Grid-Based Projector Augmented Wave (GPAW) Implementation of Quantum Mechanics/Molecular Mechanics (QM/MM) Electrostatic Embedding and Application to a Solvated Diplatinum Complex


Published in:
Journal of Chemical Theory and Computation

Link to article, DOI:
10.1021/acs.jctc.7b00621

Publication date:
2017

Document Version
Peer reviewed version

Link back to DTU Orbit

Citation (APA):
A GPAW Implementation of QM/MM Electrostatic Embedding and Application to a Solvated Diplatinum Complex
A GPAW Implementation of QM/MM

Electrostatic Embedding and Application to a Solvated Diplatinum Complex


†Contributed equally to this work
‡Faculty of Physical Sciences and Science Institute, University of Iceland
¶Department of Chemistry, Technical University of Denmark
§CAMD, Department of Physics, Technical University of Denmark
‖Department of Applied Physics, Aalto University

E-mail: asod@hi.is

Abstract

A multiscale DFT-QM/MM scheme is presented, based on an efficient electrostatic coupling between the electronic density obtained from a Grid-based Projector Augmented Wave (GPAW) implementation of density functional theory and a classical potential energy function. The scheme is implemented in a general fashion and can be used with various choices for the descriptions of the QM or MM regions. Tests on H₂O clusters, ranging from dimer to decamer show that no systematic energy errors are introduced by the coupling that exceeds the differences in the QM and MM descriptions. Over 1 nanosecond of liquid water Born-Oppenheimer QM/MM Molecular Dynamics
(MD) are sampled combining 10 parallel simulations, showing consistent liquid water
structure over the QM/MM border. The method is applied in extensive parallel MD
simulations of an aqueous solution of the diplatinum \([\text{Pt}_2(\text{P}_2\text{O}_5\text{H}_2)_4]^{4-}\) complex (Pt-
POP), spanning a total time period of roughly half a nanosecond. An average Pt-Pt
distance deviating only 0.01 Å from experimental results, and a ground state Pt-Pt
oscillation frequency deviating less than 2% from experiment were obtained. The sim-
ulations highlight a remarkable harmonicity of the Pt-Pt oscillation, while also showing
clear signs of Pt-H hydrogen bonding and directional coordination of water molecules
along the Pt-Pt axis of the complex.

1 Introduction

The combination of models of different accuracy and complexity in a multiscale simula-
tion approach that strikes a balance between precision/generality and efficiency has been
highly successful in studies of physical, chemical, and biological processes where interaction
with the environment plays an important role. By subdividing the simulated system into
subdomains based on the required level of precision, larger and more complex systems can
be studied. The basic idea is that some part of the system requires a more accurate and
fundamental description than another, larger part of the system. The former can be de-
scribed with a quantum mechanical treatment of the electronic degrees of freedom, while
the latter is described with a classical potential function, so-called molecular mechanics.
After the pioneering work of M. Levitt and A. Warshel,\(^1\) the methodology for combining
electronic structure calculations with classical potential functions has seen many further
developments.\(^2\)\textsuperscript{-20} These types of quantum mechanical / classical mechanical (QM/MM) ap-
proaches have already been used extensively within e.g. materials research,\(^21\) biology,\(^22\)\textsuperscript{-24}
enzyme-based catalysis\(^25\) and medicine,\(^26\) and fundamental photochemistry.\(^27\)

However, for many applications, the inherent computational cost inhibits more thorough
thermal samplings, and the resulting lack of statistics can e.g. lead to unfair weights on...
particular solute/solvent configurations. Explicitly coupling the QM and MM subsystems require the evaluation of electrostatic interaction terms between the electronic density of the QM subsystem, often expressed on a real space grid, and the MM point charges. These evaluations can take up a substantial amount of the total computational cost. With this work, we have taken advantage of the computational expediency of the Grid-based Projector Augmented Wave (GPAW)\textsuperscript{28,29} code to develop a QM/MM interface that only adds small computational cost, without introducing further approximations into the interaction.

A preliminary version of the QM/MM framework presented here made it possible to study both internal- and solvation dynamics,\textsuperscript{30,31} as well as thermal averages\textsuperscript{32} of the active-sites in model catalysts, with statistical quantities that made it crucial in the interpretation of femtosecond time-resolved XFEL X-Ray data.\textsuperscript{33,34} We have opted for an approach that includes electrostatic embedding of the QM subsystem in the MM subsystem.

In GPAW, the QM subsystem is described using a real space grid representation of the electronic wave function. The Projector Augmented Wave (PAW) formalism is used\textsuperscript{35,36} and the MM subsystem is described with a potential energy function based on point charge interactions. The interface between the QM and MM subsystem is represented by adding the electrostatic potential from the MM region to the electronic structure calculation of the QM region. Furthermore, a short range, non-bonded interaction between molecules in the two subsystems is included, as well as a switching function to prevent electron spillover to the point charges of the MM region. The Atomic Simulation Environment (ASE)\textsuperscript{37,38} is used for the QM/MM implementation.

We note here that other grid based, electrostatic embedding QM/MM implementations exist,\textsuperscript{39–43} where the interfacing potential is also described on real-space grid – but they differ from ours in key aspects: A major difference is the coupling to a QM calculator, where the wave functions are described in whole,\textsuperscript{41,42} or in part,\textsuperscript{39,40} by plane-waves. In some cases,\textsuperscript{42} this means that the QM and MM cells must match, which can make the direct electrostatic coupling computationally demanding even for small QM systems. In other cases the cells do
not have to be equivalent, instead the electrostatics are corrected and scaled as if the QM cell would have the dimension of the MM cell, e.g. via multigrid approximation methods.\textsuperscript{39,40}

We have performed various tests on H\textsubscript{2}O clusters, ranging from dimer to decamer, as well as liquid water, and show that our scheme has accuracy similar to commonly used generalized gradient approximation density functionals, such as PBE, and point charge models of water, in particular TIP3P\textsuperscript{44} and TIP4P.\textsuperscript{45} This methodology is applied to an aqueous solution of the diplatinum [Pt\textsubscript{2}(P\textsubscript{2}O\textsubscript{5}H\textsubscript{2})\textsubscript{4}]\textsuperscript{4−} complex, abbreviated as PtPOP. Due to the involvement of Pt in many catalytic processes, much effort is being put into understanding hydrogen bonding between Pt and water.\textsuperscript{46,47} This complex has, in particular, been studied extensively,\textsuperscript{48} both experimentally\textsuperscript{49–52} and computationally, with electronic structure methods,\textsuperscript{53,54} in spite of the computational cost brought about by its 448 electrons. The nature of the interaction of water molecules with the Pt atoms in the complex has been debated in the literature: Some argue that since no solvatochromism is observed in the emission spectra of PtPOP, there should be no axial coordination of the solvent molecules, either in the ground- or in the first excited state of PtPOP,\textsuperscript{51} since this coordination would change for different solvents. Others have hypothesized Pt-H interactions along the Pt-Pt axis as a possible mechanism of vibrational energy dissipation in the excited state.\textsuperscript{52} A characterization of the hydration shell of PtPOP and the nature of the Pt-solvent interactions may be important for understanding its properties, such as photocatalytic activity.\textsuperscript{55} Previous QM/MM MD simulations were too limited in statistical sampling because of the large computational effort to provide an answer to this question.\textsuperscript{53} The present QM/MM implementation is, however, capable of performing Born-Oppenheimer Molecular Dynamics (MD) simulations, where the classical forces on the QM atomic centers are evaluated on-the-fly from the total QM/MM potential, sampling in parallel several hundreds of picoseconds of dynamics, and collect more converged statistics on various bond angle distributions, compared to previous studies. The results show clear indication of a specific interaction between the Pt atoms of the complex and hydrogen atoms of the solvent resulting in preferred axial coordination geometry. Such a –OH···Pt hydrogen
bond has recently been identified in a joint experimental and theoretical study of a H$_2$O dimer on a Pt(111) surface.$^{56}$ The method presented here is focused on sampling of (non-equilibrium) dynamics, solvent responses, and thermal averages of smaller molecules (of up to a few hundred atoms) in solution. The current implementation is still non-adaptive, i.e. the QM and MM regions remain the same during the simulations, and there can be no charge transfer between the two regions, and/or polarization of the MM region, as with other implementations.$^{57-62}$ Currently, no methods for the QM/MM border crossing bonds$^{63}$ have been implemented. Future work includes a polarizable embedding QM/MM scheme. All density functionals available to GPAW (including meta functionals and all other functionals in libxc$^{64}$) can be used with the current implementation, however, GPAW is currently not capable of providing forces from hybrid functionals. When the capability to calculate these forces are be added to GPAW, no changes will be needed in the QM/MM implementation.

## 2 Electrostatic Embedding within the PAW formalism

Electrostatic Embedding QM/MM, originally proposed by Karplus and coworkers,$^2$ is an additive coupling scheme, meaning that a term, $E_I$, is added to the QM and MM energy to represent the interaction between the two subsystems, giving the total energy expression:

$$E_{TOT} = E_QM + E_I + E_{MM}$$  \hspace{1cm} (1)

If the MM region is described using a point charge force field, with $q_i$ denoting the charges and $\tau_i$ their spatial coordinates, the electrostatic embedding energy $E_I$ becomes (in atomic units):

$$E_I = -\sum_{i=1}^{C} q_i \int \frac{n(r)}{|r - \tau_i|} dr + \sum_{i=1}^{C} \sum_{a=1}^{A} \frac{q_i Z_a}{|R_a - \tau_i|} + E_{NB}$$  \hspace{1cm} (2)

Here, $n(r)$ is the spatial (positive) electronic density, $Z_a$ the atomic number, and $R_a$ the coordinates of the nuclei in the QM region. The $E_{NB}$ term describes the remaining, non-
Coulomb interactions.

A thorough account of the implementation of the Projector Augmented Wave (PAW) method on a real-space grid in the GPAW software can be found elsewhere, but here we will briefly outline how the quantities used in the QM/MM implementation are obtained. More details can be found in the S.I.

Within the PAW formalism the electron density of the QM system is obtained in terms of a smooth \( \tilde{n}(\mathbf{r}) \), referred to as the pseudo electron density, and one-center atomic corrections:

\[
n(\mathbf{r}) = \tilde{n}(\mathbf{r}) + \sum_{\alpha} \left( n_{\alpha}(\mathbf{r}) - \tilde{n}_{\alpha}(\mathbf{r}) \right) \tag{3}
\]

Each term \( n_{\alpha}(\mathbf{r}) - \tilde{n}_{\alpha}(\mathbf{r}) \) in the above expression is non-vanishing inside predefined augmentation spheres surrounding each nucleus \( \alpha \), while outside them, the PAW transformation ensures that \( n(\mathbf{r}) = \tilde{n}(\mathbf{r}) \). In GPAW the property of \( \tilde{n}(\mathbf{r}) \) being smooth everywhere in space is exploited by conveniently representing it on a coarse, real space grid, while all one-center quantities, which allows recovery of the true oscillatory behaviour in close proximity to the nuclei, are efficiently represented on atom-centered radial grids. In order to derive the expression used to compute the QM/MM electrostatic interaction energy within the present PAW multiscale embedding scheme, it will prove convenient to rewrite the general expression in equation 2 as:

\[
E_I = -\sum_{i=1}^{C} q_i \int \frac{n(\mathbf{r}) + \sum_{\alpha} Z_{\alpha}(\mathbf{r})}{|\mathbf{r} - \tau_i|} d\mathbf{r} + E_{NB} \tag{4}
\]

where the nuclear point charge density \( Z_{\alpha}(\mathbf{r}) \) is given by the atomic number \( Z_{\alpha} \) times a delta function operating at the nuclear site \( \mathbf{R}_{\alpha} \), \( Z_{\alpha}(\mathbf{r}) = -Z_{\alpha} \delta(\mathbf{r} - \mathbf{R}_{\alpha}) \). By inserting the PAW formulation of the electronic density (equation 3) in 4 and grouping terms that involve a summation over nuclei we obtain:

\[
E_I = -\sum_{i=1}^{C} q_i \int \frac{\tilde{n}(\mathbf{r}) + \sum_{\alpha} \left( n_{\alpha}(\mathbf{r}) + Z_{\alpha}(\mathbf{r}) - \tilde{n}_{\alpha}(\mathbf{r}) \right)}{|\mathbf{r} - \tau_i|} d\mathbf{r} + E_{NB} \tag{5}
\]
A simplification of equation 5 is achieved by, as in the original GPAW methodology,\textsuperscript{28,29} introducing a new set of smooth one-center functions $\tilde{Z}_\alpha(r)$ localized inside the augmentation spheres:

$$E_I = -\sum_{i=1}^{C} q_i \int \frac{\tilde{n}(r) + \sum_\alpha \tilde{Z}_\alpha(r) + \sum_\alpha \left( n_\alpha(r) + Z_\alpha(r) - \tilde{n}_\alpha(r) - \tilde{Z}_\alpha(r) \right)}{|r - \tau_i|} \, \text{d}r + E_{NB} \quad (6)$$

and requiring that by construction of $\tilde{Z}_\alpha(r)$ the densities $n_\alpha(r) + Z_\alpha(r)$ and $\tilde{n}_\alpha(r) + \tilde{Z}_\alpha(r)$, which vanish outside the augmentation regions, have the same electrostatic multipole moments. As a result all terms $n_\alpha(r) + Z_\alpha(r) - \tilde{n}_\alpha(r) - \tilde{Z}_\alpha(r)$ in the expansion over nuclei in 6 do not interact with the MM point charges $q_i$. Thus we obtain the simplified expression for $E_I$:

$$E_I = -\sum_{i=1}^{C} q_i \int \frac{\tilde{n}(r) + \tilde{Z}(r)}{|r - \tau_i|} \, \text{d}r + E_{NB} \quad (7)$$

where $\tilde{Z}(r) = \sum_\alpha \tilde{Z}_\alpha(r)$. This is an approximation since the multipole expansion of $\tilde{Z}_\alpha$ in GPAW\textsuperscript{28,29} is finite and carried out up to the quadrupole. Equation 7 is convenient for computations since both $\tilde{n}(r)$ and $\tilde{Z}(r)$ can be evaluated on a coarse grid in real space. By defining a pseudo total charge density $\tilde{\rho}(r)$ as $\tilde{\rho}(r) = \tilde{n}(r) + \tilde{Z}(r)$, we arrive at the central expression for the interaction energy in the GPAW QM/MM embedding scheme:

$$E_I = -\sum_{i=1}^{C} q_i \int \frac{\tilde{\rho}(r)}{|r - \tau_i|} \, \text{d}r + E_{NB} \quad (8)$$

The derivative of the right hand side of equation (8) with respect to the charge density is added to the Kohn-Sham DFT scheme as an additional external potential:

$$V_{ext}^{MM}(r) = -\sum_{i=1}^{C} \frac{q_i}{|r - \tau_i|} \quad (9)$$

so that we can redefine the total energy of the QM subsystem including the Coulomb inter-
action energy with the external potential as the following functional of the density:

\[ E'_{\text{QM}}[n(r)] = T_s[n(r)] + \int \tilde{\rho}(r)V_{\text{MM}}^{\text{ext}}(r)\,dr + E_C[\rho(r)] + E_{\text{xc}}[n(r)] \]  

(10)

The standard Kohn–Sham DFT energy functional, Hamiltonian and forces in the context of GPAW can be found elsewhere.\textsuperscript{28,29} A more detailed description of the external potential term and resulting forces in GPAW is, however, presented in the S.I.

The Coulomb interactions between grid points and point charges are the most computationally demanding part of the interfacing scheme, but utilize GPAW’s\textsuperscript{28,29} internal cost-minimizer algorithm which optimizes the domain decomposition of the real-space grid (on which the wavefunctions, density and potentials are described) among the allocated processors. The evaluation of the interface potential due to the presence of the point charges becomes a trivial task for any reasonable applications compared to the total QM SCF cycle, and hence are as efficiently scalable as the underlying QM code. We also note that the expansion \( \tilde{Z}(r) \), introduced here in the context of the GPAW QM/MM scheme, is already present in the original PAW formulation,\textsuperscript{35} where it is referred to as ‘compensation charges’. Such quantity is, in fact, added to the pseudo electron density \( \tilde{n}(r) \) and enters the definition of the PAW Hamiltonian to electrostatically decouple one-center contributions to the charge density from each other and achieve a simplification of the Coulomb terms even in the absence of an external potential. The explicit expressions that are used in GPAW to construct the compensation charge density and represent it on a real grid can be found elsewhere.\textsuperscript{28,29}

What is important here is that the present QM/MM interfacing strategy, as exemplified by equation 8, deals exclusively with the pseudo charge density \( \tilde{\rho}(r) \) allowing for a straightforward inclusion of the external MM point charge potential in the GPAW Kohn-Sham DFT scheme.

Returning to the \( E_{\text{NB}} \)-term from equation (8), we have chosen to describe the long range dispersion and short range exchange repulsion with a simple Lennard-Jones (LJ) potential
between the atoms of the QM and MM subsystems:

\[ E_{NB} = \sum_i^C \sum_{\alpha}^A 4\epsilon \left[ \left( \frac{\sigma}{|R_{\alpha} - \tau_i|} \right)^{12} - \left( \frac{\sigma}{|R_{\alpha} - \tau_i|} \right)^6 \right] \]  

(11)

where \( \epsilon \) and \( \sigma \) are the energy and size parameters, respectively, giving the depth of the energy minimum depth and the zero of the repulsive wall. This is a simple choice that does not take the electronic density distribution in the QM subsystem into account, and the modularity of the code allows for straightforward future additions of more advanced models. The LJ parameters can be chosen freely in the present implementation.

When the grid of the QM region overlaps with point charges in the MM region, special care must be taken so that electrons do not spill over to the MM point charges. Many methods have already been developed for screened and/or damped electrostatics in the relevant regions of space.\(^{63,65,66}\) Two approaches for doing this are explained in detail in the S.I., appendix A. The evaluation of the atomic forces on the Born-Oppenheimer energy surfaces are described in the S.I., appendix B, and code details on the implementation can be found in the S.I., appendix C. Appendix E of the S.I. contains an evaluation of the efficiency of the QM/MM interface.

## 3 Computational Details

The DFT calculations were carried out using functionals that are commonly used for relatively large systems, namely functionals of the GGA form, in particular PBE and BLYP. Currently, GPAW does not support the calculation of forces from hybrid functionals, but the presented model will readily accept them, when implemented. We note that for systems as simple as the water dimer and water clusters, other more advanced functionals (functionals with empirical,\(^{67}\) or from-first-principles\(^{68,69}\) van der Waals corrections) exist, that could provide more accurate descriptions of water,\(^{70,71}\) but the pure water benchmarks in this work are made to assess the coupling accuracy, not the accuracy of the water structure. In the
end, we are ultimately interested in simulating large systems where the simpler functionals are more manageable, and furthermore, one could speculate that at least in some cases, the errors introduced by forgoing an explicit solvation model would outweigh the inaccuracies brought about by simpler functionals, as has been the case for similar studies. 

For the benchmark calculations, a grid spacing of 0.15 Å was used, in conjunction with the LCAO mode of GPAW, unless specified otherwise. For the LCAO-mode calculations, a tzp size basis set was used for the QM and QM/MM calculations using the counterpoise correction to eliminate the basis set superposition error (BSSE), adding ghost atoms to the monomer- and QM/MM dimer calculations where appropriate, ensuring that the same degrees of freedom were available to the electronic wave functions in all calculations. Ghost atoms possess basis functions as their normal counterparts, but nothing else. All the QM/MM results shown have been produced with our smoothly switched (SS) short-range analytical potential, unless otherwise specified, see the S.I. appendix A for details.

Water molecules in the MM regions were represented by the TIP4P potential function in most cases, but some test calculations were also carried out using TIP3P for comparison. In all cases, water molecules were kept rigid by constraining all interatomic distances with the ASE implementation of RATTLE.

A simple but important example to test how well hydrogen bonds are represented across the QM/MM interface is the H₂O dimer. Since pure QM and pure MM descriptions of the dimer would provide different geometries, we chose to use the geometry from the s22 database throughout. The s22 geometry is optimized using numerical CCSD(T) gradients and an cc-pVQZ basis set. From this geometry, the single point water dimer binding energy was calculated, \( E_b = E_{\text{dimer}} - (E_{\text{mono}_1} + E_{\text{mono}_2}) \), changing only the distance between the two water molecules, but keeping all other structural features unchanged. The dimer binding energy curves were calculated for fully QM, fully MM, and the two possible choices for QM/MM systems.

For systems containing more than two water molecules, the interaction energy was eval-
\[ E_{\text{int}} = E_{\text{cluster}} - \sum_{m} E_{m} \]  

(12)

where \( E_m \) is the energy of monomer \( m \) in the cluster. The interaction energies are calculated for all possible \( \frac{n!}{r!(n-r)!} \) combinations of QM and MM descriptions, where \( n \) is the total number of molecules and \( r = \{1..n - 1\} \) is the number of QM molecules.

For the QM/MM MD simulations of a single QM water molecule in liquid MM water, cubic 20 Å boxes of water were prepared at the liquid density and pre-equilibrated purely with TIP3P. Then, the model was switched to QM/MM, with a single water molecule described with PBE, a 0.18 Å grid spacing and a tzp-size basis. A total of 1.2 ns of equilibrated dynamics were sampled for the QM/MM simulations, and the resulting RDF, \( g_{\text{OQM}\text{OMM}}(r) \) was sampled from the single QM Oxygen to the rest of the water oxygen atoms, modeled with TIP3P.

For the PtPOP simulations, the optimized geometry of PtPOP in its ground electronic state (GS) was placed in a cubic 35 Å box of pre-equilibrated TIP4P water molecules at 300 K. The classical subsystem included four K\(^+\) counter ions to neutralize the total charge. The size of the simulation box represents a much larger effective concentration than what is used in experiments. Therefore, during the dynamics, the position of each counter ion was restrained to regions of the simulation box outside a sphere centered at the center of the QM cell by applying the following potential (general form):

\[
V_{\text{PR}}(r_i) = \begin{cases} 
\frac{1}{2}k_{pr} (r_{i}^c - r_{i}^{pr})^2 & \text{if } r_{i}^c \leq r_{i}^{pr} \\
0 & \text{if } r_{i}^c > r_{i}^{pr} 
\end{cases}
\]

(13)

with corresponding forces:

\[
F_{i}^{\text{PR}} = \begin{cases} 
-k_{pr} \left(1 - \frac{r_{i}^{pr}}{r_{i}^c}\right) r_{i}^{pr} & \text{if } r_{i}^c \leq r_{i}^{pr} \\
0 & \text{if } r_{i}^c > r_{i}^{pr} 
\end{cases}
\]

(14)
where $k_{pr}$ is a harmonic force constant, $r_i$ is the position of counter ion $i$ and $r'_i = |r_i - r^{CQM}|$. The cutoff $r^{pr}_c$ for the harmonic restraint potential was chosen equal to 16 Å. The total number of water molecules after removing those overlapping with the solute was 1383. The ground state of PtPOP was described using the BLYP functional, a grid spacing of 0.18 Å, and with tzp$^{73}$ basis set for Pt and dzp$^{73}$ for the rest of the atoms, while non-bonded parameters were chosen from the universal force field (UFF). After solvating the complex, the entire box was further equilibrated to 300 K employing a 1 fs timestep until temperature stabilized. Thermalization was realized using the ASE Langevin thermostat applied to the solvent, while periodic boundary conditions were applied according to the minimum image convention. Stability of the simulations was ensured by constraining all OH bonds and hydrogen bonds present in PtPOP with the RATTLE algorithm. Thereafter, MD data was collected with 2 fs timestep, for at least 25 ps. From the equilibrated part of the trajectory, 48 more parallel MD production runs were started at 0.5 ps intervals, to further speed up the data collection process. When starting each trajectory, the velocities of the atoms in the solvent were randomized by imposing a Maxwell-Boltzmann distribution at 300 K, to minimize the correlation between them. To assess the impact of constraining all OH bonds in the complex on its equilibrium properties and dynamics, a single trajectory with increased mass for all hydrogen atoms but no constraints on the degrees of freedom of the solute was produced. The average of the main structural parameters of the complex and the Pt-Pt oscillating frequency obtained from this trajectory were found to be negligibly different from those obtained when employing RATTLE constraints for the OH bonds. Overall, the equilibrated trajectories amounted to 463 ps and were obtained over approximately 9750 hours of CPU time, corresponding to around 21 hours per picosecond.
4 Benchmarks

In this section, we test the accuracy of the embedding scheme in relation to the distinct models for two regions. By comparing the QM/MM interaction energies with full QM and full MM results on a system that only contains one molecular species, we can assess the accuracy of our model. Since the method has been developed to describe molecules in solution, and that one of the most ubiquitous solvents is water, we focus our tests on pure water. The criteria of success are that (1) the coupled binding energies do not over- or underbind, when compared to neither the chosen QM nor the MM models, and (2) that the various possible combinations of QM/MM geometries and regions are as similar as possible, so that there will be no orientation-induced differences in energies and forces. We note that neither the PBE functional, nor the TIP3P classical potential are perfect water models, and constructing such a model is outside the scope of this work.

4.1 Benchmark: H₂O Dimer

The calculated dimer binding energy curves for fully QM, fully MM, and the two QM/MM cases (where $E_{\text{mono}_1} \neq E_{\text{mono}_2}$), are shown in figure 1. We observe that the QM/MM and MM/QM binding curves are essentially identical.

The pure PBE energy minimum of the s22 geometry is 0.014 eV higher than the CCSD(T)/cc-pVQZ value, and the minimum energy is at a slightly larger separation, 2.99 Å. The MM force field, TIP4P, however, gives a shorter dimer bond distance and a deeper well. The two QM/MM calculations, corresponding to the different choices for the QM and MM molecules, are very similar, and are close to the pure MM binding curve. Since the implemented electrostatic embedding scheme couples the two subsystems in a "classical" fashion, i.e. via classical Coulomb interactions, this is to be expected. As seen in the figure, neither of the two QM/MM and MM/QM curves significantly over- or underbind compared to the difference between the pure QM and MM descriptions.
Figure 1: Binding energy for the H$_2$O dimer. The two dashed lines represent the QM/MM binding energy of each of the QM and MM monomer description. The PBE energy functional is used for the QM and TIP4P potential function for MM. The two QM/MM results are practically the same and close to the results where both molecules are described with MM. The vertical lines represent the optimal oxygen-oxygen distance for the s22 geometry (black) and TIP4P potential (blue).

4.2 Benchmark: Larger (H$_2$O)$_n$ Clusters

Figure 2 shows the statistical distributions of differences in interaction energies $\Delta E = E_{\text{int}}^{\text{QMMM}} - E_{\text{int}}^{\text{MM}}$, between QM/MM results and full MM results for all possible QM/MM combinations. The dataset is produced from the water cluster geometries of the Bates & Tschumper hexamers and the set of (H$_2$O)$_n$, $n = 3 - 10$, clusters provided by B. Temelso and coworkers, to achieve better statistics for the interaction energy benchmarks. Overall, for the 3-9 molecule clusters, the difference in interaction energy middle 50% of the distribution (denoted by the white boxes) is mostly within the maximum difference between a full QM and a full MM description. For the 8-10 molecule clusters, the QM/MM spread (total, as well as middle 50%) is actually smaller than for the smaller clusters, which could indicate that the difference between the QM and MM descriptions decrease as the system size increases. The water database only contains 2, 3, and 2 clusters of 8, 9, and 10 molecules, respectively. Therefore, the decreasing difference in TIP3P and PBE interaction energy with increased system size should not be considered a general trend. Indeed, if the benchmark is repeated using the BLYP functional, as shown in figure 3, the picture is different. This
Figure 2: Box plot of $\Delta E = E_{\text{int}}^{\text{QM/MM}} - E_{\text{int}}^{\text{MM}}$, the QM/MM (PBE/TIP3P) vs full MM interaction energy difference of every possible combinations of QM and MM descriptions of both the Bates-Tschumper hexamers,$^{50}$ and the (H$_2$O)$_n$, $n = 3-10$ clusters from the database by B. Temelso$^{81}$ and coworkers, for a total of 6331 single point QM/MM calculations. The boxes represent the lower and upper quartile, split by the red median line, and the whiskers the standard 1.5 times the interquartile range (IQR). The red patch is the signed maximum interaction energy difference between the full QM and full MM description. The blue patch is the maximum difference of opposite sign, so e.g. if the maximum difference is negative i.e. overbinding, the maximum difference of the opposite sign will be the largest possible underbinding result.

GGA-type functional is also often seen in QM/MM studies.$^{84-87}$ Comparing figures 2 and 3, the differences in how the two functionals (and therefore the QM/MM model) compare to TIP3P becomes evident. The absolute values for the overall full QM-full TIP3P differences (as represented by the red patches in the figures) are smaller when using the PBE functional to describe the water clusters. However, BLYP is consistent in producing smaller absolute interaction energies, giving consistently positive $\Delta E$-values, and underbinding the water clusters, compared to TIP3P. Furthermore, we observe how the QM/MM interaction energies systematically increase towards the maximum full QM vs full MM difference, when the number of QM molecules in the QM/MM calculations is increased. This systematic behavior could potentially open up for more elaborate tuning strategies for the explicit in-
Figure 3: Box plot of $\Delta E$, as in figure 2, but this time using the BLYP functional for the QM subsystem (TIP3P is still used for the MM part). Compared to the PBE results, the interaction energy difference between the full QM and the full MM description is larger, but for BLYP, it is always positive, and shows signs of systematic convergence. This also makes the QM/MM description more consistent, with a consistently increasing $\Delta E$ for increasing number of BLYP water molecules.

interaction, e.g. using constrained DFT to minimize this difference in the full QM and full MM descriptions.  

Lastly, figure 4 shows PBE/TIP4P interaction energy differences. Again, with very few exceptions, the middle 50% of the distributions of QM/MM interaction energy differences are within the maximum difference between full QM and full MM results. Here, however, the full MM TIP4P potential consistently produces smaller interaction energies than the full QM PBE description, with an absolute QM-MM difference of roughly half that of BLYP/TIP3P. Again, we observe the systematic increase of QM/MM binding energy with increasing the number of QM molecules of each cluster

In conclusion, the electrostatic embedding interface introduces no systematic energy errors exceeding the differences between the QM and MM description.

For a comparison of pure grid-based results with LCAO-mode results on the interaction energies of the Bates & Tschumper geometries, see the Supporting Info.
Figure 4: Box plot of $\Delta E$, as in figure 2, but this time using PBE/TIP4P

4.3 Benchmark: Liquid Water

Figure 5 shows how the QM/MM radial distribution functions (RDFs) have a set of features combined from the pure MM and pure QM description. The well-known over-structuring of water due to PBE$^{71,77,89,90}$ is reproduced in the QM/MM results, as is evident from the large peak-height, so the pseudostructure over the QM/MM interface is internally consistent with the electrostatic embedding interface we have developed for GPAW. The QM/MM RDFs shown have been produced using the Smoothly Switching (SS) short range potential (See Supporting Info). Results from a previously developed short-range method based on Gaussian Charge Distributions (GCD)$^{65}$, as well as results using an external MM code$^{91}$ are nearly identical (see S.I appendix D.) The same picture reveals itself when extending the liquid quasi-structure analysis to angular distributions, as seen in figure 6. The literature suggests three angular distributions of importance for the hydrogen-bond structure of water.$^{71,92-95}$ The donor angle $\alpha = \angle O_D - H_D - O_A$, the acceptor angle $\theta = \angle H_D - O_A - H_A$, and bonding angle deviation from linearity $\beta = \angle H_D - O_D - O_A$ as, illustrated in figure 6). We have used a hydrogen-bonding criterion similar to previous work$^{71,92}$ such that a hydrogen
Figure 5: MM and QM/MM Pairwise radial distribution functions (RDFs) of the O-O correlations in a cubic 20 Å box of water, produced with the implementation presented in this work. ~1.2 ns of equilibrated dynamics were sampled for the QM/MM simulations, where a single water molecule was described with PBE. The resulting RDF, \( g_{\text{OQM}O\text{MM}}(r) \) was sampled from the single QM Oxygen to the rest of the water oxygen atoms, modeled with TIP3P. The pure PBE RDFs were digitized from work presented elsewhere,\(^7^1\) which unfortunately did not contain data for the H-H correlation.

Bond is deemed to exist if \( r_{\text{OO}} < r_{\text{max}} - 0.00044\delta^2 \), where \( r_{\text{OO}} \) is the O-O distance, \( r_{\text{max}} \) is the maximum radius of the first (O-O) solvation shell, and \( \delta \) is the hydrogen bonding angle \( \beta - \beta_{\text{opt}} \). The angle \( \beta_{\text{opt}} \) is the optimal \( \beta \) angle of the water dimer from the s22 dataset.\(^7^5\)

Thus, this relation defines a geometrical, conical criterion for hydrogen bonding, which includes a demand for directionality of the bond, which increases with the O-O distance, as
Figure 6: Left, top & bottom: Pairwise molecular angular distribution functions in liquid water. The top plot shows the distribution of the $\alpha$-angle. The bottom plot is created to facilitate a comparison with pure PBE liquid water data obtained from the literature,\textsuperscript{71} where the $\beta$ and $\theta$-angles were not sampled individually, and the $\beta'$ angle was also included in the total sampling. The black curves not connected with lines show how the individual distributions make up the full curve, based on the TIP3P dataset. The hydrogen bonding criterion is visualized in the top left insets. The right bar plot shows the hydrogen bond count of each type of simulation. The slight trend of PBE overbinding when compared to TIP3P continues from the RDFs, and is therefore also observed in the QM/MM results. Again, there are no major differences between results from the two different short range cutoffs.

visualized in the top left of the figure. In the upper plot in figure 6, the distribution of $\alpha$ angles show more angular rigidity when going from the TIP3P to the PBE description, with a further QM/MM increase, mirroring the QM/MM RDF over-structuring. The bottom plot, made to facilitate direct comparison to pure PBE results from literature,\textsuperscript{71} again shows this increased rigidity in the $\beta$ angle. Since the TIP3P molecules and the QM molecule of the QM/MM simulations are completely rigid, the bottom plot angle distributions have no contributions from around 40 - 60 degrees. The QM/MM hydrogen bond count distributions are very similar to their TIP3P counterpart, indicating a good preservation of description of the hydrogen bond network over the QM/MM border.
5 Results: Solution Structure and Dynamics of PtPOP in Water

Figure 7: Left: The potential of mean force (PMF) calculated from the Pt-Pt RDF, together with the relative thermal distribution ($\Gamma(r)$) from the QM/MM MD simulations. The red line defines the average thermal energy available to the system at 300 K. Top right: Illustration of the diplatinum complex PtPOP. Pt atoms in silver, oxygens in red, phosphorus in orange and hydrogens in white. Bottom right: Fourier transformation of the Pt-Pt oscillation in the simulations.

Figure 7 shows the 300 K distribution of Pt-Pt distances in PtPOP together with the free energy surface obtained as potential of mean force (PMF):\(^{96}\)

\[
w(r) = -k_B T \ln(g_{\text{Pt-Pt}}(r))
\]  

(15)

where $g_{\text{Pt-Pt}}(r)$ is the pairwise Pt-Pt RDF. The thermally averaged Pt-Pt distance of PtPOP in aqueous solution in our simulations was 2.99 Å. A previous QM/MM MD study by Penfold et al. obtained the value of 3.06 Å.\(^{53}\) The experimental value, obtained with X-ray Diffraction (XRD) is 2.98Å.\(^{51}\)

From a Morse-potential fit to the PMF a low degree of anharmonicity of the Pt-Pt stretching vibration can be deduced. The degree of deviation from harmonicity was estimated
by calculating the anharmonicity constant $x_e$ according to $x_e = \hbar/(4D_e) \sqrt{k_0/\mu}$, with $D_e$ and $k_0$ the depth at the minimum and the force constant of the potential of mean force respectively, and $\mu$ the reduced mass of Pt$_2$. Using these parameters, a value of $x_e = 1.5 \cdot 10^{-3}$ is obtained, which is noticeably small if compared for example to that of a very harmonic diatomic system as I$_2$, $97 \cdot 2.8 \cdot 10^{-3}$. The position of the minimum of the Morse-fitted potential at 2.98 Å, a Pt-Pt distance only 0.01 Å shorter than the thermally averaged value also points to a remarkable harmonicity. As a further and last confirmation of the harmonicity, the vibrational frequency calculated from the force constant and reduced mass of Pt$_2$, 124 cm$^{-1}$ (270 fs period), is very close to the frequency obtained from the maximum of a Fourier transform (FT) of the oscillating Pt-Pt distance in the simulation and shown in figure 7. With respect to this, we also note that the value obtained from the FT analysis, 121 cm$^{-1}$ (275 fs), deviates by less than 2 % from the vibrational frequency of 119 cm$^{-1}$ (281 fs) recently obtained by van der Veen et al.$^{52}$ using femtosecond transient absorption measurements in water solution. The femtosecond optical experiments performed by van der Veen et al. highlighted an exceptional picosecond-long coherence decay for vibrational population relaxation in the first singlet excited state of PtPOP, which was interpreted as a direct consequence of the high harmonicity of the Pt-Pt bond vibrations; the latter explained as being due to the rigidity of the P-O-P bridging ligands in PtPOP. Here we show that this is not only a feature of PtPOP in the excited state, but it also applies to the Pt-Pt vibrations in the ground state.

Finally, we turn to a discussion of the solvent shell structure as extracted from the simulations. Figure 8 shows the Pt-O$_{\text{solvent}}$ and Pt-H$_{\text{solvent}}$ RDFs calculated from all the 463 ps of QM/MM MD data collected in this study. The large amount of statistics allows to fully resolve the first four peaks of solvent coordination around the Pt atoms. In particular, a peak is observed at short distances (about 3 Å in the case of the Pt-O$_{\text{solvent}}$ RDF), which arises from the presence of water molecules coordinated in axial position.

By further analyzing the orientation of the water molecules in the first peak in the RDFs,
we can address the question of the nature of the Pt-Pt axial coordination of water molecules. Figure 9 shows distributions of key angles in the solute-solvent geometry, in analogue to the angular distributions presented for neat water in figure 6. While there is no 1:1 correspondence in definitions of angles between the two different systems, all three sampled angles indicate a preference for linear geometry, or axial coordination, with one hydrogen atom pointed along the Pt-Pt axis. This is further supported by the distance between the first peaks in the Pt-O and Pt-H RDFs, which is roughly 0.96 Å, the TIP4P O-H bond length. The OH-Pt angle, comparable to the water α angle, has a very similar distribution to the one obtained from the hydrogen bonded geometry produced by the QM/MM MD simulations of neat water. Using an analogous hydrogen-bonding criterion as for neat water used in the previous section,\textsuperscript{71,92} we can quantify the average amount of hydrogen bonded water molecules per PtPOP molecule, as seen in the right part of figure 9. The extent of the first solvation shell is 3.85 Å, as observed from the \(g_{\text{PtO}}(r)\) in figure 8. Due to the symmetry of PtPOP, we assume that the optimal angle HO-Pt (akin to \(\beta_{\text{opt}}\) of water) for hydrogen
bonding criterion angle

Figure 9: Left: Angular distributions of water molecules sampled from the first peak ($r_{\text{max}} = 3.85$ Å) of the RDFs presented in figure 8. The fully drawn lines are the hydrogen bonded distributions, and the dashed lines are the total (both bonded and non-bonded) distributions. Right, top: Illustration of the angle used for the bonding criterion relation shown in figure 6. Right, bottom: Number of hydrogen bonded waters per PtPOP molecule from the simulation.

bonding is 0. It is clear that there is a predominant trend of water coordination, with one or both Pt-ends being coordinated almost all of the time. The remarkable harmonicity of the Pt-Pt oscillation in PtPOP shows that the internal dynamics are largely unperturbed by the solvent. This isolation explains that coordination is still possible without the system displaying any large solvatochromic shift, if the solvent coordination stays the same overall, in the excited state, for any solvent used.

The aforementioned previous QM/MM study of the same metal complex used a box with 1108 water molecules and collected 1 ps of equilibrated BO dynamics. In that study, simulations were performed using the CPMD package using BLYP and plane waves for PtPOP, and TIP3P for the solvent. The authors used 20 QM/MM configurations from the 1 ps trajectory to simulate the solution X-ray absorption spectrum, from which they derived a shift with respect to the spectrum calculated using the vacuum optimized geometry of PtPOP. However, as the authors point out, the limited amount of sampling precluded them from obtaining an accurate, statistically meaningful, description of the solvent structure.
and hence from drawing deeper insight into the solvent induced changes in the spectrum. Any experimental technique that is sensitive to the structural dynamics of the complex will inevitably probe the stretching along the Pt-Pt coordinate, as it was demonstrated by low-temperature and solution time-resolved optical studies.\textsuperscript{52,100} Therefore, a detailed characterization of the thermal distribution of Pt-Pt distances and motion on the underlying potential energy surface is an essential prerequisite that a computational method should meet in order to have more authority in the analysis and interpretation of these kinds of experiments. Due to the large number of degrees of freedom involved, in fact, it is quite likely that an MD simulation of limited length would attach an unfair weight to a particular solvent configuration.

6 Conclusions

We have presented a real space, grid-based Hamiltonian QM/MM coupling model, that interfaces computationally expedient finite-difference DFT calculations using GPAW to classical point charge potentials. We have also shown how it can successfully be used to model the solvent coordination of a metal complex. Further uses of the methodology includes out-of-equilibrium dynamics, solvent responses, and more.

The single point energy accuracy of the method and code has been ascertained by systematic benchmarks: Starting with a comparison of the interaction energy of water clusters, we found the LCAO-mode of GPAW achieve energies closest to the full QM or full MM description, which we ascribe to the more localized basis being better suited in conjunction with the chosen analytical potentials for short-range electronic interaction over the QM/MM border, and maybe also counteracting some of the PBE-caused overbinding. Then, we increased the size of the dataset to encompass many more water cluster geometries from three to ten molecules, and showed that on average, using PBE, the QM/MM interaction energy difference between QM/MM and pure MM descriptions are a few tens of meV per water
molecule, and that the electrostatic embedding interface introduces no systematic energy errors exceeding the differences in the QM and MM description. The QM vs MM difference is larger when using the BLYP functional, but it is systematic, which opens the possibility of a systematic error correction for future work.

The implementation is also used to carry out QM/MM MD simulations of a single QM water in a TIP3P environment, which showed that the interface produces results consistent with the QM and MM model. We also analyzed the angular distribution around hydrogen bonds over the QM/MM boundary, and obtained results supporting the RDFs and a combination of the PBE and TIP3P description of liquid water.

Lastly, we have investigated solution properties of a diplatinum complex, where our method allowed for the collection of statistically significant MD data to reliably compare the simulation results with experimental observations. Sampling the internal degrees of freedom of the molecule produced structural agreement within 0.01 Å of experimental results, and the remarkably strong harmonicity of the Pt-Pt vibrational mode was confirmed. We also quantified the solute-solvent coordination, the extent of which has been previously debated. We find a predominant trend of water coordination, with one or two ends the complex being coordinated 95 % of the time (48 % probability of double coordination, and 47% single coordination). We intend to expand on this study with future work on electronically excited states and experimental comparisons.

The electrostatic embedding implementation presented is currently exclusive to GPAW, although preliminary support for Turbomole and DFTB+ is available at the time of writing through the development version of ASE. Other DFT codes could be used by adding an interface handling I/O of the QM charge density on a grid, and the MM point charge potential. Due to the modularity of the calculator-object-based implementation, expanding on the method can be done in a straightforward manner, building upon the work presented here. We plan to expand the method to encompass adaptive QM/MM methods, polarizable embedding in more advanced and accurate force fields, and excited states via ∆SCF-
TD-DFT methods, which are already implemented in GPAW.\textsuperscript{103–106}

The QM/MM implementation is available in the official releases of GPAW and ASE, see https://wiki.fysik.dtu.dk/gpaw/ and https://wiki.fysik.dtu.dk/ase/

Acknowledgement

The authors thank S. Olsson for figure visualization discussions and advice.

This work was supported by the Icelandic Research Fund, the Academy of Finland, the Danish Council for Independent Research, and Villum Fonden.

Supporting Information Available

Short-and Long-Range electrostatic considerations, further details of the PAW formulation of the Electrostatic Embedding Scheme, including force expressions, code implementation and overview schematics, and further benchmark details, such as interface efficiency and robustness. This material is available free of charge via the Internet at http://pubs.acs.org/.
References


(27) Knorr, J.; Sokkar, P.; Schott, S.; Costa, P.; Thiel, W.; Sander, W.; Sanchez-Garcia, E.;


Laino, T.; Mohamed, F.; Laio, A.; Parrinello, M. An Efficient Linear-Scaling Electro-


Horn, H. W.; Swope, W. C.; Pitera, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.;


