Parallel and low-order scaling implementation of Hartree-Fock exchange using local density fitting

Christoph Köppl and Hans-Joachim Werner*

Institut für Theoretische Chemie, Universität Stuttgart, Pfaffenwaldring 55, 70569 Stuttgart, Germany.

E-mail: werner@theochem.uni-stuttgart.de

Abstract

Modern linear scaling electron correlation methods are often much faster than the necessary reference Hartree-Fock calculations. We report a newly implemented Hartree-Fock (HF) program that speeds up the most time consuming step, namely the evaluation of the exchange contributions to the Fock matrix. Using localized orbitals and their sparsity, local density fitting, and atomic orbital domains, we demonstrate that the calculation of the exchange matrix scales asymptotically linearly with molecular size. The remaining parts of the HF calculation scale cubically, but become dominant only for very large molecular sizes or with many processing cores. The method is well parallelized and the speedup scales well with up to about 100 CPU cores on multiple compute nodes. The effect of the local approximations on the accuracy of computed HF and local second-order Møller-Plesset perturbation theory (LMP2) energies is systematically investigated and default values are established for the parameters that determine the domain sizes. Using these values, calculations for molecules with

hundreds of atoms in combination with triple-ζ basis sets can be carried out in less than one hour, using just a few compute nodes. The method can also be used to speed up density functional (DFT) calculations with hybrid functionals that contain Hartree-Fock exchange.

1 Introduction

In the last decade the efficient determination of accurate HF orbitals and energies for large molecules has gained increasing interest, since fast linear scaling electron correlation methods[1–12] now often take less time than the determination of the HF reference function. This is particularly true for recent methods based on pair natural orbitals (PNOs),[13–21] cf. Figure 1.

In the past, a lot of progress has been made in the development of linear scaling HF programs — most notably being contributions of Ochsenfeld *et al.*[22–24], Schwegler *et al.*[25–29], Gan *et al.*[30,31], Rudberg *et al.*[32], Scuseria *et al.*[33] and He *et al.*[34]. While all of these approaches have their well justified and intrinsic amenities, it turned out that overall linear scaling with the molecular size is achieved only for quite large molecular sizes. The advantage of conventional integral-direct methods is that the exact HF result for a given basis set can be approached to any desired accuracy, depending on screening thresholds. But a severe problem is that the computational effort scales with the fourth power of the number of basis functions per atom. The computation time and the onset of linear scaling are therefore strongly dependent on the size of the basis set, and for molecules containing a few hundred atoms such calculations are still much more expensive than for example subsequent PNO-LMP2 calculations.[18,19,21] Many other approximate and low-order scaling approaches to calculate 2-electron integrals and/or the Fock matrix exist.[35–45] In contrast to standard integral-direct approaches some of these introduce additional approximations which slightly affect the energy and the optimized orbitals. For example, this is the case for the density

fitting (DF) approximation used in the current work. Since the correlation energy depends quite sensitively on the accuracy of the HF orbitals, and it is important to investigate how approximations in the HF calculation affect the overall accuracy of the results.

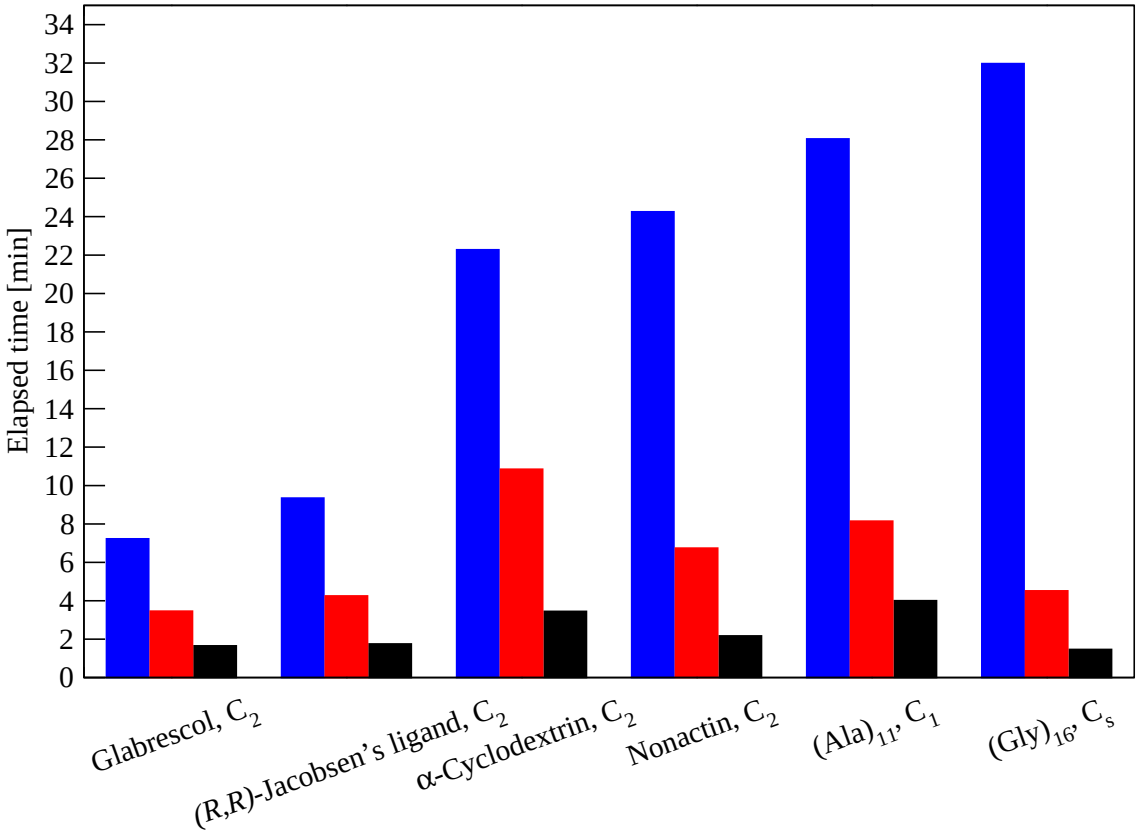
Our aim is to develop a fast Hartree-Fock program that can be applied to molecules with a few hundred atoms and large basis sets (up to about 10000 basis functions). This certainly covers most chemically interesting applications of DFT and accurate electron correlation methods. Low-order scaling is desirable for the expensive evaluation of the exchange part of the Fock matrix, while for these target molecular sizes other parts of the calculation, such as calculation of the Coulomb matrix and transformation and diagonalization of the Fock matrix, are still so fast that the conventional approaches, which scale cubically with the number of basis functions and molecular size, are sufficient.

Modern local electron correlation methods use local density fitting[46] (LDF) techniques to evaluate and transform the two-electron integrals. Density fitting (DF) is also well established in HF and DFT methods[47–55] and typically speeds up the integral calculations by at least one order of magnitude, in particular for larger basis sets. One of the important advantages of DF is that the computational effort scales only cubically with the number of basis functions per atom. The errors introduced by density fitting are usually negligible, in particular for relative energies.

In the current paper we aim at speeding up DF-HF calculations using improved and well parallelized LDF techniques, which we recently developed and applied in explicitly correlated local MP2 and coupled-cluster methods.[18–21] A DF-HF method based on LDF approximations was already developed in our group a decade ago,[56] and recently similar ideas have also been used by Mejía-Rodríguez *et al.*.[57] However, the latter method was not parallelized, and the effect of the approximations on subsequent correlation calculations was not investigated. Our previous program[56] was moderately parallelized, but for large molecules heavily I/O bound, since the half-transformed 3-index integrals were stored on disk. Furthermore, it

turned out that the correlation energies obtained in subsequent LMP2 or LCCSD calculations were quite sensitive to the accuracy of the optimized HF orbitals. The method yielded HF energies which differed from those obtained without local approximations in the exchange part by only a few μE_h, but the errors in correlation energies were sometimes significantly larger.

Figure 1: Total elapsed times of conventional DF-HF calculations without local approximations (blue), the new LDF-HF method (red), and PNO-LMP2 (black) calculations for the molecules shown in Table 2 as well as an alanine helix and a glycine chain using 20 CPU cores in one compute node with two Intel® Xeon® E5-2690, 2.8 GHz processors. The aug-cc-pVDZ basis set was used for the α-Cyclodextrin calculation. All other calculations were performed using the cc-pVTZ basis set. The indicated Abelian point group symmetries were used in the DF-HF and LDF-HF calculations as described in the text, while the PNO-LMP2 calculations used symmetry as described elsewhere.[19] The orbitals and the Fock matrix used in the PNO-LMP2 were transformed from the symmetry adapted basis to the unsymmetrized one.



In our new LDF-HF program we make use of localized intrinsic bond orbitals[58] and their approximate sparsity, local fitting domains, as well as atomic orbital (AO) domains to obtain

near linear scaling in the evaluation of the exchange matrix. Furthermore, Abelian point group symmetry can be used to reduce the computational resources. In order to achieve an efficient parallelization on systems with several compute nodes, all intermediate quantities are stored in distributed memory using the Global Arrays (GA) toolkit,[59,60] and thus disk I/O is almost entirely avoided.

In Section 2 we will introduce the local approximations and describe the technical details of our parallel implementation. We will restrict the discussion to closed-shell systems, but it is straightforward to generalize the method to spin-restricted open-shell Hartree-Fock (ROHF) or spin-unrestricted HF (UHF). In the ROHF case the closed-shell and open-shell orbitals have then to be localized separately, while in the UHF case the α-spin and β-spin orbitals are localized independently. In Section 3 we will present benchmarks for the accuracy and efficiency of the method, and establish robust default values for the various parameters that determine the local approximations.

2 Theory

2.1 Exchange integral calculation using local density fitting

In a conventional (closed-shell) HF program the most expensive contribution to the closed-shell Fock matrix

$$F_{\mu\nu} = h_{\mu\nu} + 2J_{\mu\nu} - K_{\mu\nu} \tag{1}$$

is usually the calculation of the exchange matrix $K_{\mu\nu}$ in the AO basis according to

$$K_{\mu\nu} = \sum_{\rho\sigma} D_{\rho\sigma} \int \chi_\mu(\mathbf{r}_1)\chi_\rho(\mathbf{r}_1)\frac{1}{r_{12}}\chi_\nu(\mathbf{r}_2)\chi_\sigma(\mathbf{r}_2)d\mathbf{r}_1 d\mathbf{r}_2, \tag{2}$$

where χ_μ are basis functions and

$$D_{\rho\sigma} = 2\sum_{i}^{\text{occ}} C_{\rho i}C_{\sigma i} \tag{3}$$

is the density matrix, expressed by the MO-coefficients $C_{\mu i}$. Here the index i refers to occupied molecular orbitals (MOs) $\phi_i(\mathbf{r}) = \sum_\mu \chi_\mu(\mathbf{r})C_{\mu i}$. Using the sparsity of the density matrix and screening techniques linear scaling can be achieved. Nevertheless, for large molecules and basis sets the computation of the two-electron integrals in each iteration is very expensive.

In the LDF approximation the exchange matrix is evaluated as a sum of contributions of individual MOs, without using the density matrix. Instead of the sparsity of the density matrix the sparsity of the MO coefficients in a localized orbital (LMO) basis can be exploited. In the current work we use IBOs.[58] While other localization procedures[61–66] would in principle also be possible, we use IBOs since this type of orbital localization is fast and very robust. In particular, the IBOs are insensitive to the size of the basis set and the presence of diffuse basis functions. Furthermore, the IBO partial atomic charges that are nearly independent of the basis set and can therefore be conveniently used to define orbital domains. A disadvantage of IBOs is that for some highly symmetric aromatic systems such as benzene or polycyclic aromatic hydrocarbons the LMOs are neither symmetry equivalent nor symmetry adapted. When domain approximations are introduced, this can lead to artifacts such as small dipole moments perpendicular to a symmetry plane. The same is true for Pipek-Mezey localization, but avoided with Boys localization. The disadvantage of the latter is that it breaks $\sigma - \pi$ symmetry and yields "banana" bonds. We are currently working on a systematic study about the efficiency and accuracy of various localization schemes in local correlation methods, which

will be published elsewhere.

All integrals are computed using the adaptive integral core (AIC) interface written by G. Knizia.[67,68] The basic quantities needed are two and 3-index integrals

$$J_{AB} = \int \chi_A(\mathbf{r}_1) \frac{1}{r_{12}} \chi_B(\mathbf{r}_2) d\mathbf{r}_1 d\mathbf{r}_2, \tag{4}$$

$$(A|\mu\nu) = \int \chi_A(\mathbf{r}_1) \frac{1}{r_{12}} \chi_\mu(\mathbf{r}_2) \chi_\nu(\mathbf{r}_2) d\mathbf{r}_1 d\mathbf{r}_2, \tag{5}$$

and the latter are half transformed using the LMO coefficients $\mathbf{C}$ as

$$(A|\mu i) = \sum_{\nu \in [i]_{\mathrm{LMO}}} (A|\mu\nu) C_{\nu i}, \quad \mu \in [i]_{\mathrm{AO}}, A \in [i]_{\mathrm{fit}}. \tag{6}$$

Here $\chi_A(\mathbf{r})$ are fitting basis functions, and $[i]_{\mathrm{LMO}}$, $[i]_{\mathrm{AO}}$, $[i]_{\mathrm{fit}}$ denote orbital-specific subsets (domains) of basis or fitting functions. The number of functions in each of these domains (in the following "the domain sizes") are asymptotically independent of the molecular size. The domain construction will be discussed in Section 2.3.

The exchange matrix is then approximated as

$$K_{\mu\nu} \approx \sum_i \Big[\sum_{\overline{B} \in [i]_{\mathrm{fit}}} (\overline{B}|\mu i)(\overline{B}|\nu i), \quad \mu, \nu \in [i]_{\mathrm{AO}} \Big]. \tag{7}$$

The integrals $(\overline{B}|\mu i)$ are obtained by solving for each orbital i a linear equation system

$$(A|\mu i) = \sum_{A \in [i]_{\mathrm{fit}}} G_{AB} (\overline{B}|\mu i), \quad A, B \in [i]_{\mathrm{fit}}, \ \mu \in [i]_{\mathrm{AO}}. \tag{8}$$

The triangular matrices G_{AB} are determined separately for each LMO fitting domain by a Cholesky decomposition of the Coulomb metric J_{AB}:

$$[J]_{AB} = [\mathbf{G}\mathbf{G}^\dagger]_{AB}, \quad A, B \in [i]_{\mathrm{fit}}. \tag{9}$$

The restriction of all summations to domains means that for each individual LMO i the computational effort is (asymptotically) independent of the molecular size, and thus linear scaling results by summing over all i.

Since the exchange contribution of each LMO i is restricted to an AO domain $[i]_{\mathrm{AO}}$, it must be scattered to a matrix in the full AO basis and then added to the final exchange matrix, cf. eq. (7).

In the asymptotic limit all domain sizes should become independent of the molecular size and then the algorithm scales linearly. However, as has been demonstrated in Ref. 18, the LMO-domains reach a constant size only for quite large molecules. The AO domains depend on the LMO domains (see below), and are even larger and more slowly convergent. For medium size molecules they will often include all or nearly all basis functions. In Table 1 we have summarized the theoretical asymptotic scalings of the main computational steps in the exchange matrix calculation (i) without local approximations, (ii) with local fitting only, and (iii) with all local approximations, i.e. including LMO and AO domains. The most important savings are due to the local fitting, which reduces the scaling of the fitting and assembly steps from quartic to quadratic and cubic, respectively. In particular, a quadratic dependence on the size of the fitting basis is removed. Linear scaling can only be reached if the AO domains become constant in size. Even though this will often not be the case in practice, the reduction of the CPU and storage requirements by the local fitting are significant. The scaling will then be at most cubic (with a low prefactor), as other steps in the HF calculation. Timings for various examples will be presented in Section 3.

Table 1: Asymptotic scaling of the three main steps in the exchange matrix calculation with molecular size N. It is assumed that the number of 3-index integrals $(\mu\nu|A)$ scales asymptotically quadratic.

Step	non-local	only fitting domains	All domains
Transformation, eq. (6):	N^3	N^2	N
Fitting, eq. (8):	N^4	N^2	N
Assembly, eq. (7):	N^4	N^3	N

2.2 The Coulomb matrix

The Coulomb matrix $J_{\mu\nu}$ is calculated using standard density fitting. First, the 3-index integrals are contracted with the density matrix

$$v_A = \sum_{\rho\sigma} (A|\rho\sigma) D_{\rho\sigma}, \tag{10}$$

and then the linear equations

$$v_A = \sum_B G_{AB} \bar{v}_B, \tag{11}$$

are solved to yield $\bar{v}_B$, where G_{AB} is obtained by a Cholesky decomposition of the matrix J_{AB}, similar to eq. (9), but using the whole fitting basis. This decomposition is carried out only once and reused in each HF iteration. The vector $\bar{v}_B$ is finally contracted with 3-index integrals to yield the Coulomb matrix:

$$J_{\mu\nu} = \sum_B \bar{v}_B (B|\mu\nu). \tag{12}$$

The integral evaluation and contraction steps are efficiently parallelized, but do not use local approximations and point-group symmetry. In principle, local density fitting approximations could also be used for the Coulomb part, but as already pointed out in Ref. 56 this causes much larger errors in the energy than the local exchange calculation, unless the domains are

very large. We have therefore not attempted such approximations in the current work.

The Cholesky decomposition matrix J_{AB} scales cubically with molecular size, but this has to be done only once in the beginning of the HF calculation. All other computational steps scale (in the asymptotic limit) quadratically with the molecular size.

2.3 Domains

In our current program the basis functions are treated in blocks, and one block contains all functions at one atom (center). Therefore, all domains are also center-based, i.e., they always include all basis or fitting functions at a subset of centers. While this block structure minimizes the logic and is efficient for use in matrix multiplications, the domains may be larger than absolutely necessary. Further improvements may be possible by using smaller basis function blocks, as recently implemented for computing the F12 energy contribution in explicitly correlated LMP2 methods.[21]

By default, only the starting orbitals which enter the DF-HF calculation are explicitly localized. In the subsequent iterations the LMOs are obtained by maximizing their overlap with the initial localized orbitals, as described in Ref. 56. In this case the same fitting, LMO and AO domains can be used in all iterations. We found that in most cases this is sufficient and yields excellent performance with fast and stable convergence. However, a disadvantage of this scheme is that the domains may depend on the initial starting guess, and therefore slightly different results could be obtained if for example orbitals from a neighboring geometry are used as a starting guess. This problem can be avoided by repeating the localization and domain selection in subsequent iterations or every n'th iteration. This will be further investigated in Section 3.

In order to obtain accurate orbitals and energies we use the following strategy:
(i) As proposed in Ref. 56, we recalculate the energy with full fitting domains once the

HF procedure is converged. It will be shown in Section 3 that this strongly improves the accuracy of the HF energy without much additional cost (since the full exchange matrix is not needed);
(ii) We use larger density fitting domains during the iterations than in Ref. 56 to ensure more accurate Fock-matrices and correlation energies.

The final energy can be computed using 3-index integrals $(ij|A)$, which are obtained in two transformation steps $(\mu\nu|A) \rightarrow (\mu i|A) \rightarrow (ij|A)$. Without any local approximations, the second-half transformation scales as $N_{\mathrm{AO}} N_{\mathrm{occ}}^2 N_{\mathrm{fit}}$. Furthermore, the fitting step [cf. eq. (8)] scales as $N_{\mathrm{occ}}^2 N_{\mathrm{fit}}^2$, i.e. with the fourth power of the molecular size. Although the prefactors of these two steps are small, they might become dominant in very large cases, since all other computational steps of the LDF-HF calculation scale at most cubically. This problem can be avoided by exploiting the sparsity of the LMOs, which reduces the asymptotic scaling of the two steps to $\mathcal{O}(N^2)$ and $\mathcal{O}(N^3)$, respectively.

Provided that errors caused by the LMO screening are negligible (which can be controlled by a cutoff threshold), the energy obtained in this way is an upper bound to the exact DF-HF energy (as obtained without local approximations). As will be shown in Section 3, the errors of the final energy (relative to the exact DF-HF energy) are in the sub-μH range, provided the fitting domains are reasonable large. However, due to the local approximations in the orbital optimization, the final energy is not exactly stationary with respect to infinitesimal changes of the orbitals. This might slightly affect orbital energies and other expectation values (as for example multipole moments), as well as analytical energy gradients. The latter problem has already been discussed and tested in Ref. 69, and it was found that the effect on optimized geometries is negligible (even though in that work smaller domains than in the current work were used). The effect on orbital energies, dipole moments and LMP2 correlation energies will be demonstrated in Section 3.

2.3.1 Local occupied orbitals and LMO domains

In a first step, we localize the occupied orbitals by transforming them to IBOs.[58] Since only a subset of all AOs have non-negligible contributions to each LMO, we can use this inherent sparsity as a first condition to reduce the scaling of the transformation in eq. (6). To be more precise, we set the coefficients $C_{\mu i}$ of the basis functions at one atomic center X to zero, if $\sum_{\nu \in X} |C_{\nu i}|^2 \leq T_{\mathrm{LMO}} = 10^{-6}$. The remaining non-zero coefficients are readjusted by fitting the resulting approximate LMOs to the original ones.[18,65,70] In this way only a subset of all centers contributes to the orbital coefficient vector for each LMO i, and we denote these domains as $[i]_{\mathrm{LMO}}$. The screening and fitting of the the LMO coefficients scales quadratically with the molecular size and does not take a significant amount of time.

2.3.2 Atomic orbital (AO) domains

Next, the dimension of the atomic orbital index μ in eq. (6) has to be made independent of the molecular size for each i. This is achieved by screening the 3-index integrals $(A|\mu\nu)$, which decay exponentially with the distance between the basis functions χ_μ and χ_ν. The screening is done for blocks of AOs (centers), and determines the range of indices μ which yield non-negligible integrals with any ν in the LMO domain $[i]_{\mathrm{LMO}}$. To each block of basis functions a sphere is assigned, and the integrals $(A|\mu\nu)$ are skipped if the spheres of the blocks containing μ and ν do not overlap.

The radius r_{B} assigned to a block of basis functions is the maximum radius r_μ assigned to any basis function in the block. The radius r_μ for a basis function depends on its exponent, and is chosen such that if all contributions of the basis function outside a sphere with radius r_μ are neglected, the norm changes by less than the threshold T_{B} (assuming normalized basis functions). We use a default value of $T_{\mathrm{B}} = 10^{-5}$. Larger values of this threshold may lead to significant errors, while smaller values do not improve the accuracy significantly. Therefore,

this default value was used to calculate exchange contributions during the iterations. For the accurate recalculation of the final energy without local approximations we used a value of $T_B = 10^{-8}$ in order to avoid any effect of the screening. However, a threshold of 10^{-6} should normally be sufficient to obtain microhartree accuracy.

The AO domains $[i]_{\mathrm{AO}}$ contain all basis functions ν which survive the screening with $\mu \in [i]_{\mathrm{LMO}}$. Thus, they depend on the screening threshold T_B and indirectly on the LMO threshold T_{LMO}. Asymptotically, the AO domain sizes become independent of the molecular size, even though they are much larger than the LMO domains. In attempts to reduce the AO domain sizes, we also tested distance and/or connectivity as well as overlap criteria, i.e. $S_{\mu\nu}$ with $\nu \in [i]_{\mathrm{LMO}}$, but found that the AO domain sizes cannot be significantly reduced without affecting the accuracy. Thus, the integral screening appears to be the most natural and accurate approach, providing good accuracy and efficiency. Moreover, the introduction of additional thresholds or parameters for the definition of the AO domains is avoided.

2.3.3 Fitting domains

The fitting domains are constructed using well established distance and connectivity criteria:[18,21] (i) primary domains for each LMO are defined such that all functions of a center with an IBO partial charge larger than 0.2 are included, and (ii) these primary domains are then extended by adding functions located at neighboring centers if the distance between these centers and the ones in the primary domain is smaller than a particular distance threshold R_{DF} and is connected with (or less than) I_{DF} bonds. I_{DF} determines the maximum number of bonds by which a center is separated from any center in the primary domain. Two atoms are considered to be bonded if their distance is smaller or equal to the sum of their atomic radii times 1.2. The connectivity criterion has the advantage that it takes into account different bond lengths, depending on the atoms involved. In addition, the distance criterion makes sure that centers which are not connected by $\leq I_{\mathrm{DF}}$ bonds but happen to be close anyway

(e.g. in molecular clusters or at transition states) are included in the domains. Therefore, both the connectivity criterium I_{DF} and distance criterium R_{DF} have to be fulfilled, and we used $R_{\mathrm{DF}} = (2I_{\mathrm{DF}} + 1)\, a_0$ so that only one of the two parameters has to be chosen.

2.4 Abelian point group symmetry

Given the Fock matrix, no locality or sparsity is used in the HF iterations. This part can therefore be carried out in a symmetry adapted basis, yielding canonical orbitals, which transform according to irreducible representations. The vanishing integral rule then leads to a block-diagonal structure of all matrices. This can save considerable CPU-time, in particular for the diagonalization. However, the use of localized orbitals prevents us from using a symmetry adapted basis in the exchange calculation. Furthermore, our current density fitting program for evaluating the Coulomb matrix does not support symmetry. For the Fock-matrix evaluation the MO-coefficients and the density matrix are therefore transformed from the symmetry adapted to the corresponding non-symmetrized basis, and the final Fock matrix is transformed back into the symmetry-adapted basis. In the exchange calculation one can exploit that subsets of localized orbitals and basis functions are symmetry equivalent, and generate from integrals $(\mu i|\nu i)$ the symmetry equivalent integrals $(\rho j|\sigma j)$ by mapping the respective basis functions μ, ν onto ρ, σ according to the symmetry operation that connects symmetry-equivalent orbitals i and j.[19] For efficiency reasons this approach necessitates that all symmetry-equivalent exchange contributions are calculated on the same CPU core when the program is run in parallel, cf. Section 2.7. Due to the additional logic, the speedup using this type of symmetry is quite modest, but the amount of memory needed for the density-fitted exchange integral evaluation can be significantly reduced.

2.5 Diagonalization of the Fock matrix

In contrast to a conventional HF implementation, we fully diagonalize the Fock matrix only in the first and last iterations of the HF procedure and use an efficient pseudo-diagonalization procedure described in a previous publication.[56] Similar methods have been developed earlier for large semi-empirical calculations.[71,72]

In our approach, the orbital space is partitioned into a primary orbital space containing just the orbitals within an energy window of $\pm 1 E_h$ above and below the HOMO/LUMO gap, and a secondary orbital space containing the rest of the orbitals. The Fock matrix is then exactly diagonalized only in the primary orbital space and approximately (using orbital rotations) in the secondary space. This makes the CPU time for diagonalization of the Fock matrix almost negligible during the iterations, whereas full diagonalizations would in large cases become even more costly than the calculation of the Fock matrix. As an example, the full diagonalization of the Fock matrix for $(\mathrm{Gly})_{40}$ using the VTZ basis set (6538 CGTOs) takes 76.8 s, whereas the pseudo-diagonalization takes only 17.1 s. Nevertheless, this procedure has neither an effect on the final results and nor does it slow down the speed of convergence.

2.6 Starting guess

In order to achieve fast and stable converge, a good starting guess for the orbitals is required. We use a superposition of atomic densities, constructed from a minimal basis of atomic orbitals at each atom. If a generally contracted basis set such as cc-pCnZ is employed, the contracted basis functions spanning the minimal AO basis for each atom can simply be used. This yields a minimal basis for the molecule in which each MO vector has only a single coefficient. Otherwise, for segmented basis sets, the minimal basis for each atom is created by projecting a minimal atomic basis to the current basis. The minimal basis sets are stored in the basis set library (e.g. for the cc-pVTZ basis). Each molecular orbital in the minimal

basis then has coefficients only at one atom and one angular momentum type.

From these atomic orbitals we construct atomic densities, assuming certain occupation numbers which have been guessed or roughly optimized for each atom and are stored in a library. Inner-shell core orbitals have always occupation numbers of 2, but valence orbitals have fractional occupation numbers. For example, for the atoms B through Ne the $2s$ occupation numbers n_{2s} are assumed to be 1.3 (Be), 1.0 (C), 1.7 (N), 1.8 (O), 1.9 (F), 2.0 (Ne), and the $2p$ ones are for each component $(Z - 2 - n_{2s})/3$, where Z is the nuclear charge. Thus, all atoms are assumed to be neutral and spherical. The resulting block-diagonal density matrix is used to construct a Fock matrix, which is transformed to an orthogonal basis and diagonalized as usual to obtain the starting orbitals, which are subsequently localized.

Since the initial density matrix is extremely sparse, the Fock matrix can be computed rather quickly in a standard integral-direct way. However, for large non-symmetric molecules, we found it even faster to use density fitting, with LDF approximations for exchange. In this case, the exchange matrix calculation involves the whole minimal basis of (extremely sparse) orbitals with the appropriate occupation numbers. It was found that the quality of the starting guess is very insensitive to local fitting approximations, and fitting domains which only include functions at the center where the corresponding AO is located are sufficient. Thus, the calculation of the exchange matrix is very fast and the overall total time for the starting guess is in large cases mainly dominated by the diagonalization time.

In principle, one could compute the starting guess in a smaller basis, but then the starting orbitals would have to be projected onto the current basis. This requires the inverse overlap matrix of the smaller basis and the overlap matrix between the two basis sets, and the cost for the projection largely compensates the savings by computing the Fock matrix in a smaller basis. We have therefore not used this possibility in this work.

2.7 Parallelization

The calculation of the three-index integrals in eq. (6) is dynamically parallelized by distributing blocks of fitting functions over the available processors (compute cores) and using the GA toolkit.[59,60] The GA software provides one-sided access to the global data and is efficient both within a compute node with shared memory and between different nodes with a fast network connection. The half transformed integrals $(\mu i|A)$ are written into a global array. This operation is non-blocking and does normally not cause a communication problem. In all calculations presented in this paper, the memory of one compute node (256 GB) was sufficient to store the integrals $(\mu i|A)$. For even larger molecules or basis sets two or more compute nodes may be necessary.

When all integrals have been completed and written, synchronization is necessary. In the subsequent assembly step [cf. eq. (7)] the orbitals i are dynamically distributed amongst the available CPU cores. The required integrals for a given orbital i are then recovered from the GA. This step is communication intensive, but using an Infiniband network we have no indications of a communication bottleneck. The evaluation of the J_{AB} integrals, and the solution of eq. (8) and eq. (9) is performed on the fly independently for each i and therefore does not require any communication. Since the fitting domains of different LMOs i overlap, the evaluation of the J_{AB} matrices is partially redundant. However, their calculation scales linearly with molecular size and is so fast that it is not worthwhile to extract the appropriate domains from the full J_{AB} matrix, which would then have to be kept in memory on each core. In the end, the final exchange matrix is globally summed over all cores.

An alternative implementation, where the LMOs i are statically distributed on the CPU cores, turned out to be less efficient than the one just described, because of the repeated calculation of three-index basis integrals $(A|\mu\nu)$ on each core. Since this approach avoids communication between the cores completely (apart from the final global summation), it may be a useful option for clusters with a slow communication network.

All matrix multiplications outside the parallelized Fock matrix evaluation are parallelized by evaluating blocks of the result matrix on different processors. At the end, each block must be broadcasted to all other processes. Therefore, the total speed is limited by the communication bandwidth, even with a fast Infiniband network. However, the overall computation time is strongly dominated by the Fock matrix evaluation, and therefore we did not find it worthwhile to develop more advanced parallelization schemes for solving the HF equations.

Apart from the Fock matrix evaluation, the slowest step of a HF iteration is the non-parallelized diagonalization - unless the pseudo-diagonalization described earlier is used. Full diagonalizations are still required in the starting guess, as well as in the first and last iterations. We attempted to speed up these diagonalizations using a multithreaded version of the Lapack routine `dsyevd` using all 20 CPU cores on one of the nodes (using the Intel MKL library). A similar multithreaded approach can also be used for the Cholesky decomposition (`dpotrf`) of the matrix J_{AB}, which is needed for the Coulomb fitting. Even though this may speed up individual diagonalizations or Cholesky decompositions by a factor of 4-6 (using 20 cores), the problem of this multithreaded approach is that all cores of a node have to be reassigned temporarily to one process. This makes it impossible to pin each process to a specific core. Without pinning, the remaining calculation is slowed down by a few percent. Due to this, we found only moderate overall speed-ups in our tests and therefore did not further follow this idea nor used this approach in any of the benchmarks presented in this paper.

3 Results

In the following, the determination of thresholds for our Hartree-Fock program will be exemplified using as a test system a linear $(\mathrm{Gly})_{16}$ polypeptide chain. This system was chosen since the asymptotic domain sizes are most quickly reached, and the errors caused by domain

Table 2: Visualization of selected molecular structures used for calculations presented in Table 3. More information about the structures of the other remaining molecules can be found in the Supporting Information.

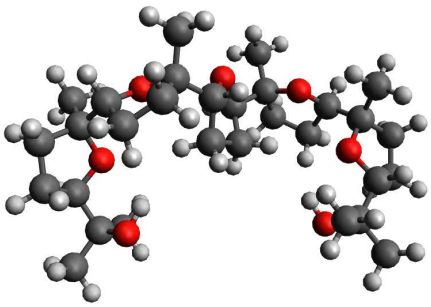
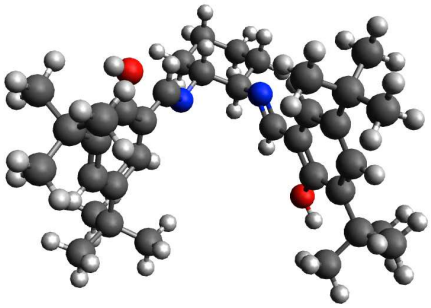
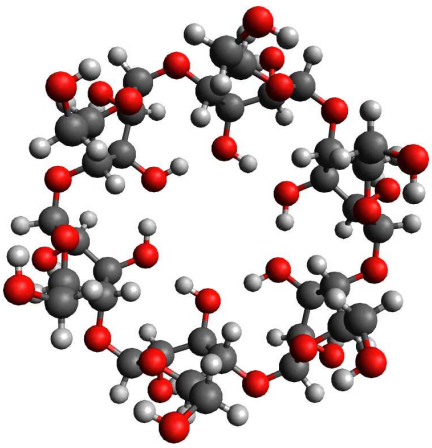
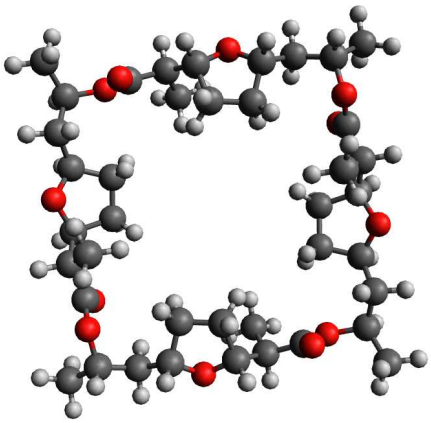
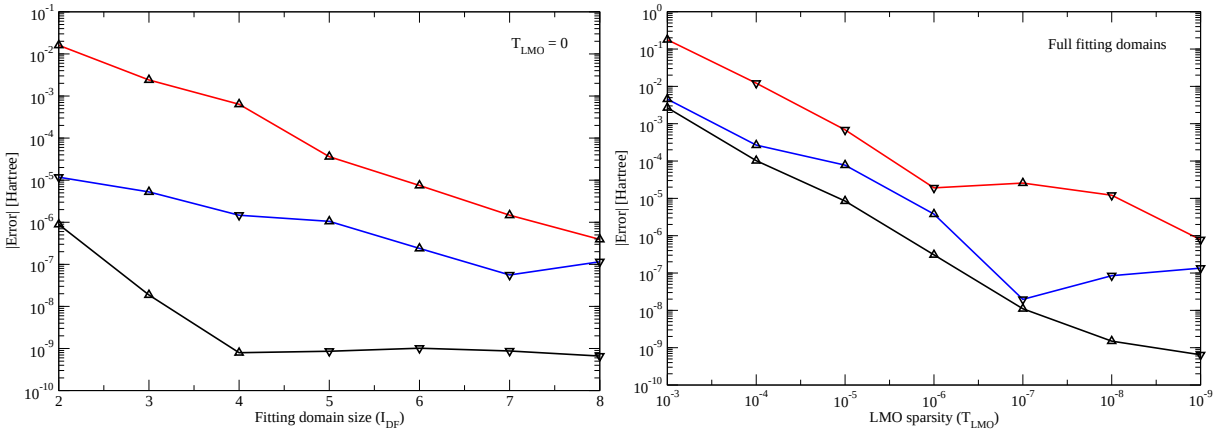
	
Glabrescol, $C_{30}H_{52}O_7$	(R,R)-Jacobsen's ligand, $C_{36}H_{54}N_2O_2$
	
α-Cyclodextrin, $C_{36}H_{60}O_{30}$	Nonactin, $C_{40}H_{64}O_{12}$

Figure 2: Visualization of the error (in E_H) in the total Hartree-Fock energy of the $(Gly)_{16}$ molecule as a function of I_{DF} (left panel) and T_{LMO} (right panel). The reference values are obtained with the canonical DF-HF program in MOLPRO.[68,73] Left panel: The red line shows the errors if the fitting domain size (I_{DF}) is the same in all iterations. The black line represents the errors if the LDF-HF energy is recalculated using full fitting domains once the HF procedure is converged. The blue line shows the errors in the PNO-LMP2 correlation energy using the Fock matrices from the LDF-HF. These errors are independent of whether or not the final HF energy is recomputed without local approximations. Triangles indicate if the calculated energy is above (triangle pointing up) or below (triangle pointing down) the corresponding reference energy. Right panel: Using the same color coding, the LDF-HF and PNO-LMP2 correlation energy errors are shown as a function of T_{LMO}.



approximations should become visible already for relatively small molecular sizes. We employ the correlation consistent cc-pVDZ and cc-pVTZ basis sets[74] (in the following denoted VDZ and VTZ, respectively), as well as the corresponding augmented sets aug-cc-pVDZ and aug-cc-pVTZ[75] (in the following denoted AVDZ and AVTZ, respectively). In all of our calculations, the cc-pVTZ/JKFIT fitting basis set[55] is used. In the starting guess and for the calculations with VDZ or AVDZ basis, the highest angular momentum functions in the fitting basis were removed.

Using the established thresholds, we will then demonstrate the efficiency and accuracy of our approach using 3-dimensional molecules with up to several hundred atoms and thousands of basis functions. Some of the molecules are shown in Table 2, all others can be found in the Supporting Information. The scaling behavior of our algorithm with respect to the molecular size will be demonstrated in Section 3.3 using extended glycine chains as well as alanine

helices. The speedup with the number of cores and compute nodes will be demonstrated in Section 3.4.

The calculations were run on a compute cluster with QDR Inifiniband communication. Each node contains two Intel® Xeon® E5-2690, 2.8 GHz processors with in total 20 CPU cores and 256 GB memory. Unless otherwise noted, all 20 cores were used in each node.

3.1 Thresholds for domains

The calculation of exchange integrals as described in Section 2.1 depends on two parameters: the threshold T_{LMO} for the LMO sparsity, cf. Section 2.3.1, and the connectivity and distance criteria I_{DF} and R_{DF}, which determine the size of the local fitting domains, cf. Section 2.3.3. As mentioned earlier, we use $R_{\mathrm{DF}} = 2I_{\mathrm{DF}} + 1\ a_0$, so that only I_{DF} is taken as an independent parameter.

We aim to obtain an accuracy of $< 10^{-5} E_{\mathrm{h}}$ for the Hartree-Fock total energy and errors of $< 10^{-4} E_{\mathrm{h}}$ for PNO-LMP2[18,19] correlation energies when compared to calculations with orbitals from the canonical density-fitted Hartree-Fock (DF-HF) implementation. The errors of relative energies should be even smaller, as will be demonstrated for some reaction energies in Section 3.2.2. Note that in the PNO-LMP2 calculations the LDF-HF Fock matrix and orbitals were used. Since the LMP2 equations are solved iteratively, the LDF-HF orbital energies are not needed in the correlation calculation.

Using a $(\mathrm{Gly})_{16}$ molecule and the VTZ basis set as test system, we found that the errors caused by the fitting domains and the LMO sparsity are largely independent from each other. We thus investigated each of the two approximations separately using an energy convergence threshold of $10^{-8}\ E_{\mathrm{H}}$. In order to avoid additional errors due to AO domains or integral screening, the screening threshold was set to $T_{\mathrm{B}} = 10^{-10}$ during the iterations. For the recalculation of the LDF-HF energy (if applicable), $T_B = 10^{-8}$ was used. Figure 2 visualizes

the absolute errors in the LDF-HF total and PNO-LMP2 correlation energies using various fitting and LMO domain sizes determined by I_{DF} (left panel) and the LMO sparsity threshold T_{LMO} (right panel), respectively.

The accuracy of the LDF-HF energy mainly depends on the fitting domain size, and the convergence to the DF-HF energy obtained with full domains is very slow. This is shown with the red line in the left panel of Figure 2. The error can be reduced by localizing the orbitals and recomputing the Fock matrix with larger domains once the HF procedure is converged, but the convergence is still slow (see Supporting Information). We found it more accurate and efficient to recalculate the final HF energy without any local approximations (black line). As demonstrated by comparison of the red and black lines of the figure, the accuracy of the final HF energy is then improved by several orders of magnitude. The calculation of the exchange only requires to compute the fully transformed integrals $(ij|A)$ and is therefore cheaper than computing the full Fock matrix with very large domains. Note that this approach does not change the Fock matrix and the orbitals, and therefore the error in the PNO-LMP2 correlation energy (blue line) only depends on the domains used during the iterative solution of the HF equations.

The LMO sparsity threshold T_{LMO} also influences the accuracy in LDF-HF total and PNO-LMP2 correlation energies (same color coding, right panel). However, the dependency of the LMO sparsity threshold on the accuracy of LDF-HF total and PNO-LMP2 correlation energies is somewhat less systematic, due to some error compensations. Again, we do not exploit the sparsity of the LMOs when recomputing the energy after convergence. More details about the dependency of the energy on the fitting domain size and the LMO sparsity threshold can be found in the Supporting Information.

According to these results and our current experience with other molecules, using an LMO sparsity threshold of $T_{\mathrm{LMO}} = 10^{-6}$ and a fitting domain parameter of $I_{\mathrm{DF}} = 3$ are normally sufficient to obtain PNO-LMP2 correlation energies in accordance with our chosen accuracy

criterion of $10^{-4}E_{\mathrm{H}}$. We cannot exclude, however, that larger fitting domains may be needed in other cases, in particular for highly delocalized electronic structures with low HOMO-LUMO gaps. This is similar as in all local correlation treatments.

3.2 Performance and accuracy

3.2.1 Absolut energies

Using the thresholds determined in Section 3.1, the performance and accuracy for a set of molecules is shown in Tables 3 and 5. As can be seen, the determined thresholds are appropriate for molecules of different size and using different basis sets. For the medium sized molecules the AO domains are in most cases hardly smaller than the total AO basis set size. Thus, the AO domains only become effective for the three largest systems, i.e. $(\mathrm{Gly})_{40}$, $(\mathrm{Ala})_{29}$, or $(\mathrm{H_2O})_{60}$. However, the fitting domains are always much smaller than the fitting basis size, and this has the largest effect on the performance.

The average AO and fitting domain sizes for alanine helices and glycine chains converge to constant values with increasing molecular size, as shown in Figure 3. While the average fitting domain sizes (solid red line) converge quite fast and stay relatively small, the average AO domain sizes (solid blue line) depend much more on the three dimensional structure of the molecule, and accordingly, these domains converge to a much larger value for the more space filling alanine helices as they do for the linear glycine chains. For comparison, the dashed red and blue lines represent the values obtained without local approximations (full domains).

As shown in Tables 3 and 5, the use of the diffuse AVDZ or AVTZ basis sets has a dramatic effect on the sparsities and consequently on the performance. The calculations with the AVTZ basis typically take 4-5 times longer than those with the VTZ basis. This is at least

Table 3: Domain sizes and timings for LDF-HF calculations using $T_{\mathrm{LMO}} = 10^{-6}$, $I_{\mathrm{DF}} = 3$, and $R_{\mathrm{DF}} = 7\,a_0$ (see text). The times are average elapsed times for a single Hartree-Fock iteration, without the starting guess and excluding the first and last iterations (see Table 5 for the total elapsed times). All calculations were performed using 20 CPU cores in one compute node.

							Time per iteration [s]		
	Basis set size		Sparsity [%]		Av. domain size		LDF-HF		DF-HF
Molecule (basis), point group	AOs	FIT	LMOs	Screen	AO	FIT	Exch.	total	total
(*S*)-BINOL (AVTZ), C_2	1334	2158	0.3	0.0	1334.0	1079.0	6.7	8.8	7.9
18-Crown-6 (AVTZ), C_i	1380	2142	1.1	0.0	1380.0	907.0	5.8	7.9	8.1
Glabrescol (VTZ), C_2	1838	4483	46.8	40.3	1814.0	1223.7	12.8	17.9	46.3
(*R*,*R*)-Jacobsen's ligand (VTZ), C_2	1956	4780	50.4	48.1	1936.2	1121.8	13.1	18.7	53.8
α-Cyclodextrin (AVDZ), C_2	2058	6000	7.9	0.7	2058.0	1097.5	38.1	52.6	143.7
$(H_2O)_{60}$ (VTZ),C_1	3480	8340	80.5	73.0	2892.0	689.9	33.1	54.7	514.6
$(Ala)_{29}$ (VTZ),C_1	6438	15944	82.6	91.9	2820.8	1156.2	95.0	203.5	4910.1
$(Gly)_{40}$ (VTZ),C_1	6538	16379	90.8	98.3	1327.1	822.7	24.3	118.9	4975.5
Basis set dependence:									
Nonactin (VDZ), C_2	1048	5112	64.1	70.3	956.2	972.9	4.6	8.0	31.4
Nonactin (AVDZ), C_2	1772	5112	10.2	1.4	1772.0	972.9	22.4	31.8	73.9
Nonactin (VTZ), C_2	2456	6028	60.9	62.0	2356.5	1139.0	18.9	28.3	129.4
Nonactin (AVTZ), C_2	3864	6028	8.2	0.4	3864.0	1139.0	88.6	116.1	284.2
Elaiophylin (VDZ), C_1	1448	7064	70.0	79.8	1006.9	1010.9	10.3	17.9	108.0
Elaiophylin (AVDZ), C_1	2448	7064	18.6	28.3	2448.0	1010.9	55.7	77.2	253.8
Elaiophylin (VTZ), C_1	3392	8328	67.2	75.0	2621.5	1181.8	44.0	67.4	468.1
Elaiophylin (AVTZ), C_1	5336	8328	17.1	24.2	5336.0	1181.8	234.9	303.5	1040.0

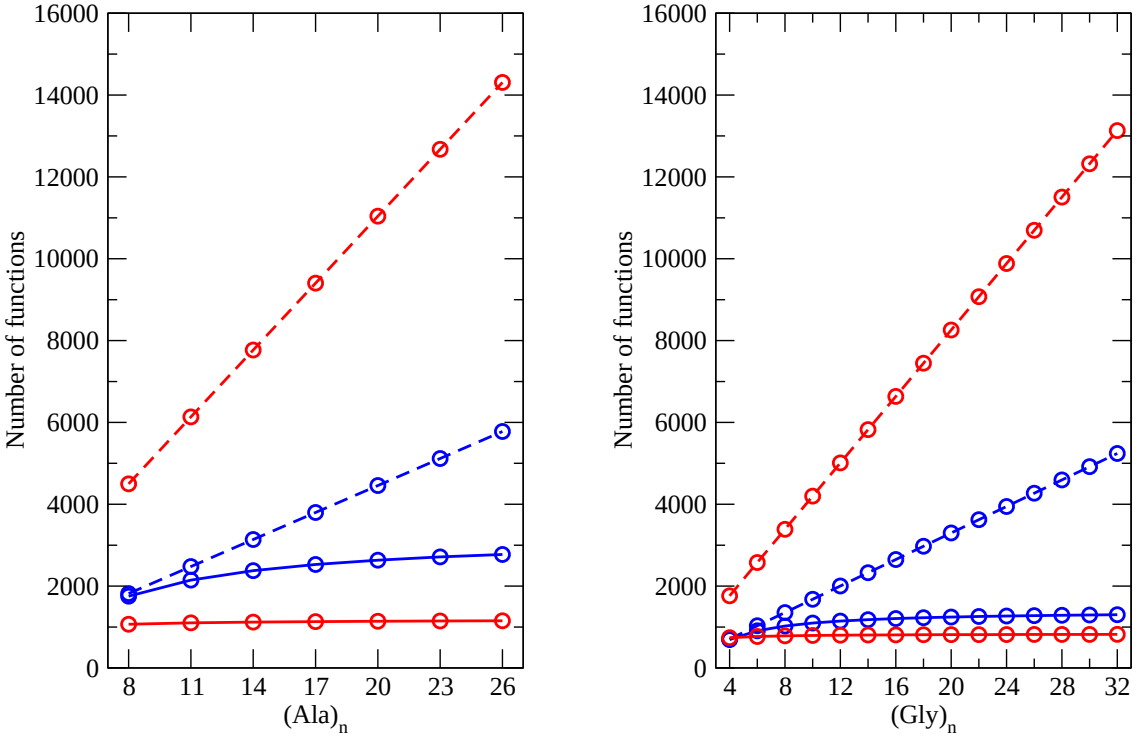
Table 4: Errors of Hartree-Fock total and PNO-LMP2 correlation energies, the RMS and MAX errors in the occupied orbital energies as well as the errors in the dipole moments μ caused by local approximations in the LDF-HF calculation. Recalculating the energy without any local approximations after the last iteration improves the LDF-HF energy by 3-4 orders of magnitude. The same parameters as in Table 3 have been used.

	Error [$m\mathrm{E_H}$]					
	$dE^{\mathrm{LDF-HF}}$			Orbital energies		
Molecule (basis), point group	last iter.	recal. energy	$dE_{\mathrm{corr}}^{\mathrm{PNO-LMP2}}$	RMS	MAX	$d\mu$ [D]
(*S*)-BINOL (AVTZ), C_2	−3.3546	−0.0006	−0.1018	0.047	0.119	0.004065
18-Crown-6 (AVTZ), C_i	−0.3683	0.0000	−0.0024	0.006	−0.008	0.000000
Glabrescol (VTZ), C_2	−1.1767	−0.0002	−0.0203	0.012	0.047	0.000299
(*R*,*R*)-Jacobsen's ligand (VTZ), C_2	−2.7699	−0.0004	−0.0593	0.034	0.195	0.003097
α-Cyclodextrin (AVDZ), C_2	−3.4036	0.0000	−0.0523	0.022	0.062	0.000500
$(H_2O)_{60}$ (VTZ),C_1	0.5181	−0.0034	−0.0801	0.024	0.117	0.000731
$(Ala)_{29}$ (VTZ),C_1	−7.3291	−0.0009	−0.1267	0.021	0.067	0.002302
$(Gly)_{40}$ (VTZ),C_1	−5.8269	−0.0007	−0.0288	0.014	0.086	0.001505
Basis set dependence:						
Nonactin (VDZ), C_2	−1.1881	−0.0012	−0.0076	0.015	−0.041	0.000004
Nonactin (AVDZ), C_2	−1.7362	0.0000	−0.0284	0.010	0.019	0.000001
Nonactin (VTZ), C_2	−1.3865	−0.0003	−0.0212	0.012	0.029	0.000001
Nonactin (AVTZ), C_2	−1.5989	−0.0001	−0.0205	0.010	0.021	0.000000
Elaiophylin (VDZ), C_1	−4.3737	−0.0017	−0.0556	0.037	−0.131	0.004725
Elaiophylin (AVDZ), C_1	−5.3083	−0.0008	−0.0649	0.033	0.218	0.005099
Elaiophylin (VTZ), C_1	−4.5013	−0.0008	−0.0556	0.037	−0.118	0.005005
Elaiophylin (AVTZ), C_1	−5.0085	−0.0015	−0.0456	0.030	0.168	0.005155

Table 5: The times are total elapsed times, including the starting guess and in the LDF-HF case the recomputation of the final energy. All calculations were performed using 20 CPU cores in one compute node.

	Elapsed time [s] (Number of iterations)		
Molecule (basis), point group	LDF-HF	DF-HF	PNO-LMP2
(S)-BINOL (AVTZ), C_2	125.7 (12)	90.9 (10)	73.8
18-Crown-6 (AVTZ), C_i	80.8 (8)	75.6 (8)	52.4
Glabrescol (VTZ), C_2	207.7 (9)	433.8 (9)	104.3
(R,R)-Jacobsen's ligand (VTZ), C_2	255.3 (11)	561.0 (10)	105.3
α-Cyclodextrin (AVDZ), C_2	651.1 (10)	1336.6 (9)	207.0
$(H_2O)_{60}$ (VTZ),C_1	795.3 (10)	4702.6 (9)	137.6
$(Ala)_{29}$ (VTZ),C_1	3882.1 (11)	54204.6 (11)	755.6
$(Gly)_{40}$ (VTZ),C_1	2557.9 (11)	55339.5 (11)	179.5
Basis set dependence:			
Nonactin (VDZ), C_2	119.7 (11)	329.7 (11)	27.5
Nonactin (AVDZ), C_2	417.4 (11)	837.1 (11)	100.4
Nonactin (VTZ), C_2	404.6 (11)	1455.2 (11)	130.3
Nonactin (AVTZ), C_2	1558.6 (11)	3225.4 (11)	494.4
Elaiophylin (VDZ), C_1	249.6 (10)	1224.2 (11)	77.4
Elaiophylin (AVDZ), C_1	928.0 (10)	2869.6 (11)	288.3
Elaiophylin (VTZ), C_1	881.9 (10)	5239.1 (11)	364.0
Elaiophylin (AVTZ), C_1	3713.0 (10)	11797.9 (11)	1470.4

Figure 3: Visualization of the average AO (blue) and fitting domain sizes (red) without local approximations (dashed lines) and with local approximations (solid lines) for a series of alanine helices and glycine chains (left and right respectively). All calculations were performed using the VTZ basis set.



partly a consequence of the fact that in our current implementation all basis functions at one atom are treated as a single block. By using smaller blocks and separating diffuse from less diffuse basis functions further improvements should be possible, as has already been discussed in Section 2.1.

Furthermore, by comparing the domains of the largest three systems, i.e. $(Gly)_{40}$, $(Ala)_{29}$ and $(H_2O)_{60}$, it is seen that the LMO and AO domains become larger as the system becomes more 3-dimensional, and the sparsity of the LMOs becomes smaller. Nevertheless, the elapsed times remain relatively low when compared to standard DF-HF calculations. For example, the total LDF-HF calculation for $(Gly)_{40}$ is about 22 times faster than the DF-HF one, and for the Elaiophylin (VTZ) calculation the speed-up is still a factor of about 6. Some examples are also shown in Figure 1.

The accuracy of the computed LDF-HF and PNO-LMP2 energies is within the expected error bounds, compared with the results obtained with a non-local DF-HF program. As expected, the errors are somewhat larger for molecules which a delocalized electronic structure.

The outlined procedure of localizing the occupied orbitals only in the first iteration and keeping all domains fixed throughout the iterations works well for all the examples reported in this paper. However, as already mentioned, a disadvantage of this method is that the results depend slightly on the choice of the initial guess. This problem can be avoided by localizing the orbitals and adjusting the domains in each (or each n'th) iteration. However, the improvements are small (cf. Table 1 of the Supporting Information), and usually not worth the additional effort. Moreover, the speed of convergence is sometimes slowed down by repeated localizations. This is particularly the case if the localization and domain selection is only repeated in each n-th iteration ($n > 1$), and therefore this is not recommended. We also tried to use the overlap between the originally localized and the projected local orbitals as a criterion to repeat the localization, but this lead to similar convergence problems.

Table 4 also reports the root mean square (RMS) and maximum (MAX) deviations of the orbital energies and the dipole moments with respect to the values obtained in respective DF-HF reference calculations. As expected, the errors in orbital energies and dipole moments are larger than the errors in the recalculated LDF-HF energies. In most cases the maximum errors of the orbital energies are below 0.5 mH, and those of dipole moments less than 0.005 D. These errors can be further reduced by increasing the fitting domains.

3.2.2 Reaction energies

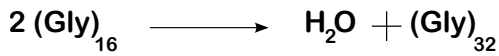
In order to test the accuracy of relative energies, reaction energies for the systems shown in Figure 4 were computed. All calculations were performed using the VTZ basis set. For the gold atom in reaction IV, the ECP60MDF effective core potential[76] for the inner 60 electrons, along with the associated cc-pVTZ-PP basis set, was employed. Abelian point group symmetry was only used in the DF-HF calculations for reaction **I**.

The PNO-LMP2 calculations were carried out using canonical DF-HF as well as LDF-HF calculations, and the differences of the reaction energies obtained with the respective orbitals are presented in Table 6. The total elapsed times correspond to the sum of the elapsed times of all molecular calculations for each reaction.

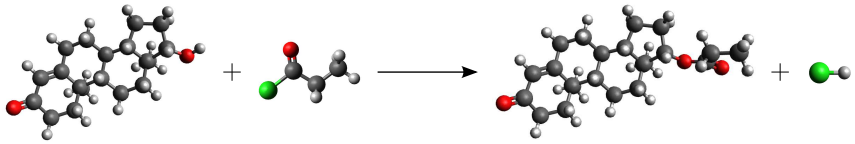
It is found that the local approximations only cause negligible errors in the HF reaction energies, and the absolute errors of the individual energies largely cancel out. To a lesser extent, this is also true for the PNO-LMP2 correlation contributions to the reaction energies. Typically, the errors in reaction energies are an order of magnitude smaller than the absolut errors.

Figure 4: Selected reactions of large molecules

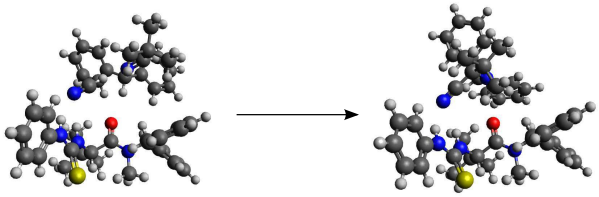
I. Glycine chain condensation



II. Esterification of testosterone[9]



III. Hydrocyanation[77]



IV. Gold(I) aminonitrene complex dissociation[20,78]

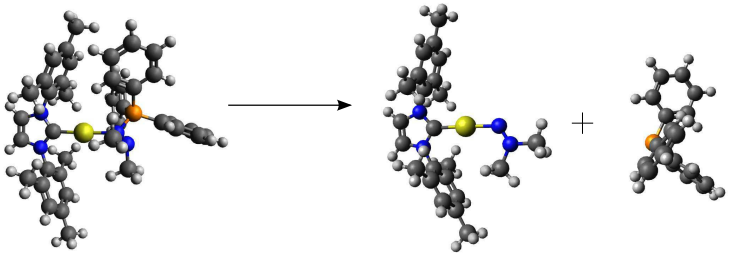


Table 6: Differences ΔE of reaction energies obtained with DF-HF and LDF-HF orbitals for the reactions shown in Figure 4. Recalculation of the energies without any local approximations significantly improves the reaction energies compared to the ones obtained in the last iteration. Timings for LDF-HF, DF-HF, and PNO-LMP2 are also given (using 20 cores).

	Error $[m\mathrm{E_H}]$					
	$\Delta E^{\mathrm{LDF-HF}}$			Total elapsed time $[s]$		
Reaction	last iter.	recal. energy	$\Delta E_{\mathrm{corr}}^{\mathrm{PNO-LMP2}}$	LDF-HF	DF-HF	PNO-LMP2
I	−0.0984	0.0001	0.0005	1717.4	28568.9	197.6
II	−0.2914	−0.0000	−0.0018	192.4	211.7	144.7
III	−1.1906	−0.0002	−0.0058	1061.3	2709.0	620.1
IV	1.0689	−0.0000	0.0181	879.4	1220.2	515.3

3.3 Scaling with molecular size

Figure 5: DF-HF and LDF-HF elapsed times for $(\mathrm{Ala})_n$ (red) and $(\mathrm{Gly})_n$ (black) as function of the chain length n. Left panel: Total elapsed times. Right panel: Average elapsed times per iteration to calculate the exchange (red) and Coulomb (black) matrices, their sum (green), and for one complete iteration (blue). Abelian point C_S group symmetry is used in all DF-HF and LDF-HF calculations for the $(\mathrm{Gly})_n$ molecules. All calculations used the VTZ basis set and were performed using 20 CPU cores in one compute node.

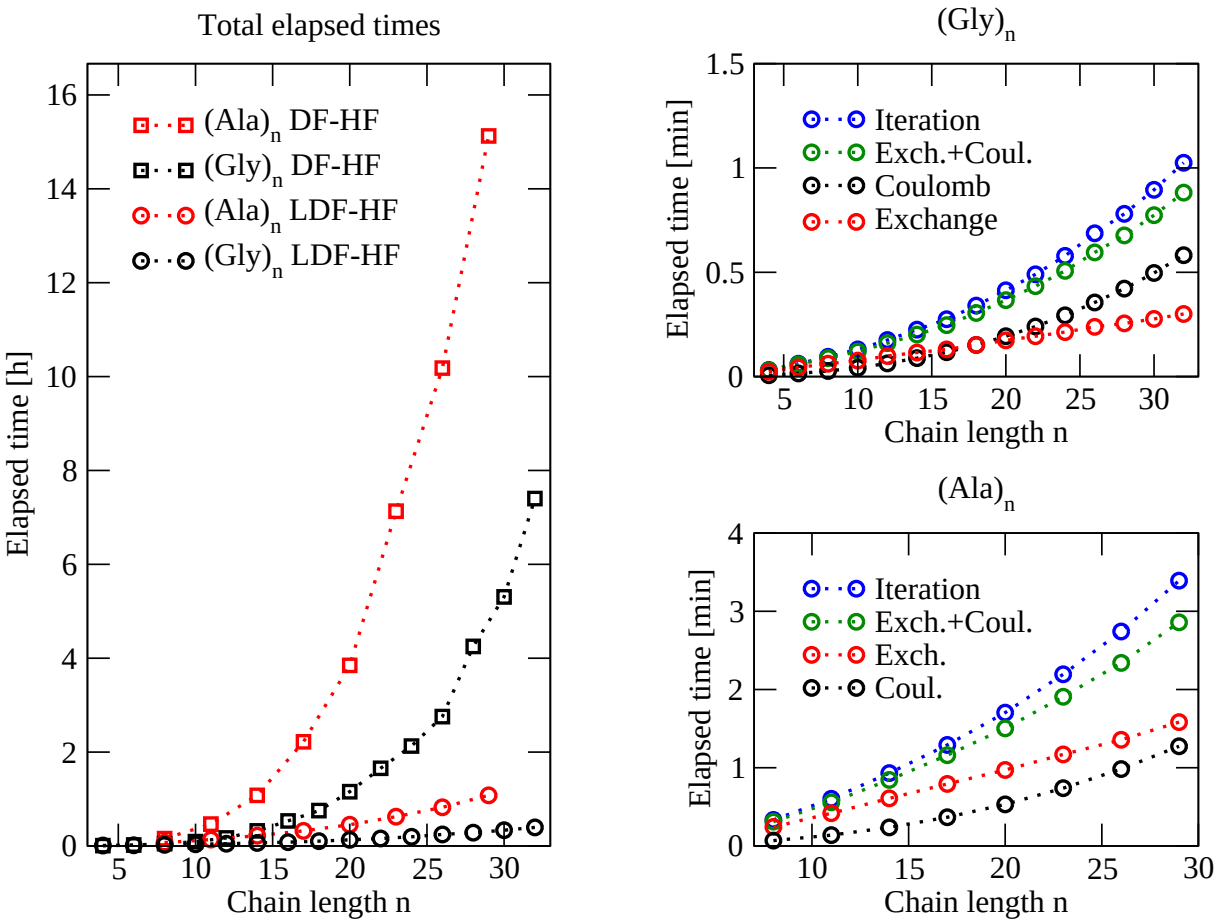
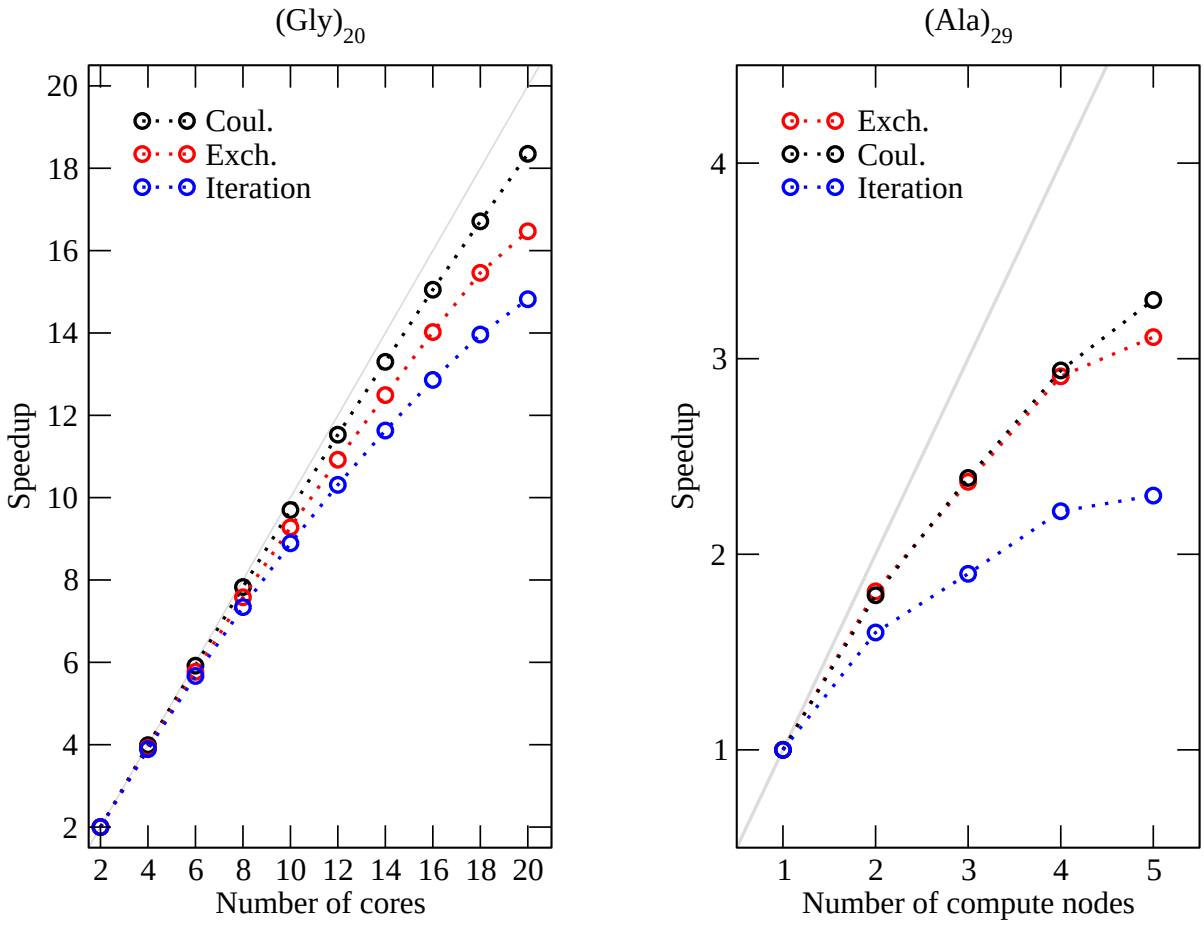


Figure 5 shows the scaling of the elapsed times as a function of the molecular size for the $(\mathrm{Gly})_n$ and $(\mathrm{Ala})_n$ polypeptides. Even though these systems may be regarded as rather unrealistic models, it is still of interest to investigate how the elapsed times scale with molecular size for such extended systems.

First of all, it is seen in the left panel of Figure 5 that the LDF-HF method scales much lower than a conventional DF-HF calculation, though of course not linearly. Therefore, the

Figure 6: Left panel: Average speedups of a single LDF-HF iteration (blue) and of the exchange (red) and Coulomb (black) contributions for $(Gly)_{20}$ using the VTZ basis set and up to 20 CPU cores (one compute node). The initial guess and the first and last iterations are not included. Right panel: similar results for $(Ala)_{29}$, using up to 100 CPU cores (20 cores per node). No Abelian point group symmetry was used throughout.



speed-up becomes larger with increasing chain lengths, and the contribution of the exchange part to the total time becomes smaller. In the right panels it is demonstrated that the computation of the exchange matrix exhibits almost perfect linear scaling, while the scaling exponent of the Coulomb contribution is approximately 2, as expected.

The figure also demonstrates that the calculations for the alanine helices are more expensive than those for the linear glycine chains. For truly 3-dimensional systems the linear scaling regime will be much more difficult to reach, and calculations for molecules with similar numbers of orbitals and basis functions are even more expensive. This is clearly seen in Tables 3 and 5. However, in practice the scaling is often less important than the absolute computation times for molecules of medium size (≈ 100 atoms), which are still significantly reduced by our local approach.

In the largest cases, the Fock matrix evaluation takes more than two thirds of the total time. The rest is used for localization, diagonalization and various matrix transformations. All of these steps scale cubically with molecular size. Since these contributions are less well parallelized they become dominant for very large molecules (due to the scaling) or when many processing cores are used.

3.4 Parallel scaling behavior

Finally, the speedup with the number of CPU cores is demonstrated in Figure 6 for $(Gly)_{20}$ with the VTZ basis set. Using a single compute node and up to 20 CPU cores (left panel), good scaling is observed for the calculation of the Coulomb (blue) and exchange (red) contributions. Also the overall Hartree-Fock procedure (blue) scales reasonably with the number of cores.

The scaling behavior on several compute nodes with up to 100 cores is demonstrated in the right panel of Figure 6 for $(Ala)_{29}$, again using the VTZ basis set. While the exchange and

Coulomb integral calculations still scale reasonably well, yielding a speedup of 3.1 and 3.3 on 5 nodes respectively, the overall Hartree-Fock program reaches only a speedup of 2.4 using 5 nodes with a total of 100 cores (relative to 1 node with 20 cores). This is due to the fact that the localization, diagonalization, and the DIIS procedure are not explicitly parallelized; in these parts only the matrix multiplications are parallel, but their speed is strongly limited by the communication bandwidth, cf. Section 2.7.

4 Conclusions

We have presented an LDF-HF implementation in which the sparsity of local orbitals as well as local fitting approximations are exploited to reach linear scaling in the calculation of the exchange matrix. For large molecules, this leads to speed-ups by typically an order of magnitude relative to canonical DF-HF calculations. The Fock matrix evaluation is well parallelized, but the overall parallel speed-up is limited by the remaining parts of the calculation, which become dominant when many processing cores are used. The method can also be used in DFT calculations with hybrid functionals.

It has been carefully investigated how the approximations made in the calculation of the Hartree-Fock orbitals affect subsequent electron correlation treatments. It was found that rather large domains are necessary to keep the errors of both the Hartree-Fock and correlation energies reasonably small. Default values for the two parameters that determine the domain sizes have been established, and with these the errors of PNO-LMP2 total energies are usually of the order of only 0.1 kJ mol^{-1}.

Despite these improvements, the LDF-HF calculations are still an order of magnitude slower than subsequent PNO-LMP2 calculations for extended 1-dimensional molecules, such as the glycine or alanine peptide chains. This is due to linear scaling of the PNO-LMP2 as well as better overall parallelization of the correlation calculation. However, for molecules with

a 2- or 3-dimensional structures, the HF and LMP2 times become more comparable. Also when explicitly correlated (F12) terms are included[21] or higher-order methods such as PNO-LCCD[20] are applied, the balance between the computational effort for the HF and correlation calculations becomes better.

In the LDF-HF calculation the exchange matrix is computed as a sum of orbital contributions, $K_{\mu\nu} = \sum_i^{\text{occ}} (\mu i|\nu i)$. The range of indices μ, ν for each i can be restricted to so-called AO domains, since the integrals $(\mu i|\nu i)$ decay exponentially with the distance between the localized orbital ϕ_i and the basis functions χ_μ or χ_ν. However, it has turned out that these domains have to be unexpectedly large in order to obtain accurate orbitals and energies. This also means that linear scaling is only obtained for rather large molecular sizes. But even without exploiting the sparsity of the LMO coefficient matrix and with full AO domains, significant savings are achieved by the local fitting. The scaling of the exchange matrix calculation is then between quadratic and cubic with molecular size, with a small prefactor for the cubic contribution.

In order to obtain accurate LDF-HF energies we recalculate the energy without any local approximations after convergence of the HF iterations. This improves the accuracy of the energy by several orders of magnitude.

A limitation of our current program is that the integral screening strongly suffers if diffuse basis functions are included. Further improvements are necessary to alleviate this problem, for example by treating diffuse functions in separate basis function blocks.

Acknowledgement

This work has been funded by the ERC Advanced Grant 320723 (ASES). We thank Gerald Knizia for providing his efficient routines for integral evaluation and IBO localization, which

have been essential for this work.

Supporting Information Available

Additional visualisations and geometries for molecules considered in this publication along with absolute LDF-HF total and PNO-LMP2 correlation energies can be found in the Supporting Information. Furthermore, a graph which shows the dependence of the accuracy on the fitting domain size and the LMO sparsity as well as a graph which exemplifies the cost and scaling to recalculate the LDF-HF energy is contained.

This material is available free of charge via the Internet at `http://pubs.acs.org/`.

References

(1) Schütz, M.; Hetzer, G.; Werner, H.-J. Low-order scaling local electron correlation methods. I. Linear scaling local MP2. *J. Chem. Phys.* **1999**, *111*, 5691–5705.

(2) Schütz, M. Low-order scaling local electron correlation methods. III. Linear scaling local perturbative triples correction (T). *J. Chem. Phys.* **2000**, *113*, 9986–10001.

(3) Hetzer, G.; Schütz, M.; Stoll, H.; Werner, H.-J. Low-order scaling local correlation methods II: Splitting the Coulomb operator in linear scaling local second-order Møller–Plesset perturbation theory. *J. Chem. Phys.* **2000**, *113*, 9443–9455.

(4) Schütz, M.; Werner, H.-J. Local perturbative triples correction (T) with linear cost scaling. *Chem. Phys. Lett.* **2000**, *318*, 370–378.

(5) Schütz, M.; Werner, H.-J. Low-order scaling local electron correlation methods. IV. Linear scaling local coupled-cluster (LCCSD). *J. Chem. Phys.* **2001**, *114*, 661–681.

(6) Schütz, M. A new, fast, semi-direct implementation of linear scaling local coupled cluster theory. *Phys. Chem. Chem. Phys.* **2002**, *4*, 3941–3947.

(7) Schütz, M. Low-order scaling local electron correlation methods. V. Connected triples beyond (T): Linear scaling local CCSDT-1b. *J. Chem. Phys.* **2002**, *116*, 8772–8785.

(8) Schütz, M.; Manby, F. R. Linear scaling local coupled cluster theory with density fitting. Part I: 4-external integrals. *Phys. Chem. Chem. Phys.* **2003**, *5*, 3349–3358.

(9) Adler, T. B.; Werner, H.-J.; Manby, F. R. Local explicitly correlated second-order perturbation theory for the accurate treatment of large molecules. *J. Chem. Phys.* **2009**, *130*, 054106.

(10) Werner, H.-J.; Schütz, M. An efficient local coupled cluster method for accurate thermochemistry of large systems. *J. Chem. Phys.* **2011**, *135*, 144116.

(11) Adler, T. B.; Werner, H.-J. An explicitly correlated local coupled cluster method for calculations of large molecules close to the basis set limit. *J. Chem. Phys.* **2011**, *135*, 144117.

(12) Kats, D. Speeding up local correlation methods. *J. Chem. Phys.* **2014**, *141*, 244101.

(13) Wennmohs, F.; Neese, F. A comparative study of single reference correlation methods of the coupled-pair type. *Chem. Phys.* **2008**, *343*, 217–230.

(14) Neese, F.; Wennmohs, F.; Hansen, A. Efficient and accurate local approximations to coupled-electron pair approaches: An attempt to revive the pair natural orbital method. *J. Chem. Phys.* **2009**, *130*, 114108.

(15) Neese, F.; Hansen, A.; Liakos, D. G. Efficient and accurate approximations to the local coupled cluster singles doubles method using a truncated pair natural orbital basis. *J. Chem. Phys.* **2009**, *131*, 064103.

(16) Hansen, A.; Liakos, D. G.; Neese, F. Efficient and accurate local single reference correlation methods for high-spin open-shell molecules using pair natural orbitals. *J. Chem. Phys.* **2011**, *135*, 214102.

(17) Riplinger, C.; Neese, F. An efficient and near linear scaling pair natural orbital based local coupled cluster method. *J. Chem. Phys.* **2013**, *138*, 034106.

(18) Werner, H.-J.; Knizia, G.; Krause, C.; Schwilk, M.; Dornbach, M. Scalable electron correlation methods I. Linear scaling local MP2-F12 using pair natural orbitals. *J. Chem. Theory Comput.* **2015**, *11*, 484–507.

(19) Köppl, C.; Werner, H.-J. On the use of Abelian point group symmetry in density-fitted local MP2 using various types of virtual orbitals. *J. Chem. Phys.* **2015**, *142*, 164108.

(20) Schwilk, M.; Usvyat, D.; Werner, H.-J. Communication: Improved pair approximations in local coupled-cluster methods. *J. Chem. Phys.* **2015**, *142*, 121102.

(21) Ma, Q.; Werner, H.-J. Scalable Electron Correlation Methods II.: Parallel PNO-LMP2-F12 with Near Linear Scaling in the Molecular Size. *J. Chem. Theory Comput.* **2015**, *11*, 5291–5304.

(22) Ochsenfeld, C.; White, C. A.; Head-Gordon, M. Linear and sublinear scaling formation of Hartree–Fock-type exchange matrices. *J. Chem. Phys.* **1998**, *109*, 1663–1669.

(23) Kussmann, J.; Beer, M.; Ochsenfeld, C. Linear-scaling self-consistent field methods for large molecules. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2013**, *3*, 614–636.

(24) Kussmann, J.; Ochsenfeld, C. Pre-selective screening for matrix elements in linear-scaling exact exchange calculations. *J. Chem. Phys.* **2013**, *138*, 134114.

(25) Schwegler, E.; Challacombe, M. Linear scaling computation of the Hartree–Fock exchange matrix. *J. Chem. Phys.* **1996**, *105*, 2726–2734.

(26) Schwegler, E.; Challacombe, M.; Head-Gordon, M. Linear scaling computation of the Fock matrix. II. Rigorous bounds on exchange integrals and incremental Fock build. *J. Chem. Phys.* **1997**, *106*, 9708–9717.

(27) Schwegler, E.; Challacombe, M. Linear scaling computation of the Fock matrix. IV. Multipole accelerated formation of the exchange matrix. *J. Chem. Phys.* **1999**, *111*, 6223–6229.

(28) Schwegler, E.; Challacombe, M. Linear scaling computation of the Fock matrix. III. Formation of the exchange matrix with permutational symmetry. *Theor. Chem. Acc.* **2000**, *104*, 344–349.

(29) Tymczak, C. J.; Weber, V. T.; Schwegler, E.; Challacombe, M. Linear scaling computation of the Fock matrix. VIII. Periodic boundaries for exact exchange at the ÎŞ point. *J. Chem. Phys.* **2005**, *122*, 124105.

(30) Gan, C. K.; Challacombe, M. Linear scaling computation of the Fock matrix. VI. Data parallel computation of the exchange-correlation matrix. *J. Chem. Phys.* **2003**, *118*, 9128–9135.

(31) Gan, C. K.; Tymczak, C. J.; Challacombe, M. Linear scaling computation of the Fock matrix. VII. Parallel computation of the Coulomb matrix. *J. Chem. Phys.* **2004**, *121*, 6608–6614.

(32) Rudberg, E.; Rubensson, E. H.; Sałek, P. Hartree–Fock calculations with linearly scaling memory usage. *J. Chem. Phys.* **2008**, *128*, 184106.

(33) Burant, J. C.; Scuseria, G. E.; Frisch, M. J. A linear scaling method for Hartree–Fock exchange calculations of large molecules. *J. Chem. Phys.* **1996**, *105*, 8969–8972.

(34) He, X.; Merz, K. M. Divide and Conquer Hartree-Fock Calculations on Proteins. *J. Chem. Theory Comput.* **2010**, *6*, 405–411.

(35) Van Alsenoy, C. Ab initio calculations on large molecules: The multiplicative integral approximation. *J. Comput. Chem.* **1988**, *9*, 620–626.

(36) Ten-no, S.; Iwata, S. Three-center expansion of electron repulsion integrals with linear combination of atomic electron distributions. *Chem. Phys. Lett.* **1995**, *240*, 578 – 584.

(37) Kendall, A. R.; Früchtl, A. H. The impact of the resolution of the identity approximate integral method on modern ab initio algorithm development. *Theor. Chem. Acc.* **1997**, *97*, 158–163.

(38) Früchtl, H. A.; Kendall, R. A.; Harrison, R. J.; Dyall, K. G. An Implementation of RI-SCF on Parallel Computers. *Int. J. Quantum Chem.* **1997**, *64*, 63–69.

(39) Weigend, F. A fully direct RI-HF algorithm: Implementation, optimised auxiliary basis sets, demonstration of accuracy and efficiency. *Phys. Chem. Chem. Phys.* **2002**, *4*, 4285–4291.

(40) Newton, M. D. Self-Consistent Molecular-Orbital Methods. II. Projection of Diatomic Differential Overlap (PDDO). *J. Chem. Phys.* **1969**, *51*, 3917–3926.

(41) Billingsley, F. P.; Bloor, J. E. Limited Expansion of Diatomic Overlap (LEDO): A Near-Accurate Approximate Ab Initio LCAO MO Method. I. Theory and Preliminary Investigations. *J. Chem. Phys.* **1971**, *55*, 5178–5190.

(42) Beebe, N. H. F.; Linderberg, J. Simplifications in the generation and transformation of two-electron integrals in molecular calculations. *Int. J. Quantum Chem.* **1977**, *12*, 683–705.

(43) O'neal, D. W.; Simons, J. Application of cholesky-like matrix decomposition methods to the evaluation of atomic orbital integrals and integral derivatives. *Int. J. Quantum Chem.* **1989**, *36*, 673–688.

(44) Izsák, R.; Neese, F. An overlap fitted chain of spheres exchange method. *J. Chem. Phys.* **2011**, *135*.

(45) Izsák, R.; Neese, F.; Klopper, W. Robust fitting techniques in the chain of spheres approximation to the Fock exchange: The role of the complementary space. *J. Chem. Phys.* **2013**, *139*.

(46) Werner, H.-J.; Manby, F. R.; Knowles, P. J. Fast linear scaling second-order Møller-Plesset perturbation theory (MP2) using local and density fitting approximations. *J. Chem. Phys.* **2003**, *118*, 8149–8160.

(47) Baerends, E. J.; Ellis, D. E.; Ros, P. Self-consistent molecular Hartree-Fock-Slater calculations. I. The computational procedure. *Chem. Phys.* **1973**, *2*, 41.

(48) Whitten, J. L. Coulombic potential energy integrals and approximations. *J. Chem. Phys.* **1973**, *58*, 4496.

(49) Dunlap, B. I.; Connolly, J. W. D.; Sabin, J. R. On the applicability of LCAO-Xα methods to molecules containing transition metal atoms: The nickel atom and nickel hydride. *Int. J. Quantum Chem.* **1977**, *12*, 81–87.

(50) Dunlap, B. I.; Connolly, J. W. D.; Sabin, J. R. On some approximations in applications of $X\alpha$ theory. *J. Chem. Phys.* **1979**, *71*, 3396.

(51) Dunlap, B. I.; Connolly, J. W. D.; Sabin, J. R. On first-row diatomic molecules and local density models. *J. Chem. Phys.* **1979**, *71*, 4993.

(52) Mintmire, J. W.; Dunlap, B. I. Fitting the Coulomb potential variationally in LCAO density-functional calculations. *Chem. Phys. Lett.* **1982**, *25*, 88.

(53) Vahtras, O.; Almlof, J.; Feyereisen, M. Integral approximations for LCAO-SCF calculations. *Chem. Phys. Lett.* **1993**, *213*, 514–518.

(54) Dunlap, B. I. Robust and variational fitting. *Phys. Chem. Chem. Phys.* **2000**, *2*, 2113–2116.

(55) Weigend, F. A fully direct RI-HF algorithm: Implementation, optimised auxiliary basis sets, demonstration of accuracy and efficiency. *Phys. Chem. Chem. Phys.* **2002**, *4*, 4285.

(56) Polly, R.; Werner, H.-J.; Manby, F. R.; Knowles, P. J. Fast Hartree–Fock theory using local density fitting approximations. *Mol. Phys.* **2004**, *102*, 2311–2321.

(57) Mejía-Rodríguez, D.; Köster, A. M. Robust and efficient variational fitting of Fock exchange. *J. Chem. Phys.* **2014**, *141*, 124114.

(58) Knizia, G. Intrinsic Atomic Orbitals: An Unbiased Bridge between Quantum Theory and Chemical Concepts. *J. Chem. Theory Comput.* **2013**, *9*, 4834–4843.

(59) Nieplocha, J.; Palmer, B.; Tipparaju, V.; Krishnan, M.; Trease, H.; Aprà, E. Advances, Applications and Performance of the Global Arrays Shared Memory Programming Toolkit. *Int. J. High Perform. Comput. Appl.* **2006**, *20*, 203–231.

(60) Krishnan, M.; Palmer, B.; Vishnu, A.; Krishnamoorthy, S.; Daily, J.; Chavarria, D. The Global Arrays User Manual. 2012.

(61) Boys, S. F. Construction of Some Molecular Orbitals to Be Approximately Invariant for Changes from One Molecule to Another. *Rev. Mod. Phys.* **1960**, *32*, 296–299.

(62) Foster, J. M.; Boys, S. F. Canonical Configurational Interaction Procedure. *Rev. Mod. Phys.* **1960**, *32*, 300–302.

(63) Edmiston, C.; Ruedenberg, K. Localized Atomic and Molecular Orbitals. *Rev. Mod. Phys.* **1963**, *35*, 457–464.

(64) Pipek, J.; Mezey, P. G. A fast intrinsic localization procedure applicable for ab initio and semiempirical linear combination of atomic orbital wave functions. *J. Chem. Phys.* **1989**, *90*, 4916–4926.

(65) Boughton, J. W.; Pulay, P. Comparison of the boys and Pipek–Mezey localizations in the local correlation approach and automatic virtual basis selection. *J. Comput. Chem.* **1993**, *14*, 736–740.

(66) Pipek, J. Localization measure and maximum delocalization in molecular systems. *Int. J. Quantum Chem.* **1989**, *36*, 487–501.

(67) Werner, H.-J.; Knizia, G.; Manby, F. R. Explicitly correlated coupled cluster methods with pair-specific geminals. *Mol. Phys.* **2011**, *109*, 407–417.

(68) Werner, H.-J.; Knowles, P. J.; Knizia, G.; Manby, F. R.; Schütz, M. Molpro: a general-purpose quantum chemistry program package. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2012**, *2*, 242–253.

(69) Schütz, M.; Werner, H.-J.; Lindh, R.; Manby, F. R. Analytical energy gradients for local second-order Møller-Plesset perturbation theory using density fitting approximations. *J. Chem. Phys.* **2004**, *121*, 737–750.

(70) Li, W.; Piecuch, P.; Gour, J. R.; Li, S. Local correlation calculations using standard and renormalized coupled-cluster approaches. *J. Chem. Phys.* **2009**, *131*, 114109.

(71) Stewart, J. J. P.; Császár, P.; Pulay, P. Fast Semiempirical Calculations. *J. Comp. Chem.* **1982**, *3*, 556.

(72) Daniels, A. D.; Scuseria, G. E. What is the best alternative to diagonalization of the Hamiltonian in large scale semiempirical calculations? *J. Chem. Phys.* **1999**, *110*, 1321–1328.

(73) Werner, H.-J.; Knowles, P. J.; Knizia, G.; Manby, F. R.; Schütz, M.; Celani, P.; Korona, T.; Lindh, R.; Mitrushenkov, A.; Rauhut, G.; Shamasundar, K. R.; Adler, T. B.; Amos, R. D.; Bernhardsson, A.; Berning, A.; Cooper, D. L.; Deegan, M. J. O.; Dobbyn, A. J.; Eckert, F.; Goll, E.; Hampel, C.; Hesselmann, A.; Hetzer, G.; Hrenar, T.;

Jansen, G.; Köppl, C.; Liu, Y.; Lloyd, A. W.; Mata, R. A.; May, A. J.; McNicholas, S. J.; Meyer, W.; Mura, M. E.; Nicklass, A.; O'Neill, D. P.; Palmieri, P.; Peng, D.; Pflüger, K.; Pitzer, R.; Reiher, M.; Shiozaki, T.; Stoll, H.; Stone, A. J.; Tarroni, R.; Thorsteinsson, T.; Wang, M. MOLPRO, development version 2015, a package of ab initio programs. 2015.

(74) Thom H. Dunning, J. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007–1023.

(75) Kendall, R. A.; Thom H. Dunning, J.; Harrison, R. J. Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions. *J. Chem. Phys.* **1992**, *96*, 6796–6806.

(76) Figgen, D.; Rauhut, G.; Dolg, M.; Stoll, H. Energy-consistent pseudopotentials for group 11 and 12 atoms: adjustment to multi-configuration Dirac–Hartree–Fock data. *Chem. Phys.* **2005**, *311*, 227 – 244.

(77) Zuend, S. J.; Jacobsen, E. N. Mechanism of Amido-Thiourea Catalyzed Enantioselective Imine Hydrocyanation: Transition State Stabilization via Multiple Non-Covalent Interactions. *J. Am. Chem. Soc.* **2009**, *131*, 15358–15374.

(78) Fedorov, A.; Batiste, L.; Couzijn, E. P. A.; Chen, P. Experimental and Theoretical Study of a Gold(I) Aminonitrene Complex in the Gas Phase. *ChemPhysChem* **2010**, *11*, 1002–1005.