

# Efficient density-fitted explicitly correlated dispersion and exchange dispersion energies

Monika Kodrycka<sup>1,\*</sup> and Konrad Patkowski<sup>1,†</sup>

<sup>1</sup>*Department of Chemistry and Biochemistry, Auburn University, Auburn, AL 36849, United States*

(Dated: January 25, 2021)

The leading-order dispersion and exchange-dispersion terms in symmetry-adapted perturbation theory (SAPT),  $E_{\text{disp}}^{(20)}$  and  $E_{\text{exch-disp}}^{(20)}$ , suffer from slow convergence to the complete basis set limit. To alleviate this problem, explicitly correlated variants of these corrections,  $E_{\text{disp}}^{(20)}$ -F12 and  $E_{\text{exch-disp}}^{(20)}$ -F12, have been proposed recently. However, the original formalism (M. Kodrycka *et al.*, *J. Chem. Theory Comput.* **2019**, 15, 5965–5986), while highly successful in terms of improving convergence, was not competitive to conventional orbital-based SAPT in terms of computational efficiency due to the need to manipulate several kinds of two-electron integrals. In this work, we eliminate this need by decomposing all types of two-electron integrals using robust density fitting. We demonstrate that the error of the density fitting approximation is negligible when standard auxiliary bases such as aug-cc-pVXZ/MP2FIT are employed. The new implementation allowed us to study all complexes in the A24 database in basis sets up to aug-cc-pV5Z, and the  $E_{\text{disp}}^{(20)}$ -F12 and  $E_{\text{exch-disp}}^{(20)}$ -F12 values exhibit vastly improved basis set convergence over their conventional counterparts. The well-converged  $E_{\text{disp}}^{(20)}$ -F12 and  $E_{\text{exch-disp}}^{(20)}$ -F12 numbers can be substituted for conventional  $E_{\text{disp}}^{(20)}$  and  $E_{\text{exch-disp}}^{(20)}$  ones in a calculation of the total SAPT interaction energy at any level (SAPT0, SAPT2+3, ...). We show that the addition of F12 terms does not improve the accuracy of low-level SAPT treatments. However, when the theory errors are minimized in high-level SAPT approaches such as SAPT2+3(CCD) $\delta$ MP2, the reduction of basis set incompleteness errors thanks to the F12 treatment substantially improves the accuracy of small-basis calculations.

## I. INTRODUCTION

Noncovalent intermolecular interactions are ubiquitous, and their quantitative understanding is essential when one investigates diverse physical and chemical phenomena from scattering resonances in cold collisions<sup>1</sup> to crystal structure and polymorphism<sup>2</sup> to chiral discrimination in biomolecular complexes<sup>3</sup>. However, accurate computations of noncovalent interaction energies are quite nontrivial due to the overwhelming importance of electron correlation. In particular, London dispersion forces, which are the main reason for the attraction between nonpolar molecules, arise entirely out of electron correlation. Actually, these forces require quite a high-level account of correlation, as the simplest dispersion estimate, contained in the supermolecular interaction energy computed using the second-order Møller-Plesset perturbation theory (MP2), often leads to significant overbinding<sup>4</sup>. In addition to the need for a high-level electronic structure treatment (such as the “gold standard”<sup>5</sup> coupled-cluster approach with single, double, and perturbative triple excitations, CCSD(T)<sup>6</sup>), capturing the dispersion effects requires large basis sets, as the correlation energy is known for its slow convergence to the complete basis set (CBS) limit. The basis set convergence of molecular correlation energies can be enhanced by extrapolations<sup>7</sup> as well by an explicitly correlated treatment, where an explicit dependence on the interelectronic distance  $r_{12}$  is inserted into the wavefunction *Ansatz*<sup>8</sup>. The most practical and successful explicitly correlated variants use a correlation factor with nonlinear  $r_{12}$  dependence, in particular, the Slater-type expression  $e^{-\gamma r_{12}}$ , where the parameter  $\gamma$  can be varied to control the spatial range of explicit correlation<sup>9</sup>. Such an approach is denoted by adding “F12” to the name of the parent electronic structure theory, and many correlated *ab initio* approaches have been “F12’ed”, including MP2<sup>9,10</sup>, CCSD(T)<sup>11,12</sup>, or even multireference methods such as the second-order complete-active-space perturbation theory (CASPT2)<sup>13</sup> or multireference configuration interaction (MRCI)<sup>14</sup>. In the specific case of intermolecular interactions, another way of speeding up the CBS convergence is an addition of “midbond” basis functions centered in between the interacting molecules<sup>15</sup>. The inclusion of bond functions is particularly helpful in recovering the dispersion energy<sup>16</sup>, and the benefits of bond functions can be combined with those of the CBS extrapolation<sup>17</sup> and of the F12 treatment<sup>18</sup>.

The most rigorous definition of dispersion energy is provided by second-order perturbation theory (some higher-order terms can also be classified as pure dispersion, but these terms are typically small)<sup>19</sup>. Thus, dispersion can be quantified, separately from other energy contributions such as electrostatics, induction, and exchange, within the framework of symmetry-adapted perturbation theory (SAPT)<sup>20,21</sup>. Many variants of SAPT have been proposed in the literature, differing in the approach used to treat the intramolecular electron correlation. Model SAPT studies for the smallest complexes have been carried out<sup>22,23</sup> with the complete, full configuration interaction (FCI) account of intramolecular correlation, however, this is obviously not possible for realistic systems. Instead, one can start the perturbation expansion from the product of the monomer Hartree-Fock (HF) determinants. The simplest way to proceed is then to ignore intramolecular correlation completely and carry out a single perturbation expansion in powers of the intermolecular interaction operator, leading to an approach termed SAPT0. The SAPT0 dispersion energy, closely related to the dispersion part of the supermolecular MP2 interaction energy<sup>24</sup>, is denoted as  $E_{\text{disp}}^{(20)}$ , where the consecutive superscripts signify that this correction is of second order in the intermolecular interaction and of zeroth order in the intramolecular correlation. If quantitative accuracy of the SAPT energy terms is required, one has to go beyond SAPT0 and include, in one way or another, the intramolecular correlation effects. The historically first method of doing so is the *many-body SAPT*<sup>25</sup>, in which SAPT becomes a double perturbation theory in powers of both the intermolecular interaction operator  $V$  and the Møller-Plesset fluctuation potential  $W = W_A + W_B$ , where  $W_X$  is the difference between the full Hamiltonian for molecule X and its HF approximation. At the highest developed level of many-body SAPT, the dispersion energy is computed as  $E_{\text{disp}}^{(20)} + E_{\text{disp}}^{(21)} + E_{\text{disp}}^{(22)}$ , thus, it includes intramolecular correlation through second order in  $W$ . Alternatively, an intramolecular correlation correction to SAPT0 dispersion can be evaluated in a nonperturbative manner, using a CC-like partial infinite-order summation of terms of different orders in  $W$ <sup>26,27</sup> or the frequency-dependent density susceptibilities from time-dependent density functional theory, as in the SAPT(DFT) approach<sup>28,29</sup>. The development of improved dispersion expressions in SAPT continues to this day: the most recent accomplishments include the computation of dispersion energies from the Bethe-Salpeter equation<sup>30</sup> and from the linear response of the complete-active-space self-consistent field (CASSCF) approach<sup>31</sup>.

Every electrostatic, induction, and dispersion correction in SAPT is accompanied by an exchange term that arises from the enforcement of the full antisymmetry upon the wave function of the complex. In particular, second-order dispersion energy has an exchange-dispersion counterpart that cancels a part of the dispersion attraction (typically on the order of 10% at the van der Waals minimum distance). While smaller than dispersion, the exchange dispersion energy converges in relative terms just as poorly as dispersion energy with both the theory level and the basis set. This is quite unfortunate as all commonly used variants of many-body SAPT approximate the exchange dispersion effects solely by their leading-order  $E_{\text{exch-disp}}^{(20)}$  term. The intramolecular correlation contributions to exchange dispersion are approximately included in SAPT(DFT)<sup>28,29</sup>, while benchmark exchange dispersion values for small complexes can be computed at the CCSD level using the approach of Ref. 32. It should be noted that the exchange dispersion

effects, just like nearly all other SAPT exchange corrections, are usually computed within the so-called *single exchange approximation* that neglects intermolecular swaps of more than one electron pair when applying the antisymmetry projector (this approach is also called the  $S^2$  approximation as it neglects terms beyond the second power in the intermolecular overlap integrals). While the full nonexpanded  $E_{\text{exch-disp}}^{(20)}$  expression has been recently derived and implemented<sup>33</sup>, all beyond- $E_{\text{exch-disp}}^{(20)}$  exchange dispersion terms are always computed within the  $S^2$  approximation.

As the explicitly correlated F12 approach has been quite successful improving the basis set convergence of supermolecular MP2 and CCSD(T) interaction energies<sup>34–37</sup>, it is worth asking if the benefits of the F12 treatment can be transferred to SAPT, in particular, to the slowly converging dispersion and exchange dispersion energies. The calculation of both effects with the aid of explicitly correlated functions has quite a long history for model four-electron systems, and the highly accurate CBS values of  $E_{\text{disp}}^{(20)}$ ,  $E_{\text{disp}}^{(21)}$ , and  $E_{\text{exch-disp}}^{(20)}$  for the helium dimer have been established<sup>38–40</sup> using Gaussian-type geminal (GTG) explicitly correlated bases<sup>41</sup>. The first step towards a system-independent F12 extension of SAPT was recently taken by Przybytek<sup>42</sup>. He proposed an algorithm to improve the CBS convergence of  $E_{\text{disp}}^{(20)}$  using an *Ansatz* similar to MP2-F12, where ordinary dispersion amplitudes are augmented by a set of F12 amplitudes approximately taking into account the excitations out of the molecular-orbital (MO) space. The F12 amplitudes in Przybytek’s approach were obtained by a minimization of a suitable dispersion-only Hylleraas functional. Thus, the best variational approximation to the CBS value of  $E_{\text{disp}}^{(20)}$  at a given basis set level was obtained, and the resulting CBS convergence improvement relative to conventional  $E_{\text{disp}}^{(20)}$  was truly impressive, comparable to increasing the orbital basis set (of the augmented correlation-consistent aug-cc-pVXZ variety<sup>43</sup>) by two cardinal numbers. However, the approach of Ref. 42 is not very practical as the complete amplitude optimization scales like the eighth power of the system size and sometimes suffers from numerical instabilities. Therefore, our collaborators and we have recently proposed<sup>44</sup> several approximate *Ansätze* for  $E_{\text{disp}}^{(20)}$ -F12 including the optimized diagonal *Ansatz* (ODA, defined precisely in the next section) that scales like  $N^5$  just like conventional  $E_{\text{disp}}^{(20)}$ , is numerically stable, and produces F12 dispersion energies nearly as accurate as the full F12 amplitude optimization. In the same work<sup>44</sup>, we have employed the F12 dispersion amplitudes (computed using any *Ansatz*) in the calculation of an explicitly correlated exchange dispersion energy,  $E_{\text{exch-disp}}^{(20)}$ -F12 (within the  $S^2$  approximation), substantially improving the convergence of conventional  $E_{\text{exch-disp}}^{(20)}$  to its CBS limit.

The F12 extensions to  $E_{\text{disp}}^{(20)}$  and  $E_{\text{exch-disp}}^{(20)}$  proposed in Refs. 42 and 44 significantly improve the basis set convergence of these SAPT0 corrections. However, in the original implementations, they are not competitive with conventional  $E_{\text{disp}}^{(20)}$  and  $E_{\text{exch-disp}}^{(20)}$  calculations in larger basis sets. The reason for the suboptimal performance of these algorithms, besides the fact that only modest effort was put into optimizing the initial proof-of-concept implementations, is the necessity to evaluate and transform multiple types of two-electron integrals over functions from the atomic-orbital (AO) basis set and the complementary auxiliary basis set (CABS) approximately spanning the orthogonal complement to the AO space<sup>45</sup>. The union of the AO and CABS bases is meant to provide a reasonably complete representation of the entire one-electron space so that all three- and four-electron integrals can be expressed by two-electron integrals using resolution of identity (RI). A further increase in the efficiency of the F12 approaches has been possible thanks to expanding the two-electron integrals and other four-index quantities in three-index auxiliary tensors using density fitting (DF)<sup>46–49</sup>. Density-fitted electronic structure theories benefit from their significantly lowered CPU time and storage requirements as the computation and transformation of two-electron integrals is avoided (even though the formal scaling of the underlying approach is rarely reduced). It should be noted that while the terms RI and DF are often used interchangeably in other contexts, they have separate and well-defined meanings within the F12 formalism. Both RI and DF have been the essential factors in turning MP2-F12 and approximate CCSD(T)-F12<sup>50–52</sup> into robust practically useful approaches that combine improved basis set convergence with only a modest increase in the computational cost relative to the parent MP2 and CCSD(T) methods. On the contrary, our initial implementation of  $E_{\text{disp}}^{(20)}$ -F12 and  $E_{\text{exch-disp}}^{(20)}$ -F12<sup>44</sup> uses RI but not DF; therefore, while the CBS convergence of these SAPT terms is vastly improved, the computational complexity is substantially higher than for the parent  $E_{\text{disp}}^{(20)}$  and  $E_{\text{exch-disp}}^{(20)}$  corrections due to the need to compute and transform several types of two-electron integrals. Moreover, these integrals involve not just AO basis indices, but up to two indices running over the RI basis that is typically several times larger than the AO set. As a result, the  $E_{\text{disp}}^{(20)}$ -F12 and  $E_{\text{exch-disp}}^{(20)}$ -F12 implementations cannot be competitive with their conventional SAPT0 counterparts until all four-index quantities present are decomposed by density fitting. This DF decomposition of the  $E_{\text{disp}}^{(20)}$ -F12 and  $E_{\text{exch-disp}}^{(20)}$ -F12 expressions, leading to an efficient and practical computer implementation, is the topic of this manuscript.

Before we proceed with the development of the density fitted SAPT-F12 expressions, it is important to note, perhaps counterintuitively, that a stand-alone CBS convergence improvement of the dispersion and exchange-dispersion terms is likely to make the total SAPT0 interaction energies worse, not better. Just like supermolecular MP2 tends to

overestimate the binding, especially in aromatic complexes, the MP2-level  $E_{\text{disp}}^{(20)}$  approximation tends to overestimate the magnitude of the true dispersion energy. As a result, it is usually preferable to perform SAPT0 calculations far away from CBS, so that the overbinding of the SAPT0 dispersion partially cancels with the underbinding due to an incomplete basis set. The “calendar” basis set jun-cc-pVDZ<sup>53</sup> has been observed to provide a particularly consistent error cancellation and is the recommended choice for SAPT0 calculations<sup>54</sup>. The replacement of small-basis  $E_{\text{disp}}^{(20)}$  and  $E_{\text{exch-disp}}^{(20)}$  values by their F12 counterparts (which are much closer to the CBS limit) disturbs the error cancellation and thus is likely to diminish the accuracy of SAPT0 interaction energies. Instead, the true utility of the  $E_{\text{disp}}^{(20)}$ -F12 and  $E_{\text{exch-disp}}^{(20)}$ -F12 corrections is in higher-level SAPT calculations, where the intramolecular correlation effects (in the form of  $E_{\text{disp}}^{(21)} + E_{\text{disp}}^{(22)}$  or, even better, the coupled-cluster-level dispersion energy<sup>27</sup>) prevent the overestimation of dispersion. While an F12 *Ansatz* for any intramolecular correlation terms in dispersion energy has not been proposed yet (except for  $E_{\text{disp}}^{(21)}$  in the specific case of interactions between two-electron systems<sup>38–40</sup>), one can combine an F12-enhanced CBS estimate of the leading-order  $E_{\text{disp}}^{(20)}$  term with an intramolecular correlation correction computed conventionally in a moderate basis set. Such a “composite” treatment, in the spirit of the popular MP2/CBS+ $\delta$ CCSD(T) approach to supermolecular interaction energies (with a moderate-basis  $\delta$ CCSD(T)=CCSD(T)–MP2 correction added on top of the MP2 CBS limit<sup>55,56</sup>) or the many successful composite approaches to thermochemistry<sup>57,58</sup>, is completely rigorous (no double counting occurs) and is likely to improve benchmark high-level SAPT dispersion energies and the resulting total SAPT interaction energies. For the latter, intramolecular correlation corrections to dispersion are included in the SAPT2+, SAPT2+(3), and SAPT2+3 levels of theory (as defined in Ref. 54), and these high-order SAPT variants are the primary target for improvement by our F12 approach to  $E_{\text{disp}}^{(20)}$  and  $E_{\text{exch-disp}}^{(20)}$ .

In this work, we derive and implement efficient expressions for the  $E_{\text{disp}}^{(20)}$ -F12(ODA) and  $E_{\text{exch-disp}}^{(20)}$ -F12(ODA) explicitly correlated SAPT0 corrections employing *robust* density fitting<sup>48,49</sup> in all two-electron quantities. The initial implementation of the new expressions (just like for the non-DF formulas of Ref. 44) is facilitated by the PSI4NUMPY framework<sup>59</sup> that combines the low-level functionality (integrals, HF vectors, ...) of the PSI4 electronic structure program<sup>60,61</sup> with the tensor manipulation and linear algebra capabilities of the NUMPY library. In the near future, the new DF- $E_{\text{disp}}^{(20)}$ -F12(ODA) and DF- $E_{\text{exch-disp}}^{(20)}$ -F12(ODA) functionality will be implemented in the development version of the PSI4 code itself. The new DF formalism enables us to extend benchmark studies of the explicitly correlated SAPT dispersion and exchange dispersion to substantially larger systems and basis sets, and we first exploit this capability by extending the SAPT-F12 calculations<sup>42,44</sup> for the A24 database of noncovalent complexes<sup>62</sup> to larger basis sets, including bases with midbond functions. Then, we investigate the CBS convergence of  $E_{\text{disp}}^{(20)}$  and  $E_{\text{exch-disp}}^{(20)}$ , and the effect of its F12 enhancement, on the total SAPT2+, SAPT2+(3), and SAPT2+3 interaction energies, on the same noncovalent databases as the SAPT benchmarking study of Ref. 54. Such an elimination of the leading basis set incompleteness error in high-level SAPT data is expected to provide a clearer picture of the directions for the further improvement of the SAPT methodology, both for the dispersion energy and for other terms.

## II. THEORY

In Sections II A and II B we summarize the main equations which define the non-DF  $E_{\text{disp}}^{(20)}$ -F12(ODA) and  $E_{\text{exch-disp}}^{(20)}$ -F12(ODA) energies, respectively. In Section II C the DF approximation is applied to obtain the corresponding expressions for DF-SAPT-F12.

Throughout this paper, the index convention from Ref. 44 is adopted. Thus, separate sets of indices are defined for monomers A and B, that is  $(i, k, m), a, r, x, \alpha$  and  $(j, l, n), b, s, y, \beta$ , respectively. The range of each orbital index is as follows: indices  $i, j, k, l, m, n$  run over occupied orbitals;  $a, b$ , virtual (unoccupied) orbitals;  $r, s$ , all molecular orbitals (MOs), both occupied and virtual;  $x, y$ , the complementary auxiliary (CA) functions approximately spanning the orthogonal complement of the MO space;  $\alpha, \beta$ , the formally complete orthonormal set, which is constructed as the union of the MO and CA subspaces; and  $A, B$ , the DF basis set. We will assume that a full dimer basis set is employed for both the orbital and CABS bases: as a result, the orbitals  $r$  and  $s$  span the same space and so do  $x$  and  $y$ . A summation over each repeated index is implied in all expressions. The index notation is summarized in Table I for easy reference.

TABLE I. Orbital spaces used in this work.

| Orbital space                    | Monomer A | Monomer B |
|----------------------------------|-----------|-----------|
| Occupied orbitals                | $i, k, m$ | $j, l, n$ |
| Virtual orbitals                 | $a$       | $b$       |
| Any molecular orbitals           | $r$       | $s$       |
| Complementary auxiliary orbitals | $x$       | $y$       |
| Complete orthonormal or RI basis | $\alpha$  | $\beta$   |
| DF basis                         | $A, B$    |           |

### A. $E_{\text{disp}}^{(20)}$ -F12(ODA)

The starting point for the development of the explicitly correlated dispersion energy is a “dispersion-only” Hylleraas functional

$$J_{\text{disp}}[\chi] = \langle \chi | \mathcal{Q}_A \mathcal{Q}_B (F^A + F^B - E_A^0 - E_B^0) \mathcal{Q}_A \mathcal{Q}_B | \chi \rangle + \langle \chi | \mathcal{Q}_A \mathcal{Q}_B (V - \langle V \rangle) | \phi_{\text{HF}}^A \phi_{\text{HF}}^B \rangle + \langle \phi_{\text{HF}}^A \phi_{\text{HF}}^B | (V - \langle V \rangle) \mathcal{Q}_A \mathcal{Q}_B | \chi \rangle \quad (1)$$

where  $F^X$  is the Fock operator,  $E_X^0$  is the sum of the occupied orbital energies  $\epsilon_{i(j)}^X$ ,  $\langle V \rangle \equiv \langle \phi_{\text{HF}}^A \phi_{\text{HF}}^B | V | \phi_{\text{HF}}^A \phi_{\text{HF}}^B \rangle$  is the first-order electrostatic correction  $E_{\text{elst}}^{(10)}$ , and  $\mathcal{Q}_X$  is the operator projecting out the ground state for a given monomer  $X \in \{A, B\}$ . It has been shown<sup>42,44</sup> that  $E_{\text{disp}}^{(20)}$ -F12 is obtained variationally by minimizing the above functional using a trial function

$$\chi = T_{ab}^{ij} |\Phi_{ij}^{ab}\rangle + T_{kl}^{ij} \mathcal{F}_{\alpha\beta}^{kl} |\Phi_{ij}^{\alpha\beta}\rangle \quad (2)$$

where  $|\Phi_{ij}^{\alpha\beta}\rangle = \hat{E}_{\alpha i} \hat{E}_{\beta j} |\phi_{\text{HF}}^A \phi_{\text{HF}}^B\rangle$  is a doubly excited (once on A, once on B) determinant. The target of F12 methods is to include the full space of doubly excited configurations, which is achieved via a set of amplitudes  $T_{kl}^{ij}$  and a suitable internal contraction

$$\mathcal{F}_{\alpha\beta}^{kl} = \langle kl | \hat{F}_{12} \hat{Q}_{12} | \alpha\beta \rangle \quad (3)$$

where  $\hat{F}_{12}$  is a correlation factor and  $\hat{Q}_{12}$  is the strong-orthogonality projector. A correlation factor plays a central role in the F12 methods since it introduces the explicit dependence of the function on the interelectronic distance  $r_{12}$ . Here, the standard exponential expression is assumed<sup>63</sup>

$$\hat{F}_{12} \equiv F(r_{12}) = \exp(-\gamma r_{12}) \quad (4)$$

where  $\gamma$  stands for a length-scale parameter. The operator  $\hat{Q}_{12}$  is chosen as *Ansatz* 3 of Ref. 10,

$$\hat{Q}_{12} = \left( 1_1 - \sum_i |i\rangle \langle i|_1 \right) \left( 1_2 - \sum_j |j\rangle \langle j|_2 \right) \left( 1_{12} - \sum_{ab} |ab\rangle \langle ab|_{12} \right) \quad (5)$$

with the subscripts indicating which electron coordinates are affected by a given part of the projector.

The final expression for the Hylleraas dispersion functional, valid for arbitrary amplitudes  $T_{ij}^{ab}$  and  $T_{ij}^{kl}$ , reads

$$J_{\text{disp}}[\chi] = 4T_{ij}^{ab} T_{ab}^{ij} (\epsilon_a^A + \epsilon_b^B - \epsilon_i^A - \epsilon_j^B) + 8T_{ab}^{ij} K_{ij}^{ab} + 4T_{ij}^{kl} T_{mn}^{ij} B_{kl,mn} - 4(\epsilon_i^A + \epsilon_j^B) T_{ij}^{kl} T_{mn}^{ij} X_{kl,mn} + 8T_{ij}^{kl} V_{kl}^{ij} + 8T_{ij}^{ab} T_{kl}^{ij} C_{ab}^{kl} \quad (6)$$

where a handful of intermediates were introduced analogously to the MP2-F12(3C) approach<sup>10</sup>

$$K_{ij}^{ab} = \langle ab | r_{12}^{-1} | ij \rangle \quad (7)$$

$$V_{kl}^{ij} = \langle ij | r_{12}^{-1} \hat{Q}_{12} \hat{F}_{12} | kl \rangle \quad (8)$$

$$B_{kl,mn} = \langle kl | \hat{F}_{12} \hat{Q}_{12} (\hat{f}_{A1} + \hat{f}_{B2}) \hat{Q}_{12} \hat{F}_{12} | mn \rangle \quad (9)$$

$$X_{kl,mn} = \langle kl | \hat{F}_{12} \hat{Q}_{12} \hat{F}_{12} | mn \rangle \quad (10)$$

$$C_{ab}^{kl} = \langle kl | \hat{F}_{12} \hat{Q}_{12} (\hat{f}_{A1} + \hat{f}_{B2}) | ab \rangle \quad (11)$$

Note that in the above equations, the Fock operators  $\hat{f}_1$  and  $\hat{f}_2$  for the first and second electron, respectively, pertain to different molecules. The presence of the projector operator  $\hat{Q}_{12}$  in Eqs. (8)–(11) leads to three- and four-electron integrals; thus, RI is employed to factorize these terms into products of two-electron integrals. In addition, we assumed the generalized Brillouin condition (GBC) and the so-called approximation 3C<sup>10</sup> to evaluate the  $C$  and  $B$  matrix elements, containing products of the correlation factor  $\hat{F}_{12}$  and the Fock operators.

The conventional ( $T_{ij}^{ab}$ ) and explicitly correlated ( $T_{ij}^{kl}$ ) dispersion amplitudes are determined by minimizing Eq. (6). Since the full minimization comes at a cost of  $N^8$ , too expensive to be practical, some approximations were proposed and tested in Ref. 44. In this work, one of these approximations, the Optimized Diagonal Ansatz (ODA)<sup>44</sup> is applied, which assumes that the  $T_{ij}^{kl}$  amplitudes are diagonal, that is,

$$T_{ij}^{kl} = T_{ij}^{ij} \delta_{ik} \delta_{jl}. \quad (12)$$

Within the ODA approximation, the dispersion amplitudes are computed from the following equations<sup>44</sup> (with the summations written explicitly this time)

$$T_{ij}^{ab} = \frac{K_{ij}^{ab} + T_{ij}^{ij} C_{ab}^{ij}}{\epsilon_i^A + \epsilon_j^B - \epsilon_a^A - \epsilon_b^B} \quad (13)$$

$$T_{ij}^{ij} \left[ B_{ij,ij} - (\epsilon_i^A + \epsilon_j^B) X_{ij,ij} + \sum_{ab} \frac{C_{ab}^{ij} C_{ab}^{ij}}{\epsilon_i^A + \epsilon_j^B - \epsilon_a^A - \epsilon_b^B} \right] = -V_{ij}^{ij} - \sum_{ab} \frac{K_{ij}^{ab} C_{ab}^{ij}}{\epsilon_i^A + \epsilon_j^B - \epsilon_a^A - \epsilon_b^B} \quad (14)$$

The ODA formalism in  $E_{\text{disp}}^{(20)}$ -F12 is closely related to the original approach to explicitly correlated MP2<sup>8,64</sup> which also retained only diagonal F12 amplitudes. Such a selection is not invariant with respect to a unitary transformation of occupied orbitals. Thus, our formalism is restricted to canonical HF orbitals on each monomer; however, the same restriction holds for conventional  $E_{\text{disp}}^{(20)}$  and  $E_{\text{exch-disp}}^{(20)}$  as well, so this is not a practical problem. Since the occupied orbitals in SAPT pertain to one monomer at a time, our ODA approach most closely relates to supermolecular MP2-F12 with an orbital-variant Ansatz using localized orbitals, which demonstrated good performance, a lack of geminal basis set superposition error, and a correct long-range behavior in the early study of weak MP2-F12 interaction energies<sup>64</sup>. In supermolecular F12 calculations, orbital invariance is commonly assured by means of a diagonal Ansatz with the explicitly correlated amplitudes fixed by cusp conditions<sup>65</sup>. We have previously tested a similar fixed-amplitude Ansatz (with an optimized common amplitude) in  $E_{\text{disp}}^{(20)}$ -F12 and  $E_{\text{exch-disp}}^{(20)}$ -F12<sup>44</sup>; however, the resulting accuracy, while clearly superior to conventional  $E_{\text{disp}}^{(20)}$  and  $E_{\text{exch-disp}}^{(20)}$ , did not match the accuracy of ODA. As both Ansätze exhibit virtually the same computational cost, only the more accurate ODA variant was used in the current work.

## B. $E_{\text{exch-disp}}^{(20)}$ -F12(ODA)

The second-order exchange dispersion correction in the F12 formalism decomposes into two terms. The first one is evaluated using the conventional exchange-dispersion formula<sup>25</sup> involving the amplitudes  $T_{ab}^{ij}$  (which, in this case, are obtained from the ODA expression (13)). The second term is an explicitly correlated correction involving the  $T_{kl}^{ij}$

amplitudes<sup>44</sup>:

$$\begin{aligned}
\delta E_{\text{exch-disp}}^{(20)}\text{-F12} = & 2T_{kl}^{ij} \left[ F_{xb}^{kl} K_{ij}^{xb'} S_i^{b'} S_{b'}^{i'} - 2F_{xb}^{kl} K_{ij}^{xb'} S_i^{b'} S_{b'}^{i'} + F_{ay}^{kl} K_{ij}^{a'y} S_j^a S_{a'}^{j'} - 2F_{ay}^{kl} K_{ij}^{a'y} S_j^a S_{a'}^{j'} \right. \\
& - 2F_{xb}^{kl} (\omega_B)_i^x S_i^b S_j^{i'} + F_{xb}^{kl} (\omega_B)_{i'}^x S_i^b S_j^{i'} - F_{ay}^{kl} (\omega_B)_i^y S_j^a \\
& - 2F_{ay}^{kl} (\omega_A)_j^y S_j^a S_i^{j'} + F_{ay}^{kl} (\omega_A)_j^y S_j^a S_i^{j'} - F_{xb}^{kl} (\omega_A)_j^x S_i^b \\
& - (K^F)_{i'j'}^{kl} S_i^{j'} S_j^{i'} + 2(K^F)_{ij'}^{kl} S_i^{j'} S_j^{i'} + 2(K^F)_{ij}^{kl} S_i^{j'} S_j^{i'} - (K^F)_{ij}^{lk} \\
& + F_{i'j'}^{lk} K_{ij}^{i'j'} + F_{aj'}^{lk} K_{ij}^{aj'} + F_{i'b}^{lk} K_{ij}^{i'b} + F_{xj'}^{lk} K_{ij}^{xj'} + F_{i'y}^{lk} K_{ij}^{i'y} \\
& + F_{xn}^{kl} K_{ij}^{ax} S_a^n + F_{my}^{kl} K_{ij}^{yb} S_b^m + F_{rs}^{kl} K_{ij}^{ab} S_a^r S_b^s \\
& + F_{xn}^{kl} K_{i'j'}^{xn} S_i^{j'} S_j^{i'} + F_{my}^{kl} K_{i'j'}^{my} S_i^{j'} S_j^{i'} + F_{rs}^{kl} K_{i'j'}^{rs} S_i^{j'} S_j^{i'} \\
& - 2F_{xn}^{kl} K_{ij}^{xn} S_i^{j'} S_j^{i'} - 2F_{my}^{kl} K_{ij}^{my} S_i^{j'} S_j^{i'} - 2F_{rs}^{kl} K_{ij}^{rs} S_i^{j'} S_j^{i'} \\
& \left. - 2F_{xn}^{kl} K_{i'j}^{xn} S_i^{j'} S_j^{i'} - 2F_{my}^{kl} K_{i'j}^{my} S_i^{j'} S_j^{i'} - 2F_{rs}^{kl} K_{i'j}^{rs} S_i^{j'} S_j^{i'} \right] \quad (15)
\end{aligned}$$

where  $F_{\alpha\beta}^{kl} = \langle kl | \hat{F}_{12} | \alpha\beta \rangle$  and  $(K^F)_{\alpha\beta}^{kl} = \langle kl | \hat{F}_{12} r_{12}^{-1} | \alpha\beta \rangle$  are two common matrix elements involving the correlation factor,  $S_j^i = \langle i | j \rangle$  is the overlap integral and  $(\omega_B)_i^a$  is the matrix element of the electrostatic potential of monomer B, that is,

$$(\omega_B)_i^a = \langle a | v_B | i \rangle + 2K_{ij}^{aj}, \quad (16)$$

and  $v_B$  is the nuclear potential of molecule B (the  $(\omega_A)_j^b$  matrix elements are defined similarly). All integral types in Eq. (15) already appear in the calculation of intermediates for the  $E_{\text{disp}}^{(20)}$ -F12 energy, however, the integrals containing  $\hat{F}_{12}$  and  $\hat{F}_{12} r_{12}^{-1}$  can now be of the exchange type:  $F_{AB}^{BA}$ . Moreover, the  $T_{ij}^{ab}$  and  $T_{ij}^{kl}$  amplitudes are obtained from the preceding  $E_{\text{disp}}^{(20)}$ -F12 calculations (Eqs. (13) and (14)); thus, in the ODA approach employed in this work, the  $T_{ij}^{kl}$  tensor is diagonal.

### C. Robust density-fitted formulas for $E_{\text{disp}}^{(20)}$ -F12 and $E_{\text{exch-disp}}^{(20)}$ -F12

We now employ density fitting (DF)<sup>48,49</sup> to decompose the 4-index integrals into contractions of 3- and 2-index quantities, thereby reducing the computational cost as well as storage requirements. The main idea of DF is replacing the one-particle orbital product densities  $|pq\rangle$  by approximated densities  $|\tilde{p}q\rangle$ . The latter are expanded in a set of auxiliary functions (denoted as  $A$ ) as follows

$$|pq\rangle \approx |\tilde{p}q\rangle = D_{pq}^A |A\rangle \quad (17)$$

The expansion coefficients  $D_{pq}^A$  can be obtained by minimizing the difference between the actual and fitted product densities<sup>46,47</sup>

$$\Delta_{pq}^w = (pq - \tilde{p}q | \hat{w}_{12} | pq - \tilde{p}q), \quad (18)$$

with a suitable positive definite  $\hat{w}_{12}$  metric. It has been demonstrated<sup>48</sup> that the Coulomb operator  $r_{12}^{-1}$  is the most convenient metric for fitting not only two-electron integrals, but also other integrals of the F12 theory, hence it will be used herein. Once the Coulomb metric is applied, the fitting coefficients are given by

$$D_{pq}^A = [\mathbf{J}^{-1}]_{AB} (B | r_{12}^{-1} | pq) \quad (19)$$

where  $[\mathbf{J}^{-1}]$  is the inverse of the two-center electron repulsion integral (ERI) matrix

$$[\mathbf{J}]_{AB} = \int A(\mathbf{r}_1) \frac{1}{r_{12}} B(\mathbf{r}_2) d\mathbf{r}_1 d\mathbf{r}_2 = (A | r_{12}^{-1} | B) \quad (20)$$

and  $(B | r_{12}^{-1} | pq)$  are the three-center ERIs. A crucial goal of density fitting is avoiding errors in integrals that are linear in the density errors. This is achieved by the so-called *robust* fit proposed by Dunlap<sup>47,48</sup>:

$$(pq | \hat{v}_{12} | rs)_{\text{robust}} \approx (\tilde{p}q | \hat{v}_{12} | rs) + (pq | \hat{v}_{12} | \tilde{r}s) - (\tilde{p}q | \hat{v}_{12} | \tilde{r}s) \quad (21)$$

where  $\hat{v}_{12}$  is any particular kernel operator for an integral. Equation (21) allows us to obtain a fitting error in the integrals that is quadratic with respect to the fitting error in the densities, regardless of whether  $\hat{v}_{12}$  and  $\hat{w}_{12}$  are the same or different<sup>48,49</sup>. When applying the DF algorithm to ordinary ERIs, they become approximated as

$$(pq|r_{12}^{-1}|rs) \approx (\tilde{p}q|r_{12}^{-1}|\tilde{r}s) = J_{pq}^A D_{rs}^A \quad (22)$$

with the  $D$  intermediate defined in Eq. (19), and the  $J$  intermediate computed as

$$J_{pq}^A = (A|r_{12}^{-1}|pq) \quad (23)$$

In the DF-SAPT-F12 theory, in addition to ERIs, we need to fit four types of explicitly correlated two-electron integrals with the operators  $\hat{F}_{12}$ ,  $\hat{F}_{12}^2$ ,  $\hat{F}_{12}r_{12}^{-1}$ , and the double commutator between the kinetic energy ( $\hat{t}_1 + \hat{t}_2$ ) and the correlation factor,  $[[\hat{F}_{12}, \hat{t}_1 + \hat{t}_2], \hat{F}_{12}]$ <sup>66</sup>. To do so, the explicitly robust formulas (Eq. (21)) combined with the Coulomb metric are employed. This leads to several additional intermediates that we need to form before we evaluate  $E_{\text{disp}}^{(20)}$ -F12(ODA) (Eqs. (7)–(11)) as well as the  $\delta E_{\text{exch-disp}}^{(20)}$ -F12(ODA) correction to the second-order exchange dispersion energy (Eq. (15)) using the DF approximation:

$$J_{AB} = (A|r_{12}^{-1}|B) \quad (24)$$

$$F_{AB} = (A|\hat{F}_{12}|B) \quad (25)$$

$$K_{AB}^F = (A|\hat{F}_{12}r_{12}^{-1}|B) \quad (26)$$

$$F_{AB}^2 = (A|\hat{F}_{12}^2|B) \quad (27)$$

$$U_{AB}^F = (A|[[\hat{F}_{12}, \hat{t}_1 + \hat{t}_2], \hat{F}_{12}]|B) \quad (28)$$

$$F_{ij}^A = (A|\hat{F}_{12}|ij) \quad (29)$$

$$K_{A,ij}^F = (A|\hat{F}_{12}r_{12}^{-1}|ij) \quad (30)$$

$$F_{A,ij}^2 = (A|\hat{F}_{12}^2|ij) \quad (31)$$

$$U_{A,ij}^F = (A|[[\hat{F}_{12}, \hat{t}_1 + \hat{t}_2], \hat{F}_{12}]|ij) \quad (32)$$

These definitions allow us to rewrite the  $V$ ,  $X$ ,  $B$  and  $C$  matrix elements, Eqs. (8)–(11), explicitly. The resulting formulas contain many instances of the  $F_{\alpha'\beta'}^{\alpha\beta}$  matrix element for different types of indices  $\alpha, \alpha', \beta, \beta'$ . To keep the expressions reasonably compact, we will list this intermediate in its non-DF, four-index form. In the actual evaluation of these expressions, the robust density-fitted form of  $F_{\alpha'\beta'}^{\alpha\beta}$  is always used:

$$F_{\alpha'\beta'}^{\alpha\beta} = D_{\alpha\alpha'}^A F_{\beta\beta'}^A + F_{\alpha\alpha'}^A D_{\beta\beta'}^A - D_{\alpha\alpha'}^A F_{AB} D_{\beta\beta'}^B \quad (33)$$

Note that, for optimal scaling and computational efficiency, this evaluation requires treating the three terms resulting from Eq. (33) separately, and optimizing the order of contractions between the individual two- and three-index tensors, both those forming a part of  $F_{\alpha'\beta'}^{\alpha\beta}$  and those arising from other matrices. The resulting expressions are

$$V_{kl}^{ij} = D_{ik}^A K_{A,jl}^F + K_{A,ik}^F D_{jl}^A - D_{ik}^A K_{AB}^F D_{jl}^B - J_{ir}^A D_{js}^A F_{rs}^{kl} - J_{ix}^A D_{jn}^A F_{xn}^{kl} - J_{im}^A D_{jy}^A F_{my}^{kl} \quad (34)$$

$$X_{mn}^{kl} = D_{km}^A F_{A,ln}^2 + F_{A,km}^2 D_{ln}^A - D_{km}^A F_{AB}^2 D_{ln}^B - F_{rs}^{kl} F_{mn}^{rs} - F_{xj}^{kl} F_{mn}^{xj} - F_{iy}^{kl} F_{mn}^{iy} \quad (35)$$

$$C_{ab}^{kl} = f_{ax}^A F_{xb}^{kl} + F_{ay}^{kl} f_{yb}^B \quad (36)$$

$$B_{kl,mn} = \frac{1}{2}(A_{kl,mn} + A_{mn,kl}) \quad (37)$$



$$\begin{aligned}
A_{kl,mn} = & D_{km}^A U_{A,ln}^F + U_{A,km}^F D_{ln}^A - D_{km}^A U_{AB}^F D_{ln}^B \\
& + (D_{xm}^A (f_{xk}^A + k_{xk}^A) + D_{rm}^A (f_{rk}^A + k_{rk}^A)) F_{A,ln}^2 + (F_{A,xm}^2 (f_{xk}^A + k_{xk}^A) + F_{A,rm}^2 (f_{rk}^A + k_{rk}^A)) D_{ln}^A \\
& - (D_{xm}^A (f_{xk}^A + k_{xk}^A) + D_{rm}^A (f_{rk}^A + k_{rk}^A)) F_{AB}^2 D_{ln}^B \\
& + D_{km}^A (F_{A,yn}^2 (f_{yl}^B + k_{yl}^B) + F_{A,sn}^2 (f_{sl}^B + k_{sl}^B)) + F_{A,km}^2 (D_{yn}^A (f_{yl}^B + k_{yl}^B) + D_{sn}^A (f_{sl}^B + k_{sl}^B)) \\
& - D_{km}^A F_{AB}^2 (D_{yn}^B (f_{yl}^B + k_{yl}^B) + D_{sn}^B (f_{sl}^B + k_{sl}^B)) \\
& - (D_{kr}^A k_{rx}^A + D_{kx'}^A k_{x'x}^A) F_{lb}^A F_{xb}^{mn} - (F_{kr}^A k_{rx}^A + F_{kx'}^A k_{x'x}^A) D_{lb}^A F_{xb}^{mn} + (D_{kr}^A k_{rx}^A + D_{kx'}^A k_{x'x}^A) F_{AB} D_{lb}^B F_{xb}^{mn} \\
& - D_{kx}^A (F_{ls}^A k_{sb}^B + F_{ly}^A k_{yb}^B) F_{xb}^{mn} - F_{kx}^A (D_{ls}^A k_{sb}^B + D_{ly}^A k_{yb}^B) F_{xb}^{mn} + D_{kx}^A F_{AB} (D_{ls}^B k_{sb}^B + D_{ly}^B k_{yb}^B) F_{xb}^{mn} \\
& - (D_{kr}^A k_{ra}^A + D_{kx}^A k_{xa}^A) F_{ly}^A F_{ay}^{mn} - (F_{kr}^A k_{ra}^A + F_{kx}^A k_{xa}^A) D_{ly}^A F_{ay}^{mn} + (D_{kr}^A k_{ra}^A + D_{kx}^A k_{xa}^A) F_{AB} D_{ly}^B F_{ay}^{mn} \\
& - D_{ka}^A (F_{ls}^A k_{sy}^B + F_{ly}^A k_{y'y}^B) F_{ay}^{mn} - F_{ka}^A (D_{ls}^A k_{sy}^B + D_{ly}^A k_{y'y}^B) F_{ay}^{mn} + D_{ka}^A F_{AB} (D_{ls}^B k_{sy}^B + D_{ly}^B k_{y'y}^B) F_{ay}^{mn} \\
& - (D_{kr}^A k_{rx}^A + D_{kx'}^A k_{x'x}^A) F_{ly}^A F_{xy}^{mn} - (F_{kr}^A k_{rx}^A + F_{kx'}^A k_{x'x}^A) D_{ly}^A F_{xy}^{mn} + (D_{kr}^A k_{rx}^A + D_{kx'}^A k_{x'x}^A) F_{AB} D_{ly}^B F_{xy}^{mn} \\
& - D_{kx}^A (F_{ls}^A k_{sy}^B + F_{ly}^A k_{y'y}^B) F_{xy}^{mn} - F_{kx}^A (D_{ls}^A k_{sy}^B + D_{ly}^A k_{y'y}^B) F_{xy}^{mn} + D_{kx}^A F_{AB} (D_{ls}^B k_{sy}^B + D_{ly}^B k_{y'y}^B) F_{xy}^{mn} \\
& - (D_{kr'}^A (f_{r'r}^A + k_{r'r}^A) + D_{kx}^A (f_{xr}^A + k_{xr}^A)) F_{ls}^A F_{rs}^{mn} - (F_{kr'}^A (f_{r'r}^A + k_{r'r}^A) + F_{kx}^A (f_{xr}^A + k_{xr}^A)) D_{ls}^A F_{rs}^{mn} \\
& + (D_{kr'}^A (f_{r'r}^A + k_{r'r}^A) + D_{kx}^A (f_{xr}^A + k_{xr}^A)) F_{AB} D_{ls}^B F_{rs}^{mn} \\
& - (D_{kr}^A (f_{rx}^A + k_{rx}^A) + D_{kx'}^A (f_{x'x}^A + k_{x'x}^A)) F_{lj}^A F_{xj}^{mn} - (F_{kr}^A (f_{rx}^A + k_{rx}^A) + F_{kx'}^A (f_{x'x}^A + k_{x'x}^A)) D_{lj}^A F_{xj}^{mn} \\
& + (D_{kr}^A (f_{rx}^A + k_{rx}^A) + D_{kx'}^A (f_{x'x}^A + k_{x'x}^A)) F_{AB} D_{lj}^B F_{xj}^{mn} \\
& - (D_{kr}^A (f_{ri}^A + k_{ri}^A) + D_{kx}^A (f_{xi}^A + k_{xi}^A)) F_{ly}^A F_{iy}^{mn} - (F_{kr}^A (f_{ri}^A + k_{ri}^A) + F_{kx}^A (f_{xi}^A + k_{xi}^A)) D_{ly}^A F_{iy}^{mn} \\
& + (D_{kr}^A (f_{ri}^A + k_{ri}^A) + D_{kx}^A (f_{xi}^A + k_{xi}^A)) F_{AB} D_{ly}^B F_{iy}^{mn} \\
& - D_{kr}^A (F_{ls'}^A (f_{s's}^B + k_{s's}^B) + F_{ly}^A (f_{ys}^B + k_{ys}^B)) F_{rs}^{mn} - F_{kr}^A (D_{ls'}^A (f_{s's}^B + k_{s's}^B) + D_{ly}^A (f_{ys}^B + k_{ys}^B)) F_{rs}^{mn} \\
& + D_{kr}^A F_{AB} (D_{ls'}^B (f_{s's}^B + k_{s's}^B) + D_{ly}^B (f_{ys}^B + k_{ys}^B)) F_{rs}^{mn} \\
& - D_{kx}^A (F_{ls}^A (f_{sj}^B + k_{sj}^B) + F_{ly}^A (f_{yj}^B + k_{yj}^B)) F_{xj}^{mn} - F_{kx}^A (D_{ls}^A (f_{sj}^B + k_{sj}^B) + D_{ly}^A (f_{yj}^B + k_{yj}^B)) F_{xj}^{mn} \\
& + D_{kx}^A F_{AB} (D_{ls}^B (f_{sj}^B + k_{sj}^B) + D_{ly}^B (f_{yj}^B + k_{yj}^B)) F_{xj}^{mn} \\
& - D_{ki}^A (F_{ls}^A (f_{sy}^B + k_{sy}^B) + F_{ly'}^A (f_{y'y}^B + k_{y'y}^B)) F_{iy}^{mn} - F_{ki}^A (D_{ls}^A (f_{sy}^B + k_{sy}^B) + D_{ly'}^A (f_{y'y}^B + k_{y'y}^B)) F_{iy}^{mn} \\
& + D_{ki}^A F_{AB} (D_{ls}^B (f_{sy}^B + k_{sy}^B) + D_{ly'}^B (f_{y'y}^B + k_{y'y}^B)) F_{iy}^{mn} \\
& - F_{ab}^{kl} f_{ax}^A F_{mn}^{xb} - F_{ab}^{kl} F_{mn}^{ay} f_{yb}^B
\end{aligned} \tag{38}$$

The  $f_{kx}^{\mathbf{X}}$  and  $k_{kx}^{\mathbf{X}}$  matrices in Eqs. (36) and (38) denote the matrix elements of the Fock and exchange operators for monomer  $\mathbf{X}$ , respectively. Note that the explicit symmetrization of Eq. (37) is not required in the ODA case as only the diagonal elements  $B_{kl,kl} = A_{kl,kl}$  are needed.

Analogously, the density-fitted explicitly correlated correction to the second-order exchange-dispersion energy,  $\delta E_{\text{exch-disp}}^{(20)}$  (Eq. (15)), is computed as

$$\begin{aligned}
\delta E_{\text{exch-disp}}^{(20)}\text{-F12} = & 2T_{kl}^{ij} \left[ F_{xb}^{kl} J_{xi'}^A D_{b'j}^A S_i^b S_{b'}^{i'} - 2F_{xb'}^{kl} J_{xi}^A D_{b'j}^A S_i^b S_{b'}^{i'} \right. \\
& + F_{ay}^{kl} J_{a'i}^A D_{yj}^A S_j^a S_{a'}^{j'} - 2F_{ay'}^{kl} J_{a'i}^A D_{yj}^A S_j^a S_{a'}^{j'} \\
& - 2F_{xb}^{kl} ((V_B)_i^x + 2J_{xi}^A D_{nn}^A) S_i^b S_j^{i'} + F_{xb}^{kl} ((V_B)_{i'}^x + 2J_{xi'}^A D_{nn}^A) S_i^b S_j^{i'} \\
& - F_{ay}^{kl} ((V_B)_i^y + 2J_{yi}^A D_{nn}^A) S_j^a S_i^{j'} - 2F_{ay}^{kl} ((V_A)_j^y + 2J_{mm}^A D_{yj}^A) S_j^a S_i^{j'} \\
& + F_{ay}^{kl} ((V_A)_{j'}^y + 2J_{mm}^A D_{yj'}^A) S_j^a S_i^{j'} - F_{xb}^{kl} ((V_A)_j^x + 2J_{mm}^A D_{xj}^A) S_i^b S_j^{i'} \\
& - (D_{ki'}^A K_{A,lj'}^F + K_{A,ki'}^F D_{lj'}^A - D_{ki'}^A K_{AB}^F D_{lj'}^B) S_i^{j'} S_j^{i'} \\
& + 2(D_{ki}^A K_{A,lj'}^F + K_{A,ki}^F D_{lj'}^A - D_{ki}^A K_{AB}^F D_{lj'}^B) S_i^{j'} S_j^{i'} \\
& + 2(D_{ki'}^A K_{A,lj}^F + K_{A,ki'}^F D_{lj}^A - D_{ki'}^A K_{AB}^F D_{lj}^B) S_i^{j'} S_j^{i'} \\
& - (D_{li}^A K_{A,kj}^F + K_{A,li}^F D_{kj}^A - D_{li}^A K_{AB}^F D_{kj}^B) \\
& + F_{i'j}^{lk} J_{i'i}^A D_{j'j}^A + F_{aj'}^{lk} J_{ai}^A D_{j'j}^A + F_{i'b}^{lk} J_{i'i}^A D_{bj}^A + F_{xj'}^{lk} J_{xi}^A D_{j'j}^A \\
& + F_{i'y}^{lk} J_{i'i}^A D_{yj}^A + F_{xn}^{kl} J_{ai}^A D_{xj}^A S_a^n + F_{my}^{kl} J_{yi}^A D_{bj}^A S_b^m \\
& + F_{rs}^{kl} J_{ai}^A D_{bj}^A S_a^s S_b^r + F_{xn}^{kl} J_{xi'}^A D_{nj'}^A S_i^{j'} S_j^{i'} + F_{my}^{kl} J_{mi'}^A D_{yj'}^A S_i^{j'} S_j^{i'} \\
& + F_{rs}^{kl} J_{ri'}^A D_{sj'}^A S_i^{j'} S_j^{i'} - 2F_{xn}^{kl} J_{xi}^A D_{nj'}^A S_i^{j'} S_j^{i'} - 2F_{my}^{kl} J_{mi}^A D_{yj'}^A S_i^{j'} S_j^{i'} \\
& - 2F_{rs}^{kl} J_{ri}^A D_{sj}^A S_i^{j'} S_j^{i'} - 2F_{xn}^{kl} J_{xi}^A D_{nj}^A S_i^{j'} S_j^{i'} - 2F_{my}^{kl} J_{mi}^A D_{yj}^A S_i^{j'} S_j^{i'} \\
& \left. - 2F_{rs}^{kl} J_{ri}^A D_{sj}^A S_i^{j'} S_j^{i'} \right] \tag{39}
\end{aligned}$$

One may notice that Eqs. (34)–(39) have many common intermediates, which are computed only once and reused. The computational cost of the  $V$ ,  $X$ ,  $B$  and  $C$  intermediates as well as of  $\text{DF-}\delta E_{\text{exch-disp}}^{(20)}\text{-F12}$  scales as  $N^5$ .

### III. COMPUTATIONAL DETAILS

The expressions for  $\text{DF-}E_{\text{disp}}^{(20)}\text{-F12(ODA)}$  and  $\text{DF-}E_{\text{exch-disp}}^{(20)}\text{-F12(ODA)}$  were implemented and tested within the PSI4NUMPY framework<sup>59</sup>. A production-level density-fitted SAPT0-F12 code is an ongoing project and will be available in the PSI4 program<sup>60,61</sup>.

It is a common practice to carry out the F12 calculations with three auxiliary basis sets in addition to the AO basis: the CABS set added to the AO one for the RI approximations, the DF basis in the calculation of correlated pair functions, and the JKFIT basis for the DF of Coulomb and exchange operators<sup>51</sup>. In this work, the orbital basis sets were the augmented correlation consistent aug-cc-pVXZ  $\equiv$  aXZ sets of Dunning and coworkers<sup>43,67</sup>, with  $X$  ranging from D to 6. The CABS bases were chosen as the aXZ-RIFIT sets, a.k.a. aXZ-MP2FIT<sup>68,69</sup>. The DF and JKFIT basis sets required for density fitting were also chosen as aXZ-MP2FIT<sup>68,69</sup>. We also examined the accuracy of the DF-SAPT-F12 calculations with the aXZ-JKFIT<sup>70</sup> sets utilized for fitting the Fock matrices, that is, with aXZ-MP2FIT as the DF set and aXZ-JKFIT as the JKFIT set. The cardinal number  $X$  of the auxiliary bases was equal or larger than the corresponding  $X$  for the orbital set. As we will write triplets such as aDZ/aDZ-RIFIT/aDZ-MP2FIT for the orbital/CABS/DF basis set combination, for clarity we will employ the notation aXZ-MP2FIT and aXZ-JKFIT rather than the more common aXZ/MP2FIT and aXZ/JKFIT. The F12 integrals were computed employing the exponential correlation factor  $\exp(-\gamma r_{12})$  with the length-scale parameter  $\gamma$  set to  $1.0 \text{ } a_0^{-1}$ . This correlation factor was fitted to a sum of 6 Gaussian terms<sup>9</sup>.

We observed that the density-fitting approximation introduces numerical instabilities in pair correlation energies for core orbitals, where the coefficient multiplying the amplitude  $T_{ij}^{ij}$  in Eq. (14) is very small, making an accurate determination of this amplitude challenging. While these instabilities were never an issue in the non-DF calculations with the ODA *Ansatz* in Ref. 44, we were not able to obtain reliable core-orbital  $T_{ij}^{ij}$  amplitudes with density fitting. Therefore, all calculations in this work were performed within the frozen core (FC) scheme. If all-electron (AE)

dispersion and exchange dispersion energies are required, one can combine the FC values of  $E_{\text{disp}}^{(20)}$ -F12 and  $E_{\text{exch-disp}}^{(20)}$ -F12 with the conventional (non-F12) contributions of pair correlation energies involving one or two core orbitals. Alternatively, the F12 pair correlation energies for core orbitals could be computed using the fixed-amplitude Ansatz from Ref. 44, which is more approximate than ODA but avoids the instability of Eq. (14). While we have never observed such instabilities in frozen-core DF- $E_{\text{disp}}^{(20)}$ -F12(ODA) calculations anywhere close to the van der Waals minimum separation, they do show up for some pair correlation energies at large distances so that the fixed-amplitude Ansatz might be preferable in this case. Fortunately, conventional  $E_{\text{disp}}^{(20)}$  is already quite accurate at these distances, and  $E_{\text{exch-disp}}^{(20)}$  is essentially zero. The FC approximation is the main reason for small discrepancies between the results presented below and the corresponding AE data of Ref. 44. The secondary reason for the discrepancies in  $E_{\text{exch-disp}}^{(20)}$ -F12 are the small errors in our original implementation of this correction<sup>44</sup> that led to minor differences (up to about 0.005 kcal/mol) with respect to the correct results – see the Erratum to Ref. 44 for more details.

The first class of systems tested are five small complexes in the same geometries as in our previous study<sup>44</sup>: He–He, Ne–Ne, Ar–Ar, H<sub>2</sub>O–H<sub>2</sub>O, and CH<sub>4</sub>–CH<sub>4</sub>. Subsequently, we performed a comprehensive analysis of DF-SAPT-F12 on the entire A24 benchmark database<sup>62</sup> with basis sets up to a5Z, comparing the results of the conventional and explicitly correlated dispersion and exchange dispersion energies. The SAPT0 calculations were carried out by means of the PSI4<sup>60,71</sup> quantum chemistry package with basis sets up to a6Z, with and without midbond functions, and utilized density fitting with the aXZ-MP2FIT auxiliary basis. The reference non-DF  $E_{\text{disp}}^{(20)}$ -F12(ODA) and  $E_{\text{exch-disp}}^{(20)}$ -F12(ODA) energies were computed using the FC scheme with all technical details presented in Ref. 44.

In this work, we increased the benchmark quality relative to Ref. 44. The CBS limits of  $E_{\text{disp}}^{(20)}$  and  $E_{\text{exch-disp}}^{(20)}$  were now obtained by the standard  $X^{-3}$  extrapolation technique<sup>7,72</sup>. Specifically, the conventional  $E_{\text{disp}}^{(20)}$  and  $E_{\text{exch-disp}}^{(20)}$  SAPT corrections were extrapolated from the a5Z atom-centered orbital basis augmented by a hydrogenic a5Z set of midbond functions, and the a6Z atom-centered orbital basis augmented by a hydrogenic a6Z set of midbond functions (such an extrapolation will be denoted as (a5Z+(a5Z),a6Z+(a6Z)). The uncertainty of the benchmark result was assumed as the absolute difference between the reference extrapolated value and the one computed in the larger of the two bases, that is, a6Z+(a6Z). In addition, to test the quality of the benchmark, we selected the water dimer and obtained an improved reference value by performing even larger calculations (also including midbond functions) using the aug-cc-pV7Z and aug-mcc-pV7Z<sup>73</sup> bases for the oxygen and hydrogen atoms, respectively. The reference value for this system employed in this large-basis convergence test (Figs. 1 and 2) was estimated by extrapolating the a6Z+(a6Z) and a7Z+(a7Z) results. However, whenever the water dimer is investigated as a part of the entire A24 dataset, we will continue using the (a5Z+(a5Z),a6Z+(a6Z)) benchmark for consistency: the improved septuple-zeta reference will only be employed when explicitly stated.

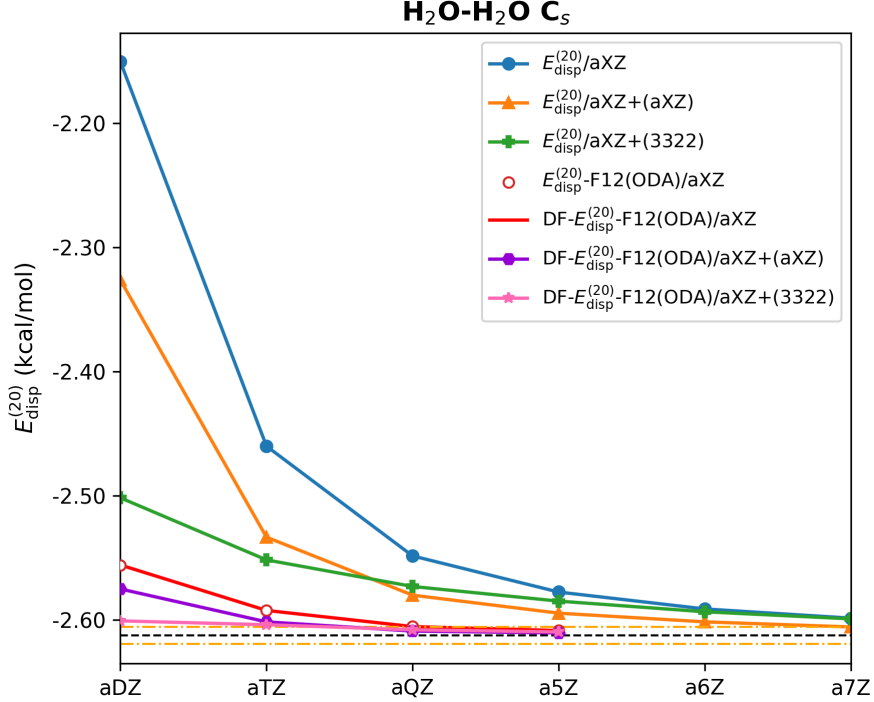
In addition to using midbond functions for benchmark calculations, their benefits for the convergence rate of DF- $E_{\text{disp}}^{(20)}$ -F12 and DF- $E_{\text{exch-disp}}^{(20)}$ -F12 were investigated for the five test case dimers and for all systems from the A24 database. In this work, we considered the constant (3s3p2d2f) set of midbond functions as well as a set of hydrogenic aXZ midbond functions with the cardinal number  $X$  that varies in accordance with the atomic basis set. Herein, the shorthand notation of aXZ+(3322) is utilized for the aXZ atom-centered basis set augmented with the (3s3p2d2f) bond functions, while the variable-midbond calculations are denoted by aXZ+(aXZ). The midbond function exponents for the (3s3p2d2f) set were (0.9, 0.3, 0.1) for sp and (0.6, 0.2) for df. The corresponding auxiliary basis, used in both the CABS and DF contexts, contains (1.8, 1.2, 0.6, 0.4, 0.2) exponents for spd functions, (1.5, 0.9, 0.5, 0.3) exponents for f, and (1.5, 0.9, 0.3) exponents for g<sup>74</sup>. The location of the midbond center was chosen as<sup>75</sup>

$$\mathbf{r}_{\text{bond}} = \frac{\sum_{a \in A} \sum_{b \in B} w_{ab} \frac{\mathbf{r}_a + \mathbf{r}_b}{2}}{\sum_{a \in A} \sum_{b \in B} w_{ab}} \quad w_{ab} = |\mathbf{r}_a - \mathbf{r}_b|^{-6} \quad (40)$$

where the summations run over all atoms  $a$  (located at  $\mathbf{r}_a$ ) in molecule  $A$  and all atoms  $b$  (located at  $\mathbf{r}_b$ ) in molecule  $B$ . Such an approach of placing midbonds has been recommended to avoid issues when one monomer is much longer than the other and the midpoint between the molecular centers of mass is still within one of the monomers<sup>76</sup>.

After examining the CBS convergence of the individual  $E_{\text{disp}}^{(20)}$  and  $E_{\text{exch-disp}}^{(20)}$  terms, in the last part of this manuscript, we illustrate how the inclusion of  $E_{\text{disp}}^{(20)}$ -F12 and  $E_{\text{exch-disp}}^{(20)}$ -F12 in place of standard  $E_{\text{disp}}^{(20)}$  and  $E_{\text{exch-disp}}^{(20)}$  affects the accuracy of total SAPT interaction energies. For this purpose, we build on the study of Ref. 54, which examined the errors of interaction energies computed using various levels of SAPT with respect to reference CCSD(T)/CBS data. Specifically, Ref. 54 investigated, in addition to SAPT0, the higher-level SAPT2, SAPT2+, SAPT2+(3), and SAPT2+3 variants including, to a various degree, the effects of intramolecular electron correlation. In addition, these higher-level SAPT treatments can be combined with a coupled cluster doubles (CCD) account of dispersion<sup>26,77</sup> in place of the conventional one based on double perturbation theory, and/or with a “ $\delta$ MP2” correction accounting for,

FIG. 1. Convergence of the frozen-core non-DF and DF  $E_{\text{disp}}^{(20)}$ -F12 results as a function of basis set for the  $\text{H}_2\text{O}-\text{H}_2\text{O}$  complex from the A24 database. The hydrogenic functions from the same aXZ orbital basis set or the constant (3s3p2d2f) set of functions are chosen for midbond functions. The reference value marked by a black dashed line was obtained at the  $E_{\text{disp}}^{(20)}/(\text{a6Z}+(\text{a6Z}),\text{a7Z}+(\text{a7Z}))$  extrapolated level. The yellow dashed lines indicate the benchmark uncertainty.



in particular, some higher-order couplings between induction and dispersion. The accuracy of all resulting variants of SAPT has been investigated in Ref. 54 using a benchmark dataset which is the union of the S22<sup>56,78</sup>, HBC6<sup>79</sup>, NBC10<sup>80</sup>, and HSG<sup>81</sup> databases. One should note that the datasets containing potential energy curves (that is, HBC6 and NBC10) were truncated in Ref. 54 to at most five points per system (the van der Waals minimum, two shorter separations, and two longer separations), and we impose exactly the same truncation. Moreover, the exchange-dispersion energies in Ref. 54 are scaled to approximately account for the effects missing in the single exchange approximation:

$$E_{\text{exch-disp}}^{(20)}(\text{scaled}) = E_{\text{exch-disp}}^{(20)}(S^2) \cdot \left( \frac{E_{\text{exch}}^{(10)}(\text{full})}{E_{\text{exch}}^{(10)}(S^2)} \right)^\alpha \quad (41)$$

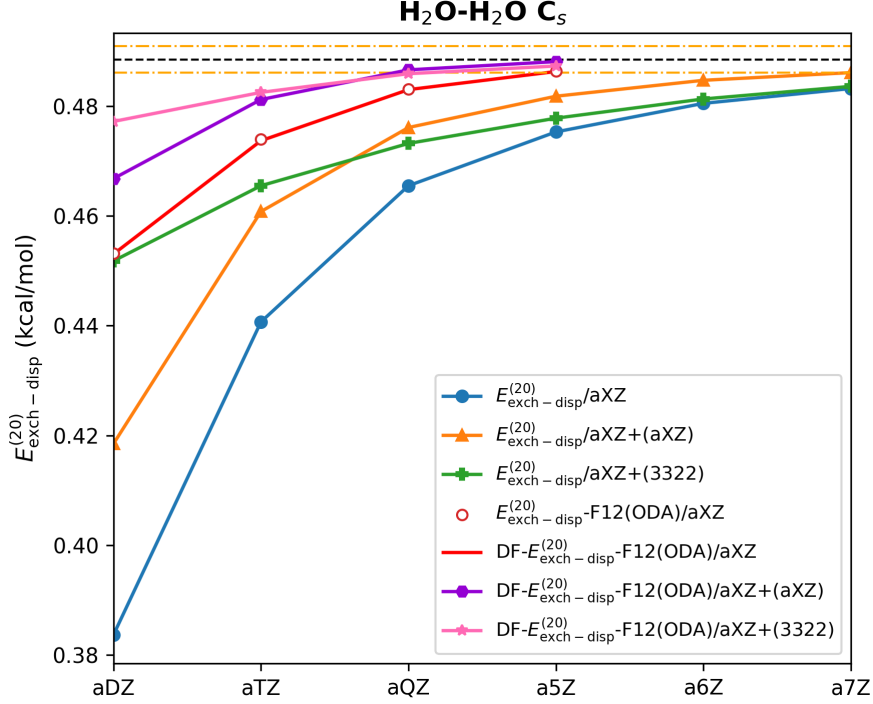
with  $\alpha = 1.0$ , and we will employ the same scaling for consistency. In this work, we consider an F12 version of each of the SAPT variants from Ref. 54, obtained from the parent method by replacing the sum  $E_{\text{disp}}^{(20)} + E_{\text{exch-disp}}^{(20)}(\text{scaled})$  by the  $E_{\text{disp}}^{(20)}\text{-F12} + E_{\text{exch-disp}}^{(20)}\text{-F12}(\text{scaled})$  value computed in the same orbital basis (aDZ or aTZ). To enable direct comparison with the error statistics of non-F12 SAPT levels presented in Ref. 54, we utilize exactly the same set of four benchmark databases. In fact, we did not even have to recalculate the non-F12 SAPT data ourselves, but imported the values obtained in Ref. 54 from the BFD database<sup>82</sup> and added the corresponding F12 corrections computed by us. As a result, the mean absolute error (MAE) values for total SAPT interaction energies, presented in the next section, are directly comparable to the MAE presented in Ref. 54.

## IV. RESULTS

### A. Convergence with the density-fitting basis set

A crucial step in the DF-SAPT-F12 development is ensuring that the density-fitting approximation does not introduce any significant errors. Following our previous study<sup>44</sup>, the set of five complexes: He-He, Ne-Ne, Ar-Ar,

FIG. 2. Convergence of the frozen-core non-DF and DF  $E_{\text{exch-disp}}^{(20)}$ -F12 results as a function of basis set for the  $\text{H}_2\text{O}-\text{H}_2\text{O}$  complex from the A24 database. The hydrogenic functions from the same aXZ orbital basis set or the constant (3s3p2d2f) set of functions are chosen for midbond functions. The reference value marked by a black dashed line was obtained at the  $E_{\text{exch-disp}}^{(20)}/(\text{a6Z}+(\text{a6Z}),\text{a7Z}+(\text{a7Z}))$  extrapolated level. The yellow dashed lines indicate the benchmark uncertainty.



$\text{H}_2\text{O}-\text{H}_2\text{O}$ , and  $\text{CH}_4-\text{CH}_4$  at their van der Waals minimum geometries was the subject of our preliminary tests. For these systems, we carried out the frozen-core DF- $E_{\text{disp}}^{(20)}$ -F12(ODA) and DF- $E_{\text{exch-disp}}^{(20)}$ -F12(ODA) calculations using the aXZ/aXZ-RIFIT ( $X=\text{D},\text{T},\text{Q},5$ ) basis sets for orbital/CABS in conjunction with the increasing cardinal number of the aYZ-MP2FIT ( $Y=\text{D},\text{T},\text{Q},5$ ) set for density fitting, with  $Y \geq X$ . In the case of the helium dimer, basis sets up to a6Z, a6Z-RIFIT, and a6Z-MP2FIT were used. The resulting DF and non-DF explicitly correlated  $E_{\text{disp}}^{(20)}$  and  $E_{\text{exch-disp}}^{(20)}$  energies are presented in Tables II and III for the water and methane dimers, respectively; the corresponding results for the rare gas dimers are given in the Supporting Information, which also contains tables displaying the convergence of the  $E_{\text{disp}}^{(20)} + E_{\text{exch-disp}}^{(20)}$  sum. We have also considered an alternative choice of auxiliary basis set for DF, in which the aYZ-MP2FIT set is used for the DF of all integrals except for those occurring in the Fock and exchange operators present in Eqs. (36) and (38), in which case the aYZ-JKFIT basis is used instead. The resulting explicitly correlated SAPT corrections are presented in the Supporting Information.

The convergence of the non-DF  $E_{\text{disp}}^{(20)}$ -F12 and  $E_{\text{exch-disp}}^{(20)}$ -F12 calculations for the same set of complexes has been thoroughly investigated in Ref. 44. In particular, it was found that while for the water and methane dimers the results are insensitive to the choice of CABS, the small CABS sets (especially aDZ-RIFIT) are not adequate for the noble gas dimers, leading to approximate dispersion energies below the variational limit. While in this work we focus on the errors introduced by density fitting rather than those arising from an incomplete RI space, to complement the findings of Ref. 44, we performed  $E_{\text{disp}}^{(20)}$ -F12 and  $E_{\text{exch-disp}}^{(20)}$ -F12 calculations for the same five complexes using another popular CABS choice, the aXZ-OPTRI sets designed specifically to describe the orthogonal complement of the aXZ orbital basis with the same  $X^{83}$ . The resulting DF-SAPT-F12 corrections (with the aYZ-MP2FIT sets,  $Y \geq X$ , used for DF) are presented in Tables SXVI-SXXX in the Supporting Information. Overall, the  $E_{\text{disp}}^{(20)}$ -F12 and  $E_{\text{exch-disp}}^{(20)}$ -F12 accuracy afforded by the two kinds of CABS bases is quite similar: while the  $E_{\text{disp}}^{(20)}$ -F12/aXZ/aXZ-MP2FIT results are more accurate than the  $E_{\text{disp}}^{(20)}$ -F12/aXZ/aXZ-OPTRI ones for the neon dimer, the opposite is true for the methane dimer. Thus, it appears that the smaller aXZ-OPTRI basis sets are also a sensible choice for CABS in SAPT-F12 calculations, however, a more extensive assessment of the performance of both auxiliary bases is required to make

TABLE II. Frozen-core non-DF and DF  $E_{\text{disp}}^{(20)}$ -F12(ODA) and  $E_{\text{exch-disp}}^{(20)}$ -F12(ODA) values (in kcal/mol) for the water dimer for different combinations of the orbital and MP2FIT auxiliary basis. The benchmark values given below the F12 data were computed as conventional DF- $E_{\text{disp}}^{(20)}/\text{DF-}E_{\text{exch-disp}}^{(20)}$  at the (a6Z+(a6Z),a7Z+(a7Z)) extrapolated level.

| Basis                           | CABS Set  | DF- $E_{\text{disp}}^{(20)}$      |              | DF- $E_{\text{disp}}^{(20)}$ -F12      |            |            |
|---------------------------------|-----------|-----------------------------------|--------------|----------------------------------------|------------|------------|
|                                 |           | DF Set:                           |              | aDZ-MP2FIT                             | aTZ-MP2FIT | aQZ-MP2FIT |
| aDZ                             | aDZ-RIFIT | -2.1504                           | -2.5652      | -2.5656                                | -2.5658    | -2.5657    |
| aTZ                             | aTZ-RIFIT | -2.4601                           |              | -2.5924                                | -2.5924    | -2.5924    |
| aQZ                             | aQZ-RIFIT | -2.5484                           |              |                                        | -2.6054    | -2.6054    |
| a5Z                             | a5Z-RIFIT | -2.5776                           |              |                                        |            | -2.6086    |
| <b>CBS(a6Z+(a6Z),a7Z+(a7Z))</b> |           | <b>-2.6126</b>                    | $\pm 0.0069$ |                                        |            |            |
| Basis                           | CABS Set  | DF- $E_{\text{exch-disp}}^{(20)}$ |              | DF- $E_{\text{exch-disp}}^{(20)}$ -F12 |            |            |
|                                 |           | DF Set:                           |              | aDZ-MP2FIT                             | aTZ-MP2FIT | aQZ-MP2FIT |
| aDZ                             | aDZ-RIFIT | 0.3837                            | 0.4530       | 0.4542                                 | 0.4543     | 0.4543     |
| aTZ                             | aTZ-RIFIT | 0.4407                            |              | 0.4737                                 | 0.4740     | 0.4740     |
| aQZ                             | aQZ-RIFIT | 0.4655                            |              |                                        | 0.4830     | 0.4831     |
| a5Z                             | a5Z-RIFIT | 0.4753                            |              |                                        |            | 0.4863     |
| <b>CBS(a6Z+(a6Z),a7Z+(a7Z))</b> |           | <b>0.4885</b>                     | $\pm 0.0024$ |                                        |            |            |

TABLE III. Frozen-core non-DF and DF  $E_{\text{disp}}^{(20)}$ -F12(ODA) and  $E_{\text{exch-disp}}^{(20)}$ -F12(ODA) values (in kcal/mol) for the methane dimer for different combinations of the orbital and MP2FIT auxiliary basis. The benchmark values given below the F12 data were computed as conventional DF- $E_{\text{disp}}^{(20)}/\text{DF-}E_{\text{exch-disp}}^{(20)}$  at the (a5Z+(a5Z),a6Z+(a6Z)) extrapolated level.

| Basis                           | CABS Set  | DF- $E_{\text{disp}}^{(20)}$      |              | DF- $E_{\text{disp}}^{(20)}$ -F12      |            |            |
|---------------------------------|-----------|-----------------------------------|--------------|----------------------------------------|------------|------------|
|                                 |           | DF Set:                           |              | aDZ-MP2FIT                             | aTZ-MP2FIT | aQZ-MP2FIT |
| aDZ                             | aDZ-RIFIT | -1.0086                           | -1.1367      | -1.1379                                | -1.1380    | -1.1380    |
| aTZ                             | aTZ-RIFIT | -1.1084                           |              | -1.1471                                | -1.1472    | -1.1471    |
| aQZ                             | aQZ-RIFIT | -1.1331                           |              |                                        | -1.1498    | -1.1498    |
| a5Z                             | a5Z-RIFIT | -1.1420                           |              |                                        |            | -1.1506    |
| <b>CBS(a5Z+(a5Z),a6Z+(a6Z))</b> |           | <b>-1.1513</b>                    | $\pm 0.0024$ |                                        |            |            |
| Basis                           | CABS Set  | DF- $E_{\text{exch-disp}}^{(20)}$ |              | DF- $E_{\text{exch-disp}}^{(20)}$ -F12 |            |            |
|                                 |           | DF Set:                           |              | aDZ-MP2FIT                             | aTZ-MP2FIT | aQZ-MP2FIT |
| aDZ                             | aDZ-RIFIT | 0.0690                            | 0.0822       | 0.0829                                 | 0.0830     | 0.0830     |
| aTZ                             | aTZ-RIFIT | 0.0787                            |              | 0.0851                                 | 0.0852     | 0.0852     |
| aQZ                             | aQZ-RIFIT | 0.0831                            |              |                                        | 0.0868     | 0.0868     |
| a5Z                             | a5Z-RIFIT | 0.0851                            |              |                                        |            | 0.0876     |
| <b>CBS(a5Z+(a5Z),a6Z+(a6Z))</b> |           | <b>0.0882</b>                     | $\pm 0.0008$ |                                        |            |            |

definite recommendations.

Out of the multitude of aXZ/aYZ-RIFIT orbital and CABS set combinations investigated in Ref. 44, we have selected only the “diagonal” aXZ/aXZ-RIFIT ones in Tables II–III, focusing on the convergence of results with the DF basis. For all considered dimers, the DF- $E_{\text{disp}}^{(20)}$ -F12 calculations in the aDZ/aDZ-RIFIT/aDZ-MP2FIT combination of orbital, CABS, and DF bases lead to errors in the range of 0.0005-0.0013 kcal/mol compared to the non-DF value, with an exception of He–He, for which the error amounts to 0.03 kcal/mol. This error is consistently reduced down to 0.0001 kcal/mol (0.0012 kcal/mol for the helium dimer) when one utilizes the aDZ/aDZ-RIFIT sets combined with the aTZ-MP2FIT set for DF. This observation confirms that the aDZ/aDZ-RIFIT/aDZ-MP2FIT combination is somewhat inaccurate for DF-SAPT-F12 calculations, which is in agreement with our earlier study<sup>44</sup> (albeit sometimes, an error cancellation may occur). For the triple-zeta and larger bases, the default aXZ-MP2FIT choice for the DF basis, with  $X$  the same as for the orbital set, is always adequate, with the DF- $E_{\text{disp}}^{(20)}$ -F12 errors not exceeding 0.0002 kcal/mol (0.0008 kcal/mol for the helium dimer).

Based on the DF- $E_{\text{exch-disp}}^{(20)}$ -F12 energies presented in Tables II–III and in the Supporting Information, we again observe that the aDZ/aDZ-RIFIT/aDZ-MP2FIT combination is the source of the largest DF error (0.0003–0.0036 kcal/mol). The error is reduced to just 0.0001–0.0002 kcal/mol (0.0015 kcal/mol for Ar–Ar) when the aDZ/aDZ-RIFIT/aTZ-MP2FIT basis sets are employed. Analyzing the bases with higher cardinal numbers, the DF- $E_{\text{exch-disp}}^{(20)}$ -F12 energies are virtually converged to their non-DF counterparts when utilizing the default aXZ-MP2FIT DF sets,

with  $X$  the same as for the orbital basis, except for the argon dimer (where the aTZ/aTZ-RIFIT/aTZ-MP2FIT combination produces the largest error of 0.0011 kcal/mol).

The results in Tables II–III demonstrate that the DF error is negligible and well-controlled in the DF- $E_{\text{disp}}^{(20)}$ -F12 and DF- $E_{\text{exch-disp}}^{(20)}$ -F12 calculations. Moreover, the choice of the auxiliary basis set family does not affect the accuracy of energies, and both MP2FIT and JKFIT sets (the results for the latter are presented in the Supporting Information) are adequate for fitting the Fock and exchange matrices in the DF-SAPT-F12 context. Since the MP2FIT basis set is usually smaller than JKFIT, and at the same time does not compromise the quality of results, it was selected for fitting all four-index quantities in all remaining calculations.

### B. Performance of DF- $E_{\text{disp}}^{(20)}$ -F12 and DF- $E_{\text{exch-disp}}^{(20)}$ -F12 on the A24 database

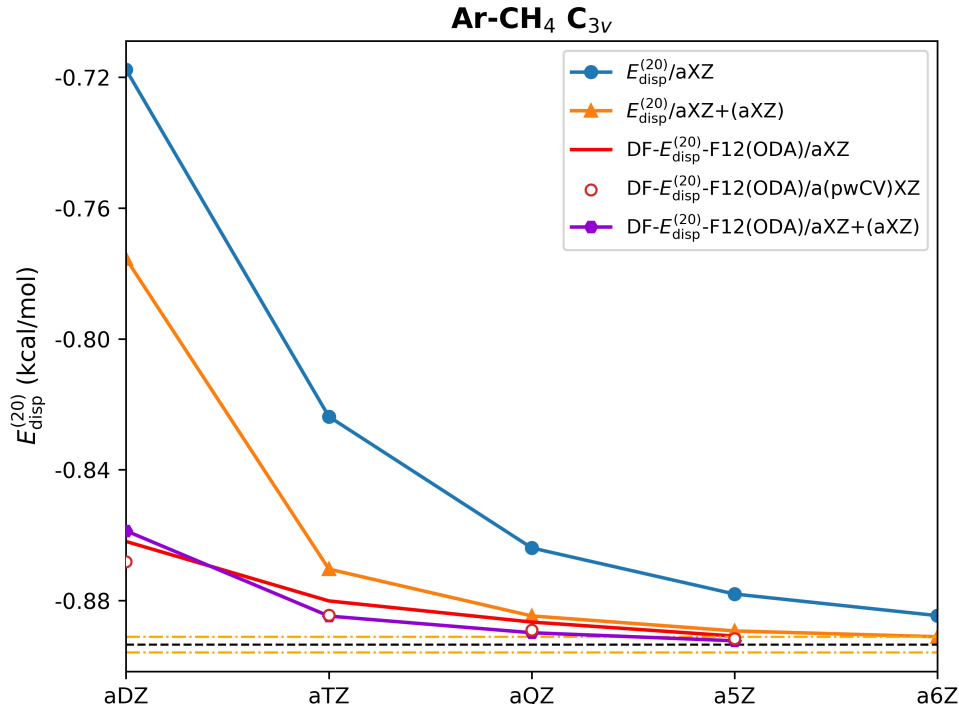
In order to investigate the performance of DF-SAPT-F12 on a broader sample of complexes, further tests were carried out on the A24 database<sup>62</sup>. In contrast to our previous work<sup>44</sup> where we were only able to run A24 calculations in basis sets up to aTZ, the current DF implementation allows us to obtain all energies in bases up to a5Z with and without midbond functions. The basis set convergence of dispersion and exchange dispersion corrections, and of the corresponding sum  $E_{\text{disp}}^{(20)} + E_{\text{exch-disp}}^{(20)}$ , for individual A24 systems is presented in Figs. 1–4 and S1–S6 (Supporting Information). Analyzing Figs. S1 and S2, it is clear that our newly proposed DF-SAPT-F12 method shows the same fast convergence rate as its non-DF counterpart: in fact, the non-DF and DF results are indistinguishable on the scale of these figures. We further see that the DF- $E_{\text{disp}}^{(20)}$ -F12/aTZ and DF- $E_{\text{exch-disp}}^{(20)}$ -F12/aQZ energies are either converged or almost converged to the reference CBS value. It turns out that the most challenging systems observed in our previous study<sup>44</sup>, namely Ar-CH<sub>4</sub> (Figure 3) and Ar-C<sub>2</sub>H<sub>4</sub> (Figure S2), require a basis set with one cardinal number larger (aQZ) to reach the benchmark level for the DF- $E_{\text{disp}}^{(20)}$ -F12 calculations. Although the DF- $E_{\text{exch-disp}}^{(20)}$ -F12/a5Z energies are still not fully converged to the reference value for these two systems, we definitely observe a substantial improvement. Another way to improve the DF- $E_{\text{disp}}^{(20)}$ -F12 and DF- $E_{\text{exch-disp}}^{(20)}$ -F12 performance for these challenging systems is to include additional compact basis functions centered on the argon atom, such as the functions from a polarized core and valence basis sets, aug-cc-pwCVXZ<sup>84</sup>. Such sets are normally used for calculations with all electrons correlated, but can provide additional flexibility to our frozen-core calculations as well. As illustrated in Figs. S7–S10 in the Supporting Information, enlarging the argon basis set from aXZ to aug-cc-pwCVXZ (and using aug-cc-pwCVXZ-MP2FIT as the CABS and DF set for argon), while keeping the aXZ basis set on all other atoms, leads to a noticeable improvement in the accuracy of DF- $E_{\text{disp}}^{(20)}$ -F12 and a small improvement in the accuracy of DF- $E_{\text{exch-disp}}^{(20)}$ -F12. Thus, it appears that the standard aXZ sets for argon might not contain sufficiently tight functions to maximize the benefits of the F12 approach.

It is worth emphasizing that the DF- $E_{\text{disp}}^{(20)}$ -F12 results are consistently converged to the CBS limit in the aQZ basis, whereas that precision level is still not reproduced by the non-F12 approach even in a very large orbital basis such as a6Z+(a6Z). To further illustrate the superior convergence of the DF- $E_{\text{disp}}^{(20)}$ -F12 and DF- $E_{\text{exch-disp}}^{(20)}$ -F12 values over their non-F12 counterparts (even those computed with midbonds), we selected the water dimer and compared both approaches against benchmark data of increased quality. For this purpose, we performed standard SAPT calculations (augmented with hydrogenic midbond functions) utilizing the aug-cc-pV7Z and aug-mcc-pV7Z bases for the oxygen and hydrogen atoms, respectively, and extrapolated the results from the a6Z+(a6Z) and a7Z+(a7Z) data. Figure 1 displays the convergence of DF- $E_{\text{disp}}^{(20)}$ -F12 H<sub>2</sub>O–H<sub>2</sub>O values to this new improved CBS reference, and it demonstrates that the DF- $E_{\text{disp}}^{(20)}$ -F12 numbers are very well converged to the extra-precise CBS value. We therefore conclude that DF- $E_{\text{disp}}^{(20)}$ -F12 converges to the actual CBS limit, which is still not reproduced to that precision by a very large orbital basis such as a7Z+(a7Z). The corresponding behavior of the exchange-dispersion energies relative to the H<sub>2</sub>O–H<sub>2</sub>O benchmark value of increased precision is illustrated in Figure 2, once again confirming superior convergence of the F12 data.

Another way of alleviating the slow basis set convergence of dispersion energy is the utilization of midbond functions. It has been demonstrated that this technique in conjunction with explicitly correlated methods works very well for supermolecular CCSD(T)-F12 interaction energies<sup>18</sup>. Thus, we decided to investigate the influence of midbond functions on the SAPT and SAPT-F12 convergence using the A24 dataset as an example. It is worth emphasizing that the midbond and F12 approaches to dispersion are complementary, not competitive.

The combination of DF- $E_{\text{disp}}^{(20)}$ -F12 with the +(aXZ) midbond set leads to a moderate improvement of the accuracy in the aDZ and aTZ bases. On the other hand, the DF- $E_{\text{disp}}^{(20)}$ -F12 values obtained with the +(3322) midbond set show

FIG. 3. Convergence of the frozen-core non-DF and DF  $E_{\text{disp}}^{(20)}$ -F12 results as a function of basis set for the Ar-CH<sub>4</sub> complex from the A24 database. The hydrogenic functions from the same aXZ orbital basis set or the constant (3s3p2d2f) set of functions are chosen for midbond functions. The reference value marked by a black dashed line was obtained at the  $E_{\text{disp}}^{(20)}/(\text{a5Z}+(\text{a5Z}), \text{a6Z}+(\text{a6Z}))$  extrapolated level. The yellow dashed lines indicate the benchmark uncertainty.



results converged to the a6Z+(a6Z) level, or even to the reference value, already in the aDZ basis with an exception of Ar-CH<sub>4</sub> and Ar-C<sub>2</sub>H<sub>4</sub>. While the superior performance of the aDZ+(3322) results might be to some extent fortuitous, it is encouraging that the accuracy is consistent, similar to the consistently high accuracy (in particular, on the same A24 database) provided by the properly selected supermolecular CCSD(T)-F12/aDZ+(3322) variants<sup>18</sup>.

The effect of hydrogenic midbond functions is more pronounced for the DF- $E_{\text{exch-disp}}^{(20)}$ -F12 correction. In Figs. 2, 4, S3, and S4 we observe a significantly improved recovery of the CBS values in aDZ+(aDZ), aTZ+(aTZ), and aQZ+(aQZ) relative to the corresponding midbondless bases. Overall, the  $E_{\text{exch-disp}}^{(20)}$ -F12/aTZ+(aTZ) energies are converged as well, or better, than standard  $E_{\text{exch-disp}}^{(20)}$  in the a5Z+(a5Z) basis. Considering the aQZ+(aQZ) basis set, the F12 energies are consistently converged above the conventional a6Z+(a6Z) level (or to that level in the case of Ar-CH<sub>4</sub> and Ar-C<sub>2</sub>H<sub>4</sub>, Figure 4 and S4) and some further enhancement is attained in the a5Z+(a5Z) basis.

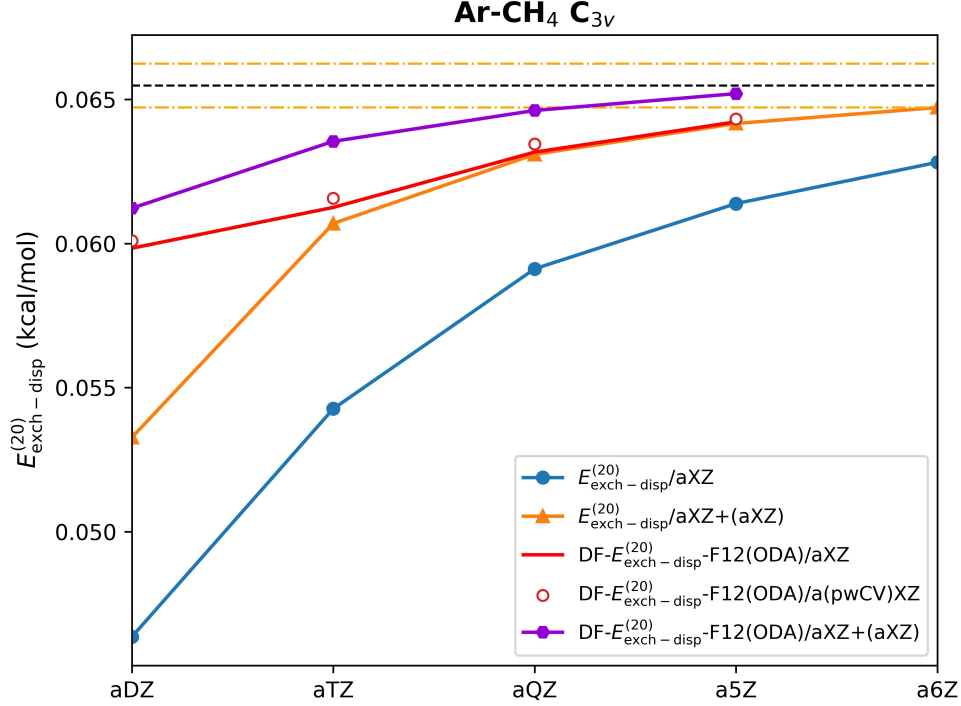
When evaluating DF- $E_{\text{exch-disp}}^{(20)}$ -F12 with the constant +(3322) set of midbond functions, we observe superior convergence at the aDZ and aTZ levels compared to DF- $E_{\text{exch-disp}}^{(20)}$ -F12 with and without the hydrogenic midbond. In a few cases (e.g., for the NH<sub>3</sub>-C<sub>2</sub>H<sub>4</sub> complex), the addition of constant midbonds provides results nearly converged to the reference value already in the aDZ+(3322) basis. The  $E_{\text{exch-disp}}^{(20)}$ -F12/aTZ+(3322) results attain the conventional  $E_{\text{exch-disp}}^{(20)}/\text{a6Z}+(\text{a6Z})$ -level accuracy across nearly the entire A24 database. The challenging systems containing the argon atom require basis sets with one cardinal number higher to reproduce the same precision.

The relative errors with respect to the benchmark for the DF- $E_{\text{disp}}^{(20)}$ -F12 and DF- $E_{\text{exch-disp}}^{(20)}$ -F12 results in a range of basis sets with and without midbond functions are illustrated in Figures 5–12. The corresponding relative errors for the sum  $E_{\text{disp}}^{(20)} + E_{\text{exch-disp}}^{(20)}$  are presented in the Supporting Information (Figs. S11–S14). One clearly sees that the DF- $E_{\text{disp}}^{(20)}$ -F12 and DF- $E_{\text{exch-disp}}^{(20)}$ -F12 energies computed in aDZ and aTZ show the same performance as their non-DF counterparts. The only systems which exhibit a visible, though still tiny, DF error for  $E_{\text{exch-disp}}^{(20)}$ -F12/aDZ are Ar-CH<sub>4</sub> and Ar-C<sub>2</sub>H<sub>4</sub>, indicating that the DF basis is not entirely converged in these cases.

Thanks to the DF algorithm, we can now investigate how well the explicitly correlated dispersion and exchange-



FIG. 4. Convergence of the frozen-core non-DF and DF  $E_{\text{exch-disp}}^{(20)}$ -F12 results as a function of basis set for the Ar-CH<sub>4</sub> complex from the A24 database. The hydrogenic functions from the same aXZ orbital basis set or the constant (3s3p2d2f) set of functions are chosen for midbond functions. The reference value marked by a black dashed line was obtained at the  $E_{\text{exch-disp}}^{(20)}/(\text{a5Z}+(\text{a5Z}), \text{a6Z}+(\text{a6Z}))$  extrapolated level. The yellow dashed lines indicate the benchmark uncertainty.

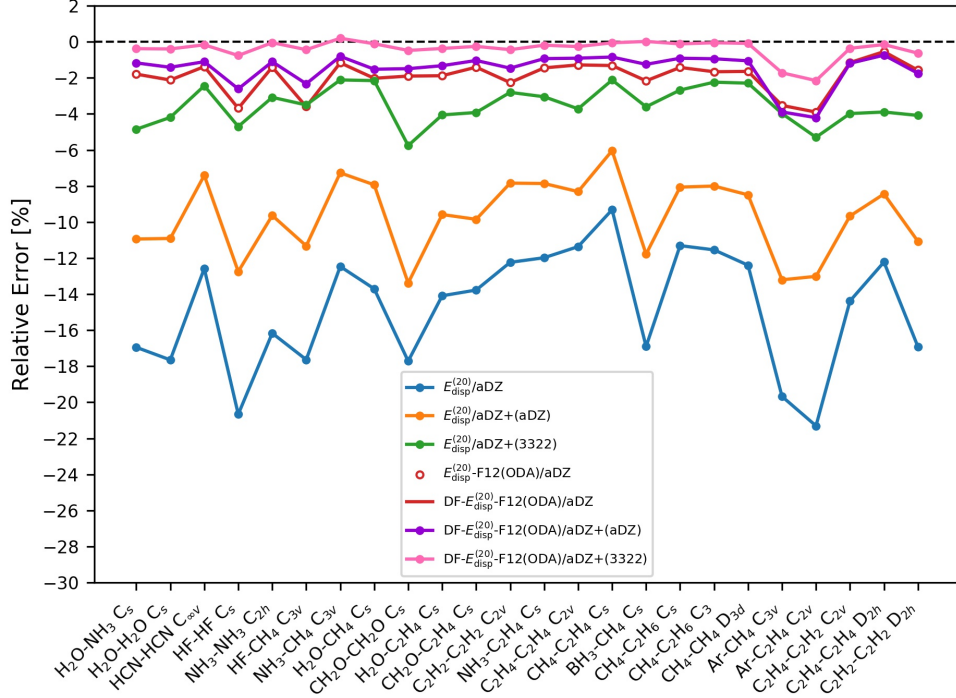


dispersion corrections work in the aQZ and a5Z basis sets (our non-DF implementation<sup>44</sup> could not handle some of the A24 systems in such large bases). Generally, the DF- $E_{\text{disp}}^{(20)}$ -F12/aQZ results closely reproduce the benchmark level, leading to the mean absolute percent error (MA%E) value of 0.23%. This error is reduced down to 0.04% in the a5Z basis. It needs to be stressed that the systems containing the argon atom are the source of largest errors.

Pursuing the same analysis for DF- $E_{\text{exch-disp}}^{(20)}$ -F12/aQZ (Fig. 11), we observe the enhanced performance compared to the aTZ level, leading to the MA%E value of 1.33%. A further improvement is obtained in the a5Z basis set (See Fig. 12), showing a small MA%E value of 0.25%. Examining the sum of DF- $E_{\text{disp}}^{(20)}$ -F12 and DF- $E_{\text{exch-disp}}^{(20)}$ -F12 at the aQZ and a5Z level (Figs. S13 and S14, respectively, in the Supporting Information), we notice that the energies phenomenally reproduce the CBS limit, leading to MA%E values of 0.08% and 0.02%, respectively.

For the last test on the A24 database, we looked at the accuracy of the explicitly correlated calculations augmented with the variable (hydrogenic) and constant (3s3p2d2f) sets of midbond functions. Figures 5 and 9 indicate that the improvement brought about by the +(3322) midbond set for the DF-SAPT-F12/aDZ calculations is nothing short of impressive. This level of theory faithfully reproduces the benchmark data with MA%E of 0.41%, 1.7%, and 0.27% for DF- $E_{\text{disp}}^{(20)}$ -F12, DF- $E_{\text{exch-disp}}^{(20)}$ -F12, and DF- $E_{\text{disp}}^{(20)}$ -F12 + DF- $E_{\text{exch-disp}}^{(20)}$ -F12, respectively. The hydrogenic midbond functions bring minimal (but still noticeable) accuracy gain for aDZ, leading to MA%E values of 1.5%, 4.3%, and 1.1% for DF- $E_{\text{disp}}^{(20)}$ -F12, DF- $E_{\text{exch-disp}}^{(20)}$ -F12, and DF- $E_{\text{disp}}^{(20)}$ -F12 + DF- $E_{\text{exch-disp}}^{(20)}$ -F12, respectively. These values should be contrasted with the respective MA%E for the midbondless aDZ data, amounting to 1.9%, 6.3%, and 1.3% for these three quantities. Moving on to the aTZ basis, an additional enhancement is observed for the explicitly correlated dispersion and exchange-dispersion corrections, as well as for their sum, when either midbond type is applied. The DF- $E_{\text{exch-disp}}^{(20)}$ -F12/aTZ+(3322) approach exhibits superior performance (a MA%E of 1.1% was found for this level) with respect to DF- $E_{\text{exch-disp}}^{(20)}$ -F12/aTZ+(aTZ) (a MA%E of 1.5%) and plain DF- $E_{\text{exch-disp}}^{(20)}$ -F12/aTZ (a MA%E of 3.0%). The DF- $E_{\text{disp}}^{(20)}$ -F12/aTZ variant performs better with the +(3322) midbond than with the hydrogenic ones, providing respective MA%E values of 0.18% and 0.30%. This behavior is very much expected: since the (3s3p2d2f) set is larger than the hydrogenic set for the aDZ and aTZ orbital basis sets, it results in more accurate values.

FIG. 5. Relative errors on the A24 database for frozen-core  $E_{\text{disp}}^{(20)}$ -F12 and  $\text{DF-}E_{\text{disp}}^{(20)}$ -F12 computed with the aDZ orbital basis set, the aTZ-RIFIT CABS set, and the aTZ-MP2FIT DF set. The notation +(aXZ) and +(3322) signifies that the hydrogenic aXZ and the constant (3s3p2d2f) set of midbond functions has been added to the aXZ atom-centered basis, respectively. The reference values were obtained at the  $E_{\text{disp}}^{(20)}/(\text{a5Z}+(\text{a5Z}), \text{a6Z}+(\text{a6Z}))$  extrapolated level.

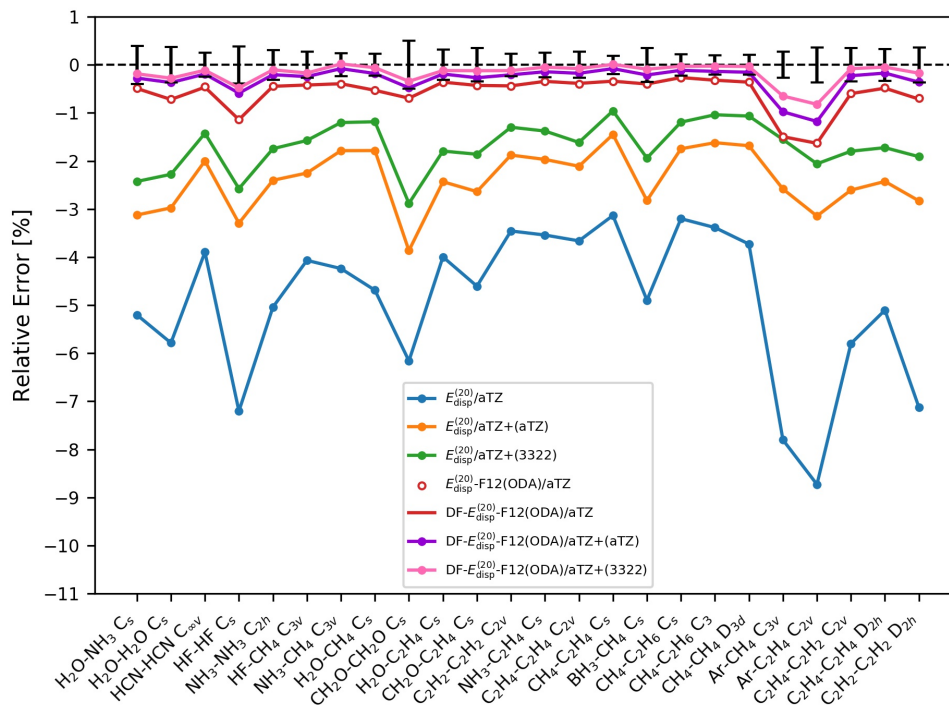


The inclusion of midbond functions still brings improvement over the midbondless approach for the  $\text{DF-}E_{\text{disp}}^{(20)}$ -F12 values in aQZ and a5Z, even though the energies are already well converged to the CBS limit. The improvement is even more pronounced for the exchange dispersion energy. For the A24 systems other than Ar-CH<sub>4</sub> and Ar-C<sub>2</sub>H<sub>4</sub>, converging the results to within the (very tight) uncertainty of the benchmark requires going up to either aQZ or aTZ+(midbond) for  $\text{DF-}E_{\text{disp}}^{(20)}$ -F12 and either a5Z or aQZ+(midbond) for  $\text{DF-}E_{\text{exch-disp}}^{(20)}$ -F12. For the two complexes containing argon, one needs to go all the way to a5Z+(midbond) to converge either correction to within the benchmark uncertainty. At the aQZ level, the +(aQZ) and +(3322) midbonds perform nearly identically, which is expected given the similar composition of the two sets of bond functions. When one goes up to a5Z, the +(a5Z) midbond set becomes larger than the +(3322) one, leading to a slightly larger improvement in accuracy. Last but not least, Figs. 8 and 12 show that the deviations between the (a5Z+(a5Z), a6Z+(a6Z)) extrapolated benchmarks and the largest-basis F12 results are significantly smaller than the benchmark uncertainties. This level of consistency is really gratifying, as it validates both the  $X^{-3}$  extrapolation and the F12 data, and even indicates that the error estimates for the former approach are quite conservative.

### C. Total SAPT-F12 interaction energies

One of the significant advantages of energy decomposition methods such as SAPT is that various perturbation corrections can be studied separately in different basis sets. This allows us to reach a CBS limit of a particular correction independently from others. Specifically, the small-basis  $E_{\text{disp}}^{(20)}$  and  $E_{\text{exch-disp}}^{(20)}$  energies can be substituted by their F12 counterparts in all wave function-based SAPT levels defined in Ref. 54, giving rise to new SAPT variants that will be denoted below as SAPT0-F12, SAPT2-F12, SAPT2+-F12, SAPT2+(3)-F12, and SAPT2+3-F12. Such a procedure is valid (no double counting occurs) and can be considered as a “focal point” approach similar to MP2/CBS+ $\delta\text{CCSD(T)}$ , where the lower-level estimate, converged to CBS (or at least much closer to CBS thanks to the F12 approach), is improved by a higher-level estimate in a moderate basis. To illustrate the effect of this F12 modification on the accuracy of total SAPT interaction energies, Figure 13 displays the mean absolute errors (MAE)

FIG. 6. Relative errors on the A24 database for frozen-core  $E_{\text{disp}}^{(20)}$ -F12 and DF- $E_{\text{disp}}^{(20)}$ -F12 computed with the aTZ orbital basis set, the aTZ-RIFIT CABS set, and the aTZ-MP2FIT DF set. The notation +(aXZ) and +(3322) signifies that the hydrogenic aXZ and the constant (3s3p2d2f) set of midbond functions has been added to the aXZ atom-centered basis, respectively. The reference values were obtained at the  $E_{\text{disp}}^{(20)}/(\text{a5Z}+(\text{a5Z}), \text{a6Z}+(\text{a6Z}))$  extrapolated level. The error bars indicate the benchmark uncertainty.



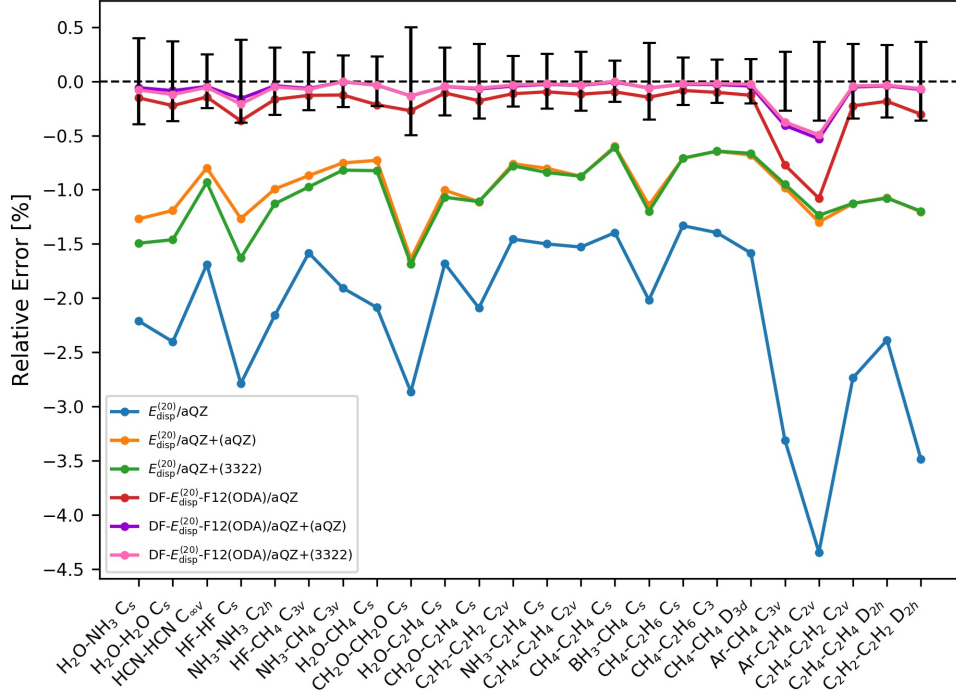
of interaction energy, relative to the CCSD(T)/CBS benchmarks, averaged over the S22<sup>56,78</sup>, HBC6<sup>79</sup>, NBC10<sup>80</sup>, and HSG<sup>81</sup> databases for standard and explicitly correlated SAPT methods, with and without the CCD dispersion and the “ $\delta$ MP2” correction. All MAE values reported below refer to this database combination and are fully compatible with the errors obtained and displayed, on the very same dataset, in Ref. 54; see Sec. III for some technical details (truncation of the HBC6 and NBC10 databases, scaling of the  $E_{\text{exch-disp}}^{(20)}$  and  $E_{\text{exch-disp}}^{(20)}$ -F12 corrections) that are essential for maintaining this compatibility. The MAE data tend to emphasize the performance on the database(s) with the largest magnitude of interaction energies (in this case, the HBC6 dataset of doubly hydrogen bonded complexes); to provide a more balanced view, Fig. 14 complements Fig. 13 by showing the respective mean absolute percent errors (MA%E). The corresponding statistical errors (MAE and MA%E) for the four individual databases are presented in Figures S15–S22 in the Supporting Information.

At first, one could think that the F12 dispersion aims to improve the SAPT0 interaction energies. However, it is just the opposite: it makes them less accurate, leading to the MAE values of 2.46 kcal/mol and 2.58 kcal/mol in aDZ and aTZ, respectively, as compared to 1.74 and 2.34 kcal/mol for conventional SAPT0. This behavior was foreseeable since the  $E_{\text{disp}}^{(20)}$  correction tends to overestimate the dispersion binding and it becomes even more negative when approaching the CBS limit. It has been demonstrated that the success of SAPT0 relies on error cancellation between the overestimation of dispersion by its leading  $E_{\text{disp}}^{(20)}$  term and the underestimation of dispersion by a small “calendar” basis set, such as jun-cc-pVDZ<sup>54</sup>. Therefore, the simplest SAPT0 approach is not recommended to be combined with the F12 dispersion, and it is excluded from further studies.

The SAPT2 variant, which extends SAPT0 by including intramolecular electron correlation up to second order for electrostatic, exchange, and induction interactions, was the subject of the first tests with the F12 treatment. The poor performance of SAPT2-F12 can be assigned to the lack of intramolecular correlation in the dispersion corrections. At this level, the F12 treatment has a negative effect on the accuracy of SAPT2, increasing the average errors in aDZ and aTZ by 0.02 kcal/mol and 0.28 kcal/mol, respectively.

When the F12 dispersion and exchange dispersion is added to SAPT2+/aDZ, the first SAPT level which includes intramolecular electron correlation for dispersion up to second order ( $E_{\text{disp}}^{(21)}$  and  $E_{\text{disp}}^{(22)}$ ), the errors increase quite

FIG. 7. Relative errors on the A24 database for frozen-core  $E_{\text{disp}}^{(20)}$ -F12 and DF- $E_{\text{disp}}^{(20)}$ -F12 computed with the aQZ orbital basis set, the aQZ-RIFIT CABS set, and the aQZ-MP2FIT DF set. The notation +(aXZ) and +(3322) signifies that the hydrogenic aXZ and the constant (3s3p2d2f) set of midbond functions has been added to the aXZ atom-centered basis, respectively. The reference values were obtained at the  $E_{\text{disp}}^{(20)}/(\text{a5Z}+(\text{a5Z}), \text{a6Z}+(\text{a6Z}))$  extrapolated level. The error bars indicate the benchmark uncertainty.



dramatically. It is worth noting that SAPT2+/aDZ was established in Ref. 54 as the silver standard of SAPT with a MAE of 0.30 kcal/mol. Such an attractive accuracy is linked to a fortuitous error cancellation between the method error and basis set incompleteness, which no longer happens for SAPT2+-F12. Interestingly, the SAPT2+-F12/aTZ approach produces the MAE value of 1.27 kcal/mol, which is the largest error across all considered SAPT-F12 options.

The SAPT2+(3)-F12 method exhibits better performance than SAPT2+-F12 in both aDZ and aTZ, affording errors of 0.62 kcal/mol and 0.88 kcal/mol, respectively. However, the F12 treatment still worsens the accuracy of standard SAPT2+(3) with MAEs of 0.39 kcal/mol and 0.54 kcal/mol in the same bases. The approach with the complete third-order correction, SAPT2+3-F12, does not outperform SAPT2+(3)-F12 either in aDZ or aTZ, leading to the MAE values of 0.86 kcal/mol and 1.22 kcal/mol, respectively. It follows that the higher accuracy of SAPT2+(3)-F12, just like for the non-F12 approach, is based on an error cancellation that occurs when the third-order mixed induction-dispersion effects are neglected.

It has been shown that the CCD-based construction of dispersion amplitudes in SAPT, even though not as remarkably accurate as the full CCSD treatment of dispersion<sup>27,85</sup>, usually provides an improved description of noncovalent interactions<sup>26,54,77</sup>. In this spirit, the amplitudes in the second-order dispersion energy terms ( $E_{\text{disp}}^{(20)} + E_{\text{disp}}^{(21)} + E_{\text{disp}}^{(22)}$ ) are replaced by the converged CCD amplitudes<sup>26</sup>, giving rise to SAPT(CCD) (this dispersion algorithm was originally denoted, more precisely, as CCD+ST(CCD) as it also includes a perturbative estimate of the singles and triples contributions to  $E_{\text{disp}}^{(22)}$  using converged doubles amplitudes). Another route to enhance the accuracy of the SAPT calculations, especially for hydrogen-bonded systems, is to include the “ $\delta$ MP2” correction. This term is computed as a difference between the counterpoise (CP) corrected MP2 interaction energy and the SAPT2 interaction energy:

$$\delta\text{MP2} = E_{\text{int}}^{\text{MP2}} - E_{\text{int}}^{\text{SAPT2}} \quad (42)$$

and it accounts for third- as well as higher-order coupling between induction and dispersion<sup>21,54</sup>. Both improvements, (CCD) and “ $\delta$ MP2”, can be employed within SAPT2+, SAPT2+(3), and SAPT2+3 together or separately. For all these SAPT variants, the effect of replacing standard  $E_{\text{disp}}^{(20)}$  and  $E_{\text{exch-disp}}^{(20)}$  by our nearly converged  $E_{\text{disp}}^{(20)}$ -F12 and  $E_{\text{exch-disp}}^{(20)}$ -F12 values was also tested.

FIG. 8. Relative errors on the A24 database for frozen-core  $E_{\text{disp}}^{(20)}$ -F12 and DF- $E_{\text{disp}}^{(20)}$ -F12 computed with the a5Z orbital basis set, the a5Z-RIFIT CABS set, and the a5Z-MP2FIT DF set. The notation +(aXZ) and +(3322) signifies that the hydrogenic aXZ and the constant (3s3p2d2f) set of midbond functions has been added to the aXZ atom-centered basis, respectively. The reference values were obtained at the  $E_{\text{disp}}^{(20)}/(\text{a5Z}+(\text{a5Z}), \text{a6Z}+(\text{a6Z}))$  extrapolated level. The error bars indicate the benchmark uncertainty.

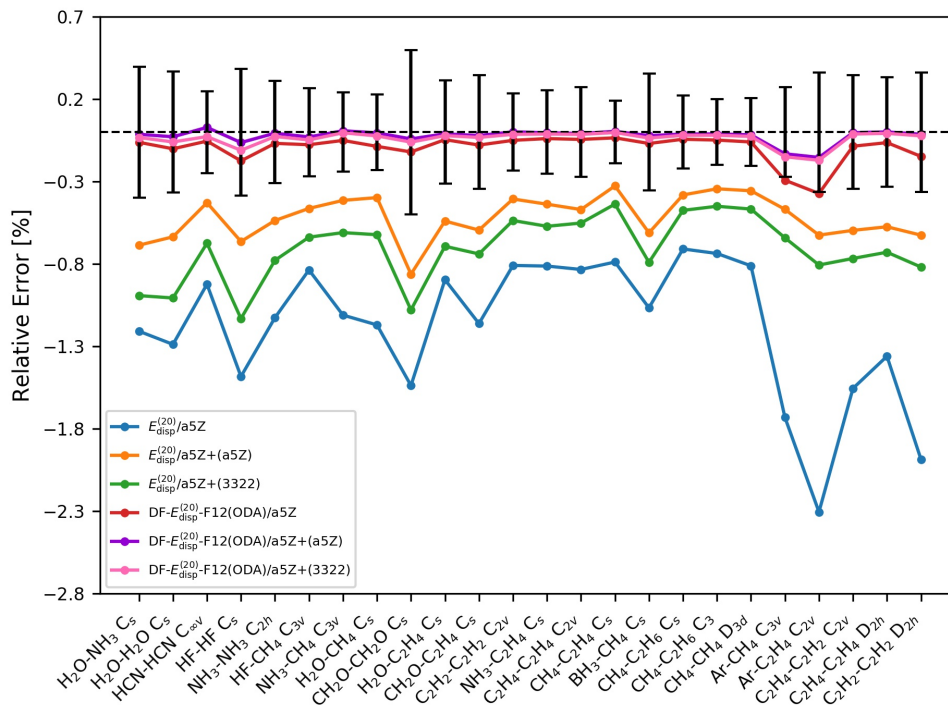


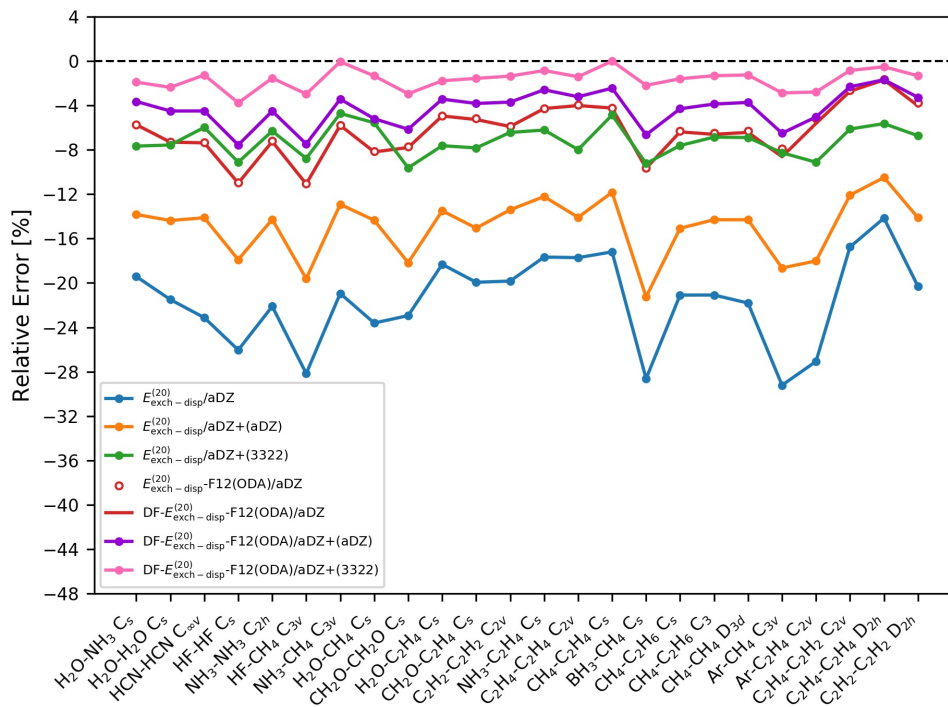
Figure 13 reveals that the more robust treatment of dispersion achieved by the CCD amplitudes reduces the overall errors at all high levels of SAPT. We again observe that SAPT2+(CCD)-F12 and SAPT2+3(CCD)-F12 perform nearly equivalent in both basis sets, and worsen results relative to SAPT2+(3)(CCD)-F12. The latter flavor leads to an enhancement in aDZ with respect to the standard counterpart, showing a MAE value of 0.45 kcal/mol, albeit the performance is deteriorated in aTZ, producing an average error of 0.69 kcal/mol.

Once the  $\delta\text{MP2}$  term is added to the SAPT calculations, we observe that SAPT2+-F12 $\delta\text{MP2}$  still slightly underperforms SAPT2+ $\delta\text{MP2}$  in aDZ, giving rise to an error of 0.54 kcal/mol, while SAPT2+(3)-F12 $\delta\text{MP2}$ /aDZ and SAPT2+3-F12 $\delta\text{MP2}$ /aDZ substantially improve the accuracy of their non-F12 counterparts, yielding the MAE values of 0.41 kcal/mol and 0.43 kcal/mol, respectively. These values should be compared with the MAE of 0.48 kcal/mol, 0.71 kcal/mol, and 0.64 kcal/mol obtained for SAPT2+ $\delta\text{MP2}$ , SAPT2+(3) $\delta\text{MP2}$ , and SAPT2+3 $\delta\text{MP2}$ , respectively. The inclusion of the F12 terms spoils the excellent accuracy of SAPT2+ $\delta\text{MP2}$ , SAPT2+(3) $\delta\text{MP2}$ , and SAPT2+3 $\delta\text{MP2}$  in aTZ, increasing the errors by 0.36 kcal/mol, 0.24 kcal/mol, and 0.33 kcal/mol, respectively. The F12 treatment combined with both CCD dispersion and the  $\delta\text{MP2}$  correction has a highly positive influence on the accuracy in the aDZ basis, reducing the non-F12 errors from 0.66 to 0.38 kcal/mol for SAPT2+(CCD)-F12 $\delta\text{MP2}$ , from 0.97 to 0.40 kcal/mol for SAPT2+(3)(CCD)-F12 $\delta\text{MP2}$ , and from 0.89 to 0.36 kcal/mol for SAPT2+3(CCD)-F12 $\delta\text{MP2}$ . Once the basis set size is enlarged, we observe that SAPT2+(CCD)-F12 $\delta\text{MP2}$ /aTZ is not nearly as accurate as SAPT2+(CCD) $\delta\text{MP2}$ /aTZ (the error increases from 0.12 kcal/mol to 0.47 kcal/mol). At the highest levels of theory, the F12 treatment still increases the MAE, but only by a small amount: from 0.36 kcal/mol to 0.38 kcal/mol for SAPT2+(3)(CCD) $\delta\text{MP2}$ /aTZ and from 0.24 kcal/mol to 0.34 kcal/mol for SAPT2+3(CCD) $\delta\text{MP2}$ /aTZ.

A closer examination of the SAPT and SAPT-F12 errors on the individual databases (presented in the Supporting Information) reveals that the overall MAE trends follow closely the behavior observed for the dataset exhibiting the strongest binding, that is, HBC6. This happens to be the subset of the overall database for which including the F12 correction in a high-level SAPT treatment such as SAPT2+3(CCD) $\delta\text{MP2}$  performs particularly poorly (note that, at this level of SAPT, the F12 inclusion is uniformly beneficial in both aDZ and aTZ for the S22, NBC10, and HSG databases). To alleviate the dominance of the HBC6 set over the overall error statistics, Fig. 14 presents the relative errors (MA%E values) computed on the entire dataset. For this statistical metric, F12 still does not provide



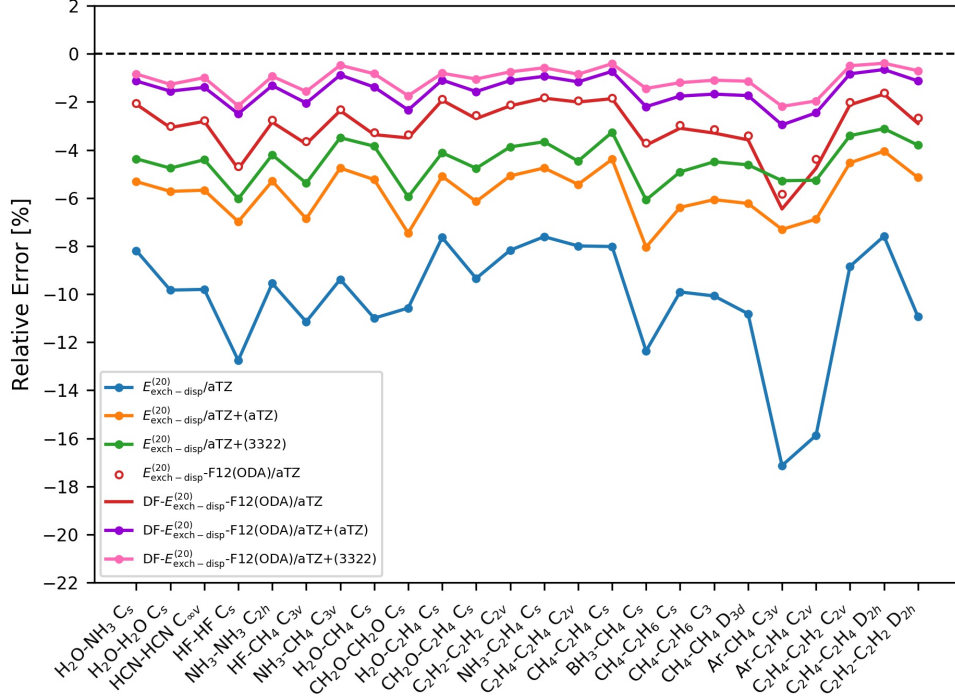
FIG. 9. Relative errors on the A24 database for frozen-core  $E_{\text{exch-disp}}^{(20)}$ -F12 and  $\text{DF-}E_{\text{exch-disp}}^{(20)}$ -F12 computed with the aDZ orbital basis set, the aTZ-RIFIT CABS set, and the aTZ-MP2FIT DF set. The notation +(aXZ) and +(3322) signifies that the hydrogenic aXZ and the constant (3s3p2d2f) set of midbond functions has been added to the aXZ atom-centered basis, respectively. The reference values were obtained at the  $E_{\text{exch-disp}}^{(20)}/(\text{a5Z}+(\text{a5Z}), \text{a6Z}+(\text{a6Z}))$  extrapolated level.



uniform improvement of the lower theory levels of SAPT. However, at the highest levels,  $\text{SAPT2}+(3)(\text{CCD})\delta\text{MP2}$  and  $\text{SAPT2}+3(\text{CCD})\delta\text{MP2}$ , the inclusion of the F12 term reduces the average relative errors in both aDZ and aTZ, and the lowest MA%E value for all approaches (albeit by a very small margin) is obtained for the formally highest-level treatment,  $\text{SAPT2}+3(\text{CCD})\text{-F12}\delta\text{MP2/aTZ}$ . The origins of the relatively poor performance of the highest-level SAPT-F12 variants on the strongly hydrogen bonded HBC6 set warrant further investigation, however, it is likely that a better basis set convergence of dispersion and exchange dispersion impedes a cancellation of errors coming from some non-dispersion SAPT terms.

Overall, the performance of different SAPT variants in the aDZ basis is a net result of the basis set incompleteness errors and the intrinsic errors of a given theory level. Only when the latter errors are sufficiently reduced (most importantly, by including the  $\delta\text{MP2}$  correction), the basis set incompleteness effects on SAPT in general, and on the dispersion and exchange-dispersion energy in particular, become the single dominant factor limiting the accuracy of the aDZ-based treatment. In such a case, the improvement of  $\text{SAPT-F12/aDZ}$  over the corresponding  $\text{SAPT/aDZ}$  variant, observed in Fig. 13, is quite remarkable. When the basis set is enlarged to aTZ, the incompleteness errors are reduced to a magnitude similar to the remaining intrinsic errors of the method. In such a case, the F12 approach, while significantly limiting one source of errors, does not provide an improvement in the accuracy of total interaction energies (it actually somewhat worsens the MAE by disturbing the error cancellation). In particular, the accuracy of  $\text{SAPT2}+(3)\delta\text{MP2/aTZ}$ , which was established as the “gold standard” of SAPT thanks to a very favorable accuracy-to-cost ratio (the accuracy is similar but the computational cost is  $\sim 50\%$  less than that of  $\text{SAPT2}+(\text{CCD})\delta\text{MP2/aTZ}^{54}$ ), cannot be beaten by the  $\text{SAPT-F12}$  treatment at this stage. However, it is likely that further enhancements, in particular, an extension of the F12 formalism to higher-order dispersion corrections including the effects of intramolecular correlation, will provide additional improvement to the  $\text{SAPT-F12}$  accuracy by strongly reducing another source of residual errors. Such an advancement is the subject of ongoing research in our group.

FIG. 10. Relative errors on the A24 database for frozen-core  $E_{\text{exch-disp}}^{(20)}$ -F12 and  $\text{DF-}E_{\text{exch-disp}}^{(20)}$ -F12 computed with the aTZ orbital basis set, the aTZ-RIFIT CABS set, and the aTZ-MP2FIT DF set. The notation +(aXZ) and +(3322) signifies that the hydrogenic aXZ and the constant (3s3p2d2f) set of midbond functions has been added to the aXZ atom-centered basis, respectively. The reference values were obtained at the  $E_{\text{exch-disp}}^{(20)}/(\text{a5Z}+(\text{a5Z}), \text{a6Z}+(\text{a6Z}))$  extrapolated level.



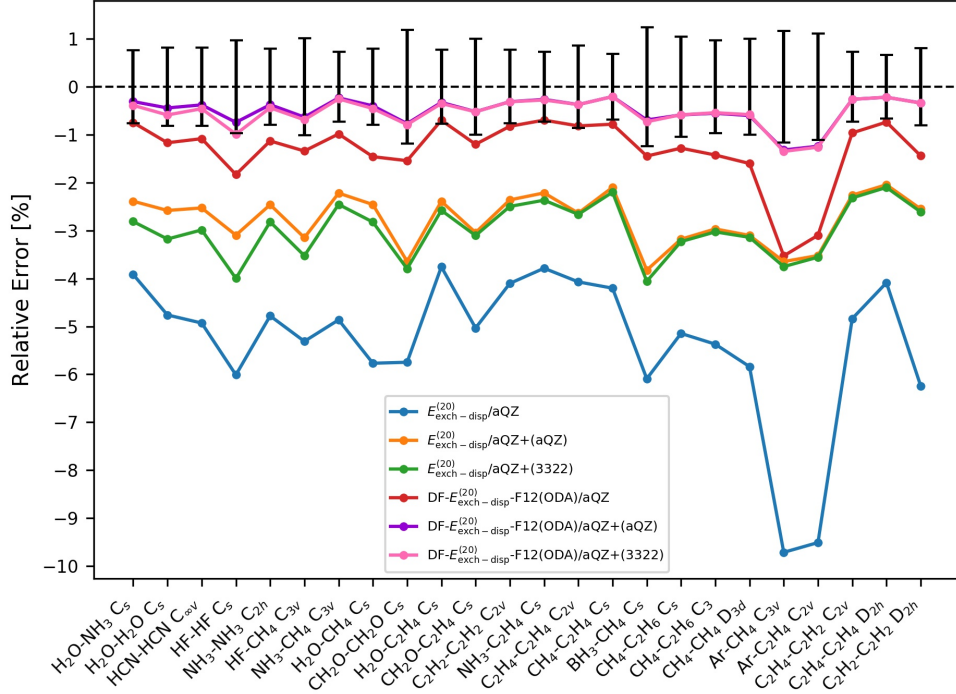
#### D. Computational cost of DF-SAPT0-F12

To conclude this section, we provide some evidence that, even though our current implementation is certainly not a fully optimized production code, a DF-SAPT0-F12 calculation is already more efficient than a standard DF-SAPT0 one in a larger orbital basis required to attain the same accuracy of  $E_{\text{disp}}^{(20)}$  and  $E_{\text{exch-disp}}^{(20)}$ . To this end, Fig. 15 compares the wall timings of DF-SAPT0 (with the highly efficient PSI4 code<sup>61</sup>) and DF-SAPT0-F12 (using our PSI4NUMPY<sup>59</sup> code for the F12 part in addition to a complete PSI4 DF-SAPT0 calculation), obtained for the benzene dimer (from the S22 database<sup>78</sup>) on a single core of a 2.4 GHz Intel Xeon E5-2680 v4 machine. The aDZ timings in Fig. 15 include both calculations with the recommended set of auxiliary bases (aTZ-MP2FIT for DF-SAPT0 and aTZ-RIFIT/aTZ-MP2FIT for DF-SAPT0-F12) and the ones utilizing reduced aDZ-RIFIT/aDZ-MP2FIT sets for CABS and DF. We see that a DF-SAPT0-F12 calculation currently takes a few times longer than a conventional DF-SAPT0 one in the same basis, with DF-SAPT0-F12/aXZ/aXZ-RIFIT slightly less expensive than DF-SAPT0/a(X+1)Z. This indicates that the F12 approach is already a viable alternative to conventional SAPT0 as far as performance is concerned. It should be stressed that our implementation, while completely density fitted, is quite preliminary, and the performance will still be improved when our code is merged into the new release of PSI4 that features a new F12 integral library<sup>61</sup>. Moreover, the overhead associated with calculating  $E_{\text{disp}}^{(20)}$ -F12 and  $E_{\text{exch-disp}}^{(20)}$ -F12 is minor compared to the cost of computing correlated SAPT corrections in case the F12 effects are meant to augment a higher-level SAPT result such as SAPT2+3.

#### V. SUMMARY

In this work, we have transformed the explicitly correlated SAPT0 corrections  $E_{\text{disp}}^{(20)}$ -F12 and  $E_{\text{exch-disp}}^{(20)}$ -F12, proposed recently in Ref. 44 (and, in part, earlier in Ref. 42), into a practical closed-shell SAPT enhancement, significantly improving the basis set convergence relative to conventional  $E_{\text{disp}}^{(20)}$  and  $E_{\text{exch-disp}}^{(20)}$ . While the proof-of-

FIG. 11. Relative errors on the A24 database for frozen-core  $E_{\text{exch-disp}}^{(20)}$ -F12 and  $\text{DF-}E_{\text{exch-disp}}^{(20)}$ -F12 computed with the aQZ orbital basis set, the aQZ-RIFIT CABS set, and the aQZ-MP2FIT DF set. The notation  $+(aXZ)$  and  $+(3322)$  signifies that the hydrogenic  $aXZ$  and the constant  $(3s3p2d2f)$  set of midbond functions has been added to the  $aXZ$  atom-centered basis, respectively. The reference values were obtained at the  $E_{\text{exch-disp}}^{(20)}/(a5Z+(a5Z),a6Z+(a6Z))$  extrapolated level. The error bars indicate the benchmark uncertainty.



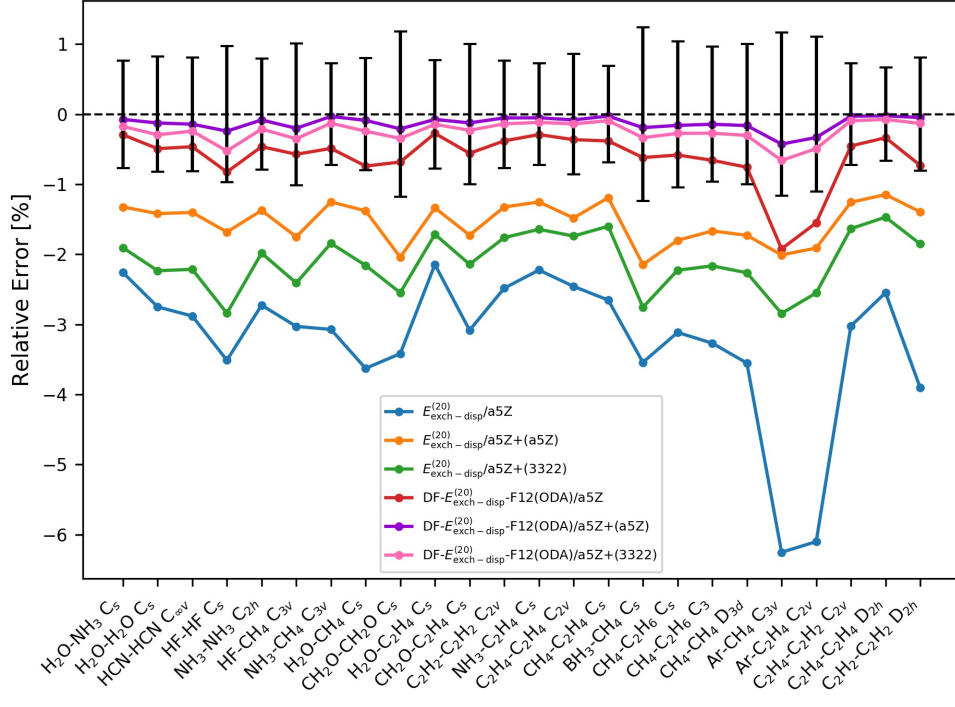
concept SAPT0-F12 implementation of Ref. 44 employed resolution of identity to avoid three-electron integrals, it was not competitive to the conventional SAPT0 approach due to the need to explicitly consider many different two-electron integrals over a combination of orbital and complementary auxiliary basis indices. In the present work, all the two-electron integrals have been decomposed into three-index quantities by means of a robust density fitting, in which the error in the fitted integral decays quadratically with the errors in the fitted densities<sup>48</sup>. Thus, our development of  $\text{DF-}E_{\text{disp}}^{(20)}$ -F12 bears many similarities to the  $\text{DF-MP2-F12}$  formalism, and we employ the so-called *approximation 3C* in the calculation of the F12 intermediates in exactly the same way as in the  $\text{MP2-F12}$  formulation of Ref. 10. Out of several *Ansätze* proposed for the explicitly correlated dispersion amplitudes in Ref. 44, we employ the ODA variant which has been shown to provide an excellent approximation to the results obtained using fully optimized amplitudes, but without the steep scaling and numerical instabilities of the latter.

Numerical tests for several small complexes indicate that the DF approximation to  $E_{\text{disp}}^{(20)}$ -F12 and  $E_{\text{exch-disp}}^{(20)}$ -F12 works very well, with virtually no loss in accuracy as long as the frozen-core approximation is applied. For an  $aXZ$  orbital basis, the standard  $aXZ$ -MP2FIT choice for the DF basis is very appropriate, except that the  $aDZ/aTZ$ -MP2FIT combination is slightly superior to  $aDZ/aDZ$ -MP2FIT. Similar to explicitly correlated electronic structure methods such as  $\text{MP2-F12}$  and  $\text{CCSD(T)-F12}$ , a separate DF basis (from the  $aXZ$ -JKFIT family) can be used for the fitting of the Fock and exchange matrices appearing in some F12 intermediates, however, the results are of exactly the same quality as when the  $aXZ$ -MP2FIT set is used in all DF contexts.

Further tests on the entire A24 database of small noncovalent complexes<sup>62</sup> confirm that the errors introduced by the DF approximation are insignificant in all cases. However, thanks to the vastly superior computational performance of  $\text{DF-SAPT0-F12}$  over the non-DF version, we were able to extend the  $aDZ$  and  $aTZ$  calculations of Ref. 44 to basis sets as large as  $a5Z+(a5Z)$ , that is,  $a5Z$  augmented by a set of midbond functions. The resulting  $E_{\text{disp}}^{(20)}$ -F12 and  $E_{\text{exch-disp}}^{(20)}$ -F12 values are so well converged to the CBS limit that better reference values than those in Ref. 44 became necessary. Thus, we compared the  $E_{\text{disp}}^{(20)}$ -F12 and  $E_{\text{exch-disp}}^{(20)}$ -F12 data for the A24 complexes with conventional  $E_{\text{disp}}^{(20)}$  and  $E_{\text{exch-disp}}^{(20)}$  values obtained by the  $(a5Z+(a5Z),a6Z+(a6Z))$  extrapolation, or even  $(a6Z+(a6Z),a7Z+(a7Z))$  for the



FIG. 12. Relative errors on the A24 database for frozen-core  $E_{\text{exch-disp}}^{(20)}$ -F12 and  $\text{DF-}E_{\text{exch-disp}}^{(20)}$ -F12 computed with the a5Z orbital basis set, the a5Z-RIFIT CABS set, and the a5Z-MP2FIT DF set. The notation +(aXZ) and +(3322) signifies that the hydrogenic aXZ and the constant (3s3p2d2f) set of midbond functions has been added to the aXZ atom-centered basis, respectively. The reference values were obtained at the  $E_{\text{exch-disp}}^{(20)}/(\text{a5Z}+(\text{a5Z}), \text{a6Z}+(\text{a6Z}))$  extrapolated level. The error bars indicate the benchmark uncertainty.



water dimer. However, the CBS convergence of the SAPT-F12 corrections is so fast that nearly all  $E_{\text{disp}}^{(20)}$ -F12/aQZ values are within the tight error bars of the benchmark, and the  $E_{\text{exch-disp}}^{(20)}$ -F12/aQZ data are not far outside the benchmark range. The only exception are the two complexes containing an argon atom for which the convergence is somewhat slower. The CBS convergence of the SAPT-F12 corrections can be further enhanced by the addition of midbond functions. In particular, the +(3322) set of bond functions makes even the  $E_{\text{disp}}^{(20)}$ -F12/aDZ values highly accurate, very often (perhaps fortuitously) within the error bars of the benchmark.

Having established that the F12 approach significantly speeds up the basis set convergence of the individual  $E_{\text{disp}}^{(20)}$  and  $E_{\text{exch-disp}}^{(20)}$  corrections, we turned our attention to the total SAPT interaction energies. In this case, a replacement of the conventional  $E_{\text{disp}}^{(20)} + E_{\text{exch-disp}}^{(20)}$  value by its F12 counterpart corresponds to combining nearly converged  $E_{\text{disp}}^{(20)}$  and  $E_{\text{exch-disp}}^{(20)}$  data with the remaining SAPT corrections computed in a moderate basis, in the spirit of the focal-point MP2/CBS+ $\delta\text{CCSD(T)}$  approach. In this way, we constructed F12-enhanced variants of all standard SAPT levels from SAPT0 to SAPT2+3(CCD) $\delta\text{MP2}^{54}$  and examined the SAPT accuracy, relative to the CCSD(T)/CBS benchmark data, on the same set of complexes as in Ref. 54. We found that low levels of SAPT do not exhibit improvement when  $E_{\text{disp}}^{(20)}$ -F12 and  $E_{\text{exch-disp}}^{(20)}$ -F12 are used in place of  $E_{\text{disp}}^{(20)}$  and  $E_{\text{exch-disp}}^{(20)}$ : in particular, the performance of SAPT0 deteriorates due to a breakdown of the error cancellation between the theory level and basis set incompleteness. Only when the method errors are minimized (which happens at high levels such as SAPT2+(3) or SAPT2+3 including the  $\delta\text{MP2}$  correction), the basis set incompleteness effects determine the overall accuracy. Under those circumstances, SAPT-F12/aDZ is much more accurate than the corresponding SAPT/aDZ approach. In the aTZ basis, incompleteness errors no longer dominate the picture and the F12 and non-F12 variants of SAPT exhibit similar average deviations from the benchmark CCSD(T)/CBS data.

FIG. 13. Mean absolute errors (MAE) (kcal/mol) of the interaction energy averaged over the S22, HBC6, NBC10, and HSG databases for SAPT and SAPT-F12 in aDZ and aTZ basis sets.

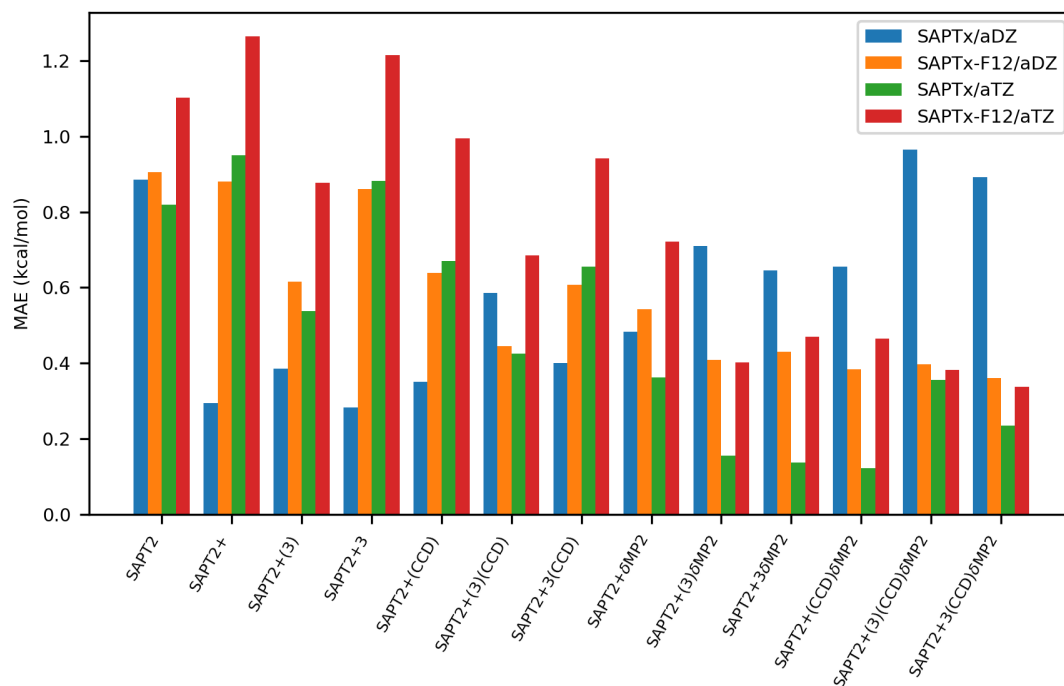


FIG. 14. Mean absolute percent errors (MA%E) of the interaction energy averaged over the S22, HBC6, NBC10, and HSG databases for SAPT and SAPT-F12 in aDZ and aTZ basis sets.

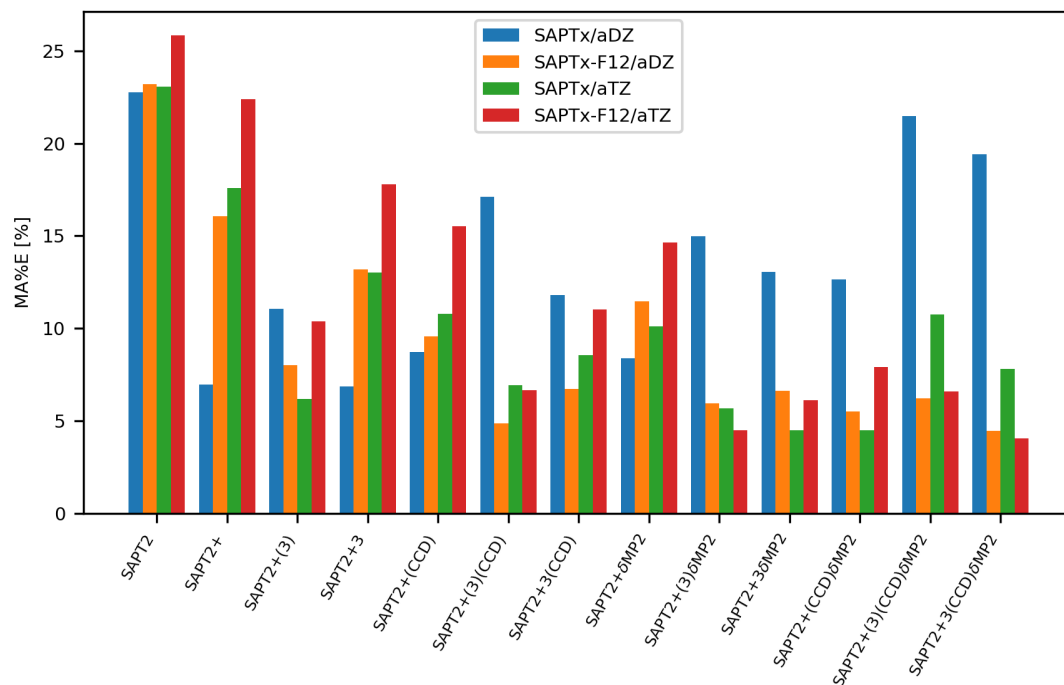
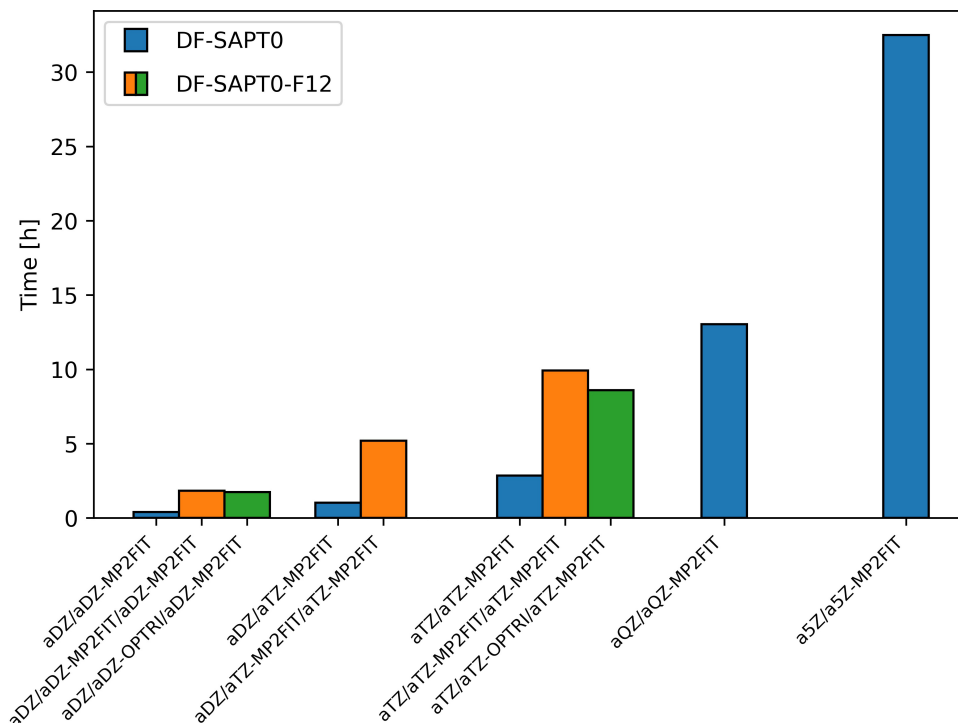


FIG. 15. The comparison of DF-SAPT0 and DF-SAPT0-F12 calculation times (in hours) for the benzene dimer (from the S22 database<sup>78</sup>) on a single core of a 2.4 GHz Intel Xeon E5-2680 v4 machine using the Psi4 quantum chemistry program and our Psi4NumPy implementation. The consecutive basis sets listed for each bar pertain to the AO and DF sets for DF-SAPT0 and to the AO, CABS, and DF sets for DF-SAPT0-F12.



## ASSOCIATED CONTENT

**Supporting Information.** Extended versions of Tables II–III including data for noble gas dimers, results with midbond functions, values obtained with the aXZ-JKFIT auxiliary basis in place of aXZ-MP2FIT in the construction of Fock and exchange matrices, values computed with the aXZ-OPTRI CABS sets, and the  $E_{\text{disp}}^{(20)} + E_{\text{exch-disp}}^{(20)}$  sums. Figures displaying the basis set convergence of  $E_{\text{disp}}^{(20)}$ -F12,  $E_{\text{exch-disp}}^{(20)}$ -F12, and the sum of the two for all individual A24 complexes. Mean absolute errors and mean absolute percent errors of different SAPT and SAPT-F12 variants for the individual databases (S22, HBC6, NBC10, and HSG). This material is available free of charge via the Internet at <http://pubs.acs.org/>.

## ACKNOWLEDGMENTS

This work was supported by the U.S. National Science Foundation (NSF) awards CHE-1351978 and CHE-1955328. M. K. acknowledges a fellowship from the Molecular Sciences Software Institute (MolSSI) under NSF Grant ACI-1547580. We thank Dr. Wim Klopper and Dr. Christof Holzer for helpful discussions.

\* Present address: Department of Chemistry, Virginia Tech, Blacksburg, Virginia 24061

† patkowski@auburn.edu

<sup>1</sup> Klein, A.; Shagam, Y.; Skomorowski, W.; Żuchowski, P. S.; Pawlak, M.; Janssen, L. M. C.; Moiseyev, N.; van de Meerakker, S. Y. T.; van der Avoird, A.; Koch, C. P.; Narevicius, E. Directly probing anisotropy in atom-molecule collisions through quantum scattering resonances. *Nature Phys.* **2017**, *13*, 35–38.

<sup>2</sup> Beran, G. J. O. Modeling Polymorphic Molecular Crystals with Electronic Structure Theory. *Chem. Rev.* **2016**, *116*, 5567–5613.

- <sup>3</sup> Kraka, E.; Freindorf, M.; Cremer, D. Chiral Discrimination by Vibrational Spectroscopy Utilizing Local Modes. *Chirality* **2013**, *25*, 185–196.
- <sup>4</sup> Hobza, P.; Selzle, H. L.; Schlag, E. W. Potential Energy Surface for the Benzene Dimer. Results of ab Initio CCSD(T) Calculations Show Two Nearly Isoenergetic Structures: T-Shaped and Parallel-Displaced. *J. Phys. Chem.* **1996**, *100*, 18790–18794.
- <sup>5</sup> Kodrycka, M.; Patkowski, K. Platinum, gold, and silver standards of intermolecular interaction energy calculations. *J. Chem. Phys.* **2019**, *151*, 070901.
- <sup>6</sup> Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M. A 5th-Order Perturbation Comparison of Electron Correlation Theories. *Chem. Phys. Lett.* **1989**, *157*, 479–483.
- <sup>7</sup> Halkier, A.; Helgaker, T.; Jørgensen, P.; Klopper, W.; Koch, H.; Olsen, J.; Wilson, A. K. Basis-Set Convergence in Correlated Calculations on Ne, N<sub>2</sub>, and H<sub>2</sub>O. *Chem. Phys. Lett.* **1998**, *286*, 243–252.
- <sup>8</sup> Kutzelnigg, W.; Klopper, W. Wave functions with terms linear in the interelectronic coordinates to take care of the correlation cusp. I. General theory. *J. Chem. Phys.* **1991**, *94*, 1985–2001.
- <sup>9</sup> Tew, D. P.; Klopper, W. New correlation factors for explicitly correlated electronic wave functions. *J. Chem. Phys.* **2005**, *123*, 074101.
- <sup>10</sup> Werner, H.-J.; Adler, T. B.; Manby, F. R. General orbital invariant MP2-F12 theory. *J. Chem. Phys.* **2007**, *126*, 164102.
- <sup>11</sup> Shiozaki, T.; Kamiya, M.; Hirata, S.; Valeev, E. F. Explicitly correlated coupled-cluster singles and doubles method based on complete diagrammatic equations. *J. Chem. Phys.* **2008**, *129*, 071101.
- <sup>12</sup> Köhn, A. Explicitly correlated connected triple excitations in coupled-cluster theory. *J. Chem. Phys.* **2009**, *130*, 131101.
- <sup>13</sup> Shiozaki, T.; Werner, H.-J. Second-order multireference perturbation theory with explicit correlation: CASPT2-F12. *J. Chem. Phys.* **2010**, *133*, 141103.
- <sup>14</sup> Shiozaki, T.; Knizia, G.; Werner, H.-J. Explicitly correlated multireference configuration interaction: MRCI-F12. *J. Chem. Phys.* **2011**, *134*, 034113.
- <sup>15</sup> Tao, F.-M. Bond functions, basis set superposition errors and other practical issues with ab initio calculations of intermolecular potentials. *Int. Rev. Phys. Chem.* **2001**, *20*, 617–643.
- <sup>16</sup> Williams, H. L.; Mas, E. M.; Szalewicz, K.; Jeziorski, B. On the effectiveness of monomer-, dimer-, and bond-centered basis functions in calculations of intermolecular interaction energies. *J. Chem. Phys.* **1995**, *103*, 7374–7391.
- <sup>17</sup> Jeziorska, M.; Cencek, W.; Patkowski, K.; Jeziorski, B.; Szalewicz, K. Complete Basis Set Extrapolations of Dispersion, Exchange, and Coupled-Clusters Contributions to the Interaction Energy: A Helium Dimer Study. *Int. J. Quantum Chem.* **2008**, *108*, 2053–2075.
- <sup>18</sup> Dutta, N. N.; Patkowski, K. Improving ‘Silver-Standard’ Benchmark Interaction Energies with Bond Functions. *J. Chem. Theory Comput.* **2018**, *14*, 3053–3070.
- <sup>19</sup> Stone, A. J. *The Theory of Intermolecular Forces, Second Edition*; Oxford University Press, 2013.
- <sup>20</sup> Jeziorski, B.; Moszyński, R.; Szalewicz, K. Perturbation Theory Approach to Intermolecular Potential Energy Surfaces of van der Waals Complexes. *Chem. Rev.* **1994**, *94*, 1887–1930.
- <sup>21</sup> Patkowski, K. Recent developments in symmetry-adapted perturbation theory. *WIREs Comput. Mol. Sci.* **2020**, *10*, e1452.
- <sup>22</sup> Korona, T.; Moszyński, R.; Jeziorski, B. Convergence of symmetry-adapted perturbation theory for the interaction between helium atoms and between a hydrogen molecule and a helium atom. *Adv. Quantum Chem.* **1997**, *28*, 171–188.
- <sup>23</sup> Patkowski, K.; Korona, T.; Jeziorski, B. Convergence behavior of the symmetry-adapted perturbation theory for states submerged in Pauli forbidden continuum. *J. Chem. Phys.* **2001**, *115*, 1137–1152.
- <sup>24</sup> Chałasiński, G.; Szczęśniak, M. M. On the connection between the supermolecular Møller-Plesset treatment of the interaction energy and the perturbation theory of intermolecular forces. *Mol. Phys.* **1988**, *63*, 205–224.
- <sup>25</sup> Rybak, S.; Jeziorski, B.; Szalewicz, K. Many-body symmetry-adapted perturbation theory of intermolecular interactions - H<sub>2</sub>O and HF dimers. *J. Chem. Phys.* **1991**, *95*, 6579–6601.
- <sup>26</sup> Williams, H. L.; Szalewicz, K.; Moszyński, R.; Jeziorski, B. Dispersion Energy in the Coupled Pair Approximation with Noniterative Inclusion of Single and Triple Excitations. *J. Chem. Phys.* **1995**, *103*, 4586–4599.
- <sup>27</sup> Korona, T.; Jeziorski, B. Dispersion energy from density-fitted density susceptibilities of singles and doubles coupled cluster theory. *J. Chem. Phys.* **2008**, *128*, 144107.
- <sup>28</sup> Hesselmann, A.; Jansen, G.; Schütz, M. Density-functional theory-symmetry-adapted intermolecular perturbation theory with density fitting: A new efficient method to study intermolecular interaction energies. *J. Chem. Phys.* **2005**, *122*, 014103.
- <sup>29</sup> Misquitta, A. J.; Podeszwa, R.; Jeziorski, B.; Szalewicz, K. Intermolecular potentials based on symmetry-adapted perturbation theory including dispersion energies from time-dependent density functional calculations. *J. Chem. Phys.* **2005**, *123*, 214103.
- <sup>30</sup> Holzer, C.; Klopper, W. Communication: Symmetry-adapted perturbation theory with intermolecular induction and dispersion energies from the Bethe-Salpeter equation. *J. Chem. Phys.* **2017**, *147*, 181101.
- <sup>31</sup> Hapka, M.; Przybytek, M.; Pernal, K. Second-Order Dispersion Energy Based on Multireference Description of Monomers. *J. Chem. Theory Comput.* **2019**, *15*, 1016–1027.
- <sup>32</sup> Korona, T. Exchange-Dispersion Energy: A Formulation in Terms of Monomer Properties and Coupled Cluster Treatment of Intramonomer Correlation. *J. Chem. Theory Comput.* **2009**, *5*, 2663–2678.
- <sup>33</sup> Schäffer, R.; Jansen, G. Single-determinant-based symmetry-adapted perturbation theory without single-exchange approximation. *Mol. Phys.* **2013**, *111*, 2570–2584.
- <sup>34</sup> Marchetti, O.; Werner, H.-J. Accurate calculations of intermolecular interaction energies using explicitly correlated wave functions. *Phys. Chem. Chem. Phys.* **2008**, *10*, 3400–3409.

- <sup>35</sup> Marchetti, O.; Werner, H.-J. Accurate Calculations of Intermolecular Interaction Energies Using Explicitly Correlated Coupled Cluster Wave Functions and a Dispersion-Weighted MP2 Method. *J. Phys. Chem. A* **2009**, *113*, 11580–11585.
- <sup>36</sup> Patkowski, K. Basis Set Converged Weak Interaction Energies from Conventional and Explicitly Correlated Coupled-Cluster Approach. *J. Chem. Phys.* **2013**, *138*, 154101.
- <sup>37</sup> Sirianni, D. A.; Burns, L. A.; Sherrill, C. D. Comparison of Explicitly Correlated Methods for Computing High-Accuracy Benchmark Energies for Noncovalent Interactions. *J. Chem. Theory Comput.* **2017**, *13*, 86–99.
- <sup>38</sup> Szalewicz, K.; Jeziorski, B. Symmetry-adapted double-perturbation analysis of intramolecular correlation effects in weak intermolecular interactions: the He-He interaction. *Mol. Phys.* **1979**, *38*, 191–208.
- <sup>39</sup> Korona, T.; Williams, H. L.; Bukowski, R.; Jeziorski, B.; Szalewicz, K. Helium dimer potential from symmetry-adapted perturbation theory calculations using large Gaussian geminal and orbital basis sets. *J. Chem. Phys.* **1997**, *106*, 5109–5122.
- <sup>40</sup> Jeziorska, M.; Cencek, W.; Patkowski, K.; Jeziorski, B.; Szalewicz, K. Pair potential for helium from symmetry-adapted perturbation theory calculations and from supermolecular data. *J. Chem. Phys.* **2007**, *127*, 124303.
- <sup>41</sup> Mitroy, J.; Bubin, S.; Horiuchi, W.; Suzuki, Y.; Adamowicz, L.; Cencek, W.; Szalewicz, K.; Komasa, J.; Blume, D.; Varga, K. Theory and application of explicitly correlated Gaussians. *Rev. Mod. Phys.* **2013**, *85*, 693–749.
- <sup>42</sup> Przybytek, M. Dispersion Energy of Symmetry-Adapted Perturbation Theory from the Explicitly Correlated F12 Approach. *J. Chem. Theory Comput.* **2018**, *14*, 5105–5117.
- <sup>43</sup> Kendall, R. A.; Dunning Jr., T. H.; Harrison, R. J. Electron Affinities of the 1st-Row Atoms Revisited - Systematic Basis Sets and Wave Functions. *J. Chem. Phys.* **1992**, *96*, 6796–6806.
- <sup>44</sup> Kodrycka, M.; Holzer, C.; Klopper, W.; Patkowski, K. Explicitly correlated dispersion and exchange dispersion energies in symmetry-adapted perturbation theory. *J. Chem. Theory Comput.* **2019**, *15*, 5965–5986.
- <sup>45</sup> Valeev, E. F. Improving on the resolution of the identity in linear R12 ab initio theories. *Chem. Phys. Lett.* **2004**, *395*, 190–195.
- <sup>46</sup> Whitten, J. L. Coulombic potential energy integrals and approximations. *J. Chem. Phys.* **1973**, *58*, 4496–4501.
- <sup>47</sup> Dunlap, B. I.; Connolly, J. W. D.; Sabin, J. R. On first-row diatomic molecules and local density models. *J. Chem. Phys.* **1979**, *71*, 4993–4999.
- <sup>48</sup> Manby, F. R. Density fitting in second-order linear-R12 Møller-Plesset perturbation theory. *J. Chem. Phys.* **2003**, *119*, 4607–4613.
- <sup>49</sup> May, A. J.; Manby, F. R. An explicitly correlated second order Møller-Plesset theory using a frozen Gaussian geminal. *J. Chem. Phys.* **2004**, *121*, 4479–4485.
- <sup>50</sup> Adler, T. B.; Knizia, G.; Werner, H.-J. A Simple and Efficient CCSD(T)-F12 Approximation. *J. Chem. Phys.* **2007**, *127*, 221106.
- <sup>51</sup> Knizia, G.; Adler, T. B.; Werner, H.-J. Simplified CCSD(T)-F12 Methods: Theory and Benchmarks. *J. Chem. Phys.* **2009**, *130*, 054104.
- <sup>52</sup> Hättig, C.; Tew, D. P.; Köhn, A. Accurate and efficient approximations to explicitly correlated coupled-cluster singles and doubles, CCSD-F12. *J. Chem. Phys.* **2010**, *132*, 231102.
- <sup>53</sup> Papajak, E.; Zheng, J.; Xu, X.; Leverentz, H. R.; Truhlar, D. G. Perspectives on Basis Sets Beautiful: Seasonal Plantings of Diffuse Basis Functions. *J. Chem. Theory Comput.* **2011**, *7*, 3027–3034.
- <sup>54</sup> Parker, T. M.; Burns, L. A.; Parrish, R. M.; Ryno, A. G.; Sherrill, C. D. Levels of Symmetry Adapted Perturbation Theory (SAPT). I. Efficiency and Performance for Interaction Energies. *J. Chem. Phys.* **2014**, *140*, 094106.
- <sup>55</sup> Riley, K. E.; Pitoňák, M.; Jurečka, P.; Hobza, P. Stabilization and Structure Calculations for Noncovalent Interactions in Extended Molecular Systems Based on Wave Function and Density Functional Theories. *Chem. Rev.* **2010**, *110*, 5023–5063.
- <sup>56</sup> Marshall, M. S.; Burns, L. A.; Sherrill, C. D. Basis Set Convergence of the Coupled-Cluster Correction,  $\delta_{\text{MP2}}^{\text{CCSD(T)}}$ : Best Practices for Benchmarking Non-Covalent Interactions and the Attendant Revision of the S22, NBC10, HBC6, and HSG Databases. *J. Chem. Phys.* **2011**, *135*, 194102.
- <sup>57</sup> Tajti, A.; Szalay, P. G.; Császár, A. G.; Kállay, M.; Gauss, J.; Valeev, E. F.; Flowers, B. A.; Vázquez, J.; Stanton, J. F. HEAT: High accuracy extrapolated ab initio thermochemistry. *J. Chem. Phys.* **2004**, *121*, 11599–11613.
- <sup>58</sup> Karton, A.; Rabinovich, E.; Martin, J. M. L.; Ruscic, B. W4 theory for computational thermochemistry: In pursuit of confident sub-kJ/mol predictions. *J. Chem. Phys.* **2006**, *125*, 144108.
- <sup>59</sup> Smith, D. G. A. et al. PSI4NUMPY: An Interactive Quantum Chemistry Programming Environment for Reference Implementations and Rapid Development. *J. Chem. Theory Comput.* **2018**, *14*, 3504–3511.
- <sup>60</sup> Parrish, R. M. et al. PSI4 1.1: An Open-Source Electronic Structure Program Emphasizing Automation, Advanced Libraries, and Interoperability. *J. Chem. Theory Comput.* **2017**, *13*, 3185–3197.
- <sup>61</sup> Smith, D. G. A. et al. PSI4 1.4: Open-source software for high-throughput quantum chemistry. *J. Chem. Phys.* **2020**, *152*, 184108.
- <sup>62</sup> Řezáč, J.; Hobza, P. Describing Noncovalent Interactions beyond the Common Approximations: How Accurate Is the 'Gold Standard,' CCSD(T) at the Complete Basis Set Limit? *J. Chem. Theory Comput.* **2013**, *9*, 2151–2155.
- <sup>63</sup> Hättig, C.; Klopper, W.; Köhn, A.; Tew, D. P. Explicitly Correlated Electrons in Molecules. *Chem. Rev.* **2012**, *112*, 4–74.
- <sup>64</sup> Tew, D. P.; Klopper, W. A comparison of linear and nonlinear correlation factors for basis set limit Møller-Plesset second order binding energies and structures of He<sub>2</sub>, Be<sub>2</sub>, and Ne<sub>2</sub>. *J. Chem. Phys.* **2006**, *125*, 094302.
- <sup>65</sup> Ten-no, S. Initiation of explicitly correlated Slater-type geminal theory. *Chem. Phys. Lett.* **2004**, *398*, 56–61.
- <sup>66</sup> Höfener, S.; Bischoff, F. A.; Glöß, A.; Klopper, W. Slater-type geminals in explicitly-correlated perturbation theory: application to *n*-alkanols and analysis of errors and basis-set requirements. *Phys. Chem. Chem. Phys.* **2008**, *10*, 3390–3399.

- <sup>67</sup> Dunning Jr., T. H. Gaussian-Basis Sets for Use in Correlated Molecular Calculations. 1. The Atoms Boron through Neon and Hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007–1023.
- <sup>68</sup> Weigend, F.; Köhn, A.; Hättig, C. Efficient use of the correlation consistent basis sets in resolution of the identity MP2 calculations. *J. Chem. Phys.* **2002**, *116*, 3175–3183.
- <sup>69</sup> Hättig, C. Optimization of auxiliary basis sets for RI-MP2 and RI-CC2 calculations: Core-valence and quintuple-zeta basis sets for H to Ar and QZVPP basis sets for Li to Kr. *Phys. Chem. Chem. Phys.* **2005**, *7*, 59–66.
- <sup>70</sup> 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.
- <sup>71</sup> Hohenstein, E. G.; Sherrill, C. D. Density fitting and Cholesky decomposition approximations in symmetry-adapted perturbation theory: Implementation and application to probe the nature of pi-pi interactions in linear acenes. *J. Chem. Phys.* **2010**, *132*, 184111.
- <sup>72</sup> Helgaker, T.; Klopper, W.; Koch, H.; Noga, J. Basis-set convergence of correlated calculations on water. *J. Chem. Phys.* **1997**, *106*, 9639–9646.
- <sup>73</sup> Mielke, S. L.; Garrett, B. C.; Peterson, K. A. A hierarchical family of global analytic Born-Oppenheimer potential energy surfaces for the H+H<sub>2</sub> reaction ranging in quality from double-zeta to the complete basis set limit. *J. Chem. Phys.* **2002**, *116*, 4142–4161.
- <sup>74</sup> Podeszwa, R.; Bukowski, R.; Szalewicz, K. Potential energy surface for the benzene dimer and perturbational analysis of  $\pi - \pi$  interactions. *J. Phys. Chem. A* **2006**, *110*, 10345–10354.
- <sup>75</sup> Akin-Ojo, O.; Bukowski, R.; Szalewicz, K. Ab Initio Studies of He-HCCCN Interaction. *J. Chem. Phys.* **2003**, *119*, 8379–8396.
- <sup>76</sup> Patkowski, K. Benchmark Databases of Intermolecular Interaction Energies: Design, Construction, and Significance. In *Annual Reports in Computational Chemistry*; Dixon, D. A., Ed.; Elsevier, Amsterdam, 2017; Vol. 13; pp 3–91.
- <sup>77</sup> Parrish, R. M.; Hohenstein, E. G.; Sherrill, C. D. Tractability gains in symmetry-adapted perturbation theory including coupled double excitations: CCD+ST(CCD) dispersion with natural orbital truncations. *J. Chem. Phys.* **2013**, *139*, 174102.
- <sup>78</sup> Jurečka, P.; Šponer, J.; Černý, J.; Hobza, P. Benchmark database of accurate (MP2 and CCSD(T) complete basis set limit) interaction energies of small model complexes, DNA base pairs, and amino acid pairs. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1985–1993.
- <sup>79</sup> Thanthiriwatt, K. S.; Hohenstein, E. G.; Burns, L. A.; Sherrill, C. D. Assessment of the Performance of DFT and DFT-D Methods for Describing Distance Dependence of Hydrogen-Bonded Interactions. *J. Chem. Theory Comput.* **2011**, *7*, 88–96.
- <sup>80</sup> Sherrill, C. D.; Takatani, T.; Hohenstein, E. G. An Assessment of Theoretical Methods for Nonbonded Interactions: Comparison to Complete Basis Set Limit Coupled-Cluster Potential Energy Curves for the Benzene Dimer, the Methane Dimer, Benzene-Methane, and Benzene-H<sub>2</sub>S. *J. Phys. Chem. A* **2009**, *113*, 10146–10159.
- <sup>81</sup> Faver, J. C.; Benson, M. L.; He, X.; Roberts, B. P.; Wang, B.; Marshall, M. S.; Kennedy, M. R.; Sherrill, C. D.; Merz, Jr., K. M. Formal Estimation of Errors in Computed Absolute Interaction Energies of Protein-Ligand Complexes. *J. Chem. Theory Comput.* **2011**, *7*, 790–797.
- <sup>82</sup> Burns, L. A.; Faver, J. C.; Zheng, Z.; Marshall, M. S.; Smith, D. G. A.; Vanommeslaeghe, K.; MacKerell, Jr., A. D.; Merz, Jr., K. M.; Sherrill, C. D. The BioFragment Database (BFDdb): An open-data platform for computational chemistry analysis of noncovalent interactions. *J. Chem. Phys.* **2017**, *147*, 161727.
- <sup>83</sup> Yousaf, K. E.; Peterson, K. A. Optimized complementary auxiliary basis sets for explicitly correlated methods: aug-cc-pVnZ orbital basis sets. *Chem. Phys. Lett.* **2009**, *476*, 303–307.
- <sup>84</sup> Peterson, K. A.; Dunning, Jr., T. H. Accurate correlation consistent basis sets for molecular core-valence correlation effects: The second row atoms Al-Ar, and the first row atoms B-Ne revisited. *J. Chem. Phys.* **2002**, *117*, 10548–10560.
- <sup>85</sup> Korona, T. A coupled cluster treatment of intramonomer electron correlation within symmetry-adapted perturbation theory: benchmark calculations and a comparison with a density-functional theory description. *Mol. Phys.* **2013**, *111*, 3705–3715.

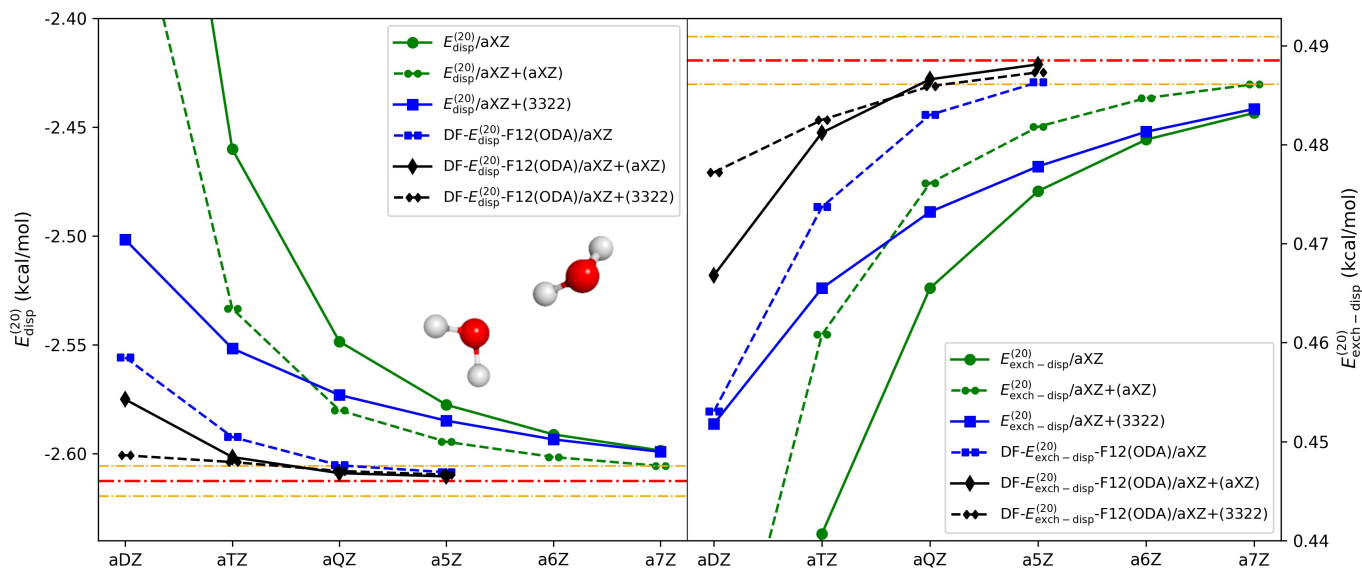


FIG. 16. TOC Graphic