaction energies as the size of the QM region increases. This work demonstrates that incorporating quantum mechanics in the description of the MM subsystem leads to dramatically improved models. By assigning the MM subsystem a physically meaningful electron density, and using ab-initio density functionals for the QM-MM interaction (accounting for exchange, correlation, and Pauli repulsion), chemical accuracy in QM/MM models of aqueous solutions is achieved for the first time.

Introduction

QM/MM (standing for quantum mechanics/molecular mechanics) has revolutionized computational biochemistry.[1] Since the pioneering work of Honig and Karplus,[2] the combination of a quantum mechanical (QM) description for a subsystem with a classical point-charge description of its environment has led to major breakthroughs in fields such as enzymatics,[3,4] drug development,[5,6] and materials design.[7] Since its conception, methods handling the QM and MM subsystems have dramatically evolved. Today's MM force fields can integrate data-driven potentials,[8] polarizable models,[9–15] and even machine learning techniques.[16,17] QM methods have also evolved dramatically, from DFT methods to wavefunction theory methods routinely used in conjunction with QM/MM.[18]

The nature of the QM-MM interaction has also evolved. Initially, these were handled mechanically,[3] with Coulomb interactions calculated *a posteriori*, influencing only the forces and total energy, but without affecting the QM wavefunction or density. The advent of

electrostatic embedding improved accuracy by incorporating MM partial charges directly into the QM Hamiltonian.[19] Ultimately, mutual QM-MM polarization was achieved using polarizable force fields,[9,13] a concept anticipated in early QM/MM work.[20]

The computational cost of QM/MM simulations is greatly reduced compared to fully quantum mechanical treatments. However, incorporating QM / polarizable-MM interactions in an efficient manner remains challenging. Methods based on judicious partitioning of the induction response have demonstrated excellent scalability.[14,21] Additionally, algorithmic advances have been supported by steady progress in software development.[17,22–28] To extend QM/MM simulations to condensed phases, periodic boundary conditions (PBC) have been implemented. Ewald summation techniques are commonly used, and adaptations for molecular condensed phases[29–32] and material systems[26,28,33,34] are now widely available.

How accurate are QM/MM models? A common way to address this question is to evaluate the convergence of the results with respect to the size of the QM subsystem. Unfortunately, generally such a convergence is slow. Protein environments, for example, are exceptionally complex, and the search for effective ways to include relevant protein regions in QM/MM simulations continues.[35–38] Particularly challenging has been capturing charge transfer interactions on larger scales.[37,39]

A slow convergence of the QM/MM setup has also affected those systems where partitioning in QM and MM subsystems involves no bond breaking. For example, water solvation.[40] Ironically, independent QM-only or MM-only treatments of liquid water can provide accurate results, but their combination in QM/MM workflows results in an overall reduced accuracy.[26,41] Accurately modeling aqueous environments with QM/MM is essential due to the need to consider large water environments to properly account for the static and dynamic responses at water-material interfaces.[42,43] Therefore, representing these polarization effects in water bulk, which can extend for several nanometers, is crucial for capturing significant effects on the energetics of solvated species.[44,45]

The culprit is the difficulty to accurately capture QM-MM interactions with a compu-

tationally efficient method. The MM subsystem is typically described using methods that largely (or completely) neglect its electronic structure. Point charges or, at times, point polarizable dipoles do not faithfully represent any electronic structure! The polarizable embedding method[46] tackles this problem by dividing the MM subsystem into two regions: one near the QM subsystem is assigned a QM density derived from isolated fragment calculations, while the remaining MM atoms are treated using conventional point charge or dipole models. This embedding approach improves the accuracy of the QM Hamiltonian by capturing both electrostatic and non-electrostatic interactions, leading to better results than traditional QM/MM setups.[47] Similar methods, such as QM/ESP,[48] QM/GEM,[49] and QXD,[50] as well as approaches in density embedding[51] and many-body expansions,[52] further explore these concepts. While attempts to account for the purely quantum mechanical Pauli repulsion within QM/MM have yielded mixed outcomes,[53] some methods address this by parametrizing the mechanical embedding interaction energy without introducing new terms in the QM Hamiltonian.[54–56]

Thus, our approach in this work is to treat QM and MM subsystems on a more equal basis, aiming to reduce the impact of an imbalanced QM-MM interface and an imbalanced treatment of the internal energy of QM and MM subsystems. We propose "density-functionalizing" the MM subsystem, assigning it an electron density such that it can be handled like an electronic subsystem within the rigorous framework of subsystem DFT.[57–61] This standardization of QM and MM subsystems allows for the use of first-principles density functionals for evaluating the QM-MM interaction, inherently capturing all relevant physical effects, such as exchange, correlation, Pauli repulsion, electrostatics, and charge penetration. The next section details the theoretical framework for this density-functionalized QM/MM approach, with additional, less critical details provided in the supplementary materials.[62]

A density-functionalization of QM/MM

The central idea is to assign an electron density to both the QM subsystem, $\rho_{QM}(\mathbf{r})$, and to the MM subsystem, $\rho_{MM}(\mathbf{r})$, with the total electron density given by their sum and the energy functional borrowed from rigorous subsystem DFT[57–59] (sDFT, hereafter),

$$\rho(\mathbf{r}) = \rho_{QM}(\mathbf{r}) + \rho_{MM}(\mathbf{r}), \tag{1}$$

$$E[\rho_{QM}, \rho_{MM}] = E[\rho_{QM}] + E[\rho_{MM}] + E^{nad}[\rho_{QM}, \rho_{MM}]. \tag{2}$$

Although formally the electronic energy is strictly a density functional, in practice, the external potential (electron-nuclear attraction) is known ahead of time and is thus specified for the QM subsystem, $v_{QM}(\mathbf{r})$, and for the MM subsystem, $v_{MM}(\mathbf{r})$, such that the additive part of the energy is given by (disregarding for the time being the nuclear-nuclear repulsion)

$$E[\rho_{QM}] = T_s[\rho_{QM}] + E_H[\rho_{QM}] + E_{xc}[\rho_{QM}] + \int v_{QM}(\mathbf{r})\rho_{QM}(\mathbf{r})d\mathbf{r}, \tag{3}$$

$$E[\rho_{MM}] = T_s[\rho_{MM}] + E_H[\rho_{MM}] + E_{xc}[\rho_{MM}] + \int v_{MM}(\mathbf{r})\rho_{MM}(\mathbf{r})d\mathbf{r}, \tag{4}$$

where T_s, E_H and E_{xc} are the noninteracting kinetic energy, the classical electron-electron repulsion (Hartree) and the exchange-correlation functionals, respectively. The nonadditive energy is thus given by

$$E^{nad}[\rho_{QM}, \rho_{MM}] = T_s^{nad}[\rho_{QM}, \rho_{MM}] + E_{xc}^{nad}[\rho_{QM}, \rho_{MM}] + \tag{5}$$

$$+ \int \frac{\rho_{QM}(\mathbf{r})\rho_{MM}(\mathbf{r})}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}d\mathbf{r}' + \int \rho_{MM}(\mathbf{r})v_{QM}(\mathbf{r})d\mathbf{r} + \int \rho_{QM}(\mathbf{r})v_{MM}(\mathbf{r})d\mathbf{r}$$

where $T_s^{nad}[\rho_{QM}, \rho_{MM}] = T_s[\rho_{QM} + \rho_{MM}] - T_s[\rho_{QM}] - T_s[\rho_{MM}]$ and equivalently for E_{xc}^{nad}.

The equations above can be exploited for a variational minimization of the energy functional with respect to variations in both ρ_{QM} and ρ_{MM}, provided that suitable approximations for the relevant density functionals are available. When both subsystems are treated

at the Kohn-Sham DFT level, this approach yields sDFT, which was found to be accurate in the limit of weak inter-subsystem interactions. This is the case for systems such as water molecules in liquid water[63] or CO_2 molecules in fluid CO_2.[64] In fact, sub-1 kcal/mol accuracy is now routinely achieved in sDFT calculations either with Kohn-Sham subsystems,[65–67] orbital-free subsystems,[68] and proper multiscale simulations.[69–72] Inter-subsystem interactions involving hydrogen bonds, are the focus of this work and are particularly well described by the nonadditive PBE exchange-correlation and revAPBEk nonadditive kinetic energy functionals[65] (and in general many GGA nonadditive functionals[73]).

In a QM/MM framework, Eq. (4) is usually replaced by a classical force-field expression, which involves electrostatic interactions between atom-centered point charges and possibly atom-centered polarizable dipoles, as well as (typically empirical) expressions for short range dispersion-repulsion interactions and bonded terms. In common implementations of QM/MM approaches, the tool associated with the classical component is responsible for the calculation of this energy term. However, the electrostatic multipoles present in the MM force-field are also involved in the electrostatic terms of the QM-MM interaction energy term in Eq. (5). Usually, ad-hoc corrections to describe short-range dispersion-repulsion interactions and avoid unphysical behaviors when the QM and MM subsystems are too close are introduced to approximate the first two terms of Eq. (5). Once the analytic dependence of Eq. (2) on the atomic positions of the MM subsystem is defined, the functional derivative of Eq. (2) with respect to $\rho_{QM}(\mathbf{r})$ allows the optimization of the QM subsystem via a self-consistent field (SCF) approach.

Point-charges and point-dipoles introduced in the definition of the MM part of the system can be tuned to fit empirical results for the MM system, or they can be optimized to reproduce properties connected to the electronic density of the MM components computed from first principles (e.g., binding energies or the behavior of the electrostatic potential). Because the permanent charges in MM force fields are independent of the geometry of the system and of its surrounding environment, atomic polarizabilities may be included in the

force field in order to capture the system's electrostatic response to external fields. The resulting polarizable force fields will, therefore, respond to the non-additive term in Eq. (5) as it contributes to the induced dipoles in the MM subsystem. This results in mutually polarized QM and MM subsystems.

The typical approximations in the above approach lead to significant inaccuracies when the boundary between QM and MM regions varies systematically. Drawing from existing QM/MM methods that account for electron densities in the MM subsystem[46,74] and the successes of the sDFT framework, we hypothesize that a QM/MM framework based on Eqs. (2–5) can accurately model solute-solvent systems. This extends to weakly interacting QM and MM subsystems, provided consistent forward and backward mappings between first-principles electronic densities and force field multipoles are established.

For the forward mapping between electronic-structure calculations and accurate classical electrostatics, the many-body polarizable approach by Paesani and collaborators showed that it is possible to effectively describe statistical properties of bulk systems by carefully parametrizing a classical force-field on a large database of accurate few-body first-principles simulations.[75,76] While the initial developments of MB-Pol focused on high-accuracy coupled-cluster simulations of water dimers and trimers, the approach can be applied to DFT-based calculations[77] giving rise to, e.g., the MB-PBE force field. The resulting force-fields describe electrostatic interactions in terms of atomic charges and polarizabilities. As in similar approaches in the literature, short-range corrections are introduced to avoid self-polarization and the unphysical divergence of the induced dipoles when different polarizable systems are too close to each other. In this work, we hypothesize that using an MM subsystem whose electrostatic interactions are fully consistent with the level of theory of the QM component will allow the seamless convergence of QM-MM calculations as the boundary between the two regions is varied.

For the backward mapping between the classical force-field used in the MM region and an effective electronic density to use in the density functionalized QM-MM interaction term,

7

we introduce the following approach. We treat each atomic nuclei and core electrons using pseudopotentials. To keep the approach computationally efficient for large MM subsystems, we use the local part of the ultrasoft pseudopotentials from Garrity and Vanderbilt (GBRV),[78] which are designed for fast, high-throughput simulations that require low plane wave cutoffs. This assignment can be efficiently handled in a computationally linear scaling manner using the particle-mesh Ewald method.[79] For the valence electrons, we convert the point-like charge multipoles computed and used by the classical force-field into a smooth electronic charge density. Namely, for each atom i, the valence electron density is represented by a Gaussian centered at the ion's position, $\mathbf{R}_i$, with an adjustable width σ_i: $\rho_{q_i}(\mathbf{r}) = (N_i - q_i)g_{\sigma_i}(\mathbf{r} - \mathbf{R}_i)$, where g_{σ_i} is a normalized Gaussian and the prefactor is crucial for accurately representing the ion's permanent charges, q_i, while accounting for the isolated atom's number of valence electrons, N_i. If the force-field involves higher multipoles in the description of its electrostatic interactions, they can be included in the reconstructed electron density by using derivatives of the normalized Gaussian. In particular, we can map a point-like dipole into a corresponding smooth density using the gradient of a Gaussian function as $\rho_{\mu_j}(\mathbf{r}) = -\vec{\mu_j} \cdot \vec{\nabla} g_{\sigma_j}(\mathbf{r} - \mathbf{R}_j)$, where $\vec{\mu_j}$ is the induced dipole at the dipole site j. Similarly, higher-order multipoles can be mapped using higher-order derivatives. The total valence electron density becomes

$$\rho_{MM}(\mathbf{r}) = \sum_{i \in \text{MM charges}} \rho_{q_i}(\mathbf{r}) + \sum_{j \in \text{MM dipoles}} \rho_{\mu_j}(\mathbf{r}). \tag{6}$$

We stress that the proposed formulation relies on a set of element-specific widths (the σ_i) that can significantly affect the final shape of the reconstructed electronic density (both permanent charges and polarization density due to the induced dipoles).

For the forward mapping, we hypothesize that using the Gaussian widths that provide the best match between QM/MM vs QM/QM (sDFT) interaction energies will allow for the accurate description of the mutual polarization effects and the seamless convergence of

QM/MM calculations as a function of an increasing size of the QM subsystem.

The functional derivatives of the non-additive interaction energy with respect to the QM or MM subsystem densities yields the embedding potentials

$$v_{emb}^{QM}(\mathbf{r}) = \frac{\delta E^{nad}[\rho_{QM}, \rho_{MM}]}{\delta \rho_{QM}(\mathbf{r})} = v_{MM}(\mathbf{r}) + \int d\mathbf{r}' \frac{\rho_{MM}(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} + \frac{\delta T_s^{nad}}{\delta \rho_{QM}(\mathbf{r})} + \frac{\delta E_{xc}^{nad}}{\delta \rho_{QM}(\mathbf{r})}, \qquad (7)$$

$$v_{emb}^{MM}(\mathbf{r}) = \frac{\delta E^{nad}[\rho_{QM}, \rho_{MM}]}{\delta \rho_{MM}(\mathbf{r})} = v_{QM}(\mathbf{r}) + \int d\mathbf{r}' \frac{\rho_{QM}(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} + \frac{\delta T_s^{nad}}{\delta \rho_{MM}(\mathbf{r})} + \frac{\delta E_{xc}^{nad}}{\delta \rho_{MM}(\mathbf{r})}, \qquad (8)$$

that can be used to optimize the QM or MM degrees of freedom. In particular, Eq. (8) enters the Hamiltonian of the QM subsystem(s) at each SCF step and it is used for the calculation of the ground state QM electronic densities as well as for the optimization of QM atomic positions. The sDFT framework and the implementation of this approach in the eDFTpy software[80] allows for the coupling of the QM/MM embedding potential into multiple QM subsystems.

Eq. (7) instead can be used by the MM engine to compute the QM effects on the induced dipoles and on the interatomic forces. For the former, the gradient of the MM embedding potential needs to be added to the classical electric field used to compute induced dipoles. While all the terms in Eq. (7) would ideally be included in the embedding field that polarizes the MM component, the current definition of atomic polarizabilities within polarizable force fields is only meaningful for a classical description of long-range polarization effects. Thus, in the current implementation, we only keep the first two terms, which are related to classical electrostatic interactions, and we neglect the effect of the non-additive kinetic and exchange-correlation terms, which are more relevant for short-range interactions.

The evaluation of the QM additive energy in (3) can utilize the sDFT implementation of eDFTpy with either a single QM subsystem or multiple QM subsystems. When N_S QM subsystems are involved, the QM electron density is represented as the sum of contributions from each subsystem, and the QM energy is decomposed into additive and nonadditive terms,
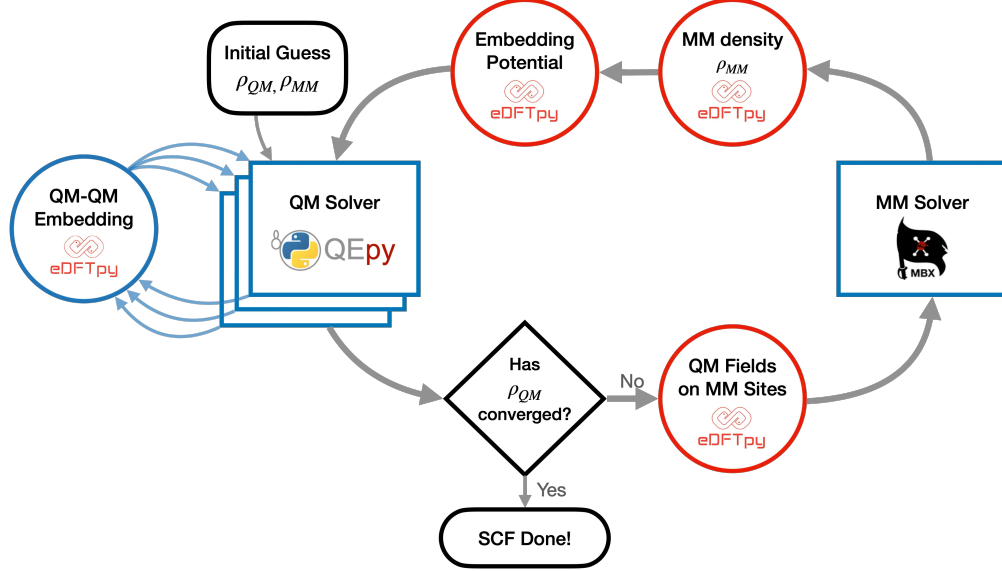
9

Figure 1: Workflow of the QM/MM method with emphasis on the software implementation. See details in section .

following the structure of (2). Namely,

$$\rho_{QM}(\mathbf{r}) = \sum_{I=1}^{N_S} \rho_I(\mathbf{r}), \tag{9}$$

$$E[\rho_{QM}] = \sum_{I=1}^{N_S} E[\rho_I] + E^{nad}[\{\rho_I\}]. \tag{10}$$

In the results Section , we will refer to a one-subsystem treatment of the QM region as fragmentation type 1 (or Frag. 1) while a many-subsystem treatment as fragmentation type 2 (or Frag. 2).

Details of the implementation

We now discuss the details of the implementation of the QM/MM algorithm in the eDFTpy software.[80] Figure 1 illustrates the QM/MM workflow of the SCF cycle in the presence of a MM subsystem. It begins with initial guesses for the MM and QM density and wavefunctions. In eDFTpy, we use QEpy[81] as QM solver (a python implementation of Quantum ESPRESSO 7.2[82]) and a python interface to MBX as the MM solver.[83] The initial guess density for the

QM subsystem is usually taken from the sum of atomic densities from the pseudopotential files. For the MM subsystem, a valence electron density consistent with the value of the permanent charges (which are fixed, independent of geometry) is used.

eDFTpy then directs QEpy, the QM solver, to solve for the electronic structure of the QM system. This is done either by a single QEpy instance when a single QM subsystem is considered, or by multiple QEpy instances when several QM subsystems are considered. The QM subsystems require the computation of QM-in-QM embedding potentials which are handled also by eDFTpy as indicated by the blue arrows connecting the QEpy solvers to the eDFTpy circle in the figure. The QM electrostatic field formally given by the negative gradient of the MM embedding potential in Eq. (7), $-\vec{\nabla} v_{emb}^{MM}(\mathbf{r})$ is then represented on the MM polarizable dipole sites. For large systems, representing the QM electric field on the MM sites is a task of non-negligible computational complexity.[33,34,84,85] The QM electric field is then represented on the grid point closest to each of the MM polarizable dipole sites. This is accurate enough given the exceptional smoothness of the QM field in the MM region[34] and saves the inconvenient step of splining the field on the exact position of the ions. Spline would require an `MPI_GATHER` operation which carries a high wall-time cost. In eDFTpy, the MM cell is spanned by a grid fine enough to properly represent the ionic pseudopotentials centered on each QM ion and MM site. In our simulations, we use a 100 Ha cutoff for this grid, which is fine enough to produce real-space grid points spaced by about 0.23 a_0.

Once the electric field is represented on the MM sites, the MM solver is tasked with solving the Coulomb problem in the MM cell to yield the MM polarizable dipoles. The MM density is then built using Eq. (6). Having the MM density and the density of all QM subsystems allows eDFTpy to compute the required embedding potentials for each subsystem, including the MM subsystem. The SCF cycles then continue until convergence is achieved, which is calibrated on the convergence of the QM density. At convergence, QM density and MM dipoles are fully mutually polarized.

In the eDFTpy implementation we take full advantage of plane wave reduction techniques

that were developed for sDFT.[68,69,86–88] Specifically, MM and QM subsystems are represented by different simulation cells, grids and thus plane wave basis sets. The QM simulation cell is smaller compared to the physical cell which coincides with the MM cell. Coulomb fields and other long-ranged energy functionals are evaluated on the large, physical cell. Employing such a multi-cell/grid approach is crucial in the context of QM/MM simulations as the MM subsystem is usually dramatically more extended than any of the QM subsystems. In the supplementary materials[62] we further discuss our parallelization strategy.
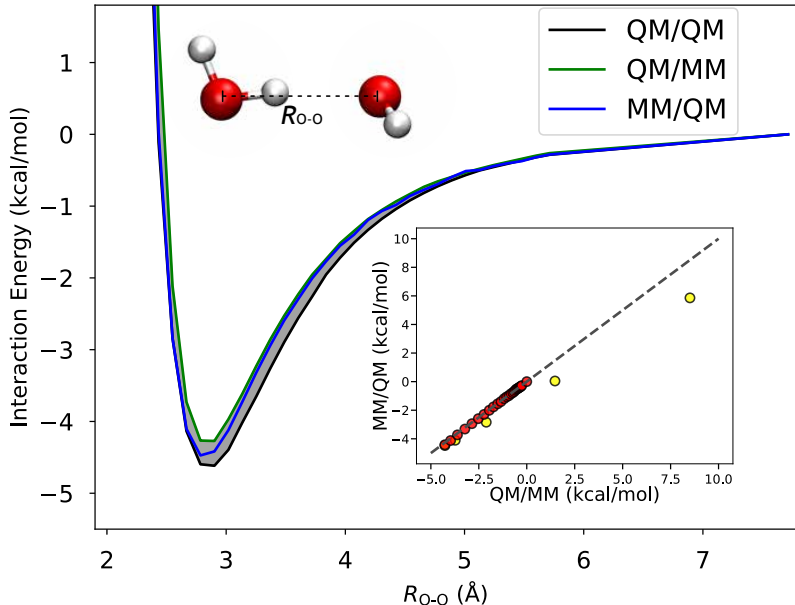

Figure 2: Water dimer energy curve (structure shown) for O–O distances ranging from 2.3 to 7.7 Å. The black line represents the QM/QM result, where both the donor and acceptor monomers are treated at the QM level. The shaded area represents the deviation of the QM/MM result from QM/QM. Green and blue lines correspond to the QM hydrogen bond donor (QM/MM) and acceptor (MM/QM), respectively. The inset shows a correlation plot of QM/MM and MM/QM interaction energies. Yellow markers represent geometries in the repulsive region of the curve ($R_{O\text{-}O} < 2.9$ Å), while red markers correspond to those in the attractive region ($R_{O\text{-}O} > 2.9$ Å).

The current implementation also features analytic energy gradients for the atoms in the QM region. In the supplementary materials we devote a section to the implementation of the forces and results are shown in Tables **??** and **??**.

Computational Details

Pseudopotentials, Density Functionals and Plane Wave Cutoffs

GBRV pseudopotentials[78] are used for all elements considered of both the QM and MM subsystems. The PBE exchange-correlation functional[89] and the revAPBEk noninteracting kinetic energy functional[90] are employed for approximating additive E_{xc} and nonadditive E_{xc} and T_s functionals. All calculations include Grimme's D3 correction.[91] We choose plane wave basis sets for the QM subsystems with a cutoff energy of 20 Ha for wavefunctions and 200 Ha for charge density and potential unless otherwise stated (see supplementary Table **??**). The energy convergence threshold for the SCF was set to 10^{-8} Ha/atom. The Brillouin zone was sampled by a $5 \times 5 \times 1$ k-point mesh for MoS_2 and the Γ point for all other calculations. MM calculations were conducted using the MBX package[83] using the MB-PBE and MB-Pol water models, which were developed to quantitatively reproduce PBE or CCSD(T) water,[77,92] respectively.

All inputs/output files, Jupyter notebooks needed to analyze the data and reproduce the figures in this work, as well as links to tagged versions of the software (MBX, eDFTpy, QEpy, and DFTpy) used for the simulations, are available as reported in the supplementary materials.[62]

Parameters Defining the MM Density

As described in the sections above, the proposed framework relies on the conversion of classical permanent charges and polarizable dipole sites into a smooth electronic density. For each atomic charge and polarizable dipole, this involves the fit of the width, σ, of the corresponding normalized Gaussian functions that contribute to the expansion in Eq. (6). In our applications to solvated systems, MM sites only involve oxygen and hydrogen atoms, for a total of 4 parameters that need to be fitted.

We also considered an additional MM-induced dipole self-energy correction. Although

the induced point-dipole self-energy is given by $\frac{1}{2}\mu^2\alpha^{-1}$,[93] where α is the isotropic dipole polarizability, we recognize that the dipoles considered in our work are not point dipoles, e.g., their charge density overlaps with the QM density at the QM/MM interface. Thus, we added an additional term to the self-energy for each site equal to $k_i^{SE}|\vec{\mu} - \vec{\mu}'|^2$, where $\vec{\mu}'$ is the induced dipole at the same site when only the MM subsystem is considered, and k_i^{SE} are element-dependent proportionality constants. Additional details about this correction and about how the permanent charge density was generated for the MB-PBE and MB-Pol force fields (which use the so-called M-site) are available in the supplementary materials document.[62]

The parameters defining the MM density were fitted so as to reproduce the QM-QM interaction energies for a single water molecule in bulk water. Ten snapshots of 64-water molecule cubic systems were taken from Ref. 94. These provided 640 water-bulk interaction energies. More details about this system will be given in the results section. The final values of the parameters for both MB-PBE and MB-Pol force fields are listed in Table **??**.

The use of bulk simulations as a reference for the parametrization of the density functionalization approach provides a robust and general strategy that reduces the need for application-specific benchmarks. Results on small water clusters and solvation effects on molecules and materials (reported in the following) highlight the transferability of the obtained parameters.

Results and Discussion

Water Dimer

We start by comparing the QM/MM potential energy curves shown in Figure 2 and the polarization density depicted in Figure 3 for water dimers. The dimer structures, sourced from Ref. 95, were placed in a 20-Å cubic simulation cell and evaluated against benchmark QM/QM simulations performed using sDFT.
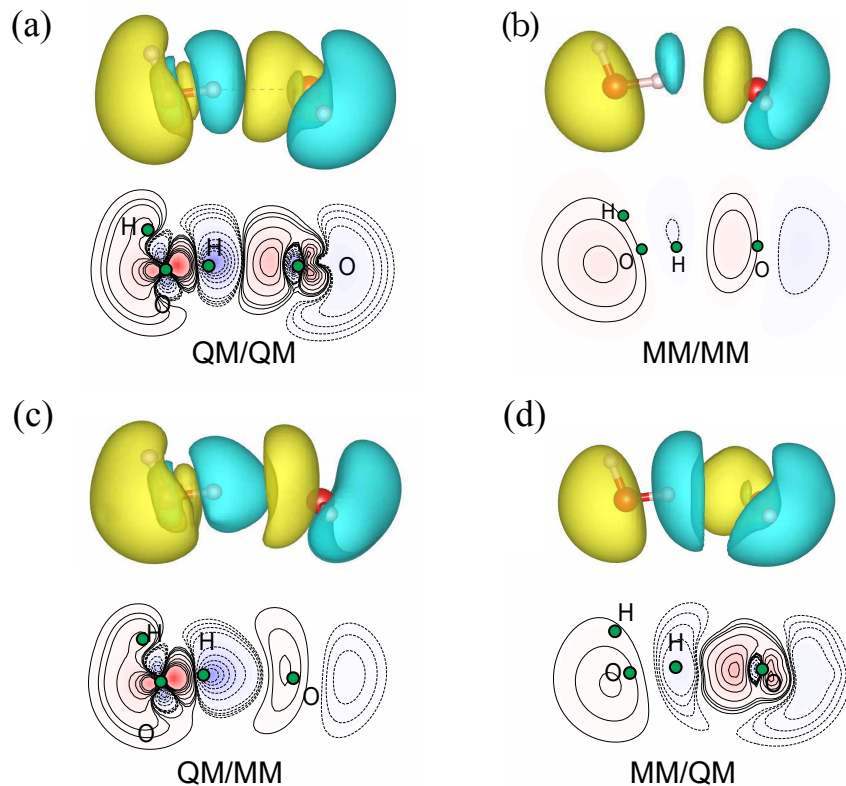
Figure 3: Polarization density, defined as $\rho(\mathbf{r}) - \rho_{\mathrm{iso}}(\mathbf{r})$, where ρ_{iso} is the sum of the electron densities of the isolated water monomers, for the water dimer at the equilibrium O–O distance. In each panel, we present isosurfaces (top) and contour plots (bottom) generated with a cutoff of $\pm 0.0007\, e \cdot a_0^{-3}$.

The key aspect of this system lies in its asymmetry, as one monomer is a hydrogen bond donor, and the other is an acceptor. Even though both QM and MM monomers are modeled by PBE (the MM subsystem is described by the MB-PBE model), the nature of the QM-MM interface is dramatically different whether one considers a QM hydrogen bond donor (QM/MM) or acceptor (MM/QM). Despite this, the curves are remarkably similar. The QM/MM and MM/QM minima (at 2.90 Å O–O distance) are less than 0.15 kcal/mol away from each other, and both deviate from the reference by a similar measure. These results show that the QM/MM interface is extremely well characterized by the nonadditive functionals. Crucially, the repulsive part of the curve is also well reproduced. Our ∼0.2 kcal/mol error compares well with the ∼0.6 kcal/mol error for the dimer reported in Ref. 41 for their QM/MM (AMOEBA) method.

In the inset of Figure 2, we report the correlation between the interaction energies of the QM/MM and MM/QM systems, showing that they are in extremely good agreement. The interaction energies in the attractive part of the curve follow the ideal trend, i.e., the points sit on the diagonal line. However, those in the repulsive part (yellow color coded markers) show a slightly deviated trend. A behavior indicative of an asymmetry in accounting for charge penetration effects of the H_{QM}-O_{MM} vs H_{MM}-O_{QM}. To ensure our results are not affected by possible artifacts due to the use of PBCs, we recalculated the interaction energy minima in a larger simulation cell with lattice constant $a = 50$ Å finding a shift of less than 10^{-3} kcal/mol.

Figure 3 shows the polarization density (defined in the caption) of the water dimer system. It is clear that the MM polarization is only qualitatively similar to the QM polarization. Specifically, comparing the QM/QM (panel a) with the MM/MM (panel b) shows that the latter misses several potentially important details near the ion's core and along the O–O line. These are well-known deficiencies in the electrostatic response of dipole-only polarizable force fields.[96,97] Interestingly, the QM/MM treatment improves the polarization of both monomers. Panels (c) and (d) show that the MM molecule polarization gains features that are prominent in the QM response of that monomer but that are absent (or weakly present) in the MM/MM case. This shows that the handling of the QM/MM interface by our method is accurate not only by the interaction energy measure (recall Figures 2) but also by the much more stringent test of subsystem mutual polarization.

Water Hexamer

Hexamer water structures play a crucial role in quantum chemistry due to their complex hydrogen-bond topologies. Accurately predicting their relative energies is often seen as a benchmark for a model's ability to represent water across its various stable phases.[98] Here, following Ref. 8, we focus on the most stable hexamer, the prism hexamer, to examine the effect of the QM-MM boundary. In the prism hexamer all water monomers act as both

hydrogen bond donors and acceptors, with either 1/2 or 2/1 accepted/donated hydrogen bonds.

The root mean square errors (RMSEs) of the computed interaction energies defined as the energy of the hexamer minus the energy of the 6 isolated water molecules are collected in Figures **??** and **??** for the MM treated with MB-PBE and MB-Pol, respectively. When multiple water molecules are treated at the QM level, the sDFT framework allows flexibility in how the QM waters are grouped: they can be combined into a single subsystem (Frag. 1 in the figure) or divided into separate subsystems (Frag. 2). For each partition type with k QM or MM water molecules, there are $\binom{6}{k}$ possible members of the partition (i.e., ways to split the hexamer into QM and MM subsystems).

Ideally, all members of each partition should yield the same interaction energy. Therefore, a larger RMSE for the predicted interaction energy indicates a lower accuracy of the method. Recognizing that no practical fragmentation method is perfect, we find a relation between the RMSE for the k-th partition ($\sigma_{\mathrm{QM/MM}}(k)$) and the square root of the number of members in the partition[99] ($\sqrt{\binom{6}{k}}$). This relationship is confirmed by R^2 values of 0.99 for all QM/MM and fragmentation methods considered (see Figure **??**). The slope of this linear relationship is proportional to the average error per QM/MM boundary in the partitions (error = slope $\cdot \sqrt{\frac{\pi}{2}}$), yielding a predicted error per boundary ranging between 0.2 and 0.3 kcal/mol for the methods considered. This error is consistent with the results for water dimers and, as we will see in the next section, also aligns with the RMSEs for the pentamer clusters extracted from bulk liquid structures.

Bulk and First Solvation Shell Water Environment

To further assess our method, we consider the dipole moment and the molecule-environment interaction energy of water molecules embedded in an MM environment of bulk liquid water. The benchmark is once again sDFT. Dipole moments of embedded molecules are accessible in sDFT, calculated with the subsystem electron density. This procedure was found to be

accurate for a variety of embedded molecular species.[57,94,100,101] We consider 10 snapshots of an ab-initio dynamics trajectory of 64 independent water molecules in a cubic cell of lattice constant 12.42 Å that was presented elsewhere[63,94] (also available in the supplementary materials[62]). We use the nomenclature n/m to denote a simulation where n molecules are treated at the QM level and m are treated at the MM level. In all cases, the reported quantity is the interaction of a single water molecule with the environment. For example, 1/63 means there are 63 MM water molecules and one QM water, for a total of 64 water molecules. We use sDFT to generate benchmark values for the water-environment interaction energy. Within a sDFT framework, one may treat all 63 environment molecules as a single subsystem or as 63 coupled subsystems (for the 1/4 system we only consider 4 environment water molecules). Owing to the accuracy of sDFT for bulk water simulations,[45,63] we employed both approaches, which resulted in very similar trends, as reported in Figure ?? of the supplementary materials. For the discussion in the following, we consider the benchmark where all 63 molecules in the environment are grouped in a single subsystem. Leveraging the available bulk configurations, inspired by Ref. 41, we also considered the performance of our QM-MM approach on "first shell" structures, where each water molecule of the bulk is embedded only by the nearest 4 water molecules of the environment, giving rise to a collection of water pentamer structures (denoted by 1/4). We also test our QM/MM scheme on a partition "minimal solvation" where each water molecule and the nearest 4 water molecules of the environment are included in the QM subsystem and the remaining 59 water molecules are kept in the MM subsystem (system 5/59). Because in each snapshot there are 64 water molecules and there are 10 snapshots, 640 total structures / data points are available.

Panels (a) and (b) of Figure 4 show that the interaction energy between water molecules and their environment in liquid water ranges from -35 to -7 kcal/mol for the bulk calculations and from -20 to -1 kcal/mol for the first shell calculations. Such energy scales are in line with comparable calculations in the literature.[41] As indicated in the figure, the root mean
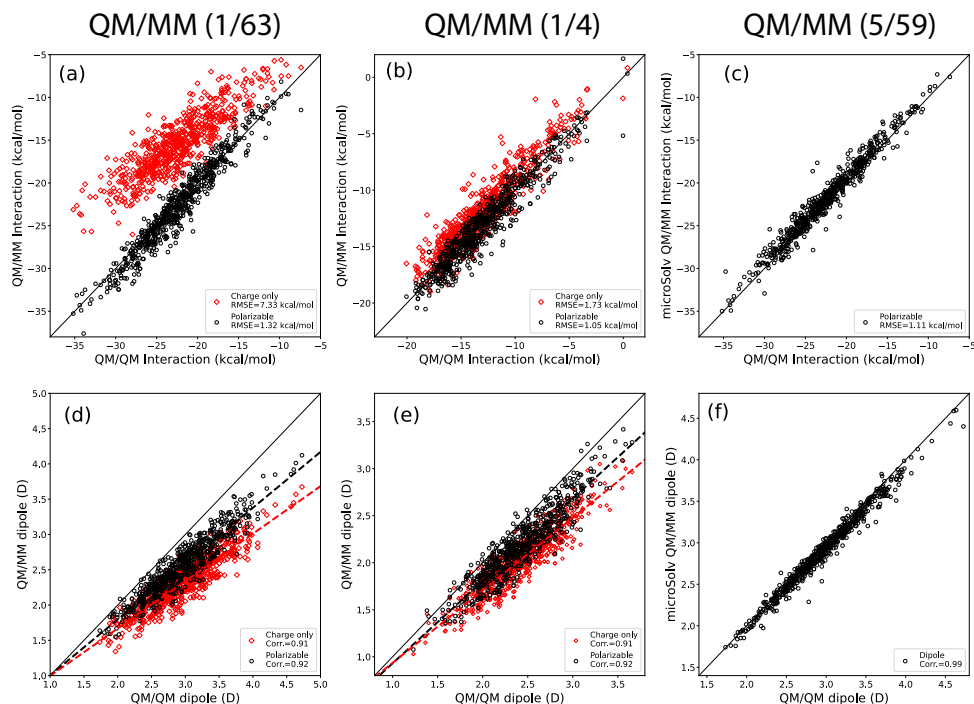
Figure 4: Panels (a), (b) and (c): correlation plots of the interaction energy (in kcal/mol) of a single water molecule with its environment in a model of liquid water. (a) Bulk: QM/MM with 1 QM water molecule and 63 MM water molecules. (b) First shell: 1 QM water molecule and 4 MM water molecules (only the first solvation shell). (c) Minimal solvation: 5 QM water molecules and 59 MM water molecules. Panels (d), (e), and (f): correlation plots of the dipole moment length of the embedded molecule (in Debye) for the same systems as for panels (a–c).

square errors (RMSEs) of the QM/MM interaction energy compared to the QM/QM ones are 1.32 kcal/mol for the bulk and 1.05 kcal/mol for the first shell calculations. The minimal solvation setup improves upon the bulk with an RMSE of 1.11 kcal/mol. The RMSE of the first shell simulations compares well with 1.4 kcal/mol from a similar simulation from Ref. 41 where MM was treated with the AMOEBA force field. To our knowledge, a similar comparison for the bulk or minimal solvation results is not available.
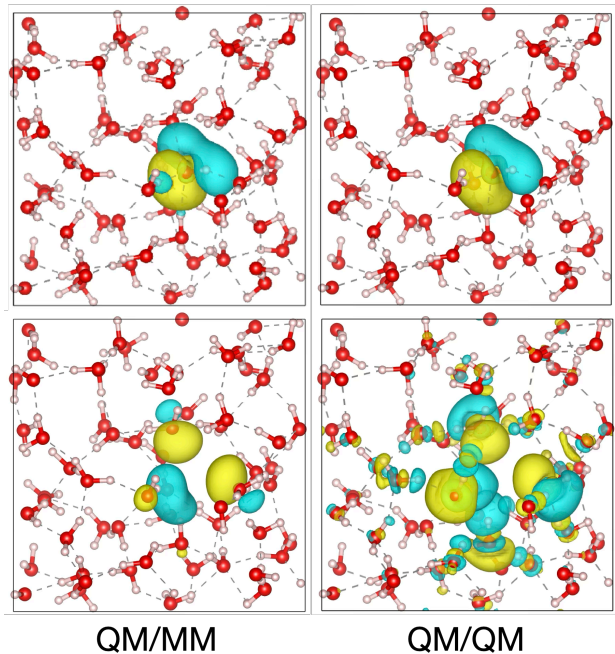

QM/MM   **QM/QM**

Figure 5: Polarization density (defined as the difference of the embedded and isolated molecular electron density) of an embedded water molecule employing the methods indicated in the figure. Top panels: single water molecule polarization. Bottom panels: environment polarization. The isosurface value is set to ± 0.0025 e·a_0^{-3}.

Such RMSEs are also consistent with the accuracy of the hexamer structure discussed before. However, ~ 1 kcal/mol accuracy for bulk systems is even more impressive. In the hexamer, the interaction energies (mbpol) average about 8.00 kcal/mol per water molecule. In the bulk water, the QM–QM interaction energy averages to a much larger value (~ 22.42 kcal/mol) due to a stronger polarization. In fact, the ~ 1 kcal/mol accuracy leads us to conclude that the QM-MM mutual polarization predicted by the QM/MM simulation in bulk and minimal solvation setups is well represented. As we can see in Figure 5, both

the polarization of the embedded water molecule (see panels a, b, and c) as well as of the environment (see panels e, f, and g) are fairly well represented by the QM/MM calculations. An interesting aspect of the MM polarization in panels (e) and (d) of Figure 5 is that, in comparison to the QM/QM polarization they only feature a small component from the second solvation shell. An explanation for this effect is the fact that polarizable force fields typically employ slightly damped atomic dipole polarizabilities compared to the reference QM value.[102,103] The damping is crucial to fend off overpolarization, which in our framework would stems from the neglect of the nonadditive components of the embedding potentials, see Eq. (7). In follow-up work, it will be interesting to analyze the effect of these additional terms on the MM polarization and the chosen values of the atomic dipole polarizabilities.

Panels (d) (e) and (f) of Figure 4 present the dipole lengths of the central, embedded water molecule. These range from 1.6 D to 4.5 D for both 1/63 and 5/59 bulk simulations, and from 1.3 D to 3.5 D for the 1/4 first-shell calculations. Generally, we notice that the dipole lengths are underestimated in the QM/MM treatment compared to the reference QM/QM. The QM/MM dipoles, however, correlate very well with the reference QM/QM (sDFT) dipoles showing Pearson's correlation coefficients above 0.9. We also computed an RMSE for the dipoles of the 1/4 calculations of 0.29 D which compare well with the 0.62 D of Ref. 41. The dipole results and the results from the interaction energies sustain our claim that our method is the most accurate QM/MM result for liquid water available to date.

Figure 4 also shows (red diamonds) results labeled as "charge only," which are from QM/MM simulations where the polarization of the MM subsystem is neglected (i.e., the polarizable dipoles are set to zero). While for the first shell systems the results are still acceptable (dipole correlation of 0.91, dipole RMSE of 0.42 D, and for the interaction energy of 1.73 kcal/mol), the results for the bulk system are qualitatively incorrect. In the bulk, the RMSE for the interaction energy jumps to 7.33 kcal/mol. An unacceptable deviation is, however, in line with the expected accuracy of a typical electrostatic embedding QM/MM simulation of condensed phases.[104,105]

When employing 63 coupled subsystems to represent the QM/QM benchmark, the interaction energies and dipole moment correlations are consistent with what is presented above, see supplementary Figure **??**. We also note that MB-Pol provides us with a very accurate benchmark for the interaction energies. We thus include a comparison of the computed QM/MM interaction energies for the 1/63 system against an MB-Pol reference in supplementary Figure **??**.

Convergence with respect to the QM size

An important test for QM/MM simulations of solvated species is the convergence with respect to the number of water molecules included in the QM subsystem. As mentioned in the introduction, having an accurate model for the QM-MM interface and employing an MM force field that is consistent with the QM method (we use the MB-PBE force field) should allow us to showcase strong QM/MM convergence. The typical target is a ± 2 kcal/mol from the reference QM calculation.[40] In Figure 6, we present the interaction energies of glucose (left panel) and the $[Pd(H_2O)_4]^{2+}$ aqua ion, counterbalanced by 2 Cl$^-$ ions (right panel), with the water bulk environment as a function of the number of water molecules included in the QM subsystem. For glucose (a neutral molecule), the figure clearly shows that the ± 2 kcal/mol target is reached (in fact, a ± 1 kcal/mol is achieved) already when only 14 water molecules are included in the QM subsystem. This amounts to including less than the first solvation shell. Interestingly, the convergence for this system is much worse when a charge-only MM model is used (i.e., as done before, we simply neglect the polarization of the MM subsystem) for which 26 QM water molecules are needed to reach convergence. For PdCl$_2$, despite the presence of a doubly charged cation, we find an essentially identical behavior with convergence being reached already with 13 QM water molecules. Conversely, the charge only QM/MM method does not reach the ± 2 kcal/mol goal even when 46 QM water molecules are included. Therefore, we conclude that similar to the glucose system, including polarization in the MM subsystem dramatically improves the convergence of the

22

interaction energy with the size of the QM subsystem.

Our results on the need to include polarization in the MM subsystem find justification in the fact that, for a condensed phase system, including the inductive response of the environment, is crucial to obtain a physical picture, particularly for charged solutes.[41]
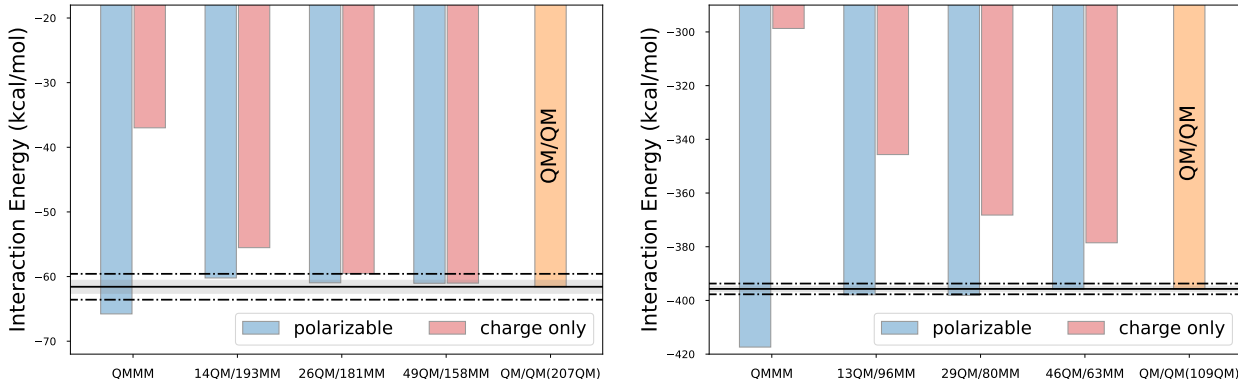

Figure 6: Convergence of the glucose (left panel) and the $[Pd(H_2O)_4]^{2+}$ aqua ion (counterbalanced by 2 Cl^- ions, right panel) solute-water interaction energy with respect to the number of water molecules included in the QM subsystem. An orange bar representing the sDFT reference is labeled as QM/QM in each plot. The dot-dashed line marks ± 2 kcal/mol, and the shade marks ± 1 kcal/mol window from the sDFT reference.

In supplementary Figure ??, we also present the polarization density of glucose and its water environment, showing once again that the polarization of the QM/MM model reproduces fairly accurately the polarization of the reference QM/QM simulation.

Wet Surfaces

In the supplementary information document,[62] we also present a QM/MM calculation of a monolayer MoS_2 solvated by water, see Figures ?? and ??. The conclusions of that analysis are consistent with what has been presented so far. That is, QM/MM simulations reproduce within a reasonable error the QM/QM result for the interaction energy. For this system, we record a QM/MM interaction energy of -0.11 kcal/mol $\cdot$ Å^{-2} at the equilibrium distance compared with a QM/QM reference of -0.23 kcal/mol $\cdot$ Å^{-2}, and QM/MM polarization densities very close to the QM/QM benchmark. The equilibrium water-MoS_2 distance of 2.1

Å is predicted by both QM/QM and QM/MM methods.

Conclusions

We introduced a novel QM/MM framework that incorporates density-functional theory (DFT) for the QM and the MM subsystems that leverages an orbital-free treatment for the MM region and the QM-MM interaction. By assigning an electron density to the MM subsystem and accurately capturing nonadditive interactions (such as exchange, correlation, Coulomb, and Pauli repulsion effects) our density-functionalized QM/MM approach achieves chemical accuracy in modeling complex solute-solvent systems. We validated the approach against a variety of water-based systems, including water clusters, bulk water, solvated ions, and a wet monolayer of MoS_2, demonstrating consistent accuracy and achieving unprecedented fast convergence to chemical accuracy with respect to increasing size of the QM subsystem. Reaching chemical accuracy with the proposed method requires only the inclusion of the first solvation shell in the QM region–a significant advancement over traditional QM/MM schemes.

Particularly striking is the performance of our density functionalized QM/MM method for H_2O in water. There we find that the excellent performance of the method for clusters (dimers, pentamers and hexamers) seamlessly translates to accurate models of bulk liquid water. A prime example is given by the dipole of the solute molecule which we predict an RMSE with QM/MM of 0.29 D, or 12%, for both bulk and pentamers compared to 0.624 D for pentamers from Ref. 41.

Our results highlight the critical role of (1) employing ab-initio density functionals for the QM-MM interactions instead of ad-hoc parametrizations; (2) properly accounting for mutual polarization at the QM-MM interface; and (3) employing MM force fields that are consistent with the QM method employed (here we use MB-PBE for MM and DFT with the PBE exchange-correlation functional for QM). We found that following these three principles

significantly improves convergence with respect to the size of the QM region compared to standard QM/MM methods. Our pilot simulations have showed that for both neutral and charged solutes, the interaction energies reached the target accuracy of within $\pm$ 2 kcal/mol using a minimal QM region, with further refinement yielding sub-1 kcal/mol errors when merely the first shell of solvent molecules is included in the QM subsystem. This level of accuracy, achieved even in bulk water systems, underscores the robustness of our method for simulating condensed-phase environments.

Overall, this work presents a significant step forward in extending the applicability of QM/MM methods to treat larger, more complex systems than typically approached by standard QM methods while maintaining chemical accuracy. Future work will explore the impact of including beyond-Coulomb, ab-initio terms in the MM embedding potential, so that the MM dipole response can more closely resemble the true electronic response of the electrons in the MM subsystem. We plan to apply the density-functionalized QM/MM framework to chemical environments other than water, for example those provided by biomolecules and materials interfaces as well as non-aqueous solvents.

Acknowledgements

Authors contributions

Pavanello, Andreussi and Paesani conceived the work and co-wrote the manuscript. Pavanello managed the students and the project. Shao and Riera developed the software. Martinez B. and Chen carried out the simulations and co-wrote the manuscript.

References

(1) Karplus, M. Development of Multiscale Models for Complex Chemical Systems: From H+H$_2$ to Biomolecules. Nobel Lecture, 2013; `https://www.nobelprize.org/uploads/2018/06/karplus-lecture.pdf`.

(2) Honig, B.; Karplus, M. Implications of torsional potential of retinal isomers for visual excitation. *Nature* **1971**, *229*, 558–560.

(3) Senn, H. M.; Thiel, W. QM/MM studies of enzymes. *Current Opinion in Chemical Biology* **2007**, *11*, 182–187.

(4) Kulik, H. J. Large-scale QM/MM free energy simulations of enzyme catalysis reveal the influence of charge transfer. *Phys. Chem. Chem. Phys.* **2018**, *20*, 20650–20660.

(5) Raghavan, B.; Paulikat, M.; Ahmad, K.; Callea, L.; Rizzi, A.; Ippoliti, E.; Mandelli, D.; Bonati, L.; De Vivo, M.; Carloni, P. Drug Design in the Exascale Era: A Perspective from Massively Parallel QM/MM Simulations. *Journal of Chemical Information and Modeling* **2023**, *63*, 3647–3658, PMID: 37319347.

(6) Lin, H.; Zhang, Y.; Pezeshki, S.; Duster, A. W.; Wang, B.; Wu, X.-P.; Zheng, S.-W.; Gagliardi, L.; Truhlar, D. G. QMMM 2023: A program for combined quantum mechanical and molecular mechanical modeling and simulations. *Computer Physics Communications* **2024**, *295*, 108987.

(7) Ma, C.; Martin-Samos, L.; Fabris, S.; Laio, A.; Piccinin, S. QMMMW: A wrapper for QM/MM simulations with Quantum ESPRESSO and LAMMPS. *Computer Physics Communications* **2015**, *195*, 191–198.

(8) Lambros, E.; Lipparini, F.; Cisneros, G. A.; Paesani, F. A Many-Body, Fully Polarizable Approach to QM/MM Simulations. *Journal of Chemical Theory and Computation* **2020**, *16*, 7462–7472.

(9) Gresh, N.; Cisneros, G. A.; Darden, T. A.; Piquemal, J.-P. Anisotropic, Polarizable Molecular Mechanics Studies of Inter- and Intramolecular Interactions and Ligand-Macromolecule Complexes. A Bottom-Up Strategy. *Journal of Chemical Theory and Computation* **2007**, *3*, 1960–1986.

(10) Lu, Z.; Zhang, Y. Interfacing ab Initio Quantum Mechanical Method with Classical Drude Osillator Polarizable Model for Molecular Dynamics Simulation of Chemical Reactions. *Journal of Chemical Theory and Computation* **2008**, *4*, 1237–1248.

(11) Ganguly, A.; Boulanger, E.; Thiel, W. Importance of MM Polarization in QM/MM Studies of Enzymatic Reactions: Assessment of the QM/MM Drude Oscillator Model. *Journal of Chemical Theory and Computation* **2017**, *13*, 2954–2961.

(12) Loco, D.; Lagardère, L.; Caprasecca, S.; Lipparini, F.; Mennucci, B.; Piquemal, J.-P. Hybrid QM/MM Molecular Dynamics with AMOEBA Polarizable Embedding. *Journal of Chemical Theory and Computation* **2017**, *13*, 4025–4033.

(13) Bondanza, M.; Nottoli, M.; Cupellini, L.; Lipparini, F.; Mennucci, B. Polarizable embedding QM/MM: the future gold standard for complex (bio)systems? *Phys. Chem. Chem. Phys.* **2020**, *22*, 14433–14448.

(14) Reinholdt, P.; Kongsted, J.; Lipparini, F. Fast approximate but accurate QM/MM interactions for polarizable embedding. *J. Chem. Theory Comput.* **2021**, *18*, 344–356.

(15) Giovannini, T.; Puglisi, A.; Ambrosetti, M.; Cappelli, C. Polarizable QM/MM Approach with Fluctuating Charges and Fluctuating Dipoles: The QM/FQFμ Model. *Journal of Chemical Theory and Computation* **2019**, *15*, 2233–2245, PMID: 30875213.

(16) Zinovjev, K. Electrostatic embedding of machine learning potentials. *J. Chem. Theory Comput.* **2023**, *19*, 1888–1897.

(17) Giese, T. J.; Zeng, J.; Lerew, L.; McCarthy, E.; Tao, Y.; Ekesan, S.; York, D. M. Software Infrastructure for Next-Generation QM/MM- ΔMLP Force Fields. *J. Phys. Chem. B* **2024**,

(18) Ratcliff, L. E.; Mohr, S.; Huhs, G.; Deutsch, T.; Masella, M.; Genovese, L. Challenges in large scale quantum mechanical calculations. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2017**, *7*, e1290.

(19) Groenhof, G. Introduction to QM/MM simulations. *Biomolecular simulations: methods and protocols* **2013**, 43–66.

(20) Warshel, A.; Levitt, M. Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *Journal of molecular biology* **1976**, *103*, 227–249.

(21) Lipparini, F. General linear scaling implementation of polarizable embedding schemes. *Journal of Chemical Theory and Computation* **2019**, *15*, 4312–4317.

(22) Bondanza, M.; Nottoli, T.; Nottoli, M.; Cupellini, L.; Lipparini, F.; Mennucci, B. The OpenMMPol library for polarizable QM/MM calculations of properties and dynamics. *J. Chem. Phys.* **2024**, *160*.

(23) Olsen, J. M. H.; Bolnykh, V.; Meloni, S.; Ippoliti, E.; Bircher, M. P.; Carloni, P.; Rothlisberger, U. MiMiC: a novel framework for multiscale modeling in computational chemistry. *Journal of chemical theory and computation* **2019**, *15*, 3810–3823.

(24) Cruzeiro, V. W. D.; Manathunga, M.; Merz Jr, K. M.; Götz, A. W. Open-source multi-GPU-accelerated QM/MM simulations with AMBER and QUICK. *Journal of Chemical Information and Modeling* **2021**, *61*, 2109–2115.

(25) Pederson, J. P.; McDaniel, J. G. PyDFT-QMMM: A modular, extensible software framework for DFT-based QM/MM molecular dynamics. *J. Chem. Phys.* **2024**, *161*.

(26) Dziedzic, J.; Mao, Y.; Shao, Y.; Ponder, J.; Head-Gordon, T.; Head-Gordon, M.; Skylaris, C.-K. TINKTEP: A fully self-consistent, mutually polarizable QM/MM approach based on the AMOEBA force field. *J. Chem. Phys.* **2016**, *145*.

(27) Lu, Y.; Sen, K.; Yong, C.; Gunn, D. S.; Purton, J. A.; Guan, J.; Desmoutier, A.; Nasir, J. A.; Zhang, X.; Zhu, L., et al. Multiscale QM/MM modelling of catalytic systems with ChemShell. *Phys. Chem. Chem. Phys.* **2023**, *25*, 21816–21835.

(28) Dohn, A.; Jonsson, E. O.; Levi, G.; Mortensen, J. J.; Lopez-Acevedo, O.; Thygesen, K. S.; Jacobsen, K. W.; Ulstrup, J.; Henriksen, N. E.; Møller, K., et al. Grid-based projector augmented wave (GPAW) implementation of quantum mechanics/molecular mechanics (QM/MM) electrostatic embedding and application to a solvated diplatinum complex. *J. Chem. Theory Comput.* **2017**, *13*, 6010–6022.

(29) Nam, K.; Gao, J.; York, D. M. An Efficient Linear-Scaling Ewald Method for Long-Range Electrostatic Interactions in Combined QM/MM Calculations. *J. Chem. Theory Comput.* **2005**, *1*, 2–13, PMID: 26641110.

(30) Holden, Z. C.; Richard, R. M.; Herbert, J. M. Periodic boundary conditions for QM/MM calculations: Ewald summation for extended Gaussian basis sets. *J. Chem. Phys.* **2013**, *139*, 244108.

(31) Bonfrate, S.; Ferré, N.; Huix-Rotllant, M. Analytic Gradients for the Electrostatic Embedding QM/MM Model in Periodic Boundary Conditions Using Particle-Mesh

Ewald Sums and Electrostatic Potential Fitted Charge Operators. *J. Chem. Theory Comput.* **2024**, *20*, 4338–4349.

(32) Pederson, J. P.; McDaniel, J. G. DFT-based QM/MM with particle-mesh Ewald for direct, long-range electrostatic embedding. *J. Chem. Phys.* **2022**, *156*.

(33) Laino, T.; Mohamed, F.; Laio, A.; Parrinello, M. An efficient linear-scaling electrostatic coupling for treating periodic boundary conditions in QM/MM simulations. *J. Chem. Theory Comput.* **2006**, *2*, 1370–1378.

(34) Laino, T.; Mohamed, F.; Laio, A.; Parrinello, M. An efficient real space multigrid QM/MM electrostatic coupling. *J. Chem. Theory Comput.* **2005**, *1*, 1176–1184.

(35) Brandt, F.; Jacob, C. R. Systematic QM region construction in QM/MM calculations based on uncertainty quantification. *J. Chem. Theory Comput.* **2022**, *18*, 2584–2596.

(36) Pedraza-González, L.; Marín, M. D. C.; Jorge, A. N.; Ruck, T. D.; Yang, X.; Valentini, A.; Olivucci, M.; De Vico, L. Web-ARM: a web-based interface for the automatic construction of QM/MM models of rhodopsins. *J. Chem. Inf. Model.* **2020**, *60*, 1481–1493.

(37) Kulik, H. J.; Zhang, J.; Klinman, J. P.; Martínez, T. J. How large should the QM region be in QM/MM calculations? The case of catechol O-methyltransferase. *J. Phys. Chem. B* **2016**, *120*, 11381–11394.

(38) Karelina, M.; Kulik, H. J. Systematic quantum mechanical region determination in QM/MM simulation. *J. Chem. Theory Comput.* **2017**, *13*, 563–576.

(39) Kulik, H. J. Large-scale QM/MM free energy simulations of enzyme catalysis reveal the influence of charge transfer. *Phys. Chem. Chem. Phys.* **2018**, *20*, 20650–20660.

(40) Pérez-Barcia, Á.; Cárdenas, G.; Nogueira, J. J.; Mandado, M. Effect of the QM size,

basis set, and polarization on QM/MM interaction energy decomposition analysis. *J. Chem. Inf. Model.* **2023**, *63*, 882–897.

(41) Dziedzic, J.; Head-Gordon, T.; Head-Gordon, M.; Skylaris, C.-K. Mutually polariz-able QM/MM model with in situ optimized localized basis functions. *The Journal of Chemical Physics* **2019**, *150*, 074103.

(42) Moscato, D.; Mandelli, G.; Bondanza, M.; Lipparini, F.; Conte, R.; Mennucci, B.; Ceotto, M. Unraveling Water Solvation Effects with Quantum Mechanics/Molecular Mechanics Semiclassical Vibrational Spectroscopy: The Case of Thymidine. *J. Am. Chem. Soc.* **2024**, *146*, 8179–8188.

(43) Gómez, S.; Giovannini, T.; Cappelli, C. Multiple facets of modeling electronic absorp-tion spectra of systems in solution. *ACS Physical Chemistry Au* **2022**, *3*, 1–16.

(44) Coons, M. P.; Herbert, J. M. Quantum chemistry in arbitrary dielectric environments: Theory and implementation of nonequilibrium Poisson boundary conditions and ap-plication to compute vertical ionization energies at the air/water interface. *Journal Chem. Phys.* **2018**, *148*.

(45) Martinez B, J. A.; Paetow, L.; Tolle, J.; Shao, X.; Ramos, P.; Neugebauer, J.; Pa-vanello, M. Which Physical Phenomena Determine the Ionization Potential of Liquid Water? *J. Phys. Chem. B* **2023**, *127*, 5470–5480.

(46) Olsen, J. M. H.; Steinmann, C.; Ruud, K.; Kongsted, J. Polarizable density embedding: A new QM/QM/MM-based computational strategy. *J. Phys. Chem. A* **2015**, *119*, 5344–5355.

(47) Kvedaraviciute, S.; Carrasco-Busturia, D.; Møller, K. B.; Olsen, J. M. H. Polarizable Embedding without Artificial Boundary Polarization. *J. Chem. Theory Comput.* **2023**, *19*, 5122–5141.

(48) Viquez Rojas, C. I.; Slipchenko, L. V. Exchange repulsion in quantum mechanical/effective fragment potential excitation energies: Beyond polarizable embedding. *J. Chem. Theory Comput.* **2020**, *16*, 6408–6417.

(49) Gokcan, H.; Kratz, E.; Darden, T. A.; Piquemal, J.-P.; Cisneros, G. A. QM/MM simulations with the Gaussian electrostatic model: A density-based polarizable potential. *J. Phys. Chem. Lett.* **2018**, *9*, 3062–3067.

(50) Kuechler, E. R.; Giese, T. J.; York, D. M. Charge-dependent many-body exchange and dispersion interactions in combined QM/MM simulations. *J. Chem. Phys.* **2015**, *143*.

(51) Treß, R. S.; Hattig, C.; Hofener, S. Employing pseudopotentials to tackle excited-state electron spill-out in frozen density embedding calculations. *J. Chem. Theory Comput.* **2022**, *18*, 1737–1747.

(52) Gillan, M.; Alfè, D.; Bygrave, P.; Taylor, C.; Manby, F. Energy benchmarks for water clusters and ice structures from an embedded many-body expansion. *J. Chem. Phys.* **2013**, *139*.

(53) Ben-Nun, M.; Martínez, T. J. Direct evaluation of the Pauli repulsion energy using classical'wavefunctions in hybrid quantum/classical potential energy surfaces. *Chem. Phys. Lett.* **1998**, *290*, 289–295.

(54) Giovannini, T.; Lafiosca, P.; Cappelli, C. A General Route to Include Pauli Repulsion and Quantum Dispersion Effects in QM/MM Approaches. *Journal of Chemical Theory and Computation* **2017**, *13*, 4854–4870, PMID: 28898079.

(55) Bedrov, D.; Piquemal, J.-P.; Borodin, O.; MacKerell, A. D. J.; Roux, B.; Schröder, C. Molecular Dynamics Simulations of Ionic Liquids and Electrolytes Using Polarizable Force Fields. *Chemical Reviews* **2019**, *119*, 7940–7995, PMID: 31141351.

(56) Hršak, D.; Olsen, J. M. H.; Kongsted, J. Optimization and transferability of non-electrostatic repulsion in the polarizable density embedding model. *Journal of computational chemistry* **2017**, *38*, 2108–2117.

(57) Krishtal, A.; Ceresoli, D.; Pavanello, M. Subsystem real-time time dependent density functional theory. *J. Chem. Phys.* **2015**, *142*.

(58) Jacob, C. R.; Neugebauer, J. Subsystem density-functional theory. *WIREs: Comp. Mol. Sci.* **2014**, *4*, 325–362.

(59) Wesolowski, T. A.; Shedge, S.; Zhou, X. Frozen-density embedding strategy for multilevel simulations of electronic structure. *Chem. Rev.* **2015**, *115*, 5891–5928.

(60) Mi, W.; Luo, K.; Trickey, S.; Pavanello, M. Orbital-free density functional theory: An attractive electronic structure method for large-scale first-principles simulations. *Chem. Rev.* **2023**, *123*, 12039–12104.

(61) Jacob, C. R.; Neugebauer, J. Subsystem density-functional theory (update). *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2024**, *14*, e1700.

(62) Supplementary Materials. See supplementary materials document at [url-to-be-inserted].

(63) Genova, A.; Ceresoli, D.; Pavanello, M. Avoiding fractional electrons in subsystem DFT based ab-initio molecular dynamics yields accurate models for liquid water and solvated OH radical. *J. Chem. Phys.* **2016**, *144*.

(64) Mi, W.; Ramos, P.; Maranhao, J.; Pavanello, M. Ab initio structure and dynamics of CO2 at supercritical conditions. *J. Phys. Chem. Lett.* **2019**, *10*, 7554–7559.

(65) Shao, X.; Mi, W.; Pavanello, M. GGA-level subsystem DFT achieves Sub-kcal/mol accuracy intermolecular interactions by mimicking nonlocal functionals. *J. Chem. Theory Comput.* **2021**, *17*, 3455–3461.

(66) Mi, W.; Pavanello, M. Nonlocal subsystem density functional theory. *J. Phys. Chem. Lett.* **2019**, *11*, 272–279.

(67) Schlüns, D.; Klahr, K.; Mück-Lichtenfeld, C.; Visscher, L.; Neugebauer, J. Subsystem-DFT potential-energy curves for weakly interacting systems. *Phys. Chem. Chem. Phys.* **2015**, *17*, 14323–14341.

(68) Shao, X.; Mi, W.; Pavanello, M. Density embedding method for nanoscale molecule–metal interfaces. *J. Phys. Chem. Lett.* **2022**, *13*, 7147–7154.

(69) Shao, X.; Lopez, A. C.; Khan Musa, M. R.; Nouri, M. R.; Pavanello, M. Adaptive subsystem density functional theory. *J. Chem. Theory Comput.* **2022**, *18*, 6646–6655.

(70) Chen, X.; Cifuentes-Lopez, A.; Shao, X.; Lin, L.; Prokopchuk, D.; Pavanello, M. Unraveling the Hydration Shell Structure and Dynamics of Group 10 Aqua Ions. *J. Phys. Chem. Lett.* **2024**, *15*, 5517–5528.

(71) Schmitt-Monreal, D.; Jacob, C. R. Density-based many-body expansion as an efficient and accurate quantum-chemical fragmentation method: Application to water clusters. *Journal of Chemical Theory and Computation* **2021**, *17*, 4144–4156.

(72) Focke, K.; Jacob, C. R. Coupled-cluster density-based many-body expansion. *The Journal of Physical Chemistry A* **2023**, *127*, 9139–9148.

(73) Kevorkyants, R.; Dulak, M.; Wesolowski, T. A. Interaction energies in hydrogen-bonded systems: A testing ground for subsystem formulation of density-functional theory. *J. Chem. Phys.* **2006**, *124*.

(74) Cisneros, G. A. Application of Gaussian Electrostatic Model (GEM) Distributed Multipoles in the AMOEBA Force Field. *J. Chem. Theory Comput.* **2012**, *8*, 5072–5080.

(75) Babin, V.; Leforestier, C.; Paesani, F. Development of a "First Principles" Water Potential with Flexible Monomers: Dimer Potential Energy Surface, VRT Spectrum,

and Second Virial Coefficient. *Journal of Chemical Theory and Computation* **2013**, *9*, 5395–5403, PMID: 26592277.

(76) Babin, V.; Medders, G. R.; Paesani, F. Development of a "First Principles" Water Potential with Flexible Monomers. II: Trimer Potential Energy Surface, Third Virial Coefficient, and Small Clusters. *Journal of Chemical Theory and Computation* **2014**, *10*, 1599–1607, PMID: 26580372.

(77) Riera, M.; Lambros, E.; Nguyen, T. T.; Götz, A. W.; Paesani, F. Low-order many-body interactions determine the local structure of liquid water. *Chem. Sci.* **2019**, *10*, 8211–8218.

(78) Garrity, K. F.; Bennett, J. W.; Rabe, K. M.; Vanderbilt, D. Pseudopotentials for high-throughput DFT calculations. *Computational Materials Science* **2014**, *81*, 446–452.

(79) Shao, X.; Mi, W.; Xu, Q.; Wang, Y.; Ma, Y. $\mathcal{O}(N \log N)$ scaling method to evaluate the ion–electron potential of crystalline solids. *J. Chem. Phys.* **2016**, *145*.

(80) Shao, X.; Mi, W.; Pavanello, M. eDFTpy: An Object-Oriented Platform for Density Embedding Simulations. available at http://edftpy.rutgers.edu, 2024; `http://edftpy.rutgers.edu`, (Last accessed on: 04/25/2024).

(81) Shao, X.; Andreussi, O.; Ceresoli, D.; Truscott, M.; Baczewski, A.; Campbell, Q.; Pavanello, M. QEpy: Quantum ESPRESSO in Python. Available at https://gitlab.com/shaoxc/qepy, 2024; `https://gitlab.com/shaoxc/qepy`, (Last accessed on: 04/25/2024).

(82) Giannozzi, P. et al. Advanced capabilities for materials modelling with QUANTUM ESPRESSO. *Journal of Physics: Condensed Matter* **2017**, *29*, 465901.

(83) Riera, M.; Knight, C.; Bull-Vulpe, E. F.; Zhu, X.; Agnew, H.; Smith, D. G. A.; Simmonett, A. C.; Paesani, F. MBX: A many-body energy and force calculator for

data-driven many-body simulations. *The Journal of Chemical Physics* **2023**, *159*, 054802.

(84) S. Nørby, M.; Magnus Haugaard Olsen, J.; Kongsted, J.; Aagard Jensen, H. J. Multipole moments for embedding potentials: Exploring different atomic allocation algorithms. *J. Comp. Chem.* **2016**, *37*, 1887–1896.

(85) Ferré, N.; Ángyán, J. G. Approximate electrostatic interaction operator for QM/MM calculations. *Chem. Phys. Lett.* **2002**, *356*, 331–339.

(86) Genova, A.; Ceresoli, D.; Krishtal, A.; Andreussi, O.; DiStasio Jr, R. A.; Pavanello, M. eQE: An open-source density functional embedding theory code for the condensed phase. *Int. J. Quantum Chem.* **2017**, *117*, e25401.

(87) Mi, W.; Shao, X.; Genova, A.; Ceresoli, D.; Pavanello, M. eQE 2.0: Subsystem DFT Beyond GGA Functionals. *Comp. Phys. Comm.* **2021**, *269*, 108122.

(88) Genova, A.; Pavanello, M. Exploiting the locality of periodic subsystem density-functional theory: efficient sampling of the Brillouin zone. *J. Phys.: Condens. Matter* **2015**, *27*, 495501.

(89) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple [Phys. Rev. Lett. 77, 3865 (1996)]. *Phys. Rev. Lett.* **1997**, *78*, 1396–1396.

(90) Laricchia, S.; Fabiano, E.; Constantin, L.; Della Sala, F. Generalized gradient approximations of the noninteracting kinetic energy from the semiclassical atom theory: Rationalization of the accuracy of the frozen density embedding theory for nonbonded interactions. *J. Chem. Theory Comput.* **2011**, *7*, 2439–2451.

(91) Grimme, S. Supramolecular binding thermodynamics by dispersion-corrected density functional theory. *Chemistry–A European Journal* **2012**, *18*, 9955–9964.

(92) Babin, V.; Leforestier, C.; Paesani, F. Development of a "first principles" water potential with flexible monomers: Dimer potential energy surface, VRT spectrum, and second virial coefficient. *J. Chem. Theory Comput.* **2013**, *9*, 5395–5403.

(93) BÖTTCHER, C., Ed. *Theory of Electric Polarization (Second Edition)*, second edition ed.; Theory of Electric Polarization; Elsevier: Amsterdam, 1973; p iii.

(94) Kumar P., S.; Genova, A.; Pavanello, M. Cooperation and Environment Characterize the Low-Lying Optical Spectrum of Liquid Water. *J. Phys. Chem. Lett.* **2017**, *8*, 5077–5083.

(95) Tschumper, G. S.; Leininger, M. L.; Hoffman, B. C.; Valeev, E. F.; Schaefer, I., Henry F.; Quack, M. Anchoring the water dimer potential energy surface with explicitly correlated computations and focal point analyses. *The Journal of Chemical Physics* **2002**, *116*, 690–701.

(96) Holt, A.; Boström, J.; Karlström, G.; Lindh, R. A NEMO potential that includes the dipole–quadrupole and quadrupole–quadrupole polarizability. *J. Comp. Chem.* **2010**, *31*, 1583–1591.

(97) Holt, A.; Karlström, G. Inclusion of the quadrupole moment when describing polarization. The effect of the dipole-quadrupole polarizability. *J. Comp. Chem.* **2008**, *29*, 2033–2038.

(98) Hincapié, G.; Acelas, N.; Castaño, M.; David, J.; Restrepo, A. Structural Studies of the Water Hexamer. *J. Phys. Chem. A* **2010**, *114*, 7809–7814, PMID: 20590081.

(99) Lord Rayleigh, The problem of the random walk. *Nature* **1905**, *72*, 318–319.

(100) Krishtal, A.; Pavanello, M. Revealing electronic open quantum systems with subsystem TDDFT. *J. Chem. Phys.* **2016**, *144*.

(101) Luber, S. Local electric dipole moments for periodic systems via density functional theory embedding. *J. Chem. Phys.* **2014**, *141*.

(102) Tang, K.; Toennies, J. P. An improved simple model for the van der Waals potential based on universal damping functions for the dispersion coefficients. *J. Chem. Phys.* **1984**, *80*, 3726–3741.

(103) Tkatchenko, A.; Ambrosetti, A.; DiStasio, R. A. Interatomic methods for the dispersion energy derived from the adiabatic connection fluctuation-dissipation theorem. *J. Chem. Phys.* **2013**, *138*.

(104) Boereboom, J. M.; Potestio, R.; Donadio, D.; Bulo, R. E. Toward Hamiltonian adaptive qm/mm: accurate solvent structures using many-body potentials. *J. Chem. Theory Comput.* **2016**, *12*, 3441–3448.

(105) Chen, J.; Harper, J. B.; Ho, J. Improving the accuracy of quantum mechanics/molecular mechanics (QM/MM) models with polarized fragment charges. *J. Chem. Theory Comput.* **2022**, *18*, 5607–5617.