The following article appeared in J. Chem. Phys. 142, 164108 (2015) and may be found at http://dx.doi.org/10.1063/1.4918772.

On the use of Abelian point group symmetry in density-fitted local MP2 using various types of virtual orbitals

Christoph Köppl and Hans-Joachim Werner

Citation: The Journal of Chemical Physics **142**, 164108 (2015); doi: 10.1063/1.4918772
View online: http://dx.doi.org/10.1063/1.4918772
View Table of Contents: http://scitation.aip.org/content/aip/journal/jcp/142/16?ver=pdfcov
Published by the AIP Publishing

On the use of Abelian point group symmetry in density-fitted local MP2 using various types of virtual orbitals

Christoph Köppl and Hans-Joachim Werner
Institut für Theoretische Chemie, Universität Stuttgart, Pfaffenwaldring 55, D-70569 Stuttgart, Germany

(Received 23 February 2015; accepted 10 April 2015; published online 29 April 2015)

Electron correlation methods based on symmetry-adapted canonical Hartree-Fock orbitals can be speeded up significantly in the well known group theoretical manner, using the fact that integrals vanish unless the integrand is totally symmetric. In contrast to this, local electron correlation methods cannot benefit from such simplifications, since the localized molecular orbitals (LMOs) generally do not transform according to irreducible representations of the underlying point group symmetry. Instead, groups of LMOs become symmetry-equivalent and this can be exploited to accelerate local calculations. We describe an implementation of such a symmetry treatment for density-fitted local Møller-Plesset perturbation theory, using various types of virtual orbitals: Projected atomic orbitals, orbital specific virtuals, and pair natural orbitals. The savings by the symmetry treatment are demonstrated by calculations for several large molecules having different point group symmetries. Benchmarks for the parallel execution efficiency of our method are also presented. © *2015 AIP Publishing LLC*. [http://dx.doi.org/10.1063/1.4918772]

I. INTRODUCTION

The use of molecular point group symmetry is beneficial in many ways. First and foremost, *ab initio* methods such as Møller-Plesset perturbation (MP2) theory[1,2] or coupled-cluster (CC) theory can be significantly accelerated—often by an order of magnitude or even more—dependent on the considered molecule and its underlying point group symmetry. The simplifications that permit such performance gains are possible if all orbitals are symmetry-adapted and belong to specific irreducible representations of the point group of the considered molecule. This is, for example, the case for canonical Hartree-Fock (HF) orbitals. The group theoretical vanishing integral rule states that an integral is necessarily zero if the integrand does not belong to the totally symmetric representation. Integral matrices originating from canonical orbitals thus exhibit a sparse block structure. The generality of the group theoretical approach allows its application for all canonical electron correlation methods.

The disadvantage of using canonical HF orbitals is that they usually extend over large parts of the molecule. Therefore, the short-range character of dynamical electron correlation cannot be exploited, and the computational resources (CPU-time, memory, disk) increase very steeply with the molecular size. This unphysical steep scaling can be strongly reduced by local correlation approaches such as local Møller-Plesset perturbation (LMP2) theory,[3–17] local coupled cluster with single and double excitation operators (LCCSD),[18–21] or LCCSD with perturbative treatment of triple excitations [LCCSD(T)],[22–25] and even linear scaling has been demonstrated for methods up to LCCSD(T).[17,19,22–25] However, the localized molecular orbitals (LMOs) used in these methods are no longer symmetry-adapted, and therefore, the conventional group theoretical approach is not applicable. Instead, one can use the fact that LMOs appear in groups of equivalent orbitals, which can be transformed into each other by symmetry operations. This is related to work presented long time ago by King and Dupuis[26] in the context of HF theory.

Local correlation methods naturally target at treating large molecules in an efficient manner. One might ask the question if large molecules belong to specific molecular point groups at all or whether they rather tend to be asymmetric. An answer to the latter question can, at least to some extent, be obtained if natural product synthesis is considered. In nature, symmetry is a frequently occurring pattern, which can facilitate the synthesis of natural products, e.g., by dimerization of monomers to complex molecular structures. Moreover, this construction principle is not only followed in nature, but is applied in the synthesis of natural products by chemists in the laboratory as well. In this context, it is also worthwhile to note that a recent study found that a notable number of natural product structures seem to obey molecular point group symmetry—C_2 being most abundant among bilaterally symmetrical natural compounds.[27] Some of our largest example molecules are therefore natural products, e.g., (−)-glabrescol (a precursor to steroids), or nonactin and elaiophylin (antibiotics). Also many inorganic and metallo-organic complexes possess high symmetry and this has been exploited in many previous theoretical studies, see, for example, Ref. 28.

Supported by these examples, we demonstrate in this paper that Abelian point group symmetry can be applied in local correlation methods, leading to a significant reduction of the computational cost, although the savings are less than for canonical methods. We will consider three types of local virtual orbitals, namely, projected atomic orbitals (PAOs),[3] orbital specific virtuals (OSVs),[29,30] and pair natural orbitals (PNOs),[17,31–39] all of which have previously been used in LMP2 and LCCSD calculations. Even though in the current paper only LMP2 is considered, the presented scheme is

general and can in future be applied in a similar way to speed-up higher-order methods such as LCCSD(T), for the computation of molecular properties or excitation energies using local response methods or for geometry optimizations of large symmetric molecules.

II. THEORY

In the following, we will first review some general aspects of our recently developed PAO-, OSV-, and PNO-LMP2 programs.[17] Subsequently, a detailed description of our symmetry treatment follows. Localized occupied HF orbitals (LMOs) will be denoted by indices i,j,k,l; PAOs by indices r,s,t,u; OSVs by indices p,q; PNOs by indices a,b,c,d; atomic orbitals (AOs) by indices μ,ν,σ,ρ; and fitting functions by indices A,B.

A. General aspects

1. Local occupied and virtual orbitals

The occupied Hartree-Fock orbitals can be localized using the methods of Foster and Boys,[40,41] Edmiston and Ruedenberg,[42] or Pipek and Mezey.[43–45] Alternatively, natural bonding orbitals (NBOs)[46] or intrinsic bonding orbitals[47] (IBOs) can be used. The symmetry properties of the resulting LMOs $|i\rangle \equiv |\phi_i\rangle$ obtained by these methods may be different. For example, the localization of Foster and Boys is known to often produce for double bonds equivalent "banana" bonds, while Pipek-Mezey (PM) localization normally yields non-equivalent σ- and π-bonds. This may affect the efficiency of the symmetry treatment but has no effect on the general aspects. In the current work, we use IBOs, which are easy to compute and very insensitive to the basis set.[47] Usually, IBOs resemble PM orbitals.

In local electron correlation methods, excitations from the LMOs are restricted to subspaces (domains) of local virtual orbitals. For example, the virtual orbitals r in singly excited configurations Φ_i^r are restricted to *orbital domains* $r \in [i]$. Similarly, double excitations Φ_{ij}^{rs} are restricted to *pair domains* $r,s \in [ij]$. The number of orbitals in these domains necessary to reach a certain accuracy depends crucially on the choice of the virtual orbitals.[17,48] In our program,[17] we use PAOs, OSVs, and PNOs, and these are generated one after the other by a sequence of transformations: AO → PAO → OSV → PNO. The domain sizes decrease in each of these steps. For example, in order to obtain 99.7% of the canonical correlation energy, the PAO, OSV, and PNO pair domains, which depend on the basis set size, include 180-420, 115-330, and 30-120 orbitals for the molecules investigated in the current paper, respectively.[17,48] Thus, the number of LMP2 amplitudes, which depends quadratically on the pair domain sizes, is in the PNO basis one order of magnitude smaller than in the PAO basis. This ratio is even much larger if more extended PAO domains are used, as in Ref. 17. Since the CPU time for solving the LMP2 equations depends cubically on the domain sizes, the time savings may even be three orders of magnitude. The three-step procedure is needed to obtain linear scaling and to minimize the memory and CPU requirements.[17] More detailed information concerning the domain sizes and their spatial extent for the investigated molecules of the present paper is shown in Figure 1 and Table II within the supplementary material.[49] We note that the OSV step can also be skipped, at the expense of some more CPU-time and memory. This is, for example, done in the method of Riplinger and Neese.[21,25] Leaving out the OSV and PNO steps leads to the PAO-LMP2 method and leaving out only the PNO step yields the OSV-LMP2 method. All these variants are possible with our program, and examples will be given in Sec. III.

For the symmetry treatment to be described in Sec. II C, it is necessary to define the sequence of transformations needed to generate the three types of virtual orbitals, and this is therefore summarized in the following. The exact definition of the various domains and transformation matrices can be found in previous papers.[17,48]

The PAOs $|\phi_r\rangle$ are obtained by projecting out the occupied (HF) orbital space $\{\phi_i\}$ from individual atomic orbitals $|\chi_r^{\mathrm{AO}}\rangle$,

$$|\phi_r\rangle = \left(1 - \sum_i |\phi_i\rangle\langle\phi_i|\right)|\chi_r^{\mathrm{AO}}\rangle. \tag{1}$$

The matrix of expansion coefficients of the PAOs in the AO basis is

$$[\mathbf{P}]_{\mu r} = [\mathbf{1} - \mathbf{L}\mathbf{L}^\dagger\mathbf{S}\mathbf{C}^{\mathrm{AO}}]_{\mu r}, \tag{2}$$

where $L_{\mu i}$ is the rectangular matrix of LMO coefficients, $S_{\mu\nu} = \langle\chi_\mu|\chi_\nu\rangle$ is the overlap matrix of the basis functions $\{\chi_\mu\}$, and $C_{\mu r}^{\mathrm{AO}}$ is, in general, a block-diagonal matrix describing a suitable set of AOs for each center. If generally contracted basis sets are used, $\mathbf{C}^{\mathrm{AO}}$ can be taken to be the unit matrix. PAO orbital domains $[i]_{\mathrm{PAO}}$ can be defined as described by Boughton and Pulay[44] or alternatively using NBO[46] or IBO[47] partial charges. We use the latter method, as described in detail in Ref. 17. Pair domains $[ij]_{\mathrm{PAO}}$ are then defined as the union of the respective orbital domains, i.e., $[ij]_{\mathrm{PAO}} = [i]_{\mathrm{PAO}} \cup [j]_{\mathrm{PAO}}$.[3–6,18] In terms of the orbital basis, the PAO pair domains are defined by submatrices $\mathbf{P}^{(ij)}$ which contain all columns of $\mathbf{P}$ that belong to the corresponding pair domain.

The transformation of the PAOs to OSVs is described by transformation matrices $\mathbf{Q}^{(i)}$,

$$|\phi_p^{(i)}\rangle = \sum_{r\in[i]_{\mathrm{PAO}}} |\phi_r\rangle Q_{rp}^{(i)}, \quad p \in [i]_{\mathrm{OSV}}. \tag{3}$$

The OSV orbital domains $[i]_{\mathrm{OSV}}$ are defined using an occupation number threshold T_{OSV}.[17] The next step is the generation of the OSV pair domains $[ij]_{\mathrm{OSV}}$. As for the PAO pair domains, these are obtained as the union of the respective orbital domains, i.e., $[ij]_{\mathrm{OSV}} = [i]_{\mathrm{OSV}} \cup [j]_{\mathrm{OSV}}$. The transformation from the PAO pair domains to the OSV pair domains is represented by transformation matrices

$$\mathbf{Q}^{(ij)} = (\mathbf{Q}^{(i)}|\mathbf{Q}^{(j)}). \tag{4}$$

This notation means that the matrix $\mathbf{Q}^{(ij)}$ contains the columns of both $\mathbf{Q}^{(i)}$ and $\mathbf{Q}^{(j)}$. Note that the rows of the latter matrices must be scattered so that they correspond to the rows of the PAO transformation matrix $\mathbf{P}^{(ij)}$. Since the OSVs from the domains $[i]_{\mathrm{OSV}}$ and $[j]_{\mathrm{OSV}}$ $(i \neq j)$ are non-orthogonal and possibly linearly dependent, the pair domain must be orthonormalized. Redundant vectors are eliminated by singular value

decomposition. Furthermore, the OSVs are transformed to diagonalize the Fock matrix within the domain. In total, the resulting *pseudo-canonical* (pc) OSVs $|\bar{\phi}_p^{(ij)}\rangle$ are obtained from the non-orthogonal ones by transformation matrices $\mathbf{V}^{(ij)}$,

$$|\bar{\phi}_p^{(ij)}\rangle = \sum_{q\in[ij]_{\mathrm{OSV}}} |\phi_q^{(ij)}\rangle V_{qp}^{(ij)}, \quad p \in [ij]_{\mathrm{OSV(pc)}}, \tag{5}$$

so that

$$\langle\bar{\phi}_p^{(ij)}|\bar{\phi}_q^{(ij)}\rangle = \delta_{pq}, \tag{6}$$

$$\langle\bar{\phi}_p^{(ij)}|\hat{f}|\bar{\phi}_q^{(ij)}\rangle = \delta_{pq}\varepsilon_p^{(ij)}, \quad p,q \in [ij]_{\mathrm{OSV(pc)}}. \tag{7}$$

The orbitals in different pair domains $[ij]_{\mathrm{OSV(pc)}}$ and $[kl]_{\mathrm{OSV(pc)}}$ are non-orthogonal.

The final step is to transform the orbitals in each OSV pair domain to PNOs. Again, the domains are defined by an occupation number threshold T_{PNO} and the PNOs are made pseudo-canonical. The PNOs for the diagonal pairs (ii) are identical to the OSVs for the corresponding orbitals (i). The transformation from $[ij]_{\mathrm{OSV(pc)}}$ to $[ij]_{\mathrm{PNO}}$ is described by transformation matrices $\mathbf{U}^{(ij)}$,

$$|\phi_a^{(ij)}\rangle = \sum_{p\in[ij]_{\mathrm{OSV(pc)}}} |\bar{\phi}_p^{(ij)}\rangle\, U_{pa}^{(ij)}, \quad a \in [ij]_{\mathrm{PNO}}, \tag{8}$$

so that exactly as for OSVs,

$$\langle\phi_a^{(ij)}|\phi_b^{(ij)}\rangle = \delta_{ab}, \tag{9}$$

$$\langle\phi_a^{(ij)}|\hat{f}|\phi_b^{(ij)}\rangle = \delta_{ab}\varepsilon_a^{(ij)}, \quad a,b \in [ij]_{\mathrm{PNO}}. \tag{10}$$

Note that there is no need to distinguish between domains for non-orthogonal and orthogonal PNOs, since the transformation matrices $\mathbf{U}^{(ij)}$ are unitary, and therefore, the PNOs for one pair are orthonormal from the beginning.

In terms of either the orbital basis $\{\chi_\mu\}$ or of the PAO basis $\{\phi_r\}$, the PNOs can now be expressed as

$$|\phi_a^{(ij)}\rangle = \sum_{r\in[ij]_{\mathrm{PAO}}} |\phi_r\rangle W_{ra}^{(ij)} = \sum_{\mu} |\chi_\mu\rangle C_{\mu a}^{(ij)}, \tag{11}$$

$$W_{ra}^{(ij)} = [\mathbf{Q}^{(ij)}\mathbf{V}^{(ij)}\mathbf{U}^{(ij)}]_{ra}, \tag{12}$$

$$C_{ra}^{(ij)} = [\mathbf{P}^{(ij)}\mathbf{W}^{(ij)}]_{\mu a}. \tag{13}$$

In our program, only the matrices $\mathbf{P}$, $\mathbf{Q}^{(i)}$, and $\mathbf{W}^{(ij)}$ are stored, while $\mathbf{P}^{(ij)}$, $\mathbf{Q}^{(ij)}$, and $\mathbf{U}^{(ij)}$ are generated on the fly when needed. It is also possible to neglect very small PAO coefficients $P_{\mu r}^{(ij)}$ and to restrict the AO index μ to domains.[17] Then the sizes of all transformation matrices scale linearly with the molecular size.

2. The LMP2 amplitude equations

We now turn our attention to the equations needed to calculate the LMP2 amplitudes and the correlation energy. The PNO-LMP2 Hylleraas functional can be written as

$$E^{\mathrm{PNO-LMP2}} = \sum_{i,j} \sum_{a,b\in[ij]_{\mathrm{PNO}}} \left(2T_{ab}^{ij} - T_{ba}^{ij}\right)\left(K_{ab}^{ij} + R_{ab}^{ij}\right), \tag{14}$$

where $K_{ab}^{ij} = (ai|bj)$ are two-electron repulsion integrals and R_{pq}^{ij} are the double residuals for PNO-LMP2,[17]

$$R_{ab}^{ij} = K_{ab}^{ij} + \left(\varepsilon_a^{ij} + \varepsilon_b^{ij} - f_{ii} - f_{jj}\right) T_{ab}^{ij} + G_{ab}^{ij} + G_{ba}^{ji}, \quad a,b \in [ij]_{\mathrm{PNO}} \tag{15}$$

$$G_{ab}^{ij} = -\sum_{k\neq i} f_{ik} \sum_{c,d\in[kj]_{\mathrm{PNO}}} S_{ac}^{ij,kj} T_{cd}^{kj} S_{db}^{kj,ij}. \tag{16}$$

Here, $f_{ik} = \langle\phi_i|\hat{f}|\phi_k\rangle$ are the Fock matrix elements in the LMO basis, and $S_{ab}^{ij,kl} = \langle\phi_a^{(ij)}|\phi_b^{(kl)}\rangle$ are PNO overlap matrices. These are needed because the PNOs for different orbital pairs are non-orthogonal. The PAO- and OSV-LMP2 residual equations are similar. The amplitudes T_{ab}^{ij} are obtained by solving the amplitude equations,

$$R_{ab}^{ij} = 0 \quad \forall\ i \geq j;\ a,b \in [ij]_{\mathrm{PNO}}. \tag{17}$$

Due to the non-zero off-diagonal elements f_{ik} of the Fock matrix, these equations need to be solved iteratively. This can be done by solving in each iteration Eqs. (15)–(17) with fixed matrices $\mathbf{G}^{ij}$ (the latter are computed using the amplitudes of the previous iteration).

3. Pair approximations

So far, we have only considered the orbital spaces and domain approximations. The second basic approximation in local correlation methods is that the correlation of distant electron pairs can be approximated or even neglected. This is based on the fact that the pair correlation energies decrease approximately with R_{ij}^{-6}, where R_{ij} is the distance between the charge centroids of two LMOs $|i\rangle$ and $|j\rangle$. The number of distant pairs ij increases quadratically with the molecular size. Therefore, even though the individual distant pair energies E_{ij} may be negligibly small, their sum can still be significant. The pair energies E_{ij} can be approximated by multipole expansions,[9,13] the simplest being the dipole-dipole approximation. Following Riplinger and Neese,[21] they can be computed in the pseudo-canonical OSV basis $|\bar{p}^{(i)}\rangle \equiv |\bar{\phi}_p^{(ii)}\rangle$ as

$$E_{ij} = -\frac{8}{R_{ij}^6} \sum_{\bar{p}\in[i]_{\mathrm{OSV(pc)}}} \sum_{\bar{q}\in[j]_{\mathrm{OSV(pc)}}} \frac{[\langle i|\mathbf{r}|\bar{p}^{(i)}\rangle \cdot \langle j|\mathbf{r}|\bar{q}^{(j)}\rangle]^2}{\varepsilon_p^{(i)} + \varepsilon_q^{(j)} - f_{ii} - f_{jj}}. \tag{18}$$

The calculation time for these energy contributions is negligible in practice, even for very large molecules. One can therefore employ this approximation in a twofold way:[21] First, it is used to determine which pairs need to be treated explicitly by solving Eq. (17), using a threshold T_{dist}. Typically, this threshold is chosen to be 10^{-6}. Second, the remaining pair energies are approximated in this way and added to the LMP2 energy.

B. Evaluation of the electron repulsion integrals

One of the most time consuming steps within PNO-LMP2 is the evaluation of two-electron integrals,

$$(ai|bj) = \int \phi_a^{(ij)}(\mathbf{r}_1)\phi_i(\mathbf{r}_1)\frac{1}{r_{12}}\phi_b^{(ij)}(\mathbf{r}_2)\phi_j(\mathbf{r}_2)d\mathbf{r}_1 d\mathbf{r}_2 \tag{19}$$

in the PNO basis. As explained earlier, the PNOs a, b are different for each pair ij. In large molecules, there may be thousands or even millions of PNOs, and it is therefore impractical to compute these integrals directly by transformation of the AO integrals $(\mu\nu|\rho\sigma)$. Instead, we first compute the integrals $(ri|sj)$ in the PAO basis using the local density fitting technique. Subsequently, they are transformed pairwise into the OSV or PNO basis using the matrices $\mathbf{V}^{(ij)}$ and $\mathbf{W}^{(ij)}$, respectively, cf. Sec. II A 1.

1. Local density fitting

In the local density fitting approach,[16,17,20] the integrals $(ri|sj)$ in the LMO/PAO basis are computed as

$$(ri|sj) \approx \sum_{\overline{A}\in[ij]_{\text{fit}}} (\overline{A}|ri)(\overline{A}|sj). \tag{20}$$

The fitting pair domains $[ij]_{\text{fit}}$ are the union of fitting orbital domains $[i]_{\text{fit}}$ and $[j]_{\text{fit}}$. The integrals $(\overline{A}|ri)$ occurring in these equations are determined by solving the linear equation system,[17]

$$(B|ri) = \sum_{B\in[ij]_{\text{fit}}} G_{AB}(\overline{A}|ri), \tag{21}$$

where

$$(B|ri) = \int \chi_B(\mathbf{r}_1)\frac{1}{r_{12}}\phi_r(\mathbf{r}_2)\phi_i(\mathbf{r}_2)d\mathbf{r}_1 d\mathbf{r}_2 \tag{22}$$

are three index integrals with fitting basis functions $\{\chi_B\}$. Since these integrals are needed for all pairs (ij) with a fixed i, they must be computed in the *united* domains of all $[j]_{\text{PAO}}$ and $[j]_{\text{fit}}$ for which j forms a non-distant pair with i.

The matrix $\mathbf{G}$ is defined via a Cholesky decomposition of the Coulomb metric in the fitting basis as

$$(A|B) = [\mathbf{J}]_{AB} = [\mathbf{G}\mathbf{G}^T]_{AB}. \tag{23}$$

If very small LMO and PAO coefficients $L_{\mu i}$ and $P_{\mu r}$, respectively, are neglected, linear scaling of the computational effort and the memory requirements can be reached. More details about this procedure and the choice of the fitting domains $[i]_{\text{fit}}$ can be found in Ref. 17.

C. Use of molecular symmetry

In Subsections II C 1–II C 4, we will present the symmetry properties of the *local* occupied and virtual orbitals as needed by our symmetry treatment. Using these properties, we then show how (Abelian) point group symmetry can be applied in local correlation methods without making use of the group-theoretical vanishing integral rule. Instead, we aim to reduce the number of orbital pairs (ij) for which the amplitude equations must be solved as much as possible and use symmetry-equivalency considerations to achieve this goal.

1. LMOs

A necessary condition to apply our symmetry treatment in local correlation methods is the symmetric localization of the occupied orbitals. This means that the employed localization procedure has to result in LMOs which are either symmetry-adapted and thus follow

$$|L_{\mu i}| = |L_{\nu i}| \quad \forall \quad \mu \leftrightarrow \nu, \tag{24}$$

just as canonical orbitals, where μ and ν refer to basis functions which are transformed into each other by symmetry operations. Or the LMOs are symmetry-equivalent and can be transformed into each other as

$$L_{\mu i} = \sum_{\nu} O_{\mu\nu}^{(i,j)} L_{\nu j} \quad \forall \quad i \leftrightarrow j. \tag{25}$$

For Abelian point groups, the transformation matrices $O_{\mu\nu}^{(i,j)}$ have only one element ± 1 in each row. The information can therefore be stored in vectors, merely addressing the position of particular orbital coefficients as well as the correct sign relations. A more detailed discussion concerning the properties of the orbitals and their preparation for a symmetry treatment is given in the supplementary material.[49]

Figure 1 depicts the symmetry properties of some LMOs of the ethene molecule. The LMOs are separated into two sets: two symmetry adapted orbitals transform according to Eq. (24) and represent σ- and π-CC bonds (blue), respectively, while the remaining 4 symmetry-equivalent occupied orbitals (red) each represent a localized σ-CH bond. One of the four symmetry-equivalent orbitals (red) can then be considered as a generator orbital from which the other three orbitals can be generated by making use of symmetry operations belonging to the D_{2h} point group.

The symmetry properties of the virtual orbitals are as important as the ones of the occupied orbitals. Next, we will describe their type-specific symmetry properties in some detail.

2. PAOs

Since the projector given in parentheses of Eq. (1) is totally symmetric, the symmetry properties of PAOs follow those of the AOs. The application of a symmetry operation to the AOs centered on one atom transforms them either to the ones centered on a symmetry-equivalent atom or simply multiplies them (dependent on the AO type) by ± 1 but does

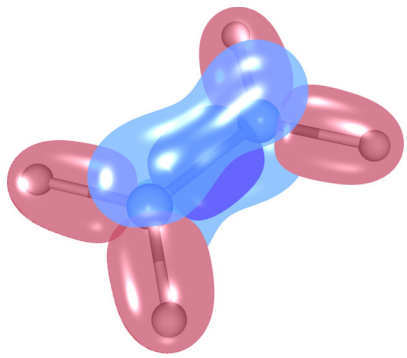


FIG. 1. Visualization[50] of the occupied valence orbitals of ethene (C_2H_4) after an IBO localization[47] using the cc-pVDZ basis set.[51] The blue orbitals represent localized symmetry-adapted σ- and π-CC bonds, whereas the red orbitals represent 4 symmetry-equivalent local σ-CH bonds. One of the red orbitals can serve as a generator: access to the remaining orbitals is then provided by application of the symmetry operations of the D_{2h} point group.

not change their location. The latter case occurs whenever the location of an atom is invariant to the symmetry operation being employed. With this, PAOs are either symmetry-adapted or symmetry-equivalent—dependent on the AOs they originate from. For Abelian point groups, the information describing the transformation properties of the AOs and PAOs can be stored in a single vector for each symmetry operation, as for the LMOs.

3. OSVs

When using symmetry, merely a symmetry-unique set of the matrices $\mathbf{Q}^{(i)}$ is needed because the ones for symmetry equivalent diagonal pairs, i.e., where $i \leftrightarrow j$ is fulfilled, can be obtained according to

$$Q^{(i)}_{rq} = \sum_s O^{(i,j)}_{rs} Q^{(j)}_{sq} \quad \forall \; i \leftrightarrow j. \tag{26}$$

If the matrix $\mathbf{C}^{\mathrm{AO}}$ in Eq. (2) is taken to be a unit matrix, the transformation matrices $O^{(i,j)}_{rs}$ are the same as those in Eq. (25), since the PAOs transform exactly as the basis functions.

As has been outlined in Sec. II A 1, the orbitals in each domain $[ij]_{\mathrm{OSV}}$ are orthogonalized and made pseudo-canonical, which are described by a second transformation $\mathbf{V}^{(ij)}$, so that the total transformation from the non-orthogonal PAO domain $[ij]_{\mathrm{PAO}}$ to the pseudo-canonical (orthonormal) OSV domain $[ij]_{\mathrm{PAO}}$ is $\mathbf{Q}^{(ij)}\mathbf{V}^{(ij)}$. Since we aim to use simple symmetry relations between symmetry-equivalent amplitudes within the residual calculation, cf. Eqs. (15) and (16), the transformation matrices $\mathbf{V}^{(ij)}$ for symmetry-equivalent non-diagonal pairs, i.e., where $ij \leftrightarrow kl$ is fulfilled, are generated such that amplitudes follow the equations presented in Table I.

4. PNOs

While the symmetry relations for diagonal OSV and PNO pairs are the same, the transformation matrices from the OSV to the PNO basis for symmetry-equivalent non-diagonal pairs ij can again be chosen such that two symmetry-related PNO spaces become exactly equivalent—similar to the former OSV case. It is interesting to note that, here, this beneficial property results from the fact that PNOs are pair-specific from the onset and, therefore, the symmetry-relations in Table I are easier to achieve for PNOs than for OSVs.

5. LMP2 iterations

Now that the symmetry-relations of the occupied and various types of virtual orbitals have been described in detail, we will draw our attention to the question how this information can beneficially be used to speed up a LMP2 calculation.

Equation (14) shows that the energy contributions for each symmetry-equivalent orbital pair $ij \leftrightarrow kl$ should be identical, and therefore, it is sufficient to only evaluate the pair energies of a symmetry-unique set of orbital pairs $ij \in \bar{P}$. Accordingly, the residual matrices $\mathbf{R}^{ij}$ need only to be computed for the unique pairs. This requires the full set of pair amplitude matrices $\mathbf{T}^{kl}$. However, it is possible to generate the amplitudes of the symmetry equivalent pairs from the amplitudes of the generator pair according to the relations summarized in Table I. These formulae result from the symmetry properties of the various virtual orbitals. Since $\mathbf{O}^{kl,ij}$ are matrices relating symmetry-equivalent PAOs for orbital pairs $i \geq j$ and $k \geq l$, the resulting amplitude matrix $\mathbf{T}^{kl}$ must be transposed if the symmetry operation applied to ij yields lk. This is achieved by the operators $\hat{P}^{kl}$. In addition, the phase factors C^{kl} respect sign changes, which may occur between symmetry-equivalent occupied orbital pairs.

The iterative optimization procedure is thus driven by a reduced unique set of orbital pairs. In order to avoid overhead by repeated generations of symmetry-equivalent amplitude matrices during the evaluation of the residuals, the full set of amplitudes is generated in the beginning of each iteration and kept in memory.

6. Other evaluation steps

From the discussion in Secs. II B and II C 1–II C 4, it follows that the integrals $K^{ij}_{ab} = (ai|bj)$ only need to be evaluated for a symmetry-unique set of orbital pairs. This also means that only a reduced set of three-index integrals needs to be evaluated, cf. Eqs. (20) and (21).

Even more so, only a symmetry-unique set of three-index integrals suffices. As an example, consider the calculation of an exchange integral $(rk|sj)$, where $k \leftrightarrow i$ holds and i, j are symmetry-unique orbitals. The necessary three-index integral $(B|rk)$ in Eq. (21) can then be generated according to

$$(B|rk) = \sum_{A \in [i\tilde{j}]_{\mathrm{fit}},\; t \in [i\tilde{j}]_{\mathrm{PAO}}} O^{kj,i\tilde{j}}_{BA} O^{kj,i\tilde{j}}_{rt} C^k (A|ti), \tag{27}$$

where $C^k = \pm 1$ represents an orbital sign factor and $[i\tilde{j}]_{\mathrm{fit}}$ and $[i\tilde{j}]_{\mathrm{PAO}}$ denote the symmetry-equivalent united fitting and united PAO domains of pair kj. These result by applying the symmetry operation which connects orbitals k and i to the corresponding domains. Note that the united PAO and united fitting domains for the calculation of the symmetry-unique three-index integrals $(A|ti)$ therefore need to be extended such

TABLE I. Formulae for the generation of symmetry-equivalent amplitude matrices $\mathbf{T}^{kl}$ from a symmetry-unique set of generator pair amplitudes $\mathbf{T}^{ij}$ for $i \geq j$ and $k \geq l$. $C^{kl} = \pm 1$ are sign factors. If the symmetry operation applied to ij yields kl, the operator $\hat{P}^{kl}$ has no effect. If, however, $ij \to lk$, then P^{kl} transposes the matrix on which it acts.

Method	Pairs	Formula	
PAO-LMP2	All	$T^{kl}_{rs} = C^{kl}\hat{P}^{kl}\sum_{tu} O^{kl,ij}_{rt} T^{ij}_{tu} O^{kl,ij}_{su}$	$\forall\, ij \leftrightarrow kl \wedge ij \in \overline{P}$
OSV-LMP2 or PNO-LMP2	ii	$T^{kk}_{pq} = T^{ii}_{pq}$	$\forall\, ii \leftrightarrow kk \wedge ii \in \overline{P}$
	ij	$T^{kl}_{pq} = C^{kl}\hat{P}^{kl} T^{ij}_{pq}$	$\forall\, ij \leftrightarrow kl \wedge ij \in \overline{P}$

that also every $\tilde{j}$ that forms a non-distant pair with i is included in the united domains for orbital i. Detailed benchmarks which reveal performance benefits for all necessary integral evaluation steps will be presented in Sec. III.

Furthermore, symmetry can be used during the multipole approximation used for distant pairs, cf. Eq. (18). In this case, the one-electron dipole integrals $\langle i|\mathbf{r}|p^{(i)}\rangle$ need only to be computed for a unique set of LMOs $|i\rangle$, and only the pair energies for symmetry-unique pairs need to be assembled. Since the time for computing the distant pair energies is very small, the CPU time savings in this step are negligible, but nevertheless, symmetry is used for consistency reasons.

Finally, a smaller number of overlap matrices are needed in PNO-LMP2 when symmetry is applied, cf. Eq. (16). In our program, we compute all $\mathbf{S}^{ij,kj}_{\mathrm{PNO}}$ and $\mathbf{S}^{ij,ik}_{\mathrm{PNO}}$ for the unique set of pairs ij.

The overlap matrices are computed as

$$\mathbf{S}_{\mathrm{PAO}} = \mathbf{P}^{\dagger}\mathbf{S}\mathbf{P}, \tag{28}$$

$$\mathbf{S}^{ij,kl}_{\mathrm{OSV}} = \mathbf{Q}^{(ij)\dagger}\mathbf{S}^{ij,kl}_{\mathrm{PAO}}\mathbf{Q}^{(kl)}, \tag{29}$$

$$\mathbf{S}^{ij,kl}_{\mathrm{PNO}} = \mathbf{W}^{(ij)\dagger}\mathbf{S}^{ij,kl}_{\mathrm{OSV}}\mathbf{W}^{(kl)}, \tag{30}$$

where ij (rows) and kl (columns) refer to the corresponding domain blocks of the overlap matrices in the PAO, OSV, or PNO basis. In order to compute all required blocks, all transformation matrices $\mathbf{Q}^{(kl)}$ and $\mathbf{W}^{(kl)}$ are needed. The former can be assembled on the fly from the matrices $\mathbf{Q}^{(k)}$ and $\mathbf{Q}^{(l)}$. Only the symmetry-unique matrices $\mathbf{Q}^{(k)}$ and $\mathbf{W}^{(kl)}$ need to be computed, the symmetry equivalent ones can then be generated using the appropriate symmetry transformation matrices.

7. Multicore implementation

The PNO-LMP2 program is fully parallelized and all large data sets are distributed over the processors (or processing cores).[17] In most places, parallelization is either over orbitals (e.g., in the OSV generation) or over pairs (in the generation of the OSV and PNO pair domains, the assembly step of the integrals, and in the iterations).

In order to distribute the workload for all available CPU cores in an (almost) equivalent fashion, only the symmetry-unique set of orbitals or orbital pairs is distributed over the CPU cores. All quantities which are symmetry-equivalent are always handled on the same CPU core, since this avoids communication overhead and is most efficient. The METIS graph partitioning technique[52] is used to optimize the load-balancing.

III. BENCHMARK RESULTS

The method described in Secs. II B and II C has been implemented within the MOLPRO package of *ab initio* programs.[53,54] In the following, we will present various benchmarks which demonstrate the performance benefits accessible through our approach to symmetry. Apart from total elapsed times, we will present detailed results for (1) the evaluation of the integrals $K^{ij}_{rs} = (ri|sj)$ and (2) the iterative solution of the residual equations, which are the most time consuming steps for the considered PAO-, OSV-, and PNO-LMP2 methods.

Before we start presenting pure numerical performance data for various example molecules, it is worthwhile to analyze the conditions under which our symmetry treatment should be most effective: In contrast to conventional expectations, a specific point group of the nucleic framework is not necessarily a well-qualified measure for the expected speedup. Instead, the ratio of symmetry-equivalent to symmetry-adapted orbitals should be as large as possible, since this allows the number of symmetry-unique orbital pairs to become sufficiently small and high performance benefits should result. This condition is best met whenever no atoms or bonds are in the centre of symmetry and many orbitals are located at (isolated) symmetry-equivalent nuclei.

We carried out benchmark calculations for 16 molecules using the correlation-consistent basis sets of Dunning and co-workers.[51,55] The benchmark includes 8 medium sized molecules, which were run on a single CPU core and 8 larger molecules, which were computed using 20 CPU cores in parallel. Ball-and-stick drawings for the 8 large molecules are shown in Table II. The optimized structures can be found in the supplementary material.[49]

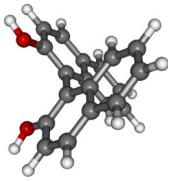
A summary of computational details for each molecule (symmetry, basis set, number of basis functions, number of pairs treated explicitly in the LMP2) as well as timing data on the single or multiple processing cores is presented in Table III. If available, the additional number of orbital pairs, which is approximated using the dipole-dipole approximation, is shown in parentheses.

For each of the molecules, two rows are given: the first contains the data and timings without symmetry and the second with symmetry used. The elapsed times for the integral evaluation and transformation are nearly the same for the PAO-LMP2, OSV-LMP2, and PNO-LMP2 methods, since in all cases, the integrals are first evaluated in the PAO basis and only finally transformed into the OSV basis. Therefore, these times are shown only once. On the other hand, due to the different pair domain sizes (cf. Sec. II A 1), the iteration times differ strongly for the three methods, and therefore, the iteration and total times are shown for each method separately. The total timings do not include the preceding density fitted Hartree-Fock (DF-HF) calculation, the localization of the orbitals, and the setup of the symmetry relations, since these parts are not yet fully optimized and parallelized. In particular, the HF step is usually much more expensive than the LMP2 one.[17] The implementation of a new improved local DF-HF program that can also exploit molecular symmetry is in progress and will be described in a future paper. Detailed absolute energies of the HF and LMP2 correlation energies for the investigated molecules can be found in Table I within the supplementary material.[49] For all molecules, the differences of the total energies with and without symmetry were smaller than 10^{-8} E_h (i.e., smaller than the convergence energy threshold in the LMP2 iterations).

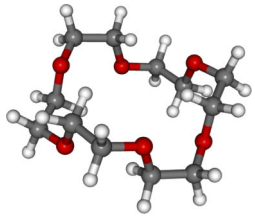
A. Integral evaluation

The performance benefits of our symmetry treatment within the density-fitted integral evaluation using local density

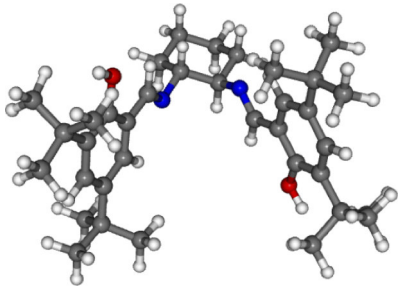
TABLE II. Molecular structures of the 8 largest molecules listed in Table III. The largest molecule (elaiophylin) consists of more than 150 atoms and extends over more than 25 Å. In addition to the actual molecular point group which is given in parentheses, the Abelian point group used within our (Abelian) symmetry treatment is given.



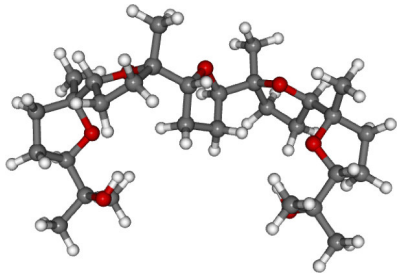
$C_{20}H_{14}O_2$, (*S*)− BINOL, C_2



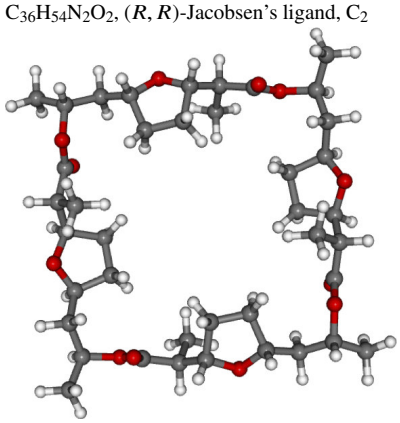
$C_{12}H_{24}O_6$, 18-Crown-6, (S_6) C_i



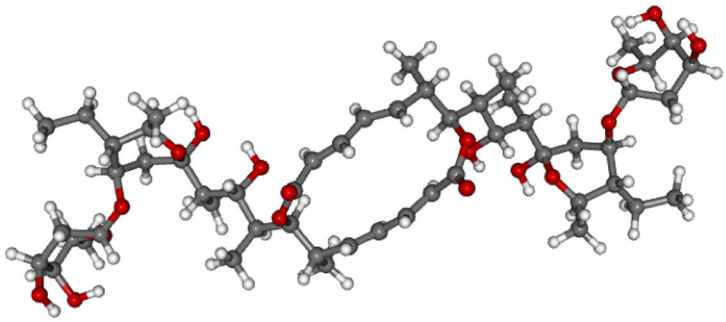
$C_{36}H_{54}N_2O_2$, (*R*, *R*)-Jacobsen's ligand, C_2



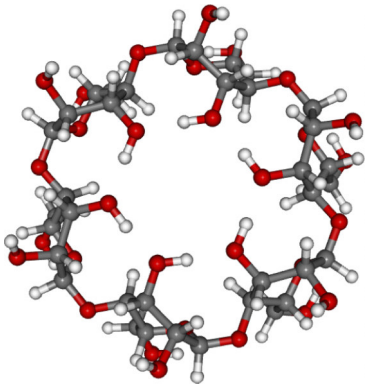
$C_{30}H_{52}O_7$, (−)-Glabrescol, C_2



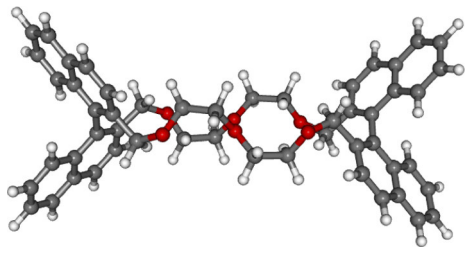
$C_{40}H_{64}O_{12}$, Nonactin, (S_4) C_2



$C_{54}H_{88}O_{18}$, Elaiophylin, C_2



$C_{36}H_{60}O_{30}$, α-cyclodextrin, (C_6) C_2



$C_{52}O_6H_{48}$, D_2

fitting (LDF) and LMO/PAO sparsity[17] are exemplified in Figure 2 for the $C_{52}O_6H_{48}$ molecule (D_2 symmetry), using the cc-pVTZ basis set. In this calculation, 20 CPU cores in a single node were used.

The total elapsed times for the integral evaluation are reported on top of the respective bars. The two smaller bars on the left represent the timings to evaluate the integrals $(ri|si)$ in the PAO basis for diagonal pairs. These integrals are computed initially in order to generate the OSVs. The two larger bars on the right visualize the timings necessary to evaluate all integrals $(pi|qj)$, including their transformation from the PAO to the OSV basis within the assembly step.

It can clearly be seen that the most time consuming evaluation step within the integral evaluation for diagonal pairs is the calculation of three-index integrals in the AO/fitting basis (brown). This is even more so the case when symmetry is used, because all other evaluation steps are efficiently accelerated if merely symmetry-unique orbitals need to be considered.

TABLE III. Overview of the total elapsed times to calculate the correlation energy using PAO-, OSV-, and PNO-LMP2 programs with and without symmetry. For each of the molecules, data obtained without the use of symmetry are shown first, while the subsequent row shows data obtained with symmetry usage. In addition to the total elapsed times to calculate the correlation energy (not including HF, orbital localization, and setup of symmetry-relations), the elapsed times to evaluate the exchange integrals (using LDF with fitting domains extended by two neighbouring shells of atoms compared to the primary PAO orbital domains and using LMO/PAO sparsity with thresholds of 10^{-5} and 10^{-6}, respectively) as well as the elapsed times spent to iteratively solve the LMP2 equations are also included. Distant pairs, i.e., pairs with energy contributions below 1×10^{-6} Hartree, are treated by multipole approximation (reported in parentheses). Moreover, extended PAO domains are used for realistic relative performance differences between the various LMP2 methods (`iext = 1`). The calculations were carried out on a single compute node with 20 CPU cores in two Intel® Xeon® E5-2690, 2.8 GHz processors, and 256 GB memory.

				Elapsed time (s)						
		Pairs		PAO-LMP2			OSV-LMP2		PNO-LMP2	
Molecule (basis), point group	AOs	Optimized	Distant	Integrals (PAO)	Iterations	Total	Iterations	Total	Iterations	Total
				Using 1 CPU core						
Ethylbenzene (cc-pVTZ), C_s	380	231		15.8	129.1	151.8	82.3	102.8	3.1	27.6
		127		11.1	70.6	88.1	46.6	60.6	1.7	18.0
2,3-dichlorobutane (cc-pVQZ), C_i	578	190		37.7	322.2	376.0	120.7	166.1	7.0	58.4
		100		27.1	171.7	214.7	64.3	96.1	3.7	38.8
Trans-2-butene (aug-cc-pVTZ), C_{2h}	368	78		7.5	35.2	46.4	18.8	28.7	0.9	12.4
		33		5.6	14.9	24.1	8.4	15.2	0.4	7.8
1,4-benzochinone (cc-pVTZ), D_{2h}	296	207	(3)	8.9	66.5	79.4	43.6	55.3	2.7	16.8
		68	(2)	4.8	22.1	30.7	14.7	21.0	0.9	7.7
Cyclohexane (aug-cc-pVQZ), C_{2h}	1032	171		154.9	1524.1	1751.3	274.4	445.4	8.5	190.2
		52		96.3	465.7	633.4	88.8	191.2	2.8	110.1
Adamantane (aug-cc-pVTZ), D_2	828	406		140.1	1631.2	1823.3	356.5	509.5	13.0	179.4
		112		71.6	458.5	580.8	101.7	177.5	3.5	84.1
Twistane (aug-cc-pVTZ), D_2	828	406		138.0	1625.8	1815.0	363.0	515.2	14.0	179.7
		113		72.0	456.6	578.7	107.5	184.4	3.9	85.3
$(C_2F_4)_2$ (aug-cc-pVTZ), D_{2h}	552	650	(16)	46.7	996.1	1067.0	237.1	289.9	9.6	69.4
		99	(3)	18.7	168.0	210.6	40.3	60.4	1.5	23.4
				Using 20 CPU cores						
(*S*)-BINOL (aug-cc-pVTZ), C_2	1334	1297	(134)	39.6	909.6	960.6	205.7	250.4	5.7	57.5
		658	(71)	26.2	490.4	528.3	113.3	143.8	3.7	38.4
18-Crown-6 (aug-cc-pVTZ), C_i	1380	1422	(63)	28.4	595.8	631.8	113.0	144.3	2.3	37.0
		720	(36)	23.5	314.7	346.3	62.3	87.7	1.2	28.5
(*R*, *R*)-Jacobsen's ligand (cc-pVTZ), C_2	1956	3181	(2 924)	51.1	719.7	778.7	262.0	318.4	7.1	70.8
		1601	(1 480)	32.4	393.5	434.0	139.4	175.8	3.8	44.6
(−)-Glabrescol (cc-pVTZ), C_2	1838	3454	(2 324)	50.9	832.7	892.0	275.5	331.8	6.1	69.2
		1737	(1 181)	32.3	421.9	462.5	142.6	178.7	3.4	43.9
Nonactin (cc-pVTZ), C_2	2456	4534	(6 270)	63.7	837.3	911.3	282.5	352.0	7.3	84.5
		2268	(3 166)	41.2	426.9	477.6	155.2	201.8	3.6	55.0
Elaiophylin (cc-pVTZ), C_2	3392	7079	(7 868)	113.5	1501.6	1630.7	525.1	647.7	14.3	151.2
		3547	(3 951)	71.7	773.1	859.9	301.3	381.1	7.4	95.4
α-cyclodextrin (aug-cc-pVDZ), C_2	2058	7659	(10 833)	129.5	766.2	903.8	200.4	335.2	7.6	148.4
		3834	(5 460)	87.0	394.5	489.3	111.1	202.2	4.1	100.1
$C_{52}O_6H_{48}$ (cc-pVTZ), D_2	2412	4856	(4 495)	81.4	1479.6	1577.4	523.9	613.8	13.2	117.5
		1237	(1 143)	35.7	390.8	440.4	134.9	176.9	4.0	50.8

When all integrals $(pi|qj)$ are evaluated, the transformation of the three-index integrals $(\mu\nu|A) \rightarrow (ri|A)$ from the AO into the LMO/PAO basis and the solution of Eq. (21) becomes the most time consuming steps. Using symmetry, however, the first and second integral transformations are accelerated significantly, because only a symmetry-unique set of $(A|ri)$ integrals needs to be calculated.

The remaining fitting and assembly steps are also strongly accelerated by symmetry, since these are only done for the symmetry unique orbital pairs *ij*. For these steps, symmetry-equivalent three-index integrals are generated on the fly from the symmetry-unique ones whenever needed. Equation (21) is then solved and the integrals assembled according to Eq. (20).

B. LMP2 iterations and total timings

The LMP2 iteration times differ enormously for the PAO-LMP2, OSV-LMP2, and PNO-LMP2 methods. For PAO-LMP2, the iterative solution of the LMP2 equations is by far most time consuming and accounts for a large percentage of the total elapsed times. Since the iterations can efficiently be accelerated by evaluating only a reduced set of symmetry-unique pair residuals, the accessible speedup is limited by the ratio between the number of all pairs and the number of symmetry-unique pairs. In practice, this limit is not completely reached, because all symmetry-equivalent amplitudes still need to be generated from the symmetry-

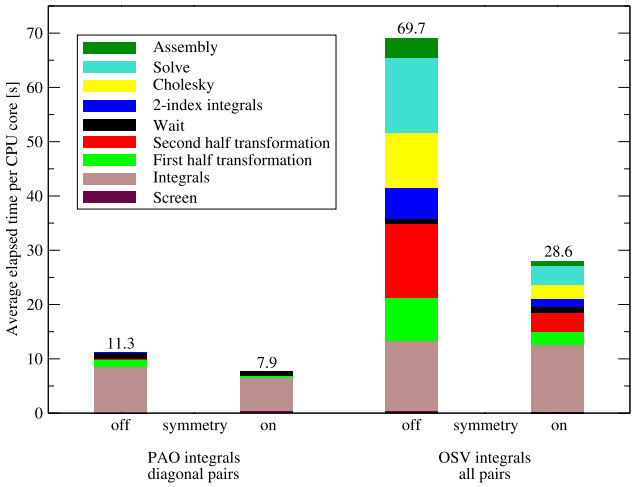


FIG. 2. Performance of the density fitted exchange integral evaluation using $C_{52}O_6H_{48}$, the cc-pVTZ basis set,[51,55] a single compute node with a total of 20 CPU cores in two Intel Xeon E5-2690, 2.8 GHz processors, and 256 GB of memory. Averaged elapsed times (in s) for integrals $(ri|si)$ (PAO basis, diagonal pairs ii) needed to generate the transformation matrices from the PAO to the OSV basis are shown on the left, whereas timings for the generation of the full set of integrals $(pi|qj)$ (OSV basis, all non-distant pairs ij) are shown on the right. In both cases, LDF sparsity and LMO/PAO sparsity are used. The total elapsed times (in s) are presented as numbers on top of the bars.

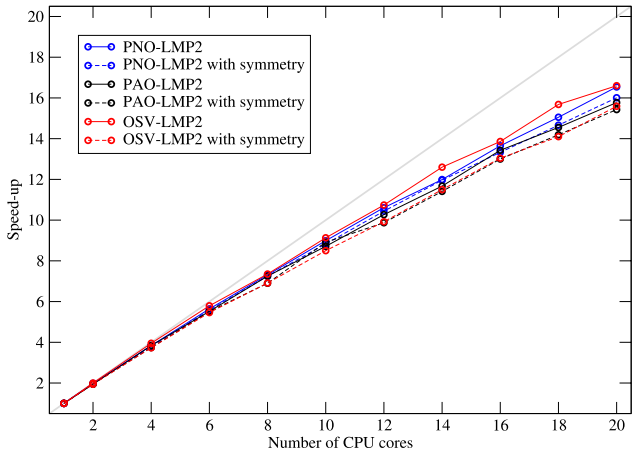


FIG. 3. Parallel efficiency based on total elapsed times for a LMP2 calculation using nonactin and the cc-pVTZ basis set. A single compute node with a total of 20 CPU cores in two Intel Xeon E5-2690, 2.8 GHz processors.

unique ones in each iteration. Nevertheless, significant savings in the elapsed times can be observed for PAO-LMP2.

The elapsed times needed to calculate the integrals and to solve the LMP2 equations become more comparable for OSV-LMP2, because the OSV domain sizes in OSV-LMP2 are much smaller than the PAO ones. However, since the integral evaluation does generally not profit as much from the use of symmetry as the iterations (cf. Sec. III A), the total speedups are smaller than for the corresponding PAO-LMP2 cases. Nevertheless, the OSV-LMP2 calculations are typically a factor of 2-3 faster than PAO-LMP2 ones (using the same initial PAO domains) and a relatively tight OSV threshold $T_{\mathrm{OSV}} = 10^{-9}$. This ratio would even be much larger if the PAO domains were further extended. In our program, such extensions are determined by a parameter `IEXT`. `IEXT` = 1, as used in the present calculations, which means that the *primary* PAO domains, which are determined on the basis of IBO partial charges (threshold $T_{\mathrm{PAO}} = 0.2$), are extended by all PAOs at the next shell of neighboring atoms (for details see Ref. 17). In our PNO-LMP2 program, the default is `IEXT` = 2, which means that the PAO domains are even much further extended. Since this would lead to extremely large PAO-LMP2 iteration times, this default has not been used in the present benchmark calculations.

In contrast to PAO-LMP2, the most time consuming evaluation step in PNO-LMP2 now becomes the evaluation of the integrals, whereas the iterative solution of the LMP2 equations is very fast—a result of the small PNO domain sizes. In the current work, their size is determined via an energy threshold of $T_{\mathrm{PNO}} = 0.997$. This means that the PNO domains are chosen to recover at least 99.7% of the semi-canonical OSV-LMP2 energy.[17] Although the absolute speedups for the total LMP2 elapsed times (which include the PNO generation, the computation of the overlap matrices, and all other intermediates) are now rather small, it is to be expected that the gains in a subsequent PNO-LCCSD calculation would be significant, both for the CPU times and the memory requirements. The implementation of a new PNO-LCCSD program is in progress.

C. Parallel scalability

Figure 3 illustrates the parallel scalability of our LMP2 implementations with and without the use of symmetry. As can be seen, all presented methods, i.e., PAO-, OSV-, and PNO-LMP2, virtually show the same speedups with the number of cores, regardless of whether symmetry is used or not. Most likely, the deviations from the ideal speedup are partly due to non-ideal load balancing and partly to the limited memory bandwidth. Nevertheless, a speedup of about 16 using 20 cores is quite satisfactory.

IV. CONCLUSION

A disadvantage of local correlation methods is that the usual symmetry treatment based on symmetry-adapted orbitals is not possible. However, we have shown in this paper that molecular symmetry can effectively be used also in local methods by exploiting that in symmetric molecules, groups of orbitals become either symmetry equivalent or symmetry adapted. This reduces the number of correlated orbital pairs that must be explicitly be treated and correspondingly, the number of integrals, amplitudes, and amplitude equations to be solved. This use of symmetry is most beneficial when the ratio of symmetry-equivalent to symmetry-adapted orbitals becomes large. Many larger molecules belong to the C_2 point group, and then speedups of nearly a factor of two can be achieved. Larger speedups are possible for molecules that possess higher symmetry. The speedups are most significant for the PAO-LMP2, since the PAO domains are large and solving the amplitude equations takes most of the time. In contrast, the time savings are much smaller for the very much faster PNO-LMP2 method.

Nevertheless, we believe that this type of symmetry treatment will be particularly useful for the PNO-LCCSD methods currently under development in our laboratory. Perhaps most importantly, the memory and storage requirements can be significantly reduced, which is important for a scalable parallel implementation.[17]

At the same time, efficiency improvements in LMP2 or LCCSD naturally generate another bottleneck: the preceding Hartree-Fock calculation. However, as shown in previous work,[56,57] local orbitals can also be used to overcome this problem, and then a similar symmetry treatment as described here should be possible. The development of a new local density-fitted Hartree-Fock program is therefore a high priority in our group.

ACKNOWLEDGMENTS

This work has been funded by the ERC Advanced Grant No. 320723 (ASES).

[1]C. Møller and M. S. Plesset, Phys. Rev. **46**, 618 (1934).
[2]E. A. Hylleraas, Z. Phys. **65**, 209 (1930).
[3]P. Pulay, Chem. Phys. Lett. **100**, 151 (1983).
[4]S. Saebø and P. Pulay, Chem. Phys. Lett. **113**, 13 (1985).
[5]P. Pulay and S. Saebø, Theor. Chim. Acta **69**, 357 (1986).
[6]S. Saebø and P. Pulay, J. Chem. Phys. **86**, 914 (1987).
[7]R. B. Murphy, M. D. Beachy, R. A. Friesner, and M. N. Ringnalda, J. Chem. Phys. **103**, 1481 (1995).
[8]R. B. Murphy, W. T. Pollard, and R. A. Friesner, J. Chem. Phys. **106**, 5073 (1997).
[9]G. Hetzer, P. Pulay, and H.-J. Werner, Chem. Phys. Lett. **290**, 143 (1998).
[10]M. Schütz, G. Hetzer, and H.-J. Werner, J. Chem. Phys. **111**, 5691 (1999).
[11]P. Y. Ayala and G. E. Scuseria, J. Chem. Phys. **110**, 3660 (1999).
[12]R. A. Friesner, R. B. Murphy, M. D. Beachy, M. N. Ringnalda, W. T. Pollard, B. D. Dunietz, and Y. Cao, J. Phys. Chem. A **103**, 1913 (1999).
[13]G. Hetzer, M. Schütz, H. Stoll, and H.-J. Werner, J. Chem. Phys. **113**, 9443 (2000).
[14]M. S. Lee, P. E. Maslen, and M. Head-Gordon, J. Chem. Phys. **112**, 3592 (2000).
[15]S. Saebø and P. Pulay, J. Chem. Phys. **115**, 3975 (2001).
[16]H.-J. Werner, F. R. Manby, and P. J. Knowles, J. Chem. Phys. **118**, 8149 (2003).
[17]H.-J. Werner, G. Knizia, C. Krause, M. Schwilk, and M. Dornbach, J. Chem. Theory Comput. **11**, 484 (2015).
[18]C. Hampel and H.-J. Werner, J. Chem. Phys. **104**, 6286 (1996).
[19]M. Schütz and H.-J. Werner, J. Chem. Phys. **114**, 661 (2001).
[20]H.-J. Werner and M. Schütz, J. Chem. Phys. **135**, 144116 (2011).
[21]C. Riplinger and F. Neese, J. Chem. Phys. **138**, 034106 (2013).
[22]M. Schütz and H.-J. Werner, Chem. Phys. Lett. **318**, 370 (2000).
[23]M. Schütz, J. Chem. Phys. **113**, 9986 (2000).
[24]M. Schütz, J. Chem. Phys. **116**, 8772 (2002).
[25]C. Riplinger, B. Sandhoefer, A. Hansen, and F. Neese, J. Chem. Phys. **139**, 134101 (2013).
[26]M. Dupuis and H. F. King, Int. J. Quantum Chem. **11**, 613 (1977).
[27]T. Voloshchuk, N. S. Farina, O. R. Wauchope, M. Kiprowska, P. Haberfield, and A. Greer, J. Nat. Prod. **67**, 1141 (2004).
[28]R. Ahlrichs, S. Elliot, and U. Huniar, *Modern Methods and Algorithms of Quantum Chemistry—Proceedings*, NIC Series Vol. 3, 2nd ed., edited by J. Grotendorst (John von Neumann Institute for Computing, Jülich, 2000), pp. 7-25, ISBN: 3-00-005834-6.
[29]J. Yang, Y. Kurashige, F. R. Manby, and G. K. L. Chan, J. Chem. Phys. **134**, 044123 (2011).
[30]J. Yang, G. K.-L. Chan, F. R. Manby, M. Schütz, and H.-J. Werner, J. Chem. Phys. **136**, 144105 (2012).
[31]C. Edmiston and M. Krauss, J. Chem. Phys. **42**, 1119 (1965).
[32]W. Meyer, Int. J. Quantum Chem. **5**, 341 (1971).
[33]W. Meyer, J. Chem. Phys. **58**, 1017 (1973).
[34]R. Ahlrichs, F. Driessler, H. Lischka, V. Staemmler, and W. Kutzelnigg, J. Chem. Phys. **62**, 1235 (1975).
[35]V. Staemmler and R. Jaquet, Theor. Chim. Acta **59**, 487 (1981).
[36]F. Neese, F. Wennmohs, and A. Hansen, J. Chem. Phys. **130**, 114108 (2009).
[37]F. Neese, A. Hansen, and D. G. Liakos, J. Chem. Phys. **131**, 064103 (2009).
[38]A. Hansen, D. G. Liakos, and F. Neese, J. Chem. Phys. **135**, 214102 (2011).
[39]M. Schwilk, D. Usvyat, and H.-J. Werner, J. Chem. Phys. **142**, 121102 (2015).
[40]S. F. Boys, Rev. Mod. Phys. **32**, 296 (1960).
[41]J. M. Foster and S. F. Boys, Rev. Mod. Phys. **32**, 300 (1960).
[42]C. Edmiston and K. Ruedenberg, Rev. Mod. Phys. **35**, 457 (1963).
[43]J. Pipek and P. G. Mezey, J. Chem. Phys. **90**, 4916 (1989).
[44]J. W. Boughton and P. Pulay, J. Comput. Chem. **14**, 736 (1993).
[45]J. Pipek, Int. J. Quantum Chem. **36**, 487 (1989).
[46]R. Mata and H.-J. Werner, Mol. Phys. **105**, 2753 (2007).
[47]G. Knizia, J. Chem. Theory Comput. **9**, 4834 (2013).
[48]C. Krause and H.-J. Werner, Phys. Chem. Chem. Phys. **14**, 7591 (2012).
[49]See supplementary material at http://dx.doi.org/10.1063/1.4918772 for further information and optimized geometries.
[50]U. Varetto, Molekel 5.4.0.8, Swiss National Supercomputing Centre, Lugano, Switzerland.
[51]T. H. Dunning, Jr., J. Chem. Phys. **90**, 1007 (1989).
[52]G. Karypis and V. Kumar, SIAM J. Sci. Comput. **20**, 359 (1998).
[53]H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, and M. Schütz, Wiley Interdiscip. Rev.: Comput. Mol. Sci. **2**, 242 (2012).
[54]H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, M. Schütz *et al.*, MOLPRO, version 2014.1, a package of *ab initio* programs, 2014, see http://www.molpro.net.
[55]R. A. Kendall, T. H. Dunning, Jr., and R. J. Harrison, J. Chem. Phys. **96**, 6796 (1992).
[56]R. Polly, H.-J. Werner, F. R. Manby, and P. J. Knowles, Mol. Phys. **102**, 2311 (2004).
[57]D. Mejía-Rodríguez and A. M. Köster, J. Chem. Phys. **141**, 124114 (2014).