Embedding for bulk systems using localized atomic orbitals

F. Libisch,[1] M. Marsman,[2] J. Burgdörfer,[1] and G. Kresse[2]

[1] *Institute for Theoretical Physics Vienna University of Technology, A-1040 Vienna, Austria, EU*

[2] *University of Vienna, Faculty of Physics and Center for Computational Materials Science, Sensengasse 8/12, A-1090 Vienna, Austria*

(Dated: 30 June 2017)

We extend projection-based embedding techniques to bulk systems to treat point defects in semiconductors and insulators. To avoid non-additive kinetic energy contributions we construct the density partition using orthogonal subsets of orbitals. We have implemented our approach in the popular `VASP` software package. We demonstrate its power for defect structures in silicon and polaron formation in titania, two challenging cases for conventional Kohn-Sham density functional theory.

I. INTRODUCTION

Ab-initio electronic structure theory for bulk materials has experienced tremendous advances in many areas such as density functional theory,[1–4] improved post-DFT[5–11] and, e.g., van der Waals functionals,[12] as well as highly accurate quantum chemical[13–15] and Monte-Carlo approaches[16]. However, many problems are still out of reach of an advanced theoretical description due to their size: the accurate description of, for example, defect structures requires both a highly accurate treatment of the local defect region, as well as the treatment of a large number of atoms of the environment[17]. It is often challenging for a single method to meet both requirements. Embedding is therefore a suitable strategy to overcome this hurdle. Its underlying idea is to treat the local structure or, more generally, the subsystem of interest by a high-level method while treating the environment with the help of a numerically less demanding lower level method. Consistently combining different electronic structure methods within the same calculation is both the advantage and the challenge of the embedding approach[18].

Several embedding schemes have been proposed[19–32], often relying on some form of a local embedding potential $V_{emb}(\mathbf{r})$ that mediates the interaction between the subsystem referred in the following as the cluster and the surrounding environment. Typically, subsystems are treated in the presence of $V_{emb}(\mathbf{r})$ [or $\hat{V}_{emb}(\mathbf{r}, \mathbf{r}')$] using a high-level method while the environment is handled by density functional theory (DFT). The individual subsystem densities are then added to obtain an approximation for the total density of the entire system. While conceptually simpler, local embedding potentials feature the distinct disadvantage that no set of mutually orthogonal orbitals of the entire system exists. Consequently, evaluation of the total energy becomes challenging: in particular the kinetic energy needs to be approximated. Several groups[19,26,30,32] used an optimized effective potential method to recover the kinetic energy given a total electron density.

Another embedding strategy is based on the density matrix rather than the density. Fornace et al. presented an embedded mean-field theory[33] partitioning the one-particle density matrix of the system based on its natural orbitals. A single Hamiltonian then describes the entire system, avoiding any issues with evaluating the kinetic energy for cluster and environment separately. Additionally, this approach, by construction, allows for charge exchange between the cluster of interest and the environment. However, a direct extension to plane-

wave basis sets used in periodic solid state computations seems challenging. Density matrix embedding theory[34] introduces bath states that allow for representing the embedded clusters as an open quantum system: charges are distributed between bath sites and embedded sites allowing for arbitrary (i.e., also fractional) partitioning of charges between the embedded subsystems.

Manby et al.[35] use a projection-based scheme to ensure mutual orthogonality of orbitals belonging to different subsystems. Such an approach was explored in various non-periodic settings[22–25]. In the present article, we adapt such a strategy to extended systems with the specific application to a frequently used plane-wave code. We determine Wannier-like orbitals localized within the cluster by performing unitary rotations within the subspace of fully occupied Kohn-Sham orbitals while the orthogonal complement of remaining orbitals resides within the environment[34]. During the optimization cycle for the cluster involving an advanced functional, the environment orbitals remain frozen and thus orthogonality is preserved. Such an approach avoids the inaccuracies associated with approximating the kinetic energy.

In the present paper we demonstrate the power of our embedding scheme in a proof-of-principle calculation addressing two problems for which standard Kohn-Sham DFT is known to be inadequate: defects in silicon and polarons in titania. We use the following hierarchy of methods: the cluster is treated by the (expensive) hybrid functional PBEh while the environment is treated only by the PBE functional. We show that this embedding scheme implemented in the Vienna Ab Initio Simulation Package (`VASP`) is robust and efficient. We emphasize that the present embedding scheme is not limited to hybrid-DFT in DFT embeddings. Future extensions will address the treatment of the cluster by RPA or correlated wavefunction approaches such as coupled cluster[36] or many-body perturbation theory based on a complete active space calculation[37].

II. TECHNIQUE

We partition a system into two parts: a cluster of interest A with atomic sites $r_{j,A}$, $(j = 1, \ldots, M_A)$ with M_A the number of atomic sites included in the cluster, and the surrounding environment B, containing M_B atomic sites $r_{j,B}$, $(j = 1, \ldots, M_B)$. In a first step, the entire system $(A + B)$ is solved using a single, comparatively cheap exchange-

correlation functional, e.g., PBE[38],

$$H \left| \psi_i \right\rangle = \bar{\varepsilon}_i \left| \psi_i \right\rangle, \tag{1}$$

yielding Kohn-Sham orbitals $\left| \psi_i \right\rangle$ with orbital energies $\bar{\varepsilon}_i$ and the density matrix

$$\gamma(\vec{r}, \vec{r}') = \sum_{i=1}^{N_{\text{tot}}} f_i \left\langle \psi_i | \vec{r} \right\rangle \left\langle \vec{r}' | \psi_i \right\rangle, \tag{2}$$

with occupation numbers $f_i \in [0, 1]$, where the index $i = 1, \ldots, N_{\text{tot}}$ goes over all orbitals and physical spin. Note that we we have not included k-point sampling in the present ansatz, since it is not straightforward to treat the transformations at different k-points independently. We aim to find a unitary rotation within the subspace of fully occupied orbitals ($f_i = 1$, $i = 1, \ldots, N$) that yields a set of orbitals aligned with the atomic orbitals $\left| \alpha_k \right\rangle$ localized around the atomic sites of the cluster. The index $k = 1, \ldots, N_A$ of the atomic orbitals includes both the site index as well as radial and angular momentum quantum numbers. To this end, we first calculate the orbital overlap matrix W,

$$W_{ki} := \left\langle \alpha_k | \psi_i \right\rangle, \quad k = 1 \ldots N_A, \quad i = 1 \ldots N, \tag{3}$$

and then perform a singular value decomposition of W according to

$$W = U \cdot D \cdot V^\dagger \tag{4}$$

with $D = \text{diag}(\sigma_i)$. The unitary matrix V represents the rotation in the space of the N occupied orbitals that optimally aligns N_A of these orbitals with the atomic orbitals $\left| \alpha_i \right\rangle$ keeping the remaining $N - N_A$ orbitals orthogonal to each other and to the optimally alligned orbitals. The singular values σ_i provide a measure for the degree of overlap between the rotated atomic orbitals $\left| \beta_i \right\rangle$ and the rotated orbitals $\left| \phi_i \right\rangle$,

$$\left| \phi_i \right\rangle = \sum_{ij} V_{ij} \left| \psi_j \right\rangle \quad \left| \beta_i \right\rangle = \sum_{ij} U_{ij} \left| \alpha_j \right\rangle$$

$$\left| \left\langle \beta_i | \phi_i \right\rangle \right| = \begin{cases} \left| \sigma_i \right| \leq 1, \ i \leq N_A \\ 0, \quad i > N_A \end{cases}. \tag{5}$$

Orbitals with indices $i > N$ outside the space of occupied orbitals are not included in the singular value decomposition and thus unaffected by the rotation, $\left| \phi_i \right\rangle = \left| \psi_i \right\rangle$, $\forall i : N < i < N_{\text{tot}}$. Using the rotated $\left| \phi_i \right\rangle$, we can thus partition the occupied space into orbitals that

have an overlap with the $|\alpha_i\rangle$, and those that do not. Ideally, if the Kohn Sham orbitals are well covered by the atomic wavefunctions, we expect the singular values σ_i to be close to 1.

After the orbital rotation, a subset $\left|\phi_i^A\right\rangle$ of those $i = 1, \ldots, N_{\text{emb}}^A$ with $N_{\text{emb}}^A \leq N_A$ is optimized using a more expensive exchange-correlation (XC) functional E_{XC}^A, e.g., a hybrid functional[39,40], while the orthogonal complement of the $N_{\text{emb}}^B = N - N_{\text{emb}}^A$ environment orbitals $\left|\phi_i^B\right\rangle$ remains frozen. The latter are still fully included in the action of the Hamiltonian on the orbitals A, i.e., in the electrostatic and XC interactions and in enforcing the Pauli principle. In principle, this implies that one determines the variational point of the projected Hamiltonian

$$\mathbf{H}^A = \left(1 - \sum_{i \in B}\left|\phi_i^B\right\rangle\left\langle\phi_i^B\right|\right)\mathbf{H}\left(1 - \sum_{i \in B}\left|\phi_i^B\right\rangle\left\langle\phi_i^B\right|\right), \tag{6}$$

where $\mathbf{H}$ is the Hamiltonian of the energy functional to be introduced below (Eq. 9). In practice, the implementation in most plane-wave codes is straightforward: since the number of plane wave basis functions precludes exact diagonalization of the full KS Hamiltonian, iterative Krylov-subspace methods or conjugate gradient techniques are employed to refine the orbitals.[41] In such schemes, Kohn-Sham orbitals are usually explicitly orthogonalized in each step or kept orthogonal using appropriate constraints. Consequently, one must only constrain the optimization to the subset A and make sure that the optimized orbitals A are kept orthogonal to the frozen B orbitals. Such an approach required only minimal modifications to the VASP code, which are discussed below.

In general, the number of orbitals N_{emb}^A used in the embedding procedure may be smaller than the number of atomic basis functions N_A: in principle, one may choose any subset of the localized orbitals $0 < N_{\text{emb}}^A \leq N_A$. In practice, we sort the rotated orbitals by their singular values σ_i and choose the N_{emb}^A orbitals corresponding to the largest σ_i, where N_{emb}^A is chosen according to the number of orbitals of interest within the cluster. A typical threshold will be $\sigma_i > 0.5$. We find that our results do not strongly depend on N_{emb}^A, as long as the number of optimized orbitals is sufficiently large as to properly describe the local bonding. When considering the change in energy as a function of a smooth deformation of the geometry, for example when studying the energy along a reaction coordinate, N_{emb}^A should be kept constant for all points along the trajectory to avoid artefacts due to changes in the number of embedded orbitals. We will discuss the choice of N_{emb}^A in more detail in the results section below.

We note that after the orbital rotation the $|\phi_i\rangle$ are no longer eigenvectors of the Kohn-Sham Hamiltonian H. The diagonal matrix elements of the Hamiltonian are given by the expectation values

$$\varepsilon_i := \langle \phi_i | H | \phi_i \rangle. \tag{7}$$

These values are obviously different from the original eigenvalues, yet they observe the condition

$$\sum_{i=1}^{N} \varepsilon_i = \sum_{i=1}^{N} \overline{\varepsilon}_i, \tag{8}$$

since traces are invariant under unitary rotations.

If fractionally occupied orbitals are present in the original calculations of the entire system, we proceed along the following line. First, the fractionally filled orbitals are not included in the singular value decomposition. Second, and contrary to unoccupied orbitals, fractionally filled orbitals are included in the subset A. The invariance condition of Eq. (8) is only used in the subspace of the $N \leq N_{\text{tot}}$ fully occupied orbitals. This procedure might require more careful considerations if quantum-chemistry based correlated wavefunction approaches are used.

Taking into consideration that different exchange-correlation functionals will be employed for the cluster A and the environment B we write the energy functional for the entire system $A + B$ as

$$E = \sum_{i=1}^{N_{\text{tot}}} f_i \langle \phi_i | T | \phi_i \rangle + \frac{1}{2} E_H[\rho] + E_{\text{XC}}^{AB}[\gamma_A, \gamma_B], \tag{9}$$

where $\rho(\vec{r}) = \gamma(\vec{r}, \vec{r})$ is the density, and $E_H[\rho]$ the Hartree energy

$$E_H[\rho] = \frac{1}{2} \sum_{i,j=1}^{N_{\text{tot}}} f_i f_j \langle i\ j | \mathcal{C} | i\ j \rangle, \tag{10}$$

with $\mathcal{C}$ denoting the Coulomb operator. The mixed exchange-correlation functional $E_{\text{XC}}^{AB}[\gamma_A, \gamma_B]$ containing both the lower (B) and higher (A) level functionals can be written as

$$E_{\text{XC}}^{AB}[\gamma_A, \gamma_B] = E_{XC}^{A}[\gamma_A] + E_{\text{XC}}^{B}[\gamma_B] + E_{\text{XC}}^{\text{int}}[\gamma_A, \gamma_B], \tag{11}$$

where E_{XC}^{A} is the high level exchange correlation functional, E_{XC}^{B} is the low level exchange correlation functional, and $E_{\text{XC}}^{\text{int}}$ is an approximate exchange correlation functional used to treat the interaction between subsystem A and B.

6

To use Eq. (11) in practice within our embedding approach, the contribution due to the interaction between the two subsystems, $E_{\mathrm{XC}}^{\mathrm{int}}$, can be, for instance, approximated by the lower-level functional (B) applied also to the environment, i.e., $E_{\mathrm{XC}}^{\mathrm{int}}[\gamma_A, \gamma_B] \to E_{\mathrm{XC}}^{B}[\gamma_A, \gamma_B]$. More formally, this approximation yields

$$E_{\mathrm{XC}}^{\mathrm{int}}[\gamma_A, \gamma_B] = E_{\mathrm{XC}}^{B}[\gamma_A + \gamma_B] - E_{\mathrm{XC}}^{B}[\gamma_A] - E_{\mathrm{XC}}^{B}[\gamma_B]. \tag{12}$$

Such a simple approximation allows for very expensive functionals to be used in the cluster A, including the RPA or high-level correlated wavefunction approaches. The drawback is that the error introduced in such a mixed approach is difficult to quantify a priori.

In the present work, we have chosen a more precise description for the interaction between the two subsystems, namely we adopt the high level functional $E_{\mathrm{XC}}^{\mathrm{int}}[\gamma_A, \gamma_B] \to E_{\mathrm{XC}}^{A}[\gamma_A, \gamma_B]$. This is possible, since we presently consider hybrid functionals as the high level functional A, that can be evaluated for the entire system with quite reasonable computational effort. In this case, the interaction term $E_{\mathrm{XC}}^{\mathrm{int}}$ can be formally written as

$$E_{\mathrm{XC}}^{\mathrm{int}}[\gamma_A, \gamma_B] = E_{\mathrm{XC}}^{A}[\gamma_A + \gamma_B] - E_{\mathrm{XC}}^{A}[\gamma_A] - E_{\mathrm{XC}}^{A}[\rho_B]. \tag{13}$$

This is equivalent to the formally simple approximation

$$E_{\mathrm{XC}}^{AB}[\gamma_A, \gamma_B] = E_{\mathrm{XC}}^{A}[\gamma_A + \gamma_B].$$

For hybrid functionals, this approximation is still very efficient when optimizing the orbitals in the subset A and keeping the orbitals in B frozen. The key point is that the contribution of terms involving $(i, j \in B)$ is constant for the Hartree (compared Eq. 10) and exact exchange energy $E_x[\gamma]$

$$E_x[\gamma] = \frac{1}{2} \sum_{i,j=1}^{N_{\mathrm{tot}}} f_i f_j \langle i\, j | \mathcal{C} | j\, i \rangle. \tag{14}$$

Hence these terms do not need to be calculated for the optimization of orbitals in A. In plane wave codes, we can make use of this fact easily, by simply restricting the action of the Hamiltonian on the subset of optimized orbitals $|\phi_i^A\rangle$, $i = 1, \ldots, N_{\mathrm{emb}}^A$. For the exchange term this means that we only need to calculate

$$V_x \left| \phi_i^A \right\rangle,$$

where V_x is the exact non-local exchange potential. This speeds up the calculations by a factor $N_A/(N_A + N_B)$ greatly reducing the numerical effort compared to full hybrid functional

calculations. The orbitals in A can now be efficiently optimized by minimizing the energy functional (Eq. (9)), while the orbitals in B are kept frozen. The main drawback of such a procedure is that during the optimization of the orbitals in A, any change in the electronic structure of A due to a more accurate XC functional cannot lead to a redistribution of charge in B.

To improve upon this point, the embedding can be made self-consistent by alternating between subsystems A and B in freeze-and-thaw cycles: after the initial solution of the entire system using the lower-level functional, an orbital rotation is performed to partition into orbital sets A and B. Then, starting with A, alternatingly one of the subsystems is optimized while the other one is kept frozen. Each cycle that optimizes the orbitals A uses the higher-level XC functional A while each cycle that optimizes the orbitals in B employs the lower-level functional B. For the example of polarons in titania discussed below, we find rapid convergence after about six freeze-and-thaw cycles[18].

Since the interaction of the cluster region (e.g. defect) with its periodic image needs to be minimized, conventional defect modeling is hampered by the requirement of large supercells. If, for example, the bulk material could well be described by generalized gradient approximation (GGA) XC functionals and only the defect structure requires more advanced techniques such as hybrid functional approaches, conventional techniques still require an expensive evaluation of the entire exchange contribution. The embedding procedure outlined above is ideally suited to significantly reduce computational effort while retaining high accuracy. Note, however, that the approximation to the total energy (Eq. (13)) is tractable only in case of the embedding of hybrid functionals into DFT. A future application of our formalism to the embedding of correlated wavefunction methods into hybrid functionals will require the use of the approximation of Eq. (12) to the total energy, the quality of which remains to be explored.

III. IMPLEMENTATION IN VASP

We have implemented the embedding scheme outlined above in the Vienna ab initio simulation package (`VASP`) using the projector augmented wave method of Blöchl in the implementation of Kresse and Joubert[42,43,47]. Usage is simple: in a first step, a GGA DFT calculation of a system is performed. In a second step, the localized atomic basis functions

8

ing) are defined. To determine the localized orbitals, our implementation currently supports projection onto the PAW basis functions (the pseudo partial waves) and standard spherical harmonics (including hybrid orbitals such as sp^3) with the radial dependence taken from suitably scaled hydrogenic functions. VASP then starts an embedding calculation, performs the orbital rotation and optimizes the set of N_{emb} orbitals localized on the cluster A, while the remaining fully occupied Kohn-Sham orbitals of the environment B are frozen (for spin-polarized calculations the two spin components are treated independently in terms of the rotation and the number of optimized orbitals). The orbitals of A are now further optimized using a more expensive hybrid functional. During the freeze-and-thaw cycles, no further localization procedure according to Eq. (5) is required, and orbital sets A and B are interchanged. To enforce orthogonality in practice, we only reorder orbitals so that the frozen B orbitals have the lowest orbital indices, and then simply freeze them. Typical plane wave DFT implementations are based on Krylov-space or conjugate gradient methods (the plane wave basis is too large for exact diagonalization to be efficient). Convergence is accelerated by using procedures like direct inversion in the iterative subspace,[47] that always require mutual orthogonalization of the current Kohn-Sham orbitals in each calculation step to guarantee numerical stability irrespective of any embedding considerations. VASP uses a modified Gram Schmidt procedure,[41] where the lowest orbitals are normalized and orbitals with greater band indices are then made orthogonal to the orbitals below. Due to reordering of the orbitals, all A orbitals (as well as all virtual orbitals) possess greater orbital indices than the B orbitals and will thus be kept orthogonal to the frozen B orbitals during optimization.

We currently avoid k-point sampling in the embedding calculation since the orbital rotations at different k-points are not independent of each other. Likewise, forces are not yet implemented in our formalism. The geometries used in this work were taken from Ref. 44 for the defects in silicon and where relaxed using the HSE functional similar to Ref. 45 for the polarons in titania.

To compare energies calculated from the functional Eq. (13) the full exchange energy with all orbitals has to be evaluated only once, after self-consistently converging the orbitals of A. Note that for the case of hybrid functionals, this is equivalent to evaluating Eq. (9) for the final set of orbitals. We benchmark our embedding approximation against a fully self consistent optimization of all orbitals using the hybrid functional. Additionally, we also

energies [eV] (error [meV])

Defect	Hybrid	GGA		GGA*		Embedding	
H	3.00	2.42	(-586)	2.99	(-17)	3.01	(10)
X	3.01	2.44	(-571)	3.01	(1)	3.04	(27)
C_{3V}	3.05	2.44	(-609)	3.03	(-19)	3.06	(18)
T	3.77	2.86	(-904)	3.34	(-423)	3.75	(-20)
VJT	4.14	3.13	(-1010)	4.52	(377)	4.19	(44)
V	4.23	3.12	(-1113)	4.83	(599)	4.26	(25)

TABLE I. Comparison of defect formation energies for different defect types in silicon. All energies in eV, errors in brackets [meV] are deviation from full hybrid benchmark calculation. The right GGA column denoted with an asterisk represent a "single-shot" evaluation of the full hybrid energy functional using the corresponding non-hybrid GGA PBE orbitals.

compare to calculations using the pure GGA functional, and to the hybrid energy functional evaluated with orbitals obtained from using a GGA functional (PBE). Obviously, both the embedded and the PBE orbitals are, by construction, not self-consistent with respect to the hybrid functional. In comparison with a full self-consistent optimization, a single evaluation step using the full hybrid functional with non-self consistent orbitals takes a small amount of time while substantially improving the accuracy: errors in the approximate evaluation of the interaction between subsystems in Eq. (12) are eliminated. For pure GGA, we denote corresponding energies by an asterisk (*) in the following.

IV. RESULTS

A. Point defects in silicon

As a first practical test of our new algorithm, we consider point defect structures in silicon[44]. We use a 64 atom supercell using Γ point only sampling. Due to the localized nature of the covalent bonds involved, GGA Kohn-Sham DFT fails to correctly reproduce experimental observations. By using hybrid functionals or even more advanced RPA formulations[44], these problems are mitigated. However, comparison with more accurate correlation functionals such as RPA and experiment show that currently available methods

10

yield a wide range of predictions depending on the employed functional[46], highlighting the necessity to move towards higher level correlated wavefunction approaches.

Due to the large supercells required to avoid interaction of the defect sites with their periodic images, embedding the orbitals close to the defect site seems desirable. As benchmark for our embedding scheme, we consider defect formation energies of a set of common interstitial defects and vacancies. We aim to reproduce the energetics of full hybrid functional calculations based on PBEh by a cheaper embedding calculation in which only a few (six to ten) orbitals localized in the immediate vicinity of the defect (taken to be the A orbitals of the cluster) are treated using the hybrid functional, while the remaining 118-122 orbitals (taken to be the B orbitals) are only treated by PBE. We also compare our results to purely DFT-based predictions.

For Si defect calculations, we choose as atomic orbitals $|\alpha_i\rangle$ the PAW pseudo-partial waves of the Si atoms at and directly adjacent to the defect site, resulting in $N_A = 16$ (vacancies) or 20 atomic basis functions (one s and three p per atom) for most cases. Increasing the number of basis functions per atom increases the overall overlap of the occupied Kohn-Sham orbitals with the defect site at the cost of a larger N_A, and thus a larger overlap matrix $W \in \mathbb{C}^{N_A \cdot N}$ of Eq. (3). Consequently, the number of singular values σ_i increases. To obtain a set of orbitals well localized at the defect site, we choose all orbitals with singular values $\sigma_i > 0.5$ as embedded orbitals. This procedure yields a number of selected embedded orbitals N_{emb}^A from 9 (X defect) to 16 (T-defect), in line with the number of Si-Si bonds one would expect for the respective defect sites. For example, each of the two defects atoms of the dumpbell defect (X) interacts strongly with four close neighbors in the surrounding lattice and with the other atom in the dumpbell, yielding a total of nine covalent bonds. Indeed, we find nine singular values substantially larger than 0.5 for this defect. The chosen PAW basis functions yield a set of orbitals with a bimodal distribution: a significant number of orbitals with $\sigma_i \approx 1$, well seperated from delocalized orbitals with small overlap $\sigma_i \approx 0$ with the defect site. The threshold of 0.5 is therefore most of the time a good compromise between choosing all N_A possible orbitals (which will include orbitals with very small singular values) and too few orbitals that will not allow for reasonable optimization. Note, however, that care must be taken to check that there are no singular values close to the chosen threshold, to avoid arbitrarily including (or discarding) orbitals upon small fluctuations in σ_i. Due to weakly broken spatial symmetries around the defect site, σ_i values often cluster, with several near-

degenerate values of σ_i corresponding to localized orbitals related by symmetry operations of the bulk crystal. Using only a part of the subset of such near-degenerate orbitals would result in an artificial breaking of symmetry and should be avoided.

As mentioned in Sec. II, calculating the final energies for our hybrid functional embedding approach and Eq. (13) will be equivalent to evaluating Eq. (9) for the final set of orbitals. We therefore also provide values for a single evaluation of the hybrid energy functional using all converged DFT GGA orbitals (i.e., not just the ones localized at the defect), which we denote by an asterisk (*).

Our results for various defect structures are summarized in Tab. I. Overall, we find excellent agreement between the hybrid functional benchmark and our embedding approach. For simple, non-metallic defects such as the dumbbell configuration (X), the hexagonal hollow (H) and a lower-symmetry variant (C_{3V}) we find that both the embedding as well as the evaluation of the hybrid functional with the low-level DFT orbitals produces good agreement with benchmark calculations (see second column of Tab. I). By contrast, the metallic tetragonal site is badly described by DFT: it features one interstitial Si atom coordinated to its four nearest neighbors, so that the local coordination of the interstitial is identical to the other Si atoms. This position is unique insofar that the highest occupied orbital is threefold degenerate (t_2 symmetry) but only occupied by two electrons. This degeneracy is preserved in DFT yielding three fractionally occupied orbitals with occupation numbers $f_i = 2/3$. Consequently, the evaluation of the hybrid energy functional based on these orbitals fails to yield reasonable formation energies. By contrast, the embedding method locally breaks the degeneracy as does the full hybrid calculation, leading to good agreement of the embedding results with the benchmark (see T, VJT and V lines in Tab. I). We note that on our hardware, the full hybrid calculations takes ten times as long as the embedded one, with a relative error of 0.4% in total energy.

Our results compare poorly with experimental data: one important reason is the interaction of periodic images of the defects to the supercell size. We therefore consider a larger supercell of 512 atoms, still with a single defect. We find a substantial change in results for the larger cell (compare Tab. II), that now fit well to experimental results for the H defect. To achieve better agreement also for vacancies (V) requires a more accurate treatment of electronic correlation (e.g., RPA) or inclusion of Van der Waals contributions[46].

12

energies [eV] (error [meV])

Defect	Hybrid	GGA		GGA*		Embedding		Experiment
T	4.97	3.65	(1314)	5.17	(198)	5.13	(165)	
H	4.22	3.51	(712)	4.19	(38)	4.25	(28)	4.2-4.7
V	5.06	1.86	(3202)	5.55	(482)	5.08	(22)	2.1-4.0

TABLE II. Same comparison as in Tab. I for an eight times larger supercell containing 512 atoms. Rightmost column shows the range of available experimental data taken from [48–57].

B. Polarons in titania

As a second demonstration of our method, we consider the formation of polarons in titania. We consider a $2 \times 2 \times 2$ supercell with 24 Ti and 28 O atoms in the rutile structure. In an accurate hybrid functional description, an additional electron localizes, distorting the lattice and forming a small polaron. The distortion decreases the energy compared to a delocalized charge. A full hybrid functional calculation yields a decrease in energy by 514 meV for the distorted geometry. We will use this value in the following as benchmark for our embedded description of the polaron. Concerning the atomic basis functions $|\alpha_i\rangle$ used for the initial localization, we typically use the one s and five d orbitals of the Ti atom centered at the small polaron deformation, as well as all s and p orbitals of the six nearest neighbor oxygen atoms, yielding a total of $N_A = 30$ orbitals. Geometries for the distorted structure were relaxed using HSE calculations.

Our results for the small polaron formation energy are summarized in Tab. III. Density functional theory invoking only a local density approximation or a GGA functional is not capable of reproducing small polaron formation predicting even a negative energy gain (i.e. energy costs) of -355 meV to form the polaron. Inserting the DFT orbitals in the hybrid energy functional leads to a correction of the sign. However the energy gain is underestimated by a factor of two (270 meV), see Tab. III. A single-cycle embedding calculation yields a slightly larger error predicting 190 meV. The origin of this error is obvious: while the hybrid functional tries to localize the charge in the cluster region A, the surrounding region B cannot react to the substantial change in the electrostatics, since all B orbitals are frozen.

Subsequent freeze-and-thaw cycles rapidly improve the result: we alternate between optimizing the two sets of orbitals A and B, one with the expensive hybrid, the other with

13

(a) Iteration 1   (c) Iteration 3   (e) Iteration 5

(b) Iteration 2   (d) Iteration 4   (f) Iteration 6

Cluster (A): Hybrid

Environment (B): DFT

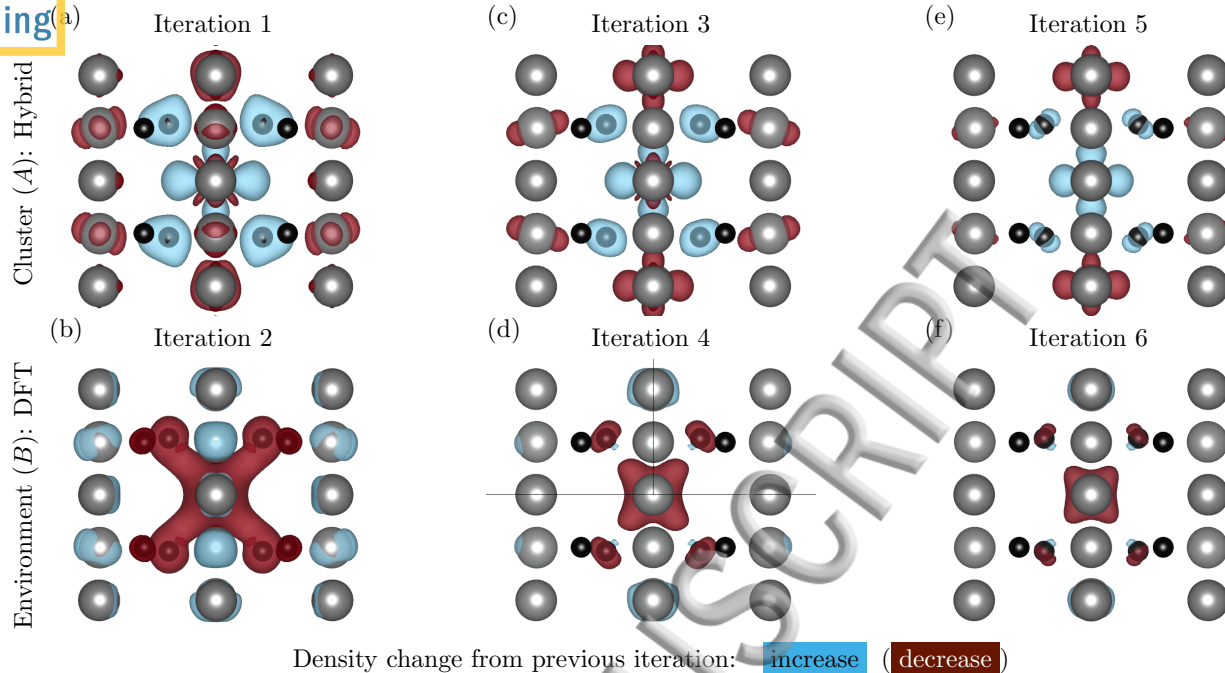Density change from previous iteration: increase (decrease)

FIG. 1. Isosurface plot of charge variation $\Delta\rho$ in the density compared to the previous iteration (for iteration 1 the difference to DFT is shown). Red (blue) denotes density decrease (increase). Odd iteration numbers (top row) correspond to optimization in the cluster A using the hybrid functional, while even iteration numbers (bottom row) correspond to PBE optimizations of the environment B.

pure GGA (PBE). We find convergence in about seven iterations, quite independent of the number of embedded orbitals N_{emb}^{A} (Tab. III). As minimum requirement for N_{emb}^{A}, the Ti atom at the center of the distortion and the surrounding oxygen atoms need to be treated accurately, which is already achieved with as few as six orbitals (Tab. III). Note that choosing only the central Ti atom as atomic basis, $N_A = N_{\mathrm{emb}}^{A} = 6$, yields a smaller polaron energy than chosing the six orbitals with the highest singular values from the $N_A = 30$ localized orbitals including also the closest oxygen atoms. The reason is that in the latter case the response of the surrounding shell of oxygen atoms is - to some degree - also treated by the hybrid functional. However, further increasing the number of localized orbitals N_A by, e.g., also including a shell of neighboring Ti atoms does not result in a stronger overlap of A orbitals on the central Ti atom as the localization procedure does not distinguish between the different atomic basis functions $|\alpha_i\rangle$. Consequently, such a large N_A would require a comparatively large N_{emb}^{A} to ensure that orbitals close to the central site are included in the

Method	cycles	E_{dist} [eV]	E_{ideal} [eV]	ΔE [meV]
Hybrid	-	-972.85	-972.33	514
DFT	-	-688.17	-687.81	-355
DFT*	-	-962.81	-962.54	270
Embedding*	1	-963.70	-963.52	174
$N_A = 30$ $N_{\text{emb}}^A = 14$	2	-963.36	-963.12	240
	3	-963.47	-963.16	313
	4	-963.41	-962.98	426
	5	-963.45	-962.99	459
	6	-963.44	-962.98	462
	7	-963.45	-962.98	475
$N_A = 30$ $N_{\text{emb}}^A = 6$	7	-963.40	-962.94	460
$N_{\text{emb}}^A = 8$	7	-963.41	-962.95	459
$N_{\text{emb}}^A = 12$	7	-963.45	-963.00	454
$N_{\text{emb}}^A = 20$	7	-963.50	-963.02	474
$N_{\text{emb}}^A = 30$	7	-963.59	-963.08	514
$N_A = 90$ $N_{\text{emb}}^A = 30$	7	-963.55	-963.08	462
$N_A = 6$ $N_{\text{emb}}^A = 6$	7	-963.38	-962.96	420

TABLE III. Energies of the distorted E_{dist} and ideal E_{ideal} lattice structure of charged rutile titania. The energy ΔE is gained by forming a small polaron. Energies after seven iterations are given for different sizes of the embedded region. Different methods are labeled as follows. Hybrid: full hybrid functional calculation used as benchmark; DFT: direct evaluation of the energies using the PBE functional; DFT*: evaluation of the hybrid energy functional using the orbitals from the PBE calculation; embedding: embedding calculations as function of the number of freeze-and-thaw cycles at fixed number of localized atomic orbitals N_A. After the unitary rotation to localize orbitals around the defect [Eq. (5)], we optimize a subset $N_{\text{emb}}^A \leq N_A$ of orbitals in the embedded region A as noted.

(a) Hybrid

(b) DFT

(c) Embedding

Unpaired spin density

positive   negative

Difference to hybrid benchmark

positive   negative

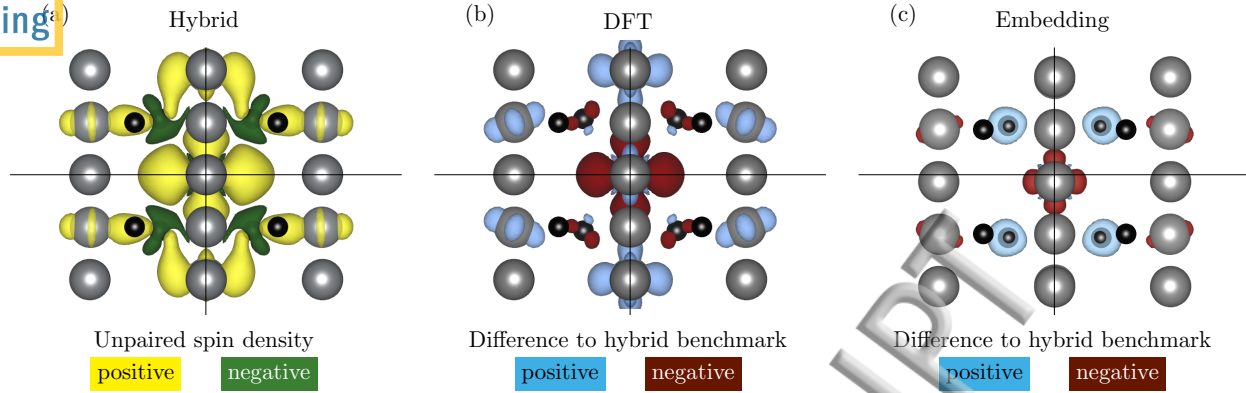Difference to hybrid benchmark

positive   negative

FIG. 2. (a) Isosurface plot of the spin density $(\rho_+ - \rho_-)$ of the converged polaron wavefunction in titania, using a full hybrid functional calculation, seen from the (100) direction. Grey (black) spheres correspond to Ti (O) atoms. The plot is centered around the Ti atom at the center of the distortion. Tourquois (yellow) denotes positive (negative) signs. (b,c) Isosurface plot of the difference in unpaired spin density between (b) DFT [(c) embedding] and the hybrid benchmark (a). Blue (red) corresponds to a density increase (decrease) compared to (a).

embedded calculation. Otherwise accuracy is lost. Indeed, we find a better agreement with the benchmark for $N_A = N_{\text{emb}}^A = 30$ than for $N_A = 90$, $N_{\text{emb}}^A = 30$ (see Tab. III). Since the numerical effort of the embedding calculation scales linearly with N_{emb}^A, in practice a small N_A that allows for $N_A \ll N$ is preferable.

It is instructive to follow the charge density variations along the freeze-and-thaw cycles [Fig. 1]. Additional charge is localized in the A cycles using the hybrid functional in the cluster (top row in Fig. 1). The density spreads out again and the environment relaxes in the B cycles when the orbitals B of the environment are optimized using the DFT functional. However, the magnitude of these changes quickly decreases with the iteration number and yields a well-converged density (and well-converged energy) within 7 iterations.

The full hybrid and the converged embedded unpaired spin densities closely match (Fig. 2) (b,c). By contrast, the DFT density does not show a strong localization of the surplus electron at all (Fig. 2) (a). Indeed, projecting the converged polaron orbital (i.e., the occupied majority spin Kohn-Sham orbital with the highest energy) onto the central Ti atom of the distortion yields quite small values for the overlap (0.39) for DFT, while the full hybrid (0.69) and embedded calculations (0.65) agree quite well. This underlines that despite the correct sign for the energy gain when using the DFT orbitals in the hybrid energy functional,

16

the DFT description of the charge density is qualitatively deficient.

V. CONCLUSIONS

We have demonstrated an embedding technique for extended periodic structures based on a suitable rotation in the subspace of fully occupied Kohn-Sham orbitals. Using a projection on local basis functions, a set of orbitals may be localized at a site of interest, for example a defect. Subsequently, these localized orbitals inside the cluster can now be optimized based on a more expensive exchange-correlation functional, such as a hybrid functional involving the exact evaluation of Fock exchange. Since exchange interactions within the frozen environment are neglected, the computation time is drastically reduced. The response of the environment to the charge rearrangement in the cluster can be self-consistently included by freeze-thaw cycles in which alternatingly the orbitals in the embedded cluster or in the environment are optimized.

We have implemented our ansatz in the popular `VASP` software package. As proof of principle, we have applied our method to two problems of current interest: a set of defects in bulk silicon, and small polarons in bulk titania. We find excellent agreement with (much more expensive) benchmark bulk hybrid calculations.

REFERENCES

[1] W. Kohn and L. J. Sham, Phys. Rev. **140**, 1133 (1965).

[2] K. Burke, J. Chem. Phys. **136**, 150901 (2012).

[3] A. J. Cohen, P. Mori-Sánchez, and W. Yang, Chem. Rev. **112**, 289-320 (2012).

[4] A. Pribram-Jones, D. A. Gross, and K. Burke, Ann. Rev. Phys. Chem. **66**, 283 (2015).

[5] P. Nozières, and D. Pines, Phys. Rev. **111**, 442 (1958).

[6] D. Langreth, and J. Perdew, Phys. Rev. B **15**, 2884 (1977).

[7] T. Miyake, F. Aryasetiawan, T. Kotani, M. van Schilfgaarde, M. Usuda, and K. Terakura, Phys. Rev. B **66**, 245103 (2002).

[8]M. Fuchs, Y.-M. Niquet, X. Gonze, and K. Burke, J. Chem. Phys. **122**, 094116 (2005)

[9]F. Furche, J. Chem. Phys. **129**, 114105 (2008).

[10]J. Paier, B. G. Janesko, T. M. Henderson, G. E. Scuseria, A. Grüneis, and G. Kresse, J. Chem. Phys. **132**, 094103 (2010).

[11]J. Paier, X. Ren, P. Rinke, G. E. Scuseria, A. Grüneis, G. Kresse, and M. Scheffler, New J. Phys. **14**, 043002 (2012).

[12]A. Tkatchenko, R. A. DiStasio Jr., R. Car, and M. Scheffler, Phys. Rev. Lett. **108**, 236402 (2012).

[13]M. Marsman, A. Grüneis, J. Paier, and G. Kresse, J. Chem. Phys. **130**, 184103 (2009).

[14]A. Grüneis, M. Marsman, and G. Kresse, J. Chem. Phys. **133**, 074107 (2010).

[15]A. Grüneis, G. H. Booth, M. Marsman, J. Spencer, A. Alavi, and G. Kresse, J. Chem. Theor. Comput. **7**, 2780 (2011).

[16]G. Booth, A. Grüneis, G. Kresse, and A. Alavi, Nature **493**, 365-370 (2013).

[17]F. Libisch, C. Huang, P. Liao, M. Pavone, and E. A. Carter, Phys. Rev. Lett. **109**, 198303 (2012).

[18]T. Dresselhaus, J. Neugebauer, S. Knecht, S. Keller, Y. Ma, and M. Reiher, J. Chem. Phys. **142**, 044111 (2015).

[19]J. Nafziger, Q. Wu, and A. Wasserman, J. Chem. Phys. **135**, 234101 (2011).

[20]T. Vreven and K. Morokuma, Ann. Rep. Comput. Chem. **2**, 35 (2006).

[21]M. Svensson, S. Humbel, R. D.Froese, T. Matsubara, S. Sieber, and K. Morokuma, Keiji, J. Phys. Chem. **100**, 19357 (1996).

[22]J. D. Goodpaster, T. A. Barnes, F. R. Manby, and T. F. Miller III, J. Chem. Phys. **140**, 18A507 (2014).

[23]D. V. Chulhai, and L. J. Jensen, J. Chem. Theor. Comput. **11**, 3080-3088 (2015).

[24]B. Hegely, P. R. Nagy, G. G. Ferenczy, and M. J. Kallay, Chem. Phys. **145**, 064107 (2016).

[25]D. V. Chulhai, and J. D. Goodpaster, J. Chem. Theor. Comput. **13**, 1503 (2017)

[26]J. D. Goodpaster, N. Ananth, F. R. Manby, and T. F. Miller, J. Chem. Phys. **1133**, 084103 (2010).

[27]J. D. Goodpaster, T. A. Barnes, and T. F. Miller, J. Chem. Phys. **134**, 164108 (2011).

[28]J. D. Goodpaster, T. A. Barnes, F. R. Manby, and T. F. Miller, J. Chem. Phys. **137**, 224113 (2012).

[29]T. A. Wesolowski and A. Warshel, J. Phys. Chem. **97**, 8050 (1993).

[30]S. Fux, C. R. Jacob, J. Neugebauer, L. Visscher, and M. Reiher, J. Chem. Phys. **132**, 164101 (2010).

[31]F. Libisch, C. Huang, and E. A. Carter, Acc. Chem. Res. **47**, 2768 (2014).

[32]C. Huang, M. Pavone, and E. A. Carter, J. Chem. Phys. **134**, 154110 (2011).

[33]M. E. Fornace, J. Lee, K. Miyamoto, F. R. Manby, and T. F. Miller, III, J. Chem. Theory Comput. **11**, 568-580 (2016).

[34]G. Knizia, and G. K.-L. Chan, Phys. Rev. Lett. **109**, 186404 (2012).

[35]F. R. Manby, M. Stella, J. D. Goodpaster, and T. F. Miller, J. Chem. Theory Comput. **8**, 2564 (2012).

[36]J. Čížek, J. Chem. Phys. **45**, 4256 (1966).

[37]K. Andersson, P.-A. Malmqvist, and B. O. Roos, J. of Chem. Phys. **96**, 1218 (1992)

[38]J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).

[39]A. D. Becke, J. Chem. Phys. **98**, 1372 (1993).

[40]J. P. Perdew, M. Ernzerhof, and K. Burke, J. Chem. Phys. **105**, 9982 (1996).

[41]G. Kresse and J. Furthmüller, Phys. Rev. B **54**, 11169 (1996).

[42]P. E. Blöchl, Phys. Rev. B **50**, 17953 (1994).

[43]G. Kresse and D. Joubert, Phys. Rev. B **59**, 1758 (1999).

[44]M. Kaltak, J. Klim'es, and G. Kresse, Phys. Rev. B **90**, 054115 (2014).

[45]A. Janotti, C. Franchini, J. B. Varley, G. Kresse, C. G. Van de Walle, Phys. Stat. Sol. Lett. **7**, 199203 (2013).

[46]W. Gao and A. Tkatchenko, Phys. Rev. Lett. **111**, 045501 (2013).

[47]G. Kresse, and J. Furthmüller, Comp. Mat. Science **6**, 15 (1996).

[48]D. Riedel et al., Phys. Rev. Lett. **109**, 226402 (2012).

[49]S. Dannefaer, P. Mascher, and D. Kerr, Phys. Rev. Lett. **56**, 2195 (1986).

[50]P. M. Fahey, P. B. Griffin, and J. D. Plummer, Rev. Mod. Phys. **61**, 289 (1989).

[51]Y. Shimizu, M. Uematsu, and K. M. Itoh, Phys. Rev. Lett. **98**, 095901 (2007).

[52]H. Bracht et al., Phys. Rev. Lett. **91**, 245502 (2003).

[53]H. Bracht, E. E. Haller, and R. Clark-Phelps, Phys. Rev. Lett. **81**, 393 (1998).

[54]A. Ural, P. B. Griffin, and J. D. Plummer, J. Appl. Phys. **85**, 6440 (1999).

[55]A. Ural, P. B. Griffin, and J. D. Plummer, Phys. Rev. Lett. **83**, 3454 (1999).

[56]H. Bracht, J. F. Pedersen, N. Sangenberg, A. N. Larsen, E. E. Haller, G. Lulli, and M. Posselt, Phys. Rev. Lett. **91**, 245502 (2003).

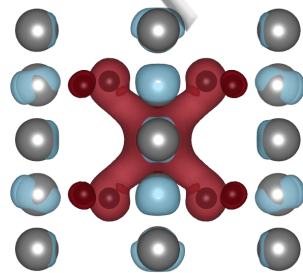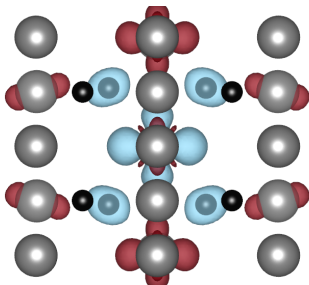57V. Ranki, and K. Saarinen, Phys. Rev. Lett. **93**, 255502 (2004).

(a) Iteration 1   (c) Iteration 3   (e) Iteration 5
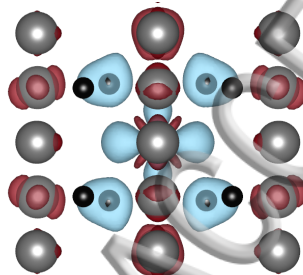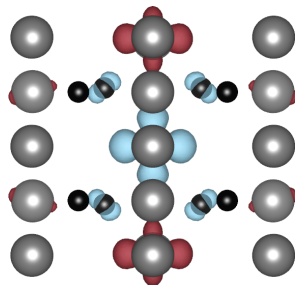
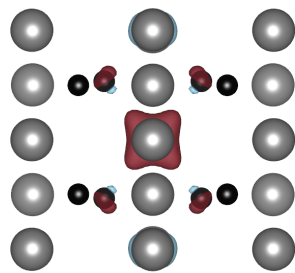(b) Iteration 2   (d) Iteration 4   (f) Iteration 6

Cluster (A): Hybrid
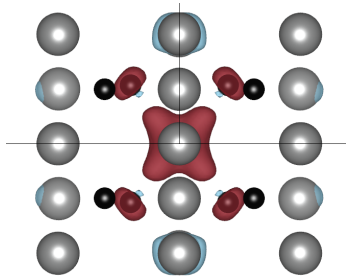
Environment (B): DFT

Density change from previous iteration: increase   (decrease)
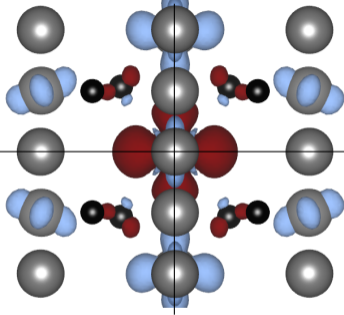
(a) Hybrid   (b) DFT   (c) Embedding
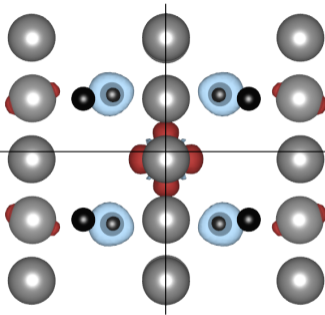
Unpaired spin density
positive   negative

Difference to hybrid benchmark
positive   negative

Difference to hybrid benchmark
positive   negative