Periodic and fragment models based on the local correlation approach

Periodic and fragment models based on the local correlation approach

Denis Usvyat[*], Lorenzo Maschio[†], Martin Schütz[‡]

Article Type:

Advanced Review

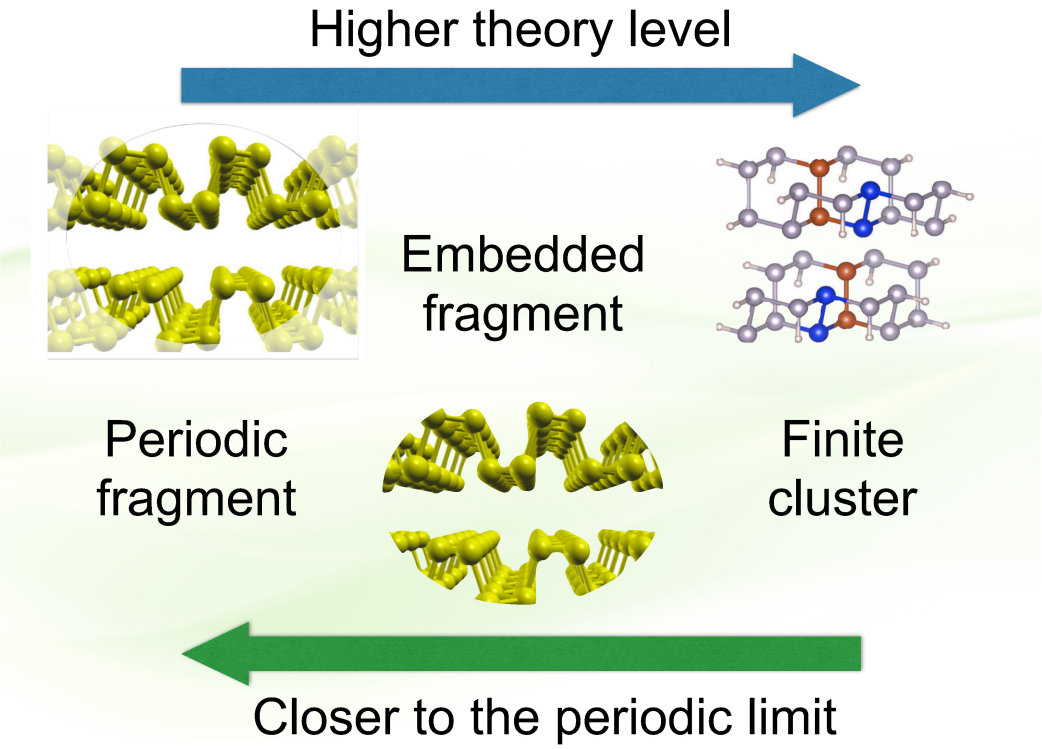
Abstract

Rigorous treatment of dynamical electron correlation in crystalline solids is one of the main challenges in today's materials quantum chemistry and theoretical solid state physics. In this work, we address this problem by using the local correlation approach and exploring a variety of methods, ranging from the full periodic treatment through embedded fragments to finite clusters. Besides the computational advantages, the direct-space local representation for the occupied space allows one to partition the system into fragments and thus forms a natural basis for a hierarchy of embedding models. Furthermore, a subset of localized orbitals in a cluster or a fragment can be chosen to mimic the unit cell of the reference periodic system. Introduction of such subsets allows one to define a formal quantity "the correlation energy per unit cell", which is directly related to the correlation energy per unit cell in the crystal. The orbital pairs, where neither of the two localized orbital indices belongs to the "unit cell" do not explicitly contribute to the "energy per cell": Their role is to provide correlated embedding via the couplings in the amplitude equations. The periodic, fragment and finite-cluster approaches can be combined in a form of high precision computational protocols, where progressively higher-level corrections are evaluated using lower-level embedding models. We apply these techniques to investigate the importance of Coulomb screening in dispersively interacting systems on the examples of phosphorene bilayer and adsorption of water on 2D silica.

[*]Institut für Chemie, Humboldt-Universität zu Berlin, Brook-Taylor-Strasse 2, D-12489 Berlin, Germany

[†]Dipartimento di Chimica, and NIS (Nanostructured Interfaces and Surfaces) Centre, Università di Torino, via Giuria 5, I-10125 Torino, Italy

[‡]Institut für Chemie, Humboldt-Universität zu Berlin, Brook-Taylor-Strasse 2, D-12489 Berlin, Germany



1 INTRODUCTION

Solid state physics and chemistry is the realm of (Kohn-Sham) density functional theory (DFT); the vast majority of theoretical studies in that field rely on this method. Mostly, relatively inexpensive local exchange-correlation functionals are generally used.[1] Standard DFT, however, is known to have certain inherent deficiencies, such as the self-repulsion and self-correlation of electrons, or the lack of long-range van der Waals dispersion. Furthermore, in contrast to wavefunction (wf) based methods, where a clear methodological hierarchy exists, there is no way to systematically improve the accuracy of DFT results. In that sense, DFT offers a take it or leave it deal.

The lack of long-range van der Waals dispersion in DFT is a particularly severe weakness of DFT in the context of 3D extended systems like solids. It can noticeable influence the structure and energetics of a solid and may even decide the relative stability of different polymorphs[2–4]. The lack of dispersion in DFT can be rigorously repaired by employing the random phase approximation as the correlation functional.[5,6] However, such DFT calculations are computationally much more expensive than standard DFT calculations, and therefore applications to solids or adsorption on surfaces employing RPA are much more rare. Another, more economical way to circumvent the lack of dispersion in standard DFT is to add the interatomic pairwise C_6R^{-6} contributions to the DFT energy[7–10]. Such ap-

proaches are presently widely used in the theoretical solid state community; nevertheless, the empiricism and the double counting of the correlation energy can derogate accuracy and predictive power of the method.

In recent years, wf-based post Hartree-Fock methods started to infiltrate the field. There are two main approaches to dynamical electron correlation in solids: (i) a periodic approach, that represents the solid as a system with translational symmetry, and (ii) a finite-cluster approach, where the solid is approximated by a cluster (optionally embedded), a fragment, or a series of those. The periodic model itself is usually more reliable for crystals, but periodic implementations of high-level quantum chemical models are scarce and not yet available for routine applications. The cluster models are also problematic, but in a different way: since the computational cost grows very rapidly with the cluster size, computationally tractable clusters might be not sufficient for an accurate description of the crystal.

Actually, *finite cluster based* approaches to tackle 3D-periodic systems in the framework of wf-based methods, like the incremental scheme proposed by Stoll,[11–13] have been around since the 1990s. The latter has been successfully applied to many different systems over the past years[14–26]. The incremental scheme in principle converges to the periodic result, but the convergence rate with respect to the locality and the order of the many-body expansion can vary very strongly depending on the system and quantity in question (*e.g.* total vs correlation energy). The desired accuracy might therefore be not reachable with a computationally affordable number of increments, if the corresponding clusters are not properly embedded. This adds another level of complexity to this already quite a technically involved method. Besides, mismatch of the orbital spaces in different clusters can have an influence on the corresponding increments, which can finally accumulate to an appreciable and virtually uncontrollable error.

During the past decade, truly periodic wf-based post Hartree-Fock methods, *i.e.*, methods employing periodic boundary conditions, started to emerge.[27–47] Wf-based methods do not share abovementioned deficiencies of standard Kohn-Sham DFT and constitute a hierarchy of well-controlled approximations. In molecular quantum chemistry they are widely used as a highly reliable and accurate alternative to (the cheaper) DFT. Yet wf-based methods are more complicated and consequently computationally more expensive than DFT. Especially the high scaling of the computational cost with system size is problematic and precluded implementations of such methods in the periodic format: already the simplest post Hartree-Fock method, Møller-Plesset perturbation theory (MP) of second order (MP2) has a scaling

of $\mathcal{O}(\mathcal{N}^5)$ with system size $\mathcal{N}$. This increases even to $\mathcal{O}(\mathcal{N}^7)$ for the coupled cluster singles doubles with perturbative triples method, CCSD(T), which represents the "gold standard" of quantum chemistry. This scaling wall is in fact an artifact of the utilization of delocalized canonical orbitals in the construction of the Slater determinants; by instead using localized orbitals the short-range character of electron correlation effects can be efficiently exploited. On that basis during the past decades in molecular quantum chemistry a number of efficient methods were devised with low or even $\mathcal{O}(\mathcal{N})$ scaling[48–62] up to the level of CCSD(T) or even beyond.[63] This gave impetus also to the development of periodic local wf-based post Hartree-Fock methods.

The first real periodic correlation method, applicable to 3D as well as to 2D, 1D and 0D periodic crystals, was the local MP2 (LMP2) method implemented in the Cryscor program.[28,34,35,39,44,64–67] The Hartree-Fock reference orbitals are provided by the Crystal program,[68] that uses a basis set of atom-centered Gaussian-type orbitals (GTOs). From these crystalline orbitals localized Wannier functions (WFs)[69,70] and projected atomic orbitals (PAOs)[66,71–76] are then generated. WFs span the occupied space of the Hartree-Fock determinant and PAOs the virtual orbital space; the WFs are mutually orthogonal, whereas the PAOs are non-orthogonal and redundant. In order to reduce the computational cost essentially two local approximations are introduced: (i) a hierarchical treatment of WF pairs on the basis of the mutual separation between the two WFs of a pair, and (ii) pair specific restricted excitations spaces (pair domains), allowing occupied $\rightarrow$ virtual substitutions out of the WF pair into the related pair domain only.

As an alternative to PAOs, recently several other choices of local virtual orbitals have been proposed, mainly in the molecular context. The highest possible compactness of pair-specific virtual spaces is provided by pair natural orbital (PNOs), which were initially proposed a few decades ago[77] and recently regained interest in the local context.[51,55,56,60–62,78–86] In addition to the computational savings, PNOs naturally adjust themselves to the changes in geometry, leading to smooth potential surfaces. For PAOs, for example, alternation of the domain sizes along the potential surface might cause noticeable discontinuities. At the same time, in contrast to PAOs the overall amount of PNOs is vast (each pair has its own unique orbitals), leading to a surge in the size of certain intermediate tensors. The so called orbital specific virtuals (OSVs)[44,53,54,87–89] offer a convenient compromise between the pair-specific virtual space and an overall amount of orbitals. For diagonal pairs they are equivalent to PNOs, while for the off-diagonal ones the virtual space is formed as a union of the respective spaces corresponding to the individual WFs forming the pair. OSVs were recently introduced also

in the periodic LMP2 implementation in CRYSCOR.[3,44,90]

Apart from the local MP2 method just sketched above, also non-local *canonical* MP2 and CCSD methods have been presented in the literature.[30,31,33,45,47,91] Furthermore, massively parallel implementations of canonical MP2, and even MP2 nuclear energy gradients have been recently reported.[36,38,42,46] Finally we also mention Quantum-Monte-Carlo (QMC) method, which stochastically samples the wavefunction,[92] and can be presently applied to quite complicated periodic systems. Conventionally QMC samples the wavefunction in the position vector representation and requires the fixed-node approximation to gain efficiency. Recently a QMC-based full configuration interaction method was proposed,[93] which samples the wavefunction in the slater-determinant representation. It has been already applied to periodic systems, employing both canonical[41] and local[94] orbitals.

As mentioned above, the cluster approach to electron correlation in solids is attractive because it automatically allows to adopt all the software arsenal available for molecular calculations. The price to pay is the slow and difficult convergence of results with respect to increasing cluster size, needed to ensure that the model cluster indeed provides an adequate representation of the infinite solid. A natural way to alleviate this problem is to embed the cluster in an environment, such that the influence of the missing part of the solid will be effectively represented by the latter. Through years numerous embedding schemes have been proposed. They can be divided in two categories: (i) an explicit embedding, where the cluster (or a fragment) experiences an explicit field from the embedding environment; and (ii) hierarchical embedding, where the cluster is treated as is, but its lower-order energy component is substituted by that, computed using the periodic model or a larger cluster. Such an implicit embedding can be equivalently formulated as a correction scheme, where a periodic lower-order energy is augmented by a higher-order correction, evaluated on a cluster. One of the simplest embedding schemes of the first kind is point-charge electrostatic embedding, where the cluster is placed inside a matrix of point charges. For higher accuracy, however, such a simple scheme is in many cases not sufficient, and various more elaborate embedding approaches have been proposed.[25,94–110]

The methods of the second type have also been employed quite extensively. For example, in the abovementioned method of increments for crystals, the finite-cluster calculations are usually carried out for the correlation energy only, which is added to the periodic Hartree-Fock.[11,17] Another popular choice of the lower-order periodic treatment is DFT.[102,111–114] Recently, with the development of the periodic correlated techniques, the periodic local MP2 was used as the lower-level method.[25,115–117]

In this contribution we overview the periodic, embedded-fragment and finite cluster techniques, based on the local correlation scheme, as implemented in the CRYSTAL/CRYSCOR and MOLPRO[118,119] program packages. We demonstrate that by combining these techniques one can reach a very high accuracy in applications to periodic systems.

We start from the periodic local MP2, which formally can also be seen as a limiting case of an LMP2 treatment for a fragments self-consistently embedded in an LMP2 environment. The next level in the embedding hierarchy is LMP2 or local direct ring coupled cluster doubles (LdrCCD) for a fragment, embedded in the periodic HF mean field. The final embedding model is a large finite cluster treated at the CC level up to LCCSD(T), in which a subset of the localized occupied orbitals define a formal unit cell, which is embedded in the rest of the cluster. We will refer to such an approach as "unit cell in cluster".

Furthermore, we will combine this embedding hierarchy with the method correcting scheme. The lower order energy component will be provided by the periodic LMP2 or embedded fragment LdrCCD. The correction as the energy difference between LCCSD(T) and LMP2 or LdrCCD will be evaluated using finite clusters and the "unit-cell-in-cluster" approach.[3] The recently developed LCCD[S]-R^{-6} scheme,[58,59] applicable to large clusters, allows for accurate description of the weak intermolecular interactions, making possible to explore the convergence of the correcting scheme with the cluster size. As an application of this technique we will investigate the importance of Coulomb screening effects in dispersively bound periodic systems. Phosphorene bilayer and water adsorbed on 2D silica will be taken as examples.

2 Theory

In this section we introduce the hierarchy of embedding models and provide the underlying formalism.

2.1 Preliminaries: definitions, notation and nomenclature

The local correlation scheme is the essential integral part of all the approaches discussed in this work. In the considered models the occupied space will be spanned by either Wannier functions (WFs), or, in case of finite clusters, by localized molecular orbitals (LMOs). The occupied orbitals will be denoted by the indices i, j, k, ... In the periodic case, these indices alone refer to the WFs belonging to the reference unit cell, while for the WFs outside the reference cell an additional calligraphic index $\mathcal{I}$, $\mathcal{J}$, $\mathcal{K}$, ... will denote the corresponding

translation vector. As the virtual orbitals in this work we will use PAOs, denoted as a, b, c, ..., and in the periodic case the PAO cells will be again identified by the calligraphic indices $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$. Within the local approximation the virtual space is restricted to pair domains, denoted $[ij]$, which are unions of the individual PAO domains of the orbitals i and j. Due to the time reversal symmetry all the orbitals will be chosen real. The electron repulsion integrals (ERIs) will be given in the chemical notation, *e.g.*:

$$(ia|jb) \equiv \int d\mathbf{r}_1 \int d\mathbf{r}_2 \psi_i(\mathbf{r}_1)\psi_a(\mathbf{r}_1)\frac{1}{|\mathbf{r}_1-\mathbf{r}_2|}\psi_j(\mathbf{r}_2)\psi_b(\mathbf{r}_2), \quad (1)$$

where $\psi(\mathbf{r})$ represents a direct-space local orbital spanning the occupied or virtual space. Such integrals can be efficiently approximated by a suitable multipolar expansion or the density-fitting factorization.

In this work we also introduce several specific terms, which will allow us to describe and discuss the embedding approaches within the local correlation scheme. These terms reflect the different ways to partition a periodic system or a cluster into local fragments. The first term we introduce is a "*pair fragment*". A pair fragment comprises a subset of the occupied localized orbital pairs, where one of the indices belongs to the reference cell and the other is restricted according to some criterion, for example a cutoff distance. The set of the individual localized occupied orbitals appearing in the pair fragment form the "*orbital fragment*". Finally the atoms that correspond to the union of all the domains of the orbitals from the *orbital fragment* will be referred to as the "*atomic fragment*".

Next, as mentioned above, the term *"unit cell"* we will use not only in the context of periodic system, where it has a well defined meaning. In a large enough *orbital fragment*, cut out of a periodic system, one can identify a subset of the fragment's occupied orbitals, which in the parent periodic system constitute the orbital unit cell. This subset of orbitals we will denote as a *"unit cell in a fragment"*. Introduction of the unit cell in a fragment allows for defining and calculating the *"energy per unit cell"* in a fragment. In contrast to truly periodic systems, such energy is not invariant with respect to the choice of the unit cell. However, the larger the fragment the less sensitive this energy to the choice of the cell and the closer to the periodic result. The quickest convergence to the periodic energy is expected for most compact unit cells, located in the center of the fragment, as the boundary effects are then minimized.

Furthermore, an analogous approach is possible for molecular clusters, provided they sufficiently well mimic the periodic system in question. In this case, the localized orbitals of the cluster resemble those of the periodic system in terms of their localization centers and orientation. This permits for interrelating between the relative sets of orbitals and,

consequently, defining a subset of the cluster's orbitals, whose periodic-system counterparts form the unit cell. We well refer to this subset as a *"unit cell in cluster"*. Similarly to the case of a fragment, considered above, the *"energy per cell"* in a cluster does depend on the choice of the "cell", but if the latter is located in the center of the cluster, the convergence of the corresponding energy to the periodic value with expansion of the cluster is expected to be the fastest.

2.2 Periodic local MP2 method

We start with the fully periodic direct-space LMP2 formalism.[65] The LMP2 energy *per cell* is given by an expression:

$$E_{\mathrm{LMP2}} = \sum_{i\,j\mathcal{J}\in(i\,j\mathcal{J})} e_{i\,j\mathcal{J}}, \tag{2}$$

where $(i\,j\mathcal{J})$ denotes the pair list and $e_{i\,j\mathcal{J}}$ are the pair energies:

$$e_{i\,j\mathcal{J}} = \sum_{a\mathcal{A}\,b\mathcal{B}\in[i\,j\mathcal{J}]} \widetilde{T}^{i\,j\mathcal{J}}_{a\mathcal{A}\,b\mathcal{B}}\,(i\,a\mathcal{A}|j\mathcal{J}\,b\mathcal{B})\,. \tag{3}$$

Here $[i\,j\mathcal{J}]$ is the PAO pair-domain for the pair i-$j\mathcal{J}$ and $\widetilde{T}$ are the contravariant doubles amplitudes

$$\widetilde{T}^{i\,j\mathcal{J}}_{a\mathcal{A}\,b\mathcal{B}} = 2T^{i\,j\mathcal{J}}_{a\mathcal{A}\,b\mathcal{B}} - T^{i\,j\mathcal{J}}_{b\mathcal{B}\,a\mathcal{A}}. \tag{4}$$

Importantly, the pair list restricts the summation over one of the indices in (2) (here the first index) to the reference cell only. One can also symmetrize the expression for the correlation energy per cell with respect to this restriction:

$$E_{\mathrm{LMP2}} = \frac{1}{2}\sum_{i\,j\mathcal{J}\in(i\,j\mathcal{J})} (e_{i\,j\mathcal{J}} + e_{j\mathcal{J}\,i})\,. \tag{5}$$

Another important aspect of eq. (2) is that formally the sum over $\mathcal{J}$ runs to infinity, but in real calculations this index obviously has to be restricted, for example according to a certain cutoff radius R_{cut}. The convergence of the LMP2 energy with respect to R_{cut} is rather slow; the missing energy contribution in 3D solids fades off just as R_{cut}^{-3}. In the reciprocal space approach this manifests as the slow convergence of the correlation energy with the k-mesh. In the direct space, however, one can avoid seeking the explicit convergence with R_{cut}. To obtain the converged periodic E_{LMP2} it is sufficient to choose R_{cut} such that the pair energies $e_{ij\mathcal{J}}$ for each pair of orbitals i and j attain the long-range R^{-6} decay regime:

$$e_{i\,j\mathcal{J}} \approx \frac{-C_6^{ij}}{R^{-6}_{i\,j\mathcal{J}}}, \tag{6}$$

where $R_{i\,j\mathcal{J}}^{-6}$ is the distance between the centroids of the WFs i and $j\mathcal{J}$ Then the pair-specific C_6^{ij} coefficients can be fitted using the explicitly computed pair energies and the missing correlation energy beyond R_{cut} can be straightforwardly recovered using (6).

In the local direct-space representation the Fock matrix is not diagonal and the LMP2 amplitudes equations are to be solved iteratively in order to obtain the amplitudes $T_{a\mathcal{A}\,b\mathcal{B}}^{i\,j\mathcal{J}}$. In the case of non-orthogonal virtual orbitals (*i.e.* PAOs) the periodic LMP2 residual takes the form:[65]

$$
\begin{aligned}
{}^{\mathrm{LMP2}}R_{i\,j\mathcal{J}}^{a\mathcal{A}b\mathcal{B}} = (i\,a\mathcal{A}|j\mathcal{J}\,b\mathcal{B}) + \sum_{\substack{a'\mathcal{A}',b'\mathcal{B}'\\ \in[i\,j\mathcal{J}]}} \Big[F_{a\mathcal{A}\,a'\mathcal{A}'} T_{a'\mathcal{A}'b'\mathcal{B}'}^{ij\mathcal{J}} S_{b'\mathcal{B}'\,b\mathcal{B}} + S_{a\mathcal{A}\,a'\mathcal{A}'} T_{a'\mathcal{A}'\,b'\mathcal{B}'}^{i\,j\mathcal{J}} F_{b'\mathcal{B}'\,b\mathcal{B}} \Big] \\
- \sum_{k\mathcal{K}} \left[\sum_{\substack{a'\mathcal{A}',b'\mathcal{B}'\\ \in[i\,k\mathcal{K}]}} S_{a\mathcal{A}\,a'\mathcal{A}'} T_{a'\mathcal{A}'b'\mathcal{B}'}^{ik\mathcal{K}} F_{k\mathcal{K},j\mathcal{J}} S_{b'\mathcal{B}'\,b\mathcal{B}} + \sum_{\substack{a'\mathcal{A}',b'\mathcal{B}'\\ \in[k\mathcal{K}\,j\mathcal{J}]}} S_{a\mathcal{A}\,a'\mathcal{A}'} F_{i,k\mathcal{K}} T_{a'\mathcal{A}'b'\mathcal{B}'}^{j\mathcal{J}k\mathcal{K}} S_{b'\mathcal{B}'\,b\mathcal{B}} \right], \qquad (7)
\end{aligned}
$$

where F denotes the internal or external Fock matrix and S is the PAO overlap matrix. The domain approximation restricts the virtual-index ranges to the respective pair domains.

All the quantities entering eq. (7) are translationally invariant:

$$
\begin{aligned}
S_{a\,a'\mathcal{A}'} &= F_{a\mathcal{A}\,a'(\mathcal{A}'\oplus\mathcal{A})}, && (8)\\
F_{a\,a'\mathcal{A}'} &= F_{a\mathcal{A}\,a'(\mathcal{A}'\oplus\mathcal{A})}, && (9)\\
F_{i\,j\mathcal{J}} &= F_{i\mathcal{I}\,j(\mathcal{J}\oplus\mathcal{I})}, && (10)\\
(i\,a\mathcal{A}|j\mathcal{J}\,b\mathcal{B}) &= (i\mathcal{I}\,a\,(\mathcal{A}\oplus\mathcal{I})\,|j\,(\mathcal{J}\oplus\mathcal{I})\;b\,(\mathcal{B}\oplus\mathcal{I}))\,, && (11)\\
T_{a\mathcal{A}\,b\mathcal{B}}^{i\,j\mathcal{J}} &= T_{a(\mathcal{A}\oplus\mathcal{I})\,b(\mathcal{B}\oplus\mathcal{I})}^{i\mathcal{I}\,j(\mathcal{J}\oplus\mathcal{I})}, && (12)
\end{aligned}
$$

where the symbolic operation $\oplus$ applied to the cell indices implies summation of the respective translation vectors.

The LMP2 amplitude equations

$$
{}^{\mathrm{LMP2}}R_{i\,j\mathcal{J}}^{a\mathcal{A}\,b\mathcal{B}} = 0 \qquad (13)
$$

are solved only for the pairs from the pair list $(i\,j\mathcal{J})$. The amplitudes for the pairs outside this list are assumed zero unless they are translationally equivalent [*cf.* eq. (12)] to those included in the list. This means that the pair list restricts also the $k\mathcal{K}$-summations in the last two terms of (7).

A periodic LMP2 with a restricted pair list can be interpreted as a pair-fragment LMP2 (the pair fragment is determined by the pair list), embedded in the periodic HF plus the

periodically translated fragment LMP2 [via the translation of the fragment's amplitudes (12) in the last term of (7)]. By varying the fragment size one can inspect the decay of the finite-size error in the correlation energy, while still remaining in the periodic framework. This type of embedding will be the highest in our the hierarchy of embedding models, and we will refer to it as "periodic fragment".

We stress again that in order to obtain the full periodic correlation energy per cell, one does not need to expand the *pair fragment* till convergence (which is rather slow), as the extrapolation technique (6) after a certain fragment's size becomes quite accurate. In the canonical correlation calculations, an approximation, related but not exactly equivalent to the pair-list truncation, is introduction of a finite k-mesh (or a single Γ-point). This approximation does not allow for a straight-forward embedding interpretation, as in our case, or a pair-wise extrapolation according to eq. (6). However other extrapolation schemes are possible.[117,120]

2.3 Periodic-Hartree-Fock-embedded local direct ring-CCD

As the next level model we consider an *orbital fragment* cut from a crystal and subjected to a molecular local correlation treatment. The localized occupied and virtual orbitals, as well as the Fock matrix in the basis of these orbitals are those of the periodic Hartree-Fock calculation. The PAOs, as usually, belong to the domains of the fragment's orbitals and thus are centered only on the atoms of the *atomic fragment*. Nevertheless, by construction they are orthogonal to the whole periodic occupied manifold rather than that of the fragment. The electron repulsion integrals are also computed in the periodic format, and although they involve only the orbitals of the fragment, they correspond to a subset of the real periodic ERIs. Such an embedding scheme corresponds to a correlated level treatment of the *orbital fragment*, embedded the periodic Hartree Fock solution.

As the methodological level for the fragment-based calculation we choose the local direct ring-CCD (LdrCCD) method.[110] It is similar to the SOSEX variant of the random phase approximation,[121–123] but contain additional diagrams that originate from the antisymmetricity of the amplitudes in the spin-orbital basis. The LdrCCD residual for the *orbital fragment*

reads:

$$
\begin{aligned}
{}^{\text{LdrCCD}}R_{i\mathcal{I}\,j\mathcal{J}}^{a\mathcal{A}\,b\mathcal{B}} &= {}^{\text{LMP2}}R_{i\mathcal{I}\,j\mathcal{J}}^{a\mathcal{A}\,b\mathcal{B}} \\
&+ \sum_{k\mathcal{K}} \sum_{\substack{a'\mathcal{A}',c\mathcal{C} \\ \in[i\mathcal{I}\,k\mathcal{K}]}} S_{a\mathcal{A}\,a'\mathcal{A}'} \widetilde{T}_{a'\mathcal{A}'\,c\mathcal{C}}^{i\mathcal{I}\,k\mathcal{K}} \,(k\mathcal{K}\,c\mathcal{C}|b\mathcal{B}\,j\mathcal{J}) + \sum_{k\mathcal{K}} \sum_{\substack{c\mathcal{C},b'\mathcal{B}' \\ \in[k\mathcal{K}\,j\mathcal{J}]}} (a\mathcal{A}\,i\mathcal{I}|k\mathcal{K}\,c\mathcal{C}) \widetilde{T}_{c\mathcal{C}\,b'\mathcal{B}'}^{k\mathcal{K}\,j\mathcal{J}} S_{b'\mathcal{B}'\,b\mathcal{B}} \\
&+ \sum_{k\mathcal{K}} \sum_{\substack{a'\mathcal{A}',c\mathcal{C} \\ \in[i\mathcal{I}\,k\mathcal{K}]}} S_{a\mathcal{A}\,a'\mathcal{A}'} \widetilde{T}_{a'\mathcal{A}'\,c\mathcal{C}}^{i\mathcal{I}\,k\mathcal{K}} \sum_{l\mathcal{L}} \sum_{\substack{d\mathcal{D},b'\mathcal{B}' \\ \in[l\mathcal{L}\,j\mathcal{J}]}} (k\mathcal{K}\,c\mathcal{C}|l\mathcal{L}\,d\mathcal{D}) \widetilde{T}_{d\mathcal{D}b'\,\mathcal{B}'}^{l\mathcal{L}\,j\mathcal{J}} S_{b'\mathcal{B}'\,b\mathcal{B}}. \qquad (14)
\end{aligned}
$$

Since the orbitals come from the periodic calculation, they still hold the information about the unit cells, they are centered in, and so the quantities in (14) are indexed correspondingly.

Formally, the fragment's correlation energy accumulates the contributions from *all* possible pairs $i\mathcal{I}j\mathcal{J}$, where the both $i\mathcal{I}$ and $i\mathcal{J}$ belong to *orbital fragment*. It is however not straight-forward to relate this energy to the periodic energy per cell. Firstly, this energy grows with the fragment size, rather than converge to the periodic value. Secondly, for any realistic fragment, a substantial share of the pairs would have both WFs located near the fragment's boundaries, which would contaminate the resulting energy with the boundary effects directly via the corresponding pair energy contributions.

In order to alleviate these problems and reach a smooth convergence of the energy obtained in the fragment calculation to the periodic result, one can keep the unit-cell-based definition of the energy also for the fragment. As noted above, for each orbital of the fragment the information on its centering cell in the parent periodic system is still available, so one can identify the pairs where (i) both orbitals, (ii) only one orbital or (iii) none of the orbitals belong to the unit cell. With this information at hand one can define the LMP2 or LdrCCD correlation energy per "unit cell in a fragment" $E_{\text{corr.}}^{\text{u.c.}}$ as [*cf.* eq. (5)]:

$$
E_{\text{corr.}}^{\text{u.c.}} = \sum_{ij} \left(e_{ij} + \frac{1}{2} \left[\sum_{\mathcal{J}\neq 0} e_{i\,j\mathcal{J}} + \sum_{\mathcal{I}\neq 0} e_{i\mathcal{I}\,j} \right] \right). \qquad (15)
$$

For the sake of clarity, the index permutation invariance of the pair energies is not taken into account in eq. (15). The energy expression, where this invariance is exploited can be found in Ref.[110] [*cf.* eq. (7) therein].

The pairs directly contributing to the energy (15) form the *pair fragment* as defined in the beginning of sect. 2. We denote these pairs as reference-cell pairs (*r.c.-pairs*), as at least one of the indices belongs to the reference cell. The remaining pairs, *i.e.* those with both indices outside the unit cell, are nevertheless processed in the calculation. The amplitudes of these pairs are essential as they couple to r.c.-pairs in the amplitude equations: see the

last two terms in eq. (7) and the last three terms terms in eq. (14). In this sense, these pairs provide the correlated (in this case LdrCCD) embedding for the *pair fragment*. Therefore we will call them embedding pairs (*emb.-pairs*).

Figure 1 qualitatively compares the periodic and fragment's pair energies and their contribution to the overall correlation energy per cell, in the case of a model bilayer system with two WFs per cell, which are depicted as circles. The unit cell indices are shown in a calligraphic font. Each arrow represents an inter-layer orbital pair (up to the second-nearest neighbors), with the first index denoted by a square and the second one by an arrowhead.

The correlation energy expression for the periodic case (2) contains only the pairs shown in Figure 1a. This can serve as guide, showing how to obtain the energy per cell in fragment. In the fragment-based treatment all occurring pairs (*cf.* Figure 1b) are explicitly evaluated. Therefore, in order to get the actual energy-per-cell value [*i.e.* eq. (15)], one has to properly scale their contribution in the energy expression to match the pattern of Figure 1a. A direct comparison of the panels of Fig. 1 reveals that the number of pairs with both indices in the reference cell is the same in both periodic and fragment cases, suggesting the factor 1 in the first term of (15). The pairs with only one index in the reference cell are doubled in the fragment calculation compared to the periodic case (note that $e_{ij\mathcal{J}} = e_{j\mathcal{J}i}$). This double counting of the pairs is taken into account by the factor 1/2 for the second and third terms of eq. (15). Finally the emb.-pairs do not have periodic counterparts in Figure 1a, and thus are excluded in (15) entirely.

In order to reveal the difference between the embedding model of this section and the "periodic-fragment" approach of section 2.2, we investigate the translational invariance (or a lack thereof) in the involved quantities. As noted above, the orbitals, the Fock and overlap matrices and the ERIs come from the periodic calculations, so they are inherently translationally invariant. In other words, the relations (9)–(12) are valid, although not explicitly used in the fragment calculations. For the amplitudes, which are evaluated in the fragment-based format, the translation symmetry (12) is not explicitly imposed and thus is lost in the iterations, even if the starting amplitudes are taken from a periodic LMP2 calculation. This manifests the actual principal difference between these two embedding models: in the first case the *pair fragment* is embedded in the translational images of the r.c. amplitudes, while in the second case the emb.-pair amplitudes are evaluated separately. Nevertheless, the latter model still very adequately represents the periodic system, especially so for larger fragments, where the distinction between the embedding regions becomes progressively less significant.

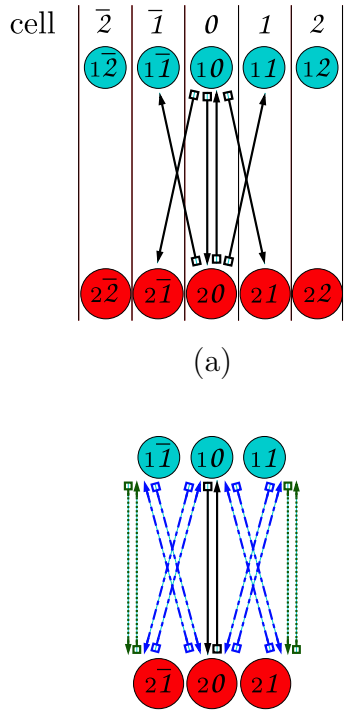


Figure 1: Comparison of (a) periodic and (b) fragment classification of inter-slab-adsorbate pairs, in the case of a model bilayer system with two WFs per cell. The WFs of the lower and upper layers are displayed by smaller blue and larger red circles, respectively, and are denoted by combined intra-cell $(1, 2)$ and cell-counting $(\bar{2}, \bar{1}, \ldots, 2)$ indices. The latter are given by the calligraphic font and decorated with an over-bar if the cell corresponds to a "negative" translation. Panel (a) presents the full periodic system and panel (b) a fragment cut out from it. The pairs are schematically represented by arrows: squares denote the first orbital index in a pair and arrowheads the second. In panel (b) the black solid arrows refer to pairs with *both* orbitals inside the reference cell, the blue dashed arrows to pairs with just one WF inside it, and the green dotted arrows to pairs with both WFs outside the reference cell. Reprinted with permission from Ref.[110]. Copyright 2016, American Chemical Society.

The periodic and embedded fragment approaches have their advantages and disadvantages. Obviously the periodic description is more closely related to the studied periodic system. Besides, the efficiency of the periodic approach benefits from the translational symmetry, such that for any quantity involved in the calculation only its translationally unique subset needs to be explicitly evaluated. The translational invariance, however, imposes also certain difficulties. Firstly the algorithms have to be adapted for translating the intermediates on the fly to make their contractions efficient. This often requires that the batches of the intermediate quantities, that have to be kept in memory simultaneously, should contain the complete cell-index ranges.[39,64]

Secondly, the long-range inter-pair couplings might lead to extremely extended ranges for the intermediates involved in the corresponding contractions. For example, an efficient evaluation of the LdrCCD terms in the amplitude equations is carried out via 3-index intermediates A:[57,110]

$$ {}^{\mathrm{LdrCCD}}R^{a\mathcal{A}\,b\mathcal{B}}_{i\mathcal{I}\,j\mathcal{J}} \;\leftarrow\; \sum_{P\mathcal{P}} A^{i\mathcal{I}\,a\mathcal{A}}_{P\mathcal{P}} A^{j\mathcal{J}\,b\mathcal{B}}_{P\mathcal{P}}, \tag{16} $$

defined in eq. (18) of Ref.[110]. The index $P\mathcal{P}$ denotes here the auxiliary functions, used for the density fitting representation of the ERIs of the first term in eq. (7) and last three terms in eq. (14). In the periodic format this implies that the range for the auxiliary functions $P\mathcal{P}$ should cover all the possibly involved product densities $i\mathcal{I}a\mathcal{A}$, $j\mathcal{J}b\mathcal{B}$, $k\mathcal{K}c\mathcal{C}$, $l\mathcal{L}d\mathcal{D}$ appearing in these terms. For the r.c.-pairs the range of the second orbital in a pair and the corresponding product density is restricted by the pair list. However, the couplings involve also emb.-pairs, which implies additional translation of both indices. This becomes especially severe in the quadratic LdrCCD term [the fifth term in eq. (14)], which contains two additional translational extensions into the emb.-pair region. The range for the fitting functions, which has to follow the range of the involved densities, thus expands vastly.

In a fragment-based calculation, where the translational invariance of the amplitudes is not preserved anyway, such problems are absent. The fitting orbital range is simply determined by the *atomic fragment* and does not grow beyond the range defined by the initial R_{cut}. Another advantage of the fragment approach is that its implementation is technically much easier, as it requires only an appropriate interface between the periodic and molecular programs. The actual calculation is processed using the existing well optimized and tested molecular implementation. This becomes especially relevant for higher-level coupled cluster approaches, such as LCCSD or LCCSD(T). Implementation of interfaces allowing for such calculations in the embedded fragment framework is currently in progress.

2.4 Large-finite-cluster LCCSD(T)

The third model we consider here is a large cluster that is obtained by cutting an *atomic fragment* out of a crystal and, in case dangling bonds occurring after the cut, saturating them by hydrogen atoms. Since we will seek for reasonably accurate finite-cluster representations of the periodic system, the clusters should be sufficiently large. In particular they will be assumed to be larger than the unit cell of the parent system.

In analogy to the embedded fragment approach we define a formal unit cell, now in the cluster, as this will allow us to obtain a quantity directly related to the periodic calculation, namely the energy per unit cell. Obviously the energy per unit cell in cluster is not invariant with respect to choice of such a unit cell. But we again use this lack of invariance for reaching the fastest possible convergence to the periodic result by placing the unit cell in the center of the cluster.

In contrast to the fragment, however, the localized orbitals, that define the unit cell, cannot be directly transferred from the periodic calculation to the cluster. Therefore they have to be identified among the cluster's localized orbitals. On the other hand, since the atomic and bonding structure of the cluster has to closely resemble that of the crystal, the principal localization and orientation of the localized orbitals in the crystal and cluster have to match, at least when they can clearly be assigned to bonding and/or lone-pair orbitals. Technically the procedure for the choice of the unit cell in cluster is described below in sect. 2.5.

Once the unit cell is chosen, the cluster's pairs can be grouped into the three categories depending on how many orbitals of a pair belong to the unit cell: (i) both orbitals (r.c.-pairs A), only one orbital (r.c.-pairs B), or none of them (emb.-pairs.). Following the discussion in sect. 2.3 the standard molecular expression for the correlation energy $E_{\text{corr.}} = \sum_{ij} e_{ij}$ has to be modified to provide the correlation energy per cell:

$$E_{\text{corr.}}^{\text{u.c.}} = \sum_{ij\in(\text{r.c.}-\text{A})} e_{ij} + \frac{1}{2} \sum_{ij\in(\text{r.c.}-\text{B})} e_{ij} + 0 \sum_{ij\in(\text{emb.})} e_{ij}. \tag{17}$$

Similarly to the fragment case, the amplitudes of all the pairs are updated and then used in the couplings, but only those of the r.c.-pairs directly contribute to the energy. The correlated embedding model is therefore similar in both approaches: the *pair fragment* is embedded in the complete set of fragment's or cluster's amplitudes. What differs now is the HF embedding level: in the fragment approach it is the periodic HF, while in the finite-cluster one it is just cluster's HF. It is the lowest-level embedding model considered here. However, it is still expected to provide reasonably accurate results, provided the cluster

is large enough. Comparable or even lower-level embedding models are commonly used in incremental calculations for solids,[25,99,124,125] which nevertheless have been demonstrated to deliver rather accurate results. Besides, in contrast to the standard incremental scheme calculations, we will use the finite cluster model not for obtaining its full correlation energy, but rather for corrections to the correlated periodic or embedded fragment results (*cf.* sect. 2.6).

The substantial advantage of the finite cluster model is that the highly efficient implementations of molecular local coupled cluster methods can be readily exploited. In particular, it opens a possibility to reach the LCCSD(T) level of accuracy for periodic systems. However, if the weak intermolecular interactions play an essential role in the studied system, the standard PAO-based LCCSD(T)[52] treatment could provide unsatisfactory description. Indeed, the actual LCCSD residual equations are so complicated and computationally intensive, that in the PAO-based LCCSD(T) version of MOLPRO they are conventionally processed for the so called strong pairs only,[52] *i.e.* the pairs with a very short interorbital distance. The reasoning behind this approximation is that these pairs provide the major fraction of the total correlation energy. The remaining pairs, *i.e.* the close and weak ones, are treated at the LMP2 level. The only difference between the close and weak pairs is that the former are included in the L(T) residuals and optionally could provide a feedback to the strong-pair LCCSD residuals.

In many applications, especially those dealing with intermolecular interactions, the short-range correlation energy cancels to a large extent in the energy differences and the interaction energy is dominated by the long-range interactions, which within the conventional scheme remains virtually at the LMP2 level. Recently we proposed a new formulation of the PAO-based LCCSD method, that treats close and weak pairs much more accurately, but whose computational cost is comparable to that of the initial scheme.[58,59] The underlying principle used in the formulation of this model, denoted as LCCD[S]-R^{-6}, was the following: the close and weak pairs should be included in a common coupled cluster amplitude equation framework, but the quickly decaying diagrams could be omitted. Importantly, the computationally expensive ladder diagrams, that do decay quite fast,[58–60,62,126] could be excluded from the close and weak pair amplitude equations.

The specification of the LCCSD(T)|LCCD[S]-R^{-6} model can be found in Ref.[59]. Here is a brief overview:

- The strong-pair residuals contain:
  - (i) all the LCCSD diagrams (*cf.* Refs.[49,73]) involving singles and/or strong-pair

doubles amplitudes;

  - (ii) all the close-pair-doubles diagrams, whose leading energy contribution decays as R^{-6} (the LMP2-terms plus the diagrams D1-D9 and D13-D16 of Ref.[58]). These diagrams correspond to the LCCD-R^{-6} model;
  - (iii) all the linear weak-pair-doubles diagrams, whose leading energy contribution decays as R^{-6} (the LMP2-terms plus the diagrams D1, D3 and D6 of Ref.[58]). These diagrams correspond to the LrCCD3 model.

- The close-pair residuals contain:
  - (i) the LCCD-R^{-6} diagrams involving strong/close-pair amplitudes;
  - (ii) the LrCCD3 diagrams of weak-pair amplitudes.

- The weak-pair residuals contain:
  - the LrCCD3 diagrams of all strong-, close- or weak-pair amplitudes.

- The R^{-6}-decaying singles contribution, involving the close-pair doubles amplitudes, is calculated perturbatively at the end in the flavor of the "[T]" approach (and therefore denoted as "[S]"), *i.e.* including only the 4th order terms with the converged LCCSD amplitudes (the diagrams D18-D21 of Ref.[59]).

- The triples as before are computed for the close/strong pairs[48,52] (among the three pairs that form a triple two are allowed to be close).

- Since the inter-monomer pairs are of major importance for the intermolecular interactions, we assign them to be close pairs regardless of their inter-orbital distance or connectivity.

The LCCSD(T)|LCCD[S]-R^{-6} model was benchmarked on a wide range of intermolecular systems and was found to be very accurate. We therefore apply it now to study intermolecular interactions in solids. The correlation energy per cell within this approach requires additional regularization. Firstly, the LCCSD correlation energy expression has a quadratic singles term, which can be represented as a formal sum over pairs. The LCCSD pair energy can be written as

$$e_{ij}^{\mathrm{LCCSD}} = \sum_{ab\in[ij]} \widetilde{T}_{ab}^{ij}\,(ia|jb) + \frac{1}{2}\sum_{\substack{a\in[i]\\ b\in[j]}} \left(2T_a^iT_b^j - T_b^iT_a^j\right)(ia|jb) \tag{18}$$

and plugged in eq. (17). Here T_a^i are the singles amplitudes and $[i]$ – the singles domains.

Secondly, the "[S]" and "(T)" contributions are computed as *a posteriori* energy corrections and thus require separate expressions. The "[S]"-correction per cell [*cf.* eq. (5) of Ref.[59]] originates from the feedback from the singles to the close-pair doubles, so its value per cell takes the form:

$$E^{\text{u.c.}}_{\text{L[S]}} = 2 \sum_{\substack{ij \\ \in(\text{r.c.}-\text{A})}} \sum_{a\in[i]} T_i^a \sum_{bc\in[ij]} \widetilde{T}_{bc}^{ij}\,(ab|jc) + \sum_{\substack{ij \\ \in(\text{r.c.}-\text{B})}} \sum_{a\in[i]} T_i^a \sum_{bc\in[ij]} \widetilde{T}_{bc}^{ij}\,(ab|jc)) + \sum_k \sum_{c\in[k]} T_k^c \left[2 \sum_{\substack{ij \\ \in(\text{r.c.}-\text{A})}} \sum_{ab\in[ij]} S_{ca} \widetilde{T}_{ab}^{ij}\,(bj|ik) + \sum_{\substack{ij \\ \in(\text{r.c.}-\text{B})}} \sum_{ab\in[ij]} S_{ca} \widetilde{T}_{ab}^{ij}\,(bj|ik) \right] . \quad (19)$$

For the perturbative triples energy[48] per cell $E^{\text{u.c.}}_{\text{L(T)}}$ we have to partition the occupied orbital triples into 4 subsets: (i) all three orbitals are in the reference cell (r.c.-triples A), (ii) only two orbitals are in the reference cell (r.c.-triples B), (iii) only one orbital is in the reference cell (r.c.-triples C), (iv) all three orbitals are outside the reference cell (emb.-triples). A procedure, analogous to that of sect. 2.3 for the doubles, gives

$$E^{\text{u.c.}}_{\text{L(T)}} = \frac{1}{3} \left[1 \sum_{\substack{ijk \\ \in(\text{r.c.}-\text{A})}}{}' e_{ijk} + \frac{2}{3} \sum_{\substack{ijk \\ \in(\text{r.c.}-\text{B})}} e_{ijk} + \frac{1}{3} \sum_{\substack{ijk \\ \in(\text{r.c.}-\text{C})}} e_{ijk} + 0 \sum_{\substack{ijk \\ \in(\text{emb.})}}{}' e_{ijk} \right] . \quad (20)$$

where a prime over a summation sign means exclusion of the $i = j = k$ terms. The individual triples energies e_{ijk} are given by the following expression:[48]

$$e_{ijk} = \sum_{abc\in[ijk]} \left(X_{abc}^{ijk} \left[W_{abc}^{ijk} + \sum_{d\in[i]} T_d^i S_{ad}\,(jb|kc) + \sum_{d\in[j]} T_d^j S_{bd}\,(ia|kc) + \sum_{d\in[k]} T_d^k S_{cd}\,(ia|jb) \right] \right) , \quad (21)$$

where the intermediates X_{abc}^{ijk} and W_{abc}^{ijk} are given by the equations (30) and (19) of Ref.[48], respectively, and $[ijk]$ denote the triples domains. We note that the energy expression (20) does not exploit the index-permutation symmetry in the triples energies e_{ijk}. In order to include it, one has to restrict the occupied orbital summation ranges to $i \geq j \geq k$ and substitute the uniform factor of $1/3$ with a term-specific one $2 - \delta_{ij} - \delta_{kj}$ [*cf.* eqs. (28) and (29) of Ref.[48]].

Finally we note that in the L(T0) approximation,[48] which omits the inter-pair couplings in the triples residuals (*i.e.* the off-diagonal elements of the internal Fock matrix in the localized orbital basis are neglected), the emb.-triples become unnecessary (they provide

neither contribution to the energy nor embedding through the couplings). This can noticeably reduce the overall number of triples as well as of the intermediates (*e.g.* 3-external ERIs associated exclusively with emb.-triples), leading to substantial computational savings.

2.5 Technical details

The molecular local coupled cluster method, used in this work, originates from the local correlation theory of Pulay and coworkers.[71,72,126–131] The formalism of local CCSD was presented in Ref.[73] and its efficient implementation in the MOLPRO code in Refs.[49,50,52] (the local density fitting approach for 4-external ERIs can be found in Ref.[132]). Ref.[63] describes the triples correction L(T). Finally the recent LCCSD(T)|LCCSD-R^{-6} technique was proposed and tested in Refs.[58,59].

The periodic local MP2 method, implemented in the CRYSCOR code,[65] also stems from Pulay's local correlation scheme. It uses the AO-based periodic HF reference[133] computed by the CRYSTAL program.[134] The initial periodic LMP2 implementation had the ERIs calculated by a 4-index transformation.[28] However only when the density fitting representation of ERIs[64,67,135] was implemented, the method became practically usable. The currently used version of the local density fitting is described in Refs.[39,67]. In Refs.[64,65] one can find the algorithms for solving the periodic LMP2 equations [*cf.* eq. (13)], and in Refs.[35,65] – the description of the C_6/R_{ij}^6-type extrapolation technique [*cf.* eq. (6)]. Construction of the occupied WFs is processed partially in the reciprocal and partially in the direct space according to the procedure of Zicovich-Wilson et al.[70], while the PAOs are generated in the reciprocal space and then transformed into the direct space.[66] Very recently we presented the embedded LdrCCD method.[110]

The implementational details of these methods are well documented and we will not reiterate them here. We will focus only on two technical aspects, which are relevant for the present study, but have not yet been fully described. First, we explain how the pair- and orbital fragments are built for adsorbate-slab or two-slab systems. In such calculations there are two very distinct groups of pairs: intra-slab (or intra-adsorbate) and inter-slab-adsorbate. Within a given R_{cut}, the intra-pairs are usually much more numerous than the inter-pairs. At the same time, when the interaction energy is calculated, the intra-pair-energy to a large extent cancels out in the energy difference, while the inter-pair energy remains entirely intact. Hence, the direct application of a single R_{cut} tolerance is inefficient, since it is strongly biased towards an unnecessary expansion of the intra-pair list, while the essential inter-pair list remains poor unless very large R_{cut} is used.

One can use of course two different tolerances $R_{\text{cut}}^{\text{intra}}$ and $R_{\text{cut}}^{\text{inter}}$ as in Ref.[110], but with two parameters the convergence test becomes much more difficult. Besides, in the Molpro LdrCCD implementation, each orbital from the *orbital fragment* has to have a corresponding diagonal pair in the pair list. This means that even if an orbital $j\mathcal{J}$ appears in the *orbital fragment* only because of some inter-pair $ij\mathcal{J}$, it will automatically generate a diagonal embedding intra-pair $j\mathcal{J}\,j\mathcal{J}$. However, in the slab-only calculation, which is needed to compute the energy difference, this $ij\mathcal{J}$ inter-pair will be obviously absent and thus $j\mathcal{J}\,j\mathcal{J}$ will not appear either. In such a situation the *pair fragments* and thus the energies in the slab-adsorbate and isolated-slab calculations could be unbalanced, ruining the reliability of the energy difference.

In order to avoid these problems, for defining the fragment we employ a three-step procedure based on a single tolerance $R_{\text{cut}}^{\text{intra}}$. We start with determining all the intra-pairs for the slab. The r.c.-intra-pairs are obtained directly using $R_{\text{cut}}^{\text{intra}}$. That yields the intra-slab part of the *pair fragment* and thus of the *orbital fragment*. Then the emb.-intra.-pairs are added, which have the following properties: (i) both orbitals of each emb.-pair must belong to the *orbital fragment*, already specified by the r.c.-pairs, and (ii) each emb.-pair must have a translational image among the r.c.-pairs.

As a second step, the same procedure is applied to the intra-adsorbate fragment. After that the *orbital fragment* is considered to be defined. With that as the third step the r.c.-inter-pairs are obtained, as *all possible* inter-r.c.-pairs within this *orbital fragment*. Finally the inter-emb.-pairs are added, which are defined as the pairs that fulfill the above-mentioned two criteria, now applied to the inter-pairs.

With such a procedure we achieve a rich list of inter-pairs regardless of $R_{\text{cut}}^{\text{intra}}$. Therefore, even small $R_{\text{cut}}^{\text{intra}}$ calculations provide meaningful results and the convergence with the fragment size is expected to be fast, which saves a considerable computational effort of processing the otherwise very numerous intra-pairs. At the same time the whole procedure is driven by a single tolerance making expansion of the fragment straightforward. Finally, since the $R_{\text{cut}}^{\text{intra}}$ is always applied to the intra-pairs, no imbalance can occur in the slab-adsorbate vs isolated-slab or isolated-adsorbate calculations.

The second topic of this section refers to the finite-cluster case and describes how the "unit cell in cluster" is determined. As was explained in sect. 2.4, such a "unit cell" is defined by a set of localized occupied orbitals, which are similar to those forming the unit cell in the crystal. If the cluster is a good representation of the solid, such orbitals can always be found. Unfortunately, a manual search for these orbitals is very tedious and has to be

automatized to make the approach practicable. One of the ways to simplify this process is to relate the orbital unit cell to the atomic unit cell. The latter can easily be located in the cluster even before any calculation is taking place. However, a naive inclusion of all localized orbitals, that are centered near the atoms of the unit cell, could still lead to a wrong result.

We will illustrate this problem on an example: consider a linear chain (-A-B-A-B-A-B-) with two atoms in the cell A and B, each having one lone-pair WF and one bonding WF connecting it with the neighboring atom to the right. The orbital unit cell consists of 4 WFs: the lone-pair WFs of both atoms and the bonding WFs, connecting A with B both inside the zeroth cell, and B from the cell -1 with A from the zeroth cell. However, if we assign all the WFs that belong to the atoms of the zeroth cell, we well get 5 WFs: two lone pair WFs, but three bonding WFs, since the orbital that connects the B atom from zeroth cell and A from the first cell will be also added, although this WF is just a translational image of one of the in-cell WFs.

In order to avoid this double counting we employ the following procedure, which correctly determines the orbital unit cell provided all the localized orbitals have a pronounced bonding (two atoms in the core domain) or lone-pair (one atom in the core domain) character. We divide all unit cell atoms in two classes: class A and class B. For each atom of the class A we include all WFs that have this atom in their core domains. This means that these WFs are allowed to connect the atoms of class A with atoms outside the unit cell. At the same time, for the atoms of class B this is not allowed: if an orbital has an atom of class B in its domain, the other atom in the domain must be also inside the unit cell. Otherwise this orbital will not be included in the list of the unit cell orbitals. Returning to our example, the atom A should be then of the class A and the atom B of the class B. Then the unit cell will correctly consist of two lone-pair WFs and two bonding ones, while the spurious WF, that connects the atom B with the atom A from the cell 1 will be omitted.

We note, that this scheme does not provide a black box solution, as finding the class A and class B atoms has to be carried out in the manual regime. Though it is way simpler than picking up the proper LMOs, for complicated periodic systems, it could still be a formidable task. Presently we are working on a completely automated scheme, but it turns out to be technically rather complex and cumbersome.

2.6 Hierarchical correction scheme

One of the essential advantages of the wavefunction-based approach is a possibility for a systematic improvement of the result. In particular, within the CCSD(T) level one attains

a very high accuracy, at least in single-reference cases, such that these results can serve as a benchmark for both experiment and theory. For solids, however, CCSD(T) calculations without additional approximations are virtually impossible due to the excessive computational cost. At the same time, the presented hierarchy of embedding approximations allows one to effectively approach the CCSD(T) level by accumulating different energy components and refining corrections.

Quite a number of protocols based on energy correcting schemes that combine a lower-order description at a periodic level with higher-order correcting energy increments evaluated on a cluster or set of clusters can be found in literature. They commonly use the following form of the energy expression:

$$E = E_{\mathrm{LL,per.}} + \delta E_{\mathrm{HL-LL,cl.}} \tag{22}$$

with

$$\delta E_{\mathrm{HL-LL,cl.}} = \left(E_{\mathrm{HL,cl.}} - E_{\mathrm{LL,cl.}}\right), \tag{23}$$

where "LL" and "HL" stand for "low-level" and "high-level", and "per." and "cl." for "periodic" and "finite cluster", respectively. The cluster can be optionally embedded. The target quantity E is usually an interaction energy per some structure unit: for example, cohesive energy per atom, lattice energy per molecule, adsorption energy per adsorbate molecule, etc. Sometimes the total energy per cell is also considered.

The most common choices for the low-level treatment are periodic HF or DFT. One of the reasons is that both HF and DFT capture reasonably well the electrostatic interactions, which are indeed very long-range and do require the correct periodic treatment. Besides, these methods are computationally inexpensive, especially so non-hybrid DFT, and are implemented in publicly available quantum chemical periodic program packages. A correction to periodic HF or DFT is expected to be of a less long-range character and thus a finite-cluster treatment could be sufficient. For example, the widely used incremental scheme[11,17] of Stoll and coworkers usually adds a CCSD(T) correlation energy, computed via finite-cluster many-body expansion, to the periodic HF [*i.e.* LL=HF, HL=CCSD(T)]. Another well known scheme of Sauer and coworkers[111,113,114] employs periodic DFT and a finite cluster MP2-DFT correction [*i.e.* LL=DFT, HL=MP2]. As a final refinement a CCSD(T)-MP2 energy difference calculated on a smaller cluster can be added.

The embedding hierarchy for correlated methods, presented in this work, offers a very convenient platform for the correcting scheme (22) of a much higher level, as it permits for

more accurate models both at LL and HL stages. Firstly, LL is in this work increased to the correlated level. Secondly, by employing the LCCSD(T)|LCCD[S]-R^{-6} method for HL, which is applicable to large systems, it becomes possible to control convergence of the results with the cluster size.

We start with the periodic LMP2 method as LL. In contrast to HF and standard DFT, MP2 does capture dispersion, which means that the HL in this case provides only a correction to the latter, rather than the whole effect. With that one can expect that in this case the $\delta E_{\mathrm{HL-LL,cl.}}$ converges with the cluster size much faster, than when LL is HF or DFT. For a certain class of systems this is indeed the case, *e.g.* for small not very polarizable molecules forming a crystal or adsorbed on a surface, for which MP2 itself is not too far off. The correction on top of periodic LMP2 is then much smaller than the overall correlation energy contribution and can be captured already with very small cluster models.

For example, for the CO2 or HCN crystals[25] the standard incremental scheme converges quite slowly with the cluster size, which reflects the slow R_{cut}^{-3} decay rate of the unaccounted dispersion (*vide supra*). Furthermore, in the HCN crystal the electrostatics plays an important role (the individual molecules have a strong dipole moment), which cannot be adequately captured by a finite-cluster representation. Here the periodic HF seems to be insufficient for accurate electrostatic interaction energy, making a scheme with LL=HF too crude. At the same time, with LL=LMP2, the incremental scheme converges already within a few most compact clusters to a result, which agrees very well with the experimental one.

In highly polarizable systems, however, MP2 notoriously overestimates dispersion. In such cases the post-MP2 correction is not small. This means that for such systems, MP2 as the LL method offers very little or even no advantage with respect to LL=HF. One of the seemingly easy ways to correct this problem is to use the empirically corrected spin-component-scaled (SCS) MP2. It can be evaluated within an MP2 calculation without any additional cost, and for dispersively bound dimers of highly polarizable molecules provides a better interaction energy than MP2. Unfortunately, SCS-MP2 cannot cure the MP2 problem of the unscreened description of the Coulomb interaction (*vide infra*). The screening of the Coulomb interaction is very efficient in small-gap bulk systems and could thus have a profound effect on dispersion.

A third LL method we propose in this work is LdrCCD. It is in many cases less accurate than MP2 or SCS-MP2, but its error is usually less dependent on the polarizability and, very importantly, it does include the Coulomb screening via the Coulomb-ring diagrams.[110,136–139] Presently, the periodic version of LdrCCD is not yet available. However by progressively

expanding the embedded fragment one can closely approach the converged periodic result.

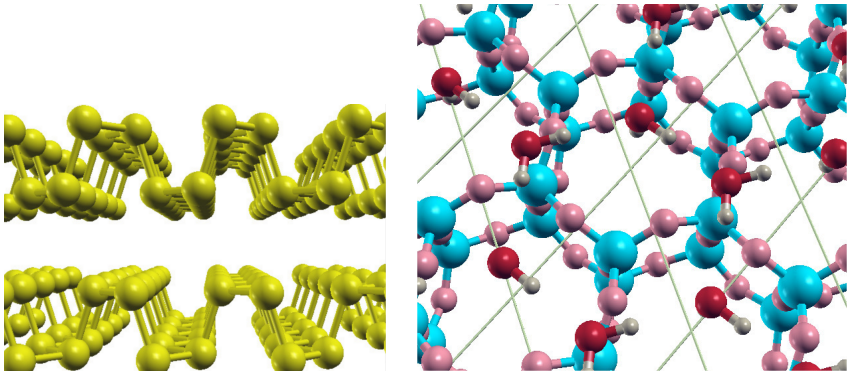
We note that in practical applications, high accuracy can be achieved only if other errors, in addition to the method error discussed so far, are also corrected. This can include the basis set incompleteness error, frozen core approximation error, local approximation error, CCSD(T) method error, etc.[25,26,115–117,140,141] In this respect the $\delta E_{\mathrm{HL-LL,cl.}}$ correction could actually consist of several components, evaluated with different clusters. It is possible to obtain the LMP2 result close to the basis set limit (using LMP2-F12[39]) as well as the core correlation at the LMP2 level still remaining in the periodic regime.[115] The basis set correction is, however, of a relatively short-range, provided the periodic LMP2 basis set is sufficiently good, say, of augmented-triple-zeta level. The interactions due the core electron are usually non-negligible also only at short distances. Therefore the corresponding corrections to a good accuracy could also be calculated using moderately sized clusters.

In this paper we will not discuss other corrections apart from the method error correction. For an overview we refer to Refs.[116,141], where extended correcting protocols are presented and discussed in detail. Here we only focus on the scheme (22) with the LL=MP2, SCS-MP2 or LdrCCD, carried out using the periodic (sect. 2.2) or embedded-fragment models (sect. 2.3), and HL=LCCSD(T)|LCCD[S]-R^{-6} with large clusters and the unit-cell-in-cluster approach.

3 Calculations and discussion

In this work we apply the presented methodology to investigate the importance of the Coulomb screening effects for van-der-Waals interactions in solids and to which extent that imposes a challenge for the theory. The term "screening" is often associated with weakening of the Coulomb interaction between a pair of electrons, holes or an electron and a hole due to the influence of other electrons and/or holes through mutually-coupled correlated Coulomb interactions. It is essentially a many-body effect and can formally be represented by the ring diagrams. Therefore the direct random-phase approximation (RPA), which sums up the cascades of these diagrams to infinity,[136,139] captures exactly this effect. CCSD also includes these diagrams but in addition many other types of correlated interactions. The LdrCCD model is based on the direct RPA,[121] but adds to it a few exchange-type diagrams.[58,110] The MP2 method does not contain Coulomb-ring diagrams and thus describes unscreened correlated Coulomb interactions only.

In the molecular context it is not common to study screening as a separate effect. In



(a) (b)

Figure 2: The systems studied in this work: phosphorene bilayer (side view, panel a), and a water monolayer adsorbed on 2D silica (top view, panel b). The phosphorus atoms are given in yellow color, the silicon atoms in blue, the water's oxygen in red, the silica's oxygens in pink, the hydrogens in grey. In the 2D silica – water system, half of the water molecules are located over the silicon atoms, while the other half over the centers of the silica's 6-membered rings. The green lines indicate the unit cells.

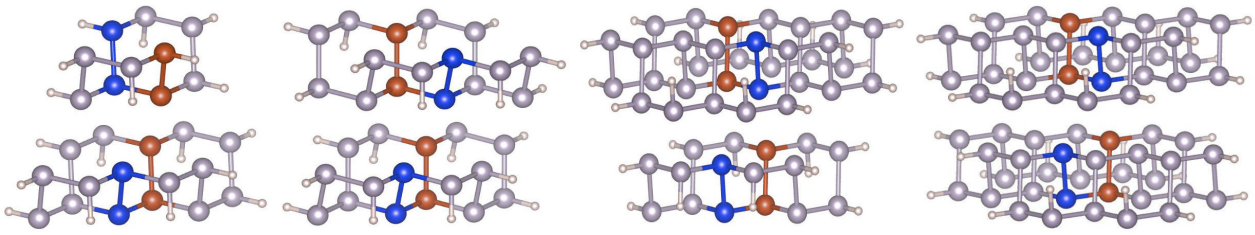
solids, on the contrary, screening plays an essential role, for example, in exciton binding and other contexts, and is thus a subject of extensive investigation and modeling. However, the solid state theory only since recently has started focusing on van der Waals dispersion interactions systematically (*cf.* *e.g.* Refs.[2,3,14,21–26,104,115–117,124,125,141–158]) and many aspects of this effect including the influence of the screening are currently under active research.

3.1 Specification of the systems

The effectiveness of the HL/LL protocol (Section 2.6), based on the embedding models of sections (2.2-2.4) and the role of screening is studied on the interaction between the phosphorene sheets in a phosphorene bilayer (Fig. 2a) and adsorption of water monolayer on 2D-silica (Fig. 2b). Both black phosphorus,[90,141,155,159–171] which is a small gap semiconductor, and 2D silica,[172–180] an insulator, are promising systems from the technological point of view and are extensively studied experimentally and theoretically.

The structure of phosphorene bilayer was taken from Ref.[90] (p-LMP2/TZVPP2 entry of table S2), which is quite close to the experimental one. As mentioned, we will focus on evaluating the interaction energy per phosphorus atom, given by:

$$\Delta E = \frac{1}{8}\left(E_{\text{bilayer}} - E_{\text{layer 1}} - E_{\text{layer 2}}\right) \tag{24}$$



$P_{10}H_8$–$P_{16}H_{10}$ $P_{16}H_{10}$–$P_{16}H_{10}$ $P_{16}H_{10}$–$P_{32}H_{14}$ $P_{32}H_{14}$–$P_{32}H_{14}$

Figure 3: The cluster models used for the phosphorene bilayer. The grey, blue and red atoms are phosphorus, and yellowish – hydrogens. The red and blue atoms denote the unit cell atoms of type A and B, respectively (*cf.* sect. 2.5).

where 8 is the number of atoms in a unit cell for the bilayer.

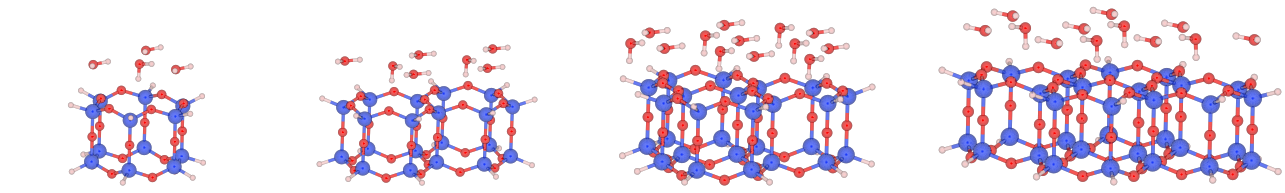
In the case of 2D silica, we are interested in evaluating the interaction energy of water molecules on the substrate; unfortunately, it is experimentally not known, in which particular structure water adsorbs on 2D silica. A B3LYP study of Ref.[172] proposed a water monolayer structure with two water molecules per silica unit cell both located above surface Si atoms. In our own optimization (at the B3LYP-D3 level with the POB-TZVP basis set[181]) we found a structure with one of the water molecules located above the center of the ring (Fig. 2b). According to a periodic LMP2 calculation this geometry is energetically preferable by around 1 kcal/mol per water molecule. We note, however, that the formation of the water layer is driven by the competition between the water-silica and relatively strong water-water interactions. Therefore the actual structure of the adsorbate is likely to show a strong dependence on the water coverage, whose detailed study goes beyond our present purposes. In this study we consider the structure of Fig. 2b as a model structure rather than a physical one, aiming to test the accuracy of different computational approaches for such a system. Water within this structure forms a hexagonal hydrogen-bond network. The interaction energy per water molecule for the 2D silica – water system is defined as:

$$\Delta E = \frac{1}{2}\left(E_{\text{silica}-\text{water}} - E_{\text{silica}} - E_{\text{water monolayer}}\right) \tag{25}$$

and within the water monolayer itself:

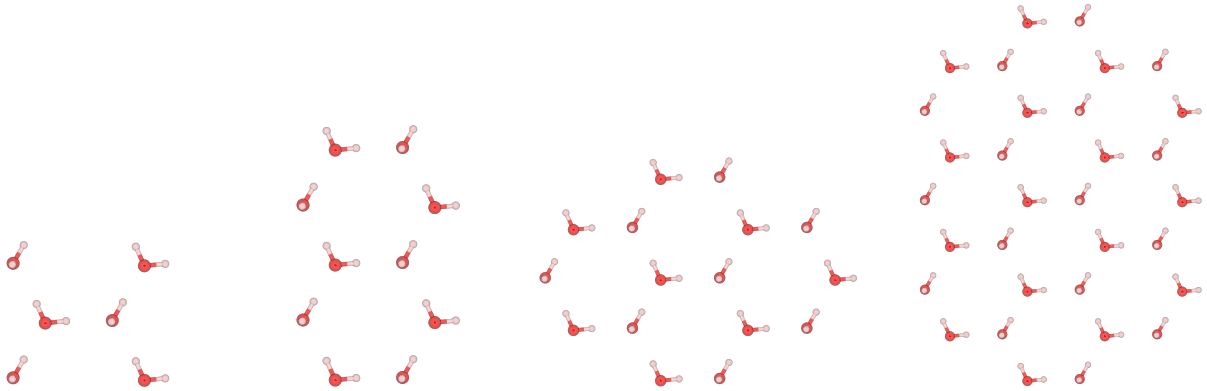
$$\Delta E = \frac{1}{2}\left(E_{\text{monolayer}} - E_{\text{molecule 1}} - E_{\text{molecule 2}}\right). \tag{26}$$

The energies E in eqs. (24)-(26), apart from E_{molecule}, can refer to the total (HF+correlation) or correlation energy per unit cell. In all these expressions the relaxation of the monomers is not taken into account, and the interaction energies are counterpoise corrected. This means



$Si_{12}O_{18}H_{12}$–$(H_2O)_4$ $Si_{20}O_{32}H_{16}$–$(H_2O)_7$ $Si_{26}O_{43}H_{18}$–$(H_2O)_{13}$ $Si_{32}O_{54}H_{20}$–$(H_2O)_{12}$

(a)



$(H_2O)_6$ $(H_2O)_{10}$ $(H_2O)_{16}$ $(H_2O)_{32}$

(b)

Figure 4: The cluster models used for the 2D silica – water system (a), and water monolayer (b). The silica atoms are shown in blue, oxygen in red and hydrogen in pink.

that the isolated phosphorene sheets in (24), 2D silica and the water monolayer in (25) and the water molecules 1 and 2 in (26) are taken in the same geometry as in the corresponding interacting systems and with the presence of the other monomer [or monomers in case of the lateral interaction energy (26)] in a form of ghost atoms, holding the basis functions.

As the unit cell and the energy per unit cell are defined for the fragments and finite clusters, the corresponding interaction energies were also computed using eqs. (24)-(26). For the phosphorene bilayer and silica-water system the embedded fragments were obtained according to the procedure described in sect. 2.5 with $R_{\mathrm{cut}}^{\mathrm{intra}}$ varying from 1 bohr to 12 bohr. For the water monolayer a global R_{cut} was used ranging from 6 bohr to 20 bohr.

For each system 4 different clusters of progressively increasing size were used. They are depicted in Fig. 3 (phosphorene bilayer) and Fig. 4 (water – 2D silica and water monolayer). They were cut out from the periodic structures with the dangling bonds (in the phosphorene bilayer and silica) saturated with hydrogens: with the P–H and Si–H distance of 1.42 and 1.48 Å, respectively. For the water molecules no bonds were cut. The four silica clusters were chosen such that they contained one, two, three and four 6-membered rings, respectively (*cf.* Fig. 4a).

All the geometries are explicitly given in the supplementary information[182] in a form of the CRYSTAL input (for the periodic structures) or xyz-files (for the finite clusters).

3.2 Computational parameters

The periodic HF calculations and subsequent generation of WFs were performed using the CRYSTAL code.[134] Periodic local MP2 calculations and preparation of the ERIs and other intermediates for the fragment calculations were done with the CRYSCOR code.[35] For the fragment-based and finite-cluster calculations the local correlation program of MOLPRO[52,59,63,110,118] was employed.

Basis sets of augmented-triple-zeta quality were used in the periodic, fragment and most of the finite-cluster calculations. For phosphorene bilayer at the periodic HF stage it was a TZVPP2 basis set from Ref.[90], which was then augmented by diffuse d- and f-orbitals from aug-cc-pVTZ[183] for the correlated periodic and embedded-fragment calculations via the dual basis set technique.[66] In the silica-water system, the periodic and fragment calculations employed the basis consisting of s- and p- (only s- for hydrogen) orbitals from def2-TZVPP[184] and d-and f- (p- and d- for hydrogen) orbitals from cc-pVTZ[185]. Again the at the correlated level, the diffuse d- and f- orbitals from aug-cc-pVTZ were added for O and Si. Some of the exponents for the silicon atoms were upscaled to avoid HF convergence problems. The basis

sets are given in the supplementary information[182] in a form of the CRYSTAL input.

The finite cluster calculations used mostly aug-cc-pVTZ[183] (cc-pVTZ for H) basis sets. The calculations for the largest phosphorene bilayer cluster ($P_{32}H_{14}$–$P_{32}H_{14}$), largest water cluster ($(H_2O)_{32}$) and two largest silica-water clusters ($Si_{26}O_{43}H_{18}$–$(H_2O)_{13}$ and $Si_{32}O_{54}H_{20}$–$(H_2O)_{12}$) were performed with the aug-cc-pVDZ basis, and the aug-cc-pVTZ value was approximated by adding the corresponding aug-cc-pVTZ − aug-cc-pVDZ energy difference, evaluated on the second largest (phosphorene bilayer and water monolayer) or third largest (silica–water) clusters.

The periodic HF calculations were carried out with the TOLINTEG[186] tolerances of 10 10 10 25 75. The Brillouin zone was sampled with an 12×12 k-mesh. In the periodic, fragment and finite-cluster local correlation calculations the virtual space was restricted to the core PAO domains. In order to identify the lone-pair and bonding LMOs in the finite-cluster calculations for the automatic unit cell determination (*cf.* sect. 2.5) the Boughton-Pulay criterion, used for the definition of such core domains, had to be adjusted to 0.97 – 0.975 (depending on the cluster). Since in this paper we focus on the method correction within the HL/LL correcting scheme, the domain- and basis-set errors were not explored.

In periodic LMP2, the pairs up to 12 Å were explicitly processed, beyond that the corresponding contribution was obtained using the $C_6^{ij}/R^6_{ij\mathcal{J}}$ extrapolation[65] [*cf.* eq. (6)]. In the finite cluster calculations all the inter-monomer pairs were treated as close (*cf.* sect. 2.4). The intra-pairs were ranked as the strong pairs (one common atom in the corresponding domains), close pairs (one connecting bond between the atoms in the corresponding domains) and week pairs (the rest). As was explained in sect. 2.4, in the LCCSD(T) method the strong pairs enjoy the full LCCSD(T) treatment, the close pairs the LCCD[S](T) one, and the weak pairs the LrCCD3 treatment.[58,59] For the perturbative triples the L(T0) approximation[63] was employed. The ERIs were approximated using the density fitting technique with Weigend's aug-cc-pVTZ fitting basis sets.[187] The exceptions were the periodic HF, where the ERIs were computed exactly or via the multipole approximation[188], and the periodic LMP2 ERIs for the pairs between 8 and 12 Å, evaluated by the multipole approximation.[65]

3.3 Results and discussion

3.3.1 Phosphorene bilayer

We start the discussion of the results with the phosphorene bilayer. Figure 5a compiles the interaction energies (24), computed using different schemes, ranging from periodic LMP2 to

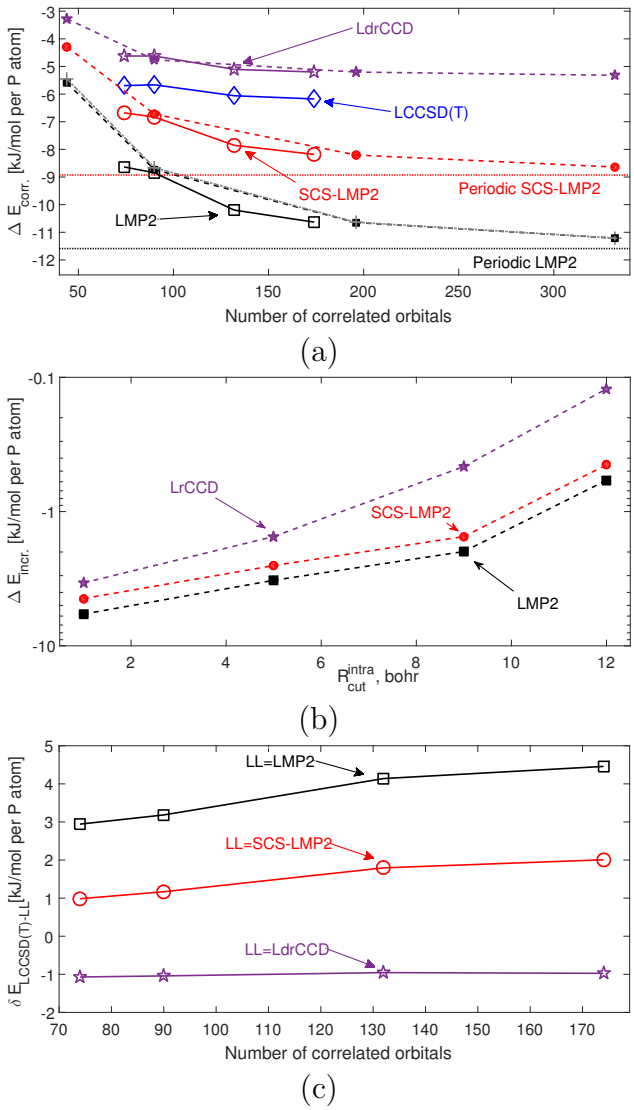


Figure 5: Correlation interaction energies $\Delta E_{\text{corr.}}$(a), the increments thereof $\Delta E_{\text{incr.}}$ by the fragment expansion (logarithmic scale) (b), and HL–LL corrections $\delta E_{\text{HL}-\text{LL}}$ (c) for the phosphorene bilayer as functions of fragment's or cluster's size. $\Delta E_{\text{corr.}}$ was computed via eq. (24). Here the "number of correlated orbitals" directly reflects the size of the fragment or cluster. $\Delta E_{\text{incr.}}$ are the correlation interaction energy increments, obtained within stepwise extensions of the fragments by increasing $R_{\text{cut}}^{\text{intra}}$. The corrections $\delta E_{\text{HL}-\text{LL}}$ were defined by eq. (23) with LL=LMP2, SCS-LMP2 and LrdCCD, and HL=LCCSD(T0)|LCCD[S]-R^{-6}. The periodic LMP2 and SCS-LMP2 results are shown by black and red dotted lines, respectively. The fragment results are given with dashed lines and filled symbols, while the finite-cluster ones with solid lines and open symbols: LMP2 – squares, SCS-LMP2 – circles, LrdCCD – asterisks, LCCSD(T0)|LCCD[S]-R^{-6} – diamonds. The fragment LMP2 and SCS-LMP2 were computed in the periodic-fragment model (*cf.* sect. 2.2), while LdrCCD in the embedded-fragment model (*cf.* sect. 2.3). The embedded-fragment LMP2 results are also given in the panel (a) by the grey dashed line and crosses. In all panels the symbols indicate the actually computed values.

finite-cluster LCCSD(T). Firstly, we note that the fragment-based LMP2 and SCS-LMP2 energies smoothly approach the periodic values. Furthermore the LMP2 energies, obtained using the periodic-fragment model of sect. 2.2 and embedded-fragment model of sect. 2.3 virtually coincide: at the scale of Fig. 5a the deviation between the grey dashed curve with crosses (embedded-fragment LMP2) and the black dashed curve with filled squares (periodic fragment LMP2) can be seen only for the smallest cluster. That suggests that the embedded-fragment model is a very good approximation to the periodic treatment at least at the LMP2 level. The LdrCCD couplings are stronger and of longer range than in LMP2, and thus the convergence of the embedded-fragment LdrCCD to the periodic result could be slower than for LMP2. We don't have the periodic or periodic-fragment LdrCCD implementation ready yet, so the direct comparison to those is not possible. However, from the convergence pattern of the embedded-fragment LdrCCD in Fig. 5a one can conclude that, in fact, it converges even faster than LMP2.

This faster convergence of LdrCCD with respect to LMP2 and SCS-LMP2 is a very important observation, because this is exactly how the Coulomb screening is supposed to manifest itself: for distant electron pairs, the dispersive interaction becomes progressively less efficient due to the influence (screening) of the electrons in between. That leads to a faster decay of dispersion with the pair distance compared to LMP2 as well as SCS-LMP2.

SCS-LMP2 uses the very same LMP2 amplitudes, which are then scaled with a uniform factor. This scaling obviously has an influence on the long range interaction as well as on the short range one, but it cannot correctly reproduce the screening effect. This becomes evident in the logarithmic plot of the gain in the interaction energy by expansion of the fragment (Fig. 5b). LMP2 and SCS-LMP2 contributions have a very similar decay rate, while LdrCCD curve differs from them quite substantially, revealing a noticeable contraction of the effective range of dispersive interactions.

Now we analyze the results of the finite-cluster model, where the methodological level can be increased further up to LCCSD(T). We note, firstly, that the finite-cluster LMP2, SCS-LMP2 and LdrCCD correlation interaction energies are quite close to those of the fragment approach and the principal pattern of convergence to the periodic result is reproduced. Importantly, the finite curves demonstrate the same distinction in the decay rate between LdrCCD on the one hand and LMP2 or SCS-LMP2 on the other.

With finite clusters we can finally compare these curves with the LCCSD(T) reference. This comparison reveals, that although LdrCCD interaction energy is not very accurate, the LCCSD(T) curve is virtually parallel to the LdrCCD one. This indicates that the electron

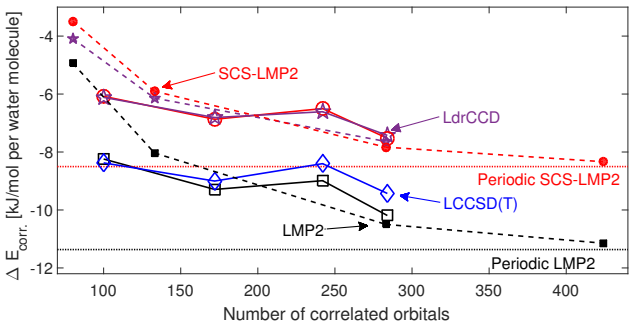
screening is the most important effect in the long range, at least for this system, shaping the strength of the distant-pair interaction. Indeed, the difference between LdRCCD and LCCSD(T) is virtually independent of the cluster size, as follows from $\delta E_{\mathrm{LCCSD(T)-LdrCCD}}$ in Fig. 5c, suggesting that it originates mainly from the short and medium range. Therefore LdRCCD can be a very efficient choice for LL, as its correction to a very good accuracy can be evaluated on a small cluster. At the same time, the $\delta E_{\mathrm{LCCSD(T)-LL}}$ correction to LMP2 or SCS-LMP2 converges very slowly - not faster than the individual LMP2 or SCS-LMP2 energies themselves. This makes these methods absolutely useless as LL for this system: the choice LL=HF would be as efficient.

To summarize, for systems of such type both periodic LMP2 and SCS-LMP2 can hardly be useful. The reason is not only the high polarizability of the monomers, as SCS-LMP2 can cope with that. The main problem comes from the high effectiveness of the Coulomb screening, which weakens the long-range part of the dispersion substantially. The unscreened LMP2 or SCS-LMP2 long-range correlation provides then a lot of fictitious attractive interaction, causing a gross overbinding in the thermodynamic limit. This overestimation of the correlation contribution to the interaction energy leads to even more pronounced relative error in the total interaction energy, in cases when the HF contribution is repulsive, as, for example, in the phosphorene bilayer, where it amounts to around +4 kJ/mol per P atom.[141] To conclude, it is highly advisable for such systems to choose the LL treatment such, that it captures the screening effect, *e.g.* LdrCCD.

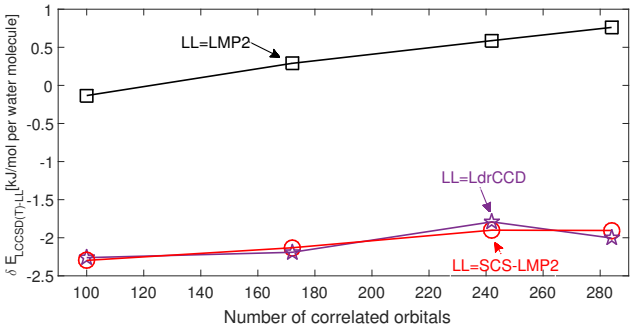
3.3.2 Water–silica adsorption

Next we consider van der Waals interactions in a system with a larger band gap: in water adsorbed on 2D silica, the interacting species are all insulators. For the interaction energy between the water monolayer and silica, LMP2 is no longer as faulty as for phosphorene bilayer. Indeed, as can be seen from Fig. 6a, which compiles the correlation adsorption energies for this system, for a small cluster it is even very close to LCCSD(T). However, despite the wide gap, the screening effects still show up, as the LMP2 interaction grows with the system size noticeably faster than that of LdrCCD or LCCSD(T). In the SCS-LMP2 this effect can also be observed in the fragment calculations, but the uniform downscaling of the pair energies in this method brings it close to LdrCCD.

Due to absence of screening in LMP2, the $\delta E_{\mathrm{LCCSD(T)-LMP2}}$ correction again converges slowly with the cluster size (*cf.* Fig. 6b), making the HL/LL scheme with choice of LL=LMP2 inefficient. LMP2 for this system could in principle be used alone without the HL



(a)



(b)

Figure 6: Correlation interaction energies $\Delta E_{\text{corr.}}$(a) and HL–LL corrections $\delta E_{\text{HL}-\text{LL}}$ (b) for the water monolayer adsorbed on 2D silica as functions of fragment's or cluster's size. $\Delta E_{\text{corr.}}$ was computed via eq. (25). The corrections $\delta E_{\text{HL}-\text{LL}}$ are defined by eq. (23) with LL=LMP2, SCS-LMP2 and LrdCCD, and HL=LCCSD(T0)|LCCD[S]-R^{-6}. The periodic LMP2 and SCS-LMP2 results are shown by black and red dotted lines, respectively. The fragment results are given with dashed lines and filled symbols, while the finite-cluster ones with solid lines and open symbols: LMP2 – squares, SCS-LMP2 – circles, LrdCCD – asterisks, LCCSD(T0)|LCCD[S]-R^{-6} – diamonds. The fragment LMP2 and SCS-LMP2 values were computed in the periodic-fragment model (*cf.* sect. 2.2), while LdrCCD in the embedded-fragment model (*cf.* sect. 2.3). The symbols indicate the actually calculated values. The "number of correlated orbitals" directly reflects the size of the fragment or cluster.

correction as its error at the end does not seem to exceed 2 kJ/mol per adsorbate molecule, which is in many cases an acceptable accuracy for such an inexpensive method as MP2.

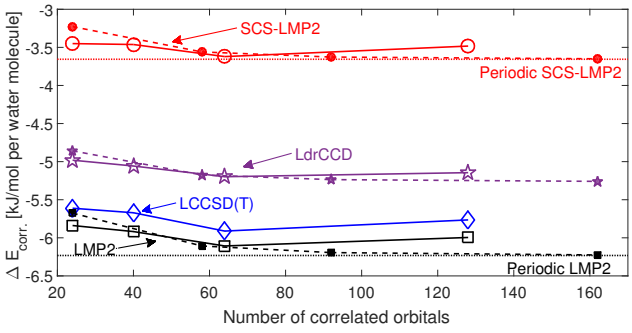
In contrast to LMP2, LdrCCD behaves very similarly to the phosphorene bilayer case. It again noticeably underestimates the binding energy. But at the same time, the convergence of the $\delta E_{\mathrm{LCCSD(T)-LdrCCD}}$ correction is quite insensitive to the fragment size, again suggesting that LdrCCD is an appropriate model for LL in the HL/LL scheme.

Finally we focus on the interactions inside the water monolayer. The results for the lateral interaction energy are compiled in Fig. 7. LMP2 is very accurate here, which is in agreement with a commonly known success of this method for hydrogen-bonded intermolecular complexes. Screening effects are very small here, so the LMP2 is accurate for the whole range of studied clusters. LdrCCD is less accurate than LMP2: it again underestimates the interaction energy. However as an LL method for HL/LL scheme it is again perfectly suitable. The worst method for this system among studied here is apparently SCS-LMP2.

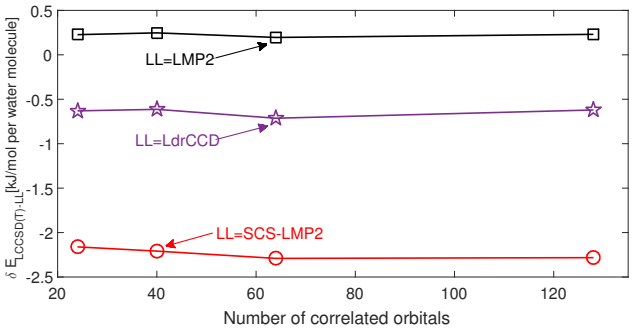
We also note that the finite cluster interaction energy does not entirely follow the convergence pattern of the fragment approach both in the case of water-silica and water-water interactions. Moreover the finite cluster results, in the water-water interaction case, does not seem to converge to the periodic result at all. We attribute this to the influence of the electrostatics. In the embedded-fragment approach the correct electrostatics is incorporated in the periodic HF solution and appears in the correlated formalism (7) or (14) via the correct periodic Fock matrix. The finite-cluster electrostatics is however always influenced by the form of the cluster and border effects. A similar behavior was observed in periodic vs finite cluster correlated calculations on the HCN crystal (*cf.* Ref.[25] and discussion in sect. 2.6). At the same time, and again in close analogy to the HCN case, the HL–LL correction is virtually immune to the border effects, as it depends on the choice of the cluster only very slightly (*cf.* Fig. 7b).

4 CONCLUSIONS

In this work we presented three local models to calculate the dynamic correlation in periodic systems: (i) periodic fragment, (ii) embedded fragment and (iii) finite cluster. The main features of these models are summarized in Fig. 8. All three models incorporate a concept of unit cell, despite the lack of translational symmetry in the latter two. By expansion of the fragment or cluster the corresponding "energy per unit cell" in all three cases converges to the periodic one, even though electrostatic effects can be somewhat troublesome in the



(a)



(b)

Figure 7: Correlation interaction energies $\Delta E_{\mathrm{corr.}}$(a) and HL–LL corrections $\delta E_{\mathrm{HL-LL}}$ (b) for the water monolayer as functions of fragment's or cluster's size. $\Delta E_{\mathrm{corr.}}$ was computed via eq. (26). The corrections $\delta E_{\mathrm{HL-LL}}$ are defined by eq. (23) with LL=LMP2, SCS-LMP2 and LrdCCD, and HL=LCCSD(T0)|LCCD[S]-R^{-6}. The periodic LMP2 and SCS-LMP2 results are shown by black and red dotted lines, respectively. The fragment results are given with dashed lines and filled symbols, while the finite-cluster ones with solid lines and open symbols: LMP2 – squares, SCS-LMP2 – circles, LrdCCD – asterisks, LCCSD(T0)|LCCD[S]-R^{-6} – diamonds. The fragment LMP2 and SCS-LMP2 values were computed in the periodic-fragment model (*cf.* sect. 2.2), while LdrCCD in the embedded-fragment model (*cf.* sect. 2.3). The symbols indicate the actually calculated values.

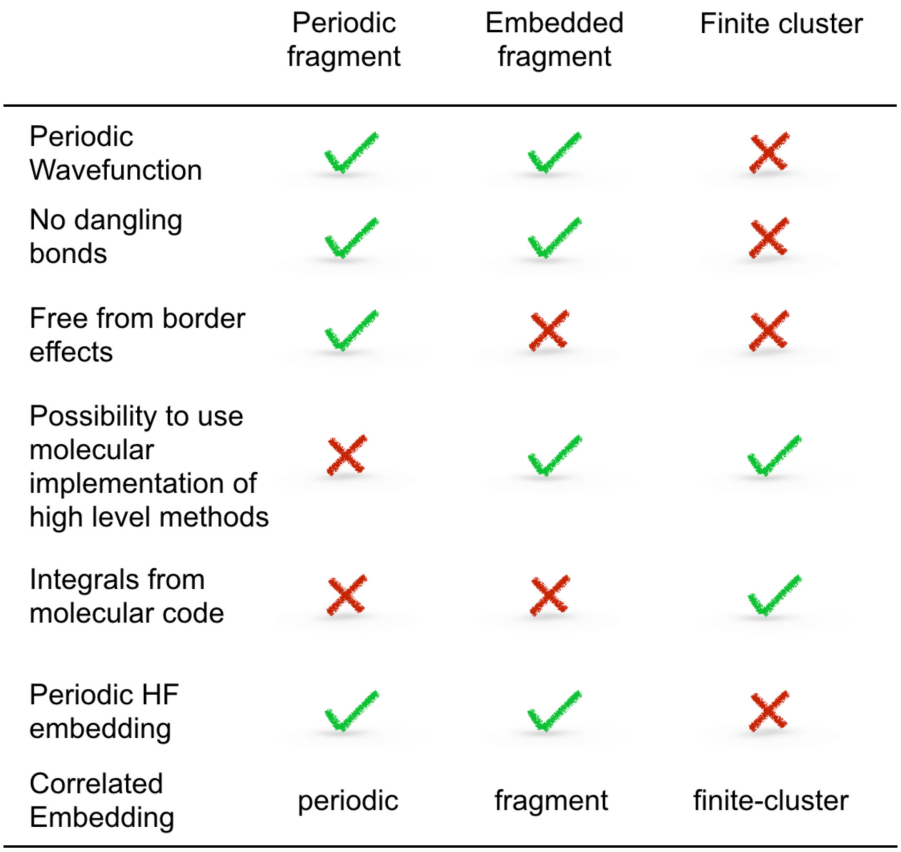


Figure 8: Comparison of the three embedding models, presented in the sections 2.2, 2.3 and 2.4, respectively.

finite cluster case.

The periodic fragment model requires the full periodic implementation and thus is presently restricted to the LMP2 (or SCS-LMP2) level of treatment. The embedded cluster model uses the periodic HF solution, periodic occupied and virtual orbitals and ERIs computed in the periodic framework, but the actual correlated calculation is carried out by a molecular program. This model can be presently used for LdrCCD calculations. For the finite cluster model one can employ any of the implemented local correlation techniques, in particular the LCCSD(T)|LCCD[S]-R^{-6} one, which provides close to CCSD(T) accuracy for large systems, including van-der-Waals-bound complexes.[59]

We use the embedding models in conjunction with the hierarchical high-level/low-level (HL/LL) scheme, that allows one to effectively reach a very high level of description for complex systems, in our case solids. For the LL treatment we tested periodic LMP2, SCS-LMP2, or embedded-fragment LdrCCD, while for the HL treatment the finite-cluster LCCSD(T) model. Importantly, due to the concept of a unit cell in fragment or cluster, and the local

correlation scheme, it becomes possible to systematically converge the interaction energies to the thermodynamic limit.

For a comparative testing of the three approaches we have chosen three prototypical periodic systems where binding is governed by the van der Waals interactions: (i) a bilayer of phosphorene, a small gap semiconductor, (ii) water monolayer adsorbed on 2D silica, an insulator, and (iii) the water monolayer a 2D hexagonal arrangement of molecules. These examples demonstrate that the reliability of periodic LMP2 as well as SCS-LMP2 strongly depends on how polarizable the system is and how much the Coulomb interactions are screened.

For the phosphorene bilayer, MP2 grossly overestimates the interaction energy even for small clusters, which becomes much more pronounced in the thermodynamic limit. SCS-LMP2 for a small cluster is seemingly accurate, but the lack of screening effects leads to substantial growth of SCS-LMP2 error with expansion of the cluster. This reveals that both LMP2 and SCS-LMP2 for systems like a phosphorene bilayer neither can provide reliable results on their own, nor can be an efficient LL method within the HL/LL scheme, since the error (*i.e.* HL–LL energy difference) depends severely on the size of the finite cluster and thus converges very slowly.

For water on silica, LMP2 is much better: for a small cluster it is even very close to LCCSD(T). However again due to the inability to capture the screening effects, the LMP2 error grows with the cluster size. The screening here is weaker than in the phosphorene bilayer and the final error of the LMP2 in the thermodynamic limit can be considered as acceptable (1-2 kJ/mol per adsorbate molecule). This suggests that LMP2 alone as a computational tool could be used for such systems, but not when aiming at very high accuracy within the HL/LL scheme, as the HL–LL correction again would converge slowly with the cluster size. Finally, for the water monolayer LMP2 is quite accurate on its own, and can further be very effectively corrected via the HL/LL scheme already on small clusters, as the screening effects are negligible here. On the other hand, the dipolar nature of water reveals a shortcoming of the finite cluster approach, that is how to deal with electrostatics at the cluster borders when aiming at the total or correlation interaction energy. While this problem might in principle be attenuated by embedding the cluster in a suitable point charge surrounding such procedure can be nontrivial and error-prone.

In contrast to LMP2, LdrCCD demonstrates a very systematic behavior, at least for these three systems: Regardless of the band gap and type of the system, LdrCCD always somewhat underestimates the interaction. But most importantly, it captures very accurately

the long-range part of dispersion, whether there is screening or not. This makes the HL–LL correction with LL=LdrCCD converging very fast with the clusters size. So although LdrCCD is not a sufficiently accurate method on its own, it seems a very promising candidate for the LL model, as its error can be virtually completely eliminated by the HL/LL scheme using already quite compact clusters.

In conclusion, while highly reliable and efficient periodic high-level methods and not yet mature, a HL/LL fragment combination approach as outlined in this work allows for studying the details of Coulomb screening, cluster size and shape, correlation level and so on, in a robust and consistent way, by synergically combining the best of available tools in periodic and molecular codes. Furthermore, it provides a powerful protocol for high precision ab initio calculations of interaction energies in periodic systems. The local correlation approach is the backbone of the whole scheme, as it allows to identify the fragments, recognize the same "unit cell" entities across the different models, and make high-level quantum chemical calculations feasible for extended systems.

Acknowledgments

The authors would like to thank Daniel Kats, Alex Tkatchenko, Luke Shulenburger and David Tomanek for useful discussions. Financial support form the Deutsche Forschungsgemeinschaft DFG (Grants Schu 1456/12-1 and US 103/1-2) is gratefully acknowledged. The graphical representation of the studied systems was created using programs Vesta[189] and Xcrysden.[190]

References

1. M. G. Medvedev, I. S. Bushmarinov, J. Sun, J. P. Perdew, and K. A. Lyssenko, Science **355**, 49 (2017).

2. M. Halo, C. Pisani, L. Maschio, S. Casassa, M. Schütz, and D. Usvyat, Phys. Rev. B **83**, 035117 (2011).

3. A. J. Karttunen, D. Usvyat, M. Schütz, and L. Maschio, Phys. Chem. Chem. Phys. **19**, 7699 (2017).

4. G. J. O. Beran, Chem. Rev. **116**, 5567 (2016).

5. J. Harl and G. Kresse, Phys. Rev. Lett. **103**, 056401 (2009).

6. A. Grüneis, M. Marsman, J. Harl, L. Schimka, and G. Kresse, J. Chem. Phys. **131**, 154115 (2009).

7. S. Grimme, J. Comput. Chem. **27**, 1787 (2006).

8. S. Grimme, J. Antony, S. Ehrlich, and H. Krieg, J. Chem. Phys. **132**, 154104 (2010).

9. A. Tkatchenko and M. Scheffler, Phys. Rev. Lett. **102**, 073005 (2009), URL http://link.aps.org/doi/10.1103/PhysRevLett.102.073005.

10. A. Tkatchenko, R. A. DiStasio, R. Car, and M. Scheffler, Phys. Rev. Lett. **108**, 236402 (2012).

11. H. Stoll, Phys. Rev. B **46**, 6700 (1992).

12. H. Stoll, Chem. Phys. Lett. **191**, 548 (1992).

13. H. Stoll, J. Chem. Phys. **97**, 8449 (1992).

14. K. Rościszewski, B. Paulus, P. Fulde, and H. Stoll, Phys. Rev. B **60**, 7905 (1999).

15. K. Rosciszewski, K. Doll, B. Paulus, P. Fulde, and H. Stoll, Phys. Rev. B **57**, 14667 (1998).

16. B. Paulus, K. Rosciszewski, N. Gaston, P. Schwerdtfeger, and H. Stoll, Phys. Rev. B **70**, 165106 (2004).

17. B. Paulus, Phys. Rep. **428**, 1 (2006).

18. E. Voloshina and B. Paulus, J. Chem. Phys. **124**, 234711 (2006).

19. N. Gaston, B. Paulus, K. Rościszewski, P. Schwerdtfeger, and H. Stoll, Phys. Rev. B **74**, 094102 (2006).

20. E. Voloshina and B. Paulus, Phys. Rev. B **75**, 245117 (2007).

21. A. Hermann and P. Schwerdtfeger, Phys. Rev. Lett. **101**, 183005 (2008).

22. R. Podeszwa, B. M. Rice, and K. Szalewicz, Phys. Rev. Lett. **101**, 115503 (2008).

23. P. Schwerdtfeger, B. Assadollahzadeh, and A. Hermann, Phys. Rev. B **82**, 205111 (2010).

24. O. Sode, M. Keceli, K. Yagi, and S. Hirata, J. Chem. Phys. **138**, 074501 (2013).

25. C. Müller and D. Usvyat, J. Chem. Theory Comput. **9**, 5590 (2013).

26. J. Yang, W. Hu, D. Usvyat, D. Matthews, M. Schütz, and G. K.-L. Chan, Science **345**, 6197 (2014).

27. P. Y. Ayala, K. N. Kudin, and G. E. Scuseria, J. Chem. Phys. **115**, 9698 (2001).

28. C. Pisani, M. Busso, G. Capecchi, S. Casassa, R. Dovesi, L. Maschio, C. Zicovich-Wilson, and M. Schütz, J. Chem. Phys. **122**, 094113 (2005).

29. D. Usvyat, L. Maschio, F. R. Manby, S. Casassa, M. Schütz, and C. Pisani, Phys. Rev. B **76**, 075102 (2007).

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

31. A. Grüneis, M. Marsman, and G. Kresse, The Journal of Chemical Physics **133**, 074107 (2010), URL http://dx.doi.org/10.1063/1.3466765.

32. T. Shiozaki and S. Hirata, J. Chem. Phys. **132**, 151101 (2010).

33. J. J. Shepherd, A. Grüneis, G. Booth, M. Marsman, G. Kresse, and A. Alavi, Phys. Rev. B **86**, 035111 (2012).

34. L. Maschio, J. Chem. Theory Comput. **7**, 2818 (2011).

35. C. Pisani, M. Schütz, S. Casassa, D. Usvyat, L. Maschio, M. Lorenz, and A. Erba, Phys. Chem. Chem. Phys. **14**, 7615 (2012).

36. M. D. Ben, J. Hutter, and J. VandeVondele, J. Chem. Theory Comput. **8**, 4177 (2012).

37. M. Lorenz, L. Maschio, M. Schütz, and D. Usvyat, J. Chem. Phys. **137**, 204119 (2012).

38. M. D. Ben, J. Hutter, and J. VandeVondele, J. Chem. Theory Comput. **9**, 2654 (2013).

39. D. Usvyat, J. Chem. Phys. **139**, 194101 (2013).

40. A. Grüneis, J. J. Shepherd, A. Alavi, D. P. Tew, and G. H. Booth, J. Chem. Phys. **139**, 084112 (2013).

41. G. H. Booth, A. Grüneis, G. Kresse, and A. Alavi, Nature **493**, 365 (2013).

42. M. D. Ben, J. Hutter, and J. VandeVondele, J. Chem. Phys. **143**, 102803 (2015).

43. A. Grüneis, Phys. Rev. Lett. **115**, 066402 (2015).

44. D. Usvyat, L. Maschio, and M. Schütz, J. Chem. Phys. **143**, 102805 (2015).

45. G. H. Booth, T. Tsatsoulis, G. K.-L. Chan, and A. Grüneis, J. Chem. Phys. **145**, 084111 (2016).

46. V. V. Rybkin and J. VandeVondele, J. Chem. Theory Comput. **12**, 2214 (2016).

47. J. McClain, Q. Sun, G. K.-L. Chan, and T. C. Berkelbach, J. Chem. Theory Comput. **13**, 1209 (2017).

48. M. Schütz, J. Chem. Phys. **113**, 9986 (2000).

49. M. Schütz and H.-J. Werner, J. Chem. Phys. **114**, 661 (2001).

50. M. Schütz, Phys. Chem. Chem. Phys. **4**, 3941 (2002).

51. F. Neese, A. Hansen, and D. G. Liakos, J. Chem. Phys. **131**, 064103 (2009).

52. H.-J. Werner and M. Schütz, J. Chem. Phys. **135**, 144116 (2011).

53. J. Yang, G. K. L. Chan, F. R. Manby, M. Schütz, and H.-J. Werner, J. Chem. Phys. **136**, 144105 (2012).

54. M. Schütz, J. Yang, G. K. L. Chan, F. R. Manby, and H.-J. Werner, J. Chem. Phys. **138**, 054109 (2013).

55. C. Riplinger and F. Neese, J. Chem. Phys. **138**, 034106 (2013).

56. C. Riplinger, B. Sandhoefer, A. Hansen, and F. Neese, J. Comp. Phys. **139**, 134101 (2013).

57. D. Kats and F. R. Manby, J. Chem. Phys. **138**, 144101 (2013).

58. O. Masur, D. Usvyat, and M. Schütz, J. Chem. Phys. **139**, 164116 (2013).

59. M. Schütz, O. Masur, and D. Usvyat, J. Chem. Phys. **140**, 244107 (2014).

60. M. Schwilk, D. Usvyat, and H.-J. Werner, J. Chem. Phys. **142**, 121102 (2015).

61. F. Pavosevic, C. Peng, P. Pinski, C. Riplinger, F. Neese, and E. F. Valeev, J. Chem. Phys. **146**, 174108 (2017).

62. M. Schwilk, Q. Ma, C. Köppl, and H.-J. Werner, J. Chem. Theory Comput. **13**, Article ASAP, DOI: 10.1021/acs.jctc.7b00554 (2017).

63. M. Schütz, J. Chem. Phys. **116**, 8772 (2002).

64. L. Maschio, D. Usvyat, F. R. Manby, S. Casassa, C. Pisani, and M. Schütz, Phys. Rev. B **76**, 075101 (2007).

65. C. Pisani, L. Maschio, S. Casassa, M. Halo, M. Schütz, and D. Usvyat, J. Comput. Chem. **29**, 2113 (2008).

66. D. Usvyat, L. Maschio, C. Pisani, and M. Schütz, Z. Phys. Chem. **224**, 441 (2010).

67. M. Schütz, D. Usvyat, M. Lorenz, C. Pisani, L. Maschio, S. Casassa, and M. Halo, in *Accurate Condensed Phase Quantum Chemistry*, edited by F. R. Manby (CRC Press, Boca Raton, FL, 2010), p. 29.

68. R. Dovesi, R. Orlando, A. Erba, C. M. Zicovich-Wilson, B. Civalleri, S. Casassa, L. Maschio, M. Ferrabone, M. De La Pierre, P. D'Arco, et al., International Journal of Quantum Chemistry (2014), ISSN 1097-461X.

69. C. M. Zicovich-Wilson, R. Dovesi, and V. R. Saunders, J. Chem. Phys. **115**, 9708 (2001).

70. S. Casassa, C. M. Zicovich-Wilson, and C. Pisani, Theor. Chem. Acc. **116**, 726 (2006).

71. S. Saebø and P. Pulay, Chem. Phys. Lett. **113**, 13 (1985).

72. J. W. Boughton and P. Pulay, J. Comput. Chem. **14**, 736 (1993).

73. C. Hampel and H.-J. Werner, J. Chem. Phys. **104**, 6286 (1996).

74. M. Schütz, G. Hetzer, and H.-J. Werner, J. Chem. Phys. **111**, 5691 (1999).

75. D. Kats, J. Chem. Phys. **141**, 244101 (2014).

76. D. Kats, J. Chem. Phys. **145**, 014103 (2016).

77. W. Meyer, Int. J. Quantum Chem. **S5**, 341 (1971).

78. F. Neese, F. Wennmohs, and A. Hansen, J. Chem. Phys. **130**, 114108 (2009).

79. A. Hansen, D. G. Liakos, and F. Neese, J. Chem. Phys. **135**, 214102 (2011).

80. D. P. Tew, B. Helmich, and C. Hättig, J. Chem. Phys. **135**, 074107 (2011).

81. C. Hättig, D. P. Tew, and B. Helmich, J. Chem. Phys. **136**, 204105 (2012).

82. Z. Rolik, L. Szegedy, I. Ladjanszki, B. Ladoczki, and M. Kallay, J. Chem. Phys. **139**, 094105 (2013).

83. H.-J. Werner, G. Knizia, C. Krause, M. Schwilk, and M. Dornbach, J. Chem. Theory Comput. **11**, 484 (2015).

84. Q. Ma and H.-J. Werner, J. Chem. Theory Comput. **11**, 5291 (2015).

85. Y. Guo, K. Sivalingam, E. F. Valeev, and F. Neese, J. Chem. Phys. **144**, 094111 (2016).

86. F. Menezes, D. Kats, and H.-J. Werner, J. Chem. Phys. **145**, 124115 (2016).

87. J. Yang, Y. Kurashige, F. R. Manby, and G. K. L. Chan, J. Chem. Phys. **134**, 044123 (2011).

88. Y. Kurashige, J. Yang, G. K. L. Chan, and F. R. Manby, J. Chem. Phys. **136**, 124106 (2012).

89. C. Krause and H.-J. Werner, Phys. Chem. Chem. Phys. **14**, 7591 (2012).

90. G. Sansone, L. Maschio, D. Usvyat, M. Schütz, and A. Karttunen, J. Phys. Chem. Lett. **7**, 131 (2016).

91. A. Grüneis, G. H. Booth, M. Marsman, J. Spencer, A. Alavi, and G. Kresse, J. Chem. Theory Comput. **7**, 2780 (2011).

92. W. Foulkes, L. Mits, R. Needs, and G. Rajagopal, Rev. Mod. Phys. **73**, 33 (2001).

93. N. S. Blunt, S. D. Smart, J. A. F. Kersten, J. S. Spencer, G. H. Booth, and A. Alavi, J. Chem. Phys. **142**, 184107 (2015).

94. D. Usvyat and A. Alavi (2017 (in preparation)), full configuration interaction for a fragment embedded in periodic mean field.

95. B. Herschend, M. Baudin, and K. Hermansson, J. Chem. Phys. **120**, 4939 (2004).

96. U. Birkenheuer, P. Fulde, and H. Stoll, Theor. Chem. Acc. **116**, 398 (2006).

97. C. Müller and K. Hermansson, Surf. Sci. **603**, 3329 (2009).

98. G. J. O. Beran and K. Nanda, J. Phys. Chem. Lett. **1**, 3480 (2010).

99. C. Müller, D. Usvyat, and H. Stoll, Phys. Rev. B **83**, 245136 (2011).

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

101. S. Wen, K. Nanda, Y. Huanga, and G. J. O. Beran, Phys. Chem. Chem. Phys. **14**, 7578 (2012).

102. C. R. Taylor, P. J. Bygrave, J. N. Hart, N. L. Allan, and F. R. Manby, Phys. Chem. Chem. Phys. **14**, 7739 (2012).

103. F. R. Manby, M. Stella, J. D. Goodpaster, and T. F. M. III, J. Chem. Theory Comput. **8**, 2564 (2012).

104. P. J. Bygrave, N. L. Allan, and F. R. Manby, J. Chem. Phys. **137**, 164102 (2012).

105. G. Knizia and G. K. L. Chan, Phys. Rev. Lett. **109**, 186404 (2012).

106. O. Sode, M. Keçili, K. Yagi, and S. Hirata, J. Chem. Phys. **138**, 074501 (2013).

107. G. Knizia and G. K. L. Chan, J. Chem. Theory Comput. **9**, 1428 (2013).

108. T. Lan, A. A. Kananenka, and D. Zgid, J. Chem. Phys. **143**, 241102 (2015).

109. M. E. Fornace, J. Lee, K. Miyamoto, F. R. Manby, and T. F. M. III, J. Chem. Theory Comput. **11**, 568 (2015).

110. O. Masur, M. Schütz, L. Maschio, and D. Usvyat, J. Chem. Theory Comput. **12**, 5145 (2016).

111. C. Tuma and J. Sauer, Phys. Chem. Chem. Phys. **8**, 3955 (2006).

112. M. Rubes and O. Bludsky, Phys. Chem. Chem. Phys. **10**, 2611 (2008).

113. S. Tosoni and J. Sauer, Phys. Chem. Chem. Phys. **12**, 14330 (2010).

114. A. D. Boese and J. Sauer, Phys. Chem. Chem. Phys. **15**, 16481 (2013).

115. R. Martinez-Casado, D. Usvyat, L. Maschio, G. Mallia, S. Casassa, J. Ellis, M. Schütz, and N. M. Harrison, Phys. Rev. B **89**, 205138 (2014).

116. D. Usvyat, J. Chem. Phys. **143**, 104704 (2015).

117. T. Tsatsoulis, F. Hummel, D. Usvyat, M. Schütz, G. H. Booth, S. S. Binnie, M. J. Gillan, D. Alfe, A. Michaelides, and A. Grüneis, J. Chem. Phys. **146**, 204108 (2017).

118. H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, and M. Schütz, Comput. Mol. Sci. **2**, 242 (2012).

119. H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, M. Schütz, et al., *Molpro, version 2015.1, a package of ab initio programs* (2015), see.

120. K. Liao and A. Grüneis, J. Chem. Phys. **145**, 141102 (2016).

121. G. E. Scuseria, T. M. Henderson, and D. C. Sorensen, J. Chem. Phys. **129**, 231101 (2008).

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

123. J. Toulouse, W. Zhu, A. Savin, G. Jansen, and J. G. Ángyán, J. Chem. Phys. **135**, 084119 (2011).

124. L. Hammerschmidt, C. Müller, and B. Paulus, J. Chem. Phys. **136**, 124117 (2012).

125. K. G. Steenbergen, N. Gaston, C. Müller, and B. Paulus, J. Chem. Phys. **141**, 124707 (2014).

126. S. Saebø and P. Pulay, J. Chem. Phys. **86**, 914 (1987).

127. P. Pulay, Chem. Phys. Lett. **100**, 151 (1983).

128. P. Pulay, S. Saebø, and W. Meyer, J. Chem. Phys. **81**, 1901 (1984).

129. P. Pulay and S. Saebø, Theor. Chim. Acta **69**, 357 (1986).

130. S. Saebø and P. Pulay, J. Chem. Phys. **88**, 1884 (1988).

131. S. Saebø and P. Pulay, Annu. Rev. Phys. Chem. **44**, 213 (1993).

132. M. Schütz and F. R. Manby, Phys. Chem. Chem. Phys. **5**, 3349 (2003).

133. C. Pisani, R. Dovesi, and C. Roetti, *Hartree-Fock Ab Initio Treatment of Crystalline solids*, vol. 48 of *Lecture Notes in Chemistry Series* (Springer Verlag, Berlin, 1988).

134. R. Dovesi, R. Orlando, A. Erba, C. M. Zicovich-Wilson, B. Civalleri, S. Casassa, L. Maschio, M. Ferrabone, M. D. L. Pierre, P. DÁrco, et al., Int. J. Quantum Chem. **114**, 1287 (2014).

135. L. Maschio and D. Usvyat, Phys. Rev. B **78**, 073102 (2008).

136. R. D. Mattuck, *A Guide to Feynman Diagrams in the Many-Body Problem* (Dover Publications, New York, 1992).

137. J. J. Shepherd and A. Grüneis, Phys. Rev. Lett. **110**, 226401 (2013).

138. A. Ambrosetti, A. M. Reilly, R. A. D. Jr., and A. Tkatchenko, J. Chem. Phys. **140**, 18A508 (2014).

139. D. Kats, J. Chem. Phys. **144**, 044102 (2016).

140. S. J. Nolan, M. J. Gillan, D. Alfè, N. L. Allan, and F. R. Manby, Phys. Rev. B **80**, 165109 (2009).

141. M. Schütz, L. Maschio, A. J. Karttunen, and D. Usvyat, J. Phys. Chem. Lett. **8**, 1290 (2017).

142. A. Hermann and P. Schwerdtfeger, J. Chem. Phys. **131**, 244508 (2009).

143. N. Marom, J. Bernstein, J. Garel, A. Tkatchenko, E. Joselevich, L. Kronik, , and O. Hod, Phys. Rev. Lett. **105**, 046801 (2010).

144. L. Maschio, D. Usvyat, M. Schütz, and B. Civalleri, J. Chem. Phys. **132**, 134706 (2010).

145. G.-X. Zhang, A. Tkatchenko, J. Paier, H. Appel, and M. Scheffler, Phys. Rev. Lett. **107**, 245501 (2011).

146. B. Santra, J. Klimes, D. Alfe, A. Tkatchenko, B. Slater, A. Michaelides, R. Car, and M. Scheffler, Phys. Rev. Lett. **107**, 185701 (2011).

147. D. Usvyat, K. Sadeghian, L. Maschio, and M. Schütz, Phys. Rev. B **86**, 045412 (2012).

148. W. Liu, J. Carrasco, B. Santra, A. Michaelides, M. Scheffler, and A. Tkatchenko, Phys. Rev. B **86**, 245405 (2012).

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

150. A. M. Reilly and A. Tkatchenko, J. Chem. Phys. **139**, 024705 (2013).

151. V. V. Gobre and A. Tkatchenko, Nature communications **4**, 2341 (2013).

152. J. Li, O. Sode, G. Voth, and S. Hirata, Nature communications **4**, 2647 (2013).

153. R. A. DiStasio, V. V. Gobre, and A. Tkatchenko, J. Phys.: Condens. Matter **26**, 213202 (2014).

154. C. Müller and D. Spångberg, J. Comput. Chem. **36**, 1420 (2015).

155. L. Shulenburger, A. Baczewski, Z. Zhu, J. Guan, and D. Tománek, Nano Lett. **15**, 8170 (2015).

156. A. Ambrosetti, N. Ferri, R. A. DiStasio, and A. Tkatchenko, Science **351**, 1171 (2016).

157. J. Hermann, R. A. DiStasio, and A. Tkatchenko, Chem. Rev. **117**, 4714 (2017).

158. A. Ambrosetti, P. L. Silvestrelli, and A. Tkatchenko, Phys. Rev. B **95**, 235417 (2017).

159. L. Li, Y. Yu, G. J. Ye, Q. Ge, X. Ou, H. Wu, D. Feng, X. H. Chen, and Y. Zhang, Nat. Nanotech. **9**, 372 (2014).

160. H. Liu, A. T. Neal, Z. Zhu, Z. Luo, X. Xu, D. Tomnek, and P. D. Ye, ACS Nano **8**, 4033 (2014).

161. J. Yang, R. Xu, J. Pei, Y. W. Myint, F. Wang, Z. Wang, S. Zhang, Z. Yu, and Y. Lu (2014), unambiguous identification of monolayer phosphorene by phase-shifting interferometry. Preprint at https://arxiv.org/abs/1412.6701.

162. V. Tran, R. Soklaski, Y. Liang, and L. Yang, Phys. Rev. B **89**, 235319 (2014).

163. A. Castellanos-Gomez, J. Phys. Chem. Lett. **6**, 4280 (2015).

164. J. Qiao, X. Kong, Z.-X. Hu, F. Yang, and W. Ji, Nat. Commun, **4**, 4475 (2014).

165. L. Kou, C. Chen, and S. C. Smith, J. Phys. Chem. Lett. **6**, 2794 (2015).

166. L. Li, F. Yang, G. J. Ye, Z. Zhang, Z. Zhu, W. Lou, X. Zhou, L. Li, K. Watanabe, T. Taniguchi, et al., Nat. Nanotech. **11**, 593 (2016).

167. H. Kim, J. Korean Phys. Socs. **64**, 547 (2014).

168. M. Cascella, I.-C. Lin, I. Tavernelli, and U. Rothlisberger, J. Chem. Theory Comput. **5**, 2930 (2009).

169. F. Bachhuber, J. von Appen, R. Dronskowski, P. Schmidt, T. Nilges, A. Pfitzner, and R. Weihrich, Z. Kristallogr. **230**, 107 (2015).

170. M. Wu, H. Fu, L. Zhou, K. Yao, and X. C. Zeng, Nano Lett. **15**, 3557 (2015).

171. F. Bachhuber, J. von Appen, R. Dronskowski, P. Schmidt, T. Nilges, A. Pfitzner, and R. Weihrich, Angew. Chem. Int. Ed. **53**, 11629 (2014).

172. S. Tosoni, B. Civalleri, and P. Ugliengo, J. Phys. Chem. C **114**, 19984 (2010).

173. D. Löffler, J. J. Uhlrich, M. Baron, B. Yang, X. Yu, L. Lichtenstein, L. Heinke, C. Büchner, M. Heyde, S. Shaikhutdinov, et al., Phys. Rev. Lett. **105**, 146104 (2010).

174. L. Lichtenstein, C. Büchner, B. Yang, S. Shaikhutdinov, M. Heyde, M. Sierka, R. Wlodarczyk, J. Sauer, and H.-J. Freund, Angew.Chem. Int. Ed. **51**, 404 (2012).

175. L. Lichtenstein, M. Heyde, and H.-J. Freund, J. Phys. Chem. C **116**, 20426 (2012).

176. T. Björkman, S. Kurasch, O. Lehtinen, J. Kotakoski, O. V. Yazyev, A. Srivastava, V. Skakalova, J. H. Smet, U. Kaiser, and A. V. Krasheninnikov, Sci. Rep. **3**, 3482 (2013).

177. P. Y. Huang, S. Kurasch, J. S. Alden, A. Shekhawat, A. A. Alemi, P. L. McEuen, J. P. Sethna, U. Kaiser, and D. A. Muller, Science **342**, 224 (2013).

178. X. Yu, E. Emmez, Q. Pan, B. Yang, S. Pomp, W. E. Kaden, M. Sterrer, S. Shaikhutdinov, H.-J. Freund, I. Goikoetxea, et al., Phys. Chem. Chem. Phys. **18**, 3755 (2016).

179. P. Schlexer, G. Pacchioni, R. Wlodarczyk, and J. Sauer, Surf. Sci. **648**, 2 (2016).

180. K. M. Burson, C. Büchner, M. Heyde, and H.-J. Freund, J. Phys.: Condens. Matter **29**, 035002 (2017).

181. M. F. Peintinger, D. V. Oliveira, and T. Bredow, J. Comp. Chem. **34**, 451 (2013).

182. See Supplementary Material Document No._____ for specification of the structures and basis sets.

183. R. A. Kendall, T. H. Dunning, Jr., and R. J. Harrison, J. Chem. Phys. **96**, 6796 (1992).

184. F. Weigend and R. Ahlrichs, Phys. Chem. Chem. Phys. **7**, 3297 (2005).

185. T. H. Dunning, Jr., J. Chem. Phys. **90**, 1007 (1989).

186. R. Dovesi, V. R. Saunders, R. Roetti, R. Orlando, C. M. Zicovich-Wilson, F. Pascale, B. Civalleri, K. Doll, N. M. Harrison, I. J. Bush, et al., *CRYSTAL14 User's Manual*, University of Torino, Torino (2014).

187. F. Weigend, A. Köhn, and C. Hättig, J. Chem. Phys. **116**, 3175 (2002).

188. V. Saunders, C. Freyria-Fava, R. Dovesi, L. Salasco, and C. Roetti, Mol. Phys. **77**, 629 (1992).

189. K. Momma and F. Izumi, J. Appl. Crystallogr. **44**, 1272 (2011).

190. A. Kokalj, Comp. Mater. Sci. **28**, 155 (2003).