Technical Report

Improvements in periodic representation of solvated systems with FHI-aims

Abstract

The FHI-aims software package can model molecular and periodic systems on the same numerical infrastructure and is known to be particularly good for modeling of surface structure and related adsorption/reaction processes. Due to FHI-aim's representation of electronic structures with state-of-art basis sets, a good balance of accuracy, performance, and memory requirements is achieved for simulation of surface models. Through this eCSE project, the capabilities of FHI-aims' to simulate on-surface adsorption have been improved. Specifically: the robustness of the Basis Set Superposition Error (BSSE) correction has been increased for periodic systems, including surface slabs; the implementation of the MPE (Multipolar Potential Expansion) implicit solvation model has been extended to support periodic systems, such as surface slabs; and the sparsity pattern of a large matrix used by the MPE subroutines has been dramatically improved.

Keyword: ARCHER2, eCSE, Density Functional Theory, FHI-aims, Implicit Solvent Model

Motivation

The FHI-aims code [1] is widely used for simulation of various on-surface processes including heterogeneous catalysis and adsorption. However, until this eCSE project there were notable functionality gaps that complicated typical computational studies of this kind, or hampered their accuracy. The key advantage of FHI-aims for surface simulations is the real-space localized basis set, which allows efficient representation of any thickness of a vacuum layer above the simulated surface without additional computational expense. Indeed, many density functional theory (DFT) codes tailored to simulate periodic systems use a plane-wave basis set, which introduces substantial memory requirements for surface models. The drawback of the real-space localized basis set is the difficulty involved in implementing periodic boundary conditions, and functionality gaps which existed only for periodic systems, including:

1. The so-called "ghost" atoms were only robustly supported for non-periodic systems.
2. There was no implicit solvent model for periodic systems.

Both omissions from the original code have been addressed in this project.

"Ghost" atoms

Basis functions localized in real space are commonly parametrized by coordinates of a point, called an expansion center. In DFT codes, atomic nuclei are typically used as expansion centers, but additional basis functions with expansion centers different from atomic coordinates are sometimes favorable. Such basis expansion centers are called "ghost" atoms; they are used for electron structure description, but unlike "real" atoms they

do not add charge or mass to the system. Therefore, they require special treatment in DFT codes. For electronic structure implementations with an atom-centred basis, the “ghost” atoms are required calculation of a Basis Set Superposition Error (BSSE) correction [2]; in FHI-aims, they also provide specific functionality necessary for some QM/MM schemes implemented in the ChemShell framework [3].

Implicit solvent model

Many physical and chemical processes important for industry and medicine occur in a solvent environment. Explicit simulation of the solvent phase is extremely expensive because it is both highly mobile and is composed of many constituent solvent molecules.. An exhaustive number of configurational samples is necessary for explicit calculation of solvent’s entropic contribution into the free energy; however, thanks to the mobility and disorder of solvent molecules, their contribution can be time-averaged and approximated with implicit models. Implicit solvent models represent the solvent as a continuous media with specific dielectric permittivity that responds to the electrostatic potential created by the solute.

FHI-aims implements two implicit solvent models: Modified Poisson-Boltzmann (SMPB) and Multipole Expansion (MPE). Both of these models use the concept of *solute cavity*, *i.e.*, a space region unavailable for solvent molecules due to spatial repulsions of the solvated species or solids. The solute cavity region has dielectric permittivity that differs from the solvent region. The key difference between SMPB and MPE models is the method to represent the cavity and its boundary; the MPE model implies sharp change of the dielectric permittivity on the boundary, whilst SMPB uses smooth transition.

Implementations of both models in FHI-aims are actively developed, and before this project neither provided support for periodic boundary conditions.

MPE solvent model

The detailed description of the MPE model can be found Ref. [4]. An important recent modification that improved scalability and accuracy of the model has been introduced in Ref. [5]. Here, we provide a birds-eye view of the model with focus on key features of its implementation in FHI-aims.

The MPE model modifies the electronic potential in the Kohn-Sham equation by adding the electrostatic potential that “results from polarization of the surrounding dielectric medium” [4]. This additional potential (further referred to as MPE potential) is described as a linear combination of regular and irregular solid harmonics centered on nuclei of heavy atoms (non-hydrogen). The regular harmonics are used to represent the MPE potential within the solute cavity, and irregular harmonics are used for the solvent region. The total potential should be continuous on the cavity boundary. This continuity requirement can be written as a system of linear equations (SLE) with the coefficients of the solid harmonics as the unknown variables (see Eq. 15 in Ref. 4). The problem can be represented in the matrix form thus:

$$Ac = b \tag{1}$$

Each linear equation in (1), *i.e.*, each row of matrix *A*, corresponds to a point on the cavity interface, where the boundary conditions are defined (In practice, there are two rows for each point, to account for both the potential and potential gradient continuity constraints). **c** is the vector of coefficients of solid harmonics, and **b** is the vector of constants that depend on the potential induced by the solute at the corresponding point. FHI-aims provides multiple methods to define the mesh of interfacial points; some define a static mesh before DFT calculations, whilst other options update the mesh dynamically between the SCF (self-consistent field) loop iterations, adjusting it to the equilibrating electronic density. In contrast, the vector **b** changes on every SCF step due to the change in the potential. Overall, the key computational task of the MPE model is the solution of Eq (1).

In the original MPE formulation, the matrix *A* in Eq (1) was dense. To increase the scalability and numerical stability of the MPE model, it was recently proposed [5] to split the solute cavity onto subcavities around each heavy atom. In that way, the MPE potential in a subcavity was represented using solid harmonics centered at *only* the atom in that subcavity. As a consequence, the boundary conditions on each subcavity interface directly affects only a specific subset of the multipolar expansion. The modification adds boundaries between subcavities, and therefore new boundary conditions and rows to the matrix *A* in Eq (1), but simultaneously *A* becomes sparser (See mpe-nc pattern, Figure 1). The matrix sparsity provides opportunity for performance gains, but *A* still has an undesirable dense block that corresponds to the MPE potential in the solvent region.

To make the whole of matrix *A* sparse, and therefore to open the way to usage of high-performant sparse linear solvers, the locality of solid harmonics was exploited by further splitting of the solvent region to increase the matrix sparsity, similar to the implementation for the cavity region. Moreover, the piecewise representation of the solvent region would provide a natural way to introduce periodic boundary conditions, which is a highly desirable feature for molecular modelling.

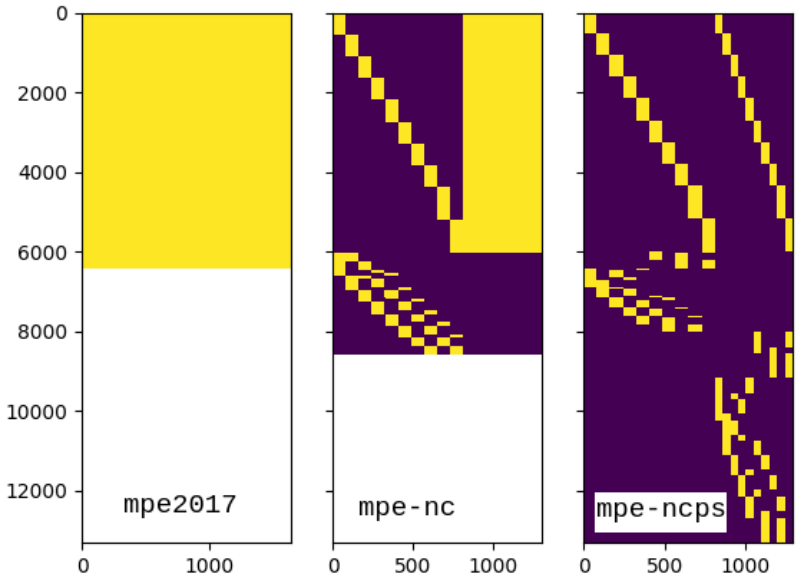

Figure 1. Sparsity patterns of the MPE matrix [see Eq. (1)] for different implementations. Blue and yellow colours represent zero and non-zero elements correspondingly. The implementations considered are: `mpe2017` - the original MPE version [4]; `mpe-nc` – MPE with solute subcavities [5]; `mpe-ncps` – new MPE implementation with piecewise solvent.

Results

Piecewise solvent for MPE model

In this project we have implemented the piecewise representation of the solvent region for the MPE model in the FHI-aims code. To achieve this outcome, the MPE matrix construction subroutines were significantly refactored, with new solute-solute interfaces introduced between solute subregions. Furthermore, the solute-solvent boundary conditions have been adopted to use a single-atom basis set for a solvent side.

To realise the complementary objective of adding support of periodic boundaries for the MPE model, we decided to employ a designated library for space tessellation. After a review of available packages, Voro++ by Chris H. Rycroft [6] was adopted as it is capable to build Voronoi tessellation of both periodic and non-periodic systems on an equal footing. As the space tessellation inside the solute cavity must coincide with partitioning of the outer solvent region, to prevent complications in building the solute-solvent interface, we introduced Voro++ functionality for both solute-solute and solvent-solvent interfaces.

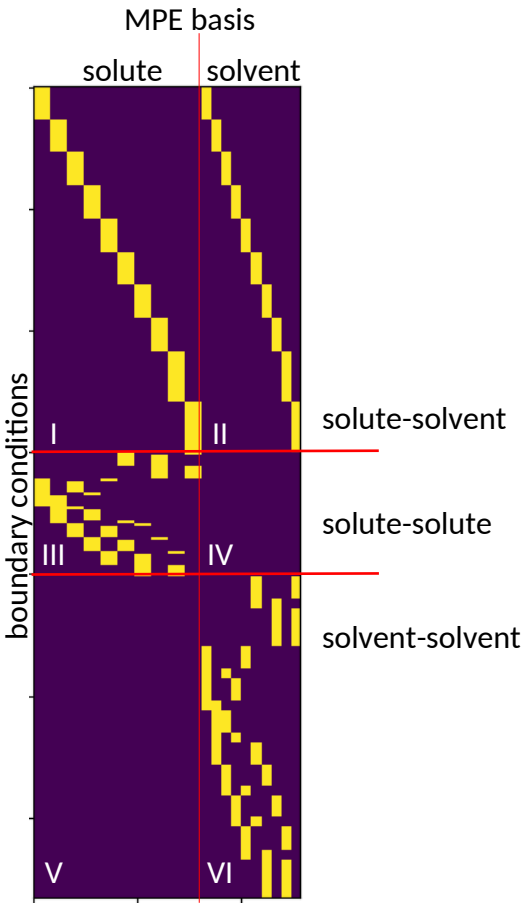

Figure 2. Structure of the MPE matrix [Eq. (1)], with labels I-VI representing the different blocks for solute and solvent interactions.

For a better numerical stability of the MPE algorithm, it is desirable to place a boundary between subcavities as close to the minimum of electronic density as possible. Therefore, the free-atom electron density is used for boundary definition, with the boundary between subcavities shifted appropriately to account for differences in the isodensity radii of the atomic species. The Voro++ library provides a radical (*i.e.* weighted) Voronoi tesselation algorithm that can set a size for each atom, making it capable of producing the space partitioning with boundaries over low electron density regions.
The achieved sparsity pattern of the MPE matrix *A* is shown in Figure 1 as `mpe-ncps`, with the physical meaning of the matrix blocks illustrated in Figure 2. In Figure 2, block II becomes sparse in the new implementation, corresponding to the solvent side of the boundary condition of the solute-solvent interface. Also, blocks V and VI are introduced because there were no solute-solute interfaces previously (similar to the absence of blocks III and IV for the `mpe2017` model). Blocks IV and V are always zero by construction.

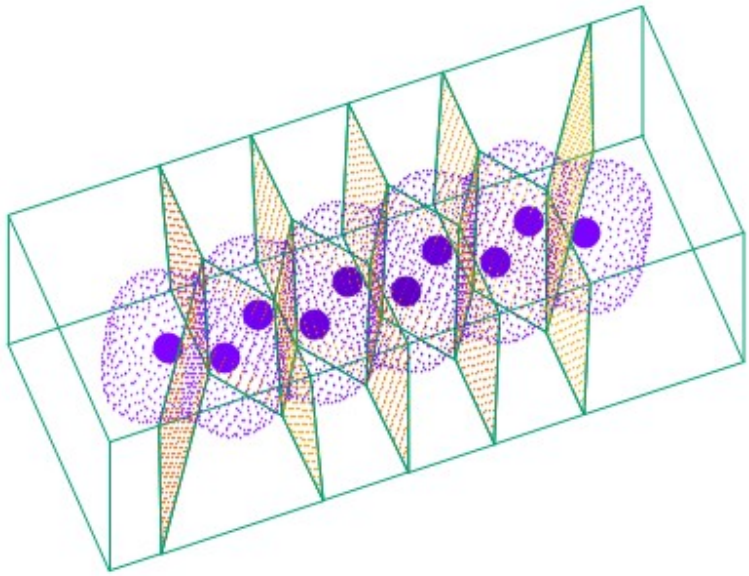

Figure 3. Interfaces between the MPE model subregions of the decane molecule. Big blue circles - carbon atoms, hydrogens are not shown. Blue dots – solvent-solute interface mesh. Yellow dots – solute-solute and solvent-solvent interfaces mesh. Green lines – limits for solute-solute interface mesh (*not* a periodic lattice cell).

The size of matrix *A* was increased because of the added solute-solute interfaces, and corresponding boundary conditions, but the number of added rows can be reduced, because the number of points on the solute-solute boundaries is probably excessive now. Currently it is limited by the cuboid box which size is defined by a 3 Å gap to the nearest heavy atom; in Figure 3, this box is shown by green lines. Benchmarks are necessary to define reasonable and sufficient default settings that will define the number of points on the solute-solute interfaces, and this work is part of our future plans.

Despite the increased size of the MPE matrix, the number of non-zero elements decreased dramatically due to greater sparsity. Table 1 demonstrates the difference between MPE matrices used by the available MPE versions in FHI-aims. For a decane molecule, the `mpe-ncps` matrix has 3.5 times less non-zero elements than the original

implementation. For a larger system - a graphite slab of 128 atoms - the difference is even larger at 14 times (relatively 0.76% vs. 15.53%). This comparison is done for non-periodic slab, because earlier MPE implementations do not support periodic boundaries. Looking closely at the matrix structure (Figure 2), one can see that the number of non-zero entries in each row is constant. Indeed, the matrix sparsity does not depend on the number of columns (or atoms), as each row always has only two non-zero blocks. The number of interfacial points, and therefore the number of matrix rows, grows linearly with the number of atoms; hence, the total number of non-zero elements in the matrix is $O(n)$, where n is the number of heavy atoms in the system.

Currently, exploitation of the introduced sparsity for performance gain is infeasible because the SLE solvers supported by FHI-aims for the MPE model are not compatible with the new sparsity scheme. This was an unforeseen issue in the project realisation and will be resolved by integration of such solvers in future work.

Table 1. Dimensions of the MPE matrix for a decane molecule (non-periodic system) with different MPE implementations.

MPE version	# rows	# columns	# of non-zero elements	Portion of non-zero elements, %
`mpe2017` [4]	6 442	1 649	10 619 637	99.97
`mpe-nc` [5]	8 574	1 299	3 833 865	34.42
`mpe-ncps` (Current project)	13 316	1 290	1 605 871	9.35

Support of periodic boundaries for MPE model

Building on the new piecewise solution, subsequent work achieved support for periodic boundaries in the MPE model. We successfully implemented the capability to build interfaces between periodic images of subcavities and solvent subregions, with minor complications noted herein.

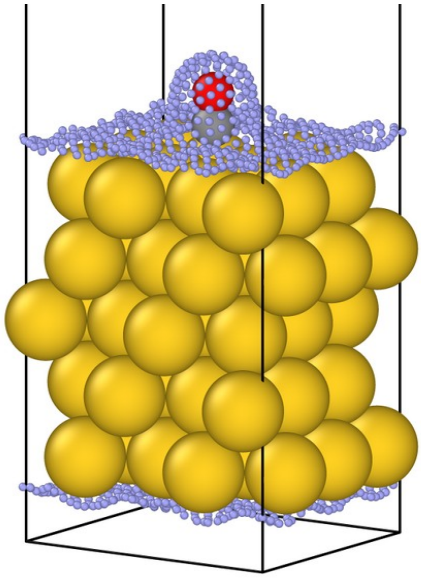

Figure 4. The Au(111) slab (yellow spheres) with adsorbed CO molecule (black and red spheres). Small blue spheres represent the point mesh of solute-solvent interface built by the MPE model. Black lines represent the unit cell.

With the work performed, multiple interfaces between the same subregions can now exist, *i.e.* across periodic boundaries. The change to support periodic MPE solvation required further refactoring of the MPE matrix construction subroutines. Furthermore, selection of the minimum length vector among all periodic images had to be introduced in numerous places of the code to manage periodicity; and a new algorithm for calculating the cavity volume was introduced, because the previous algorithm did not work appropriately with an infinite periodic system (Figure 4). The new algorithm uses a numerical integration method, so careful convergence testing has become necessary, with benchmarking to define a reasonable default integration parameter crucial in our future work.

Extension of regression test suite

To complement the outlined new functionality, a range of regression tests have been added to guarantee the sustainability of the newly implemented and earlier existing functionality. The tests for the "ghost" atoms in periodic systems have been implemented for all Hamiltonians robustly supported by FHI-aims: DFT, exact exchange (hybrid DFT), MP2 (double hybrid DFT), and PBE-GW. A special test suite was also added that ensures proper calculation of forces acting on "ghost" atoms that is necessary for some QM/MM schemes, specifically supported by the ChemShell framework [3]. Several tests have been added for the newly implemented `mpe-ncps` solvent model for both periodic and non-periodic systems.

Conclusions and future work

The new important functionality has been added to the FHI-aims code and reliability of existing functionality has been ensured. Solvent effects now can be included in calculations of adsorption and reaction processes on material surfaces, as well as properties of materials themselves in solvated environments. The BSSE correction can be calculated for any periodic model and most popular Hamiltonians, which is particularly beneficial for adsorption on surfaces and in porous periodic systems like zeolites. The work provides a foundation for substantial acceleration of calculations with the MPE solvent model, *via* the dramatically increased sparsity of the MPE matrix. Future efforts should be focused on integration of the sparse linear solvers for MPE SLE, and on benchmarking of periodic systems with MPE solvents in order to determine optimal default settings for numerous parameters of the new method.

Acknowledgments

This work was funded under the embedded CSE programme of the ARCHER2 UK National Supercomputing Service (http://www.archer2.ac.uk). G.B. and A.L. gratefully appreciate funding by the UKRI Future Leaders Fellowship program (MR/T018372/1).

Bibliography

1: Volker Blum, Ralf Gehrke, Felix Hanke, Paula Havu, Ville Havu, Xinguo Ren, Karsten Reuter, Matthias Scheffler, Ab initio molecular simulations with numeric atom-centered orbitals. Computer Physics Communications2009; 180; 2175-2196.
2: Roman M. Balabin, Enthalpy difference between conformations of normal alkanes: Intramolecular basis set superposition error (BSSE) in the case of n-butane and n-hexane. J. Chem. Phys.2008; 129; 164101.
3: Y. Lu, M. R. Farrow, P. Fayon, A. J. Logsdail, A. A. Sokol, C. R. A. Catlow, P. Sherwood and T. W. Keal, Open-Source, Python-Based Redevelopment of the ChemShell Multiscale QM/MM Environment. J. Chem. Theory Comput.2019; 15; 1317-1328.
4: Markus Sinstein, Christoph Scheurer, Sebastian Matera, Volker Blum, Karsten Reuter, and Harald Oberhofer, Efficient Implicit Solvation Method for Full Potential DFT. J. Chem. Theory Comput2017; 13; 5582–5603.
5: Jakob Filser, Karsten Reuter, Harald Oberhofer, Piecewise Multipole-Expansion Implicit Solvation for Arbitrarily Shaped Molecular Solutes. J. Chem. Theory Comput.2022; 18; 461–478.
6: Chris H. Rycroft, A three-dimensional Voronoi cell library in C++. Chaos2009; 19; 041111.