The Journal of Chemical Physics

# Embedded-cluster calculations in a numeric atomic orbital density-functional theory framework

Cite as: J. Chem. Phys. **141**, 024105 (2014); https://doi.org/10.1063/1.4885816
Submitted: 08 April 2014 • Accepted: 18 June 2014 • Published Online: 10 July 2014

**Daniel Berger, Andrew J. Logsdail, Harald Oberhofer, et al.**

View Online

Export Citation

CrossMark

**ARTICLES YOU MAY BE INTERESTED IN**

A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu
The Journal of Chemical Physics **132**, 154104 (2010); https://doi.org/10.1063/1.3382344

CP2K: An electronic structure and molecular dynamics software package - Quickstep: Efficient and accurate electronic structure calculations
The Journal of Chemical Physics **152**, 194103 (2020); https://doi.org/10.1063/5.0007045

The ORCA quantum chemistry program package
The Journal of Chemical Physics **152**, 224108 (2020); https://doi.org/10.1063/5.0004608

AIP Publishing



© 2014 AIP Publishing LLC.

# Embedded-cluster calculations in a numeric atomic orbital density-functional theory framework

Daniel Berger,[1,a)] Andrew J. Logsdail,[2,b)] Harald Oberhofer,[1] Matthew R. Farrow,[2] C. Richard A. Catlow,[2] Paul Sherwood,[3] Alexey A. Sokol,[2] Volker Blum,[4] and Karsten Reuter[1]

[1]*Chair for Theoretical Chemistry and Catalysis Research Center, Technische Universität München, Lichtenbergstr. 4, D-85747 Garching, Germany*
[2]*Department of Chemistry, Kathleen Lonsdale Materials Chemistry, University College London, 20 Gordon St., London WC1H 0AJ, United Kingdom*
[3]*Scientific Computing Department, STFC Daresbury Laboratory, Daresbury, Warrington, United Kingdom*
[4]*Department of Mechanical Engineering and Materials Science, Duke University, Durham, North Carolina 27708, USA*

(Received 8 April 2014; accepted 18 June 2014; published online 10 July 2014)

We integrate the all-electron electronic structure code `FHI-aims` into the general `ChemShell` package for solid-state embedding quantum and molecular mechanical (QM/MM) calculations. A major undertaking in this integration is the implementation of pseudopotential functionality into `FHI-aims` to describe cations at the QM/MM boundary through effective core potentials and therewith prevent spurious overpolarization of the electronic density. Based on numeric atomic orbital basis sets, `FHI-aims` offers particularly efficient access to exact exchange and second order perturbation theory, rendering the established QM/MM setup an ideal tool for hybrid and double-hybrid level density functional theory calculations of solid systems. We illustrate this capability by calculating the reduction potential of Fe in the Fe-substituted ZSM-5 zeolitic framework and the reaction energy profile for (photo-)catalytic water oxidation at $TiO_2$(110). © *2014 AIP Publishing LLC*. [http://dx.doi.org/10.1063/1.4885816]

## I. INTRODUCTION

Periodic boundary conditions (PBC) and their use in supercell geometries constitute a most efficient approach to bulk and surface electronic structure calculations of solids. While focusing the computational effort on the finite supercell, this approach elegantly captures, e.g., long-range electrostatic effects and in particular the delocalized electronic bonding in metallic systems. Much of this elegance and efficiency is lost though in the application to localized perturbations of the lattice periodicity, for instance, in form of defects or adsorbates. Large supercells are then needed to suppress spurious interactions with periodic images, and for explicitly charged systems intricate correction schemes are required to reach convergence with respect to supercell size at all.[1–3] For non-metallic systems, this has long motivated an alternative solid-state embedding approach in form of hybrid quantum and molecular mechanical (QM/MM) calculations.[4] Here, a finite "quantum region" is embedded in a surrounding environment, generally modeled using classical molecular mechanics ("MM region"). Such aperiodic embedded-cluster models do not suffer from spurious interactions between defects, adsorbates, and charges. Simultaneously, the extended MM environment ensures correct long-range electrostatics and elasticity, and mitigates quantum confinement effects in comparison to approaches based on bare or hydrogen-saturated clusters.

The exploration of advanced density-functional theory (DFT) exchange-correlation (xc) functionals for solid-state applications adds another motivation for the QM/MM ansatz. With computationally particularly efficient implementations presently achieved for localized basis sets and finite systems, solid-state embedding promises unprecedented access to extended systems. With this objective, we here present a QM/MM implementation integrating the `FHI-aims` program package[5] into the general `ChemShell` framework.[6,7] Exploiting tailored numeric atomic orbital (NAO) basis sets,[5,8] the resolution-of-identity technique[9] and massively parallel algorithms[10–12] `FHI-aims` offer a wealth of tractable functionality beyond semi-local DFT. This includes in particular exact exchange and second order perturbation theory (PT2),[13] rendering `FHI-aims` an ideal platform to test the performance of hybrid and double hybrid xc functionals. On the other side, `ChemShell` is a powerful computational chemistry environment, which takes over the communication and data handling in hybrid QM/MM calculations. For solid state embedding in particular, Chemshell allows the use of polarizable core-shell model potentials to fit the electrostatic potential to that of a periodic calculation.[6,7]

In order to prevent spurious charge leakage out of the QM and into the MM region, solid-state embedding often requires the introduction of an intermediate shell at the QM boundary. In this shell, all cations are replaced by effective core potentials, which prevent the spurious overpolarization of the electron density through adjacent bare positive MM charges, i.e., a spurious charge leakage into the MM region. A

a)Electronic mail: daniel.berger@ch.tum.de
b)Electronic mail: a.logsdail@ucl.ac.uk

 © 2014 AIP Publishing LLC

major part of the FHI-aims–ChemShell coupling therefore concerns the implementation of efficient pseudopotential (PP) functionality into the all-electron full-potential code FHI-aims. Tailored to the NAO basis sets of FHI-aims we specifically achieve this through norm-conserving PPs of Kleinman-Bylander type.[14]

The performance and validity of the resulting QM/MM implementation is demonstrated using Fe-substituted ZSM-5 zeolites and the $TiO_2$(110) surface as examples for semi-covalent and ionic systems, respectively. For the former, the purely electrostatic embedding is successfully benchmarked against previous QM/MM calculations using the Gamess-UK package for the QM calculations.[15] For the latter, we exploit the FHI-aims capability of calculating finite and periodic systems within the same numerical framework and systematically compare the solid-state embedding against PBC slab calculations. Both for reduction potentials of Fe in Fe-substituted ZSM-5 and adsorption energies of reaction intermediates in the (photo-)catalytic water oxidation at $TiO_2$(110) the obtained results underscore the importance of capturing the long-range electrostatics of the extended systems and the need to scrutinize more readily available semi-local DFT energetics.

## II. THEORY

### A. Solid-state embedding with ChemShell

The embedded-cluster models used in the QM/MM ChemShell calculations generally consist of multiple concentric regions, each of which is described using more approximate methods as one moves further from the center of the cluster, as illustrated in Figure 1. The central region of the embedded cluster is described quantum mechanically, with the applied level of theory being DFT, Hartree-Fock (HF), or higher post-HF as offered by the employed QM calculator. Atoms in the outer MM regions are treated at the level of interatomic potentials and are represented within the QM calculations as external point charges. The MM region is typically divided into an inner "active" part, where atoms are allowed to relax their positions, and an outer part where atoms are constrained to their lattice positions. Shell model potentials can be employed in the MM region for an accurate treatment of polarizable materials. A final exterior shell of point charges is added to the embedded cluster, with the point charges fitted to reproduce the full electrostatic embedding potential of an infinite bulk reference calculation within the regions of physical and/or chemical interest (the "active" region).[7]

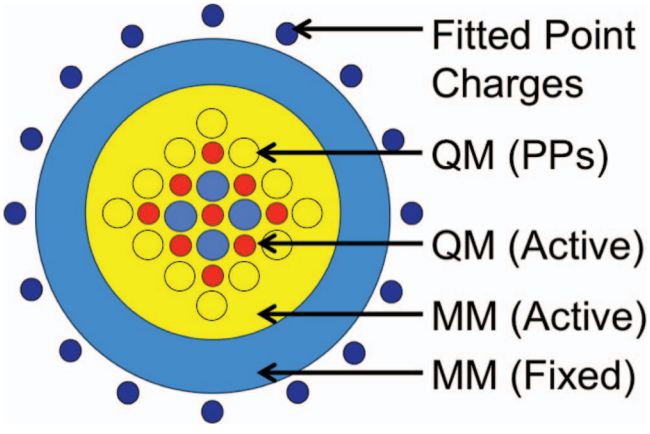


FIG. 1. Schematic representation of the embedded-cluster models used in the QM/MM calculations: The central QM atoms, represented by individual red and blue circles, is surrounded by regions of active (yellow) and frozen (light blue) MM atoms. If necessary, such as for ionic systems, an additional embedding potential (yellow circles) is included around the QM region in the form of atom-centered pseudopotentials (PPs). The outer shell of point charges (dark blue) is fitted to ensure reproduction of the correct long-range electrostatic potential within the active region.

This zero-reference potential for a given extended "perfect" system can be, e.g., the bulk of a three-dimensional periodic solid, i.e., a crystal, or one of its surfaces or interfaces with another material or solvent. In this way, a common absolute potential reference is provided for all embedded-cluster calculations. The latter is of special importance when modeling atomic and electronic processes involving a change in the charge (oxidation) state or comparing systems in different charge states. For example, calculation of defect levels in semiconductors and insulators that represent energies of defect ionization with electrons or holes becoming delocalized over conduction states depend crucially on this property. Another example is given by complementary oxidation-reduction cycles in acid-base and redox chemistry in solutions that involve ions or electrons to be removed or added to the system from remote locations. In all of these cases, a straightforward summation of the electrostatic contributions to the potential over a neutral extended system is typically performed using Ewald-like techniques.[7] The values of this potential on a grid of points spanning the active region serve then as reference for the embedded cluster model, i.e., the values of the point charges around the embedded cluster are fitted to reproduce optimally the potential of the extended reference calculation.

The energy of the QM region, $E^{\mathrm{QM}}$, thus combines the actual QM Hamiltonian ($\hat{H}^{\mathrm{QM}}$) with the surrounding environment,

$$E^{\mathrm{QM}} = \langle\Psi|\hat{H}^{\mathrm{QM}} + \hat{V}^{\mathrm{MM}}|\Psi\rangle + E^{\mathrm{QM}}_{\mathrm{nuc}} + E^{\mathrm{QM/MM}}_{\mathrm{nuc}}, \qquad (1)$$

where $\hat{V}^{\mathrm{MM}}$ is the external embedding potential acting on $|\Psi\rangle$ from the surrounding point charges, $E^{\mathrm{QM}}_{\mathrm{nuc}}$ is the Coulombic interaction among QM nuclei, and $E^{\mathrm{QM/MM}}_{\mathrm{nuc}}$ is the Coulombic interaction between the QM nuclei and the MM charges. For covalently bound, or semi-covalent/semi-ionic systems such as siliceous frameworks, saturation of dangling bonds on the exterior of the QM region with "link"-atoms is a commonly used methodology.[16,17] Such a setup is sufficient for strongly covalent materials with directional bonds, for example, of a $\sigma$ character. However, in the case of heteropolar semiconductors and ionic insulators, direct linkage of the field of MM point charges with the QM region as in Eq. (1) results in a spurious overpolarization of the wavefunction, aka a charge leakage towards the attractive cations of the immediately surrounding MM region. In this instance, the introduction of a boundary region is necessary between QM and MM atoms. In the present work, this is realized by replacing the cationic point charges in an intermediate shell surrounding the QM

region with norm-conserving PPs of Kleinman-Bylander type mimicking the corresponding ions. The energy of the QM region then reads

$$E^{\mathrm{QM}} = \langle \Psi | \hat{H}^{\mathrm{QM}} + \hat{V}^{\mathrm{PP}} + \hat{V}^{\mathrm{MM}} | \Psi \rangle + E^{\mathrm{QM}}_{\mathrm{nuc}} + E^{\mathrm{QM/MM}}_{\mathrm{nuc}}, \tag{2}$$

where $\hat{V}^{\mathrm{PP}}$ is the potential acting on $|\Psi\rangle$ in form of the PPs.

The MM energy term ($E^{\mathrm{MM}}$), discussed in detail previously,[7] depends on the problem being addressed, with the `ChemShell` package accommodating various forms: the General Utility Lattice Program (`GULP`) offers comprehensive MM functionality, especially in the implementations for strongly polarizable materials such as ionic oxides;[18] an alternative such as `DL_POLY` provides excellent parallelisation of larger molecular dynamics problems, and contains forcefield implementations more suited for semi-covalent systems.[19] The total energy ($E^{\mathrm{tot}}$) of the embedded-cluster system is finally given by simple additive definition,

$$E^{\mathrm{tot}} = E^{\mathrm{QM}} + E^{\mathrm{MM}}. \tag{3}$$

`ChemShell` manages the properties of an embedded cluster during calculations, outsourcing the calculation of energy and force terms through the varied QM and MM interfaces. The program couples a number of quantum chemistry and classical forcefield software packages: In a given geometry, the QM driver computes the energy as given in Eq. (1) or (2), and the forces acting on all particles (atoms, PPs, and point charges). The forcefield software then evaluates further forces on point charges, and geometry optimization is performed using the `DL-FIND` routine.[20] Iteratively, this leads to self-consistent embedding and polarization. In this work, we use the `FHI-aims` package for the QM calculations and modified it to be run as a library package rather than runtime executable. For the MM calculations, `GULP` and `DL_POLY` are used for the ionic and semi-covalently bound showcase systems, respectively.

### B. `FHI-aims` as QM calculator

`FHI-aims` is a full-potential all-electron electronic structure theory package providing both DFT and "beyond-DFT" functionality.[5,9,13] Notably, this comprises efficient treatments beyond standard semi-local DFT, such as bare or screened single-determinant exchange; quantum-chemical perturbation theory for the Coulomb interaction, e.g., second-order Møller-Plesset (MP2) perturbation theory, bare (MP2) or screened (GW) self-energies for single-electron excitation energies; or the random-phase approximation (RPA) in the adiabatic connection fluctuation dissipation theorem. It is based on hierarchical sets of all-electron atom-centered NAO basis functions of the form

$$\langle \mathbf{r} | \phi_\alpha \rangle = \phi_\alpha(\mathbf{r}) = \frac{u_l(r)}{r} Y_{lm}(\Omega), \tag{4}$$

where $Y_{lm}(\Omega)$ are the spherical harmonics, and the radial functions $u_l(r)$ are numerically tabulated and therefore fully flexible. Each basis function is strictly localized inside a given radius, which enables a highly efficient computation of both finite and (within PBCs) infinite systems within the same numerical framework. The regular basis set levels (called tiers, i.e., tier1, tier2,...), as well as the recent valence-correlation consistent NAO-VCC-nZ basis sets for light elements[8] are constructed to enable systematic accuracy improvements from fast qualitative to meV per atom. Radial functions are integrated on a dense one-dimensional grid with logarithmic spacing. Global integration is efficiently achieved on a sparser three-dimensional concatenation of atom-centered grids[21,22] (see Ref. 10 for details of the present implementation). Converting between logarithmic and global integration grid is done by using cubic splines. When working with embedding potentials (particularly, bare monopoles) inside the range of other integrands (e.g., the Kohn-Sham Hamiltonian matrix elements), these singularities must be integrated accurately, e.g., by placing an additional atom-centered grid on the site of the embedding potential.

With analytic forces provided, the major effort in the interfacing of `FHI-aims` with the `ChemShell` environment is the implementation of PP-functionality to describe cations at the boundary between the QM and MM region. For numerical efficiency reasons described below, we opt for norm-conserving PPs of Kleinman-Bylander type.[14] In general, PPs are constructed to replace the atomic all-electron potential such that core states are eliminated and the then missing orthogonalization constraint of valence and core wavefunctions is achieved through appropriate scattering properties of the PP.[23,24] In a real-space representation convenient to use with localized basis sets, this energy-dependent scattering is expressed by a dependence on the angular momentum,

$$\hat{V}^{\mathrm{PP}} = \sum_{lm} |Y_{lm}\rangle V_l(r) \langle Y_{lm}|. \tag{5}$$

For every angular momentum $l$ up to a maximum $l_{\mathrm{max}}$, usually 2 or 3, there is thus a different spherically symmetric PP channel $V_l(r)$ with corresponding eigenfunctions,

$$\langle \mathbf{r} | \psi_{lm} \rangle = \psi_{lm}(\mathbf{r}) = \frac{v_l(r)}{r} Y_{lm}(\Omega). \tag{6}$$

The radial functions $v_l(r)$ as well as the potentials $V_l(r)$ are commonly tabulated on a logarithmic grid. Our implementation is specifically tailored to read the tabulated format of the *.cpi PP files provided with the `FHI98PP` package.[24,25] In principle, any kind of PP can be used though, as long as it is made available in the *.cpi format of `FHI98PP`. An almost complete database of PPs is, e.g., available on the `abinit` webpage.[26]

Mimicking an ion with a net charge, all PP channels $V_l(r)$ must embody the same Coulomb behavior and are thus independent of the angular momentum in the far field, i.e., all $V_l(r)$ are the same outside the core radius $r_{\mathrm{core}}$ of the PP by construction. One channel can thus be chosen to embody all ionic long-range behavior. This local potential $V^{\mathrm{loc}}(r)$ then acts on all electrons independent of their angular momentum, while the remaining channels $\delta V_l = V_l(r) - V^{\mathrm{loc}}(r)$ are now short-ranged with $\delta V_l(r) = 0$ for $r > r_{\mathrm{core}}$. Mathematically, the choice of the local component is hereby arbitrary. In practice, a proper choice is crucial for the performance and transferability of the PP.[27] For instance, the routine `pswatch` in

the FHI98PP package provides full analysis capabilities according to rigorous criteria by Gonze *et al.*[27]

Numerically, it is most convenient to further transform the PP into the fully separable form of Kleinman-Bylander, where the short-ranged part $\delta V^{\mathrm{KB}}$ is a fully nonlocal operator in **r**-space,[14]

$$\begin{aligned}\langle \mathbf{r}|\hat{V}^{\mathrm{PP}}|\mathbf{r}'\rangle &= \langle \mathbf{r}|\hat{V}^{\mathrm{loc}}|\mathbf{r}'\rangle + \langle \mathbf{r}|\hat{V}^{\mathrm{KB}}|\mathbf{r}'\rangle \\ &= V^{\mathrm{loc}}(r)\delta(\mathbf{r}-\mathbf{r}') + \sum_{l=0}^{l_{\max}}\sum_{m=-l}^{l}\langle \mathbf{r}|\chi_{lm}\rangle E_l^{\mathrm{KB}}\langle \chi_{lm}|\mathbf{r}'\rangle.\end{aligned} \tag{7}$$

Here, the projector functions $|\chi_{lm}\rangle$ are defined as

$$|\chi_{lm}\rangle = \frac{1}{\sqrt{\langle \psi_{lm}|(\delta V_l)^2|\psi_{lm}\rangle}}\delta V_l|\psi_{lm}\rangle \tag{8}$$

with energies

$$E_l^{\mathrm{KB}} = \frac{\langle \psi_{lm}|(\delta V_l)^2|\psi_{lm}\rangle}{\langle \psi_{lm}|\delta V_l|\psi_{lm}\rangle}. \tag{9}$$

The local potential part is straightforward to implement. It is mapped from the (finite) logarithmic grid around the PP center onto the global integration grid with the help of cubic splines,[5,22] and its long-range part beyond the limits of the atom-centered grid is extrapolated by its Coulombic behavior (formal charge over distance). In the evaluation of the Hamiltonian matrix, $V^{\mathrm{loc}}$ is then evaluated exactly like all other local potentials. The general matrix expression of the nonlocal potential involves the evaluation of a sum over the projections of every basis function $|\phi_\alpha\rangle$ on to every Kleinman-Bylander projector function $|\chi_{lm}\rangle$

$$\langle \phi_\alpha|\hat{V}^{\mathrm{KB}}|\phi_\beta\rangle = \sum_{l=0}^{l_{\max}}\sum_{m=-l}^{l}\langle \phi_\alpha|\chi_{lm}\rangle E_l^{\mathrm{KB}}\langle \chi_{lm}|\phi_\beta\rangle. \tag{10}$$

For a given geometry, this matrix is computed once and is added to the Hamiltonian matrix $H_{\alpha\beta}$ in every iteration in the self-consistent field (SCF) cycle.

The strength of the Kleinman-Bylander formalism lies generally in its scaling behavior. Rather than having to evaluate and store all $\langle \phi_\alpha|\delta V_l|\phi_\beta\rangle$ matrix elements, resulting in $\mathcal{O}(N^2)$ scaling with the number of basis functions, the fully separable KB form only requires evaluation and storage of $N \times M$ projections with $M$ the (small) number of KB projector functions. In case of atom-centered basis sets, the projections $\langle \phi_\alpha^{(j)}|\chi_{lm}^{(i)}\rangle$ correspond furthermore to a two-center integral of basis function $|\phi_\alpha^{(j)}\rangle$ of atom $j$ at $\mathbf{R}_j$ and $|\chi_{lm}^{(i)}\rangle$ of PP $i$ at $\mathbf{R}_i$. The explicit expression to be evaluated is then

$$\begin{aligned}\langle \phi_\alpha^{(j)}|\chi_{lm}^{(i)}\rangle = &\frac{1}{\int dr(\delta V_l(r)v_l(r))^2} \\ &\times\int d\mathbf{r}\phi_\alpha(\mathbf{r}-\mathbf{R}_j)\frac{\delta V_l(|\mathbf{r}-\mathbf{R}_i|)v_l(|\mathbf{r}-\mathbf{R}_i|)}{|\mathbf{r}-\mathbf{R}_i|}Y_{lm} \\ &\times\left(\frac{\mathbf{r}-\mathbf{R}_i}{|\mathbf{r}-\mathbf{R}_i|}\right),\end{aligned} \tag{11}$$

which can be most efficiently computed with the help of spherical Bessel transformations[28–30] already implemented in FHI-aims.[9] Again, the radial functions $v_l(r)$ and the potential $V_l(r)$, respectively, potential differences $\delta V_l(r)$ of the projector are pretabulated in *.cpi format and provided as input for each relevant element. The total number of overlap integrals that needs to be computed at all is further reduced by locality, as overlap integrals are exactly zero if the distance of the involved atom centers exceeds a maximum value. This maximum value is given by the maximum extension of any basis function plus the maximum extension of any KB projector function, and is typically in the range of 8 Å.

For geometry optimization, molecular dynamics or vibrational analysis, force contributions on the PP with formal charge $q_i$ and centered at position $\mathbf{R}_i$ also need to be evaluated. Hellman-Feynman contributions arise from the embedding of the PP into the electrostatic fields of the electron density and all other nuclei in the QM region at positions $\mathbf{R}_j$ and with charges $Z_j$,

$$\begin{aligned}\boldsymbol{F}_i^{\mathrm{loc}} &= -\sum_j^{n_{\mathrm{atoms}}}\frac{Z_j q_i}{|\boldsymbol{R}_j-\boldsymbol{R}_i|^3}(\boldsymbol{R}_j-\boldsymbol{R}_i) \\ &\quad -\int d^3r\rho(\boldsymbol{r})\nabla_i V^{\mathrm{loc}}(\boldsymbol{r}).\end{aligned} \tag{12}$$

Here, $\nabla_i V^{\mathrm{loc}}(\boldsymbol{r})$ is the gradient of $V^{\mathrm{loc}}$ with respect to the position $\boldsymbol{R}_i$ and is needed on every global integration grid point. This is efficiently computed from a cubic spline interpolation of $V^{\mathrm{loc}}$ created before entering the SCF cycle. Analogous to the first term in Eq. (12), the Coulomb interaction between nuclei and PPs gives rise to a force term of the same shape that complements the general QM-force expression for atoms in FHI-aims (Eqs. (69) and (71)– (75) in Ref. 5). As written and implemented, Eq. (12) will work for non-periodic geometries. For periodic geometries, a formalism analogous to Eq. (70) in Ref. 5 would have to be adopted.

The overlap, Eq. (11), between any basis function and a KB-projector function $|\chi_{lm}^{(i)}\rangle$ of the PP $i$ at $\boldsymbol{R}_i$ is a function of $\boldsymbol{R}_i$. This gives rise to a force acting on the PP,

$$\begin{aligned}\boldsymbol{F}_i^{\mathrm{KB}} = -\sum_k f_k\sum_{j,h}^{n_{\mathrm{atoms}}}\sum_{\alpha\in\mathcal{J},\beta\in\mathcal{H}}c_{k\alpha}c_{k\beta}\sum_{\chi\in\mathcal{I}}\sum_{lm} \\ \left(\left(\frac{\partial}{\partial\boldsymbol{R}_i}\langle\phi_\alpha^{(j)}|\chi_{lm}^{(i)}\rangle\right)E_l^{\mathrm{KB}}\langle\chi_{lm}^{(i)}|\phi_\beta^{(h)}\rangle \right. \\ \left. +\langle\phi_\alpha^{(j)}|\chi_{lm}^{(i)}\rangle E_l^{\mathrm{KB}}\left(\frac{\partial}{\partial\boldsymbol{R}_i}\langle\chi_{lm}^{(i)}|\phi_\beta^{(h)}\rangle\right)\right),\end{aligned} \tag{13}$$

where $f_k$ is the occupation number of Kohn-Sham state $k$, and $c_{k\alpha}c_{k\beta}$ is the density matrix. $\mathcal{I}$, $\mathcal{J}$, and $\mathcal{H}$ are the subspaces of projector, respectively, basis functions belonging to PP $i$ and atoms $j$ and $h$, respectively. Again, derivatives of overlaps only have to be calculated for those pairs of functions, which have a non-zero overlap. Since the overlap, Eq. (11), is also a function of the position of atom $j$, this gives reciprocally rise to a negative force acting on atom $j$ complementing the general force expression on atoms.

In order to compute the Hartree potential, FHI-aims follows the strategy of an atom-centered multipole (MP) decomposition as introduced by Delley.[22] Here, the density difference to a sum-over-free-atom density is partitioned into atom-centered components on the integration grid shells centered around individual atoms.[5] Applying a MP decomposition to these atom-centered components leads (together with the known MPs of the free-atom density components) to the total MP components of the Hartree potential.[5,22] Although the ionic PPs do not introduce any electron density of their own, they do act as integration grid centers and are thus included into the partitioning. As a result, some electron density components are assigned to the PPs, even when no basis functions are centered at the position of the PP. The MP expansion always introduces an error $\rho(\boldsymbol{r}) - \rho_{\mathrm{MP}}(\boldsymbol{r})$ though, as the expansion is truncated beyond a maximum angular momentum $l_{\mathrm{MP}}^{\mathrm{max}}$. This error leads to a net force,[5]

$$\boldsymbol{F}_i^{\mathrm{MP}} = -\int d^3r[\rho - \rho_{\mathrm{MP}}]\nabla_i V^{\mathrm{loc}}(\boldsymbol{r}) \tag{14}$$

as the missing multipole terms move with the position of the center.

## III. RESULTS

### A. Reduction potentials in Fe–ZSM-5

For a first demonstration of the established QM/MM embedded-cluster framework, we address the zeolite ZSM-5 as an example for a semi-covalent system, where neither embedding PPs nor a polarizable MM environment are required. In problems of current interest, it is essential to describe the time averaged response in a zeolitic system to a perturbation near the center of the cluster, which is accounted for already by the current model. Long-range polarization effects here are less important than in many other classes of dielectric materials due to a combination of their relatively low static permittivity ($\epsilon \sim 4$) with low density. The local electronic and ionic polarization fields are fully captured within the QM part of the total system, and the smaller static dielectric response from the MM vicinity is included as point charges can relax. ZSM-5 is commonly used in catalytic applications,[31,32] and has previously been the subject of investigations with the `ChemShell` package coupled with the QM calculator `Gamess-UK`.[15] Belonging to the MFI structural family, ZSM-5 is a porous aluminium silicate framework ($SiO_2$ building units with Al-substitutions), which can be further doped with catalytically active transition metals such as Fe and Ti.[33] Quantifying the reaction energetics at such transition metal centers is then an obvious milestone on the route to improved catalyst design. The accurate calculation of corresponding energies has proven problematic though, due to the difficulty in accounting for both the extended bulk framework and the electron localization around the active site. As semi-local DFT does not appropriately accomplish the latter (*vide infra*),[34,35] calculations at hybrid xc level are at least required. Existing work for Fe–ZSM5 at this level of theory has hitherto been restricted to finite fragments that only reproduce the active site of the catalyst.[36–40] As the surrounding framework undoubtedly does have an influence on the electrostatic potential at the active site,[6,41,44] efficient QM/MM calculations are an ideal tool that in combination with powerful QM calculators such as `FHI-aims` also allows to assess the performance of advanced functionals. Here, we illustrate this approach by calculating the reduction potential for intra-framework $Fe^{3+}$ sites in MFI using semi-local, hybrid, and double-hybrid functionals. Double-hybrid functionals such as XYG3[45] are still rather uncommon in the application to extended systems, while their implicit ability to improve electron localization could render them particularly appealing for systems such as Fe–ZSM-5.

We construct three different embedded clusters from the undoped parent MFI framework, each centered on a tetrahedral Si atom (T-sites) of differing structural nature within a framework active site:[6] a T-site at a straight channel (M7-T1), at a sinusoidal channel (Z6-T4), and at a channel intersection (I2-T12), cf. Fig. 2. Around this central atom the QM region includes all other members of the active site, as well as any extra Si atoms that are first neighbours to the active site, and all the corresponding linking oxygen atoms. Bigger clusters are not necessary for such simple chemisorption processes, but only for larger molecules interacting with the entire channel.[42,43] All incomplete bonds from the oxygen atoms are saturated with H link-atoms, and a bond-dipole correction is added to the MM representation.[44] Concentric regions of MM atoms are then added; first, an active region up to a radius of 10.58 Å (20 bohrs) and then a fixed region up to a radius of 21.17 Å (40 bohrs). This results in a total number of atoms in the embedded clusters in the range 2155–2180. The MM interactions are represented using the forcefield of Hill and Sauer,[46] and energy/force calculations are performed using the `DL_POLY` package. The accuracy of the `FHI-aims–ChemShell` coupling is validated by calculating the same embedded clusters also with `Gamess-UK` as QM calculator, and using the same TZVP Gaussian basis sets of Ahlrichs and Taylor[47] in both QM codes. At the PBE[48] xc

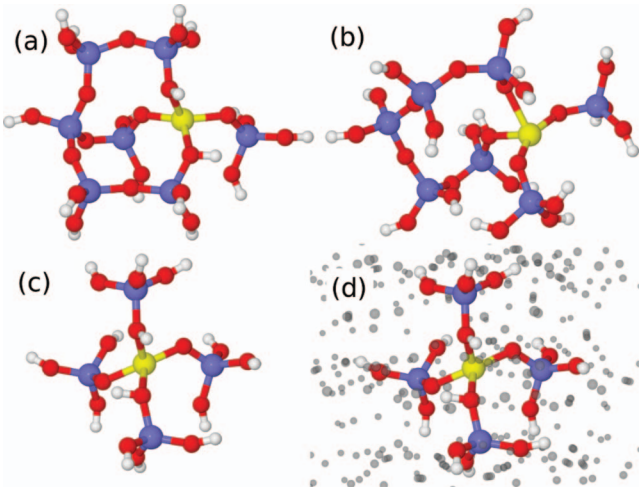


FIG. 2. Illustration of the QM clusters used in our calculations of Fe-ZSM-5: (a) M7-T1; (b) Z6-T4; and (c) I2-T12. Fe, Si, O, and H atoms are represented in yellow, mauve, red, and white, respectively. The extended embedding environment included in the QM calculation of (c) is illustrated in (d), with the MM point charges represented in grey.

level of theory and using converged integration grid settings in both codes differences in total energies are below 10 meV for each cluster, while forces on QM atoms differ by less than 5 meV/Å.

With this confidence, production calculations with `FHI-aims` are henceforth performed with the code's own numerical basis sets. For the three embedded clusters at PBE level, already tier1 leads hereby to total energies lower by 1 eV than for the TZVP basis set, demonstrating the better description of, e.g., the near-nuclear potential and kinetic energy integrals by the purposely designed NAO basis sets. The $Fe^{3+}$ active center is introduced to the embedded clusters by replacing the central Si atom. The reduced $Fe^{2+}$ state is described by coordinating an additional H atom to an O atom directly adjacent to the central Fe species. The reduction potential can then be defined as

$$E^{\mathrm{Red}}(\mathrm{Fe}^{3+/2+}) = E^{\mathrm{tot}}(\mathrm{Fe}^{2+}\text{-MFI}) + \frac{1}{2}E^{\mathrm{tot}}(\mathrm{H}_2) - E^{\mathrm{tot}}(\mathrm{Fe}^{3+}\text{-MFI}), \quad (15)$$

where $E^{\mathrm{tot}}(\mathrm{H}_2)$ is the energy of a gas-phase hydrogen dimer. Calculations are performed for the semi-local PBE and BLYP, for the hybrid PBE0 and B3LYP, as well as for the double-hybrid XYG3 functionals.[45,48–53] All geometries are fully relaxed at the corresponding semi-local level, with hybrid and double-hybrid calculations performed on these geometries, and specifically those of the according correlation treatment (i.e., B3LYP and XYG3 on the BLYP geometry, PBE0 on the PBE geometry). The XYG3 calculations are furthermore performed post-SCF on the optimized B3LYP Kohn-Sham orbitals. At tier2 basis set level the obtained reduction potentials are already numerically converged to within 1 meV for the semi-local and hybrid functionals. At the double hybrid-level a sufficient convergence to within ±20 meV can instead only be reached using the tier3 basis set for Fe and for all other (light) species the NAO-VCC-4Z basis set that specifically converges the unoccupied-space sums.[8] The more diffuse functions contained in this basis set then require an ensuing counterpoise correction[54] though.

The efficiency of this QM/MM approach can be seen by comparing the computational cost for one SCF iteration with that of the reference PBC setup. The I2-T12 cluster geometry comprises 21 atoms, while the MFI unit cells consist of 288 atoms. On the level of PBE, the average SCF iteration for the QM/MM setup takes 30 s using 4 CPUs (AMD Opteron Processor 250, 2393 MHz), compared to 1200 s for the periodic calculation (k-grid reduced to Γ-point) on the same machine. Periodic references on the hybrid and double hybrid level, moreover, demand a prohibitive amount of memory. Such levels of theory are thus only accessible through a QM/MM approach.

Table I compiles the obtained reduction potentials for Fe at the three structural sites in the MFI framework. At all levels of theory the differences between these three sites are very small, indicating a low structure sensitivity for the catalytic properties of the active center. This structure insensitivity warrants to directly compare to previous experimental voltammetry by Pérez-Ramírez *et al.*,[55] which gave very small reduction energies for intra-framework $Fe^{3+}$ of the order 0.05 eV $\gtrsim V_{\mathrm{Red}}(\mathrm{Fe}^{3+/2+}) \gtrsim$ −0.15 eV. There is, thus, a huge effect induced by the siliceous framework, when considering that the corresponding reduction potential of $Fe^{3+}$ in a hexagonally coordinated $[Fe(H_2O)_6]^{3+}$ complex was, for instance, calculated as 1.07 eV at PBE level.[56] This framework effect is well captured by essentially all xc levels studied, which in good overall agreement with the experimental findings yield small reduction potentials scattering around zero. More specifically, the majority of the xc-treatments agrees that the reduction potential is negative, indicating stability for Fe(II) in the material. Interestingly, the double-hybrid functional XYG3 yields reduction potentials closer to the ones at the semi-local (BLYP) level, and thus makes up for most of the change observed at the hybrid (B3LYP) functional level. At least within the small scatter of Table I there is thus no clear trend in the calculated reduction potentials that would prominently reflect the allegedly gradually increasing electron localization achieved in the sequence BLYP → B3LYP → XYG3.

TABLE I. Computed reduction potentials (in eV) for $Fe^{3+/2+}$ embedded intra-framework within a MFI siliceous structure. Active sites are modeled as the M7-T1, Z6-T4, and I2-T12 sites shown in Fig. 2. Calculations performed using semi-local and hybrid functionals used the tier2 basis level for all atoms; at the double hybrid-level the NAO-VCC-4Z basis set was used for all atoms except Fe, for which a tier3 basis set was used.

| | M7-T1 | Z6-T4 | I2-T12 |
|---|---|---|---|
| $E^{\mathrm{Red}}_{\mathrm{PBE}}(\mathrm{Fe}^{3+/2+})$ | −0.18 | −0.16 | −0.18 |
| $E^{\mathrm{Red}}_{\mathrm{BLYP}}(\mathrm{Fe}^{3+/2+})$ | −0.29 | −0.39 | −0.17 |
| $E^{\mathrm{Red}}_{\mathrm{PBE0}}(\mathrm{Fe}^{3+/2+})$ | −0.05 | 0.04 | −0.03 |
| $E^{\mathrm{Red}}_{\mathrm{B3LYP}}(\mathrm{Fe}^{3+/2+})$ | −0.09 | 0.01 | −0.02 |
| $E^{\mathrm{Red}}_{\mathrm{XYG3}}(\mathrm{Fe}^{3+/2+})$ | −0.34 | −0.20 | −0.18 |

### B. Water splitting reaction energetics at $TiO_2$ (110)

As a show case application to a polarizable material, where both effective core potentials at the QM/MM boundary and core-shell potentials in the MM region are essential, we next consider the $TiO_2$(110) surface. Among many other applications, this semiconductor catalyst is particularly known for its ability to oxidize water using light.[57–59] The desire for a molecular-level understanding of this intriguing property motivates detailed mechanistic studies unraveling the chemical steps involved. In surface science studies especially, the rutile $TiO_2$(110) surface has been frequently studied,[57,60] with recent theoretical work considering a reaction mechanisms including an OOH intermediate on a defect-free surface.[61,62] Characteristic for the field of surface catalysis, these calculations were performed at the semi-local DFT level, specifically with the rPBE[63] functional. Arguably because of its acceptable description of hydrogen bonds, this functional has been frequently employed in the context of water dissociation and photocatalysis at $TiO_2$,[62,64–68] but, of course, as a GGA functional it still suffers from the well-known electron delocalization problems. The possibility to perform hybrid and double-hybrid level calculations efficiently with our QM/MM setup

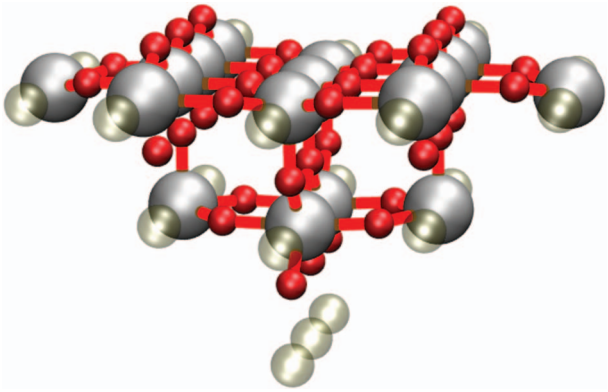

FIG. 3. Perspective view of the $Ti_{17}O_{34}$ cluster, exposing at its top center the fivefold coordinated Ti adsorption site of the $TiO_2$(110) surface. Ti atoms are shown as large white spheres, O atoms as small red spheres, and semi-transparent grey spheres mark the positions where PPs represent the immediately surrounding Ti-cations.

offers therefore an ideal platform to assess how much this affects the reaction energetics.

For the QM region and systematic size convergence tests, we employ a series of three clusters proposed by Anmal and Heyden,[69] $Ti_{17}O_{34}$, $Ti_{29}O_{58}$, and $Ti_{33}O_{66}$, with the $Ti_{17}O_{34}$ cluster including the employed PPs at the boundary region illustrated in Fig. 3. At each level of theory, the embedded clusters are constructed on the basis of optimized rPBE lattice positions. Formal charges are used for the embedding point charges (+4 for Ti and −2 for oxygen), whereby any positive point charge in a 8 Å vicinity of the QM region was replaced by $Ti^{4+}$ PPs. The MM region extends up to a distance of 27 Å from the central adsorption site, involving a total number of 4480 point charges. The interactions in the MM region are described with the well-established polarizable $TiO_2$ forcefield from Ref. 70, where only the spring potential between oxygen shell and oxygen core has been modified to $V = kd^2(\cosh(\frac{r}{d}) - 1)$ with $k = 20$ eV/Å$^2$ and $d = 0.1$ Å. This improves the representation of the dielectric constant of the MM region when setting it up with the DFT lattice parameters. In order to describe the intrinsic surface polarization properly, the position of the top-row O MM atoms is fully optimized for the clean $TiO_2$(110) surface. For the neutral systems studied here, further geometry optimization and self-consistent polarization of the MM region in response to the adsorbates is neglected. The position of the adsorbates on top of the central fivefold-coordinated Ti site in the QM region, cf. Fig. 3, is optimized at the rPBE level; higher-level calculations are then performed for these fixed geometries. As reference we also compute supercell geometries with a 5 O–$Ti_2O_2$–O trilayer slab, a 50 Å vacuum, and applying the identical surface geometry optimization protocol as for the embedded clusters. For (1 × 2) and (2 × 4) surface unit cells (4 × 4 × 1) and (2 × 2 × 1) Monkhorst-Pack k-point grids were used, respectively, while for the density of states (DOS) calculations of the clean (1 × 2) cell this grid was increased to (40 × 40 × 1).

Following earlier theoretical work,[61,62,71] we assume the water oxidation pathway at defect-free $TiO_2$(110) to proceed along four electron-coupled proton transfer steps,

$$\begin{aligned}
H_2O + (^*) &\rightarrow OH^* + H^+ + e^-,\\
OH^* &\rightarrow O^* + H^+ + e^-,\\
H_2O + O^* &\rightarrow OOH^* + H^+ + e^-,\\
OOH^* &\rightarrow O_2 + (^*) + H^+ + e^-,
\end{aligned}$$

where the asterisk stands for the fivefold coordinated Ti centers offered by the catalytic surface (*) and particles attached to them (e.g., O*), respectively. Central energetic quantities for this pathway are correspondingly the binding energies of O, OH, and OOH, defined as

$$E_b[X] = E^{tot}[X@TiO_2] - E^{tot}[TiO_2] - E^{tot}[X]. \quad (16)$$

Here, $E^{tot}[X@TiO_2]$ and $E^{tot}[TiO_2]$ are the total energies of the $TiO_2$(110) surface with and without adsorbate $X$, respectively, and $E^{tot}[X]$ is the total energy of the isolated adsorbate. At the hybrid and double-hybrid functional levels, identical spin states are obtained in corresponding QM/MM and slab calculations. At the semi-local level, which significantly suffers from self-interaction, this needed to be explicitly ensured by fixing the spin states to those of the higher-level calculations, i.e., doublet for OH and OOH, and triplet for O. This is an important issue, as most semi-local functionals yield the wrong spin-polarization for the O adsorbate, and already due to this produce a large deviation in the corresponding binding energy with respect to the higher-rung calculations.[71] Similar to the findings made for the zeolitic system, tier2 basis sets readily converge all $E_b$ to within 10 meV at the semi-local and hybrid level, while at the double-hybrid level the valence-correlation consistent NAO-VCC-4Z basis set for all O atoms together with a counterpoise correction[54] was required to reach this level of convergence.

Table II compares the rPBE binding energies computed with the embedded clusters against results obtained for a bare cluster and from supercell calculations with two coverages, 0.5 monolayer (ML) in a (2 × 1) surface unit-cell and 0.25 ML in a (4 × 2) surface unit-cell. In the QM/MM setup, convergence with respect to the size of the QM region is rapidly reached. Already the $Ti_{17}O_{34}$ cluster yields

TABLE II. Calculated rPBE binding energies (in eV) of O, OH, and OOH on the rutile $TiO_2$(110) surface. Compared are results from different clusters with and without embedding against results obtained from a periodic supercell setup and differing surface unit-cells.

| $E_b$ | O | OH | OOH |
|---|---|---|---|
| PBC | | | |
| (2 × 1) | −0.46 | −0.53 | −0.17 |
| (4 × 2) | −0.46 | −0.52 | −0.26 |
| Free cluster | | | |
| $Ti_{17}O_{34}$ | −0.52 | −0.51 | 0.01 |
| QM/MM without PP | | | |
| $Ti_{17}O_{34}$ | −0.58 | −0.76 | −1.28 |
| QM/MM | | | |
| $Ti_{17}O_{34}$ | −0.45 | −0.49 | −0.26 |
| $Ti_{29}O_{58}$ | −0.45 | −0.50 | −0.25 |
| $Ti_{33}O_{66}$ | −0.44 | −0.50 | −0.27 |

binding energies converged to within 10 meV, which we attribute to the correct polarization treatment of our solid-state embedding approach. In contrast, in the supercell calculations large (4 × 2) surface unit-cells are needed to reach the low-coverage limit modeled in the QM/MM approach for the OOH adsorbate. This, in turn, we attribute to the high dipole moment of this adsorbate, which sensitively feels nearby periodic images.

In case of the bare cluster or an embedded cluster without using PPs, the adsorption energies are significantly affected by spurious polarization of the QM-region boundary. If only point charges are used to saturate the cluster, electron density can spill over onto the Coulomb singularities of the positive point charges. Then, the band gap breaks down and, as demonstrated in Table II, adsorption energies are off by 1 eV. This highlights the importance of employing PPs in the transition region.

For all three adsorbates, representing a wide range of electron affinities and ionization potentials, low-coverage supercell and size-converged QM/MM calculations agree within 20 meV, demonstrating the accuracy of the established solid-state embedding approach and its huge potential for future work on explicitly charged systems (without need for intricate charge-compensation schemes).

Most importantly, this agreement between supercell and QM/MM calculations is not only restricted to the *binding energetics, but extends, for instance, also to the more sensitive underlying electronic structure.* This is illustrated in Fig. 4,

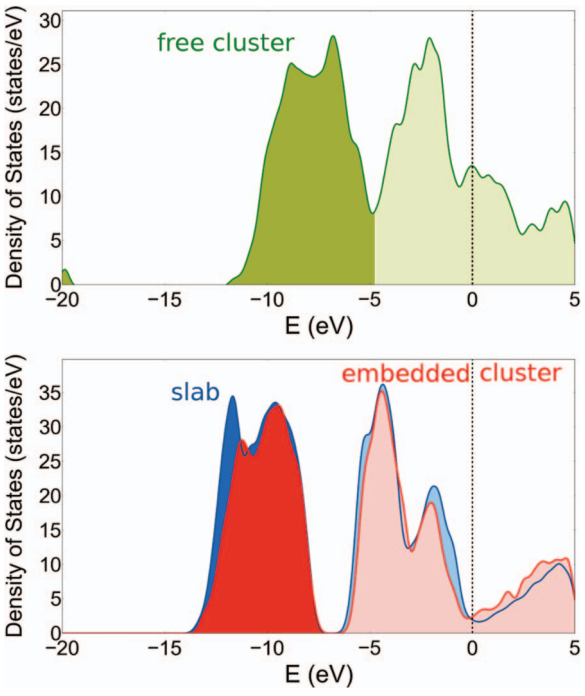


FIG. 4. Comparison of the total density of states (DOS) for the clean $TiO_2$(110) surface at the rPBE level: Even a relatively large free $Ti_{54}O_{108}$ cluster (upper panel) still shows huge deviations with respect to the corresponding embedded $Ti_{54}O_{108}$ cluster (lower panel), whereas the latter achieves almost perfect agreement with the corresponding PBC supercell calculations (lower panel), in particular in the valence and lower conduction band region. The vacuum level represents the zero reference throughout, and filled states are depicted in darker colour.

TABLE III. Calculated binding energies (in eV) of O, OH, and OOH on the rutile $TiO_2$(110) surface. Results are obtained from an embedded $Ti_{17}O_{34}$ cluster at different levels of theory (GGA, hybrid, and double-hybrid), see text.

| $E_b$ | O | OH | OOH |
|---|---|---|---|
| rPBE | −0.45 | −0.49 | −0.26 |
| BLYP | −0.60 | −0.62 | −0.30 |
| PBE | −0.71 | −0.79 | −0.43 |
| B3LYP | −0.36 | −0.56 | −0.35 |
| PBE0 | −0.39 | −0.64 | −0.41 |
| HSE06 | −0.41 | −0.66 | −0.43 |
| XYG3 | −0.36 | −0.64 | −0.53 |

which compares the total DOS for the clean $TiO_2$(110) surface. Whereas the DOS of an already quite large free $Ti_{54}O_{108}$ cluster still exhibits huge deviations with respect to the DOS obtained for the same-size embedded cluster, there is essentially perfect agreement between the latter and the DOS obtained from a PBC supercell calculation. This agreement extends not only to the band gap, i.e., relative difference of highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital (LUMO), but also to the absolute band edge positions, which are, of course, of utmost importance for photoelectrochemical applications like the light-driven water oxidation reaction studied here. The position of the delocalized LUMO is hereby more sensitive to the finite QM-region of the solid-state embedding approach. For the smaller $Ti_{17}O_{34}$ cluster, it is 0.3 eV higher in energy compared to its converged position in the large $Ti_{54}O_{108}$ cluster, whereas the position of the localized HOMO is essentially independent of the employed cluster size.

Table III compiles the results obtained at GGA, hybrid, and double-hybrid level. Generally, we observe a somewhat larger scatter among the three GGAs tested as compared to the group of hybrid and double-hybrid functionals. As widely perceived, in particular the PBE functional seems to be more on the overbinding side. Interestingly, hybrid and double-hybrid xc level reduces the bond strength of the O adsorbate and increases the bond strength of the OOH intermediate as compared to the semi-local description. Together with its overall weaker binding, rPBE mimics this best, which to some extent seems to support the arguments made in its favor in preceding work at the semi-local level. The inclusion of PT2 correlation in the XYG3 double-hybrid functional primarily changes the most polarization-affected OOH binding as compared to its parent B3LYP hybrid functional. This breaks the trend in bond strength OH > O > OOH that was consistently obtained for the semi-local and hybrid functionals.

Notwithstanding, these differences are primarily due to the description of the bare radical in the gas phase, which is used as reference in the definition of $E_b$ in Eq. (16) and which is, of course, most sensitively affected by a varying degree of electron localization achieved at the different xc-levels. To circumvent this, adsorption energies in photoelectrochemical calculations are often given with respect to the computational standard hydrogen electrode (SHE).[61] Table IV reproduces the same binding energies of Table III with

TABLE IV. Calculated binding energies (in eV) of O, OH, and OOH on the rutile $TiO_2$(110) surface, but referenced against the standard hydrogen electrode (SHE) following Eqs. (10) – (12) of Ref. 72. Results are obtained from the same computational setup and settings as in Table III.

| $E_b^{SHE}$ | O | OH | OOH |
|---|---|---|---|
| rPBE | 4.80 | 2.43 | 4.60 |
| BLYP | 4.75 | 2.34 | 4.59 |
| PBE | 4.90 | 2.34 | 4.61 |
| B3LYP | 4.86 | 2.37 | 4.78 |
| PBE0 | 4.92 | 2.36 | 4.86 |
| HSE06 | 4.87 | 2.32 | 4.84 |
| XYG3 | 4.88 | 2.39 | 4.89 |

respect to this reference and demonstrates that this indeed removes most of the scatter and leads to consistent trends at all xc functional levels studied. Under the above described constraint that the GGA functionals are enforced to yield the correct spin polarization, these results thus suggest that despite the known electron localization problems, a description at the semi-local level seems indeed sufficient for computational screening work relying on trends and correlations between reaction intermediate binding energies,[72,73] rather than quantitative differences within 0.2 eV.[72] Of course, this may be quite different when addressing adsorption at defects, in charged states, or when calculating reaction barriers – with the present QM/MM-setup then forming an ideal tool to conduct higher-level calculations either directly for production or as reference.

## IV. SUMMARY AND CONCLUSIONS

We have presented a general-purpose QM/MM-setup integrating the all-electron electronic structure theory code `FHI-aims` into the `ChemShell` package. In this course, pseudopotential functionality was implemented into `FHI-aims`, which in the solid-state embedding context is employed to prevent spurious charge leakage out of the QM zone into nearby cationic MM charges. The `FHI-aims` NAO basis sets enable particularly efficient access to exact exchange and second order perturbation theory. The here established QM/MM-approach is therefore ideally suited for hybrid and double-hybrid DFT calculations of extended systems.

The performance of our approach was demonstrated by the application to two different examples: The calculation of the Fe reduction potential in a Fe-substituted ZSM-5 zeolitic framework and the calculation of adsorption energies of reaction intermediates in the water oxidation at defect-free $TiO_2$(110). In both cases, our results confirm the crucial importance of the appropriate description of the long-range electrostatics achieved through the solid-state embedding. Systematic comparison to PBC supercell calculations for the more demanding $TiO_2$ system even shows that the possibility to adequately capture its polarizability through core-shell potentials in the MM region yields reliable absolute band edge positions, which predestines the approach for (photo-)electrochemical applications.

The calculation of the reduction potentials and adsorption energies at hybrid and double-hybrid xc level demonstrates the value of being able to scrutinize more readily available semi-local DFT data. In case of the reduction potentials, the higher-level reference data confirm the low structure sensitivity and the huge effect of the siliceous framework, with the small remaining scatter increasing the confidence in the quantitative numbers. In case of the adsorption energies, the higher-level reference data reveal that the electron delocalization problem of the semi-local description can lead to wrong spin polarizations, which in turn would lead to qualitatively wrong results. If the correct spin polarization is enforced, the benchmark against the hybrid and double-hybrid numbers indicates that the semi-local binding energetics of the neutral reaction intermediates is reliable enough for computational screening work, in particular if referencing is done against the standard hydrogen electrode rather than delicate bare radicals. Whether this prevails for more demanding cases like adsorption at (charged) defects remains to be seen and is the topic of on-going work – with the here established solid-state embedding approach providing a powerful tool that enables such calculations at higher xc-level efficiently and without the need for intricate charge compensation schemes.

## ACKNOWLEDGMENTS

We thank Jürgen Wieferink (now Aleo Solar) for help with his logarithmic spherical Bessel transform infrastructure in `FHI-aims`, Igor Zhang for his advice regarding numerical polarisation basis sets, and Thomas W. Keal for discussions regarding the implementation of the `FHI-aims/ChemShell` interface. A.J.L., C.R.A.C., P.S., and A.A.S. acknowledge funding from Engineering and Physical Sciences Research Council (U.K.) (EPSRC(GB)) (Grant No. EP/IO30662/1), M.R.F. and C.R.A.C. acknowledge funding from EPSRC (Grant No. EP/I03014X/1), and D.B. acknowledges a scholarship from the TUM International Graduate School of Science and Engineering. Further support came from the Solar Technologies Go Hybrid initiative of the State of Bavaria. The authors acknowledge the use of the following high-performance computing facilities, and associated support services: HECToR, via our membership of the UK HPC Materials Chemistry Consortium (EP/L000202); IRIDIS, provided by the Centre for Innovation; HYDRA, provided by the Max-Planck Supercomputing Center. Parts of this work were carried out under the HPC- EUROPA2 project (Project No. 228398), with the support of the European Community - Research Infrastructure Action of the FP7.

## APPENDIX: FULL PP-FUNCTIONALITY IN `FHI-AIMS`

In the QM/MM-context, the PP-functionality is exclusively used to prevent overpolarization of the QM charge due to immediately adjacent MM cationic monopoles. As such, the monopole term of the cation is replaced by a bare ionic PP with equal formal charge. In contrast, when using the PP in regular electronic structure calculations, valence electrons would be considered for the pseudoized atom and would then contribute to the total electronic energy. As the xc energy and potential are not additive with respect to the electron density, a nonlinear core correction (NLCC)[74,75] may be required to

make up for the missing core density of the pseudoized atom and therewith allow to properly capture the xc contribution of the added valence electrons. The essential idea of the NLCC is to simply add the core density $\rho_{\text{core}}$ of a free atom in the calculation of the pertinent parts of the total energy expression.

`FHI-aims` calculates the total energy, referred to as $E^{\text{QM}}$ in the main text, as[5]

$$\begin{aligned} E^{\text{QM}} = &\sum_{i}^{N_{\text{states}}} f_i \epsilon_i - \int dr^3 [\rho(\boldsymbol{r}) v_{\text{xc}}[\rho](\boldsymbol{r})] \\ &+ E_{\text{xc}}[\rho] \\ &- \frac{1}{2} \int dr^3 [\rho(\boldsymbol{r}) V_{\text{H}}[\rho](\boldsymbol{r})] + E_{\text{nuc-nuc}} \end{aligned} \tag{A1}$$

with $f_i$ the occupation number of Kohn-Sham state $\epsilon_i$ and $V_{\text{H}}$ the Hartree potential. The eigenvalues $\epsilon_i$ are hereby obtained in the standard manner from diagonalization of the Hamiltonian of the Kohn-Sham system, $\hat{h}_{\text{KS}} = \hat{t}_{\text{s}} + \hat{v}_{\text{ex}} + \hat{v}_{\text{es}} + \hat{v}_{\text{xc}}$. Here, $\hat{t}_{\text{s}}$ is the kinetic energy operator, $\hat{v}_{\text{ex}}$ the external potential, $\hat{v}_{\text{es}}$ the electrostatic potential, and $\hat{v}_{\text{xc}}$ the xc potential for the single electrons. The NLCC correspondingly affects the first three terms in Eq. (A1), which are thus replaced as follows:

$$\begin{aligned} E_{\text{xc}}^{\text{NLCC}}[\rho + \rho_{\text{core}}] = &\sum_{i}^{N_{\text{states}}} f_i \langle \psi_i | \hat{v}_{\text{xc}}[\rho + \rho_{\text{core}}] | \psi_i \rangle \\ &- \int d^3 r \ \rho \ V_{\text{xc}}[\rho + \rho_{\text{core}}] \\ &+ \int d^3 r \ (\rho + \rho_{\text{core}}) \epsilon_{\text{xc}}[\rho + \rho_{\text{core}}]. \end{aligned} \tag{A2}$$

Even though not of concern for the QM/MM focus, this expression was also implemented into `FHI-aims`. Thus, if an "empty" site (no nucleus, but including basis functions and integration grids) is placed atop a PP in `FHI-aims`, our present implementation also allows to perform norm-conserving PP calculations in `FHI-aims`. As smoothness of the core density is not an issue for NAO basis sets and integration grids designed for an all-electron code, we hereby employ the full atomic core density $\rho_{\text{core}}$ and not a smoothed auxiliary representation as is commonly done in plane-wave codes.[24]

As the added core density is atom centered, a Pulay-type force term on the pseudoized atom $i$ arises when adding the NLCC.[76–78] Differentiating Eq. (A2) determines this term as

$$\begin{aligned} \boldsymbol{F}_i^{\text{NLCC}} &= -\frac{\partial E_{\text{xc}}^{\text{NLCC}}[\rho]}{\partial \boldsymbol{R}_i} \\ &= -\int \frac{\delta((\rho + \rho_{\text{core}}) \epsilon_{\text{xc}}[\rho + \rho_{\text{core}}])}{\delta(\rho + \rho_{\text{core}})} \frac{\partial \rho_{\text{core}}}{\partial \boldsymbol{R}_i} \\ &= -\int dr^3 V_{\text{xc}}[\rho + \rho_{\text{core}}] \frac{\partial \rho_{\text{core}}}{\partial \boldsymbol{R}_i} \end{aligned} \tag{A3}$$

at the level of LDA.


[1] G. Makov and M. C. Payne, Phys. Rev. B **51**, 4014 (1995).
[2] S. Lany and A. Zunger, Model. Simul. Mater. Sci. Eng. **17**, 084002 (2009).
[3] C. Freysoldt, J. Neugebauer, and C. G. Van de Walle, Phys. Rev. Lett. **102**, 016402 (2009).
[4] N. Bernstein, J. R. Kermode, and G. Csányi, Rep. Prog. Phys. **72**, 026501 (2009).
[5] V. Blum, R. Gehrke, F. Hanke, P. Havu, V. Havu, X. Ren, K. Reuter, and M. Scheffler, Comput. Phys. Commun **180**, 2175 (2009).
[6] P. Sherwood, A. H. de Vries, M. F. Guest, G. Schreckenbach, C. R. A. Catlow, S. A. French, A. A. Sokol, S. T. Bromley, W. Thiel, A. J. Turner, S. Billeter, F. Terstegen, S. Thiel, J. Kendrick, S. C. Rogers, J. Casci, M. Watson, F. King, E. Karlsen, M. Sjövoll, A. Fahmi, A. Schäfer, and C. Lennartz, J. Mol. Struct.: THEOCHEM **632**, 1 (2003).
[7] A. A. Sokol, S. T. Bromley, S. A. French, C. R. A. Catlow, and P. Sherwood, Int. J. Quantum Chem. **99**, 695 (2004).
[8] I. Y. Zhang, X. Ren, P. Rinke, V. Blum, and M. Scheffler, New J. Phys. **15**, 123033 (2013).
[9] X. Ren, P. Rinke, V. Blum, J. Wieferink, A. Tkatchenko, A. Sanfilippo, K. Reuter, and M. Scheffler, New J. Phys. **14**, 053020 (2012).
[10] V. Havu, V. Blum, P. Havu, and M. Scheffler, J. Comput. Phys. **228**, 8367 (2009).
[11] T. Auckenthaler, V. Blum, H.-J. Bungartz, T. Huckle, R. Johanni, L. Krämer, B. Lang, H. Lederer, and P. R. Willems, Parallel Comput. **37**, 783 (2011).
[12] A. Marek, V. Blum, R. Johanni, B. Lang, V. Havu, T. Auckenthaler, A. Heinecke, H.-J. Bungartz, and H. Lederer, J. Phys.: Condens. Matter **26**, 213201 (2014); Psi-k Scientific Highlight for December, 2013, see http://www.psi-k.org/newsletters/News_120/Highlight_120.pdf, 2014.
[13] X. Ren, P. Rinke, C. Joas, and M. Scheffler, J. Mater. Sci. **47**, 7447 (2012).
[14] L. Kleinman and D. M. Bylander, Phys. Rev. Lett. **48**, 1425 (1982).
[15] M. F. Guest, I. J. Bush, H. J. J. Van Dam, P. Sherwood, J. M. H. Thomas, J. H. Van Lenthe, R. W. A. Havenith, and J. Kendrick, Mol. Phys. **103**, 719 (2005).
[16] U. C. Singh and P. A. Kollman, J. Comput. Chem. **7**, 718 (1986).
[17] M. J. Field, P. A. Bash, and M. Karplus, J. Comput. Chem. **11**, 700 (1990).
[18] J. D. Gale, J. Chem. Soc., Faraday Trans. **93**, 629 (1997).
[19] T. R. Forester, W. Smith, and I. T. Todorov, *The DL_POLY 2.0 User Manual* (Daresbury Laboratory, STFC Daresbury Laboratory, Cheshire, UK, 2008).
[20] J. Kästner, J. M. Carr, T. W. Keal, W. Thiel, A. Wander, and P. Sherwood, J. Phys. Chem. A **113**, 11856 (2009).
[21] A. D. Becke, J. Chem. Phys. **88**, 2547 (1988).
[22] B. Delley, J. Chem. Phys. **92**, 508 (1990).
[23] B. J. Austin, V. Heine, and L. J. Sham, Phys. Rev. **127**, 276 (1962).
[24] M. Fuchs and M. Scheffler, Comput. Phys. Commun. **119**, 67 (1999).
[25] Fhi98pp source, see http://th.fhi-berlin.mpg.de/th/fhi98md/download/index.html#fhipp; accessed 2013-09-30.
[26] Abinit web page, see http://www.abinit.org/downloads/psp-links/pseudopotentials; accessed 2013-09-29.
[27] X. Gonze, R. Stumpf, and M. Scheffler, Phys. Rev. B **44**, 8503 (1991).
[28] J. D. Talman, Int. J. Quantum Chem. **93**, 72 (2003).
[29] J. D. Talman, Int. J. Quantum Chem. **95**, 442 (2003).
[30] J. D. Talman, Comput. Phys. Commun. **180**, 332 (2009).
[31] H. M. Barrer, *Hydrothermal Chemistry of Zeolites* (Academic Press, 1982).
[32] J. M. H. Thomas, Sci. Am. **266**, 112 (1992).
[33] T. W. Keal, P. Sherwood, G. Dutta, A. A. Sokol, and C. R. A. Catlow, Proc. R. Soc. London, Ser. A **467**, 1900 (2011).
[34] J. To, A. A. Sokol, S. A. French, N. Kaltsoyannis, and C. R. A. Catlow, J. Chem. Phys. **122**, 144704 (2005).
[35] G. Pacchioni, F. Frigoli, D. Ricci, and J. A. Weil, Phys. Rev. B **63**, 054102 (2000).
[36] J. A. Ryder, A. K. Chakraborty, and A. T. Bell, J. Catal. **220**, 84 (2003).
[37] H. Guesmi, D. Berthomieu, B. Bromley, B. Coq, and L. Kiwi-Minsker, Phys. Chem. Chem. Phys. **12**, 2873 (2010).
[38] M. F. Fellah, I. Onal, and R. A. van Santen, J. Phys. Chem. C **114**, 12580 (2010).
[39] M. F. Fellah, E. A. Pidko, R. A. van Santen, and I. Onal, J. Phys. Chem. C **115**, 9668 (2011).
[40] M. F. Fellah and I. Onal, J. Phys. Chem. C **116**, 13616 (2012).
[41] A. H. de Vries, P. Sherwood, S. J. Collins, A. M. Rigby, M. Rigutto, and G. Jan Kramer, J. Phys. Chem. B **103**, 6133 (1999).
[42] U. Eichler, M. Brändle, and J. Sauer, J. Phys. Chem. B **101**, 10035 (1997).
[43] X. Solans-Monfort, V. Branchadell, M. Sodupe, M. Sierka, and J. Sauer, J. Chem. Phys. **121**, 6034 (2004).
[44] P. Sherwood, A. H. de Vries, S. J. Collins, S. P. Greatbanks, N. A. Burton, M. A. Vincent, and I. H. Hillier, Faraday Discuss. **106**, 79 (1997).

[45]Y. Zhang, X. Xu, and W. A. Goddard, Proc. Natl. Acad. Sci. U.S.A. **106**, 4963 (2009).
[46]J.-R. Hill and J. Sauer, J. Phys. Chem. **99**, 9536 (1995).
[47]R. Ahlrichs and P. R. Taylor, J. Chim. Phys. **78**, 315 (1981).
[48]J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).
[49]C. Adamo and V. Barone, J. Chem. Phys. **110**, 6158 (1999).
[50]A. D. Becke, Phys. Rev. A **38**, 3098 (1988).
[51]A. D. Becke, J. Chem. Phys. **98**, 5648 (1993).
[52]C. Lee, W. Yang, and R. G. Parr, Phys. Rev. B **37**, 785 (1988).
[53]P. J. Stephens, F. J. Devlin, C. F. Chabalowski, and M. J. Frisch, J. Phys. Chem. **98**, 11623 (1994).
[54]S. F. Boys and F. Bernardi, Mol. Phys. **19**, 553 (1970).
[55]J. Pérez-Ramírez, G. Mul, F. Kapteijn, J. A. Moulijn, A. R. Overweg, A. Doménech, A. Ribera, and I. W. C. E. Arends, J. Catal. **207**, 113 (2002).
[56]L. Guimarães, H. A. de Abreu, and H. A. Duarte, Chem. Phys. **333**, 10 (2007).
[57]U. Diebold, Surf. Sci. Rep. **48**, 53 (2003).
[58]A. Fujishima, X. Zhang, and D. A. Tryk, Surf. Sci. Rep. **63**, 515 (2008).
[59]M. A. Henderson, Surf. Sci. Rep. **66**, 185 (2011).
[60]T. L. Thompson and J. T. Yates, Jr., Top. Catal. **35**, 197 (2005).
[61]J. K. Nørskov, J. Rossmeisl, A. Logadottir, L. Lindqvist, J. R. Kitchin, T. Bligaard, and H. Jónsson, J. Phys. Chem. B **108**, 17886 (2004).
[62]Á. Valdés, Z.-W. Qu, G.-J. Kroes, J. Rossmeisl, and J. K. Nørskov, J. Phys. Chem. C **112**, 9872 (2008).
[63]B. Hammer, L. B. Hansen, and J. K. Nørskov, Phys. Rev. B **59**, 7413 (1999).
[64]P. J. D. Lindan and C. Zhang, Phys. Rev. B **72**, 075439 (2005).
[65]Z.-W. Qu and G.-J. Kroes, J. Phys. Chem. B **110**, 23306 (2006).
[66]R. Ferrando, personal communication (2009).
[67]M. Garciá-Mota, A. Vojvodic, H. Metiu, I. C. Man, H.-Y. Su, J. Rossmeisl, and J. K. Nørskov, ChemCatChem **3**, 1607 (2011).
[68]M. Garciá-Mota, A. Vojvodic, F. Abild-Pedersen, and J. K. Nørskov, J. Phys. Chem. C **117**, 460 (2013).
[69]S. C. Ammal and A. Heyden, J. Chem. Phys. **133**, 164703 (2010).
[70]A. Jentys and C. R. A. Catlow, Catal. Lett. **22**, 251 (1993).
[71]H. Oberhofer and K. Reuter, J. Chem. Phys. **139**, 044710 (2013).
[72]I. C. Man, H.-Y. Su, F. Calle-Vallejo, H. A. Hansen, J. I. Martínez, N. G. Inoglu, J. Kitchin, T. F. Jaramillo, J. K. Nørskov, and J. Rossmeisl, ChemCatChem **3**, 1159 (2011).
[73]D. van der Vliet, D. S. Strmcnik, C. Wang, V. R. Stamenkovic, N. M. Markovic, and M. T. M. Koper, J. Electroanal. Chem. **647**, 29 (2010).
[74]S. G. Louie, S. Froyen, and M. L. Cohen, Phys. Rev. B **26**, 1738 (1982).
[75]D. Porezag, M. R. Pederson, and A. Y. Liu, Phys. Rev. B **60**, 14132 (1999).
[76]P. Pulay, Mol. Phys. **17**, 197 (1969).
[77]J. Gerratt and I. M. Mills, J. Chem. Phys. **49**, 1719 (1968).
[78]J. Gerratt and I. M. Mills, J. Chem. Phys. **49**, 1730 (1968).