

Hybrid DFT quality thermochemistry and environment effects at GGA cost via local quantum embedding

József Csóka,^{*,†,‡,¶} Dénes Berta,^{†,‡,¶} and Péter R. Nagy^{*,†,‡,¶}

[†]*Department of Physical Chemistry and Materials Science, Faculty of Chemical Technology and Biotechnology, Budapest University of Technology and Economics, Műegyetem rkp. 3., H-1111 Budapest, Hungary*

[‡]*HUN-REN-BME Quantum Chemistry Research Group, Műegyetem rkp. 3., H-1111 Budapest, Hungary*

[¶]*MTA-BME Lendület Quantum Chemistry Research Group, Műegyetem rkp. 3., H-1111 Budapest, Hungary*

E-mail: csoka.jozsef@vbk.bme.hu; nagy.peter@vbk.bme.hu

Abstract

Reliable thermochemical modeling of reaction mechanisms requires hybrid DFT or higher-level models as well as inclusion of environment, conformer, thermal, etc. effects. Quantum embedding, such as the Huzinaga-equation and projection-based models employed here, can make such computations more accessible by focusing the use of the more costly models to the atoms involved in forming and breaking the bonds or residing in interacting surfaces, etc. Here, we further accelerate these embedding computations by combining local approximations in the atomic orbital and auxiliary function space of the hybrid DFT part with a new in-core density fitting implementation optimized for multi-layer DFT. The so introduced local embedded subsystem (LESS) framework, when increasing the size of the environment, leads to asymptotically constant cost for the hybrid DFT layer. We demonstrate on reaction and activation energies of practical homogeneous, heterogeneous and enzymatic catalysis reactions that the intrinsic accuracy of hybrid DFT is retained, with a few tenths of a kcal/mol error including all (embedding and local) approximations. Compared to the same complete (density fitted) hybrid DFT reference, the LESS hybrid DFT-in-GGA runtimes are 30–90 times faster on systems with up to 171–238 atoms. Achieving energetics with practically hybrid DFT quality and GGA cost is a significant step toward predictive thermochemistry including reliable sampling, dynamics, etc. as well as quantum environment effects.

1 Introduction

Electronic structure calculations are required and thus routinely employed to model a large variety of chemical properties, including molecular interactions, catalytic reaction mechanisms, light-matter interactions, etc. However, accurate simulations incorporating, e.g., thermochemical and environment effects in homogeneous, heterogeneous or enzyme reactions require the quantum mechanical (QM) treatment of up to a few hundred atoms. This poses a limitation on our most accurate electronic structure methods suitable for reactivity modeling, especially when the exploration of the conformational space or dynamic properties is also of interest. Moreover, chemically accurate modeling of thermodynamic and especially kinetic properties often requires models beyond the capabilities of pure density functional theory (DFT) models. Although efficient local correlation based wave function approaches, such as the local natural orbital (LNO) based and other coupled cluster (CC) methods can now routinely scale up to a few hundred atoms,¹⁻³ their application is at the moment limited to single point energy calculations. Hence, DFT remains the dominating electronic structure method due to its affordable computational cost and the availability of a wide range of properties beyond energetics. However, the accurate modeling of chemical reactions often require hybrid (or higher rung) functionals and sufficiently large basis sets.^{4,5} These methods have a computational cost that often necessitates compromises on the size of the chemical model system, the treatment of its environment or the quality of thermochemical corrections.

Multi-level modeling techniques are popular tools to reduce the computational cost by combining different theoretical methods for the subsystem in focus and its environment. For example, QM/MM⁶⁻⁸ methods have a pivotal role in describing chemical processes in biochemical,^{9,10} material,^{11,12} surface,¹³ cavity,¹⁴ explicit solvent,¹⁵ etc. environments. To improve the quality of the environment model or to decrease the size of the high-level QM model, multi-level QM-QM embedding methods are also available. These algorithms divide the QM system into two or more subsystems, which are modeled with different QM models, such as via different rungs of DFT functionals, semi-empirical QM models or even wave

function approaches. Typically, a more accurate level of theory is used for the chemically active region (describing bonds breaking and forming) and a faster method is applied for the environment regions (e.g. solvent, protein structure, surface of solids).

The family of QM-QM methods include ONIOM,^{16–18} density matrix embedding,¹⁹ embedded mean-field theory,^{20,21} frozen density embedding,^{22–24} subsystem DFT,^{25–27} fragmentation methods,^{28–32} orthogonality constrained basis set expansion,³³ wave function-in-wave function embedding,^{1,34–38} reduced density matrix embedding.³⁹ Here, we focus on the projection-based^{40–44} and the numerically very similar, but formally exact Huzinaga-equation-based embedding methods.^{43,45,46}

The Huzinaga and projector embedding schemes divide the system into two parts, an active region and an environment. The active region is typically described with a (double) hybrid DFT functional or a wave function method. In both cases, a computationally demanding self-consistent (SCF) field calculation [Hartree–Fock (HF) or hybrid DFT] is performed on the active subsystem. For hybrid DFT level high-level calculations, the performance bottleneck is the exact exchange term in the Fock (or Kohn–Sham) matrix, and thus our goal here is to accelerate its computation.

In four-center integral based hybrid DFT (or HF) algorithms, the Coulomb and exchange matrices are built using the density matrix expanded in atomic orbital (AO) basis. In their original form, projection and Huzinaga based DFT-in-DFT embedding schemes do not reduce the number of AO functions in the high-level calculation, thus, the high-level method remains fourth-order scaling with the size of the complete QM system. One direction of improvement on the high-level execution times could be the use of accelerated HF and hybrid DFT approaches. For example, advanced screening^{47–49} and incremental SCF^{50–52} methods can greatly reduce the number of integral evaluations and contractions. Alternatively, seminumerical methods perform one of the integration numerically and the other analytically to compute the four-center two-electron integrals, including the pseudospectral method,^{53–57} chain of spheres exchange,^{58–61} and other low-scaling seminumerical^{62,63} approaches. More-

over, the auxiliary density matrix method^{64,65} and tensor factorization ideas,^{66,67} including tensor hypercontraction^{68,69} and the multipole approaches⁷⁰⁻⁷⁶ are also effective. Here, we employ density fitting (DF),⁷⁷⁻⁸² which can be used to expand the AO product densities on an auxiliary (or fitting) basis. Although DF does not change the theoretical scaling of the SCF algorithm (with exact exchange), in practice, it can be significantly faster than four-center algorithms, especially with reliable AO basis sets. The DF scheme can be made even more efficient, for example, via (pair) atomic resolution of identity (RI) approaches,^{83,84} occupied-orbital RI,⁸⁵ or local DF methods,⁸⁶⁻⁸⁹ the latter employed also in the present work.

Although the above methods are efficient and generally applicable to speed up SCF calculations, they are not yet combined with the ideas exploited in embedding methods, such as partitioning to low- and high-level subsystems. The first step along these lines within the projection-based and Huzinaga embedding schemes was the development of approximations to decrease the size of the AO basis. A dual basis set approach was combined with embedding by Hégyely and co-workers,⁴⁵ an AO basis set reduction scheme was published by Barnes and co-workers⁹⁰ and was later further developed by Bennie *et. al.*⁹¹ and Bensberg and Neugebauer.⁹² In AO basis reduction, one exploits that only a subset of the AOs are required for the accurate expansion of the active MOs, as shown also in various chemical applications.⁹³⁻⁹⁶ This reduces the scaling of the high-level calculations, which, at least for four-center integral based methods, could become (asymptotically) independent of the size of the full system.

Our aim is to extend the use of reduced-cost and embedding approaches beyond four-center algorithms, i.e., in combination with DF. Already without additional approximations, DF-SCF benefits from embedding as the cost of the exchange matrix computation is proportional to the number of active MOs, thereby reducing the overall scaling to third order. Furthermore, here we introduce the new local embedded subsystem (LESS) framework, employing AO and new DF auxiliary basis set reduction approaches for the active subsystem. First, extending local AO reduction to DF-SCF decrease its scaling already to quadratic, due

to the quadratic scaling of some of the rate-determining steps of DF-SCF with the number of DF auxiliary functions. Second, we generalize this local AO approximation to DF auxiliary basis reduction in the high-level calculation. Combining these two developments, the computer time needed for the high-level DF-SCF calculation does not depend on the size of the full system in the limit of much larger environments.

An additional advantage of the new LESS method is that the combined reduction on both the AO and DF auxiliary bases makes it possible to store the three-center DF electron-repulsion integrals (ERI) in the primary memory, even for hundreds of atoms using at least triple- ζ quality basis sets. Note that, storing the integrals in standard four-center and DF algorithms would not be possible due to the size of the ERI tensor, highlighting the benefits of combining LESS and DF-SCF. To utilize this, we also developed an efficient in-core DF-SCF algorithm, which reduces the runtime by a factor of 4–5 compared to the integral-direct implementation. The combined approximations still have low errors in the few tenths of a kcal/mol range, which does not alter the intrinsic accuracy of hybrid DFT methods. However, the DFT-in-DFT computations using the LESS, DF-based and in-core developments lead to only about twice as long computation times as required by pure, DF accelerated GGA rung DFT models. The achieved accuracy and efficiency properties are demonstrated on real-world applications, including biochemical, organo- and surface catalytic reactions with realistic environments of about 200 QM atoms. Reaching DF-GGA level speed represents a significant step toward efficient conformational sampling and dynamics studies in accurate reactivity predictions, including reliable environment models.

In this article, in section 2, we start by presenting the theory of the Huzinaga and projection-based embedding methods as well as the new algorithms for the atomic and auxiliary basis reduction. Section 3 details the representative chemical reactions on which the methods accuracy (section 4) and computational affordability (section 5) are demonstrated. Finally, in section 6, we present discussion and general advice for applying the LESS framework in practical computational chemistry.

2 Theory

2.1 Huzinaga- and projection-based embedding

In this section, we give an overview of the Huzinaga-equation-based⁴³ and the projector embedding⁴¹ methods. Both methods determine the low- and high-level models and a set of atoms constituting the active region similarly. In the first step, the low-level calculation is performed for the whole QM system. Next, the low-level occupied molecular orbitals (MOs) are localized. The MOs residing on the subsystem atoms are selected as active MOs and the rest constitute the environment orbitals. The MO selection can be performed in several ways, for example, by using the Mulliken population based^{41,43} or the SPADE algorithm.^{97,98} In the last step, the high-level MOs are optimized with the high-level SCF method, while the environment MOs are kept frozen. In order to avoid the use of non-additive kinetic energy potential, the high-level MOs are made to be orthogonal to the environment orbitals. This can be achieved by a level-shift operator⁴¹ as used in projector embedding or by the adoption of the Huzinaga equations,^{43,99} which we used in this work. In brief, the Huzinaga equations result from the optimization of a single determinant with the side conditions enforcing the orthogonality of the active MOs to the environment MOs. Note that wave function method-based correlation calculation can also be executed for the active subsystem (after a Hartree–Fock calculation) and even multiple levels of LNO CC layers can be embedded into DFT environments.^{43,45}

The total energy for DFT-in-DFT embedding is written as

$$E = E_{\text{LL}}[\mathbf{D}^{AB}] - E_{\text{LL}}[\mathbf{D}^A] + E_{\text{HL}}[\tilde{\mathbf{D}}^A] + \text{Tr} \left\{ (\tilde{\mathbf{D}}^A - \mathbf{D}^A) \frac{\partial E_{\text{LL-HL}}}{\partial \mathbf{D}^A} \right\}, \quad (1)$$

where LL and HL indices are used for the low- and high-level calculations, respectively, while AB , A and B label the full, active and environment systems, respectively. Moreover, $\tilde{\mathbf{D}}^A$

denotes the subsystem density calculated with the high-level method and

$$E_{\text{LL-HL}} = E_{\text{LL}}[\mathbf{D}^{AB}] - E_{\text{LL}}[\mathbf{D}^A] + E_{\text{HL}}[\tilde{\mathbf{D}}^A] . \quad (2)$$

The orthogonality of the high-level subsystem and environment MOs can be enforced by either a level-shift operator in projector embedding or the Huzinaga equations. In the case of projector embedding, a projector of

$$P_{\mu\nu} = \sum_{\rho\sigma} S_{\mu\rho} D_{\rho\sigma}^B S_{\sigma\nu} \quad (3)$$

is added to the Fock or Kohn–Sham matrix as

$$\tilde{F}_{\mu\nu} + \alpha P_{\mu\nu} . \quad (4)$$

Here \mathbf{S} is the AO overlap, α is the level-shift and the high-level Fock matrix is defined as

$$\tilde{F}_{\mu\nu} = \frac{\partial E}{\partial \tilde{D}_{\mu\nu}^A} . \quad (5)$$

In the $\alpha \rightarrow \infty$ limit, the projector based embedding becomes exact. In practice, only finite α values can be used, which have an often negligible effect on the total energy but it may be more difficult to handle for other properties.

The use of the finite level-shift can be avoided by adding orthogonality constraints to the Lagrangian of the high-level SCF method. The effective Kohn–Sham matrix of the high-level SCF (called the Huzinaga matrix) can be derived from this Lagrangian^{99,100} and reads as

$$H_{\mu\nu} = \tilde{F}_{\mu\nu} - \sum_{\rho\sigma} S_{\mu\rho} D_{\rho\sigma}^B \tilde{F}_{\sigma\nu} - \sum_{\rho\sigma} \tilde{F}_{\mu\rho} D_{\rho\sigma}^B S_{\sigma\nu} . \quad (6)$$

The SCF procedure is performed as usual with \mathbf{H} at the place of the Fock or Kohn–Sham matrix.

In general, the high-level SCF calculation is the rate determining step during the energy evaluation due to the two-electron part (especially the exact exchange contribution) of the Kohn–Sham matrix. With density fitting, the two-electron contribution reads as

$$G_{\mu\nu}[\tilde{\mathbf{D}}^A] = \sum_{\sigma\lambda PQ} \tilde{D}_{\lambda\sigma}^A(\mu\nu|P)(P|Q)^{-1}(Q|\lambda\sigma) - c_{\text{HFx}} \sum_{i\sigma\lambda PQ} \tilde{C}_{\sigma i}^A \tilde{C}_{\lambda i}^A(\mu\lambda|P)(P|Q)^{-1}(Q|\sigma\nu), \quad (7)$$

where the Greek letters denote the AOs, P, Q are fitting basis functions, c_{HFx} is the weight of the Hartree–Fock exchange contribution and $\tilde{\mathbf{C}}^A$ is the high-level MO coefficient matrix. The construction of the exchange matrix [second term in Eq. (7)] takes advantage of the embedding approximation as the MO index i is restricted to the active orbitals, and thus its overall scaling is reduced to third power. Further savings can be achieved by introducing local approximations to reduce the AO basis, which improves the scaling to second-order by limiting the AO indices ($\mu, \nu, \lambda, \sigma$) in Eq. (7). In addition, in the LESS framework we introduce new algorithms, which retain the most important auxiliary functions and thus restrict the summation over the fitting functions (P, Q). This way, the operation count of the Coulomb and exchange matrix formation becomes asymptotically independent of the full system size, and depends only on the size of the active subsystem.

2.2 Reduction of the atomic orbital basis

We can exploit that only a small set of AOs is significant for the expansion of the active MOs.⁹⁰ The selection of the important functions is based on the net Mulliken population⁹¹

$$q_{\mu} = \sum_i C_{\mu i}^A C_{\mu i}^A S_{\mu\mu}, \quad (8)$$

where \mathbf{C}^A is the low-level active subsystem MO coefficients and \mathbf{S} is the AO overlap. The basis function μ is kept in the high-level calculation if $q_{\mu} > \varepsilon_{\text{AO}}$, where ε_{AO} is a predefined threshold. It is important to note that if a single function is retained from a given shell then

the other functions of the shell are also kept.

The low-level SCF is solved in the complete AO basis. If any AOs are dropped after this point, the AO expansion of the low-level MOs is also truncated, but only when they are used to build the embedding potential or projector in Eqs. 6 and 3, respectively. After the AO basis is reduced, the orthogonality of these low-level MOs are lost. To maintain consistency in the high-level calculation, both the environment and the active low-level orbitals are (re-)orthogonalized in the reduced AO metric (separately from each other), using Löwdin canonical orthogonalization. Namely, the overlap of the MOs in the reduced basis is diagonalized and the MOs are transformed with the eigenvectors. As the retained AOs are concentrated around the active subsystem, some of the environment MOs far from the active atoms may not be accurately expanded in the reduced AO basis. These MOs can be detected via their small MO overlap eigenvalues and then they are removed from the space of the environment orbitals. Next, to restore the orthogonality of the environment and all active MOs, the subspace of the retained environment MOs is projected out from all active orbitals.

2.3 Reduction of the DF auxiliary function basis

Here, we propose three alternative methods for the reduction of the DF auxiliary function basis for the high-level calculation. First, a straightforward way to construct a reduced fitting basis set is the use of natural auxiliary functions (NAFs).¹⁰¹ The three-center integrals with NAFs in the place of the auxiliary functions is the best approximation to the original integral tensor in the least-squares sense.¹⁰¹ The NAF approach uses the singular value decomposition to construct this compressed representation. When Huzinaga embedding is combined with NAFs, first, the active low-level occupied MOs (\mathbf{C}^A) are used to build the

$$J_{\mu i}^P = \sum_{Q\nu} C_{\nu i}^A(\mu\nu|Q)L_{QP} \tag{9}$$

half-transformed integrals, where \mathbf{L} is the Cholesky factor of the inverse of the two-center Coulomb integrals. The right singular vectors (\mathbf{N}) of \mathbf{J} define the NAFs (optimized for the exchange matrix evaluation). In practice, the singular vectors and values are calculated as the eigendecomposition of $W_{PQ} = \sum_{\mu i} J_{\mu i}^P J_{\mu i}^Q$. The compressed integrals can be expressed as

$$J_{\mu\nu}^{\bar{P}} = \sum_{QR} (\mu\nu|Q) L_{QR} N_{R\bar{P}} , \quad (10)$$

where \bar{P} labels the retained NAFs corresponding to singular values larger than a threshold ε_{NAF} .

In theory, the matrix \mathbf{W} could be built from the AO integrals, however, it would result in a larger NAF space.¹⁰¹ The NAFs from the \mathbf{C}^A -dependent integrals are a good choice, because \mathbf{C}^A and $\tilde{\mathbf{C}}^A$ can be expected to span a similar space. Alternatively, the NAFs could be determined from the high-level MOs ($\tilde{\mathbf{C}}^A$) in each iteration of the high-level SCF, however, the overhead introduced by Eqs. (9) and (10) would be much larger than the performance gains. Consequently, the NAF method is not as well suited for integral-direct SCF algorithms either. Although the NAF approach yields the most compact compression of the auxiliary space, a potential drawback is that the calculation of the analytic gradients is also more complicated with the NAF approach.¹⁰²

In our second approach, Mulliken-like charges were defined based on the analogy to the AO reduction. To define these generalized Mulliken charges, we exploit the analogy, that the NAFs are the system specifically optimal linear combination of the fitting functions, just like the MOs are the linear combination of the AOs. Thus, inspired by Eq. (8), the generalized Mulliken charges are defined as

$$\hat{q}_P = \sum_{\bar{Q}} N_{P\bar{Q}} N_{P\bar{Q}} S_{PP}^{\text{DF}} , \quad (11)$$

where \mathbf{S}^{DF} is the overlap of the auxiliary functions and \hat{q}_P is the generalized Mulliken charge

of the auxiliary function P . The fitting functions with \hat{q} values above a threshold $\varepsilon_{\text{Mull}}$ are retained, while the rest of the fitting functions are not used in the high-level computation. Similarly to the AO basis reduction, if an auxiliary function is kept in a shell then none of the shell's functions are dropped. The advantage of this method is that it can be used together with integral-direct algorithms as the auxiliary functions are selected before the high-level calculation and the transformation step in Eq. (10) is not required. However, the computational overhead of the calculation and diagonalization of \mathbf{W} is still present, even though it has to be performed only once.

In our third approach, a reduced fitting basis set is defined through the generalization of the local DF (LDF) domains used in LDF-SCF calculations.^{86,87,103} In the original LDF algorithm, a set of atoms (a domain) is selected specifically for each localized MO, and only the auxiliary functions centered on the selected atoms are used to fit the (μi) -type orbital products in the exchange term, that is

$$\sum_i \sum_{P,Q \in [i]} (\mu i | P)(P | Q)^{-1}(Q | \nu i) , \quad (12)$$

where $[i]$ is the domain of the i th MO and $P \in [i]$ denotes that the summation runs over the auxiliary functions centered on the atoms in $[i]$. Our LDF-SCF algorithm^{75,87} could be used without any modification, however, it would not reduce the fitting basis set for the Coulomb term. Additionally, it would complicate the development of analytic derivatives⁸⁹ as the LDF-SCF is not exactly variational due to the MO dependence of the second summation in Eq. (12).

To overcome these issues, we employ the fitting functions in the union of the domains to fit both the exchange and Coulomb terms. This way, the variational property of the SCF method is retained. Moreover, to increase the efficiency of our previous domain construction scheme,⁷⁵ the domains are built from shells rather than atoms. First, Löwdin atomic charges are computed for each MO and atom. If the charge of an MO exceeds 0.05 on a given atom,

then all shells of this atom are added to the so called primary domain of the MO. This primary domain is then extended based on the Cauchy–Schwarz screening. We retain the fitting functions P centered on atom B with angular momentum ℓ if

$$\max_{\mu \in A, \nu \in B} \sqrt{(\mu\nu|\mu\nu)} \max_{P \in B_\ell} \sqrt{(P|P)} > \varepsilon_{\text{dom}} \quad (13)$$

holds for any atom A from the primary domain, where μ and ν are AOs on atoms A and B and ε_{dom} is a predefined threshold.

The LDF domain based fitting reduction is well suited for both integral-direct, disk-based and in-core SCF algorithms as the set of important fitting functions is built prior to the high-level algorithm using the low-level MO orbitals. Although not as compact as NAFs, this LDF based approximation is fast and introduces negligible overhead.

2.4 In-core algorithm

It is well-known that for large molecules and basis sets, it is not possible to store the three-center Coulomb integrals in the primary memory or even in the secondary memory (i.e., on disk). In such cases integral-direct approaches are used, which do not store but rather re-evaluate all required integrals in each SCF iteration. Our integral-direct DF-SCF implementation was presented in detail previously.⁷⁵ However, by combining embedding methods with basis reduction and efficient screening, we can drastically reduce the number of integrals, and it becomes possible to store the integrals needed for the high-level computation in primary memory for practical applications.

In more details, let us use the notation

$$J_{\mu\nu}^P = \sum_Q (\mu\nu|Q) L_{QP} \quad (14)$$

for the three-center Coulomb integrals. The Coulomb matrix \mathbf{F}^J is calculated in two steps

$$X_P = \sum_{\mu\nu} D_{\mu\nu} J_{\mu\nu}^P \quad \text{and} \quad F_{\mu\nu}^J = \sum_P X_P J_{\mu\nu}^P \quad (15)$$

and similarly, the exchange matrix \mathbf{F}^K is built as

$$J_{\mu i}^P = \sum_{\nu} C_{\nu i} J_{\mu\nu}^P \quad \text{and} \quad F_{\mu\nu}^K = -c_{\text{HFx}} \sum_{iP} J_{\mu i}^P J_{\nu i}^P. \quad (16)$$

We optimize for the evaluation of the exchange matrix as this is the rate determining step. Therefore, the $J_{\mu\nu}^P$ integrals are stored as follows: for a given shell S and AO function ν , the quantity

$$J_{\nu}^S = \max_{\mu \in S} \max_P J_{\mu\nu}^P \quad (17)$$

is calculated and if it is below threshold, the integrals are dropped. Otherwise, the integrals are stored in such an order that P is the fastest index, ν is the slowest and μ is in the middle restricted to S . Additionally, a mapping is stored which keeps a list of ν - S pairs for which J_{ν}^S of Eq. (17) is above the screening threshold. The Coulomb and exchange matrices are evaluated shellwise, that is, the integrals for a given shell S are retrieved (either from primary memory or disk), the density and the MO coefficients are sorted according to the mapping and the first steps in Eqs. (15) and (16) are performed. Thus, the repeated recalculation of the integrals are avoided, moreover, the storage scheme makes it possible to use a single matrix-matrix and matrix-vector multiplication to get the $J_{\mu i}^P$ and X_P intermediates, respectively. Next, the second steps in Eqs. (15) and (16) are executed and the final result is sorted to the Fock or Kohn–Sham matrix. The implementation of the above in-core algorithm exploit the parallel processing power of modern CPUs by using shared memory parallel, optimized linear algebra subroutines and it does not rely on explicit parallelization.

We note, that the main storage space savings, which make the in-core approach feasible, come from the AO and fitting basis reduction. The screening described above screens out

a relatively small number of integrals (less than 30%) when combined with the basis set reduction schemes in LESS for typical systems, like those with up to 100-200 atoms considered here.

3 Chemical applications

To assess the performance of our approximations, demonstrative reactions were investigated that are typically used to benchmark embedding methods:^{20,43,45,92,104,105} the deprotonation of decanoic acid, the hydrogenation of pentacene, the Diels–Alder reaction of octadecanonaene and 1,3-butadiene, the substitution reaction of 1-chlorodecane with a hydroxide anion to form 1-decanol and a chloride anion. The reactions are depicted in Fig. 1.

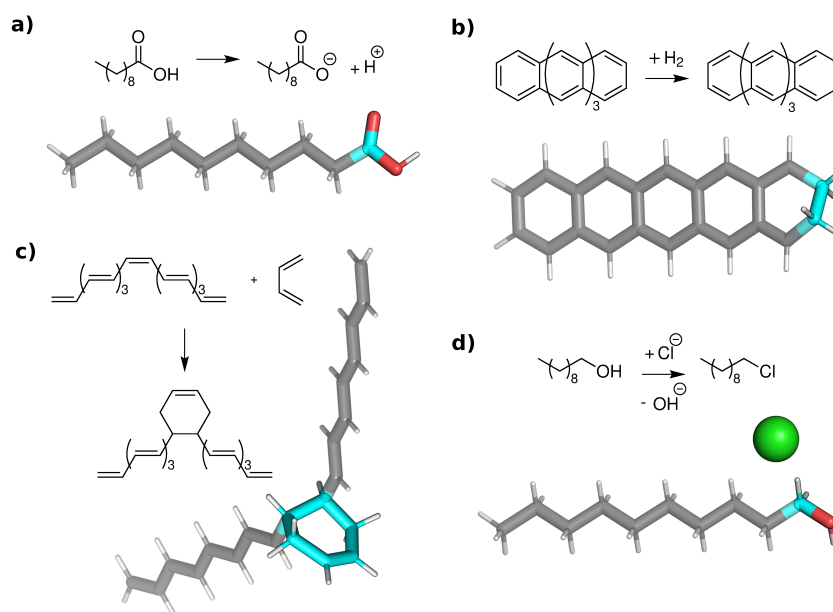


Figure 1: Reactions in standard tests. The active atoms are highlighted with color and can be found in the SI as well. a) the deprotonation of decanoic acid b) hydrogenation of pentacene c) Diels–Alder reaction of octadecanonaene and 1,3-butadiene d) substitution reaction of 1-chlorodecane with a hydroxide anion to form 1-decanol and a chloride anion.

Next, three larger test systems of 171–238 atoms were selected (see Fig. 2). The first one is a zeolite catalyzed methylation reaction of propylene.^{100,106} It is known, that the reaction energy and the barrier are sensitive to the applied DFT functional and hybrid functionals

are considerably more accurate.¹⁰⁶

Second, the GTP hydrolysis catalyzed by the KRas small GTPase activated by the p120-GAP is a typical example of phosphate biochemistry.¹⁰⁷ The accurate description of the S_N2 -type phosphate cleavage requires hybrid DFT description,¹⁰⁸ while calculated properties are notoriously slow to converge with the size of the QM region.^{109,110} Finally, the organocatalytic chlorolactonization involving a cationic chlorine transfer¹¹¹ poses further challenge to GGA methods,¹¹² while the large organocatalyst required for stereoselectivity hinders the use of more expensive methods.

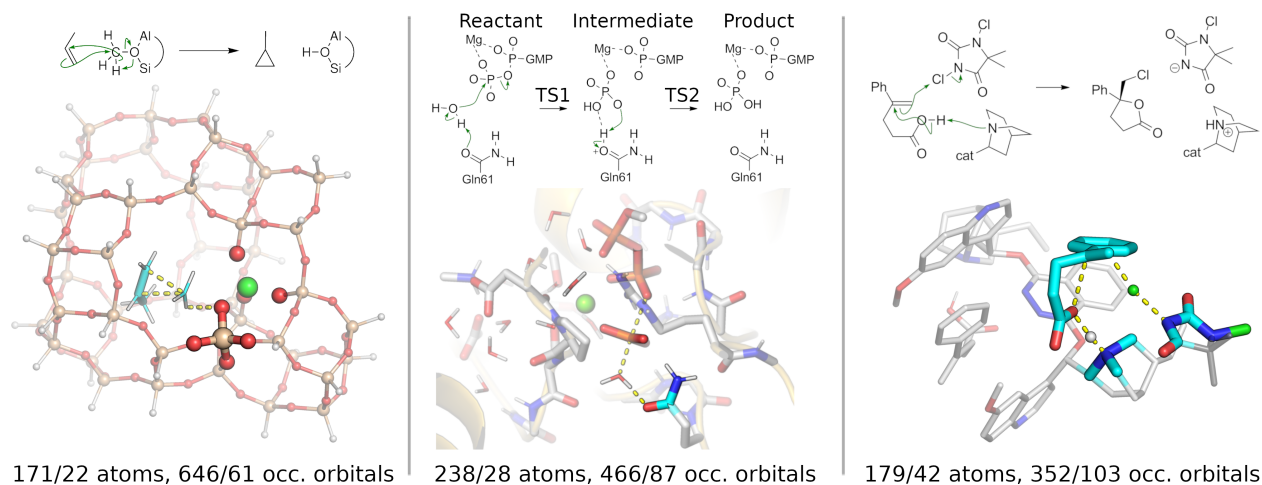


Figure 2: Zeolite catalyzed methylation reaction¹⁰⁶ (left), enzymatic phosphate hydrolysis¹⁰⁷ (middle), halocyclization reaction¹¹¹ (right). The skeletal structures above show the reactions, while the figures below depict the selected QM environment (transparent) and active (solid colors) regions in the TS structures. The numbers on the bottom denote the total number of QM/active atoms and occupied orbitals.

For the above reactions, (meta-)GGA functionals typically yield too high errors. In the case of the enzymatic reaction,¹⁰⁸ the GGA errors are in the -5.7 to -4.3 kcal/mol range. PBE¹¹³ is among the best at the (meta-)GGA rungs, while PBE0¹¹⁴ is among the best hybrid and range-separated hybrid functionals compared to LNO coupled-cluster singles, doubles and perturbative triples [CCSD(T)]^{1,2,115} reference calculations. Therefore, we used the PBE0-in-PBE embedding method in our tests. For the zeolite-catalyzed reaction, the errors are between -9 to -15 kcal/mol with GGA functionals. The PBE family of functionals are

among the best performers, thus we opted for the PBE0-in-PBE method again. In the case of the organocatalytic reaction, we found errors around -13.5 to -16.3 kcal/mol with GGA functionals with respect to LNO-CCSD(T).¹¹² At the meta-GGA rung, M06-L^{116,117} gave the best results with B97M-V¹¹⁸ being almost as good. Among the hybrid functionals, the CAM-B3LYP¹¹⁹ was the most accurate, thus we used CAM-B3LYP-in-M06-L embedding. For the small test reactions, the PBE0-in-PBE embedding was employed. We applied the D3 dispersion correction¹²⁰ with Becke–Johnson damping¹²¹ for all reactions, both for the low- and high-level calculations.

Dunning’s cc-pVXZ ($X=D,T,Q$) basis set^{122,123} were used for the small and zeolite reactions with the corresponding cc-pVXZ-RI-JK¹²⁴ fitting basis sets. For the other enzyme and organocatalytic reactions, a mixed basis set was used with def2-TZVP on the active and def2-SVP¹²⁵ on the environment atoms. The fitting basis set was a mixed def2-QZVP-RI-JK¹²⁴/def2-QZVP-RI-J¹²⁶ for the enzyme reaction and def2-QZVP-RI-JK for the halocyclization as implemented in MRCC.^{127–129} We used the AO reduction threshold of $\epsilon_{AO} = 10^{-4}$ unless stated otherwise.

The molecular structures of all test systems can be found in the Supplementary Information (SI). The enzyme system was studied in a QM/MM setting. The MM point charges are also listed after the Cartesian coordinates, based on the CHARMM36m protein and TIP3P water force fields.^{130,131} All calculations were performed with the MRCC program package.^{127–129} Our implementation uses OpenMP parallelization for all algorithms. Additionally, the Message Passing Interface (MPI) can also be used with integral direct calculations. The presented algorithms will be available in the next release free for academic use. A sample input file, additional timing, scaling, and numerical results are provided in the SI.

4 Accuracy

4.1 Standard test reactions

The effect of AO basis reduction was already benchmarked for these test reactions.⁹² Based on these results, we used the threshold value of $\varepsilon_{\text{AO}} = 10^{-4}$, as it already gives reliable convergence and acceptable accuracy (around 1 kcal/mol). Thus focusing on the new fitting basis reduction approach, we calculated the reaction energy errors of the auxiliary basis reduction with respect to the results obtained with AO reduction only. In our experience, the choice of basis set does not alter our results significantly (see the Figs. S1 and S2 of the SI), therefore, only the errors corresponding to the cc-pVTZ basis is shown in detail in Fig. 3.

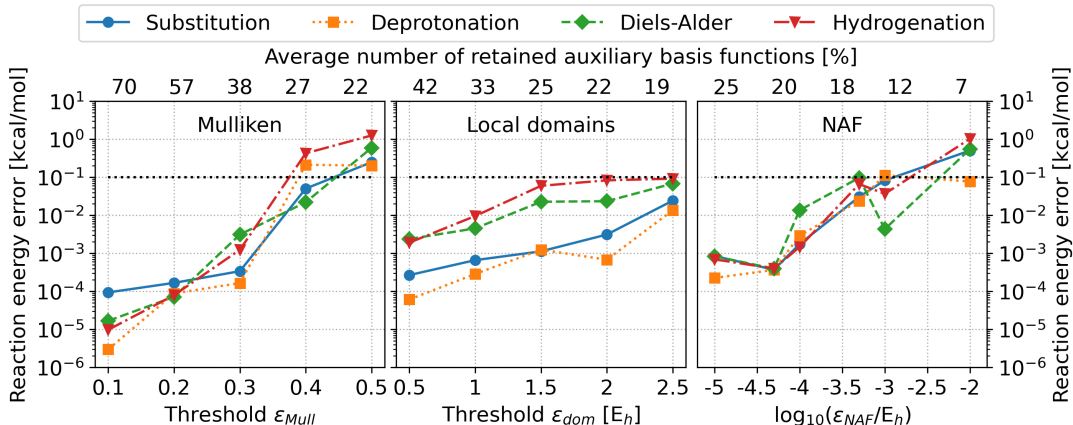


Figure 3: Reaction energy error of standard benchmark reactions (see text for details). The reference calculations use AO basis reduction only. The bottom x axis shows the applied thresholds, while the top x axis display the average ratio of kept fitting basis functions compared to the size of the complete auxiliary basis in %.

For all three tested fitting basis reduction methods, the errors are typically below 0.1 kcal/mol for affordably tight threshold values. The generalized Mulliken charges method retains around 40–70% of the fitting functions without notable loss in accuracy (errors are smaller than 0.01 kcal/mol) and around 30% of the fitting functions is enough if close to 0.1 kcal/mol errors are acceptable. The LDF domain based scheme performs slightly better as it can drop

around 10% more functions with similar accuracy. The NAF approach retains the least number of fitting functions with below 0.1 kcal/mol error. In conclusion, around 70–90% of the fitting functions can be dropped depending on the chosen algorithm with sub-0.1 kcal/mol errors in the reaction energies.

Importantly, these conclusions hold independently of the reaction type in these 4 test cases, including single, double and aromatic bond breaking. However, these reactions involve quasi-one-dimensional molecules in order to study embedding methods with limited system size. For this reason, it is important to assess the performance of our algorithms on real-life, three-dimensional target applications with larger active subsystem and realistic amount of bonds on the border of the active and environment regions, which is presented in the next sections.

4.2 Catalysis in zeolite cavity

The barrier height of the zeolite catalyzed reaction is assessed in Fig. 4. In addition to the effect of the auxiliary basis reduction, the performance of the AO reduction with respect to the complete, reduction-free embedding calculation is also presented.

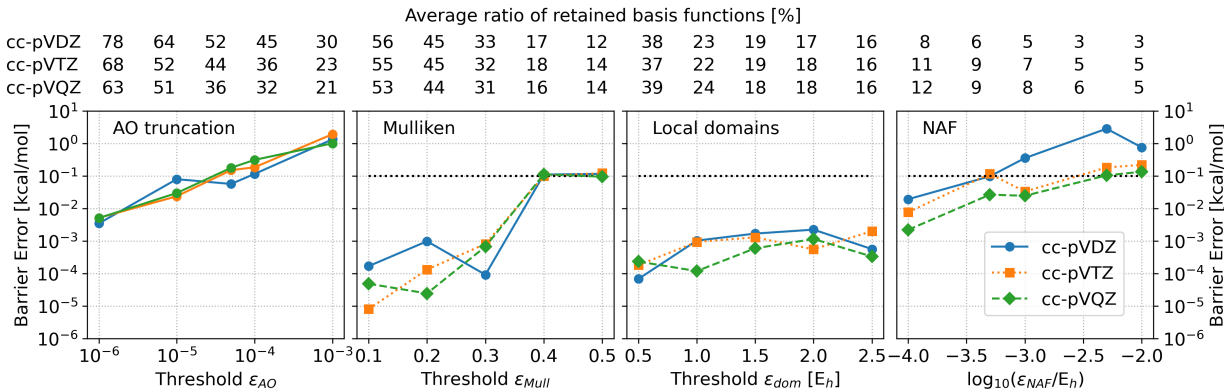


Figure 4: Barrier height errors for the zeolite system. The first panel shows the error of the AO reduction compared to approximation free embedding, while the other three panels show the error of the fitting basis reduction.

It is apparent from the results obtained with the cc-pVXZ ($X=D,T,Q$) basis sets that the errors of either the AO or fitting reduction schemes do not depend significantly on the

basis set, as also observed for the smaller test molecules. The error of the AO reduction is around 0.1 kcal/mol for most threshold values. To get below 0.1 kcal/mol for all basis sets, one can opt for $\varepsilon_{\text{AO}} < 10^{-5}$, retaining around 60% of the AOs.

Regarding the fitting basis reduction schemes, the generalized Mulliken charges approach gives somewhat higher, 0.1 kcal/mol error for looser (0.4–0.5) thresholds and a significantly smaller, 10^{-4} – 10^{-3} kcal/mol error for the tighter ones, while keeping around 30–40% of the auxiliary functions. The LDF domains based method performs more consistently. The errors are negligible, as they stay below 0.01 kcal/mol for all tested parameters and it only requires 15–20% of the original fitting functions. The NAF approach provides similar errors (around 0.01–0.1 kcal/mol) with a more compressed auxiliary space (below 10% of the original dimension) than with the other two algorithms.

4.3 Enzyme catalysis

We show the effect of both the AO and auxiliary basis reduction on the intermediate, TS and product states in Fig. 5. The results do not differ significantly, thus the basis reduction schemes can be safely applied for both barriers, intermediates and reaction energies. The results are similar to that obtained for the zeolite reaction, although the errors are slightly higher for the generalized Mulliken charge and LDF domain based methods, but still under 0.1 kcal/mol. On the other hand, the NAF-based approach becomes more accurate and the errors stay below 0.1 kcal/mol even with the loosest thresholds retaining 11–17% of the dimension of the original space.

The energy differences between the two barrier heights and between the intermediate and product states are also accurate (see Fig. 6). For the AO basis reduction, the errors are typically below or around 0.1 kcal/mol except for very loose thresholds. The fitting basis reduction schemes give errors below or around 0.01 kcal/mol in general. The LDF domain based approach performs the best, as the errors stay close to 0.01 kcal/mol for all tested thresholds. Thus differences of energy differences often required for chemical conclusions are

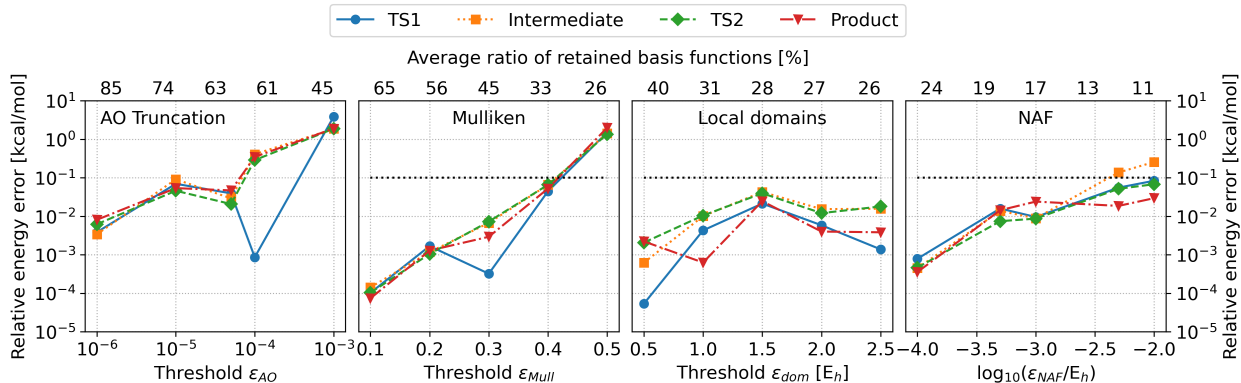


Figure 5: Relative energy errors for the enzyme reaction. For further details see the description of Fig. 4.

also highly reliable.

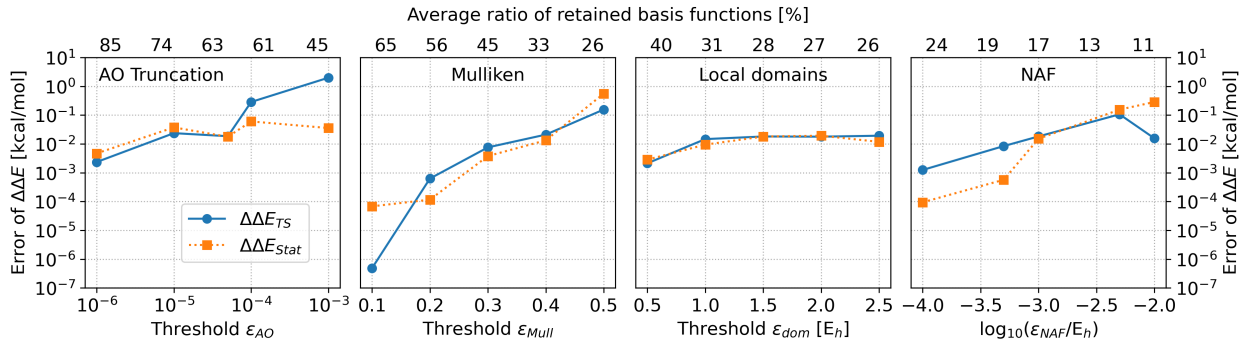


Figure 6: Error in the difference of the TS1 and TS2 transition states energies ($\Delta\Delta E_{TS}$) and the intermediate and product stationary states ($\Delta\Delta E_{stat}$). For reference, the energy differences calculated with basis reduction free embedding are 5.05 and 4.22 kcal/mol for $\Delta\Delta E_{TS}$ and $\Delta\Delta E_{stat}$, respectively. For further details see the description of Fig. 4.

The number of retained fitting basis functions in the enzyme reaction is around 4–10% larger than for the zeolite with the same thresholds. It is the result of the relatively larger number of active atoms (87 vs. 61). The NAF approach is still the most effective requiring about half as many fitting functions as the other two. Just as for the zeolite, the local domains based algorithm is more accurate than the generalized Mulliken charge based scheme if similarly sized fitting spaces are compared.

4.4 Organocatalysis

For the halocyclization, larger part of the substrates and the catalyst are required to be active (163 active MOs). The results are similar as for the zeolite and enzyme reactions (see Fig. S3 of the SI). In brief, the AO basis reduction gives around 0.1 kcal/mol error at the 10^{-4} threshold value, where it drops approximately 33% of the AO functions. The generalized Mulliken charge based algorithm gives satisfactory accuracy of 0.1 kcal/mol below the threshold value of 0.3 and it discards 40–65% of the auxiliary functions. The local domain based algorithm is more accurate as the errors stay below 10^{-3} kcal/mol, while it keeps similar number of functions as the generalized Mulliken charge based method.

5 Timing and primary memory requirements

To assess the performance of the AO and auxiliary basis reduction, runtime measurements were carried out on the three larger test systems with both integral-direct and the new in-core algorithm, using 16 cores (and 16 threads, with hyperthreading turned off) of an Intel Xeon Gold 6448H CPU. To provide multiple perspectives, the total speedup, the speedup in the high-level SCF and the ratio of high-level SCF and total runtime are plotted. The reference method is always the complete basis, reduction-free embedding calculation using integral-direct DF-SCF algorithm with the same basis set. That is, all reference hybrid and GGA computations are accelerated via DF and none of the timings reported in the paper were obtained via four-center algorithms. Timing measurements were performed in a production environment where competing computational load was present and not controlled.

Starting with the high-level SCF for the zeolite system, the AO reduction is around 2.6-times faster than the reference DF hybrid DFT (see Fig. 7a). Using the in-core algorithm, the speedup can be increased to 12. Fitting basis reduction outperforms this with a speedup of around 20, even with the integral-direct generalized Mulliken charge or LDF domain based methods. The NAF and in-core LDF domain based schemes can reach speedups around

40–50 or even higher.

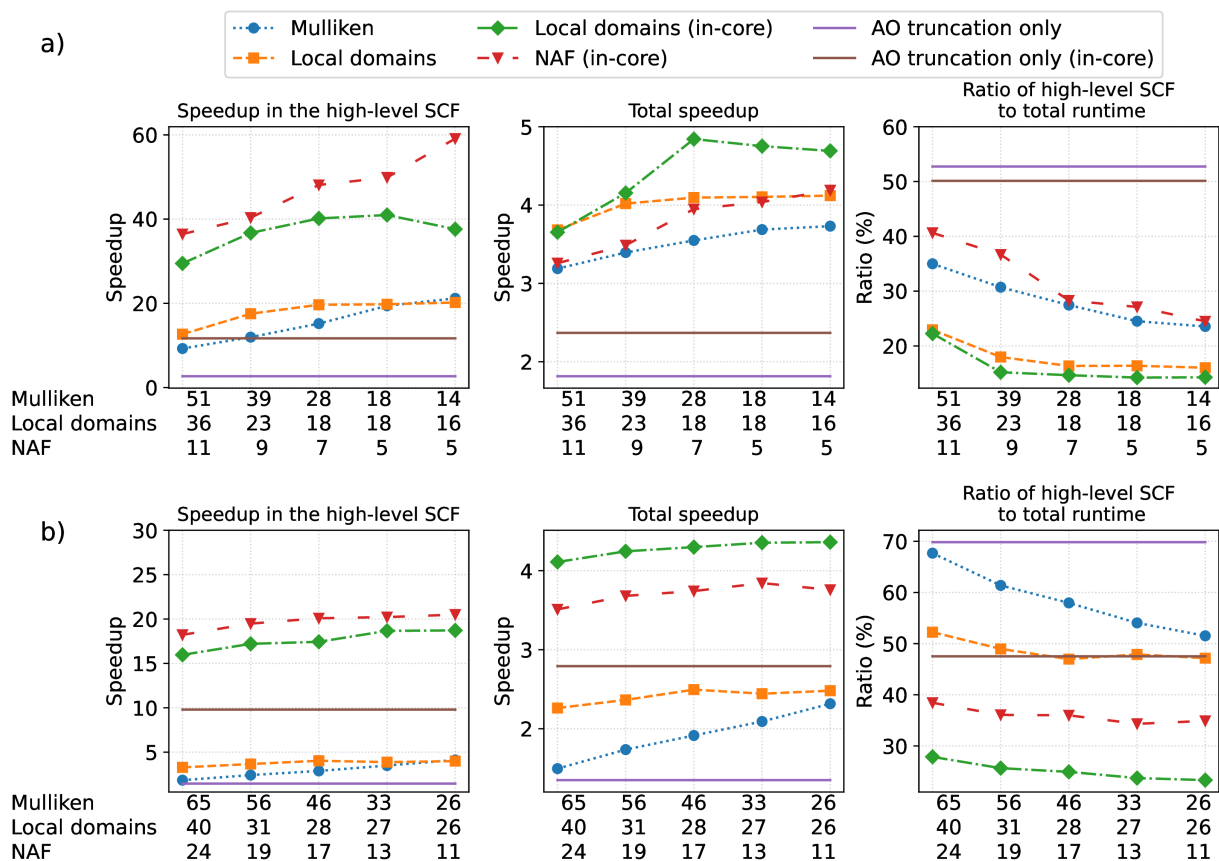


Figure 7: a) Runtime data for the zeolite system using the cc-pVTZ basis set and b) the enzyme reaction using the def2-TZVP-in-def2-SVP basis set. The x axis shows the ratio of the retained basis functions in %.

The speedups in the total embedding calculations (middle panels) are a bit smaller than 2 via the AO basis reduction alone, which can be improved to around 2.5 by the in-core algorithm. Turning on the fitting basis reduction schemes leads to better performance of 3–4 times speedups in general. Moreover, an about 5-fold acceleration in the total runtime is measured for the LDF domain and in-core algorithm. It is worth noting that the NAF approach is slower than the local domains based scheme, despite its higher level of compression. The reason is the relatively large overhead of the NAF (and thus also of the generalized Mulliken charge based) approach.

Finally, considering the ratio of the high-level SCF to the total runtime, it is 16–20% for

the in-core LDF domain and NAF approaches and 25–40% for the integral-direct algorithms. Thus, we achieved that the most time consuming step is now the low-level DF-based GGA calculation for the entire molecule, and further improvement in the high-level method will not result in significant additional acceleration.

In the high-level SCF calculation for the enzyme reaction (Fig. 7b, left), the AO basis reduction achieves a speedup of 1.5, while the LDF domains and generalized Mulliken charge based methods show 4-fold acceleration. The in-core algorithms are significantly faster: for AO basis reduction a 10-fold speedup, and for fitting basis reduction around 15–20-fold speedups were measured. The smaller values compared to the zeolite system can be explained by the larger active MO list and thus the higher proportion of the kept basis functions.

The total speedups are similarly satisfactory as for the zeolite system. It is below 3 for the AO reduction only and integral-direct algorithms, while it is around 4 for the in-core fitting basis reduction methods. The high-level method takes around 50–70% of the total runtime for the integral-direct case, while the low-level DF-based GGA part dominates with 60–80% for the in-core algorithm.

For the halocyclization reaction, we get similar results as for the zeolite system (Fig. S4 of the SI). With AO basis reduction alone, smaller than 2-fold speedups are measured. Using the auxiliary basis reduction too, around 10–12-fold speedup can be achieved for the high-level calculation and 6 times speedup in the total runtime. Here, the runtime of the high- and low-level methods are comparable. In the case of AO basis reduction only, almost 90% of the total runtime is spent in the high-level method, while with fitting basis set reduction, this is reduced to around 60%. The somewhat higher proportion of the high-level runtime is caused by the larger number of active MOs and the range-separated functional, for which two exchange matrices are built.

Based on the benchmark results above, we recommend threshold values of $5 \cdot 10^{-4} E_h$ for the NAF, $2.0 E_h$ for the LDF domain and 0.3 for the generalized Mulliken charge based algorithms. These values provide both high performance gains and accuracy. The runtimes

Table 1: Timing data in minutes for the three large test systems. TZ labels full triple- ζ (cc-pVTZ is used for all zeolite computations), TZ-in-DZ means mixed def2-TZVP-in-def2-SVP basis set (used for all embedded enzyme and halocyclization reactions, including those where not labeled explicitly). All computations, including the non-embedded reference hybrid and GGA measurements, consistently employ DF. Thresholds: $\varepsilon_{\text{AO}} = 10^{-4}$, $\varepsilon_{\text{dom}} = 2.0 E_h$, $\varepsilon_{\text{NAF}} = 5 \cdot 10^{-4} E_h$.

Number of	Zeolite	Enzyme	Halocyclization
Atoms	171	238	179
Active atoms	22	28	42
Occupied MOs	646	466	352
Active MOs	61	87	103
Wall times in minutes			
Hybrid/TZ	1327	527	862
Hybrid/TZ-in-DZ	-	139	273
GGA/TZ	12	15	23
GGA/TZ-in-DZ	-	6	19
Hybrid-in-GGA	68	45	204
AO reduction	30	34	122
Local domains	17	19	32
generalized Mulliken	19	24	40
Local domains (in-core)	14	10	-
NAF (in-core)	19	12	-

with these threshold are presented in Table 1. On 16 cores, starting from 527–1327 minutes without approximations, the new LESS approach achieves 10–32 minutes wall time. This is equivalent to around 2–8 CPU core hour for a single energy evaluation depending on the modeled system and DFT functional. The 10–32 minutes runtimes are comparable to the DF-accelerated GGA computations with full triple- ζ basis (12–23 minutes) and thus reaches the speed required for large scale sampling and (with forces under development) dynamics simulations in the future.

The above timings clearly show that storing integrals in primary memory has an advantage over integral-direct calculations. However, it is not feasible to store in primary memory the complete set of three-center integrals even in sparse format for large basis sets and/or molecules. For typical active region sizes, the AO reduction can roughly halve the number of integrals, while the fitting basis reduction omits around 80% of the auxiliary functions.

As a result, approximately 10% of the integrals is kept, which can very often be stored in the primary memory or at least on disk.

To show the viability of this idea, we find in Table 2 significant, 9–70 times primary memory savings for the zeolite and enzyme systems. Using the proposed approximations, the memory of a single workstation (i.e., here 20–40 GiB) is enough for our new in-core algorithm.

Table 2: Total primary memory requirement of the in-core algorithm for the zeolite and enzyme systems in GiB. Thresholds: $\varepsilon_{\text{AO}} = 10^{-4}$, $\varepsilon_{\text{dom}} = 2.0E_h$, $\varepsilon_{\text{NAF}} = 5 \cdot 10^{-4}E_h$.

Memory requirement [GiB]			
	Zeolite		Enzyme
	cc-pVDZ	cc-pVTZ	TZ-in-DZ
Hybrid-in-GGA	192.8	1400*	181.8
AO reduction	66.2	220.0	105.1
LDF domains	11.4	40.5	28.2
NAF	4.1	19.4	19.9

*Estimated value

6 Discussion and Conclusions

To summarize, we introduced and assessed local approximations to embedded subsystems, i.e., the LESS framework in combination with the Huzinaga embedding and DF algorithms. We implemented AO basis reduction and three new algorithms to select the important fitting basis functions in a Huzinaga or projection-based embedding calculation. The LDF domains based algorithm is accurate in all cases (below 0.1 kcal/mol error in both reaction energies and barrier heights), while it drops around 70–80% of the fitting function depending on the system. The NAF approach is also very accurate and can discard even more functions, approximately 80–90%. The LDF domains based scheme is usually the fastest of the three in total runtime as it drops a large portion of the fitting functions, while it has a negligible computational overhead. Moreover its generalization to, e.g., derivative properties is also relatively straightforward and it can be used together with integral-direct or in-core

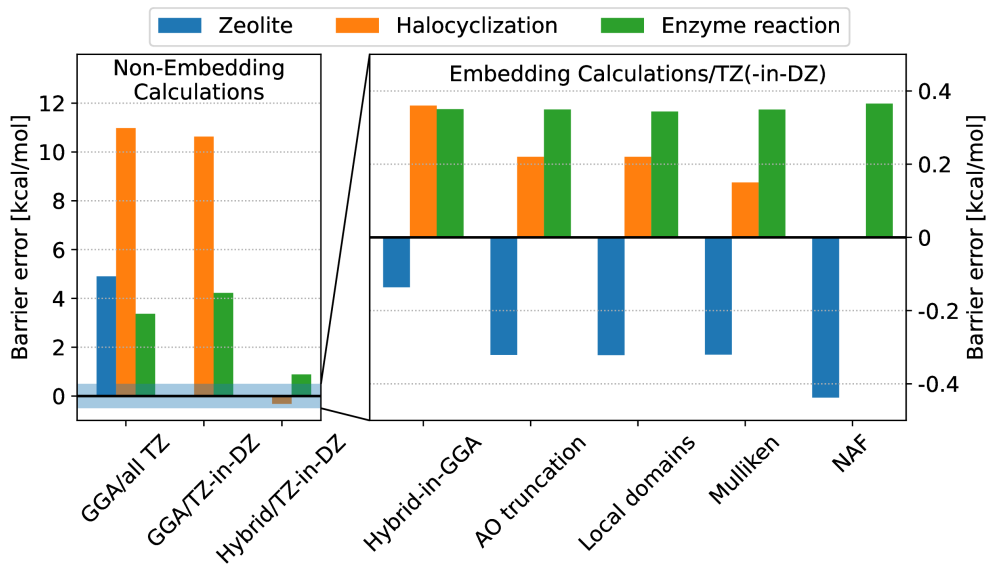


Figure 8: Summary of the accuracy of the different levels of theory for the three larger test systems with the recommended thresholds with respect to the hybrid DFT/triple- ζ results on the complete system. Barrier heights are shown here, while the corresponding reaction energy errors are given in Fig. S6 of the SI. For the enzyme reaction, the barrier height of the first TS was used here.

algorithms.

If one aims at an error on the order of 0.1 kcal/mol in relative energies, such as reaction energies and barriers (i.e. not affecting the overall accuracy of hybrid DFT), we recommend the thresholds of $5 \cdot 10^{-4} E_h$ for the NAF, $2.0 E_h$ for the LDF domain and 0.3 for the generalized Mulliken charge based algorithms. The errors with these settings for the three larger demonstrative reactions are displayed in Table 3, as well as in Fig. 8 for the barriers and in Fig. S6 for the reaction energies. In all of these, comparisons are made to the complete DF-based hybrid DFT calculations (without embedding and basis reduction) using a triple- ζ quality basis set. The LESS approximated reaction energies, especially for the zeolite and halocyclization reactions, are even more accurate than the barriers. The errors in the absolute energies are also highly satisfactory, $10^{-5}\%$ and $10^{-6}\%$ – $10^{-7}\%$ for the AO and auxiliary basis reduction, respectively, and cancel excellently in energy differences (Table S1). The corresponding absolute and relative runtimes are collected in Tables 1 and 3, respectively. For example, comparing the first and last row of Table 3, we find that the DF-accelerated GGA

Table 3: Summary of speedups and errors [kcal/mol] with the recommended thresholds. For the zeolite, the cc-pVTZ basis set was used on all atoms in all cases. For the other two molecules triple- ζ quality basis set was used on the active and double- ζ on the environment atoms. The reference is the hybrid DFT calculation with triple- ζ basis set for both errors and speedups.

	Zeolite		Enzyme		Halocyclization	
	Speedup	Error	Speedup	Error	Speedup	Error
GGA/TZ(-in-DZ)	111	4.90	86	4.23	46	10.63
Hybrid/TZ-in-DZ	-	-	4	0.89	3	-0.32
Hybrid-in-GGA	20	-0.14	12	0.35	4	0.36
AO reduction	45	-0.32	16	0.35	7	0.22
LESS (LDF domains)*	93	-0.32	51	0.34	27	0.22

*The speedup values correspond to the in-core algorithm for the zeolite and enzyme systems, and to the integral-direct for the halocyclization.

and high-level hybrid DFT computations with the LESS approach are 46–111- and 27–93-times faster than the approximation free, DF-based hybrid DFT computation, respectively. Compared to their similar speed, the GGA and hybrid-DFT-in-GGA errors are 4.2–10.6 kcal/mol and 0.2–0.3 kcal/mol, respectively. Thus, we can conclude that practically hybrid DFT accuracy can be achieved (including all the embedding, mixed AO basis, AO and fitting basis reduction approximations), while the overhead of the hybrid DFT part is similar or less than that of a DF-accelerated GGA calculation for the full system (cf. Table 1). In other words, the total LESS hybrid DFT-in-GGA computation time is comparable to that of about two DF-GGA runs. The errors are mostly caused by the embedding and AO basis reduction, the fitting basis reduction introduces errors in the order of 0.01 kcal/mol. If motivated by tighter accuracy requirements in certain applications, naturally, all approximations can be systematically improved even further.

The runtime acceleration is the combined effect of several techniques. On their own, each of the applied approximations (mixed basis, embedding, AO and fitting basis reduction) yield 2–4-fold speedups in the high-level computation. If used together, the effect of the approximations amplify each other and ca. 30–90-fold speedups are achieved. This performance is close to that of ONIOM (if one uses an active region that is tiny compared to the usual practice in ONIOM calculations), while the errors are decreased for these reactions

by a factor of 10–50 compared to that of GGA calculations.

Due to this performance, we find the here introduced LESS framework highly promising for DFT-in-DFT embedding. We recommend as default the use of the LDF domain based algorithm, due to the advantageous properties, such as the high compression rate, the computational affordability and its compatibility with both integral-direct and in-core (or disk-based) algorithms. For these reasons, this LDF domain based algorithm is also the prime candidate for accelerating the analytic derivatives for embedding methods. The extension of the local embedded subsystem (LESS) framework is underway in our laboratory to not just DFT-in-DFT but also wave function-in-DFT type schemes including both single point energies and analytic gradients. Such developments should enable large scale structure optimization or dynamics applications including both proper environment effects and accurate treatment of reactivity.

Supporting Information Available

See the file SI_geometries.xyz of the supporting information for the Cartesian coordinates and the applied active subspaces of the benchmarked molecules. The results obtained for the larger test systems (zeolite, halocyclization, enzyme reaction) are tabulated in SI.xls. Example input and further information can be found in SI.pdf.

Acknowledgment

The authors gratefully acknowledge support by ERC Starting Grant No. 101076972, “aC-Curacy”, János Bolyai Research Scholarship of the Hungarian Academy of Sciences, the National Research, Development, and Innovation (NRDI) Office (Starting 149434, FK142489), University Research Excellence Program (EKÖP-24-4-II-BME-321, EKÖP-24-4-I-BME-266). The Hungarian Governmental Information-Technology Development Agency is acknowledged for awarding access to the Komondor and the LEONARDO supercomputer, owned by the

EuroHPC Joint Undertaking, hosted by CINECA (Italy) and the LEONARDO consortium.

References

- (1) Nagy, P. R. State-of-the-art local correlation methods enable accurate and affordable gold standard quantum chemistry up to a few hundred atoms. *Chem. Sci.* **2024**, *15*, 14556.
- (2) Nagy, P. R.; Kállay, M. Approaching the basis set limit of CCSD(T) energies for large molecules with local natural orbital coupled-cluster methods. *J. Chem. Theory Comput.* **2019**, *15*, 5275.
- (3) Guo, Y.; Riplinger, C.; Becker, U.; Liakos, D. G.; Minenkov, Y.; Cavallo, L.; Neese, F. Communication: An improved linear scaling perturbative triples correction for the domain based local pair-natural orbital based singles and doubles coupled cluster method [DLPNO-CCSD(T)]. *J. Chem. Phys.* **2018**, *148*, 011101.
- (4) Goerigk, L.; Hansen, A.; Bauer, C.; Ehrlich, S.; Najibi, A.; Grimme, S. A look at the density functional theory zoo with the advanced GMTKN55 database for general main group thermochemistry, kinetics and noncovalent interactions. *Phys. Chem. Chem. Phys.* **2017**, *19*, 32184.
- (5) Mardirossian, N.; Head-Gordon, M. Thirty years of density functional theory in computational chemistry: an overview and extensive assessment of 200 density functionals. *Mol. Phys.* **2017**, *115*, 2315.
- (6) Warshel, A.; Levitt, M. Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.* **1976**, *103*, 227.
- (7) Tzeliou, C. E.; Mermigki, M. A.; Tzeli, D. Review on the QM/MM methodologies and their application to metalloproteins. *Molecules* **2022**, *27*, 2660.

- (8) Chow, M.; Lambros, E.; Li, X.; Hammes-Schiffer, S. Nuclear–Electronic Orbital QM/MM Approach: Geometry Optimizations and Molecular Dynamics. *J. Chem. Theory Comput.* **2023**, *19*, 3839–3848.
- (9) van der Kamp, M. W.; Mulholland, A. J. Combined Quantum Mechanics/Molecular Mechanics (QM/MM) Methods in Computational Enzymology. *Biochemistry* **2013**, *52*, 2708–2728.
- (10) Berta, D.; Buigues, P. J.; Badaoui, M.; Rosta, E. Cations in motion: QM/MM studies of the dynamic and electrostatic roles of H⁺ and Mg²⁺ ions in enzyme reactions. *Current Opinion in Structural Biology* **2020**, *61*, 198–206.
- (11) Bernstein, N.; Kermode, J. R.; Csányi, G. Hybrid atomistic simulation methods for materials systems. *Reports on Progress in Physics* **2009**, *72*, 026501.
- (12) Dong, K.; Liu, X.; Dong, H.; Zhang, X.; Zhang, S. Multiscale Studies on Ionic Liquids. *Chemical Reviews* **2017**, *117*, 6636–6695.
- (13) Sushko, M. L.; Sushko, P. V.; Abarenkov, I. V.; Shluger, A. L. QM/MM method for metal–organic interfaces. *Journal of Computational Chemistry* **2010**, *31*, 2955–2966.
- (14) Van der Mynsbrugge, J.; Bell, A. T. Challenges for the theoretical description of the mechanism and kinetics of reactions catalyzed by zeolites. *Journal of Catalysis* **2021**, *404*, 832–849.
- (15) Garcia Carcamo, R. A.; Shi, J.; Estejab, A.; Xie, T.; Bhattacharjee, S.; Biswas, S.; Bodenschatz, C. J.; Chen, X.; Maurya, M.; Zhang, X.; Getman, R. B. A Perspective on Multiscale Modeling of Explicit Solvation-Enabled Simulations of Catalysis at Liquid–Solid Interfaces. *ACS Catalysis* **2025**, *15*, 7448–7457.
- (16) Maseras, F.; Morokuma, K. IMOMM – A new integrated ab-initio plus molecular

- mechanics geometry optimization scheme of equilibrium structures and transition-states. *J. Comput. Chem.* **1995**, *16*, 1170.
- (17) Vreven, T.; Byun, K. S.; Komáromi, I.; Dapprich, S.; Montgomery, J. A. J.; Morokuma, K.; Frisch, M. J. Combining Quantum Mechanics Methods with Molecular Mechanics Methods in ONIOM. *J. Chem. Theory Comput.* **2006**, *2*, 815–826.
- (18) Svensson, M.; Humbel, S.; Froese, R. D. J.; Matsubara, T.; Sieber, S.; Morokuma, K. ONIOM: A Multilayered Integrated MO + MM Method for Geometry Optimizations and Single Point Energy Predictions. A Test for Diels–Alder Reactions and Pt(P(t-Bu)₃)₂ + H₂ Oxidative Addition. *J. Chem. Phys.* **1996**, *100*, 19357–19363.
- (19) Knizia, G.; Chan, G. K.-L. Density Matrix Embedding: A Strong-Coupling Quantum Embedding Theory. *J. Chem. Theory Comput.* **2013**, *9*, 1428.
- (20) Fornace, M. E.; Lee, J.; Miyamoto, K.; Manby, F. R.; Miller III, T. F. Embedded Mean-Field Theory. *J. Chem. Theory Comput.* **2015**, *11*, 568.
- (21) He, N.; Li, C.; Evangelista, F. A. Second-Order Active-Space Embedding Theory. *J. Chem. Theory Comput.* **2022**, *18*, 1527–1541.
- (22) Wesolowski, T. A.; Warshel, A. Frozen density functional approach for ab initio calculations of solvated molecules. *J. Phys. Chem.* **1993**, *97*, 8050.
- (23) Wesolowski, T. A.; Shedge, S.; Zhou, X. Frozen-Density Embedding Strategy for Multilevel Simulations of Electronic Structure. *Chem. Rev.* **2015**, *115*, 5891.
- (24) De Santis, M.; Sorbelli, D.; Vallet, V.; Gomes, A. S. P.; Storchi, L.; Belpassi, L. Frozen-Density Embedding for Including Environmental Effects in the Dirac-Kohn–Sham Theory: An Implementation Based on Density Fitting and Prototyping Techniques. *J. Chem. Theory Comput.* **2022**, *18*, 5992–6009.

- (25) Cortona, P. Self-consistently determined properties of solids without band-structure calculations. *Phys. Rev. B* **1991**, *44*, 8454.
- (26) Jacob, C. R.; Neugebauer, J. Subsystem density-functional theory. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2014**, *4*, 325–362.
- (27) Jacob, C. R.; Neugebauer, J. Subsystem density-functional theory (update). *WIREs Computational Molecular Science* **2024**, *14*.
- (28) Kitaura, K.; Ikeo, E.; Asada, T.; Nakano, T.; Uebayasi, M. *Chem. Phys. Lett.* **1999**, *313*, 701.
- (29) Le, H.-A.; Tan, H.-J.; Ouyang, J. F.; Bettens, R. P. A. Combined Fragmentation Method: A Simple Method for Fragmentation of Large Molecules. *J. Chem. Theory Comput.* **2012**, *8*, 469–478.
- (30) Gadre, S. R.; Shirsat, R. N.; Limaye, A. C. Molecular Tailoring Approach for Simulation of Electrostatic Properties. *J. Phys. Chem.* **1994**, *98*, 9165–9169.
- (31) Zhang, D. W.; Zhang, J. Z. H. Molecular fractionation with conjugate caps for full quantum mechanical calculation of protein–molecule interaction energy. *J. Chem. Phys.* **2003**, *119*, 3599–3605.
- (32) Meitei, O. R.; Heßelmann, A. Molecular energies from an incremental fragmentation method. *J. Chem. Phys.* **2016**, *144*, 084109.
- (33) Culpitt, T.; Brorsen, K. R.; Hammes-Schiffer, S. Density functional theory embedding with the orthogonality constrained basis set expansion procedure. *J. Chem. Phys.* **2017**, *146*, 211101.
- (34) Li, S.; Shen, J.; Li, W.; Jiang, Y. An efficient implementation of the “cluster-in-molecule” approach for local electron correlation calculations. *J. Chem. Phys.* **2006**, *125*, 074109.

- (35) Mata, R. A.; Werner, H.-J.; Schütz, M. Correlation regions within a localized molecular orbital approach. *J. Chem. Phys.* **2008**, *128*, 144106.
- (36) Li, W.; Piecuch, P. Multilevel Extension of the Cluster-in-Molecule Local Correlation Methodology: Merging Coupled-Cluster and Møller–Plesset Perturbation Theories. *J. Phys. Chem. A* **2010**, *114*, 6721.
- (37) Myhre, R. H.; Sánchez de Merás, A. M. J.; Koch, H. Multi-level coupled cluster theory. *J. Chem. Phys.* **2014**, *141*, 224105.
- (38) Sparta, M.; Retegan, M.; Pinski, P.; Riplinger, C.; Becker, U.; Neese, F. Multilevel Approaches within the Local Pair Natural Orbital Framework. *J. Chem. Theory Comput.* **2017**, *13*, 3198.
- (39) Pernal, K. Reduced density matrix embedding. General formalism and inter-domain correlation functional. *Phys. Chem. Chem. Phys.* **2016**, *18*, 21111–21121.
- (40) Lee, S. J. R.; Welborn, M.; Manby, F. R.; Miller III, T. F. Projection-Based Wavefunction-in-DFT Embedding. *Acc. Chem. Res.* **2019**, *52*, 1359.
- (41) Manby, F. R.; Stella, M.; Goodpaster, J. D.; Miller III, T. F. A Simple, Exact Density-Functional-Theory Embedding Scheme. *J. Chem. Theory Comput.* **2012**, *8*, 2564.
- (42) Goodpaster, J. D.; Barnes, T. A.; Manby, F. R.; Miller III, T. F. Accurate and systematically improvable density functional theory embedding for correlated wavefunctions. *J. Chem. Phys.* **2014**, *140*, 18A507.
- (43) Hégyely, B.; Nagy, P. R.; Ferenczy, G. G.; Kállay, M. Exact density functional and wave function embedding schemes based on orbital localization. *J. Chem. Phys.* **2016**, *145*, 064107.
- (44) Parravicini, V.; Jagau, T.-C. Embedded equation-of-motion coupled-cluster theory for

- electronic excitation, ionisation, electron attachment, and electronic resonances. *Mol. Phys.* **2021**, *119*, e1943029.
- (45) Hégyely, B.; Nagy, P. R.; Kállay, M. Dual basis set approach for density functional and wave function embedding schemes. *J. Chem. Theory Comput.* **2018**, *14*, 4600.
- (46) Chulhai, D. V.; Goodpaster, J. D. Improved Accuracy and Efficiency in Quantum Embedding through Absolute Localization. *J. Chem. Theory Comput.* **2017**, *13*, 1503.
- (47) Thompson, T. H.; Ochsenfeld, C. Integral partition bounds for fast and effective screening of general one-, two-, and many-electron integrals. *J. Chem. Phys.* **2019**, *150*, 044101.
- (48) Irmeler, A.; Pauly, F. Multipole-based distance-dependent screening of Coulomb integrals. *J. Chem. Phys.* **2019**, *151*, 084111.
- (49) Kussmann, J.; Ochsenfeld, C. Pre-selective screening for matrix elements in linear-scaling exact exchange calculations. *J. Chem. Phys.* **2013**, *138*, 134114.
- (50) Cremer, D.; Gauss, J. An unconventional scf method for calculations on large molecules. *J. Comput. Chem.* **1986**, *7*, 274–282.
- (51) Häser, M.; Ahlrichs, R. Improvements on the direct SCF method. *J. Comput. Chem.* **1989**, *10*, 104.
- (52) 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.
- (53) Friesner, R. A. Solution of the Hartree–Fock equations by a pseudospectral method: Application to diatomic molecules. *J. Chem. Phys.* **1986**, *85*, 1462.
- (54) Friesner, R. A. Solution of self-consistent field electronic structure equations by a pseudospectral method. *Chem. Phys. Lett.* **1985**, *116*, 39.

- (55) Friesner, R. A. Solution of the Hartree–Fock equations for polyatomic molecules by a pseudospectral method. *J. Chem. Phys.* **1987**, *86*, 3522.
- (56) Greeley, B. H.; Russo, T. V.; Mainz, D. T.; Friesner, R. A.; Langlois, J.-M.; Goddard, W. A.; Donnelly, R. E.; Ringnalda, M. N. New pseudospectral algorithms for electronic structure calculations: Length scale separation and analytical two-electron integral corrections. *J. Chem. Phys.* **1994**, *101*, 4028.
- (57) Ringnalda, M. N.; Belhadj, M.; Friesner, R. A. Pseudospectral Hartree–Fock theory: Applications and algorithmic improvements. *J. Chem. Phys.* **1990**, *93*, 3397.
- (58) Neese, F.; Wennmohs, F.; Hansen, A.; Becker, U. Efficient, approximate and parallel Hartree–Fock and hybrid DFT calculations. A ‘chain-of-spheres’ algorithm for the Hartree–Fock exchange. *Chem. Phys.* **2009**, *356*, 98.
- (59) 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*, 094111.
- (60) Helmich-Paris, B.; de Souza, B.; Neese, F.; Izsák, R. An improved chain of spheres for exchange algorithm. *J. Chem. Phys.* **2021**, *155*, 104109.
- (61) Plessow, P.; Weigend, F. Seminumerical calculation of the Hartree–Fock exchange matrix: Application to two-component procedures and efficient evaluation of local hybrid density functionals. *J. Comput. Chem.* **2012**, *33*, 810.
- (62) Laqua, H.; Kussmann, J.; Ochsenfeld, C. Efficient and Linear-Scaling Seminumerical Method for Local Hybrid Density Functionals. *J. Chem. Theory Comput.* **2018**, *14*, 3451.
- (63) Laqua, H.; Thompson, T. H.; Kussmann, J.; Ochsenfeld, C. Highly Efficient, Linear-

- Scaling Seminumerical Exact-Exchange Method for Graphic Processing Units. *J. Chem. Theory Comput.* **2020**, *16*, 1456.
- (64) Guidon, M.; Hutter, J.; VandeVondele, J. Auxiliary Density Matrix Methods for Hartree–Fock Exchange Calculations. *J. Chem. Theory Comput.* **2010**, *6*, 2348.
- (65) Merlot, P.; Izsák, R.; Borgoo, A.; Kjærgaard, T.; Helgaker, T.; Reine, S. Charge-constrained auxiliary-density-matrix methods for the Hartree–Fock exchange contribution. *J. Chem. Phys.* **2014**, *141*, 094104.
- (66) Dong, K.; Hu, W.; Lin, L. Interpolative Separable Density Fitting through Centroidal Voronoi Tessellation with Applications to Hybrid Functional Electronic Structure Calculations. *J. Chem. Theory Comput.* **2018**, *14*, 1311.
- (67) Qin, X.; Liu, J.; Hu, W.; Yang, J. Interpolative Separable Density Fitting Decomposition for Accelerating Hartree–Fock Exchange Calculations within Numerical Atomic Orbitals. *J. Phys. Chem. A* **2020**, *124*, 5664.
- (68) Parrish, R. M.; Hohenstein, E. G.; Martínez, T. J.; Sherrill, C. D. Tensor hypercontraction. II. Least-squares renormalization. *J. Chem. Phys.* **2012**, *137*, 224106.
- (69) Lu, J.; Ying, L. Compression of the electron repulsion integral tensor in tensor hypercontraction format with cubic scaling cost. *J. Comput. Phys.* **2015**, *302*, 329.
- (70) Watson, M. A.; Salek, P.; Macak, P.; Helgaker, T. Linear-scaling formation of Kohn–Sham Hamiltonian: Application to the calculation of excitation energies and polarizabilities of large molecular systems. *J. Chem. Phys.* **2004**, *121*, 2915.
- (71) Challacombe, M.; Schwegler, E.; Almlöf, J. Fast assembly of the Coulomb matrix: A quantum chemical tree code. *J. Chem. Phys.* **1996**, *104*, 4685.
- (72) White, C. A.; Johnson, B. G.; Gill, P. M. W.; Head-Gordon, M. The continuous fast multipole method. *Chem. Phys. Lett.* **1994**, *230*, 8.

- (73) 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.
- (74) Le, H.-A.; Shiozaki, T. Occupied-Orbital Fast Multipole Method for Efficient Exact Exchange Evaluation. *J. Chem. Theory Comput.* **2018**, *14*, 1228.
- (75) Csóka, J.; Kállay, M. Speeding up density fitting Hartree–Fock calculations with multipole approximations. *Mol. Phys.* **2020**, *118*, e1769213.
- (76) Neese, F.; Colinet, P.; DeSouza, B.; Helmich-Paris, B.; Wennmohs, F.; Becker, U. The “Bubblepole” (BUPO) Method for Linear-Scaling Coulomb Matrix Construction with or without Density Fitting. *J. Phys. Chem. A* **2025**, *129*, 2618–2637.
- (77) 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.
- (78) Dunlap, B. I. *Phys. Chem. Chem. Phys.* **2000**, *2*, 2113.
- (79) Dunlap, B. I. Robust and variational fitting: Removing the four-center integrals from center stage in quantum chemistry. *J. Mol. Struct. (THEOCHEM)* **2000**, *529*, 37.
- (80) Eichkorn, K.; Treutler, O.; Öhm, H.; Häser, M.; Ahlrichs, R. *Chem. Phys. Lett.* **1995**, *240*, 283.
- (81) 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.
- (82) Stocks, R.; Palethorpe, E.; Barca, G. M. J. Multi-GPU RI-HF Energies and Analytic Gradients—Toward High-Throughput Ab Initio Molecular Dynamics. *J. Chem. Theory Comput.* **2024**, *20*, 7503–7515.

- (83) Manzer, S. F.; Epifanovsky, E.; Head-Gordon, M. Efficient Implementation of the Pair Atomic Resolution of the Identity Approximation for Exact Exchange for Hybrid and Range-Separated Density Functionals. *J. Chem. Theory Comput.* **2015**, *11*, 518.
- (84) Sodt, A.; Head-Gordon, M. Hartree–Fock exchange computed using the atomic resolution of the identity approximation. *J. Chem. Phys.* **2008**, *128*, 104106.
- (85) Manzer, S.; Horn, P. R.; Mardirossian, N.; Head-Gordon, M. Fast, accurate evaluation of exact exchange: The occ-RI-K algorithm. *J. Chem. Phys.* **2015**, *143*, 024113.
- (86) Polly, R.; Werner, H.-J.; Manby, F. R.; Knowles, P. J. Fast Hartree–Fock theory using local fitting approximations. *Mol. Phys.* **2004**, *102*, 2311.
- (87) Nagy, P. R.; Samu, G.; Kállay, M. An integral-direct linear-scaling second-order Møller–Plesset approach. *J. Chem. Theory Comput.* **2016**, *12*, 4897.
- (88) Delesma, F. A.; Geudtner, G.; Mejía-Rodríguez, D.; Calaminici, P.; Köster, A. M. Range-Separated Hybrid Functionals with Variational Fitted Exact Exchange. *J. Chem. Theory Comput.* **2018**, *14*, 5608.
- (89) Csóka, J.; Kállay, M. Analytic gradients for local density fitting Hartree–Fock and Kohn–Sham methods. *J. Chem. Phys.* **2023**, *158*, 024110.
- (90) Barnes, T. A.; Goodpaster, J. D.; Manby, F. R.; Miller III, T. F. Accurate basis set truncation for wavefunction embedding. *J. Chem. Phys.* **2013**, *139*, 024103.
- (91) Bennie, S. J.; Stella, M.; Miller III, T. F.; Manby, F. R. Accelerating wavefunction in density-functional-theory embedding by truncating the active basis set. *J. Chem. Phys.* **2015**, *143*, 024105.
- (92) Bensberg, M.; Neugebauer, J. Automatic basis-set adaptation in projection-based embedding. *J. Chem. Phys.* **2019**, *150*, 184104.

- (93) Bennie, S. J.; van der Kamp, M. W.; Pennifold, R. C. R.; Stella, M.; Manby, F. R.; Mulholland, A. J. A Projector-Embedding Approach for Multiscale Coupled-Cluster Calculations Applied to Citrate Synthase. *J. Chem. Theory Comput.* **2016**, *12*, 2689.
- (94) Bennie, S. J.; Curchod, B. F. E.; Manby, F. R.; Glowacki, D. R. Pushing the Limits of EOM-CCSD with Projector-Based Embedding for Excitation Energies. *J. Phys. Chem. Lett.* **2017**, *8*, 5559.
- (95) Chapovetsky, A.; Welborn, M.; Luna, J. M.; Haiges, R.; Miller, T. F. I.; Marinescu, S. C. Pendant Hydrogen-Bond Donors in Cobalt Catalysts Independently Enhance CO₂ Reduction. *ACS Cent. Sci.* **2018**, *4*, 397–404.
- (96) Zhang, X.; Bennie, S. J.; van der Kamp, M. W.; Glowacki, D. R.; Manby, F. R.; Mulholland, A. J. Multiscale analysis of enantioselectivity in enzyme-catalysed ‘lethal synthesis’ using projector-based embedding. *Royal Soc. Open Sci* **2018**, *5*, 171390.
- (97) Claudino, D.; Mayhall, N. J. Automatic Partition of Orbital Spaces Based on Singular Value Decomposition in the Context of Embedding Theories. *J. Chem. Theory Comput.* **2019**, *15*, 1053.
- (98) Kolodzeiski, E.; Stein, C. J. Automated, Consistent, and Even-Handed Selection of Active Orbital Spaces for Quantum Embedding. *J. Chem. Theory Comput.* **2023**, *19*, 6643.
- (99) Huzinaga, S.; Cantu, A. A. Theory of Separability of Many Electron Systems. *J. Chem. Phys.* **1971**, *55*, 5543.
- (100) Csóka, J.; Hégyel, B.; Nagy, P. R.; Kállay, M. Development of analytic gradients for the Huzinaga quantum embedding method and its applications to large-scale hybrid and double hybrid DFT forces. *J. Chem. Phys.* **2024**, *160*, 124113.

- (101) Kállay, M. A systematic way for the cost reduction of density fitting methods. *J. Chem. Phys.* **2014**, *141*, 244113.
- (102) Petrov, K.; Csóka, J.; Kállay, M. Analytic gradients for density fitting MP2 using natural auxiliary functions. *J. Phys. Chem. A* **2024**, *128*, 6566.
- (103) Köppl, C.; Werner, H.-J. Parallel and Low-Order Scaling Implementation of Hartree–Fock Exchange Using Local Density Fitting. *J. Chem. Theory Comput.* **2016**, *12*, 3122.
- (104) Lee, S. J. R.; Miyamoto, K.; Ding, F.; Manby, F. R.; Miller III, T. F. Density-based errors in mixed-basis mean-field electronic structure, with implications for embedding and QM/MM methods. *Chem. Phys. Lett.* **2017**, *683*, 375.
- (105) Miyamoto, K.; Miller III, T. F.; Manby, F. R. Fock-Matrix Corrections in Density Functional Theory and Use in Embedded Mean-Field Theory. *J. Chem. Theory Comput.* **2016**, *12*, 5811.
- (106) Goncalves, T. J.; Plessow, P. N.; Studt, F. On the Accuracy of Density Functional Theory in Zeolite Catalysis. *ChemCatChem* **2019**, *11*, 4368.
- (107) Berta, D.; Gehrke, S.; Nyíri, K.; Vértessy, B. G.; Rosta, E. Mechanism–Based Redesign of GAP to Activate Oncogenic Ras. *J. Am. Chem. Soc.* **2023**, *145*, 20302.
- (108) Berta, D.; Csóka, J.; Hégyel, B.; Nagy, P. R. in preparation.
- (109) Mehmood, R.; Kulik, H. J. Both Configuration and QM Region Size Matter: Zinc Stability in QM/MM Models of DNA Methyltransferase. *J. Chem. Theory Comput.* **2020**, *16*, 3121.
- (110) Clemente, C. M.; Capece, L.; Martí, M. A. Best Practices on QM/MM Simulations of Biological Systems. *Journal of Chemical Information and Modeling* **2023**, *63*, 2609.

- (111) Yousefi, R.; Sarkar, A.; Ashtekar, K. D.; Whitehead, D. C.; Kakeshpour, T.; Holmes, D.; Reed, P.; Jackson, J. E.; Borhan, B. Mechanistic Insights into the Origin of Stereoselectivity in an Asymmetric Chlorolactonization Catalyzed by (DHQD)2PHAL. *J. Am. Chem. Soc.* **2020**, *142*, 7179.
- (112) Laczkó, G.; Pápai, I.; Nagy, P. R. Understanding DFT uncertainties for more reliable reactivity predictions by advancing the analysis of error sources. *J. Chem. Theory Comput.* **2025**, under revision.
- (113) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865.
- (114) Perdew, J. P.; Ernzerhof, M.; Burke, K. Rationale for mixing exact exchange with density functional approximations. *J. Chem. Phys.* **1996**, *105*, 9982.
- (115) Nagy, P. R.; Samu, G.; Kállay, M. Optimization of the linear-scaling local natural orbital CCSD(T) method: Improved algorithm and benchmark applications. *J. Chem. Theory Comput.* **2018**, *14*, 4193.
- (116) Zhao, Y.; Truhlar, D. G. A new local density functional for main-group thermochemistry, transition metal bonding, thermochemical kinetics, and noncovalent interactions. *J. Chem. Phys.* **2006**, *125*, 194101.
- (117) Zhao, Y.; Truhlar, D. G. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor. Chem. Acc.* **2006**, *120*, 215.
- (118) Mardirossian, N.; Head-Gordon, M. Mapping the genome of meta-generalized gradient approximation density functionals: The search for B97M-V. *J. Chem. Phys.* **2015**, *142*, 074111.

- (119) Yanai, T.; Tew, D. P.; Handy, N. C. A new hybrid exchange-correlation functional using the Coulomb-attenuating method (CAM-B3LYP). *Chem. Phys. Lett.* **2004**, *393*, 51.
- (120) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* **2010**, *132*, 154104.
- (121) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* **2011**, *32*, 1456.
- (122) Dunning Jr., T. H. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007.
- (123) Woon, D. E.; Dunning Jr., T. H. Gaussian basis sets for use in correlated molecular calculations. III. The atoms aluminum through argon. *J. Chem. Phys.* **1993**, *98*, 1358.
- (124) Weigend, F. Hartree–Fock Exchange Fitting Basis Sets for H to Rn. *J. Comput. Chem.* **2008**, *29*, 167.
- (125) Weigend, F.; Ahlrichs, R. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297.
- (126) Weigend, F. Accurate Coulomb-fitting basis sets for H to Rn. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1057.
- (127) Mester, D.; Nagy, P. R.; Csóka, J.; Gyevi-Nagy, L.; Szabó, P. B.; Horváth, R. A.; Petrov, K.; Hégyel, B.; Ladóczki, B.; Samu, G.; Lőrincz, B. D.; Kállay, M. An overview of developments in the MRCC program system. *J. Phys. Chem. A* **2025**, *129*, 2086.
- (128) Kállay, M.; Nagy, P. R.; Mester, D.; Gyevi-Nagy, L.; Csóka, J.; Szabó, P. B.; Rólik, Z.; Samu, G.; Csontos, J.; Hégyel, B.; Ganyecz, Á.; Ladjánszki, I.; Szegedy, L.;

- Ladóczki, B.; Petrov, K.; Farkas, M.; Mezei, P. D.; Horváth, R. A. MRCC, *a quantum chemical program suite*. See <https://www.mrcc.hu/> **Accessed Jan 1, 2025**,
- (129) Kállay, M.; Nagy, P. R.; Mester, D.; Rolik, Z.; Samu, G.; Csontos, J.; Csóka, J.; Szabó, P. B.; Gyevi-Nagy, L.; Hégyel, B.; Ladjánszki, I.; Szegedy, L.; Ladóczki, B.; Petrov, K.; Farkas, M.; Mezei, P. D.; Ganyecz, Á. The MRCC program system: Accurate quantum chemistry from water to proteins. *J. Chem. Phys.* **2020**, *152*, 074107.
- (130) Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E. M.; Mittal, J.; Feig, M.; MacKerell, A. D. Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone ϕ , ψ and Side-Chain χ_1 and χ_2 Dihedral Angles. *J. Chem. Theory Comput.* **2012**, *8*, 3257–3273.
- (131) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935.

Graphical TOC Entry

