The orbital-specific virtual local triples correction: OSV-L(T)

Martin Schütz' , Jun Yang' , Garnet Kin-Lic Chan' , Frederick R. Manby' , and Hans-Joachim Werner'

Citation: J. Chem. Phys. **138**, 054109 (2013); doi: 10.1063/1.4789415
View online: http://dx.doi.org/10.1063/1.4789415
View Table of Contents: http://aip.scitation.org/toc/jcp/138/5
Published by the American Institute of Physics

The orbital-specific virtual local triples correction: OSV-L(T)

Martin Schütz,[1,a)] Jun Yang,[2,b)] Garnet Kin-Lic Chan,[2,c)] Frederick R. Manby,[3,d)] and Hans-Joachim Werner[4,e)]
[1]*Institut für Physikalische und Theoretische Chemie, Universität Regensburg, Regensburg D-93040, Germany*
[2]*Department of Chemistry, Princeton University, Princeton, New Jersey 08544, USA*
[3]*Center for Computational Chemistry, School of Chemistry, University of Bristol, Bristol BS8 1TS, United Kingdom*
[4]*Institut für Theoretische Chemie, Universität Stuttgart, Stuttgart D-70569, Germany*

(Received 23 November 2012; accepted 11 January 2013; published online 4 February 2013)

A local method based on orbital specific virtuals (OSVs) for calculating the perturbative triples correction in local coupled cluster calculations is presented. In contrast to the previous approach based on projected atomic orbitals (PAOs), described by Schütz [J. Chem. Phys. **113**, 9986 (2000)], the new scheme works without any *ad hoc* truncations of the virtual space to domains. A single threshold defines the pair and triple specific virtual spaces completely and automatically. It is demonstrated that the computational cost of the method scales linearly with molecular size. Employing the recommended threshold a similar fraction of the correlation energy is recovered as with the original PAO method at a somewhat lower cost. A benchmark for 52 reactions demonstrates that for reaction energies the intrinsic accuracy of the coupled cluster with singles and doubles excitations and a perturbative treatment of triples excitations method can be reached by OSV-local coupled cluster theory with singles and doubles and perturbative triples, provided a MP2 correction is applied that accounts for basis set incompleteness errors as well as for remaining domain errors. As an application example the interaction energies of the guanine-cytosine dimers in the Watson-Crick and stacked arrangements are investigated at the level of local coupled cluster theory with singles and doubles and perturbative triples. Based on these calculations we propose new complete-basis-set-limit estimates for these interaction energies at this level of theory. © *2013 American Institute of Physics*. [http://dx.doi.org/10.1063/1.4789415]

I. INTRODUCTION

Coupled cluster (CC) theory is probably the most successful post-Hartree-Fock method to treat dynamic electron correlation; CC with singles and doubles excitations and a perturbative treatment of triples excitations (CCSD(T)) provides chemical accuracy for single reference cases, provided that either extended basis sets in conjunction with extrapolation techniques are used, or explicitly correlated terms are included that strongly reduce the basis set incompleteness error.[1–6] The success of CC theory originates from the exponential ansatz $|\Psi_{CC}\rangle = \exp(\mathbf{T})$ for the CC wavefunction, where $\mathbf{T}$ is the cluster operator, truncated, e.g., beyond doubles excitations in the case of CCSD. This exponential ansatz ensures that the method is size-extensive, and, more generally, introduces all excited determinants of full configuration interaction (CI) beyond doubles (for the CCSD case) into the wavefunction, with coefficients factorized into singles and doubles amplitudes (disconnected products). From that angle, CC theory is a full CI approximation, which imposes a special tensor factorization on the CI coefficients of determinants beyond the truncation level of $\mathbf{T}$ (in contrast to many other tensor factorization schemes there is no contraction over auxiliary indices here). It is well known that connected triples are still important to reach chemical accuracy, yet for excitations beyond triples this factorization of the full CI coefficients works very well. Hence, it reduces the scaling of the computational cost with molecular size $\mathcal{N}$ enormously from factorial to polynomial, i.e., to $\mathcal{O}(\mathcal{N}^7)$, when assuming that the perturbative (T) correction[7–9] is sufficiently accurate.

However, a scaling of $\mathcal{O}(\mathcal{N}^7)$ is still very high, restricting the application range of CCSD(T) to rather small molecules. It is therefore tempting to impose a further tensor factorization onto the CC doubles and triples amplitudes.

In quantum chemistry, matrix factorizations are frequently used. Obvious examples are the Cholesky[10–14] and singular value decompositions,[15] and density fitting.[16–22] Also schemes defining new occupied and virtual orbital sets, such as the projected atomic orbital (PAO) based local correlation,[23–27] the pair natural orbital (PNO),[28–37] and the orbital specific virtual (OSV)[38,39] methods can be interpreted as tensor factorizations. All these methods substitute the delocalized occupied Hartree-Fock orbitals by (still mutually orthogonal) localized molecular orbitals (LMOs), obtained by a unitary transformation of the former. Yet they differ in the handling of the virtual space: PAO methods define a single global set of virtuals (PAOs) and specify LMO-pair- and triple-specific subsets thereof. PNO methods, on the other hand, define a distinct adapted set of virtuals for

a)Electronic mail: martin.schuetz@chemie.uni-regensburg.de.
b)Electronic mail: juny@princeton.edu.
c)Electronic mail: gkchan@princeton.edu.
d)Electronic mail: fred.manby@bris.ac.uk.
e)Electronic mail: werner@theochem.uni-stuttgart.de.

each pair. This has the advantage that a considerably smaller number of PNOs than PAOs per pair is needed to reach the same accuracy, but due to the huge total number of PNOs it leads to complications in the integral transformation. OSV methods represent a compromise: here each LMO rather than each LMO-pair has its own adapted set of virtuals (OSVs). This leads to important simplifications relative to the PNO method, at the price of a less compact virtual orbital basis for each pair. A direct comparison of the various schemes can be found in Ref. 37. Hybrid PAO/OSV/PNO schemes, which exploit and combine the advantages of the different schemes and avoid the $\mathcal{O}(\mathcal{N}^4)$ scaling of the OSV and PNO generation, have also been proposed[37] and are currently being efficiently implemented.[40] Similar methods are also being developed by other groups.[41]

Recently, we presented an efficient implementation of OSV-local coupled cluster theory with singles and doubles (LCCSD).[39] In the present work, we extend the method to OSV-LCCSD(T). To this end, we adapt our previous PAO based triples program[42–44] according to the OSV-L(T) formalism. While a high-spin open-shell PAO-LCCSD(T) implementation is already available,[45] the current work is restricted to the closed-shell case.

II. THEORY

In this section, we briefly describe the OSV approach and then provide the relevant OSV-L(T) equations for the closed-shell case. In the following, indices i, j, k, l will denote LMOs, a, b, c, d canonical virtual orbitals. These orbital spaces are orthonormal and mutually orthogonal. For OSV orbitals, we will employ indices r, s, t, u. As the name implies, OSVs are specific for different LMOs. Where appropriate the respective LMO is indicated by a superscript, e.g., $|r^i\rangle \equiv |\phi_r^{(i)}\rangle$, etc. Electron repulsion integrals are written in the physicists' notation, i.e.,

$$\langle rs|ij\rangle = \int d\mathbf{r}_1 \int d\mathbf{r}_2\ \phi_r(\mathbf{r}_1)\phi_s(\mathbf{r}_2) r_{12}^{-1} \phi_i(\mathbf{r}_1)\phi_j(\mathbf{r}_2). \quad (1)$$

Since we are using a real spin-free formulation all indices refer to spatial orbitals $\phi_r(\mathbf{r})$ and there is no antisymmetrization of the integrals. In the following, ϵ_a are the eigenvalues of the virtual block of the Fock matrix in the canonical orbital basis, and f_{ij} are the elements of the occupied-occupied block of the Fock matrix in the LMO basis.

A. The OSV approach

The OSV orbitals are typically generated by diagonalization of the *diagonal* MP2 pair amplitudes,[38,39]

$$T_{ab}^{ii} = -\frac{\langle ab|ii\rangle}{\epsilon_a + \epsilon_b - 2f_{ii}} \quad (2)$$

as

$$\sum_{ab} Q_{ar}^i Q_{bs}^i T_{ab}^{ii} = t_r^{ii}\delta_{rs}, \quad |r^i\rangle = \sum_a |a\rangle Q_{ar}^i. \quad (3)$$

It is also possible to generate them by direct optimization of a Lagrangian, e.g., the Hylleraas functional, as discussed previously.[46] In either case, the resulting unitary (orthogonal) transformation matrices $\mathbf{Q}^i$ transform from canonical virtuals $|a\rangle$ to the OSVs $|r^i\rangle$. The OSVs $|r^i\rangle$ for a given LMO i form an orthogonal basis, but the OSV sets for different LMOs are in general mutually non-orthogonal, $\langle r^i|s^j\rangle \neq 0$ for $i \neq j$. If determined using Eq. (2), the OSVs are equal to the PNOs for the corresponding diagonal pairs. From here on we shall consider only OSVs determined from Eq. (2).

Based on the magnitude of the eigenvalues t_r^{ii} a certain subset of OSVs (domain $[i]$) can be determined for each LMO i. Alternatively, the MP2 correlation energy of the diagonal pair,

$$\epsilon_{ii} = \sum_{ab} \langle ii|ab\rangle T_{ab}^{ii} = \sum_r k_r^{ii} t_r^{ii}; \quad k_r^{ii} = \sum_{ab} \langle ii|ab\rangle Q_{ar}^i Q_{br}^i \quad (4)$$

can be used as a criterion:[39] OSVs (ordered according to decreasing t_r^{ii}) are added to $[i]$ until the deviation between $\sum_{r\in[i]} k_r^{ii} t_r^{ii}$ and the exact ϵ_{ii} in Eq. (4) becomes smaller than a certain threshold l_{OSV}. This criterion will be used throughout the current work. Here and in the following, in analogy to the PAO method, we still denote these subsets of OSVs *domains*, but point out that they do not originate from an *ad hoc* assumption on the truncation of the virtual space.

In the truncated OSV basis, the doubles amplitudes in canonical basis are approximated by the relation

$$T_{ab}^{ij} \approx \sum_{rs\in[ij]} Q_{ar}^{ij} Q_{bs}^{ij} T_{rs}^{ij}, \quad (5)$$

where the indices r, s run over the united domain $[ij] = [i] \cup [j]$, and $\mathbf{Q}^{ij}$ are composite transformation matrices

$$\mathbf{Q}^{ij} = (\mathbf{Q}^i \mathbf{Q}^j), \quad (6)$$

i.e., the two transformation matrices $[\mathbf{Q}^i]_{ar^i}$, $r^i \in [i]$ and $[\mathbf{Q}^j]_{ar^j}$ $r^j \in [j]$ are concatenated side by side. We denote the union of the OSVs $|r^i\rangle$ and $|r^j\rangle$ as a *pair domain* of virtual orbitals

$$|r^{ij}\rangle = \sum_a |a\rangle Q_{ar}^{ij}. \quad (7)$$

It may happen that the OSVs $|r^{ij}\rangle$ in a pair domain are (nearly) linearly dependent, and singular value decomposition is used to eliminate redundant functions (this is only done at intermediate stages when the amplitudes are updated in an orthogonal basis, cf. Ref. 39). For a full set of non-redundant OSVs (i.e., if their number equals the number of virtual MOs), the transformation in Eq. (5) is exact. However, for nonzero thresholds l_{OSV}, i.e., a truncated set of OSVs (less OSVs than canonical virtuals) Eq. (5) is not a transformation, but a projection; the T_{ab}^{ij} on the lhs are then not equal to the true canonical amplitudes. For a thorough discussion of the transformation/projection relations between amplitudes and integrals in canonical and local basis we refer to Refs. 37 and 47.

The orbitals $|r^{ij}\rangle$ in the pair domains are non-orthogonal, with the overlap matrix

$$S_{rs}^{ij,kl} = \langle r^{ij}|s^{kl}\rangle = \sum_a Q_{ar}^{ij} Q_{as}^{kl}. \quad (8)$$

$\mathbf{S}^{ij,\,kl}$ is a rectangular submatrix of the full OSV overlap matrix $\mathbf{S}$, where the rows run over the pair domain $[ij]$, the columns over pair domain $[kl]$. In a diagrammatic representation of the

CCSD residual equations such overlaps pop up whenever a particle line originating from an amplitude vertex on the `bra` side is directly connected to the `ket` side, i.e., not connected to a vertex corresponding to a fragment of the Hamilton operator.

For the triples amplitudes we have, by analogy to Eqs. (5) and (6),

$$T^{ijk}_{abc} \approx \sum_{rst\in[ijk]} Q^{ijk}_{ar} Q^{ijk}_{bs} Q^{ijk}_{ct} T^{ijk}_{rst}. \tag{9}$$

Here, the composite transformation matrices

$$\mathbf{Q}^{ijk} = (\mathbf{Q}^i \mathbf{Q}^j \mathbf{Q}^k), \tag{10}$$

now concatenate three OSV transformation matrices $\mathbf{Q}^i$, and the r, s, t summations in Eq. (9) now run over the triples domain $[ijk] = [i] \cup [j] \cup [k]$.

B. The OSV-L(T) equations

In order to compute the perturbative triples correction[7–9] in a non-canonical local basis, the triples equations must be formulated in an orbital-invariant way. Various and quite different local correlation methods for the perturbative triples correction have been proposed in the past,[43,48–51] the later ones based on "divide-and-conquer," and "cluster-in-molecule" approaches. The OSV-L(T) method presented here is based on the first, i.e., the PAO based local triples method discussed in Ref. 43. Since the OSV-L(T) formalism is very similar to that of the PAO-L(T) method we just give a short overview, provide the relevant equations, and refer for further details to Ref. 43. The triples amplitudes are obtained by solving the second-order triples equations

$$R^{ijk}_{abc} = \langle \tilde{\Phi}^{abc}_{ijk} | \hat{f}\hat{T}_3 | \Phi_0 \rangle + \langle \tilde{\Phi}^{abc}_{ijk} | \hat{V}\hat{T}_2 | \Phi_0 \rangle = 0, \tag{11}$$

where $\hat{f}$ and $\hat{V}$ are the normal ordered Fock- and fluctuation potential operators, and

$$\hat{T}_2 = \frac{1}{2} \sum_{ij} \sum_{ab} T^{ij}_{ab} \hat{E}_{ai} \hat{E}_{bj}, \tag{12}$$

$$\hat{T}_3 = \frac{1}{6} \sum_{ijk} \sum_{abc} T^{ijk}_{abc} \hat{E}_{ai} \hat{E}_{bj} \hat{E}_{ck}, \tag{13}$$

are the doubles and triples excitation operators, written in terms of spin-conserving one-particle excitation operators $\hat{E}_{ai}$. $|\Phi_0\rangle$ is the Hartree-Fock reference determinant, and

$$\tilde{\Phi}^{abc}_{ijk} = \frac{1}{6}\big(\Phi^{ab\bar{c}}_{ij\bar{k}} + \Phi^{\bar{a}\bar{b}c}_{\bar{i}\bar{j}k} + \Phi^{a\bar{b}c}_{i\bar{j}k} + \Phi^{\bar{a}b\bar{c}}_{\bar{i}j\bar{k}} + \Phi^{a\bar{b}\bar{c}}_{i\bar{j}\bar{k}} + \Phi^{\bar{a}bc}_{\bar{i}jk} + \Phi^{ac\bar{b}}_{ij\bar{k}} + \Phi^{\bar{a}\bar{c}b}_{\bar{i}\bar{j}k} + \Phi^{b\bar{a}c}_{i\bar{j}k} + \Phi^{\bar{b}a\bar{c}}_{\bar{i}j\bar{k}} + \Phi^{c\bar{b}\bar{a}}_{i\bar{j}\bar{k}} + \Phi^{\bar{c}ba}_{\bar{i}jk}\big) \tag{14}$$

are contravariant triply excited configuration state functions, chosen as in Ref. 43. In Eq. (14), the orbital indices on the lhs refer to spatial orbitals, those on the rhs to the corresponding spin orbitals with either α or (when decorated with an overbar) β spin.

Equation (11) straightforwardly leads to[43]

$$R^{ijk}_{abc} = V^{ijk}_{abc} + W^{ijk}_{abc} = 0, \tag{15}$$

with

$$V^{ijk}_{abc} = \mathcal{P}_1 \left(\sum_d f_{cd} T^{ijk}_{abd} - \sum_l f_{kl} T^{ijl}_{abc} \right), \tag{16}$$

$$W^{ijk}_{abc} = \mathcal{P}_1 \mathcal{P}_2 \left(\sum_d \langle bc|dk\rangle T^{ij}_{ad} - \sum_l \langle lc|jk\rangle T^{il}_{ab} \right), \tag{17}$$

and with orbital index permutation operators

$$\mathcal{P}_1 = 1 + (jk)(bc) + (ik)(ac), \quad \mathcal{P}_2 = 1 + (ij)(ab). \tag{18}$$

The (T) correction to the CCSD energy is defined as

$$\delta E = \langle 0|(\hat{T}_1 + \hat{T}_2)^\dagger \hat{V} \hat{T}_3|0\rangle, \tag{19}$$

where $\hat{T}_1 = \sum_{ia} t^i_a \hat{E}_{ai}$ is the spin-conserving singles excitation operator. This leads to the energy expression[43]

$$\delta E = \delta E(S) + \delta E(D), \tag{20}$$

with

$$\begin{aligned} \delta E(S) &= \sum_{i\le j\le k} (2 - \delta_{ij} - \delta_{jk}) \sum_{abc} X^{ijk}_{abc} \mathcal{P}_3 t^i_a \langle jk|bc\rangle, \\ \delta E(D) &= \sum_{i\le j\le k} (2 - \delta_{ij} - \delta_{jk}) \sum_{abc} X^{ijk}_{abc} W^{ijk}_{abc}, \end{aligned} \tag{21}$$

using the additional orbital index permutation operator

$$\mathcal{P}_3 = 1 + (ij)(ab) + (ik)(ac), \tag{22}$$

and defining

$$X^{ijk}_{abc} = 4T^{ijk}_{abc} - 2T^{ijk}_{acb} - 2T^{ijk}_{cba} - 2T^{ijk}_{bac} + T^{ijk}_{cab} + T^{ijk}_{bca}. \tag{23}$$

What we have derived so far is an orbital-invariant formalism of the triples correction in an orthonormal orbital basis. For a canonical orbital basis with a diagonal Fock matrix in Eq. (16), Eq. (15) can directly be inverted, yielding (for a fixed triple ijk) T^{ijk}_{abc}, then via (23) X^{ijk}_{abc}, and finally via (20) and (21) the contribution of that triple to δE.

In a non-canonical orbital basis, Eq. (15) has to be solved iteratively, which implies storage of the triples amplitudes. Yet neglecting the internal couplings via the occupied-occupied block of the Fock matrix allows for direct inversion and calculation of δE in the *pseudocanonical* orbital basis for each triple ijk separately without the necessity of storing the triples amplitudes (cf. Sec. II E in Ref. 43): the pseudocanonical basis is obtained by diagonalizing the external Fock operator projected onto the ijk-specific virtual space; the corresponding amplitudes then are computed by direct inversion, i.e., as

$$T^{ijk}_{\tilde{a}\tilde{b}\tilde{c}} = -W^{ijk}_{\tilde{a}\tilde{b}\tilde{c}} / (\epsilon_{\tilde{a}} + \epsilon_{\tilde{b}} + \epsilon_{\tilde{c}} - f_{ii} - f_{jj} - f_{kk}), \tag{24}$$

where $\tilde{a}, \ldots$ are indices of the pseudocanonical orbitals spanning the ijk-specific virtual space, $\epsilon_{\tilde{a}}, \ldots$ the related orbital energies of the projected Fock operator, and $f_{ii}, \ldots$ the diagonal elements of the internal Fock matrix in LMO basis. This specifies the so-called L(T0) approximation to the full iterative L(T) correction. Treating the internal couplings at the level of first-order perturbation theory [the L(T1)

approximation[43]] also avoids storage of the triples amplitudes, but is computationally still much more costly than the L(T0) approximation. δE in the L(T1) approximation is identical to δE of the first iteration of the full L(T) correction.

The OSV-L(T) equations are now straightforwardly derived by transforming Eqs. (15) and (21) from the canonical to the OSV basis by employing the transformation relation for integrals (3), and the converse transformation relations for amplitudes (5) and (9). For example, from Eq. (17) we obtain

$$\begin{aligned} W_{rst}^{ijk} &= \sum_{abc} W_{abc}^{ijk} Q_{ar}^{ijk} Q_{bs}^{ijk} Q_{ct}^{ijk} \\ &= \mathcal{P}_1' \mathcal{P}_2' \left(\sum_{ur'} \langle st|uk\rangle T_{r'u}^{ij} S_{r'r}^{ij,ijk} \right. \\ &\quad \left. - \sum_{lr's'} \langle lt|jk\rangle T_{r's'}^{il} S_{r'r}^{il,ijk} S_{s's}^{il,ijk} \right), \end{aligned} \tag{25}$$

with $\mathcal{P}_1'$ and $\mathcal{P}_2'$ being analogously defined as the $\mathcal{P}_1$ and $\mathcal{P}_2$ in Eq. (18), but permuting in addition also the related primed indices accordingly, i.e.,

$$\begin{aligned} \mathcal{P}_1' &= 1 + (jk)(st)(s't') + (ik)(rt)(r't'), \\ \mathcal{P}_2' &= 1 + (ij)(rs)(r's'). \end{aligned} \tag{26}$$

Equation (25) is formally identical to Eq. (19) of Ref. 43. We therefore do not provide the detailed OSV-L(T) equations obtained by transforming Eqs. (16) and (21), but refer instead to Eqs. (18), (28), and (29) of Ref. 43. Note the occurrence of the overlap matrix $\mathbf{S}^{ij,\,ijk} = \mathbf{Q}^{ij\dagger}\mathbf{Q}^{ijk}$. It is a rectangular submatrix of the full OSV overlap matrix $\mathbf{S}$, where the rows run over the pair domain $[ij]$, the columns over the triples domain $[ijk]$. As discussed in our previous paper,[39] the full OSV Fock and overlap matrices are kept in memory in our program, and appropriate submatrices extracted from it.

C. Local pair- and triples approximations

In the local CC method not all pairs or triples are treated at the same level, but additional approximations are applied to different classes of pairs or triples, depending on the mutual spatial separation between the related LMOs.[22–24,52,53] To this end, classes of *strong*, *close*, *weak*, and *very distant* pairs are distinguished. Only strong pairs are treated at the full LCCSD level, close and weak pairs only at the LMP2 level, or optionally, at the level of the local random phase approximation (LRPA) (*vide infra*). Very distant pairs are neglected. Only strong and close pairs enter the construction of the restricted triples list for the L(T) calculation: by default, only those triples are added to the list, which comprise at least one strong, and at most two close pairs.[43] It is also possible to lift the restriction that one pair has to be strong, thus, extending the triples list and the cost of the L(T) calculation. However, as Table 3 in Ref. 43 shows, the effect on the triples energy is small. Hence, in all test calculations presented in Sec. III the default setting is employed.

The spatial separation between the LMOs specifying the different pair classes can either be measured by a distance or a connectivity criterion. The latter counts the number of bonds between two LMOs. As in the previous PAO based method, each LMO still specifies a subset of relevant atoms (obtained from a Löwdin population analysis and truncation of the ordered atoms list beyond a certain population, as done in the Boughton-Pulay (BP) procedure[54]). Hence, distances or number of bonds between two LMO are defined as the distances or number of bonds between the closest atoms of the two respective subsets. Two atoms are considered to be bonded, if their distance is smaller than 1.2× the sum of their atomic radii.

In the calculations presented below, we employ the connectivity criterion. The pair approximation is defined by the three integers w, c, and k, as discussed in detail in Refs. 5 and 6. w and c specify the minimum number of bonds between the two LMOs that form weak and close pairs, respectively. Only strong pair amplitudes are optimized in the LCCSD step. $k = 1$ flags that LMP2 close pair amplitudes enter the LCCSD residuals for the strong pairs, while for $k = 0$ this is not done. For example, `wck = 321` means that in strong pairs the two orbital domains are separated by at most one bond, close pairs by two bonds, weak pairs by at least three bonds, and that close pairs enter the LCCSD strong pair residuals.

It should be noted that the above selection of the pair types is still based on *ad hoc* distance or connectivity criteria. The effect of these criteria has been extensively investigated in Refs. 5 and 6. An alternative possibility, which avoids any distance criteria altogether, is to define the pair classes based on energy thresholds, using the LMP2 pair energies. This will be explored further in future work.

Our program recognizes individual molecules in intermolecular complexes or clusters. On that basis, also intermolecular pairs are identified. These can then either be specified independently as strong, close, or weak. This feature is utilized in Sec. III C. In all calculations presented here, the class of very distant pairs (which are entirely neglected) is switched off and remains empty.

III. TEST CALCULATIONS

The OSV-L(T) method has been implemented in the MOLPRO program package[55,56] on the basis of the PAO-L(T) program described in Ref. 43. Density fitting is employed to efficiently evaluate the required two- and three-external electron repulsion integrals, as discussed in detail in Ref. 22.

In the following, we present the results of some test calculations to explore how the OSV-L(T) method performs in comparison to PAO-L(T). The preceding LCCSD calculations are carried out with the OSV-LCCSD program reported earlier.[22,39] In all calculations, the LMOs have been generated by using Pipek-Mezey localization.[57] For basis sets augmented by diffuse basis functions, the contributions of the most diffuse basis functions of each angular momentum are removed from the localization criterion. This improves localization and is the recommended way to generate LMOs for basis sets with diffuse functions.

A. Scaling behaviour with molecular size

As a first test system we employ the same linear poly-glycine peptide chains $(\mathrm{Gly})_n \equiv \mathrm{HO[C(O)CH_2NH]}_n\mathrm{H}$

($n = 1 \ldots 4$) as used already in Ref. 43 (structures and energies can be found in the supplementary material[58]). These artificial systems quickly and systematically reach the asymptotic regime and are therefore well suited to explore the scaling properties of a method. The calculations on $(Gly)_n$ were carried out in the cc-pVTZ basis (with the related `MP2FIT` fitting basis set of Weigend *et al.*[59]); the calculation of the canonical (T) correction for $n = 4$ took already about three weeks on four Intel Xeon cores X5690 @ 3.47 GHz. For the local calculations, the `wck = 321` pair approximation was employed.

Table I compiles the canonical CCSD and (T) correlation energies, along with the percentage thereof recovered by the individual local methods. The timings and correlation energies for the preceding LCCSD calculation can be found in the supplementary material.[58] For the PAO-LCCSD(T) calculations, standard BP domains with a completeness criterion of 0.98 are used. For the OSV-LCCSD(T) calculations, three different settings for the threshold l_{OSV}, i.e., $l_{OSV} = 1.0 \times 10^{-4}$, 3.2×10^{-5}, and 1.0×10^{-5} are investigated. Furthermore, for the local triples corrections the full iterative L(T) treatment, the L(T1) variant, as well as the often used L(T0) approximation are compared.

As expected, the OSV method performs very well for the LCCSD correlation energy E_{CCSD}: already a threshold of $l_{OSV} = 1.0 \times 10^{-4}$ recovers between 99.2% ($n = 1$) and 99.0% ($n = 4$) of the canonical correlation energy, which is about the same amount as recovered by the PAO based LCCSD with standard BP domains. We note that there is a slight decrease in the fraction of the recovered correlation energy on going from $n = 1$ to $n = 4$. By decreasing the OSV threshold to $l_{OSV} = 1.0 \times 10^{-5}$ even 99.9% of the canonical correlation energy is recovered.

The L(T) correction performs not quite as well, but still recovers a large percentage of the canonical triples correction. The somewhat less good performance of OSV-L(T) compared to OSV-LCCSD is not surprising, since of course the OSV basis is constructed by analyzing doubles, rather than triples amplitudes. Nevertheless, with a threshold of $l_{OSV} = 1.0 \times 10^{-4}$ between 94.6% ($n = 1$) and 92.5% ($n = 4$) of the canonical (T) energy correction $\Delta E_{(T)}$ are recovered. Most of the decrease of this fraction with increasing chain length occurs between $n = 1$ and $n = 2$; apparently, the electronic structure of $n = 1$ is somewhat different from the larger molecules. The fraction still slightly decreases from $n = 2$ to $n = 4$, but this effect is further reduced if more pairs are treated as strong or close (see below). In any case, taking into account that $\Delta E_{(T)}$ is more than an order of magnitude smaller than E_{CCSD} the effect of this decrease on the total energies is smaller for the L(T) correction than for the LCCSD correlation energy.

Decreasing the threshold to $l_{OSV} = 3.2 \times 10^{-5}$ recovers 97.2%–95.5%, which is slightly more than the fraction recovered by PAO-L(T) with default domains. With a threshold of $l_{OSV} = 1.0 \times 10^{-5}$ finally, 98.5%–97.1% of $\Delta E_{(T)}$ is recovered. On going from the full L(T) correction to the L(T1) or L(T0) approximations further ≈0.5% or ≈2.5%, respectively, of $\Delta E_{(T)}$ are lost. Nevertheless, considering the much smaller size of $\Delta E_{(T)}$ relative to E_{CCSD} our experience shows that this is even for the L(T0) approximation still unproblematic for most applications.

In order to check how the pair approximations affect the triples energy, the weak and close pair criteria have been varied, using the (T0) approximation and $l_{OSV} = 3.2 \times 10^{-5}\, E_h$. The results are shown in Table II. The fraction of LCCSD correlation energy is nearly independent of these parameters; it slightly decreases if more strong pairs are fully optimized in the LCCSD (i.e., with increasing parameter `c`). In contrast, the fraction of recovered triples energy increases if more

TABLE I. Canonical CCSD E_{CCSD} and (T) correlation energies for $(Gly)_n$ ($n = 1 \ldots 4$), and percentage thereof recovered by PAO-LCCSD(T) (with standard BP domains and completeness criterion of 0.98), and OSV-LCCSD(T) with $l_{OSV} = 1.0 \times 10^{-4}$, 3.2×10^{-5}, and 1.0×10^{-5}, respectively. The `wck = 321` pair approximation is used.

$n =$	1	2	3	4
E_{CCSD}	−1.024627	−1.786898	−2.549422	−3.312013
$\Delta E_{(T)}$	−0.041638	−0.076421	−0.111313	−0.146228
l_{OSV}/E_h		LCCSD		
1.0×10^{-4}	99.17%	99.02%	98.96%	98.95%
3.2×10^{-5}	99.66%	99.58%	99.55%	99.53%
1.0×10^{-5}	99.88%	99.85%	99.83%	99.85%
PAO	99.15%	99.05%	99.01%	98.99%
l_{OSV}/E_h		L(T)		
1.0×10^{-4}	94.61%	93.23%	92.70%	92.42%
3.2×10^{-5}	97.25%	96.13%	95.74%	95.51%
1.0×10^{-5}	98.48%	97.63%	97.26%	97.06%
PAO	96.64%	95.86%	95.55%	95.39%
l_{OSV}/E_h		L(T1)		
1.0×10^{-4}	94.19%	92.83%	92.31%	92.03%
3.2×10^{-5}	96.79%	95.69%	95.30%	95.08%
1.0×10^{-5}	98.00%	97.17%	96.80%	96.61%
PAO	96.19%	95.42%	95.12%	94.96%
l_{OSV}/E_h		L(T0)		
1.0×10^{-4}	92.01%	90.78%	90.32%	90.08%
3.2×10^{-5}	94.42%	93.44%	93.11%	92.92%
1.0×10^{-5}	95.53%	94.80%	94.49%	94.30%
PAO	93.77%	93.09%	92.83%	92.69%

TABLE II. Dependence of the LCCSD and (T0) correlation energy contributions relative to the canonical CCSD(T) values (in percent) on the choice of the close and weak pair parameters `wck` (see text). The OSV threshold was chosen to be $l_{OSV} = 3.2 \times 10^{-5}\, E_h$.

$n =$	1	2	3	4
`wck`		LCCSD		
321	99.66	99.58	99.55	99.53
431	99.64	99.52	99.52	99.50
541	99.64	99.52	99.52	99.50
421	99.66	99.57	99.54	99.52
521	99.66	99.56	99.54	99.52
`wck`		L(T0)		
321	94.42	93.44	93.11	92.92
431	94.72	94.02	93.82	93.70
541	94.72	94.13	93.96	93.86
421	94.79	94.10	93.90	93.78
521	94.79	94.21	94.04	93.94

close pair amplitudes are included in the triples calculation (parameter w). This effect increases with increasing chain length of the peptides. As already mentioned, for $(Gly)_1$ a larger fraction of correlation energy is recovered than for the larger molecules, both at the LCCSD level, as well as for the L(T0) correction. This effect remains even with extended strong and close pair lists. For the larger molecules, the decrease of the fraction of L(T0) with increasing chain length is reduced if more pairs are included in the triples calculation: for wck = 321 the fraction decreases from $n = 2$ to $n = 4$ by 0.52%, while for wck = 541 it decreases only by 0.27%. The last two rows in the table show that it is not necessary to include more strong pairs in the LCCSD calculation (which would strongly affect the computational effort[5]); it is sufficient just to use the LMP2 amplitudes (close pairs) in the triples calculation in order to recover most of the effect on the triples energy. But it should be noted that the number of integrals $\langle rs|ti\rangle$ to be computed, stored, and processed increases with increasing number of close pairs, and therefore wck = 321 appears to be a good compromise in terms of efficiency and accuracy.

Figure 1 displays the timings of the L(T) and L(T0) calculations, respectively, for $n = 1, \ldots, 4$, measured on seven Intel Xeon cores X5690 @ 3.47GHz. Evidently, the computational cost of both the PAO-L(T) and OSV-L(T) implementation scale linearly with n, whereas the cost of the canonical (T) correction (also shown) just explodes beyond $n = 2$. The L(T0) approximation is faster by a factor of 15–30 compared to the full iterative L(T) treatment. The slightly nonlinear scaling is probably due to the I/O overhead, caused by less efficient caching of integrals for the larger cases. Comparing the timings of the OSV triples with threshold $l_{\mathrm{OSV}} = 3.2 \times 10^{-5}$ to those of the PAO triples, which both recover a very similar percentage of $\Delta E_{(\mathrm{T})}$, one can see that OSV-L(T0) is slightly faster than PAO-L(T0), and that full OSV-L(T) is significantly faster (by more than a factor of two) than PAO-L(T). This factor of two in the latter case is most likely due to somewhat more compact triples domains in the OSV case (average triples domain: 95 OSVs) than in the PAO case (average triples domain: 118 PAOs). This implies a factor of about 1/2 in the size of the triples amplitudes block for a fixed LMO triple. In less artificial systems involving, e.g., aromatic rings, the ratio in the amount of non-vanishing triples between OSV- and PAO-L(T) is even much smaller, as is demonstrated in Sec. III C.

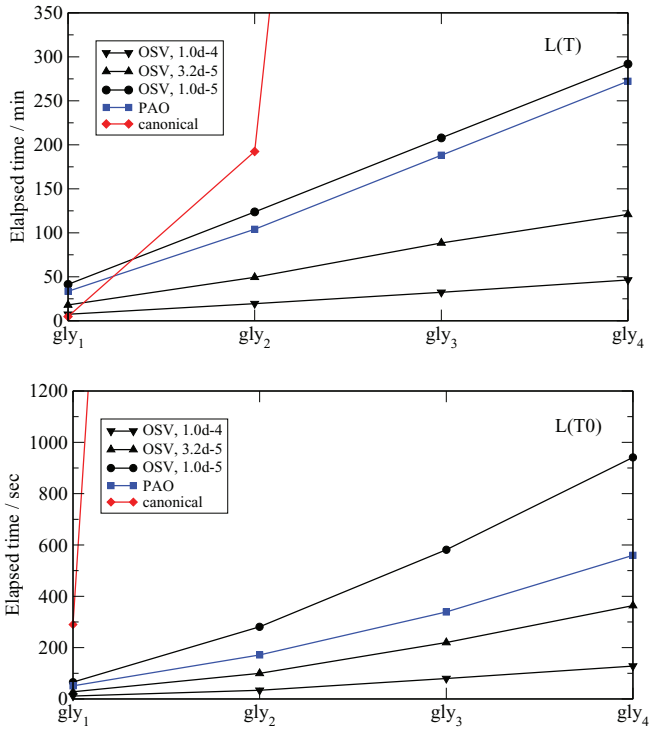

FIG. 1. Elapsed times for the full L(T) correction (upper panel, in min) and the L(T0) approximation approximation (lower panel, in s) for the $(Gly)_n$ ($n = 1 \ldots 4$) calculations. The elapsed times for the canonical calculations are also shown, as much as visible on the respective scale. The calculations were performed on seven Intel Xeon cores X5690 @ 3.47GHz with 512 MW memory per process.

In any case, the OSV triples calculation with a threshold of $l_{\mathrm{OSV}} = 3.2 \times 10^{-5}$ is certainly competitive to a default PAO triples calculation, with respect to efficiency, as well as with respect to accuracy.

B. A benchmark for reaction energies

In this section, we present benchmark results for reaction energies of 52 reactions. The reactions and molecular geometries are taken from Ref. 6, and this benchmark has already been used in Refs. 6, 37, and 39. In order to facilitate the comparison with those data, the same basis sets and reference values as in Ref. 6 have been used. Table III shows results relative to the canonical CCSD and CCSD(T) values obtained with the VTZ-F12 basis set,[60] while in Table IV the deviations of the VTZ-F12 results from the canonical estimated complete basis set (CBS) limits are shown. The CBS values[6] have been obtained by extrapolating CCSD(T)-F12b/VTZ-F12 and CCSD(T)-F12b/VQZ-F12 results as described in Ref. 61. In all local calculations, the pair selection criteria were wck

TABLE III. Maximum (MAX), mean absolute (ABS) and root-mean-square (RMS) deviations (in kJ/mol) of OSV-LCCSD and OSV-LCCSD(T0) reaction energies from the corresponding canonical CCSD and CCSD(T) results for 52 reactions, basis VTZ-F12 (wck = 321). ΔMP2 = MP2−LMP2 is a MP2 correction that accounts for domain errors, see text. The reactions, basis sets and reference values are the same as in Ref. 6.

	LCCSD			LCCSD+ΔMP2		
l_{OSV}/E_h	MAX	ABS	RMS	MAX	ABS	RMS
1.0×10^{-4}	12.5	3.3	4.3	5.2	1.1	1.4
3.2×10^{-5}	4.3	1.4	1.8	3.3	0.6	0.8
1.0×10^{-5}	3.0	0.7	0.9	2.4	0.5	0.7
PAO[a]	13.5	2.9	3.9	3.0	0.7	1.0
	LCCSD(T0)			LCCSD(T0)+ΔMP2		
l_{OSV}/E_h	MAX	ABS	RMS	MAX	ABS	RMS
1.0×10^{-4}	14.9	3.3	4.3	4.6	0.8	1.2
3.2×10^{-5}	6.8	2.1	2.7	2.6	0.7	0.9
1.0×10^{-5}	3.8	1.2	1.5	2.9	0.7	0.9
PAO[a]	15.5	3.5	4.8	3.0	0.8	1.1

[a]PAO-LCCSD and PAO-LCCSD(T0) calculations as described in the text.

TABLE IV. Maximum (MAX), mean absolute (ABS) and root-mean-square (RMS) deviations (in kJ/mol) of OSV-LCCSD and OSV-LCCSD(T0) reaction energies from canonical CCSD and CCSD(T) complete basis set estimates, respectively, for 52 reactions, basis VTZ-F12 (wck = 321). ΔMP2 = MP2/CBS[45]−LMP2/VTZ-F12 is a MP2 correction that accounts for domains and basis set incompleteness errors, see text. The reactions, basis sets and reference values are the same as in Ref. 6.

	LCCSD			LCCSD + ΔMP2		
l_{OSV}/E_h	MAX	ABS	RMS	MAX	ABS	RMS
1.0×10^{-4}	14.2	5.3	6.5	4.7	0.9	1.3
3.2×10^{-5}	13.9	4.2	5.1	3.8	0.9	1.3
1.0×10^{-5}	13.5	3.5	4.5	3.8	0.9	1.2
PAO[a]	18.9	5.2	6.6	5.1	1.2	1.7
Canonical[b]	12.2	3.4	4.4	3.6	0.9	1.3
	LCCSD(T0)			LCCSD(T0) + ΔMP2		
l_{OSV}/E_h	MAX	ABS	RMS	MAX	ABS	RMS
1.0×10^{-4}	18.6	6.4	8.0	3.6	1.2	1.5
3.2×10^{-5}	14.5	4.9	5.9	3.4	1.2	1.5
1.0×10^{-5}	13.8	4.0	5.0	3.4	1.2	1.4
PAO[a]	21.0	5.9	7.4	5.3	1.3	1.8
Canonical[b]	12.2	3.4	4.4	3.0	0.9	1.2

[a]PAO-LCCSD and PAO-LCCSD(T0) calculations as described in the text.
[b]Results of canonical CCSD and CCSD(T) calculations relative to the corresponding CBS values.

= 321, as explained in Sec. II C. The cc-pVTZ/JKFIT[62] and aug-cc-pVTZ/MP2FIT[59] fitting basis sets, respectively, have been used for fitting the Fock matrix and transformed electron repulsion integrals. In the PAO-LCCSD calculations, standard BP domains with a completeness criterion of 0.985 have been used. This is exactly as in Ref. 6, but note that in our previous paper on OSV-LCCSD[39] the PAO domains have been chosen in a different way.

The first three columns in Table III demonstrate that with decreasing threshold l_{OSV} the OSV-LCCSD and OSV-LCCSD(T) results converge monotonically towards the corresponding canonical ones. The maximum errors for the OSV-LCCSD(T) are up to 2.5 kJ/mol larger than those without triples, but this difference decreases with decreasing threshold. Even with the largest threshold ($l_{OSV} = 10^{-4} E_h$), the OSV results are slightly more accurate than the PAO-LCCSD ones, but of course the PAO results could also be improved by extending the domains.[39,63]

In the last three columns of Table III, the MP2 correction is applied, i.e., $\Delta E_{MP2} = E_{MP2} - E_{LMP2}$ is added to all LCCSD or LCCSD(T) energies. As already found in previous work,[22,39,63] this very effectively reduces the domain errors, and now the statistical data are nearly independent of the OSV threshold used. Even with the largest threshold the maximum errors are reduced to about 1 kcal/mol, and the root-mean-square deviations are even 3-4 times smaller.

Table IV shows similar results, but in this case the CCSD/CBS and CCSD(T)/CBS values are taken as the reference. As expected, the deviations from the reference values are now larger, since they also include the basis set incompleteness errors. As can be seen by comparison of the canonical CCSD and CCSD(T) values in Table IV with the OSV-LCCSD(T) values in Table III the basis set errors are about as large as the local errors with the largest threshold (without MP2 correction). This means that both sources of error must be reduced simultaneously in order to reach the intrinsic accuracy of the CCSD(T) method. The last three columns of Table IV demonstrate that this is very effectively achieved by adding the ΔMP2(CBS) correction $\Delta E_{MP2(CBS)} = E_{MP2/CBS} - E_{LMP2}$, where $E_{MP2/CBS}$ is obtained by extrapolating MP2/aug-cc-pVQZ ($n = 4$) and MP2/aug-cc-pV5Z ($n = 5$) correlation energies to the CBS limit (using $E_n = E_{CBS} + A \cdot n^{-3}$, where n is the cardinal number of the basis set), while E_{LMP2} is computed with the same basis set as the LCCSD(T) energies. This correction reduces both the basis set incompleteness errors as well as the domain errors, and chemical accuracy (<1 kcal/mol) is obtained, independent of the OSV threshold used. The deviations of the local results from the reference values are only marginally larger than those of the underlying canonical methods.

We note that similar accuracy can be achieved just with the VTZ-F12 basis set by including explicitly correlated terms in the wavefunction, as demonstrated in Refs. 6 and 37. This eliminates the need to carry out canonical MP2 calculations with very large basis sets. These so-called LCCSD-F12 methods very effectively reduce both the basis set incompleteness errors, as well as the errors due to the domain approximation,[6,37,64,65] and near linear cost scaling with molecular size can be achieved.[66] An efficient OSV-LCCSD(T)-F12 implementation will be presented elsewhere.

In conclusion, this benchmark demonstrates that the intrinsic accuracy of the CCSD(T) method can be reached using OSV-LCCSD(T0) with a threshold $l_{OSV} = 3.2 \times 10^{-5} E_h$, provided that either the MP2 correction is applied or F12 terms are included.

C. Intermolecular interaction energies

As an example for an application of the new method we consider the intermolecular interaction energies of the guanine-cytosine Watson-Crick (G-C/WC) and stacked (G-C/S) dimers. The geometries were taken from the JSCH-2005 benchmark set presented in Ref. 67, and are also provided in the supplementary material[58] for convenience. The aVTZ basis set (cc-pVTZ on hydrogen atoms, aug-cc-pVTZ on all other atoms) was used (in conjunction with the corresponding MP2FIT basis set of Weigend *et al.*[59]). The local approximations (domains, pair lists, number of redundancies in the pair specific virtual spaces) were determined at large intermolecular separation and kept fixed, as recommended for the treatment of intermolecular interactions.[68,69] For the LCCSD(T0) calculations, the wck = 320 pair approximation was employed. All intermolecular pairs were specified as close pairs and therefore enter the L(T0) calculation, but not the preceding LCCSD calculation. In order to improve the interaction energy contributions of close and weak pairs beyond LMP2 (the latter considerably overestimates these contributions for the case of the stacked dimer, cf. Ref. 39), we prefer to treat close and weak pairs at the level of the direct local random phase approximation, d-LRPA. d-LRPA includes higher-order ring diagrams, yet disregards exchange-type diagrams and thus contains exclusion principle violation

TABLE V. Counterpoise-corrected LCCSD(T0)/aVTZ interaction energies in (kcal mol^{-1}) for the Watson-Crick (G-C/WC) and the stacked (G-C/S) G-C base pairs. All intermolecular pairs are specified as close pairs and enter the L(T0) calculation. Close and weak pair amplitudes are calculated using the local direct random phase approximation, d-LRPA. For comparison, interaction energies for LMP2, d-LRPA, LCCSD, and canonical MP2 are included as well. Moreover, ΔMP2 corrected LCCSD and LCCSD(T0) values are also given (see text). The individual BSSE estimates according to the CP correction are provided in parenthesis.

l_{OSV}/E_h	LMP2	d-LRPA	LCCSD	LCCSD(T0)	LCCSD +ΔMP2	LCCSD(T0) +ΔMP2
			Watson-Crick			
1.0×10^{-4}	−27.42 (0.39)	−26.60 (1.15)	−26.25 (0.49)	−26.53 (0.43)	−28.52	−28.81
3.2×10^{-5}	−28.26 (0.86)	−27.18 (1.88)	−26.90 (0.97)	−27.31 (0.92)	−28.34	−28.75
1.0×10^{-5}	−28.71 (1.27)	−27.49 (2.28)	−27.26 (1.32)	−27.74 (1.34)	−28.25	−28.73
PAO	−28.52 (−0.01)	−27.18 (0.19)	−26.95 (0.15)	−27.35 (0.09)	−28.12	−28.52
Canonical	−29.70 (1.86)					
			Stacked			
1.0×10^{-4}	−16.06 (1.46)	−11.57 (2.73)	−12.57 (1.44)	−13.74 (1.36)	−16.68	−17.85
3.2×10^{-5}	−17.58 (1.98)	−12.52 (3.45)	−13.67 (1.87)	−15.10 (1.84)	−16.25	−17.68
1.0×10^{-5}	−18.48 (2.42)	−13.10 (3.93)	−14.34 (2.25)	−15.95 (2.27)	−16.02	−17.62
PAO	−18.88 (0.36)	−13.31 (0.60)	−14.48 (0.41)	−16.17 (0.32)	−15.76	−17.45
Canonical	−20.16 (3.09)					

diagrams. However, the effect of these missing exchange-type diagrams must be very small when using d-LRPA just for the treatment of close and weak pairs. The accuracy of such a d-LRPA treatment for close and weak pairs was checked by additional calculations with the intermolecular pairs specified as strong pairs, and employing just the lowest OSV threshold $l_{OSV} = 1.0 \times 10^{-4}$. It turns out that the deviations between the full LCC and the approximate d-LRPA treatment are rather small, amounting to 0.52, and to −0.24 kcal/mol for the Watson-Crick, and the stacked dimer, respectively. For a thorough discussion about utilizing d-LRPA in the context of local coupled cluster calculations on intermolecular complexes and clusters, we defer to a later publication.[70]

Table V compares the interaction energies obtained with OSV-LCCSD(T0) for the three thresholds $l_{OSV} = 1.0 \times 10^{-4}$, 3.2×10^{-5}, and 1.0×10^{-5} to those calculated with PAO-LCCSD(T0) and standard BP domains The corresponding LMP2, d-LRPA, LCCSD, and canonical MP2 values are also included. For the hydrogen-bonded G-C/WC dimer, all methods perform quite well, as expected. LMP2 just overshoots slightly, and d-LRPA is in very good agreement with LCCSD. The effect of the triples correction is pretty small, i.e., less than 0.5 kcal/mol. For the π-stacked G-C/S dimer, on the other hand, LMP2 overshoots the LCCSD value by 3–4 kcal/mol, while d-LRPA underestimates it by about one kcal/mol. The triples correction here amounts to about 1.5 kcal/mol.

The basis set superposition error (BSSE), calculated according to the counterpoise procedure, is considerably larger for the OSV than for the PAO based methods, as already observed previously.[39] This is of course not surprising, since the selected OSV orbitals try to optimally reproduce the canonical MP2 amplitudes. We note in passing that the OSV-d-LRPA method has a larger BSSE than the other local methods, which might be related to above mentioned absence of exchange diagrams (here, all the pairs are treated by dLRPA). In any case, we would like to stress the point that in contrast to PAO-based methods, OSV-based methods should not be employed for intermolecular interactions without BSSE correction.

As already demonstrated in Sec. III B, the domain error of the local approximation virtually vanishes when adding to the LCCSD energies the ΔMP2 correction, evaluated in the AO basis of the LCCSD calculation. The corresponding LCCSD+ΔMP2 and LCCSD(T0)+ΔMP2 interaction energies are included in Table V, in addition. Evidently, these energies become virtually independent of the domain approximation or the l_{OSV} threshold: the LCCSD(T0)+ΔMP2 values range between −28.8 and −28.5, and between −17.9 and −17.5 kcal/mol for G-C/WC and G-C/S, respectively. This is consistent with the results for reaction energies in Sec. III B.

Adding instead the ΔMP2(CBS) correction (cf. Sec. III B) yields a good estimate for the LCCSD(T0) complete basis set limit. Based on an aVTZ/aVQZ extrapolation we obtain, on the basis of the OSV-LCCSD(T0) calculations, interaction energies of −29.9 and −18.4 kcal/mol for G-C/WC, and G-C/S, respectively. Adding furthermore to these values, the small corrections obtained by comparison of the full LCC vs. d-LRPA treatment of intermolecular pairs we finally arrive at −30.4 and −18.2 kcal/mol for G-C/WC and G-C/S, respectively. These values are to our knowledge the best estimates for the CCSD(T) basis set limit presently available. They are in excellent agreement with corresponding interaction energies reported in a previous DFT-SAPT study[71] (−30.5 and −17.8 kcal/mol), and compare quite well also with the CCSD(T) CBS limits reported in Ref. 67 (−32.1 and −19.0 kcal/mol). The latter are based on CCSD(T) calculations in small double-ζ basis sets.

Table VI compiles, for the individual OSV- and PAO-L(T0) calculations on G-C/WC, some key numbers determining the efficiency of the method. Evidently, the average and maximum triples domain sizes are considerably larger for the PAO-L(T) method, i.e., the average domain size is about twice as large, the maximum domain size even almost three

TABLE VI. Average/maximum triples domain size, number of 3-ext integrals and triples amplitudes (both in GW), overall number of read operations (in GW), and elapsed times (in h) for the individual G-C/WC calculations, performed on an AMD Opteron 6180 SE server with 128 GB memory and eight scratch disks, forming two striped file systems.

l_{OSV}/E_h	$[ijk]_{\mathrm{avg}}$	$[ijk]_{\mathrm{max}}$	$N(\langle rs\|ti\rangle)$	$N(W_{rst}^{ijk})$	Read ops.	Elapsed time
1.0×10^{-4}	92	153	8.2	9.9	3959.3[a]	2.1[a]
3.2×10^{-5}	127	215	21.5	26.3	14242.8[a]	24.1[a]
1.0×10^{-5}	163	286	45.8	56.5	29991.5[b]	86.2[b]
PAO	248	585	15.1	230.8	11408.6[a]	26.2[a]

[a]On 7 cores, 1024 MW allocated memory per process.
[b]On 4 cores, 2048 MW allocated memory per process.

times as large as for the OSV-L(T) method with $l_{\mathrm{OSV}} = 3.2 \times 10^{-5}$. Consequently, the number of non-vanishing triples amplitudes, or, equivalently, the number of triples intermediates $N(W_{rst}^{ijk})$ is about an order of magnitude larger. On the other hand, the number of required 3-ext integrals $N(\langle rs|ti\rangle)$ is substantially smaller (15.1 vs. 21.5 gigawords (GW)). The algorithm, as it is presently implemented, avoids storage of any triples intermediates and thus is driven by the triples *ijk*. Due to the permutation operators appearing in Eq. (25) the 3-ext integrals are read many times. In order to avoid excessive disk access, a rather sophisticated, internal cache mechanism is employed. However, for calculations with a large number of 3-ext integrals $\langle rs|ti\rangle$ per LMO *i* this internal cache is easily overrun. Table VI shows the overall number of read operations (in GW). These numbers are indeed huge! For the OSV-L(T0) calculation on G-C/WC with $l_{\mathrm{OSV}} = 3.2 \times 10^{-5}$ an amount of about 14 terawords of 3-ext integrals is read, somewhat more than for the PAO calculation, where virtually no memory is left for the integral caching due to the large memory requirements of the triples intermediates. Fortunately, the external global system read buffer still caches most of these integrals such that for a parallel calculation on 7 cores the central processing units run at almost 100% most of the time. Yet for the OSV-L(T0) calculation with $l_{\mathrm{OSV}} = 3.2 \times 10^{-5}$ with its larger 3-ext integral file the disk reads start to become a problem. Overall, the elapsed times for the PAO-L(T0) calculation, and the OSV-L(T0) calculation with $l_{\mathrm{OSV}} = 3.2 \times 10^{-5}$ are comparable, with the OSV-L(T0) calculation being somewhat faster (26.2 vs. 24.1 h).

Taking into account the much more compact triples intermediates of the OSV-L(T) method on the one hand, and the horrendous amount of read operations on the other hand, an alternative algorithm driven by the 3-ext integrals could be envisaged. Such an algorithm stores the triples intermediates W_{rst}^{ijk}, but just passes once through the 3-ext integral list. For the triples intermediates, on the other hand, in total six input/output operations (three reads, three writes) are required. Such an algorithm would save considerably in input/output operations, particularly so for the OSV-L(T) method. The required memory to hold all required W_{rst}^{ijk} to which the actual integral batch contributes, amounts to $N(W_{rst}^{ijk})/N_{\mathrm{occ}}$, where N_{occ} is the number of LMOs. Furthermore, running in parallel on several cores the required memory can be distributed evenly over the individual cores by *a priori* assigning individual triples to individual cores. Hence, the required memory scales properly on parallel computers, and there is no need for synchronization between the individual processes during the evaluation of the triples intermediates. We will report in detail elsewhere on such a new integral driven OSV-L(T) algorithm.[72]

IV. CONCLUSIONS

In this work, we have extended the OSV approach from LCCSD to LCCSD(T), which now constitutes an alternative to the well-established PAO-LCCSD(T) method. In contrast to the latter, no *ad hoc* truncations of the virtual space to domains on the basis of locality is necessary. A single threshold, l_{OSV}, specifies the pair and triple specific virtual spaces completely and automatically. By setting l_{OSV} to a sufficiently small value, the result of the full virtual space is approached arbitrarily closely (although such a calculation would become expensive). A disadvantage of the OSV approach is that BSSE is not as completely avoided as in PAO methods, but turns out to be of similar magnitude as in canonical methods.

Using a threshold of $l_{\mathrm{OSV}} = 3.2 \times 10^{-5}$, the OSV-L(T) method and its L(T0) approximation recover a similar percentage of the canonical (T) correction as the corresponding default PAO methods at a somewhat lower computational cost. This is particularly true for full OSV-L(T) calculations, since the average number of OSVs per triple is usually smaller than the average size of the triples domains in PAO-L(T). The domain error in the OSV-LCCSD(T) calculations can be substantially reduced by adding a MP2 correction. If the ΔMP2(CBS) correction is applied, the canonical CCSD(T)/CBS values can be closely approached. This has been demonstrated for a benchmark of 52 reaction energies.

As an application example, we present OSV-LCCSD(T0) calculations on the Watson-Crick and stacked guanine-cytosine dimers. In contrast to previous LCCSD calculations on that system, the close and weak pairs are not treated at the level of LMP2, but with d-LRPA. The latter includes higher-order ring diagrams and generally provides more reliable intermolecular pair energies than LMP2. As new estimates of the CCSD(T) basis set limit of the interaction energies of these complexes, we propose values of −30.4 and −18.2 kcal/mol for the hydrogen-bonded and the π-stacked dimer, respectively.

ACKNOWLEDGMENTS

G.K.C. acknowledges support from the (U.S.) Department of Energy (DOE), Office of Science Award No. DE-FG02-07ER46432. M.G.S. acknowledges support from the Deutsche Forschungsgemeinschaft (DFG). H.J.W. acknowledges support from the DFG excellence cluster SimTech at the University of Stuttgart.

[1]T. B. Adler, G. Knizia, and H.-J. Werner, J. Chem. Phys. **127**, 221106 (2007).
[2]D. P. Tew, W. Klopper, C. Neiss, and C. Hättig, Phys. Chem. Chem. Phys. **9**, 1921 (2007).
[3]G. Knizia, T. B. Adler, and H.-J. Werner, J. Chem. Phys. **130**, 054104 (2009).
[4]C. Hättig, D. P. Tew, and A. Köhn, J. Chem. Phys. **132**, 231102 (2010).

[5]H.-J. Werner, G. Knizia, and F. R. Manby, Mol. Phys. **109**, 407 (2011).
[6]T. B. Adler and H.-J. Werner, J. Chem. Phys. **135**, 144117 (2011).
[7]M. Urban, J. Noga, S. J. Cole, and R. J. Bartlett, J. Chem. Phys. **83**, 4041 (1985).
[8]J. A. Pople, M. Head–Gordon, and K. Raghavachari, J. Chem. Phys. **87**, 5968 (1987).
[9]K. Raghavachari, G. W. Trucks, J. A. Pople, and M. Head-Gordon, Chem. Phys. Lett. **157**, 479 (1989).
[10]N. H. F. Beebe and J. Linderberg, Int. J. Quantum Chem. **12**, 683 (1977).
[11]D. W. O'Neal and J. Simons, Int. J. Quantum Chem. **36**, 673 (1989).
[12]H. Koch, A. S. de Merás, and T. B. Pedersen, J. Chem. Phys. **118**, 9481 (2003).
[13]F. Aquilante, T. B. Pedersen, and R. Lindh, J. Chem. Phys. **126**, 194106 (2007).
[14]F. Aquilante and T. B. Pedersen, Chem. Phys. Lett. **449**, 354 (2007).
[15]T. Kinoshita, O. Hino, and R. J. Bartlett, J. Chem. Phys. **119**, 7756 (2003).
[16]J. L. Whitten, J. Chem. Phys. **58**, 4496 (1973).
[17]F. Weigend, M. Kattannek, and R. Ahlrichs, J. Chem. Phys. **130**, 164106 (2009).
[18]T. S. Chwee and E. A. Carter, J. Chem. Phys. **132**, 074104 (2010).
[19]H.-J. Werner, F. R. Manby, and P. J. Knowles, J. Chem. Phys. **118**, 8149 (2003).
[20]F. R. Manby, J. Chem. Phys. **119**, 4607 (2003).
[21]M. Schütz and F. R. Manby, Phys. Chem. Chem. Phys. **5**, 3349 (2003).
[22]H.-J. Werner and M. Schütz, J. Chem. Phys. **135**, 144116 (2011).
[23]P. Pulay, Chem. Phys. Lett. **100**, 151 (1983).
[24]S. Saebø and P. Pulay, Chem. Phys. Lett. **113**, 13 (1985).
[25]P. Pulay and S. Saebø, Theor. Chim. Acta **69**, 357 (1986).
[26]S. Saebø and P. Pulay, J. Chem. Phys. **86**, 914 (1987).
[27]S. Saebø and P. Pulay, J. Chem. Phys. **88**, 1884 (1988).
[28]C. Edmiston and M. Krauss, J. Chem. Phys. **42**, 1119 (1965).
[29]W. Meyer, Int. J. Quantum Chem. **S5**, 341 (1971).
[30]W. Meyer, J. Chem. Phys. **58**, 1017 (1973).
[31]R. Ahlrichs, F. Driessler, H. Lischka, V. Staemmler, and W. Kutzelnigg, J. Chem. Phys. **62**, 1235 (1975).
[32]V. Staemmler and R. Jaquet, Theor. Chim. Acta. **59**, 487 (1981).
[33]F. Neese, F. Wennmohs, and A. Hansen, J. Chem. Phys. **130**, 114108 (2009).
[34]F. Neese, A. Hansen, and D. G. Liakos, J. Chem. Phys. **131**, 064103 (2009).
[35]A. Hansen, D. G. Liakos, and F. Neese, J. Chem. Phys. **135**, 214102 (2011).
[36]D. P. Tew, B. Helmich, and C. Hättig, J. Chem. Phys. **135**, 074107 (2011).
[37]C. Krause and H.-J. Werner, Phys. Chem. Chem. Phys. **14**, 7591 (2012).
[38]J. Yang, Y. Kurashige, F. R. Manby, and G. K. L. Chan, J. Chem. Phys. **134**, 044123 (2011).
[39]J. Yang, G. K. L. Chan, F. R. Manby, M. Schütz, and H.-J. Werner, J. Chem. Phys. **136**, 144105 (2012).
[40]M. Schwilk, C. Krause, and H.-J. Werner, (to be published).
[41]C. Hättig, D. P. Tew, and B. Helmich, J. Chem. Phys. **136**, 204105 (2012).
[42]M. Schütz and H.-J. Werner, Chem. Phys. Lett. **318**, 370 (2000).
[43]M. Schütz, J. Chem. Phys. **113**, 9986 (2000).
[44]M. Schütz, J. Chem. Phys. **116**, 8772 (2002).
[45]Y. Liu, "A local coupled-cluster method for open-shell molecules: DF-LUCCSD(T)," M.S. thesis, University of Stuttgart, 2011.
[46]Y. Kurashige, J. Yang, G. K. L. Chan, and F. R. Manby, J. Chem. Phys. **136**, 124106 (2012).
[47]D. Kats, D. Usvyat, and M. Schütz, Phys. Chem. Chem. Phys. **10**, 3430 (2008).
[48]P. E. Maslen, A. D. Dutoi, M. S. Lee, Y. Shao, and M. Head-Gordon, Mol. Phys. **103**, 425 (2005).
[49]M. Kobayashi and H. Nakai, J. Chem. Phys. **131**, 114108 (2009).
[50]W. Li, P. Piecuch, J. R. Gour, and S. Li, J. Chem. Phys. **131**, 114109 (2009).
[51]Z. Rolik and M. Kállay, J. Chem. Phys. **135**, 104111 (2011).
[52]C. Hampel and H.-J. Werner, J. Chem. Phys. **104**, 6286 (1996).
[53]M. Schütz and H.-J. Werner, J. Chem. Phys. **114**, 661 (2001).
[54]J. W. Boughton and P. Pulay, J. Comput. Chem. **14**, 736 (1993).
[55]H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, and M. Schütz *et al.*, MOLPRO, version 2010.1, a package of *ab initio* programs, 2011, see http://www.molpro.net.
[56]H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, and M. Schütz, Comput. Mol. Sci. **2**, 242 (2012).
[57]J. Pipek and P. G. Mezey, J. Chem. Phys. **90**, 4916 (1989).
[58]See supplementary material at http://dx.doi.org/10.1063/1.4789415 for Cartesian coordinates of the test molecules, related energies, and additional timings of the test calculations.
[59]F. Weigend, A. Köhn, and C. Hättig, J. Chem. Phys. **116**, 3175 (2002).
[60]K. A. Peterson, T. B. Adler, and H.-J. Werner, J. Chem. Phys. **128**, 084102 (2008).
[61]J. G. Hill, K. A. Peterson, G. Knizia, and H.-J. Werner, J. Chem. Phys. **131**, 194105 (2009).
[62]F. Weigend, Phys. Chem. Chem. Phys. **4**, 4285 (2002).
[63]H.-J. Werner and K. Pflüger, Annu. Rep. Comp. Chem. **2**, 53 (2006).
[64]H.-J. Werner, J. Chem. Phys. **129**, 101103 (2008).
[65]T. B. Adler and H.-J. Werner, J. Chem. Phys. **130**, 241101 (2009).
[66]T. B. Adler, H.-J. Werner, and F. R. Manby, J. Chem. Phys. **130**, 054106 (2009).
[67]P. Jurečka, J. Šponer, J. Černý, and P. Hobza, Phys. Chem. Chem. Phys. **8**, 1985 (2006).
[68]M. Schütz, G. Rauhut, and H.-J. Werner, J. Phys. Chem. A **102**, 5997 (1998).
[69]J. G. Hill, J. A. Platts, and H.-J. Werner, Phys. Chem. Chem. Phys. **8**, 4072 (2006).
[70]O. Masur and M. Schütz, (to be published).
[71]A. Hesselmann, G. Jansen, and M. Schütz, J. Am. Chem. Soc. **128**, 11730 (2006).
[72]M. Schütz, (to be published).