Periodic local MP2 method employing orbital specific virtuals

Denis Usvyat, Lorenzo Maschio, and Martin Schütz

Articles you may be interested in

A new ab initio potential energy surface for the collisional excitation of HCN by para- and ortho-H2
J. Chem. Phys. **139**, 224301 (2013); 10.1063/1.4833676

Linear-scaling explicitly correlated treatment of solids: Periodic local MP2-F12 method
J. Chem. Phys. **139**, 194101 (2013); 10.1063/1.4829898

Calculating dispersion interactions using maximally localized Wannier functions
J. Chem. Phys. **135**, 154105 (2011); 10.1063/1.3647912

Approaching the theoretical limit in periodic local MP2 calculations with atomic-orbital basis sets: The case of LiH
J. Chem. Phys. **134**, 214105 (2011); 10.1063/1.3595514

Local-MP2 electron correlation method for nonconducting crystals
J. Chem. Phys. **122**, 094113 (2005); 10.1063/1.1857479
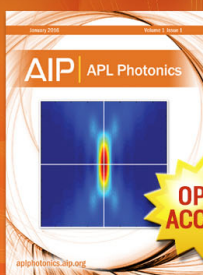
Periodic local MP2 method employing orbital specific virtuals

Denis Usvyat,[1,a)] Lorenzo Maschio,[2,b)] and Martin Schütz[1,c)]
[1]*Institute for Physical and Theoretical Chemistry, Universität Regensburg, Universitätsstraße 31, D-93040 Regensburg, Germany*
[2]*Dipartimento di Chimica, and Centre of Excellence NIS (Nanostructured Interfaces and Surfaces), Università di Torino, via Giuria 5, I-10125 Torino, Italy*

We introduce orbital specific virtuals (OSVs) to represent the truncated pair-specific virtual space in periodic local Møller-Plesset perturbation theory of second order (LMP2). The OSVs are constructed by diagonalization of the LMP2 amplitude matrices which correspond to *diagonal* Wannier-function (WF) pairs. Only a subset of these OSVs is adopted for the subsequent OSV-LMP2 calculation, namely, those with largest contribution to the diagonal pair correlation energy and with the accumulated value of these contributions reaching a certain accuracy. The virtual space for a general (non diagonal) pair is spanned by the union of the two OSV sets related to the individual WFs of the pair. In the periodic LMP2 method, the diagonal LMP2 amplitude matrices needed for the construction of the OSVs are calculated in the basis of projected atomic orbitals (PAOs), employing very large PAO domains. It turns out that the OSVs are excellent to describe short range correlation, yet less appropriate for long range van der Waals correlation. In order to compensate for this bias towards short range correlation, we augment the virtual space spanned by the OSVs by the most diffuse PAOs of the corresponding minimal PAO domain. The Fock and overlap matrices in OSV basis are constructed in the reciprocal space. The 4-index electron repulsion integrals are calculated by local density fitting and, for distant pairs, via multipole approximation. New procedures for determining the fit-domains and the distant-pair lists, leading to higher efficiency in the 4-index integral evaluation, have been implemented. Generally, and in contrast to our previous PAO based periodic LMP2 method, the OSV-LMP2 method does not require anymore great care in the specification of the individual domains (to get a balanced description when calculating energy differences) and is in that sense a black box procedure. Discontinuities in potential energy surfaces, which may occur for PAO-based calculations if one is not careful, virtually disappear for OSV-LMP2. Moreover, due to much increased compactness of the pair-specific virtual spaces, the OSV-LMP2 calculations are faster and require much less memory than PAO-LMP2 calculations, despite the noticeable overhead of the initial OSV construction procedure. © *2015 AIP Publishing LLC.* [http://dx.doi.org/10.1063/1.4921301]

I. INTRODUCTION

During the past several years, periodic quantum chemical wavefunction methods have become an increasingly important tool in solid state applications.[1–10] Furthermore, by combining the periodic quantum chemical treatment with finite-cluster approaches, like, e.g., the incremental scheme,[11,12] unprecedented accuracy has been reached in the determination of interaction energies in periodic systems.[13–18]

Presently, there are three development directions of 3D-periodic correlation techniques; (i) the reciprocal-space plane-wave approach,[19–22] reaching coupled cluster level; (ii) massively parallel canonical MP2[23] and RPA[24] methods, employing atomic orbital basis sets and plane waves as an auxiliary basis for density fitting; and (iii) direct-space local MP2[25–27] and MP2-F12[28] methods.

Employing spatially local orbitals to span the occupied and virtual Hartree-Fock (HF) spaces introduces sparsity in direct-space quantities like the MP2 amplitudes, etc. This can be exploited to achieve low (or even linear) scaling of the computational cost with the number of atoms per unit cell. For the occupied space, usually orthogonal localized orbitals are employed. If the orthogonality constraint for the occupied orbitals is released,[29,30] their localization rate can be substantially improved. On the other hand, non-orthogonality leads to additional couplings between individual amplitudes, i.e., inter-pair coupling in the local MP2 equations via the corresponding overlap matrix elements. This, in turn, again increases the computational cost and offsets the gain due to the additional sparsity. In MP2 (as well as in other perturbative methods) it is possible to avoid this complication by employing the Laplace-transform of the energy denominator,[31,32] which is used, for example, for AO-based schemes.[33–35]

For the virtual space, it is also possible to construct orthogonal localized orbitals.[36] However, since in high quality correlated calculations the number of virtual orbitals is usually much larger than the number of the occupied ones, and they are more diffuse than the latter, their mutual orthogonalization can worsen their localization. Within the local correlation

---

a)Electronic address: denis.usvyat@chemie.uni-regensburg.de
b)Electronic address: lorenzo.maschio@unito.it
c)Electronic address: martin.schuetz@chemie.uni-regensburg.de

143, 102805-1

formalism, nonorthogonality of the virtual orbitals is not a great obstacle for high computational efficiency. Therefore, more effort is invested in constructing a compact and computationally convenient, but not necessarily orthogonal virtual basis.

Projected atomic orbitals (PAOs), introduced first by Pulay,[37–39] are among the most widely used local orbital representations of the virtual space. PAOs, like the AOs from which they originate, are highly localized. Hence, the excitation space related to a certain pair of occupied orbitals (denoted as *pair virtual space* in the following) can be truncated *a priori* to pair specific *pair domains*. These pair domains comprise those (relatively few) PAOs, centered in the vicinity of the respective two occupied orbitals. Such a truncation is motivated by the rapid (exponential) decay rate of the pair amplitudes with respect to the distance between occupied orbitals and PAOs. Note that the PAOs themselves are *not pair specific*, yet the pair domains are *pair specific* in the sense that they contain pair specific subsets of PAOs.

The PAO approach has been employed in molecular local MP2[40–44] and coupled cluster[45–52] methods including explicit correlation,[53–55] in time-dependent coupled cluster linear response,[56–60] as well as in multireference techniques.[61] Also the periodic local MP2[26,62–64] and MP2-F12[28] methods as implemented in the CRYSCOR program[26] so far employ PAOs.

However, PAOs have certain shortcomings. In particular, for high accuracy, the PAO pair domains may still be relatively large, especially so in 3D densely packed systems (up to a few hundred orbitals per pair),[65] which leads to excessively large doubles amplitude sets. Furthermore, pair domains usually respond abruptly rather than smoothly to changes in the molecular structure, i.e., by inclusion or exclusion of an atom (with all PAOs centered on it) from the pair domain. This may lead to discontinuities in potential energy surfaces and to nonphysical artifacts in energy differences, unless the domains are chosen to be identical for all considered geometries. Note that geometry optimizations employ fixed domains and pair lists, which eventually have to be re-specified close to the minimum in case of large geometry changes between initial and final structures.[41,60] Head-Gordon *et al.* proposed to damp pair domains in order to avoid discontinuities in the potential energy surfaces.[66] However, such an approach would lead to even larger pair domains in 3D densely packed systems.

True pair specific orbitals, i.e., orbitals which are indeed pair specific by themselves, the so called pair natural orbitals (PNOs), were already suggested 40 years ago by Meyer.[67] Recently, they have been revived in the local context by Neese *et al.* and others.[68–75] A truncated PNO basis spans the most compact virtual space that captures a given amount of pair correlation. For a given accuracy, the number of PNOs required is many times smaller than that of PAOs.[74] Furthermore, PNO-based schemes are virtually free from the problem of potential energy surface discontinuity, since PNOs naturally adapt themselves to the change of geometry. A drawback of the PNOs approach is that the overall amount of PNOs is much larger than that of PAOs (since they are truly pair specific). Therefore, objects like the PNO Fock and overlap matrices can become sizeable and hence expensive to compute, store, and manipulate.

The orbital specific virtual (OSV) approach,[76–80] which uses the unions of diagonal pair natural orbitals as pair specific virtual space, can be regarded as a compromise between PAOs and PNOs. Indeed, substantially less OSVs than PAOs per pair are required to reach the same accuracy. Furthermore, also OSVs adapt to geometry changes and hence are less prone to cause discontinuities in the potential energy surface than PAOs. Compared to PNOs, the pair-specific OSV spaces are larger, but the dimensionality of the intermediate quantities and the associated computational cost is generally lower. Moreover, the OSV machinery can be used as the first step towards a PNO scheme, both algorithmically (PNOs are often constructed from OSVs[70,73,75]) and technically, as the OSV routines can be adapted accordingly for PNOs. A comparison between various choices of orbitals to represent pair virtual spaces in molecular systems can be found in Ref. 74.

In this paper, we present an OSV implementation of the periodic local MP2 method. In the following sections, we describe the formalism and algorithms for constructing the OSVs, for calculating the 4-index integrals involving OSVs via local density fitting and multipole approximations, and for solving the LMP2 equations in the OSV basis. Furthermore, the shortcoming of the OSVs to capture long-range van der Waals correlation is analyzed. As is demonstrated, this deficiency can be circumvented by adding the few most diffuse PAOs to each related OSV set.

The new OSV based periodic LMP2 method, in contrast to its PAO based precursor, is a *black-box method*. Moreover, it is computationally more efficient (due to a lower prefactor) than the previous (already linear scaling) PAO based method.

II. THEORY

A. Orbital specific virtuals

The concept of OSVs has initially been introduced from the angle of a tensor decomposition of the doubles amplitudes.[76] Equivalently, OSVs can be defined as PNOs for diagonal pairs, i.e., the orbitals that diagonalize the MP2 virtual pair density matrices $\mathbf{D}^{ii}$ corresponding to diagonal pairs ii of, in our case, localized occupied orbitals ϕ_i belonging to the reference unit cell,

$$D_{ab}^{ii} = \sum_c T_{ac}^{ii} T_{bc}^{ii}. \tag{1}$$

T_{ac}^{ii} are MP2 doubles amplitudes in the basis of yet unspecified normalized and mutually orthogonal virtual orbitals ϕ_a, ϕ_b, and ϕ_c. Since the amplitude matrix for a diagonal pair is symmetric, the eigenvectors of the density matrix $\mathbf{D}^{ii}$ are identical to those of the amplitudes T_{ab}^{ii} themselves, hence, the latter can be used for the OSV specification.

The orbital invariant MP2 formalism, which has to be adopted for local MP2, involves inter-pair couplings via the occupied-occupied block of the Fock matrix, i.e., the internal Fock matrix.[25,81] Hence, a linear equation system has to be solved to obtain the LMP2 amplitudes. OSVs generated from such amplitudes may be useful in the context of a subsequent coupled cluster calculation, but trivially not for the MP2 problem itself (since the problem is already solved then). In order

to generate a sensible set of OSVs for a local MP2 calculation itself, just the required *diagonal pair amplitudes* $\bar{\mathbf{T}}^{ii}$ are calculated at the *uncoupled* LMP2 level[70,73,75,79] (corresponding to the zeroth LMP2 iteration) according to

$$\bar{T}^{ii}_{ab} = \frac{(ia|jb)}{\epsilon_a + \epsilon_b - f_{ii} - f_{jj}}, \tag{2}$$

where $\mathbf{f}$ is the Fock matrix and $(ia|jb)$ is an electron repulsion integral in chemical (Mulliken) notation

$$(ia|jb) = \int d\mathbf{r}_1 \phi_i^*(\mathbf{r}_1)\phi_a(\mathbf{r}_1) \int d\mathbf{r}_2 \frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|} \phi_j^*(\mathbf{r}_2)\phi_b(\mathbf{r}_2). \tag{3}$$

Note that the $\bar{\mathbf{T}}^{ii}$ can deviate from the orbital-invariant LMP2 solution. Nevertheless, since the inter-pair couplings in LMP2 via the diagonal-dominant internal Fock matrix are weak, such an approximation is not expected to affect the OSVs significantly.

In Eq. (2), the virtual orbitals are assumed to diagonalize the Fock matrix, i.e., ϵ_a is the eigenvalue of the external Fock matrix related to virtual orbital ϕ_a. Hence, only the inter-pair coupling via the internal Fock matrix of the LMP2 equations is neglected. This is naturally achieved by employing canonical virtual orbitals ϕ_a.[73,76,79] However, the electron repulsion integrals in the canonical virtual basis are not available in the periodic LMP2 implementation. A convenient alternative is to utilize the pair-specific *pseudo-canonical orbitals* instead, which diagonalize the external Fock matrix in the spaces of the related PAO pair domains.

Pseudo-canonical orbitals are already used for generating the amplitude updates via first-order perturbation theory when solving the LMP2 equations. The amplitudes $\bar{\mathbf{T}}^{ii}$ defined in Eq. (2) thus are identical to the LMP2 amplitudes of the zeroth iteration before transforming them back from pseudo-canonical to PAO basis. The pair-specific transformation matrices $\mathbf{W}^{[ij]}$ between the pseudo-canonical and the PAO basis are normally evaluated by diagonalization of the related pair-specific piece of the PAO overlap matrix, removal of the numerically redundant orbitals, normalization of the surviving ones, and a subsequent diagonalization of the Fock matrix in this new orthogonal basis.[25,45,82] All this implies that the amplitudes $\bar{\mathbf{T}}^{ii}$ required for OSV construction can easily be computed with the existing LMP2 code.

Since only a very limited set of pairs (namely the diagonal set) is involved, much larger domains than in usual LMP2 calculations can be used without significant computational cost. This possibility of using large PAO domains for OSV construction is important because the individual virtual spaces spanned by these domains should be close to the full canonical virtual space; too small domains would re-introduce discontinuities in the potential energy surface through the backdoor.

The importance of a particular OSV $\phi_{\bar{a}}^{[i]}$ (related to occupied orbital ϕ_i) for the subsequent LMP2 calculation can be assessed by its individual contribution,

$$e^{ii}_{\bar{a}[i]} = \bar{t}^{ii}_{\bar{a}[i]}(i\tilde{a}[i]|i\tilde{a}[i]), \tag{4}$$

to the total correlation energy of the *ii*-pair. In the above equation, $\bar{t}^{ii}_{\bar{a}[i]}$ is the corresponding eigenvalue of the $\bar{\mathbf{T}}^{[ii]}$ amplitude matrix. In order to truncate the virtual space spanned by the OSVs related to ϕ_i, i.e., the eigenvectors of $\bar{\mathbf{T}}^{[ii]}$, the latter are ordered according to descending $e^{ii}_{\bar{a}[i]}$. OSVs are accepted until the difference between the sum over the corresponding $e^{ii}_{\bar{a}[i]}$ and the total *ii*-pair energy falls below a certain threshold $\mathcal{E}_{\text{OSV}}$.

The AO coefficients $\mathbf{C}^{\text{OSV}}$ for OSV $\phi_{\bar{a}}^{[i]}$ belonging to the occupied localized orbital (in the periodic context denoted as Wannier function (WF)) ϕ_i, are given by the expression

$$C^{\text{OSV}}_{\mu\mathcal{M},\bar{a}[i]} = \sum_{\bar{a}\mathcal{A}\in[ii]_{\text{PAO}}} C^{\text{PAO}}_{\mu\mathcal{M},\bar{a}\mathcal{A}} \sum_{a[i]} W^{[ii]}_{\bar{a}\mathcal{A},a[i]} Q^{ii}_{a[i],\bar{a}[i]}. \tag{5}$$

Here, greek indices denote AOs, and latin ones with an overbar—PAOs. The calligraphic indices represent the unit cells where the corresponding orbitals are centered (orbital indices without a calligraphic index lie in the reference cell), and $\mathbf{C}^{\text{PAO}}$ is the PAO coefficient matrix specifying the PAOs in the AO basis. We note that both OSVs and pseudocanonical orbitals always belong to the same cell as the corresponding WF. The PAO-summation is restricted to the (large, *vide supra*) PAO domain $[ii]$ of the diagonal *ii*-pair, $\phi_a^{[ii]}$ are the pseudo-canonical orbitals corresponding to this domain, and $Q^{ii}_{a[ii],\bar{a}[i]}$ is the eigenvector of the amplitude matrix $\bar{\mathbf{T}}^{ii}$ in the pseudo-canonical basis, corresponding to the eigenvalue $\bar{t}^{ii}_{\bar{a}[i]}$.

B. OSVs and long-range correlation

The OSVs obviously allow for a very efficient truncation of the virtual space for diagonal pairs (they are the related natural orbitals). Yet, this does not automatically imply that they also form an optimal basis for the off-diagonal pairs. In fact, an accurate description of long-range correlation (van der Waals dispersion) requires inclusion of diffuse orbitals, which are not as essential for the description of short-range correlation. Consequently, those OSVs needed to describe long-range correlation contribute very little to the diagonal pair energies and thus, are usually dropped in the OSV construction procedure described above, unless extremely tight truncation thresholds are used. Yet, such tight thresholds are not really an option, since (i) a lot of unwanted OSVs are then included in the calculation, as well, and (ii) OSVs corresponding to low eigenvalues $\bar{t}^{ii}_{\bar{a}[i]}$ become very delocalized and oscillatory due to their mutual orthogonality, implying that the resulting orbital product densities become increasingly difficult to fit, which, in turn, necessitates extensively large fit domains and rich fitting basis sets. This aspect will be further discussed and illustrated with numerical examples in Sec. III.

On the other hand, it is known that diffuse high angular momentum AOs are of prime importance for a proper description of long range correlation.[1] Therefore, the above mentioned problem can to a large extent be circumvented by augmenting the individual sets of OSVs (as obtained with a relatively loose threshold $\mathcal{E}_{\text{OSV}}$) by the few most diffuse PAOs centered on related atoms. It turns out that a single PAO shell per angular momentum for each atom within the minimal domain of a given WF is sufficient to properly describe long range correlation. These minimal domains are determined by the Boughton-Pulay (BP) procedure[83] with a criterion of 0.9. The number of PAOs in these minimal domains is rather modest; they contain

either just a single atom (for lone-pair WFs) or two atoms (for bonding WFs). Furthermore, these minimal domains are well defined and remain the same along changes in the geometry, unless the bonding pattern changes entirely. Augmenting the OSVs by such PAOs hence does not lead to discontinuities in the potential energy surface. In Sec. III, it is demonstrated that after inclusion of these extra PAOs, tight OSV truncation thresholds can be avoided, making the density fitting much more stable.

C. Electron repulsion integrals with OSVs

In order to adapt the existing PAO-based periodic LMP2 code described in Refs. 25, 63, 84, and 85 to OSVs, the latter are treated as fictitious PAOs, which belong not to atoms but rather to WFs that are considered in this context as pseudo-atoms. Above mentioned additional PAOs are also treated like OSVs, i.e., assigned to WFs rather than atoms, and the OSV coefficient matrix defined in Eq. (5) is extended with the respective piece of the PAO coefficient matrix by concatenation.

The OSV orbital pseudo domain $[i]$ for each WF ϕ_i consists only of one pseudo-atom, i.e., the WF itself. Proceeding along this avenue, the adaptation of the PAO routines to OSVs is straightforward: OSV pair domains $[ij]$ are formed by unifying the two individual orbital pseudo domains $[i]$ and $[j]$ of the WFs ϕ_i and ϕ_j. As in the PAO case, the functions spanning the pair-specific virtual space $[ij]$ can be redundant, and amplitude updates are computed in the same way as for PAOs.

However, some of the computational approaches employed for PAOs had to be modified to obtain stable results for OSVs. This mainly concerns the evaluation of electron repulsion integrals (ERIs) $(i\tilde{a}\mathcal{A}|j\mathcal{J}\tilde{b}\mathcal{B})$ involving OSVs: for "strong" or "weak" pairs $i\,j\mathcal{J}$, the calculation of the ERIs proceeds via local density fitting as described in Refs. 84 and 86, while for "distant" pairs the ERIs are approximated by a multipole expansion.[25] The latter implies that for distant pairs the respective orbital product densities $\phi_i\phi_{\tilde{a}\mathcal{A}}$ and $\phi_{j\mathcal{J}}\phi_{\tilde{b}\mathcal{B}}$ must not mutually overlap. In order to fulfill this condition, a conservative value of 8 Å for the distance criterion specifying distant pairs is employed in our PAO-based periodic LMP2 implementation. This was shown to work well in numerous applications.[1–4,14–17,87] Yet specifying the distant pairs by a single strict distance criterion for all systems has certain disadvantages, which become more important when substituting PAOs by OSVs. First, for many systems this single distance criterion is much too strict, which unnecessarily increases the computational cost. On the other hand, for other systems like e.g., small band gap semiconductors where the WFs cannot be so well localized, a distance criterion of 8 Å may still be good for PAOs, but insufficient for OSVs, which can be more delocalized than PAOs. To circumvent this problem, we propose an adaptive approach, which decides on the basis of the real spread and orientation of the WFs in the pair, if a multipole approximation is appropriate or not, i.e., if the pair should belong to the strong/weak pair class, or the distant pair class.

An estimation of the mutual penetration of the two WF-OSV product densities of a pair can be reduced to the analysis of the penetration of the WFs of this pair. To this end, we adopt an approach recently proposed by Kats:[88] the mutual penetration of two WFs is estimated by the product of the two atomic populations $q_{A\mathcal{A}}^i$ obtained from the two *individual* densities of the two WFs, i.e.,

$$O_{ij\mathcal{J}} = \sum_{A\mathcal{A}} \left|q_{A\mathcal{A}}^i\right|\left|q_{A(\mathcal{A}\ominus\mathcal{J})}^j\right|, \tag{6}$$

with

$$q_{A\mathcal{A}}^i = \sum_{\nu\in A}\sum_{\mu\mathcal{M}} C_{\mu\mathcal{M},i}^{\mathrm{WF}} S_{\mu,\nu(\mathcal{A}\ominus\mathcal{M})} C_{\nu\mathcal{A},i}^{\mathrm{WF}}. \tag{7}$$

Here, $\mathbf{C}^{\mathrm{WF}}$ is the AO coefficient matrix of the WFs, index A denotes atoms in the reference cell, index $\mathcal{A}$ the lattice vector of the atom's actual cell, and the symbolic operation $\ominus$ the actual operation on the corresponding lattice vectors. Since the Mulliken charges in Eq. (7) can be negative, their absolute values are used in the estimator $O_{ij\mathcal{J}}$ defined in Eq. (6).

$O_{ij\mathcal{J}}$ is evaluated for each pair from the initial large pair list (to define the latter, a sufficiently large distance criterion of 12 Å is used). If $O_{ij\mathcal{J}}$ is smaller than a specified threshold, the pair is assigned to the distant pair class and the multipole approximation is employed for the ERIs. Since the WFs of such distant pairs are well separated, only ERIs of the type $(i\,\tilde{a}^{[i]}|j\mathcal{J}\,\tilde{b}^{[j]}\mathcal{J})$ are calculated. For the other pairs, the density fitting (DF) procedure is used to compute all 4 types of integrals

$$\begin{aligned}(i\,\tilde{a}^{[i]}|j\mathcal{J}\,\tilde{b}^{[i]}), && (i\,\tilde{a}^{[i]}|j\mathcal{J}\,\tilde{b}^{[j]}\mathcal{J}),\\(i\,\tilde{a}^{[j]}\mathcal{J}|j\mathcal{J}\,\tilde{b}^{[i]}), && (i\,\tilde{a}^{[j]}\mathcal{J}|j\mathcal{J}\,\tilde{b}^{[j]}\mathcal{J}).\end{aligned}$$

These are computed via robust local density fitting[28,84,89,90]

$$\begin{aligned}(i\tilde{a}\mathcal{A}|j\mathcal{J}\tilde{b}\mathcal{B}) &= \sum_{P\mathcal{P}\in[i\tilde{a}\mathcal{A}]_{\mathrm{DF}}} d_{P\mathcal{P}}^{i\tilde{a}\mathcal{A}}\,(P\mathcal{P}|j\mathcal{J}\tilde{b}\mathcal{B})\\&+ \sum_{P\mathcal{P}\in[j\mathcal{J}\tilde{b}\mathcal{B}]_{\mathrm{DF}}} (i\tilde{a}\mathcal{A}|P\mathcal{P})\,d_{P\mathcal{P}}^{j\mathcal{J}\tilde{b}\mathcal{B}}\\&- \sum_{\substack{P\mathcal{P}\in[i\tilde{a}\mathcal{A}]_{\mathrm{DF}}\\Q\mathcal{Q}\in[j\mathcal{J}\tilde{b}\mathcal{B}]_{\mathrm{DF}}}} d_{P\mathcal{P}}^{i\tilde{a}\mathcal{A}}\,(P\mathcal{P}|Q\mathcal{Q})\,d_{Q\mathcal{Q}}^{j\mathcal{J}\tilde{b}\mathcal{B}}, \tag{8}\end{aligned}$$

with P and Q being indices of auxiliary basis functions. The DF coefficients are determined as

$$d_{P\mathcal{P}}^{i\tilde{a}\mathcal{A}} = \sum_{Q\mathcal{Q}\in[i\tilde{a}\mathcal{A}]_{\mathrm{DF}}} (i\tilde{a}\mathcal{A}|Q\mathcal{Q})\,(Q\mathcal{Q}|P\mathcal{P})^{-1}. \tag{9}$$

Of course translational symmetry applies to all objects in Eq. (8) so that all quantities are evaluated with one index being in the reference cell, e.g.,

$$(P\mathcal{P}|Q\mathcal{Q}) = (P|Q(\mathcal{Q}\ominus\mathcal{P})) \tag{10}$$

or

$$(P\mathcal{P}|j\mathcal{J}\tilde{b}\mathcal{B}) = (P(\mathcal{P}\ominus\mathcal{J})|j\tilde{b}(\mathcal{B}\ominus\mathcal{J})). \tag{11}$$

In the PAO case, the local fit-domains $[i\tilde{a}\mathcal{A}]_{\mathrm{DF}}$ are constructed on the basis of the quasi-populations defined in Eq. (2.13) of Ref. 84 for each pair formed by a WF and an atom. $[i\tilde{a}\mathcal{A}]_{\mathrm{DF}}$ thus is identical for each pair of the given WF and any PAO centered on this atom. For OSVs such a specification turned out not to be sufficiently stable. OSVs can be less localized and have a more complicated form than PAOs. In this work

we employ a simpler technique for the construction of the fit-domains, which turned out to be quite successful. The OSVs belonging to a certain WF $j\mathcal{J}$ are localized around this WF. Then, the product of any of these OSVs and the WF i should have its maximum somewhere in between the centers of the WF i and WF $j\mathcal{J}$. Consequently, a decent fit-domain appropriate for all ERIs involving WF i and the OSVs of WF $j\mathcal{J}$ should comprise all fitting functions centered on the atoms surrounding the mid-point between the WFs i and $j\mathcal{J}$ within a certain radius. Test calculations indeed show that such an approach works well, providing sufficient accuracy with just a few fit-domain centers (see Sec. III).

D. Fock and overlap matrices in the OSV basis

The next step after constructing the ERIs (either via DF or multipole approximation) is to solve the LMP2 equations.[25] Here again the PAO code is employed with the OSVs treated as fictitious PAOs centered on pseudo-atoms as described at the beginning of Sec. II C.

The LMP2 equations involve the external Fock and overlap matrices.[25] It is convenient to construct these matrices in the reciprocal space where one benefits from their block-diagonal structure due to the **k**-vector symmetry. To this end, the OSV coefficients are first Fourier transformed to the reciprocal space,

$$C^{\text{OSV}}_{\mu,\tilde{a}[i]}(\mathbf{k}) = \sum_{\mathcal{M}} C^{\text{OSV}}_{\mu\mathcal{M},\tilde{a}[i]} \exp\left(i\mathbf{k}\mathbf{R}_{\mathcal{M}}\right). \quad (12)$$

The reciprocal images

$$S_{\mu\nu}(\mathbf{k}) = \sum_{\mathcal{N}} S_{\mu,\nu\mathcal{N}} \exp\left(-i\mathbf{k}\mathbf{R}_{\mathcal{N}}\right) \quad (13)$$

of the AO overlap $S_{\mu\nu\mathcal{N}}$ matrices are then transformed into the OSV basis according to

$$S_{\tilde{a}[i],\tilde{b}[j]}(\mathbf{k}) = \sum_{\mu\nu} \left(C^{\text{OSV}}_{\mu,\tilde{a}[i]}(\mathbf{k})\right)^{\dagger} S_{\mu\nu}(\mathbf{k}) C^{\text{OSV}}_{\nu,\tilde{b}[j]}(\mathbf{k}) \quad (14)$$

and transformed back thereafter to the direct space

$$S_{\tilde{a}[i]\mathcal{I},\tilde{b}[j]} = \frac{1}{N_{\mathbf{k}}} \sum_{\mathbf{k}} S_{\tilde{a}[i],\tilde{b}[j]}(\mathbf{k}) \exp\left(-i\mathbf{k}\mathbf{R}_{\mathcal{I}}\right). \quad (15)$$

The external Fock matrix in OSV basis $F_{\tilde{a}[i]\mathcal{I},\tilde{b}[j]}$ is obtained analogously.

The attentive reader may note a subtlety in the signs of the exponents in Eqs. (12)–(15). For example, one may wonder why the direct Fourier transform (FT) of the overlap matrix **S** from direct to reciprocal space, Eq. (13) has the same sign in the exponential as the back FT of its Fourier image from reciprocal to direct space, Eq. (15). The reason for that is that the transformation indices differ in these two equations, i.e., in Eq. (13) it is the `ket` index, whereas in Eq. (15) it is the `bra` index that is transformed. The FT of the OSV coefficient matrix $\mathbf{C}^{\text{OSV}}$, on the other hand, has a different sign in the exponential than the one of **S**, even though both are direct FTs. As before, the transformation indices differ. Furthermore, the $(\mathbf{C}^{\text{OSV}})^{\dagger}$ transforms conversely to **S**, implying another change of sign in the exponential.[43,91] Finally, formation of the adjoint,

i.e., $(\mathbf{C}^{\text{OSV}})^{\dagger}$ from $(\mathbf{C}^{\text{OSV}})$, implies complex conjugation and hence a third change of sign in the exponential.

As already mentioned above, and as will be demonstrated in Sec. III, the representation of the pair virtual spaces by OSVs augmented with a few diffuse PAOs is much more compact than that of PAOs alone. The resulting set of amplitudes hence is substantially smaller than in the PAO basis, which leads to substantial speedups and savings in memory when solving the LMP2 equations. This also helps to avoid input-output overhead, since the whole amplitude set can usually be kept in memory, whereas in the PAO case the algorithm often has to page amplitudes.

On the other hand, the overall external overlap and Fock matrices are more cumbersome with respect to memory consumption in the OSV case, since the overall number of OSVs (plus additional diffuse PAOs) by far exceeds the number of PAOs. Indeed, the number of pseudo-atoms, i.e., WFs, are usually larger than the number of atoms per cell, while the number of OSVs per pseudo-atom, which remain after truncation with a reasonable threshold, plus additional diffuse PAOs, are also larger than the number of PAOs per atom. Consequently, the overall external overlap and Fock matrices consume considerably more memory than the PAO counterparts and may constitute a potential memory bottleneck in some calculations.

In order to avoid that, the individual blocks are pre-screened and the external Fock and overlap matrices kept in core in sparse form. The external Fock matrix is only involved in matrix multiplications involving identical pair domains on column and row side (no pair couplings), which restricts the required blocks to pairs of pseudo-atoms of the restricted WF pair list. The overlap matrix, on the other hand, also couples OSVs corresponding to different pairs, i.e., living in different pair domains.[25] The required blocks, therefore, are not confined to the restricted WF pair list. In order to reduce the number of blocks such that also the overlap matrix can be kept in core, we employ prescreening: the size of the individual contribution of a certain block of the overlap matrix to the residual is estimated on the basis of the maximal values of the zeroth-iteration amplitudes for each pair, the values of the *internal* Fock matrix (coupling individual WF pairs), and the maximum values of the OSV overlap matrix for different pseudo-atom pairs. By virtue of this prescreening procedure, the external Fock and overlap matrices in OSV basis can easily be kept in core, keeping the memory consumption of the OSV LMP2 equation solver rather low.

III. TEST CALCULATIONS

A. Test systems

For testing the performance of the OSV-based periodic LMP2 code we considered several systems, which exemplify common applications of the periodic LMP2 method. Emphasis was put on the long-range van der Waals correlation, which is problematic for DFT and thus requires a proper quantum chemical treatment for accurate description. Moreover, as already discussed above, the OSVs themselves do not provide an optimal virtual basis for long-range correlation. Hence, it is crucial to test the effectiveness of the approach on those

systems, where long-range correlation provides an essential contribution to binding. To this end, we considered the following crystals: (i) the high (β-) and low temperature (α-) phases of the GeF$_2$ crystal;[6] (ii) the molecular crystal CO$_2$;[1,15] and (iii) adsorption of argon atoms on the MgO (001) surface.[14] Additionally, we also included the germanium crystal, which is an example of a densely packed system with relatively delocalized WFs. In all the calculations, basis sets of triple-zeta quality were employed, which for CO$_2$, Ar, and GeF$_2$ were augmented with d- and f-diffuse orbitals. These diffuse polarization functions are essential for an accurate description of dispersion.[1] For a detailed specification of the computational parameters, we refer to supplementary material.[92]

The GeF$_2$ crystal has a complex binding pattern: the covalently bound GeF$_2$ chains are stacked together by intermolecular forces with a large contribution from dispersion.[6] The two studied polymorphs have different symmetries, so the relative energy between them can show how sensitive the proposed approach is to the change in the symmetry. Besides that, the calculation of the lattice energy of GeF$_2$, which can be defined with reference to an individual GeF$_2$ molecule, demonstrates the influence of changes in the bonding structure (in the crystal each Ge atom has 3 bonds with F) on the stability of the approach. Finally, GeF$_2$ is also a challenging test system in terms of the number of correlated electrons: together with the d-electrons of the Ge atoms included in the correlation treatment, it has 116 "valence" electrons per cell, which in a periodic LMP2 calculation with a standard pair cut-off distance of 12 Å are correlated with more than 4000 electrons. For β-GeF$_2$ that corresponds to 59 136 WF pairs.

B. Comparison of OSV and PAO performance

We start with the general analysis of the performance of the OSV-based LMP2 method with different truncation tolerances on the germanium crystal. The minimal PAO domains (*vide supra*) for this crystal consist of two atoms and contain, within the basis set employed, 94 PAOs. For LMP2 applications, PAO domains of this size are usually not sufficient. The next possible PAO domains (still reflecting the symmetry of the crystal) contain 8 atoms or 376 PAOs, which becomes already computationally rather demanding. E.g., the size of the doubles amplitude buffer amounts in this case to 10 GB. The next possible choices for the domains are 20 and 32 atoms, which correspond to more than 60 GB of doubles amplitudes, and cannot be handled in a PAO-based LMP2 calculation without inclusion of point-group symmetry.[25] In Fig. 1, it is shown how the correlation energy converges w.r. to the PAO domain size (horizontal dashed lines).

The LMP2 energy, calculated with OSVs constructed from 8-atom PAO domains, obviously has the lower bound of the 8-atom PAO LMP2 energy. It closely approaches this value for an OSV truncation tolerance of $\mathcal{E}_{OSV} = 10^{-7}$. On the other hand, for tolerances $\mathcal{E}_{OSV} < 10^{-6}$ it becomes increasingly difficult to fit the ERIs accurately: for $\mathcal{E}_{OSV} = 10^{-7}$ (and 8-atom initial PAO domains), the correct energy can be obtained only with a fit basis designed for a quadruple-zeta (rather than a triple-zeta) AO basis. Therefore, even though the pair-specific virtual OSV basis constructed with $\mathcal{E}_{OSV} = 10^{-7}$ is substan-
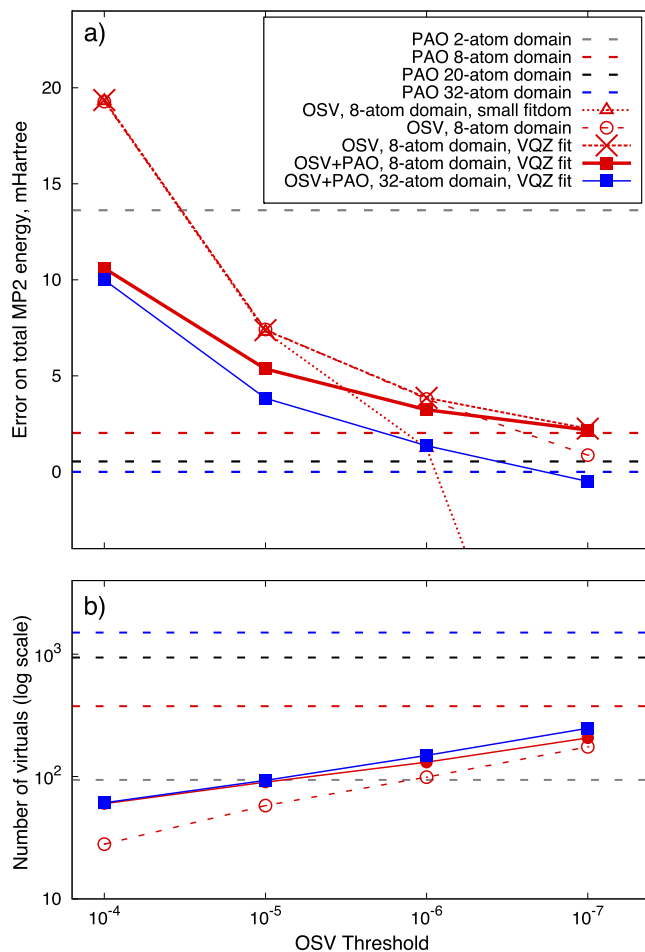

FIG. 1. The LMP2 correlation energy for the germanium crystal, calculated with PAOs and OSVs as a function of the OSV truncation tolerance (panel (a)). Panel (b) shows the corresponding number of virtual orbitals (PAOs, OSVs, or OSVs plus few PAOs) per WF. The 2-, 8-, 20-, and 32-atoms are the possible choices for the PAO domains in this high symmetry crystal. In the OSV calculations, the fit-domains were constructed by the procedure described in Sec. II C, and comprise either 4 atoms (denoted as "small fitdom"), 8 atoms for VQZ fit, or 15 atoms otherwise. OSV + PAO denotes the OSV virtual basis augmented with the most diffuse PAOs from the minimal domain as described in Sec. II B.

tially smaller than the 8-atom PAO domain, the difficulties with achieving a stable fit renders such thresholds as impractical (*vide infra*). On the other hand, with tolerances of $\mathcal{E}_{OSV} = 10^{-4}$ or $\mathcal{E}_{OSV} = 10^{-5}$ the fit even with very small fit domains is quite accurate, yet the deviation of the correlation energy from the 8-atom PAO lower bound still is rather large. A substantial improvement is achieved by addition of diffuse PAOs from the minimal (in this case 2-atom) domains. Most importantly, the additional fraction of correlation energy so obtained is to a large extent long-range van der Waals correlation (*vide infra*). At the same time, these extra PAOs become less relevant for OSV calculations with tighter truncation tolerances.

A further lowering of the correlation energy is achieved by expansion of the initial PAO domain, say, by constructing the OSVs on the basis of a 32-atom PAO calculation instead. In this case, the OSV-LMP2 energy with $\mathcal{E}_{OSV} = 10^{-6}$ tolerance (plus diffuse PAOs) is even below the 8-atom PAO energy. Yet fitting in this case becomes more difficult: even a fitting basis for quadruple zeta AO basis sets in conjunction with 8-atom

TABLE I. LMP2 correlation contributions to the interaction energy of Ar adsorbed on the MgO (100) surface (at minimum energy separation) and to the lattice energies of the CO_2 and GeF_2 crystals. For Ar–MgO, the correlation interaction energy was partitioned into intra-adsorbate ($\Delta E_{intra-Ar}$), intra-slab ($\Delta E_{intra-MgO}$), and inter-Ar–MgO (ΔE_{inter}) components.[14,93] The lattice energies of the CO_2 and GeF_2 crystals were defined with reference to the CO_2 (unrelaxed) and GeF_2 (relaxed at the MP2 level) molecules,[92] respectively. For the α and β phases of GeF_2, the LMP2 correlation part of the relative stability is also given. This quantity was calculated as the difference between either the total correlation energies E_{corr}, or the correlation parts of the lattice energies ΔE_{lat}. For GeF_2, two AO basis sets of triple- and augmented-triple-zeta quality were used (see supplementary material[92] for the exact specification). In the OSV calculations, the WF specific virtual spaces were generated according to an $\mathcal{E}_{OSV}$ tolerance of 10^{-4} or 10^{-5}, without or with additional diffuse PAOs (denoted as OSV + PAOs). The domains in the PAO calculations were defined either automatically by the Boughton-Pulay method with the tolerance $T_{BP} = 0.99$ or explicitly ("denoted as extended domains") by specifying the number of atoms for each WF (see supplementary material[92] for the exact specification). The OSVs were generated from a 25-atom PAO domain. All energies are given in kJ/mol.

	OSVs		OSVs + PAOs		PAOs	
	$\mathcal{E}_{OSV} = 10^{-4}$	$\mathcal{E}_{OSV} = 10^{-5}$	$\mathcal{E}_{OSV} = 10^{-4}$	$\mathcal{E}_{OSV} = 10^{-5}$	$T_{BP} = 0.99$	Extended domains
Ar–MgO						
ΔE_{inter}	−6.159	−7.554	−7.533	−8.000	−7.835	−7.921
$\Delta E_{intra-MgO}$	+0.967	+1.104	+1.175	+1.140	+1.205	+1.155
$\Delta E_{intra-Ar}$	+0.178	+0.187	+0.191	+0.191	+0.211	+0.211
CO_2						
ΔE_{lat}	−19.15	−22.88	−24.76	−25.71	−25.72	−26.37
GeF_2, VTZ-basis						
ΔE_{lat}^{α}	−15.93	−23.79	−25.43	−27.70	−23.79	−29.23
ΔE_{lat}^{β}	−12.48	−20.46	−21.99	−24.22	−18.00	−26.60
$\Delta E^{\alpha} - \Delta E^{\beta}$	−2.61	−3.32	−3.17	−3.43	−2.50	−1.69
$\Delta E_{lat}^{\alpha} - \Delta E_{lat}^{\beta}$	−3.45	−3.33	−3.43	−3.47	−5.79	−2.63
GeF_2, AVTZ-basis						
ΔE_{lat}^{α}				−43.49		−46.65
ΔE_{lat}^{β}				−40.64		−42.58
$\Delta E_{corr}^{\alpha} - \Delta E_{corr}^{\beta}$				−3.33		−5.58
$\Delta E_{lat}^{\alpha} - \Delta E_{lat}^{\beta}$				−2.85		−4.07

fit domains is not sufficient to obtain a correct LMP2 energy above the PAO 32-atom domain bound with $\mathcal{E}_{OSV} = 10^{-7}$ OSV tolerance.

To summarize, OSVs constructed from large PAO domains with tolerance $\mathcal{E}_{OSV} = 10^{-5}$, and augmented by additional diffuse PAOs belonging to the minimal domains, can be considered to be the optimal choice with respect to accuracy vs. efficiency. In case of Ge this yields a correlation energy close to that of the 8-atom domain PAO calculation.

Similar results are also obtained for energy differences (interaction and lattice energies), as shown in Table I. The additional PAOs are indeed essential to capture the long-range correlation: comparison with PAO calculations using extended domains shows that the van der Waals dominated interaction energies of these systems are quite well described by utilizing an OSV basis generated with $\mathcal{E}_{OSV} = 10^{-5}$ and augmented with diffuse PAOs as described above. On the other hand, without diffuse PAOs the OSV results are rather poor. Yet in GeF_2, even with inclusion of diffuse PAOs the $\mathcal{E}_{OSV} = 10^{-5}$ OSV calculations underestimate the lattice energy by about 2-3 kJ/mol, compared to the extended domain PAO calculation (which, on the other hand, is considerably more expensive, *vide infra*). This discrepancy can be considered as acceptable, especially taking into account the size of the total lattice energy, of which the HF contribution alone amounts to about −40 kJ/mol.

The relative stability between the α and β phases of GeF_2 is a much more delicate quantity than the lattice energy, since it amounts only to a few kJ/mol. And here the OSV based calculations turn out to be much more stable than those with

PAOs. As is evident from Table I, the OSV results are quite insensitive to the domain or basis set size. Furthermore, virtually the same result is achieved, if the relative energy is evaluated from the total energies [contaminated by basis set superposition error (BSSE)] or the lattice energies, where the BSSE is counterpoise corrected. At the same time, the deviations among the different PAO results are substantially larger. Whether OSVs are generally superior to PAOs for computing relative stabilities remains, however, an open question and requires further investigation.

Finally, we note that the change of the bonding structure from the crystal (each Ge atom has three bonds) to the molecule (Ge atom has two bonds) leads to different minimal domains (with the default minimal domain BP tolerance of 0.9), and thus to different numbers of additional PAOs added to the OSVs for certain pairs in crystal and molecule, respectively. This could potentially be a drawback of the OSV plus diffuse PAOs scheme, which again becomes dependent on the choice of the domains. However, our calculations show that in GeF_2 the mismatch between the virtual spaces has only a minor effect on the lattice energy, which changes only by −0.3 kJ/mol, if the minimal domains in the crystal are adjusted to those in the monomer. Yet this discrepancy has no effect on the smoothness of the potential energy surface of the crystal, since it originates from a domain mismatch between crystal and isolated monomer, rather than individual crystal geometries.

Another important aspect of the OSV scheme is the size of the initial PAO domains needed for the construction of the OSVs. As already discussed above in the context of the Ge
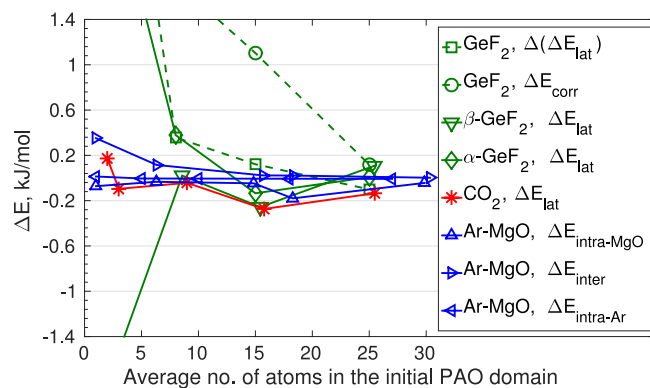
FIG. 2. The deviations of various LMP2 correlation energy differences for different choices of the initial PAO domain from the reference calculations with 35-atom initial PAO domain (or larger if dictated by symmetry[92]). The following quantities were considered: (i) the correlation parts of the lattice energies of the CO_2 and GeF_2 crystals (ΔE_{lat}), (ii) the intra-slab ($\Delta E_{intra-MgO}$), intra-adsorbate ($\Delta E_{intra-MgO}$), and inter-Ar–MgO (ΔE_{inter}) components of the correlation contribution to the adsorption energy of Ar on the MgO (100) surface, and (iii) the correlation part of the relative stability calculated either as the difference of the lattice energies $\Delta(\Delta E_{lat})$ or the total correlation energies ΔE_{corr}.

crystal, expansion of the initial PAO domains leads to a lowering of the total correlation energy. Besides that, large PAO domains are essential for smearing out the energy fluctuations due to a mismatch in the virtual space for different points on the potential surface or in the reference structures, when evaluating energy differences. The PAO correlation energy with large domains indeed approaches the canonical result (see Ref. 65 or the sequence of the PAO LMP2 energies in Fig. 1). Therefore, the larger the domains, the less harmful the mismatch in the domain sizes for the smoothness of the LMP2 potential surface.

Expansion of the PAO domains for the OSV generation is obviously much easier than for pure PAO-based LMP2 calculations (since in the former case only diagonal pairs are treated with these large PAO domains), but not entirely effortless. Therefore, it is important to find an optimal range for the domain sizes. Fig. 2 shows the deviations of various correlation energy differences, i.e., the correlation part of (i) the adsorption energy (Ar–MgO), (ii) the lattice energy (CO_2 and GeF_2), and (iii) the relative stability (GeF_2) from the related reference values. The related reference values are OSV results with large initial PAO domains of 35 atoms or larger, as dictated by crystalline structure.[92]

For CO_2 or Ar–MgO, the initial PAO domain size is not so critical, but the relative stability of the GeF_2 phases is very sensitive to it, especially that computed from the total correlation energies. This correlates with the strong dependence of this quantity on the domain size in the pure PAO calculations as reported in Table I. In any case, for domains larger than 25 atoms, all the considered quantities are converged within ±0.2 kJ/mol.

To analyze the efficiency of the new approach in comparison to the PAO-LMP2 method, we consider GeF_2 in the VTZ basis (cf. Table II). For this system, solving of the LMP2 equations constitutes the computational bottleneck of the PAO calculation: this step alone takes more than half of the overall time. Furthermore, the memory requirements become quite large: already with small domains (BP threshold of 0.99), which according to the results of Table I are by far not yet sufficient for an accurate description, the size of the amplitude buffer is nearly 5 GB. For extended domains it increases to 27 GB. In the OSV case, on the other hand, the LMP2 equations become computationally inexpensive, with relatively modest memory requirements due to the compactness of the OSV-based pair specific virtual spaces; the amplitudes require just slightly more than 3 GB. Furthermore, the overlap and Fock matrices, which become generally quite large in the OSV case, require after prescreening (cf. Sec. II D) not more than 1 GB each. The dominant step in the OSV calculation is clearly the integral evaluation, but even this step is faster than in the PAO case. That makes the OSV overall less expensive than the PAO calculations, despite the overhead due to the initial OSV generation, for which an additional large-domain PAO calculation for diagonal pairs has to be carried out.

C. Accuracy of ERIs

Next we investigate accuracy and stability of the ERI evaluation by DF or multipole expansion. The latter technique is very efficient, but applicable only to ERIs which involve non-penetrating densities. In the PAO-based periodic LMP2 method, those pairs treated by the multipole approximation (the so-called distant pairs) are determined according to a distance criterion R_{dist} related to the spacing between the two WFs. However, as seen in Fig. 3(a), a distance criterion beyond which the multipole approximation can safely be invoked, is quite system dependent. For example, for the semiconductor Ge, R_{dist} should be at least set to 11 Å, while for the molecular CO_2

TABLE II. Elapsed computational times in hours for OSV- and PAO-based LMP2 calculations (and main individual steps thereof) for α-GeF_2 with VTZ orbital basis, using a serial compilation of the code linked with explicitly parallelized BLAS libraries on an 8-core AMD Opteron 6180 SE node. The computational parameters are given in the supplementary material.[92]

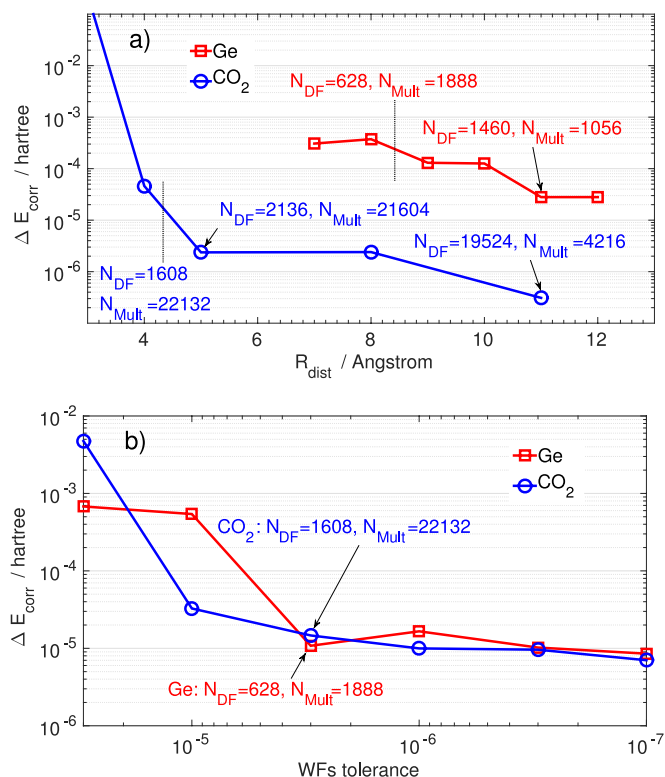|	PAOs		OSVs + PAOs
	$T_{BP} = 0.99$	Extended domains	$\varepsilon_{OSV} = 10^{-5}$
OSV or PAO generation	0.3	0.3	7.1
Multipolar integrals	0.5	1.3	0.5
DF integrals	14.5	19.9	11.0
LMP2 equations	24.0	46.6	4.5
Total	39.3	68.1	23.1

FIG. 3. The OSV LMP2 correlation energies for the germanium and CO_2 crystals as a function of the distant-pair tolerance. In panel (a), the distance criterion is used; in panel (b) the threshold for the new O_{ijJ} estimate [Eq. (6)] denoted as the WF tolerance. The amounts of pairs treated by density fitting (N_{DF}), and by the multipole approximation (N_{Mult}) (i.e., the number of distant pairs) are also given for the relevant values of the tolerances. In panel (a), the vertical lines indicate the distant pair cutoffs, providing the same number of distant pairs as in the new scheme (panel (b)) with the tolerance 3×10^{-6}.

crystal an R_{dist} of 5 Å is already sufficient. Thus, if $R_{dist} = 11$ Å would be specified as the general default value (as dictated by the case of Ge), it would be much too large for the CO_2 case and similar systems, making calculations on such systems unnecessarily costly. Besides that, it is very likely that an R_{dist} value of 11 Å is still too small for a system with even narrower band gap than Ge. Ineffectiveness of the distance criterion to define the distant-pair range has also recently been noted in the context of new local correlation approaches for molecules.[70,75]

The new distant pair criterion introduced in Sec. II C is much more adequate in this respect. It depends on the actual spread and orientation of the WFs and thus adapts to the system under study. This is illustrated in Fig. 3(b), where the same two systems CO_2 and Ge are subject to the new treatment. The error in the correlation energy drops to the 10^{-5} hartree region at approximately the same value of the threshold for the O_{ijJ} quantity defined in Eq. (6). The O_{ijJ} quantity turns out to be a much better criterion for specifying distant pairs than R_{dist}, substantially increasing the amount of ERIs, which can safely be treated by the multipole expansion.

For the density fitting scheme itself, we also have introduced a new procedure for the generation of the local fit domain, as is explained in Sec. II C. Table III compiles the total correlation energies and energy differences for all the test systems. Despite its simplicity the new scheme turns out to be very efficient. Already with 6-atom fit-domains the fitting error

becomes less than 10^{-4} hartree per atom for the total energy and falls below 0.1 kJ/mol for energy differences. The largest error in the total correlation energy is observed for GeF_2, most likely due to the d-electrons of germanium, which were included in the correlation treatment of this system. However, even in this case the error is below 10^{-4} hartree per atom, and, even more importantly, it virtually vanishes for the energy differences.

The possibility to employ such small fit-domains is important for reasons of efficiency. Otherwise, due to the unfavorable cubic scaling (with the number of fitting functions) of the linear equation solver, calculation of the fitting coefficients in Eq. (9) may become the computational bottleneck. It has already been mentioned above that the necessary size of the fit domains for a reasonable fit of the ERIs sensitively depends on the OSV threshold $\mathcal{E}_{OSV}$. Tightening this threshold implies considerably larger fit domains. For that reason, the use of tight $\mathcal{E}_{OSV}$ becomes computationally expensive and thus inconvenient in applications. As already stated above, we recommend a $\mathcal{E}_{OSV}$ value of 10^{-5} and augmentation of the OSV basis by diffuse PAOs.

D. Smoothness of potential energy surfaces

Finally, we test how well a black box OSV-LMP2 approach works for potential surfaces in comparison to PAO-LMP2. To this end we calculated the potential energy curve of the Ar–MgO interaction along the perpendicular distance between the argon monolayer and the MgO slab. The inter-argon separation within the chosen 2×2 monolayer is such that the Ar–Ar interaction is small and not relevant in the present context.[14]

The Ar–MgO interaction potential is very shallow, and even small discontinuities become apparent at this scale. Moreover, since van der Waals dispersion is the dominant attractive force (the contribution due to induction is much smaller), this test shows if van der Waals dispersion indeed is accurately captured by the OSV-LMP2 method with the default set of the parameters specified in accordance with the results of above tests, i.e., with the virtual space spanned by OSVs with truncation tolerance $\mathcal{E}_{OSV} = 10^{-5}$ plus diffuse PAOs, 25 atoms in the initial PAO domain, 8 atoms in the fit domain, and distant pair tolerance set to 10^{-6}. Fig. 4 displays the OSV-LMP2 potential energy curve and, for reference, two curves calculated with the PAO-LMP2 approach. The first one was obtained with extended and individually specified domains (1 atom for WFs on Ar, and 6 or 7 atoms for WFs on MgO, depending if the WF is located on the surface or in the internal layer, respectively).[92] Such a specification of the virtual space delivers accurate and smooth potential energy surfaces and was actually employed in a previous application, i.e., the study of geometrical frustration of an Ar monolayer on the MgO surface.[14] Yet, it is not a black-box approach since an explicit specification of individual orbital domains is necessary, rather than just providing a threshold. The second PAO curve was calculated with "black-box" PAO domains obtained from the BP procedure with a criterion of 0.99. In Fig. 4, in addition to the total interaction energies, also the intra-MgO, intra-Ar, and inter-Ar–MgO correlation components thereof are plotted.

Evidently, the "black box" PAO approach is not safe due to a possible mismatch in the corresponding pair-specific

TABLE III. The LMP2 total correlation energies (in hartree) and correlation contributions to the interaction energies (in kJ/mol), calculated with OSVs ($\mathcal{E}_{OSV} = 10^{-5}$ plus diffuse PAOs) with different fit-domains and fitting basis sets. The corresponding computational times $T_{DF\text{-}coef.}$ for solving the density fitting equations [Eq. (9)] are also given (in hours, and as the fraction of the overall computational time T_{LMP2}, for Ge and CO_2 calculations only). The specifications of the fitting basis set can be found in the supplementary material.[92]

Fit basis	VTZ						V5Z
No. of atoms in fit domain	4	6	8	10	15	20	8
Ge							
E_{corr}, hartree	−0.227 878	−0.227 645 3	−0.227 644 7	−0.227 607 3	−0.227 606 0	−0.227 566 8	−0.227 606 8
$T_{DF\text{-}coef.}$, h	0.5	1.1	1.7	6.5	9.9	16.9	5.5
$T_{DF\text{-}coef.}/T_{LMP2}$, %	9.0	19.4	27.1	55.8	65.1	76.0	39.1
CO_2							
E_{corr}, hartree	−2.486 963	−2.486 953	−2.487 055	−2.486 922	−2.486 911	−2.486 875	−2.486 867
ΔE_{lat}, kJ/mol	−25.705 0	−25.698 9	−25.710 9	−25.697 1	−25.698 1	−25.702 6	−25.710 0
$T_{DF\text{-}coef.}$, h	0.2	0.4	0.9	2.3	5.2	12.7	5.5
$T_{DF\text{-}coef.}/T_{LMP2}$, %	1.6	3.0	6.1	13.6	26.2	45.5	13.8
Ar–MgO							
E_{corr}, hartree	−1.932 688 7	−1.932 637 0	−1.932 705 3	−1.932 703 1	−1.932 702 6	−1.932 701 7	−1.932 748 0
ΔE_{inter}, kJ/mol	−7.650	−7.887	−7.979	−7.980	−8.000	−8.002	−8.002
$\Delta E_{intra\text{-}MgO}$, kJ/mol	1.139 6	1.140 5	1.139 8	1.139 8	1.139 8	1.139 5	1.135 0
$\Delta E_{intra\text{-}Ar}$, kJ/mol	0.190 8	0.190 8	0.190 8	0.190 8	0.190 8	0.190 8	0.190 5
GeF_2, VTZ-basis							
E_{corr}^{α}, hartree	−2.744 162	−2.744 191	−2.744 200	−2.744 241	−2.744 285	−2.744 295	−2.743 453
ΔE_{lat}^{α}	−18.186 1	−18.199 0	−18.196 8	−18.198 3	−18.196 1	−18.194 7	−18.151 6
$E_{corr}^{\alpha} - \Delta E_{corr}^{\beta}$	−3.543 1	−3.450 6	−3.427 2	−3.449 7	−3.453 6	−3.450 7	−3.415 6

virtual spaces of the full system vs. slab or monolayer alone. Moreover, if not kept fixed for different Ar–MgO distances (which is not done in these calculations), there is also a possible mismatch in the virtual spaces from point to point along the curve. As can be seen, this mismatch mainly affects the intra-slab correlation component of the interaction energy and causes non-physical steps on the potential surface, quite large at the scale of the interaction of this system. Obviously, this problem is absent if the domains are individually defined to be the same in all the calculations.
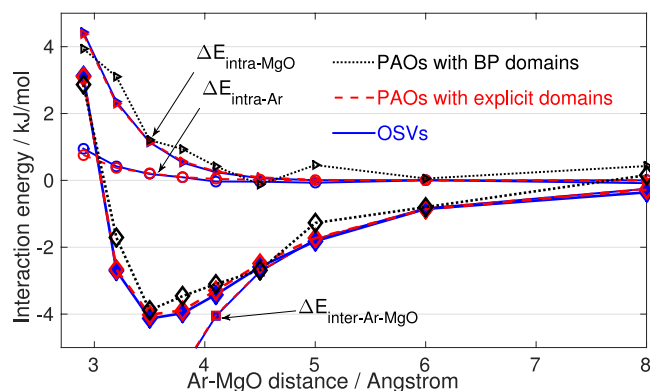

FIG. 4. The HF + LMP2 (OSV or PAO) adsorption energy (diamonds) for Ar on the MgO (100) surface (on top of the Mg atom),[14,92] as a function of the Ar–Mg distance. In the PAO LMP2 case, the PAO domains were determined by the BP procedure with the tolerance of 0.99 (black dotted lines), or defined explicitly to consist of 6 (or 7 for the internal layer) atoms for the oxygen WFs and 1 atom for the argon WFs (red dashed lines). The OSV curves (blue solid lines) were obtained with $\mathcal{E}_{OSV} = 10^{-5}$ plus diffuse PAOs from the minimal domains. The intra-slab (triangles), intra-adsorbate (circles), and inter-slab-adsorbate (squares) components of the correlation part of the interaction energy are also shown.

The OSV approach, on the other hand, delivers interaction energies very close to those obtained with the individual explicit specification of the PAO orbital domains. Indeed, the most problematic intra-slab component in Fig. 4 is in the OSV case virtually indistinguishable from that with explicit PAO domain specification, while it is full of kinks when the automatic BP PAO domain approach is employed instead. At the same time, the OSV approach remains a black-box method without the requirement for a tedious specification of individual orbital domains. We note that there are still small discontinuities in the OSV-LMP2 curve caused by slight mismatch of corresponding virtual spaces (this appears to be unavoidable in blind truncation procedures). However, these discontinuities are much smaller than those of the PAO-LMP2 calculations with BP domains and hardly visible anymore at the scale of even this small interaction energy.

IV. CONCLUSIONS

A new periodic local MP2 method has been presented. It employs orbital specific virtuals to represent the pair specific virtual spaces. OSVs are the eigenvectors of the doubles amplitude matrices of diagonal Wannier function pairs.

The OSV-LMP2 method has a number of advantages over our previous PAO-based implementation. First, virtual space truncation is controlled by a single parameter. This turns the OSV-LMP2 method into a black-box approach, which, in contrast to its PAO based precursor, no longer requires a tedious input for PAO domain specification. At the same time, PAO calculations employing the automatic Boughton-Pulay scheme can be prone to large discontinuities in potential surfaces, which originate from a mismatch of the virtual spaces

from one point to the other. This problem virtually vanishes when OSVs are used. Second, due to the higher compactness of the virtual space OSV-LMP2 is faster (especially for the LMP2 equation solver), and has a considerably smaller memory footprint than its PAO based predecessor. This is also very important for efficient parallelization, which is simpler if the amplitude buffer can be replicated.[27]

We note that OSVs also have certain deficiencies. For example, it is much more difficult to incorporate point group symmetry in the OSV treatment, which, in case of PAOs can lead to considerable savings for highly symmetric (e.g., cubic) crystals. Next, since the OSVs are natural orbitals for diagonal pairs, the truncated OSV basis has a bias towards short-range correlation. Van der Waals long-range correlation, which is often essential in solids,[2] is less well described by OSVs. We have demonstrated that this problem can be circumvented by augmenting the OSV-based virtual space associated with each Wannier function by the few most diffuse PAOs of the corresponding minimal domain.

ACKNOWLEDGMENTS

[1] L. Maschio, D. Usvyat, M. Schütz, and B. Civalleri, J. Chem. Phys. **132**, 134706 (2010).
[2] M. Halo, C. Pisani, L. Maschio, S. Casassa, M. Schütz, and D. Usvyat, Phys. Rev. B **83**, 035117 (2011).
[3] L. Maschio, D. Usvyat, and B. Civalleri, CrystEngComm **12**, 2429 (2010).
[4] R. Martinez-Casado, G. Mallia, D. Usvyat, L. Maschio, S. Casassa, M. Schütz, and N. M. Harrison, J. Chem. Phys. **134**, 014706 (2011).
[5] J. T. Tanskanen, L. Maschio, A. J. Karttunen, M. Linnolahti, and T. A. Pakkanen, ChemPhysChem **13**, 2361 (2012).
[6] D. Usvyat, C. Yin, G. Wälz, C. Mühle, M. Schütz, and M. Jansen, Phys. Rev. B **86**, 054102 (2012).
[7] A. J. Karttunen and T. F. Fässler, Chem. - Eur. J. **20**, 6693 (2014).
[8] G. Constantinescu, A. Kuc, and T. Heine, Phys. Rev. Lett. **111**, 036104 (2013).
[9] M. D. Ben, M. Schönherr, J. Hutter, and J. VandeVondele, J. Phys. Chem. Lett. **4**, 3753 (2013).
[10] F. Göltl, A. Grüneis, T. Bucko, and J. Hafner, J. Chem. Phys. **137**, 114111 (2012).
[11] H. Stoll, Phys. Rev. B **46**, 6700 (1992).
[12] B. Paulus, Phys. Rep. **428**, 1 (2006).
[13] E. Voloshina, D. Usvyat, M. Schütz, Y. Dedkov, and B. Paulus, Phys. Chem. Chem. Phys. **13**, 12041-12047 (2011).
[14] D. Usvyat, K. Sadeghian, L. Maschio, and M. Schütz, Phys. Rev. B **86**, 045412 (2012).
[15] C. Müller and D. Usvyat, J. Chem. Theory Comput. **9**, 5590 (2013).
[16] 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).
[17] R. Martinez-Casado, D. Usvyat, L. Maschio, G. Mallia, S. Casassa, J. Ellis, M. Schütz, and N. M. Harrison, Phys. Chem. Chem. Phys. **16**, 21106 (2014).
[18] J. Yang, W. Hu, D. Usvyat, D. Matthews, M. Schütz, and G. K.-L. Chan, Science **345**, 6197 (2014).
[19] M. Marsman, A. Grüneis, J. Paier, and G. Kresse, J. Chem. Phys. **130**, 184103 (2009).
[20] A. Grüneis, G. H. Booth, M. Marsman, J. Spencer, A. Alavi, and G. Kresse, J. Chem. Theory Comput. **7**, 2780 (2011).
[21] G. Booth, A. Grüneis, G. Kresse, and A. Alavi, Nature **493**, 365 (2012).
[22] A. Grüneis, J. J. Shepherd, A. Alavi, D. P. Tew, and G. H. Booth, J. Chem. Phys. **139**, 084112 (2013).
[23] M. D. Ben, J. Hutter, and J. VandeVondele, J. Chem. Theory Comput. **8**, 4177 (2012).
[24] M. D. Ben, J. Hutter, and J. VandeVondele, J. Chem. Theory Comput. **9**, 2654 (2013).
[25] C. Pisani, L. Maschio, S. Casassa, M. Halo, M. Schütz, and D. Usvyat, J. Comput. Chem. **29**, 2113 (2008).
[26] C. Pisani, M. Schütz, S. Casassa, D. Usvyat, L. Maschio, M. Lorenz, and A. Erba, Phys. Chem. Chem. Phys. **14**, 7615 (2012).
[27] L. Maschio, J. Chem. Theory Comput. **7**, 2818 (2011).
[28] D. Usvyat, J. Chem. Phys. **139**, 194101 (2013).
[29] V. P. Smirnov and D. E. Usvyat, Phys. Rev. B **64**, 245108 (2001).
[30] D. Usvyat and M. Schütz, Theor. Chem. Acc. **114**, 276 (2005).
[31] J. Almlöf, Chem. Phys. Lett. **181**, 319 (1991).
[32] M. Häser and J. Almlöf, J. Chem. Phys. **96**, 489 (1992).
[33] P. Y. Ayala, K. N. Kudin, and G. E. Scuseria, J. Chem. Phys. **115**, 9698 (2001).
[34] B. Doser, D. Lambrecht, J. Kussmann, and C. Ochsenfeld, J. Chem. Phys. **130**, 064107 (2009).
[35] S. A. Maurer, D. S. Lambrecht, J. Kussmann, and C. Ochsenfeld, J. Chem. Phys. **138**, 014101 (2013).
[36] B. Jansìk, S. Høst, K. Kristensen, and P. Jørgensen, J. Chem. Phys. **134**, 194104 (2011).
[37] P. Pulay, Chem. Phys. Lett. **100**, 151-154 (1983).
[38] S. Saebø and P. Pulay, Chem. Phys. Lett. **113**, 13 (1985).
[39] P. Pulay and S. Saebø, Theor. Chim. Acta **69**, 357-368 (1986).
[40] M. Schütz, G. Hetzer, and H.-J. Werner, J. Chem. Phys. **111**, 5691 (1999).
[41] M. Schütz, H.-J. Werner, R. Lindh, and F. R. Manby, J. Chem. Phys. **121**, 737 (2004).
[42] H.-J. Werner, F. R. Manby, and P. J. Knowles, J. Chem. Phys. **118**, 8149 (2003).
[43] D. Kats, D. Usvyat, and M. Schütz, Phys. Chem. Chem. Phys. **10**, 3430 (2008).
[44] S. Loibl and M. Schütz, J. Chem. Phys. **137**, 084107 (2012).
[45] C. Hampel and H.-J. Werner, J. Chem. Phys. **104**, 6286 (1996).
[46] M. Schütz, J. Chem. Phys. **113**, 9986 (2000).
[47] M. Schütz and H.-J. Werner, Chem. Phys. Lett. **318**, 370 (2000).
[48] M. Schütz and H.-J. Werner, J. Chem. Phys. **114**, 661 (2001).
[49] M. Schütz, Phys. Chem. Chem. Phys. **4**, 3941 (2002).
[50] H.-J. Werner and M. Schütz, J. Chem. Phys. **135**, 144116 (2011).
[51] O. Masur, D. Usvyat, and M. Schütz, J. Chem. Phys. **139**, 164116 (2013).
[52] M. Schütz, O. Masur, and D. Usvyat, J. Chem. Phys. **140**, 244107 (2014).
[53] H.-J. Werner, J. Chem. Phys. **129**, 101103 (2008).
[54] T. B. Adler and H.-J. Werner, J. Chem. Phys. **135**, 144117 (2011).
[55] T. B. Adler, H.-J. Werner, and F. R. Manby, J. Chem. Phys. **130**, 054106 (2009).
[56] D. Kats, T. Korona, and M. Schütz, J. Chem. Phys **125**, 104106 (2006).
[57] D. Kats and M. Schütz, J. Chem. Phys. **131**, 124117 (2009).
[58] K. Freundorfer, D. Kats, T. Korona, and M. Schütz, J. Chem. Phys. **133**, 244110 (2010).
[59] K. Ledermüller, D. Kats, and M. Schütz, J. Chem. Phys. **139**, 084111 (2013).
[60] K. Ledermüller and M. Schütz, J. Chem. Phys. **140**, 164113 (2014).
[61] T. S. Chwee, A. B. Szilva, R. Lindh, and E. A. Carter, J. Chem. Phys. **128**, 224106 (2008).
[62] C. Pisani, M. Busso, G. Capecchi, S. Casassa, R. Dovesi, L. Maschio, C. Zicovich-Wilson, and M. Schütz, J. Chem. Phys. **122**, 094113 (2005).
[63] L. Maschio, D. Usvyat, F. Manby, S. Casassa, C. Pisani, and M. Schütz, Phys. Rev. B **76**, 075101 (2007).
[64] D. Usvyat, L. Maschio, F. R. Manby, S. Casassa, M. Schütz, and C. Pisani, Phys. Rev. B **76**, 075102 (2007).
[65] D. Usvyat, B. Civalleri, L. Maschio, R. Dovesi, C. Pisani, and M. Schütz, J. Chem. Phys. **134**, 214105 (2011).
[66] J. E. Subotnik and M. Head-Gordon, J. Chem. Phys. **123**, 064108 (2005).
[67] W. Meyer, Int. J. Quantum Chem. **S5**, 341 (1971).
[68] F. Neese, F. Wennmohs, and A. Hansen, J. Chem. Phys. **130**, 114108 (2009).
[69] F. Neese, A. Hansen, and D. G. Liakos, J. Chem. Phys. **131**, 064103 (2009).
[70] C. Riplinger and F. Neese, J. Chem. Phys. **138**, 034106 (2013).
[71] C. Riplinger, B. Sandhoefer, A. Hansen, and F. Neese, J. Chem. Phys. **139**, 134101 (2013).
[72] D. P. Tew, B. Helmich, and C. Hättig, J. Chem. Phys. **135**, 074107 (2011).
[73] C. Hättig, D. P. Tew, and B. Helmich, J. Chem. Phys. **136**, 204105 (2012).
[74] C. Krause and H.-J. Werner, Phys. Chem. Chem. Phys. **14**, 7591 (2012).
[75] H.-J. Werner, G. Knizia, C. Krause, M. Schwilk, and M. Dornbach, J. Chem. Theory Comput. **11**, 484 (2015).
[76] J. Yang, Y. Kurashige, F. R. Manby, and G. K. L. Chan, J. Chem. Phys. **134**, 044123 (2011).
[77] Y. Kurashige, J. Yang, G. K. L. Chan, and F. R. Manby, J. Chem. Phys. **136**, 124106 (2012).
[78] M. Schütz, J. Yang, G. K. L. Chan, F. R. Manby, and H.-J. Werner, J. Chem. Phys. **138**, 054109 (2013).

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

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

[81]P. Pulay and S. Saebø, Theor. Chem. Acc. **69**, 357 (1986).

[82]P. J. Knowles, M. Schütz, and H.-J. Werner, in *Modern Methods and Algorithms of Quantum Chemistry* (John von Neumann Institute for Computing, 2000), Vol. 1, p. 69.

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

[84]M. Schütz, D. Usvyat, M. Lorenz, C. Pisani, L. Maschio, S. Casassa, and M. Halo, "Density fitting for correlated calculations in periodic systems," in *Accurate Condensed-Phase Quantum Chemistry*, Computation in Chemistry (CRC Press, 2010), Vol. 27, p. 29.

[85]D. Usvyat, L. Maschio, C. Pisani, and M. Schütz, Z. Phys. Chem. **224**, 441-454 (2010).

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

[87]M. Halo, S. Casassa, L. Maschio, C. Pisani, R. Dovesi, D. Ehinon, I. Baraille, M. Rèrat, and D. Usvyat, Phys. Chem. Chem. Phys. **13**, 4434 (2011).

[88]D. Kats, J. Chem. Phys. **141**, 244101 (2014).

[89]B. I. Dunlap, Phys. Chem. Chem. Phys. **2**, 2113 (2000).

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

[91]M. Lorenz, D. Usvyat, and M. Schütz, J. Chem. Phys. **134**, 094101 (2011).

[92]See supplementary material at http://dx.doi.org/10.1063/1.4921301 for the computational parameters of the studies systems.

[93]M. Schütz, G. Rauhut, and H. J. Werner, J. Phys. Chem. A **102**, 5997 (1998).