

# Density fitting of intramonomer correlation effects in symmetry-adapted perturbation theory

Edward G. Hohenstein and C. David Sherrill\*

Citation: [The Journal of Chemical Physics](#) **133**, 014101 (2010); doi: 10.1063/1.3451077

View online: <http://dx.doi.org/10.1063/1.3451077>

View Table of Contents: <http://aip.scitation.org/toc/jcp/133/1>

Published by the [American Institute of Physics](#)

---

---



**COMPLETELY  
REDESIGNED!**

PHYSICS  
TODAY

*Physics Today* Buyer's Guide  
Search with a purpose.

# Density fitting of intramonomer correlation effects in symmetry-adapted perturbation theory

Edward G. Hohenstein<sup>1</sup> and C. David Sherrill<sup>1,2,a)</sup>

<sup>1</sup>Center for Computational Molecular Science and Technology, School of Chemistry and Biochemistry, Georgia Institute of Technology, Atlanta, Georgia 30332-0400, USA

<sup>2</sup>College of Computing, Georgia Institute of Technology, Atlanta, Georgia 30332-0280, USA

(Received 9 April 2010; accepted 20 May 2010; published online 1 July 2010)

Symmetry-adapted perturbation theory (SAPT) offers insight into the nature of intermolecular interactions. In addition, accurate energies can be obtained from the wave function-based variant of SAPT provided that intramonomer electron correlation effects are included. We apply density-fitting (DF) approximations to the intramonomer correlation corrections in SAPT. The introduction of this approximation leads to an improvement in the computational cost of SAPT by reducing the scaling of certain SAPT terms, reducing the amount of disk I/O, and avoiding the explicit computation of certain types of MO integrals. We have implemented all the intramonomer correlation corrections to SAPT through second-order under the DF approximation. Additionally, leading third-order terms are also implemented. The accuracy of this truncation of SAPT is tested against the S22 test set of Hobza and co-workers [Phys. Chem. Chem. Phys. **8**, 1985 (2006)]. When the intramonomer corrections to dispersion are included in SAPT, a mean absolute deviation of 0.3–0.4 kcal mol<sup>−1</sup> is observed for the S22 test set when using an aug-cc-pVDZ basis. The computations on the adenine-thymine complexes in the S22 test set with an aug-cc-pVDZ basis represent the largest SAPT computations to date that include this degree of intramonomer correlation. Computations of this size can now be performed routinely with our newly developed DF-SAPT program. © 2010 American Institute of Physics. [doi:10.1063/1.3451077]

## I. INTRODUCTION

The ability to perform accurate and efficient theoretical studies of noncovalent interactions is important for many areas of chemistry. Of particular interest is the hydrogen bonding and stacking of nucleic acid base pairs in DNA,<sup>1–6</sup> the interactions between the side chains of amino acids that influence the structure of proteins,<sup>7–11</sup> drug binding,<sup>12</sup> the structure and lattice energy of organic crystals,<sup>13–16</sup> and intercalation phenomena.<sup>2,17–19</sup> In order to accurately describe noncovalent interactions, coupled-cluster with singles and doubles including perturbative triples [CCSD(T)] (Ref. 20) is relied upon. Beyond computing highly accurate interaction energies, it is often useful to obtain a decomposition of the energy. There are various energy decomposition techniques available,<sup>21–29</sup> but perhaps the most well-defined and robust is the symmetry-adapted perturbation theory (SAPT).<sup>26</sup>

A review by Jezierski *et al.*<sup>26</sup> describes the development and applications of wave function-based SAPT. In order to obtain accurate interaction energies from SAPT, there must be some account of the intramonomer electron correlation. The wave function-based formulation of SAPT including second-order intramonomer electron correlation corrections has been influential in the understanding of  $\pi$ - $\pi$  interactions,<sup>30–35</sup> XH- $\pi$  interactions,<sup>10,36–38</sup> and ion- $\pi$  interactions.<sup>39,40</sup> Despite the successes of this method, current applications are typically limited to systems no larger than substituted benzene dimers. Recent work by Singh *et*

*al.*<sup>41</sup> included a particularly extensive application of wave function-based SAPT. The largest system studied in this work was the benzene dimethyl-bipyridinium complex with a 6-31+G\* basis. This system contains 40 atoms and roughly 500 basis functions; it is likely the largest wave function-based SAPT computation performed to date that includes some account of intramonomer electron correlation. In that work, however, only the electrostatic and exchange terms are corrected to account for intramonomer electron correlation.

In order to improve the efficiency of SAPT, the density-fitting (DF) approximation (DF, also called resolution-of-the-identity or RI)<sup>42–47</sup> can be applied. Previously, this approximation has been applied to SAPT in the context of SAPT(DFT).<sup>27–29</sup> More recently, we have applied the DF approximation to the simplest form of wave function-based SAPT, SAPT0.<sup>48</sup> In both cases, the use of DF integrals has improved the performance of the SAPT programs while introducing negligible errors.

In the present work, we apply the DF approximation to the intramonomer electron correlation corrections in SAPT. We are able to use DF to increase the efficiency of the evaluation of the molecular integrals required for SAPT computations. Additionally, the DF approximation can be inserted into the SAPT energy equations to reduce the scaling or improve efficiency through reduced disk I/O. This is the first report of DF applied to the intramonomer electron correlation corrections to SAPT. With the DF-SAPT program whose development is reported in this work, we have performed what we believe are the largest SAPT computations to date

<sup>a)</sup>Electronic mail: sherrill@gatech.edu.

that include this degree of electron correlation. We apply our program to the S22 test set of Hobza and co-workers<sup>6</sup> (using the recently revised interaction energies<sup>49</sup>) in order to benchmark the accuracy of various truncations of SAPT theory.

## II. THEORETICAL METHODS

SAPT provides a means of directly computing noncovalent interactions between molecules. For each monomer, one defines a monomer Hamiltonian,  $H_A$  (for monomer  $A$ ), as  $H_A = F_A + W_A$ , where  $F_A$  is the usual Fock operator for nuclei and electrons associated with monomer  $A$ , and  $W_A$  is the fluctuation potential accounting for electron correlation on monomer  $A$ . An intermolecular operator,  $V$ , is defined as the difference between the full dimer Hamiltonian and the sum of two monomer Hamiltonians, so that

$$H = F_A + W_A + F_B + W_B + V. \quad (1)$$

For a more complete description of SAPT, we refer the reader to Ref. 26. In previous work,<sup>48</sup> we described the application of DF to SAPT0, which neglects  $W_A$  and  $W_B$  and includes terms up to second-order in  $V$ . In this work we will focus on the SAPT terms that are first and second-order in  $W$ . We will group the individual terms into physically meaningful components

$$E_{\text{electrostatic}} = E_{\text{elst}}^{(10)} + E_{\text{elst,resp}}^{(12)} + E_{\text{elst,resp}}^{(13)}, \quad (2)$$

$$E_{\text{exchange}} = E_{\text{exch}}^{(10)} + E_{\text{exch}}^{(11)} + E_{\text{exch}}^{(12)}, \quad (3)$$

$$E_{\text{induction}} = E_{\text{ind,resp}}^{(20)} + E_{\text{exch-ind,resp}}^{(20)} + {}^tE_{\text{ind}}^{(22)} + {}^tE_{\text{exch-ind}}^{(22)} + \delta E_{\text{ind,resp}}^{(\text{HF})}, \quad (4)$$

$$E_{\text{dispersion}} = E_{\text{disp}}^{(20)} + E_{\text{disp}}^{(30)} + E_{\text{disp}}^{(21)} + E_{\text{disp}}^{(22)} + E_{\text{exch-disp}}^{(20)}, \quad (5)$$

where the first number in parentheses denotes the perturbation order in  $V$ , and the second number in parentheses denotes the perturbation order in  $W$ . When three numbers appear in parentheses, the first number still denotes the perturbation order in  $V$ , while the second and third numbers denote perturbation order in  $W_A$  and  $W_B$ , respectively. Definitions and physical interpretations of the individual terms shown above are given in Ref. 50. Often, a  $\delta E_{\text{ind,resp}}^{(\text{HF})}$  term is computed from the Hartree–Fock (HF) interaction energy as

$$\delta E_{\text{ind,resp}}^{(\text{HF})} = E_{\text{int}}^{\text{HF}} - (E_{\text{elst}}^{(10)} + E_{\text{exch}}^{(10)} + E_{\text{ind,resp}}^{(20)} + E_{\text{exch-ind,resp}}^{(20)}). \quad (6)$$

To a first approximation, this term captures induced-multipole induced-multipole interactions that are not described by the  $E_{\text{ind,resp}}^{(20)}$  and  $E_{\text{exch-ind,resp}}^{(20)}$  terms, so it will be grouped as an induction term.

Additionally, we will define groupings of terms which can be used to compute interaction energies with increasing accuracy

$$E_{\text{SAPT0}} = E_{\text{elst}}^{(10)} + E_{\text{exch}}^{(10)} + E_{\text{ind,resp}}^{(20)} + E_{\text{exch-ind,resp}}^{(20)} + E_{\text{disp}}^{(20)} + E_{\text{exch-disp}}^{(20)}, \quad (7)$$

$$E_{\text{SAPT2}} = E_{\text{SAPT0}} + E_{\text{elst,resp}}^{(12)} + E_{\text{exch}}^{(11)} + E_{\text{exch}}^{(12)} + {}^tE_{\text{ind}}^{(22)} + {}^tE_{\text{exch-ind}}^{(22)}, \quad (8)$$

$$E_{\text{SAPT2+}} = E_{\text{SAPT2}} + E_{\text{disp}}^{(21)} + E_{\text{disp}}^{(22)}, \quad (9)$$

$$E_{\text{SAPT2+(3)}} = E_{\text{SAPT2+}} + E_{\text{elst,resp}}^{(13)} + E_{\text{disp}}^{(30)}. \quad (10)$$

The SAPT0 and SAPT2 groupings are commonly used in the literature, although the second-order induction terms are not always included in SAPT2. In order to apply SAPT0 and SAPT2 to chemical problems, one must be careful to pair these methods with an appropriate basis set. Additionally, the  $\delta E_{\text{ind,resp}}^{(\text{HF})}$  term should be included with SAPT0 and SAPT2 unless omitting it, through fortuitous error cancellation, results in a more accurate interaction energy. For SAPT2+ and SAPT2+(3), the dispersion energy is corrected for intramonomer electron correlation and one does not need to rely on error cancellation to compute accurate interaction energies. The  $\delta E_{\text{ind,resp}}^{(\text{HF})}$  term will always be added to the SAPT2+ and SAPT2+(3) energies in this work. A thorough discussion and derivation of third-order interactions in SAPT, of which we consider only  $E_{\text{disp}}^{(30)}$ , can be found Ref. 51. It should be noted that  $E_{\text{ind}}^{(30)}$  and  $E_{\text{exch-ind}}^{(30)}$  are captured at the supermolecular HF level and are, therefore, implicitly included in the  $\delta E_{\text{ind,resp}}^{(\text{HF})}$  term as defined above.

The most approximate of these methods, SAPT0, will scale as  $\mathcal{O}(o^2v^2n_{ri})$  in the DF integral evaluation and  $\mathcal{O}(o^3v^2)$  in the energy evaluation. Here,  $o$  refers to the number of occupied orbitals in one monomer (for convenience we will assume that both monomers have the same number of occupied orbitals);  $v$  refers to the number of virtual orbitals for one monomer. The number of auxiliary basis functions in the DF approximation is denoted  $n_{ri}$ . SAPT2 will scale as  $\mathcal{O}(o^2v^4)$  in the energy evaluation and  $\mathcal{O}(v^4n_{ri})$  in the DF integral evaluation. The more accurate SAPT2+ will scale as  $\mathcal{O}(o^3v^4)$  in the energy evaluation due to the triples correction to  $E_{\text{disp}}^{(22)}$ . SAPT2+(3) adds an additional term that scales  $\mathcal{O}(o^2v^4)$  to SAPT2+.

The two-electron integrals in the DF approximation are given by

$$(\mu\nu|\rho\sigma) \approx \sum_{PQ} (\mu\nu|P)[J^{-1}]_{PQ}(Q|\rho\sigma). \quad (11)$$

The two-index quantity,  $[J^{-1}]_{PQ}$ , is the inverse of the Coulomb metric evaluated in an auxiliary basis set

$$[J]_{PQ} = \int P(\mathbf{r}_1) \frac{1}{r_{12}} Q(\mathbf{r}_2) d^3\mathbf{r}_1 d^3\mathbf{r}_2. \quad (12)$$

Ignoring any sparsity due to large distances between centers, there are  $\mathcal{O}(n_{ri}n_{ao}^2)$  three-index integrals in the DF approach, compared to  $\mathcal{O}(n_{ao}^4)$  two-electron integrals. It is convenient to rewrite Eq. (11) with different three-index quantities,

TABLE I. Notation and common intermediates for SAPT equations. (Indices using Greek letters refer to any orbital on either monomer.  $\epsilon$  refers to HF eigenvalues, the index determines to which monomer the eigenvalues belong. ( $\nu_A$ ) and ( $\nu_B$ ) refer to the nuclear attraction integrals corresponding to the nuclei of monomers  $A$  and  $B$ , respectively.)

| Terms                       | Definition  |
|-----------------------------|---|
| $\nu_{\mu\sigma}^{\mu\rho}$ | $(\mu\nu \rho\sigma)$   |
| $(\omega_B)_{i_2}^{i_1}$    | $(\nu_B)_{i_2}^{i_1} + 2\nu_{b_1i_2}^{b_1i_1}$  |
| $(\omega_A)_{j_2}^{j_1}$    | $(\nu_A)_{j_2}^{j_1} + 2\nu_{a_1j_2}^{a_1j_1}$  |
| $t_{r_1r_2}^{a_1a_2}$       | $\nu_{r_1r_2}^{a_1a_2}/(\epsilon_{a_1} + \epsilon_{a_2} - \epsilon_{r_1} - \epsilon_{r_2})$ |
| $t_{s_1s_2}^{b_1b_2}$       | $\nu_{s_1s_2}^{b_1b_2}/(\epsilon_{b_1} + \epsilon_{b_2} - \epsilon_{s_1} - \epsilon_{s_2})$ |
| $t_{r_1s_1}^{a_1b_1}$       | $\nu_{r_1s_1}^{a_1b_1}/(\epsilon_{a_1} + \epsilon_{b_1} - \epsilon_{r_1} - \epsilon_{s_1})$ |
| $\theta_{r_1r_2}^{a_1a_2}$  | $2t_{r_1r_2}^{a_1a_2} - t_{r_1r_2}^{a_2a_1}$  |
| $\theta_{s_1s_2}^{b_1b_2}$  | $2t_{s_1s_2}^{b_1b_2} - t_{s_1s_2}^{b_2b_1}$  |
| $\tilde{X}_{a_2}^{a_1}$     | $\theta_{r_1r_2}^{a_1a_3}t_{r_1r_2}^{a_2a_3}$   |
| $\tilde{X}_{r_2}^{r_1}$     | $\theta_{r_1r_3}^{a_1a_2}t_{r_2r_3}^{a_1a_2}$   |
| $B_{\mu\nu}^P$              | $\Sigma_Q(\mu\nu Q)[J^{-1/2}]_{PQ}$   |
| $\Theta_{a_1r_1}^P$         | $\theta_{r_1r_2}^{a_1a_2}B_{r_2r_2}^P$  |
| $\Theta_{b_1s_1}^P$         | $\theta_{s_1s_2}^{b_1b_2}B_{s_2s_2}^P$  |
| $T_{a_1r_1}^P$              | $t_{r_1s_1}^{a_1b_1}B_{b_1s_1}^P$   |
| $T_{b_1s_1}^P$              | $t_{r_1s_1}^{a_1b_1}B_{a_1r_1}^P$   |

$$B_{\mu\nu}^Q = \sum_P (\mu\nu|P)[J^{-1/2}]_{PQ}, \quad (13)$$

$$(\mu\nu|\rho\sigma) \approx \sum_Q B_{\mu\nu}^Q B_{\rho\sigma}^Q. \quad (14)$$

In SAPT, the three-index terms must be transformed into the molecular orbital (MO) basis of each monomer

$$B_{ij}^Q = \sum_{\mu\nu} C_{\mu i}^M B_{\mu\nu}^Q C_{\nu j}^N, \quad (15)$$

where  $C^M$  and  $C^N$  represent the SCF coefficient matrices of monomers  $M$  and  $N$ , and where  $i$  and  $j$  are MOs resulting from the HF computations on monomers  $M$  and  $N$ , respectively. All the two-electron integrals necessary in SAPT can be formed from the  $B_{ij}^Q$  quantities through the MO basis analog of Eq. (14).

The working equations for SAPT can be found in Refs. 50–54; only equations where the introduction of DF is relevant will be presented here. Also, many of the places where DF can be introduced are similar for different terms; to avoid redundancy, we will present enough unique cases so that the others will become obvious. To remain consistent with the notation used throughout the literature describing SAPT, the indices  $a(b)$ ,  $r(s)$ , and  $i(j)$  will be used to describe the occupied, virtual, and any orbital of monomer  $A$  ( $B$ ), respectively. Table I contains the definition of many terms that we will refer to throughout.

The applicability of conventional SAPT computations has been limited, in part, by the large amount of disk I/O required.<sup>55</sup> Our philosophy for developing a DF-SAPT program was to reduce the amount of disk I/O required through the introduction of reusable DF intermediates. In principle, the intramonomer correlation corrections to SAPT can be evaluated in a dimer or monomer centered basis. While some

of the terms we have implemented are valid in either, as a whole, our SAPT implementation is only valid in the dimer basis.

The most demanding step in the evaluation of intra-monomer electron correlation effects (excluding  $E_{\text{disp}}^{(22)}$ ) is the formation of the second-order double excitation amplitudes,

$$\begin{aligned} t_{r_1r_2}^{(2)a_1a_2} = & (\nu_{r_1r_2}^{a_1a_2} + \nu_{a_3a_4}^{a_1a_2} + \nu_{r_2a_3}^{a_1a_2} + \nu_{r_1a_3}^{a_1a_2} \\ & - \nu_{a_3r_2}^{a_1a_2} - \nu_{a_3r_1}^{a_1a_2} - \nu_{a_3r_2}^{a_1a_2} \\ & - \nu_{a_3r_1}^{a_1a_2})/(\epsilon_{a_1} + \epsilon_{a_2} - \epsilon_{r_1} - \epsilon_{r_2}). \end{aligned} \quad (16)$$

The third and fourth terms can be rewritten with the DF integrals and the  $\Theta$  intermediates (see Table I) as

$$\Theta_{a_1r_1}^P B_{a_2r_2}^P + B_{a_1r_1}^P \Theta_{a_2r_2}^P, \quad (17)$$

which reduces the scaling of these terms from  $\mathcal{O}(o^3v^3)$  to  $\mathcal{O}(o^2v^2n_{ri})$ . Including the contribution from the  $v^4$  integrals in the first term scales as  $\mathcal{O}(o^2v^4)$ . A conventional SAPT program must have access to this large group of integrals in order to evaluate this term. There are two types of  $v^4$  integrals needed for SAPT ( $r^4$  and  $s^4$ ); additionally, to evaluate the  $E_{\text{disp}}^{(30)}$  term the  $r^2s^2$  integrals are required. While DF cannot improve the scaling of steps with  $v^4$  dependence, DF can still improve the efficiency of the SAPT algorithm. For example, each of the  $r^4$  integrals are needed only once during the SAPT computation (this is true of all  $v^4$  integrals); when they are approximated with DF, it is practical to form these integrals “on the fly” (from the  $B^P$  three-index DF integrals) and evaluate their contribution to  $t_{r_1r_2}^{(2)a_1a_2}$  without storing them. The temporary batches of integrals should be as large as the system’s memory will allow. With exact integrals, a similar algorithm is needed, except the batches of  $v^4$  integrals will be read from disk or formed using some type of direct or semidirect algorithm. Other, smaller, classes of DF integrals are needed multiple times during an SAPT computation. To avoid redundant work, these are formed from the three-index DF integrals and stored on disk.

#### A. $E_{\text{elst,resp}}^{(12)}$

The  $E_{\text{elst,resp}}^{(120)}$  correction is given by

$$\begin{aligned} E_{\text{elst,resp}}^{(120)} = & 4 \text{Re } Y_{r_1}^{(2)a_1} C_{a_1}^{r_1} + 2\theta_{r_1r_2}^{a_1a_2}(\omega_B)_{r_3}^{r_2}t_{r_3}^{r_1r_2} \\ & - 2\theta_{r_1r_2}^{a_1a_2}(\omega_B)_{a_2}^{a_3}t_{a_2}^{r_1r_2}, \end{aligned} \quad (18)$$

where  $C_{a_1}^{r_1}$  are the coupled-perturbed HF coefficients for monomer  $A$  solved for in the presence of the electrostatic potential of monomer  $B$ , and

$$\begin{aligned} Y_{r_1}^{(2)a_1} = & \theta_{r_4r_2}^{a_4a_2}t_{a_4a_2}^{r_4r_3}(2\nu_{r_1r_3}^{a_1r_2} - \nu_{r_1r_3}^{a_2a_1}) - \theta_{r_4r_2}^{a_4a_2}t_{a_4a_2}^{r_4r_2} \\ & \times (2\nu_{r_1a_2}^{a_1a_3} - \nu_{r_1a_2}^{a_3a_1}) + \theta_{r_3r_2}^{a_3a_1}t_{a_3r_1}^{r_3r_2} - \theta_{r_3r_1}^{a_3a_2}t_{a_3a_2}^{r_3r_2}. \end{aligned} \quad (19)$$

The second and third terms of  $E_{\text{elst,resp}}^{(120)}$  do not contain two electron integrals explicitly, so they must be evaluated as usual. By inserting the three-center DF integrals and  $\tilde{X}_{r_3}^{r_2}$ , the first  $Y_{r_1}^{(2)a_1}$  term can be written as



$$\tilde{X}_{r_3}^{r_2}(2B_{a_1r_1}^P B_{r_2r_3}^P - B_{r_1r_2}^P B_{a_1r_3}^P). \quad (20)$$

Through the factorization

$$C^P = \tilde{X}_{r_3}^{r_2} B_{r_2r_3}^P \quad (21)$$

and

$$D_{a_1r_2}^P = \tilde{X}_{r_3}^{r_2} B_{a_1r_3}^P, \quad (22)$$

this contribution can be evaluated as

$$(2B_{a_1r_1}^P C^P - B_{r_1r_2}^P D_{a_1r_2}^P). \quad (23)$$

Likewise, by introducing the three-center DF integrals and  $\tilde{X}_{a_3}^{a_2}$  into the second  $Y_{r_1}^{(2)a_1}$  term, we get

$$\tilde{X}_{a_3}^{a_2}(2B_{a_1r_1}^P B_{a_2a_3}^P - B_{a_1a_2}^P B_{a_3r_1}^P). \quad (24)$$

Using similar factorizations, the following intermediates can be defined:

$$E^P = \tilde{X}_{a_3}^{a_2} B_{a_2a_3}^P, \quad (25)$$

and

$$F_{a_2r_1}^P = \tilde{X}_{a_3}^{a_2} B_{a_3r_1}^P. \quad (26)$$

Now, we can write the second  $Y_{r_1}^{(2)a_1}$  term as

$$(2B_{a_1r_1}^P E^P - B_{a_1a_2}^P F_{a_2r_1}^P). \quad (27)$$

The contributions from the last two terms of  $Y_{r_1}^{(2)a_1}$  can be computed with the  $\Theta$  intermediates (see Table I) as

$$\Theta_{a_1r_2}^P B_{r_1r_2}^P - \Theta_{a_2r_1}^P B_{a_1a_2}^P. \quad (28)$$

We have presented equations which introduce DF intermediates into the  $E_{\text{elst,resp}}^{(120)}$  term wherever possible. In practice, it is not always advantageous to introduce these intermediates, so we will examine the scaling of this term with and without the DF intermediates. First, note that the second and third  $E_{\text{elst,resp}}^{(120)}$  terms can be efficiently evaluated by forming the  $\tilde{X}_{r_3}^{r_2}$  and  $\tilde{X}_{a_3}^{a_2}$  intermediates, respectively. The formation of these quantities scales  $\mathcal{O}(o^2v^3)$  and  $\mathcal{O}(o^3v^2)$ , respectively. Rather conveniently, we can reuse these quantities in the formation of  $Y_{r_1}^{(2)a_1}$ ; the formation of the  $\tilde{X}$  type intermediates will not be considered in the  $Y_{r_1}^{(2)a_1}$  scaling, since they will have been formed already. With a conventional algorithm, the four  $Y_{r_1}^{(2)a_1}$  terms will scale  $\mathcal{O}(ov^3)$ ,  $\mathcal{O}(o^3v)$ ,  $\mathcal{O}(o^2v^3)$ , and  $\mathcal{O}(o^3v^2)$ , respectively. With the DF equations presented above, the four  $Y_{r_1}^{(2)a_1}$  terms will scale  $\mathcal{O}(ov^2n_{ri})$ ,  $\mathcal{O}(o^2vn_{ri})$ ,  $\mathcal{O}(ov^2n_{ri})$ , and  $\mathcal{O}(o^2vn_{ri})$ , respectively. The scaling of the third and fourth terms assumes that the  $\Theta$  type intermediates are also available; this assumption will be discussed further in the context of the  $E_{\text{exch}}^{(111)}$  correction.

We can see from this analysis that the introduction of the DF intermediates makes the evaluation of the third and fourth terms more efficient, since, for most practical situations  $ov > n_{ri}$ . The second term has a strong dependence on occupied orbitals, so it is not advantageous to introduce DF intermediates because  $o \ll n_{ri}$ . Therefore, our DF-SAPT

implementation evaluates the second  $Y_{r_1}^{(2)a_1}$  term conventionally. The first term is also slightly worse when DF intermediates are introduced ( $v < n_{ri}$ ) in terms of the number of floating point operations. However, by using the DF intermediates, the amount of disk I/O is reduced and the  $ar^3$  integrals do not need to be computed explicitly in order to evaluate the  $E_{\text{elst,resp}}^{(120)}$  correction. So, our DF-SAPT implementation evaluates the first  $Y_{r_1}^{(2)a_1}$  term with DF intermediates as outlined above.

## B. $E_{\text{exch}}^{(11)}$

We take the  $E_{\text{exch}}^{(11)}$  equations from Ref. 52. There is an alternative derivation which leads to equations that can be implemented efficiently using exact two-electron integrals; however, it is not amenable to the introduction of DF intermediates.<sup>53</sup>

The  $E_{\text{exch}}^{(110)}$  correction is given by

$$E_{\text{exch}}^{(110)} = -2\theta_{a_1a_2}^{r_1r_2} \nu_{r_2b_1}^{a_2s_1} S_{s_1}^{a_1} S_{r_1}^{b_1} - 2\theta_{r_1r_2}^{a_1a_2} \nu_{a_1b_1}^{r_1s_1} S_{s_1}^{r_2} S_{a_2}^{b_1} - 4 \text{Re } \theta_{r_1r_2}^{a_1a_2} (\omega_B)_{a_1}^{r_1} S_{b_1}^{r_2} S_{a_2}^{b_1}, \quad (29)$$

where  $S$  are the overlap integrals. DF intermediates cannot be used in the evaluation of the third  $E_{\text{exch}}^{(110)}$  term, so it is evaluated conventionally. We will rewrite the first two contributions to  $E_{\text{exch}}^{(110)}$  with the  $\Theta$  type intermediates as

$$-2\Theta_{a_1r_1}^P B_{b_1s_1}^P S_{s_1}^{a_1} S_{r_1}^{b_1} - 2\Theta_{a_1r_1}^P B_{b_1s_1}^P S_{s_1}^{r_1} S_{a_1}^{b_1}. \quad (30)$$

It is possible to improve this term further by defining the following quantities:

$$C_{a_1b_1}^P = \Theta_{a_1r_1}^P S_{r_1}^{b_1}, \quad (31)$$

$$D_{a_1b_1}^P = B_{b_1s_1}^P S_{s_1}^{a_1}, \quad (32)$$

and

$$E_{a_1r_1}^P = B_{b_1s_1}^P S_{s_1}^{r_1} S_{a_1}^{b_1}. \quad (33)$$

Now we can write the first two  $E_{\text{exch}}^{(110)}$  terms as

$$-2C_{a_1b_1}^P D_{a_1b_1}^P - 2\Theta_{a_1r_1}^P E_{a_1r_1}^P. \quad (34)$$

The scaling of the DF evaluation of the  $E_{\text{exch}}^{(110)}$  correction is dependent on the formation of the  $E_{a_1r_1}^P$  quantities (again assuming that the  $\Theta$  intermediates are available). This leads to a scaling of  $\mathcal{O}(ov^2n_{ri})$ . If the equation for  $E_{\text{exch}}^{(11)}$  presented above was evaluated conventionally, it would scale  $\mathcal{O}(o^2v^3)$ . The alternative equation from Ref. 52 scales  $\mathcal{O}(o^3v^2)$ . For the form of  $E_{\text{exch}}^{(110)}$  presented above to be practical, clearly, the introduction of DF intermediates is necessary. Our DF implementation of  $E_{\text{exch}}^{(110)}$  is roughly equivalent in cost to a conventional implementation of the alternative equation.

The form of the  $E_{\text{exch}}^{(12)} K_2^u$  term is identical in structure to the  $E_{\text{exch}}^{(110)}$  term with the first-order double excitation amplitudes replaced by the second-order amplitudes. The same arguments about the scaling of the  $E_{\text{exch}}^{(110)}$  term apply to the  $E_{\text{exch}}^{(12)} K_2^u$  term. The relevant intermediates are useful in the evaluation of the  $E_{\text{elst,resp}}^{(13)}$  term.

### C. $E_{\text{exch}}^{(111)}$

The bilinear exchange term

$$E_{\text{exch}}^{(111)} = -4 \text{Re} \theta_{a_1 a_2}^{r_1 r_2} \theta_{b_1 b_2}^{s_1 s_2} \nu_{r_1 s_1}^{a_1 b_1} S_{s_2}^{a_2 b_2} - 4 \text{Re} \theta_{r_1 r_2}^{a_1 a_2} \theta_{b_1 b_2}^{s_1 s_2} \nu_{a_1 s_1}^{r_1 b_1} S_{a_2}^{r_2 b_2}, \quad (35)$$

can be evaluated with the  $\Theta$  intermediates as

$$E_{\text{exch}}^{(111)} = -4 \text{Re} \Theta_{a_1 r_1}^P \Theta_{b_1 s_1}^P (S_{s_1}^{a_1 b_1} + S_{a_1}^{b_1 r_1}). \quad (36)$$

In order to efficiently evaluate the energy contribution, the  $S$  integrals should be contracted with the  $\Theta$  intermediates in an analogous manner to the contractions used in the  $E_{\text{exch}}^{(110)}$  evaluation [Eqs. (31)–(33)]. Conventionally, the evaluation of the bilinear exchange term scales as  $\mathcal{O}(o^3 v^3)$ . Under the DF approximation, the formation of the  $\Theta$  intermediates, which scales  $\mathcal{O}(o^2 v^2 n_{ri})$ , becomes the rate determining step.

### D. $E_{\text{exch}}^{(12)} K_2^f$

Again, we take the  $E_{\text{exch}}^{(12)} K_2^f$  equations from Ref. 52. Without the introduction of DF intermediates, this form of the equations is much more computationally demanding than the equations presented in Ref. 53. However, by contracting the overlap integrals with the three-center DF integrals, the complexity of the  $E_{\text{exch}}^{(12)} K_2^f$  equations can be simplified.  $E_{\text{exch}}^{(120)} K_2^f$  is given by

$$E_{\text{exch}}^{(120)} K_2^f = -2 t_{a_1}^{(2)r_1} X_{r_1}^{a_1} - 2 t_{r_1}^{(2)a_1} Y_{a_1}^{r_1}, \quad (37)$$

where

$$t_{r_1}^{(2)a_1} = (\theta_{r_3 r_2}^{a_3 a_1} \nu_{a_3 r_1}^{r_3 r_2} - \theta_{r_3 r_1}^{a_3 a_2} \nu_{a_3 a_2}^{r_3 a_1}) / (\epsilon_{a_1} - \epsilon_{r_1}), \quad (38)$$

$$X_{r_1}^{a_1} = \nu_{r_1 b_1}^{r_2 s_1} S_{r_2}^{b_1 s_1} - \nu_{a_2 b_1}^{r_1 s_1} S_{r_1}^{b_1 s_1} + 2 \nu_{r_1 b_1}^{a_1 s_1} S_{s_1}^{a_2 b_1} + (\omega_B)_{r_1 r_2}^{r_2 s_1} S_{b_1}^{a_1} - (\omega_B)_{a_2 r_1}^{a_1 s_1} S_{b_1}^{a_2} + (\omega_A)_{b_1}^{s_1} S_{r_1}^{a_1} S_{a_1}^{b_1}, \quad (39)$$

and

$$Y_{a_1}^{r_1} = \nu_{a_1 b_1}^{r_2 s_1} S_{r_2}^{b_1 s_1} - \nu_{a_2 b_1}^{r_1 s_1} S_{r_1}^{b_1 s_1} + 2 \nu_{a_1 b_1}^{r_1 s_1} S_{s_1}^{a_2 b_1} + (\omega_B)_{a_1 r_2}^{r_2 s_1} S_{b_1}^{a_1} - (\omega_B)_{a_2 r_1}^{a_1 s_1} S_{b_1}^{a_2} + (\omega_A)_{b_1}^{s_1} S_{r_1}^{a_1} S_{a_1}^{b_1}. \quad (40)$$

The  $t_{r_1}^{(2)a_1}$  amplitudes can be computed and stored during the evaluation of the  $E_{\text{elst}}^{(12)}$  term. Only the first three terms of  $X_{r_1}^{a_1}$  and  $Y_{a_1}^{r_1}$  can benefit from the introduction of DF intermediates. The general strategy for forming the  $X_{r_1}^{a_1}$  and  $Y_{a_1}^{r_1}$  quantities is to insert the three-index DF integrals and contract them with the  $S$  integrals in order to remove virtual indices. Contractions that replace occupied indices with occupied indices or virtual with virtual will not be performed. We will show the factorization of  $X_{r_1}^{a_1}$  and the factorization of  $Y_{a_1}^{r_1}$  follows from this. For the first  $X_{r_1}^{a_1}$  term, we will define two intermediates,

$$C_{b_1 r_1}^P = B_{r_1 r_2}^P S_{r_2}^{b_1} \quad (41)$$

and

$$D_{a_1 b_1}^P = B_{s_1 b_1}^P S_{s_1}^{a_1}. \quad (42)$$

Now, this term can be evaluated as

$$D_{a_1 b_1}^P C_{b_1 r_1}^P. \quad (43)$$

The second  $X_{r_1}^{a_1}$  term will also use the  $D_{a_1 b_1}^P$  intermediate. This term is evaluated first by

$$E_{a_1 b_1} = B_{a_1 a_2}^P D_{a_2 b_1}^P, \quad (44)$$

and then

$$E_{a_1 b_1} S_{r_1}^{b_1}. \quad (45)$$

The third term can be evaluated through the formation of

$$F^P = B_{b_1 s_1}^P S_{s_1}^{a_2} S_{a_2}^{b_1}, \quad (46)$$

as

$$B_{a_1 r_1}^P F^P. \quad (47)$$

The conventional evaluation of the  $E_{\text{exch}}^{(120)} K_2^f$  correction presented above scales as  $\mathcal{O}(ov^3)$ . The DF evaluation scales as  $\mathcal{O}(ov^2 n_{ri})$ , which is slightly more expensive, but it avoids the formation of  $r^2 bs$  integrals. The conventional evaluation of the alternative formula for  $E_{\text{exch}}^{(120)} K_2^f$  scales as  $\mathcal{O}(o^3 v)$ ,<sup>53</sup> which is better than the DF formulation outlined above. However, the procedure shown above is much simpler than the alternative. Since the cost of computing this term is negligible (regardless of formulation) compared to other second-order terms, our DF-SAPT program uses the DF implementation of this term out of convenience.

### E. ${}^t E_{\text{ind}}^{(22)}$

The equation for the second-order correction to induction is taken from Ref. 54. This correction has two parts,  ${}^t E_{\text{ind}}^{(220)}(A \leftarrow B)$  and  ${}^t E_{\text{ind}}^{(220)}(A \rightarrow B)$ . Here,  $(A \leftarrow B)$  refers to the interaction of monomer  $A$  with the electric field of monomer  $B$  and  $(A \rightarrow B)$  refers to the interaction of monomer  $B$  with the electric field of monomer  $A$ . Both of these corrections can be evaluated more efficiently through the introduction of DF intermediates. The  ${}^t E_{\text{ind}}^{(220)}(A \leftarrow B)$  correction is given by

$$\begin{aligned} {}^t E_{\text{ind}}^{(220)}(A \leftarrow B) = & (2X_{a_1 a_2}^{r_1 r_2} - X_{a_1 a_2}^{r_2 r_1}) X_{r_1 r_2}^{a_1 a_2} / (\epsilon_{a_1} + \epsilon_{a_2} - \epsilon_{r_1} - \epsilon_{r_2}) \\ & + 4 t_{a_1}^{(2)r_1} z_{r_1}^{a_1} - 2 \theta_{r_1 r_2}^{a_1 a_2} I_{r_3}^{a_3} ((\omega_B)_{a_3}^{r_2} I_{a_2 a_1}^{r_3 r_1} \\ & + (\omega_B)_{a_2}^{r_3} I_{a_3 a_1}^{r_2 r_1}) - 2 \theta_{a_1 a_2}^{r_1 r_2} I_{r_3}^{a_3} (I_{r_3}^{a_2} \nu_{r_2 r_1}^{a_3 a_1} + I_{r_2}^{a_3} \nu_{r_3 r_1}^{a_2 a_1}) \\ & + 2 I_{a_1}^{r_1} I_{a_2}^{r_2} \theta_{r_1 r_2}^{(2)a_1 a_2} (\epsilon_{a_1} + \epsilon_{a_2} - \epsilon_{r_1} - \epsilon_{r_2}) \\ & - 4 I_{a_1}^{r_1} I_{a_2}^{r_2} (2 \nu_{r_1 a_3}^{r_1 r_2} - \nu_{r_1 a_3}^{r_2 r_1}) \theta_{r_3 r_2}^{a_3 a_2}, \end{aligned} \quad (48)$$

where

$$I_{r_1}^{a_1} = (\omega_B)_{r_1}^{a_1} / (\epsilon_{a_1} - \epsilon_{r_1}) \quad (49)$$

and

$$z_{r_1}^{a_1} = (\omega_B)_{r_1}^{r_2} I_{r_2}^{a_1} - (\omega_B)_{a_2}^{a_1} I_{r_1}^{a_2}. \quad (50)$$

The  $X_{r_1 r_2}^{a_1 a_2}$  quantity is given as

$$X_{r_1 r_2}^{a_1 a_2} = (\omega_B)^{r_3} t_{r_1 r_3 r_2}^{a_1 a_2} + (\omega_B)^{r_3} t_{r_2 r_1 r_3}^{a_1 a_2} - (\omega_B)^{a_1} t_{r_3 r_1 r_2}^{a_2 a_3} - (\omega_B)^{a_2} t_{r_3 r_1 r_2}^{a_1 a_3},$$

$$\nu_{r_1 r_2}^{a_1 r_3} I_{r_3}^{a_2} + \nu_{r_1 r_2}^{r_3 a_2} I_{r_3}^{a_1} - \nu_{r_1 a_3}^{a_1 a_2} I_{r_2}^{a_3} - \nu_{a_3 r_2}^{a_1 a_2} I_{r_1}^{a_3}. \quad (51)$$

DF intermediates can only be used in the first, fourth, and sixth terms of  ${}^tE_{\text{ind}}^{(220)}(A \leftarrow B)$ . First, we will examine the formation of  $X_{r_1 r_2}^{a_1 a_2}$ . The first four terms of  $X_{r_1 r_2}^{a_1 a_2}$  cannot use DF intermediates. Including the contributions from the last two terms conventionally scales as  $\mathcal{O}(o^3 v^2)$ . Using DF intermediates would lead to a scaling of  $\mathcal{O}(o^2 v^2 n_{ri})$ , therefore we will not use DF intermediates to evaluate this term. The fifth and sixth terms depend on the  $ov^3$  integrals and can be evaluated by defining an intermediate as

$$C_{a_1 r_1}^P = B_{r_3 r_1}^P I_{r_3}^{a_1}. \quad (52)$$

This contribution to  $X_{r_1 r_2}^{a_1 a_2}$  can be evaluated by forming another quantity

$$Y_{r_1 r_2}^{a_1 a_2} = C_{a_1 r_1}^P B_{a_2 r_2}^P, \quad (53)$$

and adding

$$X_{r_1 r_2}^{a_1 a_2} = Y_{r_1 r_2}^{a_1 a_2} + Y_{r_2 r_1}^{a_2 a_1}. \quad (54)$$

To evaluate the fourth  ${}^tE_{\text{ind}}^{(220)}(A \leftarrow B)$  term, the  $\Theta$  intermediates can be reused along with two new quantities

$$D_{a_2 r_2}^P = I_{a_3 r_3}^{a_2} B_{a_3 r_2}^P \quad (55)$$

and

$$E_{a_2 r_2}^P = I_{a_3 r_2}^{a_3} B_{a_2 r_3}^P. \quad (56)$$

With these quantities, this contribution to  ${}^tE_{\text{ind}}^{(220)}(A \leftarrow B)$  can be evaluated as

$$-2\Theta_{a_2 r_2}^P (D_{a_2 r_2}^P + E_{a_2 r_2}^P). \quad (57)$$

The sixth term contributing to  ${}^tE_{\text{ind}}^{(220)}(A \leftarrow B)$  can be evaluated conventionally as  $\mathcal{O}(o^2 v^2)$ . The introduction of DF intermediates leads to a scaling of  $\mathcal{O}(ov^2 n_{ri})$ , so DF intermediates are not used here.

A conventional evaluation of  ${}^tE_{\text{ind}}^{(220)}(A \leftarrow B)$  will scale  $\mathcal{O}(o^2 v^3)$ . Specifically, the first, third and fourth terms exhibit this scaling, while the fifth and sixth terms will scale  $\mathcal{O}(o^2 v^2)$ . The purpose of introducing DF intermediates into the  ${}^tE_{\text{ind}}^{(220)}(A \leftarrow B)$  evaluation was to remove the need to deal with a type of  $ov^3$  integrals explicitly. The two terms where DF intermediates were introduced scale  $\mathcal{O}(o^2 v^2 n_{ri})$ .

The  ${}^tE_{\text{ind}}^{(220)}(A \rightarrow B)$  contributions to  ${}^tE_{\text{ind}}^{(220)}$  can be written as

$${}^tE_{\text{ind}}^{(220)}(A \rightarrow B) = -8\theta_{r_1 r_2}^{t_3 a_2} t_{r_1 r_2}^{r_3 a_1} \nu_{a_2 b_1}^{a_1 s_1} I_{s_1}^{b_1} + 8\theta_{r_3 r_2}^{a_1 a_2} t_{r_3 r_1}^{r_3 a_1} \nu_{r_2 b_1}^{r_1 s_1} I_{s_1}^{b_1} + 16t_{r_1}^{(2) a_1} \nu_{r_1 b_1}^{a_1 s_1} I_{s_1}^{b_1}. \quad (58)$$

This can be evaluated by substituting some of the intermediates defined earlier

$${}^tE_{\text{ind}}^{(220)}(A \rightarrow B) = -8(\tilde{X}_{a_2}^{r_1} B_{a_1 a_2}^P)(B_{b_1 s_1}^P I_{s_1}^{b_1}) + 8(\tilde{X}_{r_2}^{r_1} B_{r_1 r_2}^P) \times (B_{b_1 s_1}^P I_{s_1}^{b_1}) + 16(t_{r_1}^{(2) a_1} B_{a_1 r_1}^P)(B_{b_1 s_1}^P I_{s_1}^{b_1}). \quad (59)$$

The quantities inside of the parenthesis should be fully contracted, then this term can be evaluated as a series of dot products. Conventionally, this term scales  $\mathcal{O}(ov^3)$ ; the DF implementation scales  $\mathcal{O}(v^2 n_{ri})$  and avoids the  $r^2 bs$  integrals. By introducing DF intermediates into the  ${}^tE_{\text{ind}}^{(220)}$  evaluation, the computation of two types of  $ov^3$  integrals has been avoided.

## F. $E_{\text{disp}}^{(22)}(S)$

The singles contribution to the second-order intramonomer correlation correction to dispersion is given by<sup>50</sup>

$$E_{\text{disp}}^{(220)}(S) = 4 \text{Re} \frac{X_{r_1}^{a_1} Y_{a_1}^{r_1}}{\epsilon_{a_1} - \epsilon_{r_1}}, \quad (60)$$

where

$$X_{r_1}^{a_1} = 2(2\nu_{a_2 r_1}^{r_3 r_2} - \nu_{r_1 a_2}^{r_3 r_2}) t_{r_2 r_3}^{a_1 a_2} - 2(2\nu_{a_2 a_3}^{a_1 r_2} - \nu_{a_3 a_2}^{a_1 r_2}) t_{r_1 r_2}^{a_2 a_3}, \quad (61)$$

and

$$Y_{a_1}^{r_1} = \nu_{r_2 s_1}^{r_1 b_1} t_{a_1 b_1}^{r_2 s_1} - \nu_{a_1 s_1}^{a_2 b_1} t_{a_2 b_1}^{r_1 s_1}. \quad (62)$$

$X_{r_1}^{a_1}$  is, to within a constant, the third and fourth terms of Eq. (19).<sup>56</sup> Therefore, it can be reused from the evaluation of the  $E_{\text{elst,resp}}^{(120)}$  term or, if necessary, formed using the equations presented earlier.  $Y_{r_1}^{a_1}$  can be rewritten with the  $T$  intermediates defined in Table I,

$$Y_{a_1}^{r_1} = B_{r_2 r_1}^P T_{a_1 r_2}^P - B_{a_2 a_1}^P T_{a_2 r_1}^P. \quad (63)$$

This factorization of  $E_{\text{disp}}^{(22)}(S)$  does not require the computation and storage of the  $ov^3$  integrals.

Conventionally, both the  $X_{r_1}^{a_1}$  and  $Y_{r_1}^{a_1}$  quantities have one contribution that scales  $\mathcal{O}(o^2 v^3)$  and another that scales  $\mathcal{O}(o^3 v^2)$ . This leads to an overall scaling of  $\mathcal{O}(o^2 v^3)$  for the conventional evaluation of  $E_{\text{disp}}^{(220)}(S)$ . The formation of  $X_{r_1}^{a_1}$  with DF intermediates scales  $\mathcal{O}(o^2 v^2 n_{ri})$  and avoids the  $ar^3$  integrals. The two  $Y_{r_1}^{a_1}$  terms scale  $\mathcal{O}(ov^2 n_{ri})$  and  $\mathcal{O}(o^2 v n_{ri})$ , respectively, assuming that the  $T$  intermediates are available. The  $T$  intermediates are stored during the evaluation of the  $E_{\text{disp}}^{(220)}(Q)$  correction.

The equations presented above are sufficient to show how DF can be introduced into SAPT to allow for a more efficient evaluation of the intramonomer correlation corrections. The remaining corrections are similar in structure to those presented above. Any equation not presented here can be found in Ref. 50; the DF intermediates described in the present work can easily be applied to the rest of the terms. Additionally, we have presented equations with respect to monomer A, the equations for monomer B can be obtained by interchanging the indices and quantities pertaining to monomer A with those for monomer B, and vice versa. Unfortunately, the limiting step in the computation of the

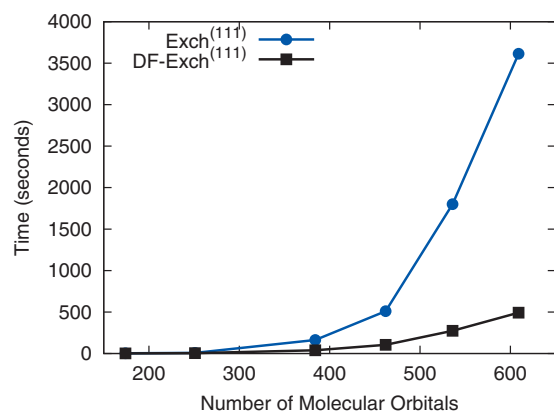


FIG. 1. Timings of the conventional and DF evaluation of  $E_{\text{exch}}^{(111)}$  with an aug-cc-pVDZ orbital basis and an aug-cc-pVDZ-RI fitting basis.

second-order correction to dispersion, the triples contribution to  $E_{\text{disp}}^{(22)}$ , cannot be improved by the introduction of DF. This term unavoidably scales as  $\mathcal{O}(o^3v^4)$ .

The second-order intramonomer correlation corrections to SAPT described above have been implemented within our DF-SAPT program.<sup>48</sup> This program has been developed within the framework of PSI 3.4.<sup>57</sup> All of the terms that are first or second-order in  $W$  have been implemented. Additionally, DF versions of the  $E_{\text{disp}}^{(30)}$  and  $E_{\text{elst,resp}}^{(13)}$  corrections have been implemented.

We will benchmark the accuracy of DF-SAPT for the S22 test set of Hobza and co-workers<sup>6</sup> using recently revised interaction energies by Takatani *et al.*<sup>49</sup> We compute interaction energies for these complexes with the aug-cc-pVDZ' and aug-cc-pVDZ basis sets.<sup>58,59</sup> The corresponding aug-cc-pVXZ-RI basis sets<sup>60</sup> are employed to approximate the two-electron integrals. Here, the aug-cc-pVDZ' basis set removes all diffuse functions from hydrogen atoms and diffuse  $d$  functions from carbon atoms. In our experience, fortuitous error cancellation occurs when this basis set is paired with MP2-like methods (i.e., SAPT0 and SAPT2).<sup>33,48</sup>

### III. RESULTS AND DISCUSSION

#### A. DF-SAPT Performance

We will begin by examining the improvement of the  $E_{\text{exch}}^{(111)}$  evaluation due to the introduction of DF intermediates because it scales  $\mathcal{O}(o^3v^3)$ , which is as costly as any energy evaluation in SAPT2 (the formation of second-order double excitation amplitudes is slightly more expensive at  $\mathcal{O}(o^2v^4)$ ). Additionally, due to the simple form of this correction, it is possible to compare the timings of the DF algorithm to a nearly optimal conventional algorithm. When DF intermediates are introduced, the scaling of the  $E_{\text{exch}}^{(111)}$  term is reduced to  $\mathcal{O}(o^2v^2n_{\text{ri}})$  and depends on the formation of the  $\Theta_{a1r1}^P$  and  $\Theta_{b1s1}^P$  intermediates that exhibit  $\mathcal{O}(o^2v^2n_{\text{ri}})$  scaling. As illustrated by Fig. 1, this is a significant improvement over the  $\mathcal{O}(o^3v^3)$  conventional algorithm. These timings show that for systems with 600 MOs, there is more than a factor of 7 speedup; this factor will continue to increase for larger systems.

Our implementation of DF-SAPT has an advantage over conventional SAPT since it reduces the number of MO four-

index integrals that must be computed or stored. In addition to removing the need to store the  $v^4$  integrals, there are contributions to the SAPT energy (evaluated as described above) from 6 types of  $ov^3$  integrals. The  $E_{\text{elst,resp}}^{(12)}$  term requires  $ar^3$  and  $bs^3$  type integrals; the  $E_{\text{exch}}^{(12)}K_2^f$  term requires  $r^2bs$  and  $s^2ar$  type integrals. The  $E_{\text{ind}}^{(22)}$  term and the singles contribution to  $E_{\text{disp}}^{(22)}$  require all four of these  $ov^3$  integrals. Through the use of DF intermediates these four types of  $ov^3$  integrals do not need to be stored or even computed. Only two types of  $ov^3$  integrals, which are needed for the  $E_{\text{exch}}^{(12)}K_{11}^u$  term must be computed and stored on disk. Thus, an obvious advantage of a DF based algorithm is that only two of the six types of  $ov^3$  integrals ever need to be computed. The  $E_{\text{elst,resp}}^{(13)}$  term requires the  $ar^3$  and  $bs^3$  type integrals; while it is possible to avoid forming these integrals, it is not advantageous to do so. However, these  $ov^3$  integrals are only needed once, so they do not need to be stored; they are computed from the three-index DF integrals ( $B^P$ ). Another advantage of the DF based algorithm is that the three types of  $v^4$  integrals do not need to be stored.

In order to understand how our DF-SAPT code performs relative to a conventional SAPT code, we compare the timings of our DF-SAPT code to the SAPT2008 program.<sup>61</sup> All of the timings reported in this work were run on a quadcore Intel® Xeon E5430 processor clocked at 2.66 GHz. Both programs were allocated a maximum of 16 Gbyte of memory, which allows the conventional integral transformation in SAPT2008 to be performed “in core.” The timings assume that the results of HF computations on the monomers are already available. In order to assure a fair comparison between the two programs, the following energy terms were included in the SAPT timing:  $E_{\text{exch-ind,resp}}^{(20)}$ ,  $E_{\text{disp}}^{(20)}$ ,  $E_{\text{exch-disp}}^{(20)}$ ,  $E_{\text{elst,resp}}^{(12)}$ ,  $E_{\text{exch}}^{(12)}$ ,  $E_{\text{exch}}^{(12)}$ ,  $E_{\text{exch}}^{(12)}$ ,  $E_{\text{ind}}^{(22)}$ , and  $E_{\text{disp}}^{(21)}$ . Additionally, the integral transformation is also included in the timing; for conventional SAPT this includes only the AO to MO transformation. For DF-SAPT this includes the formation of AO three-index quantities, the AO to MO transformation of the three-index quantities, and the formation of the necessary DF four-index MO integrals from the three-index MO integrals. SAPT2008 utilizes a CCSD program to compute the second-order double excitation amplitudes, while our program forms these quantities directly. The time spent in the CCSD program is included in the SAPT2008 timing. Although the CCSD amplitudes are not iterated until convergence, this may still be including some additional work in the SAPT2008 timing that is not included in the DF-SAPT timing; however, this time must be included in order to include the rate determining step into the SAPT timing.

At this point it should be noted that the timing differences between SAPT2008 and our DF-SAPT program are primarily a result of the introduction of DF integrals into the computation. Our conventional integral transformation and the energy terms which cannot benefit from DF perform similarly to those in SAPT2008. The timings of conventional SAPT and DF-SAPT are shown in Fig. 2. At 350 orbitals, DF-SAPT is roughly a factor of 8 faster than the conventional SAPT. This speedup will grow for larger systems, as shown in Fig. 1, since the overall scaling of certain terms is reduced. It is difficult to get meaningful timings past 350



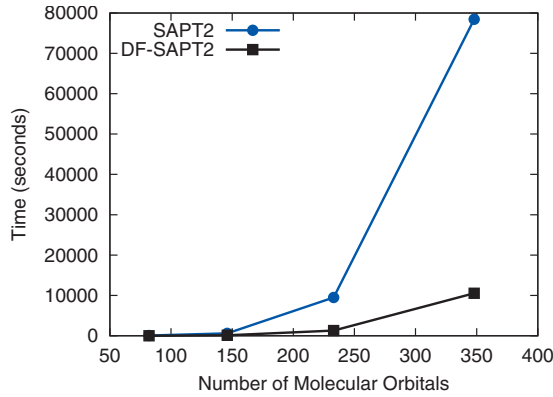


FIG. 2. Timings of the conventional and DF-SAPT2 computations of selected complexes from the S22 test set (Ref. 6) with an aug-cc-pVDZ orbital basis and an aug-cc-pVDZ-RI fitting basis.

orbitals because the conventional SAPT computations become I/O bound, therefore the timings would be strongly dependent on the available hardware.

In practice, we get additional improvements due to the threading of the DF-SAPT energy evaluation. Since most modern computers are built with multi-core processors, our DF-SAPT code exploits the availability of these additional processors. The timings of our DF-SAPT with multiple threads are shown in Fig. 3. Although to date we have only made trivial modifications to our code to allow for threading, even this minor effort has been useful for extending our code to larger systems. Since the most time consuming steps in the DF-SAPT computation are cast as matrix multiplications, we use threaded Intel<sup>®</sup> MKL BLAS routines. Other parts of our program that could not be cast as BLAS routines, such as the three-index integral evaluation, are threaded using OpenMP. The timings shown in Fig. 3 were run on dual quad-core Intel<sup>®</sup> Xeon E5430 processors clocked at 2.66 GHz. For the T-shaped indole-benzene, the SAPT/aug-cc-pVDZ (462 orbitals) computations get a factor of 1.92 speedup from one to two threads. The efficiency degrades slightly from two to four threads with only an additional factor of 1.82 speedup. When the both processors are fully utilized, we only see a factor of 1.64 moving from four to eight threads. Nevertheless, these speedups are encouraging given how straight-

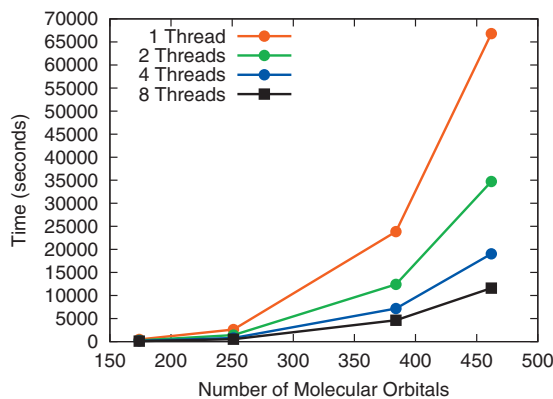


FIG. 3. Timings of the threaded DF-SAPT2 computations with an aug-cc-pVDZ orbital basis and an aug-cc-pVDZ-RI fitting basis.

forward they were to achieve. We intend to pursue larger-scale parallelization in future work.

## B. $E_{\text{disp}}^{(22)}(T)$ Performance

In order to get quantitative accuracy from SAPT, the effect of triple excitations on the  $E_{\text{disp}}^{(22)}$  term must be included. Unfortunately, evaluation of the  $E_{\text{disp}}^{(22)}(T)$  correction scales as  $\mathcal{O}(o^3v^4)$ , so applications including these contributions are limited. Additionally, the introduction of DF approximations cannot improve the formal scaling. Despite this, we have still found DF to be useful for improving the efficiency of the evaluation of the  $E_{\text{disp}}^{(220)}$  term.

The  $E_{\text{disp}}^{(220)}(T)$  is given by<sup>50</sup>

$$E_{\text{disp}}^{(220)}(T) = \frac{(4W_{r_1r_2s_1}^{a_1a_2b_1} - 2W_{r_1r_2s_1}^{a_2a_1b_1})W_{a_1a_2b_1}^{r_1r_2s_1}}{\epsilon_{a_1} + \epsilon_{a_2} + \epsilon_{b_1} - \epsilon_{r_1} - \epsilon_{r_2} - \epsilon_{s_1}}, \quad (64)$$

where

$$\begin{aligned} W_{r_1r_2s_1}^{a_1a_2b_1} = & \nu_{r_2s_1}^{r_3b_1} \nu_{r_1r_3}^{a_1a_2} + \nu_{r_1s_1}^{r_3b_1} \nu_{r_3r_2}^{a_1a_2} - \nu_{a_3s_1}^{a_2b_1} \nu_{r_1r_2}^{a_1a_3} - \nu_{a_3s_1}^{a_1b_1} \nu_{r_1r_2}^{a_3a_2} \\ & + \nu_{r_1r_2}^{r_3a_2} \nu_{r_3s_1}^{a_1b_1} + \nu_{r_1r_3}^{a_1r_3} \nu_{r_3s_1}^{a_2b_1} - \nu_{a_3r_2}^{a_1a_2} \nu_{r_1s_1}^{a_3b_1} - \nu_{r_1a_3}^{a_1a_2} \nu_{r_2s_1}^{a_3b_1}. \end{aligned} \quad (65)$$

It should be noted that there are only four unique terms that contribute to  $W_{r_1r_2s_1}^{a_1a_2b_1}$ ; the other four are symmetric with respect to the interchange of  $a_1$  and  $r_1$  with  $a_2$  and  $r_2$ . The algorithm described here computes all of the energy contributions to  $E_{\text{disp}}^{(220)}$  for each  $b_1$  and  $s_1$ . Two of the  $ov^3$  groups of integrals are required to evaluate this term; it cannot be assumed that either of these can be stored in memory. This is not problematic for the  $\nu_{r_2s_1}^{r_1b_1}$  type integrals, since if all  $r_1$  and  $r_2$  integrals are read into memory for each  $b_1$  and  $s_1$ , then the storage requirement is only  $v^2$  and this group of integrals is only read into memory once. The  $\nu_{r_2r_3}^{r_1a_1}$  type integrals are more problematic. This entire  $ov^3$  group is needed for each  $b_1$  and  $s_1$ ; naively, this would mean reading

---

**Algorithm 1.** Algorithm to evaluate the  $E_{\text{disp}}^{(220)}$  term assuming the use of exact two electron integrals.

---

**for**  $b_1 < \text{occupied B}$  **do**

**for**  $s_1 < \text{virtual B}$  **do**

$$V_{r_1r_2}^{a_1a_2} = \sum_{r_3} \nu_{r_2s_1}^{r_3b_1} \nu_{r_1r_3}^{a_1a_2}$$

$$V_{r_1r_2}^{a_1a_2} = \sum_{a_3} \nu_{a_3s_1}^{a_2b_1} \nu_{r_1r_2}^{a_1a_3}$$

$$V_{r_1r_2}^{a_1a_2} = \sum_{r_3} \nu_{r_1r_2}^{r_3a_2} \nu_{r_3s_1}^{a_1b_1}$$

$$V_{r_1r_2}^{a_1a_2} = \sum_{a_3} \nu_{a_3r_2}^{a_1a_2} \nu_{r_1s_1}^{a_3b_1}$$

$$W_{r_1r_2}^{a_1a_2} = V_{r_1r_2}^{a_1a_2} + V_{r_2r_1}^{a_2a_1}$$

$$E_{\text{disp}}^{(220)}(T) = (4W_{r_1r_2}^{a_1a_2} - 2W_{r_1r_2}^{a_2a_1})W_{a_1a_2b_1}^{r_1r_2} / (\epsilon_{a_1} + \epsilon_{a_2} + \epsilon_{b_1} - \epsilon_{r_1} - \epsilon_{r_2} - \epsilon_{s_1})$$

**end for**

**end for**

---

in an  $ov^3$  array  $ov$  times. To circumvent this problem, we replace the term that depends on the  $\nu_{r_2r_3}^{r_1a_1}$  type integrals with

TABLE II. Errors (in kcal mol<sup>-1</sup>) of DF-SAPT2/aug-cc-pVDZ for selected complexes from the S22 test set (Ref. 6) relative to conventional SAPT2/aug-cc-pVDZ. [SAPT2 energies for 15 of the 22 complexes were computed with SAPT2008 (Ref. 61). DF errors for the components of DF-SAPT0 have been analyzed in Ref. 48.]

|   | MAX <sup>a</sup> | MSE <sup>b</sup> | MUE <sup>c</sup> | rms <sup>d</sup> |
|---|------------------|------------------|------------------|------------------|
| $E_{\text{elst,resp}}^{(12)}$                     | 0.009            | 0.002            | 0.002            | 0.003            |
| $E_{\text{exch}}^{(11)} + E_{\text{exch}}^{(12)}$ | -0.001           | 0.000            | 0.000            | 0.000            |
| $E_{\text{ind,resp}}^{(22)}$                      | 0.002            | 0.000            | 0.000            | 0.001            |

<sup>a</sup>Maximum absolute error.

<sup>b</sup>Mean signed error.

<sup>c</sup>Mean unsigned error.

<sup>d</sup>Root mean square error.

three-center DF integrals. Now, two steps are required; first the formation of an intermediate for a fixed  $b_1$  and  $s_1$

$$C_{a_1 r_1}^P = B_{r_3 r_1}^P a_{r_3 s_1}^{a_1 b_1}. \quad (66)$$

Then the contribution to  $V_{r_1 r_2}^{a_1 a_2}$  for a fixed  $b_1$  and  $s_1$  can be evaluated as

$$V_{r_1 r_2}^{a_1 a_2} = C_{a_1 r_1}^P B_{a_2 r_2}^P. \quad (67)$$

The evaluation of the contribution with exact integrals scales as  $\mathcal{O}(o^3 v^4)$ . With the DF approach, the formation of  $C_{a_1 r_1}^P$  scales as  $\mathcal{O}(o^2 v^3 n_{r_i})$  and the DF evaluation of  $V_{r_1 r_2}^{a_1 a_2}$  scales as  $\mathcal{O}(o^3 v^3 n_{r_i})$ . This makes the DF approach slightly more costly in terms of floating point operations than the conventional approach. However, the storage requirements of the DF intermediates are much lower, and the redundant disk I/O that is problematic in the conventional algorithm is removed.

In practice, the existing parallel implementation of SAPT uses the algorithm described above.<sup>55</sup> However, the serial, SAPT2008 program uses a more efficient algorithm.<sup>61</sup> The algorithm described above is blocked over  $b_1$  and  $s_1$ , this leads to  $\mathcal{O}(o^2 v^4)$  disk I/O. If the loops are blocked over  $a_1$ ,  $a_2$ , and  $b_1$ , then the disk I/O is reduced to  $\mathcal{O}(o^3 v^3)$ .<sup>56</sup> Even when the more efficient algorithm is used, the DF method for evaluating  $E_{\text{disp}}^{(22)}(T)$  is preferable.

### C. Accuracy of DF-SAPT

In our previous work on DF-SAPT0, the errors introduced through the DF approximation of the two-electron integrals was negligible.<sup>48</sup> In that work, we report errors of, at most, about 0.01 kcal mol<sup>-1</sup> for any of the zeroth-order components of the interaction energy. Additionally, previous works on DF-SAPT(DFT) report only negligible errors created by the DF integrals.<sup>27-29</sup> More generally, we are not aware of any case reported in the literature where DF approximations created problematic errors for *interaction energies*. Here, we will report the DF error associated with the second-order corrections to electrostatics, exchange, and induction for 15 complexes selected from the S22 test set. The accuracy of the DF is assessed against conventional SAPT energies computed with the SAPT2008 program.<sup>61</sup> The results of this analysis are shown in Table II. It is evident that DF errors are negligibly small for the second-order corrections to SAPT. The max error for the cases considered was less than

TABLE III. Accuracy of various truncations of SAPT for the S22 test set (Ref. 6) with aug-cc-pVDZ and aug-cc-pVDZ' basis sets relative to the recently revised (Ref. 49), estimated CCSD(T) CBS limit interaction energies. (Errors given in kcal mol<sup>-1</sup>.)

|  | MAD <sup>a</sup> | rms <sup>b</sup> | MAX <sup>c</sup> |
|--|------------------|------------------|------------------|
| aug-cc-pVDZ'                                       |                  |                  |                  |
| SAPT0  | 1.33             | 1.92             | 4.98             |
| SAPT0 + $\delta E_{\text{ind,resp}}^{(\text{HF})}$ | 0.47             | 0.63             | -1.73            |
| SAPT2  | 2.74             | 4.10             | 10.59            |
| SAPT2 + $\delta E_{\text{ind,resp}}^{(\text{HF})}$ | 1.32             | 1.77             | 4.04             |
| aug-cc-pVDZ  |                  |                  |                  |
| SAPT0  | 1.01             | 1.38             | 2.94             |
| SAPT0 + $\delta E_{\text{ind,resp}}^{(\text{HF})}$ | 1.45             | 1.88             | -3.82            |
| SAPT2  | 2.04             | 3.31             | 8.74             |
| SAPT2 + $\delta E_{\text{ind,resp}}^{(\text{HF})}$ | 0.94             | 1.23             | -2.62            |
| SAPT2+ <sup>d</sup>                                | 0.33             | 0.44             | -1.24            |
| SAPT2+(3) <sup>d</sup>                             | 0.39             | 0.45             | 0.82             |

<sup>a</sup>Mean average deviation.

<sup>b</sup>Root mean square deviation.

<sup>c</sup>Maximum absolute deviation ( $E_{\text{SAPT}} - E_{\text{ref}}$ ).

<sup>d</sup>The  $\delta E_{\text{ind,resp}}^{(\text{HF})}$  term is always added to SAPT2+ and SAPT2+(3).

0.01 kcal mol<sup>-1</sup> and the average errors are on the order of a few thousandths of a kcal mol<sup>-1</sup> at the most. Due to the excellent performance of these second-order terms and the expense associated with the conventional dispersion corrections, we will not present results for the second-order dispersion corrections.

We have assessed the accuracy of SAPT for the S22 test set<sup>6</sup> with recently revised estimated CCSD(T) CBS limit interaction energies.<sup>49</sup> It is common in the literature to pair SAPT0 or SAPT2 with a truncated basis set and rely on error cancellation rather than compute the expensive corrections to the dispersion energy. Here, we will use aug-cc-pVDZ'; these results are shown in Table III. For SAPT0 and SAPT2, it is beneficial to include the  $\delta E_{\text{ind,resp}}^{(\text{HF})}$  term when an aug-cc-pVDZ' basis is used. Perhaps surprisingly, SAPT0 performs better than SAPT2. This is a useful result, since SAPT0 can be extended to very large systems. Additionally, SAPT2 performs slightly better with the larger aug-cc-pVDZ basis than with the aug-cc-pVDZ' basis. The most difficult system for these methods with the aug-cc-pVDZ' basis is the formic acid dimer. This dimer is strongly bound and has large contributions from intramonomer electron correlation. This dimer represents the max error for all the SAPT/aug-cc-pVDZ' methods considered except for SAPT0 +  $\delta E_{\text{ind,resp}}^{(\text{HF})}$ . The formic acid dimer and the hydrogen bonded uracil dimer have the largest intramonomer correlation contributions to nondispersion terms. For all the methods tested, the largest errors tend to occur for hydrogen bonded systems (partially due to the larger magnitude of these interactions). For the description of hydrogen bonded systems, it is imperative to include the  $\delta E_{\text{ind,resp}}^{(\text{HF})}$  term. For hydrogen bonded systems and the notoriously difficult  $\pi$ - $\pi$  systems, the error cancellation observed for SAPT0/aug-cc-pVDZ' is remarkable.

For systems where the triples correction to dispersion can be afforded, highly accurate energies can be computed

TABLE IV. Comparison of various truncations of SAPT with supermolecular MP2 and CCSD(T) with a aug-cc-pVDZ basis set for the S22 (Refs. 6 and 49) test set. (Errors given in kcal mol<sup>-1</sup>.)

|  | MAD <sup>a</sup> | rms <sup>b</sup> | MAX <sup>c</sup> |
|--|------------------|------------------|------------------|
| MP2/aug-cc-pVDZ                                    |                  |                  |                  |
| SAPT0  | 0.41             | 0.54             | -1.41            |
| SAPT0 + $\delta E_{\text{ind,resp}}^{(\text{HF})}$ | 1.66             | 2.52             | -6.63            |
| SAPT2  | 1.40             | 2.25             | 5.93             |
| SAPT2 + $\delta E_{\text{ind,resp}}^{(\text{HF})}$ | 0.20             | 0.28             | -0.82            |
| SAPT2+   | 0.74             | 1.05             | -2.32            |
| SAPT2+(3)  | 0.78             | 1.13             | -2.71            |
| CCSD(T)/aug-cc-pVDZ                                |                  |                  |                  |
| SAPT0  | 0.98             | 1.41             | -3.89            |
| SAPT0 + $\delta E_{\text{ind,resp}}^{(\text{HF})}$ | 2.47             | 3.10             | -6.67            |
| SAPT2  | 1.59             | 2.34             | 5.89             |
| SAPT2 + $\delta E_{\text{ind,resp}}^{(\text{HF})}$ | 0.99             | 1.43             | -3.54            |
| SAPT2+ <sup>d</sup>                                | 1.01             | 1.27             | -2.67            |
| SAPT2+(3) <sup>d</sup>                             | 0.72             | 0.98             | -2.07            |

<sup>a</sup>Mean average deviation.

<sup>b</sup>Root mean square deviation.

<sup>c</sup>Maximum absolute deviation ( $E_{\text{SAPT}} - E_{\text{ref}}$ ).

<sup>d</sup>The  $\delta E_{\text{ind,resp}}^{(\text{HF})}$  term is always added to SAPT2+ and SAPT2+(3).

with SAPT. The lowest MAD and rms error is observed for SAPT2+, where all the second-order corrections to SAPT are included. The maximum error can be improved by also including the  $E_{\text{elst,resp}}^{(13)}$  and  $E_{\text{disp}}^{(30)}$  terms, which we have denoted SAPT2+(3). The SAPT2+ and SAPT2+(3) methods are essentially identical in cost since the  $\mathcal{O}(o^3v^4)$  triples correction to dispersion is required for both. These methods tend to be underbound for most complexes with an aug-cc-pVDZ basis, with the exception of  $\pi$ - $\pi$  systems where they tend to be slightly overbound. The difficulty in studying  $\pi$ - $\pi$  interactions is evident when the magnitude of the intramonomer corrections to dispersion is examined. For stacked adenine-thymine, the  $E_{\text{disp}}^{(21)}$  term is repulsive by 6.9 kcal mol<sup>-1</sup>. This is offset by a  $E_{\text{disp}}^{(32)}$  term that is attractive by 6.3 kcal mol<sup>-1</sup>; 4.6 kcal mol<sup>-1</sup> of this are due to the triples correction. This example makes it obvious that either all of the corrections to dispersion should be included or none should be included. These results show that for studying systems of similar size to nucleic acid base dimers, SAPT+(3) provides an attractive alternative to CCSD(T) computations.

It is useful to have a comparison of SAPT methods with more familiar supermolecular methods to have a way of thinking about what terms they include. In Table IV, we show the deviation of SAPT/aug-cc-pVDZ from MP2/aug-cc-pVDZ and CCSD(T)/aug-cc-pVDZ for the S22 test set.<sup>6,49</sup> As would be expected, SAPT2, when the  $\delta E_{\text{ind,resp}}^{(\text{HF})}$  term is included, performs very similarly to MP2. The largest deviation is for the formic acid dimer, which has an exceptionally large  $\delta E_{\text{ind,resp}}^{(\text{HF})}$  term. SAPT+(3) is the closest to CCSD(T); the larger difference between SAPT+(3) and CCSD(T) than between SAPT2 and MP2 was expected. CCSD includes higher order corrections than those found in SAPT+(3). The converged CCSD amplitudes can be used in place of the second-order amplitudes to compute the exchange corrections.<sup>52,53</sup> Although this would bring the SAPT

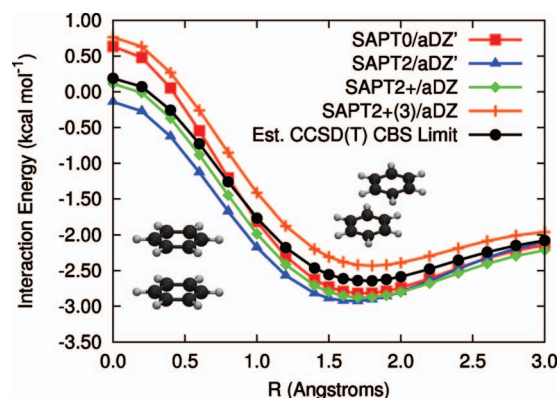


FIG. 4. Performance of various levels of SAPT for the parallel-displaced benzene dimer at a vertical separation of 3.4 Å.

results closer to CCSD(T), it would be at considerable expense, so we do not consider this possibility here. The difference between CCSD(T)/aug-cc-pVDZ and SAPT+(3)/aug-cc-pVDZ tends to put the SAPT+(3)/aug-cc-pVDZ results more in line with the estimated CCSD(T) CBS limit results.

We have also assessed the performance of several SAPT methods for the parallel-displaced benzene dimer; the results are shown in Fig. 4. Here, the  $\delta E_{\text{ind,resp}}^{(\text{HF})}$  term is included in all of these methods. The benchmark estimated CCSD(T) CBS limit potential energy curve is taken from previous work.<sup>62</sup> The results for this system are consistent with the results for the S22 test set. SAPT0/aug-cc-pVDZ' reproduces the CCSD(T) potential energy curve remarkably well; however, it underbinds the sandwich configuration ( $R=0.0$  Å) and overbinds the parallel-displaced minimum. In this respect, the SAPT2/aug-cc-pVDZ' energies are more consistent, as they always slightly overbind; however, on average, the errors for SAPT2/aug-cc-pVDZ' are larger than the SAPT0/aug-cc-pVDZ' errors. The best performance for this system comes from SAPT2+/aug-cc-pVDZ. The errors for this level of theory remain between -0.1 and -0.2 kcal mol<sup>-1</sup> for all geometries. Unlike any of the other levels of theory tested, SAPT2+/aug-cc-pVDZ maintains roughly the same amount of error for the sandwich configuration and the parallel-displaced minimum. The slightly more expensive SAPT2+(3) method consistently underbinds with an aug-cc-pVDZ basis. Since this method provides a more complete description of the interaction, it should be paired with a larger basis set than aug-cc-pVDZ.

Overall, our results suggest that for the purpose of computing accurate interaction energies with SAPT, SAPT2+/aug-cc-pVDZ or SAPT2+(3)/aug-cc-pVDZ will perform well as long as the expense of the triples correction can be afforded. If it cannot, SAPT0/aug-cc-pVDZ' is the best alternative, for both its accuracy and its scalability. Our previous work has shown that this level of theory can be applied to systems with over 100 atoms.<sup>48</sup> Surprisingly, SAPT2 is not an improvement on SAPT0 when paired with the basis sets considered. This is noteworthy since SAPT2 has been widely used in the literature with a basis set of similar quality to aug-cc-pVDZ'. However, if individual energy components are the target, rather than the total interac-



tion energy, SAPT2 is superior to SAPT0. The favorable error cancellation that leads to the apparent accuracy of SAPT0/aug-cc-pVDZ' appears for the total interaction energy, not for each term individually. The electrostatic and induction energies will be approximately equal for SAPT0 and SAPT2, however, the intramonomer corrections to exchange are usually significant and destabilizing.

#### IV. CONCLUSIONS

We have applied DF approximations to the treatment of intramonomer electron correlation in SAPT. Through the formation of a group of reusable intermediates based on the DF representation of the two-electron integrals, we are able to greatly improve the efficiency of our DF-SAPT program. In some cases, the DF intermediates are used to reduce the overall scaling of certain SAPT corrections. These intermediates can also be used to avoid the need to handle certain classes of integrals explicitly. As a result the dependence on disk I/O, which has plagued other SAPT implementations, is greatly reduced in our program. In order to get highly accurate interaction energies from SAPT, the triples contribution to dispersion must be included; this term scales as  $\mathcal{O}(o^3v^4)$ . By using DF intermediates, the evaluation of this term can be streamlined to avoid a bottleneck due to excessive disk I/O. Our implementation of DF-SAPT has allowed us to perform the largest SAPT computations to date that include all intramonomer correlation corrections through second-order.

The accuracy of SAPT at the level available with our DF-SAPT code was benchmarked against the S22 test set.<sup>6</sup> As anticipated, to compute highly accurate energies with SAPT, the dispersion energy must be corrected to account for intramonomer correlation. The expensive triples term must be included in order to improve the dispersion energy. The most complete SAPT methods tested include all the second-order terms and the leading third-order terms. This level of SAPT can be expected to predict interactions to within 0.3–0.4 kcal mol<sup>-1</sup>. Maximum errors can be expected to be less than 1 kcal mol<sup>-1</sup>. If the triples correction cannot be afforded, the best alternative is SAPT0/aug-cc-pVDZ', which can be extended to very large systems.

#### ACKNOWLEDGMENTS

We would like to thank Professor Krzysztof Szalewicz, Professor Rafal Podeszwa, and Professor Konrad Patkowski for helpful comments on the manuscript. This material is based upon work supported by the National Science Foundation (Grant No. CHE-0715268). This work is also funded in part by the U.S. Department of Energy, the division of Advanced Scientific Computing Research as a component of the Computational Chemistry Endstation project under subcontract from the University of Tennessee. The Center for Computational Molecular Science and Technology is funded through an NSF CRIF award (Grant No. CHE 04-43564) and by Georgia Tech.

<sup>1</sup>W. Saenger, *Principles of Nucleic Acid Structure* (Springer-Verlag, New York, 1984).

<sup>2</sup>H.-S. Shieh, H. M. Berman, M. Dabrow, and S. Neidle, *Nucleic Acids Res.* **8**, 85 (1980).

- <sup>3</sup>J. Šponer, J. Leszczynski, and P. Hobza, *J. Phys. Chem.* **100**, 5590 (1996).
- <sup>4</sup>M. Elstner, P. Hobza, T. Frauenheim, S. Suhai, and E. Kaxiras, *J. Chem. Phys.* **114**, 5149 (2001).
- <sup>5</sup>P. Jurečka and P. Hobza, *J. Am. Chem. Soc.* **125**, 15608 (2003).
- <sup>6</sup>P. Jurečka, J. Šponer, J. Černý, and P. Hobza, *Phys. Chem. Chem. Phys.* **8**, 1985 (2006).
- <sup>7</sup>E. A. Meyer, R. K. Castellano, and F. Diederich, *Angew. Chem., Int. Ed. Engl.* **42**, 1210 (2003).
- <sup>8</sup>S. K. Burley and G. A. Petsko, *Science* **229**, 23 (1985).
- <sup>9</sup>J. Braun, H. J. Neusser, and P. Hobza, *J. Phys. Chem. A* **107**, 3918 (2003).
- <sup>10</sup>A. L. Ringer, M. S. Figgis, M. O. Sinnokrot, and C. D. Sherrill, *J. Phys. Chem. A* **110**, 10822 (2006).
- <sup>11</sup>A. L. Ringer, A. Senenko, and C. D. Sherrill, *Protein Sci.* **16**, 2216 (2007).
- <sup>12</sup>D. L. Mobley, A. P. Graves, J. D. Chodera, A. C. McReynolds, B. K. Shoichet, and K. A. Dill, *J. Mol. Biol.* **371**, 1118 (2007).
- <sup>13</sup>J. C. Collings, K. P. Roscoe, E. G. Robins, A. S. Batsanov, L. M. Stimson, J. A. K. Howard, S. J. Clark, and T. B. Marder, *New J. Chem.* **26**, 1740 (2002).
- <sup>14</sup>A. L. Ringer and C. D. Sherrill, *Chem.-Eur. J.* **14**, 2542 (2008).
- <sup>15</sup>R. Podeszwa, B. M. Rice, and K. Szalewicz, *Phys. Rev. Lett.* **101**, 115503 (2008).
- <sup>16</sup>R. Podeszwa, B. M. Rice, and K. Szalewicz, *Phys. Chem. Chem. Phys.* **11**, 5512 (2009).
- <sup>17</sup>D. Řeha, M. Kabeláč, F. Ryjáček, J. Šponer, J. E. Šponer, M. Elstner, S. Suhai, and P. Hobza, *J. Am. Chem. Soc.* **124**, 3366 (2002).
- <sup>18</sup>K. M. Langner, P. Kedzierski, W. A. Sokalski, and J. Leszczynski, *J. Phys. Chem. B* **110**, 9720 (2006).
- <sup>19</sup>S. Li, V. R. Cooper, T. Thonhauser, B. I. Lundqvist, and D. C. Langreth, *J. Phys. Chem. B* **113**, 11166 (2009).
- <sup>20</sup>K. Raghavachari, G. W. Trucks, J. A. Pople, and M. Head-Gordon, *Chem. Phys. Lett.* **157**, 479 (1989).
- <sup>21</sup>K. Morokuma, *J. Chem. Phys.* **55**, 1236 (1971).
- <sup>22</sup>K. Kitaura and K. Morokuma, *Int. J. Quantum Chem.* **10**, 325 (1976).
- <sup>23</sup>P. S. Bagus, K. Hermann, and C. W. Bauschlicher, *J. Chem. Phys.* **80**, 4378 (1984).
- <sup>24</sup>W. Stevens and W. Fink, *Chem. Phys. Lett.* **139**, 15 (1987).
- <sup>25</sup>W. Chen and M. S. Gordon, *J. Phys. Chem.* **100**, 14316 (1996).
- <sup>26</sup>B. Jeziorski, R. Moszynski, and K. Szalewicz, *Chem. Rev. (Washington, D.C.)* **94**, 1887 (1994).
- <sup>27</sup>R. Bukowski, R. Podeszwa, and K. Szalewicz, *Chem. Phys. Lett.* **414**, 111 (2005).
- <sup>28</sup>A. Heßelmann, G. Jansen, and M. Schütz, *J. Chem. Phys.* **122**, 014103 (2005).
- <sup>29</sup>R. Podeszwa, R. Bukowski, and K. Szalewicz, *J. Chem. Theory Comput.* **2**, 400 (2006).
- <sup>30</sup>K. S. Kim, P. Tarakeshwar, and J. Y. Lee, *Chem. Rev. (Washington, D.C.)* **100**, 4145 (2000).
- <sup>31</sup>M. O. Sinnokrot and C. D. Sherrill, *J. Am. Chem. Soc.* **126**, 7690 (2004).
- <sup>32</sup>E. C. Lee, B. H. Hong, J. Y. Lee, J. C. Kim, D. Kim, Y. Kim, P. Tarakeshwar, and K. S. Kim, *J. Am. Chem. Soc.* **127**, 4530 (2005).
- <sup>33</sup>M. O. Sinnokrot and C. D. Sherrill, *J. Phys. Chem. A* **110**, 10656 (2006).
- <sup>34</sup>R. Podeszwa, R. Bukowski, and K. Szalewicz, *J. Phys. Chem. A* **110**, 10345 (2006).
- <sup>35</sup>E. C. Lee, D. Kim, P. Jurečka, P. Tarakeshwar, P. Hobza, and K. S. Kim, *J. Phys. Chem. A* **111**, 3446 (2007).
- <sup>36</sup>P. Tarakeshwar, H. S. Choi, S. J. Lee, K. S. Kim, T. Ha, J. H. Jang, J. G. Lee, and H. Lee, *J. Chem. Phys.* **111**, 5838 (1999).
- <sup>37</sup>P. Tarakeshwar, H. S. Choi, and K. S. Kim, *J. Am. Chem. Soc.* **123**, 3323 (2001).
- <sup>38</sup>T. P. Tauer, M. E. Derrick, and C. D. Sherrill, *J. Phys. Chem. A* **109**, 191 (2005).
- <sup>39</sup>D. Kim, S. Hu, P. Tarakeshwar, K. S. Kim, and J. M. Lisy, *J. Phys. Chem. A* **107**, 1228 (2003).
- <sup>40</sup>D. Kim, E. C. Lee, K. S. Kim, and P. Tarakeshwar, *J. Phys. Chem. A* **111**, 7980 (2007).
- <sup>41</sup>N. J. Singh, S. K. Min, D. Y. Kim, and K. S. Kim, *J. Chem. Theory Comput.* **5**, 515 (2009).
- <sup>42</sup>J. L. Whitten, *J. Chem. Phys.* **58**, 4496 (1973).
- <sup>43</sup>E. W. Stuebing, J. H. Weare, and R. G. Parr, *Int. J. Quantum Chem., Symp.* **11**, 81 (1977).



- <sup>44</sup>B. I. Dunlap, J. W. D. Connolly, and J. R. Sabin, *J. Chem. Phys.* **71**, 3396 (1979).
- <sup>45</sup>O. Vahtras, J. Almlöf, and W. M. Feyereisen, *Chem. Phys. Lett.* **213**, 514 (1993).
- <sup>46</sup>M. Feyereisen, G. Fitzgerald, and A. Komornicki, *Chem. Phys. Lett.* **208**, 359 (1993).
- <sup>47</sup>A. P. Rendell and T. J. Lee, *J. Chem. Phys.* **101**, 400 (1994).
- <sup>48</sup>E. G. Hohenstein and C. D. Sherrill, *J. Chem. Phys.* **132**, 184111 (2010).
- <sup>49</sup>T. Takatani, E. G. Hohenstein, M. Malagoli, M. S. Marshall, and C. D. Sherrill, *J. Chem. Phys.* **132**, 144104 (2010).
- <sup>50</sup>B. Jeziorski, R. Moszynski, A. Ratkiewicz, S. Rybak, K. Szalewicz, and H. L. Williams, in *Methods and Techniques in Computational Chemistry: METECC94*, edited by E. Clementi (STEF, Cagliari, 1993), Vol. B, p. 79.
- <sup>51</sup>K. Patkowski, K. Szalewicz, and B. Jeziorski, *J. Chem. Phys.* **125**, 154107 (2006).
- <sup>52</sup>R. Moszynski, B. Jeziorski, and K. Szalewicz, *J. Chem. Phys.* **100**, 1312 (1994).
- <sup>53</sup>R. Moszynski, B. Jeziorski, S. Rybak, K. Szalewicz, and H. L. Williams, *J. Chem. Phys.* **100**, 5080 (1994).
- <sup>54</sup>K. Szalewicz and B. Jeziorski, in *Molecular Interactions: From van der Waals to Strongly Bound Complexes*, edited by S. Scheiner (Wiley, New York, 1997), p. 3.
- <sup>55</sup>R. Bukowski, W. Cencek, K. Patkowski, P. Jankowski, M. Jeziorska, M. Kolaski, and K. Szalewicz, *Mol. Phys.* **104**, 2241 (2006).
- <sup>56</sup>K. Patkowski, personal communication (2010).
- <sup>57</sup>T. D. Crawford, C. D. Sherrill, E. F. Valeev, J. T. Fermann, R. A. King, M. L. Leininger, S. T. Brown, C. L. Janssen, E. T. Seidl, J. P. Kenny, and W. D. Allen, *J. Comput. Chem.* **28**, 1610 (2007).
- <sup>58</sup>T. H. Dunning, *J. Chem. Phys.* **90**, 1007 (1989).
- <sup>59</sup>R. A. Kendall, T. H. Dunning, and R. J. Harrison, *J. Chem. Phys.* **96**, 6796 (1992).
- <sup>60</sup>F. Weigend, A. Köhn, and C. Hättig, *J. Chem. Phys.* **116**, 3175 (2002).
- <sup>61</sup>R. Bukowski, W. Cencek, P. Jankowski, M. Jeziorska, B. Jeziorski, V. F. Lotrich, S. A. Kucharski, A. J. Misquitta, R. Moszyński, K. Patkowski, R. Podeszwa, S. Rybak, K. Szalewicz, H. L. Williams, R. J. Wheatley, P. E. S. Wormer, and P. S. Żuchowski, SAPT2008: An *ab initio* program for many-body symmetry-adapted perturbation theory calculations of intermolecular interaction energies, University of Delaware and University of Warsaw (<http://www.physics.udel.edu/~szalewic/SAPT/SAPT.html>).
- <sup>62</sup>C. D. Sherrill, T. Takatani, and E. G. Hohenstein, *J. Phys. Chem. A* **113**, 10146 (2009).